Transformation coefficients of permutation groups

Jin-Quan Chen
Physics Department, Nanjing University, Nanjing, People's Republic of China, and Department of Physics and Atmospheric Science, Drexel University, Philadelphia, Pennsylvania 19104

David F. Collinson
Department of Physics and Atmospheric Science, Drexel University, Philadelphia, Pennsylvania 19104

Mei-Juan Gao
Physics Department, Nanjing University, Nanjing, People's Republic of China

(Received 13 April 1982; accepted for publication 22 October 1982)

The eigenfunction method is used to calculate the transformation coefficients $\langle {}^{[\nu]}_{m}|[\nu],{}^{\tau[\nu_1][\nu_2]}_{m_1 m_2}\rangle$ from the Yamanouchi basis of the permutation group $S_{f_1+f_2}$ to the $S_{f_1+f_2}\supset S_{f_1}\otimes S_{f_2}$ irreducible basis. A program in FORTRAN is written, and tables of the transformation coefficients for the permutation group S_f up to $f=6$ are given. Possible applications of the transformation coefficients are sketched.

PACS numbers: 02.20.Qs, 03.65.Fd

I. INTRODUCTION

The transformation coefficients (or transformation matrix) of the permutation group (abbreviated as TC) were first introduced by Jahn[1] and Kaplan[2] in constructing orbital wavefunctions for a multishell configuration. These coefficients play an important role in nuclear structure calculations and molecular and atomic calculations, and have been studied extensively by Jahn,[1] Kaplan,[2] Horie,[3] Kramer,[4,5] and Sarma.[6]

In Refs. 7 and 8 it was shown that the calculation of many particle FPC (fractional parentage coefficients) for the group Chains $SU(mn)\supset SU(m)\times SU(n)$ and $SU(m+n)\supset SU(m)\otimes SU(n)$ needs the TC. Kramer[4,5] and Kramer and Seligman[9] showed that the Racah coefficients and 9ν coefficients of the unitary group, $SU(n)$, can be expressed in terms of the TC. Therefore, if one has a simple algorithm to calculate these transformation coefficients, one would be able to calculate the Racah and 9ν coefficient of $SU(n)$ with arbitrary n. Unfortunately, such an algorithm did not exist up to now. Kaplan[2] used the projection operator method to calculate the transformation coefficients. As pointed out by Kukulin, Smirnov, and Majling,[10] such a calculation is very cumbersome for a large number of particles. Sarma[6] suggested an algorithm for TC; however, it is still not convenient for computer calculation.

Based on a new approach to group representation theory,[11–13] a powerful and versatile technique, the eigenfunction method, has been developed. This method is easily implemented on computers. Three programs are available for evaluating the CG (Clebsch–Gordan) coefficients and ORC (outer-product reduction) coefficients[14] of the permutation group and the single-particle FPC for the group chain $SU(mn)\supset SU(m)\times SU(n)$.[15] In this paper we will describe the calculation of the TC by the eigenfunction method. An efficient program has been written and systematic tables of the TC have been obtained. The next paper will be devoted to the calculation of the Racah coefficients of $SU(n)$ for arbitrary n.

II. TRANSFORMATION COEFFICIENTS OF PERMUTATION GROUPS

Let

$$\left|[\nu],{}^{\tau[\nu_1][\nu_2]}_{m_1 m_2}\right\rangle,\quad \tau=1,2,\cdots,\{\nu_1\nu_2\nu\},\tag{1}$$

be the $S_f\supset S_{f_1}\otimes S_{f_2}$ irreducible basis which belongs to the irrep (irreducible representation $[\nu]$ of S_f with $f=f_1+f_2$, and at the same time the Yamanouchi basis $[\nu_1]m_1$ of $S_{f_1}=S_{f_1}(1,2,\ldots f_1)$ and $[\nu_2]m_2$ of $S_{f_2}=S_{f_2}(f_1+1,\ldots f)$, where m_1 and m_2 enumerate the Yamanouchi vectors of the irreps $[\nu_1]$ and $[\nu_2]$ of the groups S_{f_1} and S_{f_2}, respectively, τ is a multiplicity label, $\{\nu_1\nu_2\nu\}$ is the number of times that the representation $([\nu_1],[\nu_2])$ of S_{f_1} and S_{f_2} occurs in the irrep $[\nu]$ of S_f, and the m_i can be regarded either as Yamanouchi symbols or the indices for the basis vectors of the irrep $[\nu_i]$ in the so-called decreasing page order for the Yamanouchi symbols.[16] It was shown[13] that the $f-1$ two-cycle class operators $C(f), C(f-1),\ldots,C(2)$ of the permutation groups $S_f,S_{f-1},\ldots,S_2$ constitute the so-called second kind of complete set of commuting operators (CSCO-II) of S_f, and their simultaneous eigenvectors constitute the Yamanouchi basis of S_f. Consequently, $[\nu_i]m_i$ can also be viewed as the eigenvalues of the CSCO-II of S_{f_i}, namely,

$$\begin{aligned}(\nu_1,m_1)&=(\lambda_{f_1},\lambda_{f_1-1},\ldots,\lambda_2),\\(\nu_2 m_2)&=(\lambda_{f_2},\lambda_{f_2-1},\ldots,\lambda_2),\end{aligned}\tag{2}$$

where λ_i is the eigenvalue of $C(i)$, $\nu_1=\lambda_{f_1}$, $\nu_2=\lambda_{f_2}$, and m_1 and m_2 symbolize the rest of the respective sets of eigenvalues.

The $S_f\supset S_{f_1}\otimes S_{f_2}$ irreducible basis can be expanded in terms of the Yamanouchi basis $|{}^{[\nu]}_{m}\rangle$ of S_f:

$$\left|[\nu],{}^{\tau[\nu_1][\nu_2]}_{m_1 m_2}\right\rangle=\sum_m\left|{}^{[\nu]}_{m}\right\rangle\left\langle{}^{[\nu]}_{m}\right|[\nu],{}^{\tau[\nu_1][\nu_2]}_{m_1 m_2}\Big\rangle.\tag{3}$$

 0022-2488/83/122695-11$02.50

The TC

$$\left\langle \begin{matrix}[\nu] \\ m\end{matrix} \middle| [\nu], \begin{matrix}\tau[\nu_1][\nu_2] \\ m_1 m_2\end{matrix} \right\rangle$$

satisfy the unitarity conditions

$$\sum_{\nu_2 m_2 \tau} \left\langle \begin{matrix}[\nu] \\ m\end{matrix} \middle| [\nu], \begin{matrix}\tau[\nu_1][\nu_2] \\ m_1 m_2\end{matrix} \right\rangle \left\langle \begin{matrix}[\nu] \\ m'\end{matrix} \middle| [\nu], \begin{matrix}\tau[\nu_1][\nu_2] \\ m_1 m_2\end{matrix} \right\rangle = \delta_{mm'},$$

$$\sum_m \left\langle \begin{matrix}[\nu] \\ m\end{matrix} \middle| [\nu], \begin{matrix}\tau[\nu_1][\nu_2] \\ m_1 m_2\end{matrix} \right\rangle \left\langle \begin{matrix}[\nu] \\ m'\end{matrix} \middle| [\nu], \begin{matrix}\tau[\nu_1][\nu_2] \\ m_1 m_2\end{matrix} \right\rangle = \delta_{mm'}, \tag{4}$$

$$\sum_m \left\langle \begin{matrix}[\nu] \\ m\end{matrix} \middle| [\nu], \begin{matrix}\tau[\nu_1][\nu_2] \\ m_1 m_2\end{matrix} \right\rangle \left\langle \begin{matrix}[\nu] \\ m\end{matrix} \middle| [\nu], \begin{matrix}\tau'[\nu_1][\nu_2^1] \\ m_1 m_2^1\end{matrix} \right\rangle = \delta_{\nu_2 \nu_2'} \delta_{m_2 m_2'} \delta_{\tau\tau'}.$$

According to Theorem 4 in Ref. 13, the $S_f \supset S_{f_1} \otimes S_{f_2}$ irreducible basis must satisfy the following eigenequations:

$$\begin{pmatrix} \mathbf{C}(f) \\ C(f_1) \\ C(S_1) \\ C'(f_2) \\ C'(S_2) \end{pmatrix} \left| [\nu], \begin{matrix}\tau[\nu_1][\nu_2] \\ m_1 m_2\end{matrix} \right\rangle = \begin{pmatrix} \nu \\ \nu_1 \\ m_1 \\ \nu_2 \\ m_2 \end{pmatrix} \left| [\nu], \begin{matrix}\tau[\nu_1][\nu_2] \\ m_1 m_2\end{matrix} \right\rangle, \tag{5}$$

where $\mathbf{C}(f)$ is the CSCO-I of S_f, which in the most useful cases consists of the two-cycle and three-cycle class operators of S_f, while $(C(f_1), C(S_1))$ and $(C'(f_2), C'(S_2))$ are the CSCO-II of S_{f_1} and S_{f_2}, respectively:

$$\begin{aligned} &(C(f_1), C(S_1)) = (C(f_1), C(f_1 - 1), \ldots, C(2)), \\ &(C'(f_2), C'(S_2)) = (C'(f_2), C'(f_2 - 1), \ldots, C'(2)), \\ &C(l) = \sum_{i>j=1}^{l} (ij), \\ &C'(l) = \sum_{i>j=f_1+1}^{f_1+l} (ij), \end{aligned} \tag{6}$$

where (ij) is the transposition operator. Since the Yamanouchi basis of S_f is necessarily a Yamanouchi basis of S_{f_1}, the last $f_1 - 1$ eigenvalues in the eigenvalue set $(\nu,m) = (\lambda_f, \lambda_{f-1}, \ldots, \lambda_{f_1}, \ldots, \lambda_2)$ of the CSCO-II of S_f are just the eigenvalues (ν_1,m_1) of the CSCO-II of S_{f_1}, i.e.,

$$(\nu,m) = (\lambda_f \cdots \lambda_{f_1+1}, \nu_1 m_1). \tag{7}$$

Therefore, the TC

$$\left\langle \begin{matrix}[\nu] \\ m\end{matrix} \middle| [\nu], \begin{matrix}\tau[\nu_1][\nu_2] \\ m_1 m_2\end{matrix} \right\rangle$$

is zero unless (7) is satisfied. The summing index m in (3) is restricted to the quantum numbers $\lambda_{f-1} \cdots \lambda_{f_1+1}$. Furthermore, according to the Schur lemma, the TC is independent of m_1. Thus the label m_1 in the TC is redundant in the sense that the TC differing only in m_1 have identical values. However, it is better to retain this redundant label so that the summing index m in (3) can again be viewed as free.

As we said, each term on the rhs of (3) is already an irreducible basis of S_f and S_{f_1}; consequently, to find the TC, we merely need to diagonalize the CSCO-II of S_{f_2} in the Yamanouchi basis $|{}^{[\nu]}_{m}\rangle$ with fixed $[\nu_1]m_1$.

$$\sum_{m'} \left[\left\langle \begin{matrix}[\nu] \\ m\end{matrix} \middle| \begin{matrix}C'(f_2) \\ C'(S_2)\end{matrix} \middle| \begin{matrix}[\nu] \\ m'\end{matrix} \right\rangle - \binom{\nu_2}{m_2} \delta_{mm'} \right] \left\langle \begin{matrix}[\nu] \\ m'\end{matrix} \middle| [\nu], \begin{matrix}\tau[\nu_1][\nu_2] \\ m_1 m_2\end{matrix} \right\rangle = 0. \tag{8}$$

For fixed $[\nu]$ and $[\nu_1]m_1$, there are h_ν / h_{ν_1} different Yamanouchi basis vectors $|{}^{[\nu]}_{m}\rangle$, h_ν and h_{ν_1} being the dimensions of the irreps $[\nu]$ and $[\nu_1]$. Thus the index m' in (8) only takes h_ν / h_{ν_1} possible values. In practical calculation, m_1 can be chosen arbitrarily, for example, $m_1 = 1$ (the first component).

In Ref. 13 it was pointed out that, for computer calculation, it is more convenient to choose a single operator

$$M = \sum_{i=2}^{f_2} k_i C'(i) \tag{9}$$

as the CSCO-II of S_{f_2}, where the k_i are coefficients appropriately chosen. For example, we may choose $k_i = i + 7$. The eigenvalues λ of M and (ν_2,m_2) have a one-to-one correspondence for S_2–S_7:

$$\lambda = \sum_{i=2}^{f_2} (i+7)\lambda_i \leftrightarrow (\nu_2 m_2). \tag{10}$$

The set of eigenequations (8) can be replaced by a single eigenequation,

$$\sum_m \left[\left\langle \begin{matrix}[\nu] \\ m\end{matrix} \middle| M \middle| \begin{matrix}[\nu] \\ m'\end{matrix} \right\rangle - \lambda \delta_{mm'} \right] \left\langle \begin{matrix}[\nu] \\ m'\end{matrix} \middle| [\nu], \begin{matrix}\tau[\nu_1][\nu_2] \\ m_1 m_2\end{matrix} \right\rangle = 0. \tag{11}$$

The matrix elements of M can be calculated easily by using the Yamanouchi irreducible matrix elements.[16] In solving the eigenequation, if the eigenvalue λ [or, correspondingly, (ν_2,m_2)] has d-fold degeneracy, then $\{\nu_1 \nu_2 \nu\} = d$. For this eigenvalue λ we have d linearly independent eigenvectors which can be chosen orthogonal with respect to the multiplicity labels τ as shown in (4).

In order that the basis vectors

$$\left| [\nu], \begin{matrix}\tau[\nu_1][\nu_2] \\ m_1 m_2\end{matrix} \right\rangle$$

with different m_2 have the correct, i.e., the Yamanouchi, relative phases, we use the technique given in Ref. 13. Equation (59b) in Ref. 13 now takes the following form:

$$\left| [\nu], \begin{matrix}\tau[\nu_1][\nu_2] \\ m_1 m_2'\end{matrix} \right\rangle = \frac{1}{D^{[\nu_2]}_{m_2' m_2}(T_2)} \left[T_2' - D^{[\nu_2]}_{m_2 m_2}(T_2) \right] \left| [\nu], \begin{matrix}\tau[\nu_1][\nu_2] \\ m_1 m_2\end{matrix} \right\rangle, \tag{12}$$

where

$$T_2' = (i,i+1) \in S_{f_2}(f_1+1 \cdots f),$$
$$T_2 = (i-f_1, i-f_1+1) \in S_{f_2}(12 \cdots f_2). \tag{13}$$

From (13) we have

$$\left\langle \begin{matrix}[\nu] \\ m'\end{matrix} \middle| [\nu], \begin{matrix}\tau[\nu_1][\nu_2] \\ m_1 m_2'\end{matrix} \right\rangle = \frac{1}{D^{[\nu_2]}_{m_2' m_2}(T_2)} \sum_m [D^{[\nu]}_{m' m}(T_2') - D^{[\nu_2]}_{m_2 m_2}(T_2)\delta_{mm'}] \times \left\langle \begin{matrix}[\nu] \\ m\end{matrix} \middle| [\nu], \begin{matrix}\tau[\nu_1][\nu_2] \\ m_1 m_2\end{matrix} \right\rangle. \tag{14}$$

Therefore, for any irrep $[\nu_2]$ one only needs to find $\{\nu_1\nu_2\nu\}$ orthogonal eigenvectors of (8) or (11) corresponding to a particular m_2. By choosing appropriate adjacent permutation $(i,i+1)$,[13] from (14) we can get all the other components m_2' successively. We remove all sign arbitrariness by requiring the absolute phase convention that the first nonvanishing coefficient of a nonstandard basis vector (3) (where the Yamanouchi basis vectors are enumerated in decreasing page order[16]) with $m_2 = 1$ be positive.

III. APPLICATION OF THE TRANSFORMATION COEFFICIENT

A. Calculation of Racah coefficients and 9ν coefficients of SU(n)

Kramer[5] showed that the Racah coefficients of the group SU(n) can be expressed in terms of the TC

$$U(\nu_1\nu_2\nu\nu_3;\nu_{12}\nu_{23})^{\tau_{12}\tau}_{\tau_{23}\tau'} = \sum_{m_{12}m_{23}m}{}' \left\langle \begin{matrix}[\nu] \\ m\end{matrix} \middle| [\nu], \begin{matrix}\tau[\nu_{12}][\nu_3] \\ m_{12}m_3\end{matrix} \right\rangle \left\langle \begin{matrix}[\nu_{12}] \\ m_{12}\end{matrix} \middle| [\nu_{12}], \begin{matrix}\tau_{12}[\nu_1][\nu_2] \\ m_1 m_2\end{matrix} \right\rangle \left\langle \begin{matrix}[\nu] \\ m\end{matrix} \middle| [\nu], \begin{matrix}\tau[\nu_1][\nu_{23}] \\ m_1 m_{23}\end{matrix} \right\rangle \left\langle \begin{matrix}[\nu_{23}] \\ m_{23}\end{matrix} \middle| [\nu_{23}], \begin{matrix}\tau_{23}[\nu_2][\nu_3] \\ m_2 m_3\end{matrix} \right\rangle. \tag{15}$$

The sum is carried out with fixed m_1, m_2, and m_3. In practical calculation we may set $m_1 = m_2 = m_3 = 1$. Similarly, the 9ν coefficient of SU_n can be expressed as

$$\begin{pmatrix} \nu_1 & \nu_2 & \nu_{12} \\ \nu_3 & \nu_4 & \nu_{34} \\ \nu_{13} & \nu_{24} & \nu \end{pmatrix}^{\tau_{12}\tau_{34}\tau}_{\tau_{13}\tau_{24}\tau'} = \sum_{\substack{m_{12}m_{34}m \\ m_{13}m_{24}m'}} D^{[\nu]}_{mm'}(P) \left\langle \begin{matrix}[\nu] \\ m\end{matrix} \middle| [\nu], \begin{matrix}\tau[\nu_{12}][\nu_{34}] \\ m_{12}m_{34}\end{matrix} \right\rangle \times \left\langle \begin{matrix}[\nu_{12}] \\ m_{12}\end{matrix} \middle| [\nu_{12}], \begin{matrix}\tau_{12}[\nu_1][\nu_2] \\ m_1 m_2\end{matrix} \right\rangle \left\langle \begin{matrix}[\nu_{34}] \\ m_{34}\end{matrix} \middle| [\nu_{34}], \begin{matrix}\tau_{34}[\nu_3][\nu_4] \\ m_3 m_4\end{matrix} \right\rangle \left\langle \begin{matrix}[\nu] \\ m'\end{matrix} \middle| [\nu], \begin{matrix}\tau'[\nu_{13}][\nu_{24}] \\ m_{13}m_{24}\end{matrix} \right\rangle \times \left\langle \begin{matrix}[\nu_{13}] \\ m_{13}\end{matrix} \middle| [\nu_{13}], \begin{matrix}\tau_{13}[\nu_1][\nu_3] \\ m_1 m_3\end{matrix} \right\rangle \left\langle \begin{matrix}[\nu_{24}] \\ m_{24}\end{matrix} \middle| [\nu_{24}], \begin{matrix}\tau_{24}[\nu_2][\nu_4] \\ m_2 m_4\end{matrix} \right\rangle, \tag{16}$$

where the sum is carried out under fixed m_1, m_2, m_3, and m_4, and the permutation P is

$$P = \begin{pmatrix} f_1+1, f_1+2, \ldots, f_{12}, f_{12}+1, \ldots, f_{123} \\ f_{12}+1, f_{12}+2, \ldots, f_{123}, f_1+1, \ldots, f_{12} \end{pmatrix}, \tag{17}$$

and f_i is the number of boxes in the Young diagram $[\nu_i]$, $f_{12} = f_1 + f_2$, and $f_{123} = f_{12} + f_3$. Since the TC of permutation groups are independent of n, the Racah coefficients and 9ν coefficients of SU(n) do not depend on n explicitly. From the TC we can calculate these coefficients for arbitrary SU(n) once and for all.

B. Calculation of the outer-product reduction coefficients

Suppose there are two subsystems with particles $(\omega_1^0) = (1,2,\ldots,f_1)$ and $(\omega_2^0) = (f_1+1,\ldots,f_1+f_2=f)$. Suppose $\psi^{\nu_1}_{m_1}(\omega_1^0)$ and $\psi^{\nu_2}_{m_2}(\omega_2^0)$ are two wavefunctions of these two subsystems in some coordinate space, say x, which are the Yamanouchi bases of the permutation groups S_{f_1} and S_{f_2}, respectively. Using the so-called outer-product reduction coefficients (ORC)[12] $C^{[\nu]\tau,m}_{\nu_1 m_1,\omega_1,\nu_2 m_2 \omega} \equiv C^{[\nu]\tau,m}_{\nu_1,m_1,\nu_2 m_2,\omega}$ of the permutation group S_f, the wavefunction $\Psi^{[\nu]\tau}_m$ of the total system with a definite symmetry $[\nu]m$ with respect to S_f and be constructed as

$$\Psi^{[\nu]\tau}_m(\omega^0) = \sum_{m_1 m_2 \omega} C^{[\nu]\tau,m}_{\nu_1 m_1,\nu_2 m_2,\omega} \Psi^{[\nu_1]}_{m_1}(\omega_1)\Psi^{[\nu_2]}_{m_2}(\omega_2), \tag{18a}$$

where $\tau = 1,2,\ldots,\{\nu_1\nu_2\nu\}$ is the multiplicity label, with the multiplicity $\{\nu_1\nu_2\nu\}$ determined by the Littlewood rule, $(\omega^0) = (1,2,\ldots,f)$, and $(\omega) = (\omega_1,\omega_2)$, ω_i being the normal order sequences

$$(\omega_1) = (a_1 a_2 \cdots a_{f_1}), \quad (\omega_2) = (a_{f_1+1} \cdots a_f),$$
$$a_1 \langle a_2 \langle \cdots \langle a_{f_1}, \quad a_{f_1+1} \langle a_{f_1+2} \langle \cdots \langle a_f, \tag{18b}$$

a_i representing any one of the numbers $1,2,\ldots,f$. The ORC are also the coefficients for coupling the irreducible basis of SU(m) and SU(n) into the SU($m+n$)$\supset$SU(m)$\otimes$SU(n) irreducible basis.[12] In Ref. 12 a method is given for calculating the ORC and systematic tables of ORC for S_2–S_6 are given. Using the TC, we now can find a much simpler way to calculate the ORC.

Using the left coset decomposition with respect to the subgroup $S_{f_1} \otimes S_{f_2}$, the permutation operator R of S_f can be

written as

$$R = \binom{\omega_0}{\omega} P_1 P_2 = Q_\omega P_1 P_2, \tag{19}$$

where $P_1 \in S_{f_1}$, $P_2 \in S_{f_2}$, and $Q_\omega = \binom{\omega_0}{\omega}$ is a permutation operator which brings the natural order sequence $(\omega^0) = (1,2,\ldots,f)$ into the normal order sequence $(\omega) = (\omega_1,\omega_2)$.

Introducing the projection operators

$$\mathring{P}^{[\nu_i]m_i'}_{m_i} = \left(\frac{h_{\nu_i}}{f_i!}\right)^{1/2} \sum_{a=1}^{f_i!} \left\langle \begin{matrix}[\nu_i] \\ m_i\end{matrix} \middle| R_a \middle| \begin{matrix}[\nu_i] \\ m_i'\end{matrix} \right\rangle R_a, \tag{20}$$

$$\mathring{P}^{[\nu],\tau[\nu_1]m_1'[\nu_2]m_2'}_{m} = \left(\frac{h_\nu}{f!}\right)^{1/2} \sum_{Q_\omega \in P_1 P_2} \left\langle \begin{matrix}[\nu] \\ m\end{matrix} \middle| Q_\omega P_1 P_2 \middle| [\nu], \begin{matrix}\tau[\nu_1][\nu_2] \\ m_1' m_2'\end{matrix} \right\rangle Q_\omega P_1 P_2, \tag{21}$$

it is easy to show that

$$\mathring{P}^{[\nu],\tau[\nu_1]m_1'[\nu_2]m_2'}_{m} = \left(\frac{h_\nu}{h_{\nu_1} h_{\nu_2}}\right)^{1/2} \left(\frac{f_1! f_2!}{f!}\right)^{1/2} \sum_{m_1 m_2 Q_\omega} \left\langle \begin{matrix}[\nu] \\ m\end{matrix} \middle| Q_\omega \middle| [\nu], \begin{matrix}\tau[\nu_1][\nu_2] \\ m_1 m_2\end{matrix} \right\rangle Q_\omega \mathring{P}^{[\nu_1]m_1'}_{m_1} \mathring{P}^{[\nu_2]m_2'}_{m_2}. \tag{22}$$

Applying (21) to an f-particle product state $|\Phi_0\rangle = |i_1 i_2 \cdots i_f\rangle$ with f different single particle states $i_1 i_2 \cdots i_f$ yields

$$\begin{aligned}\Psi^{[\nu]\tau'}_{m} &= \mathring{P}^{[\nu],\tau[\nu_1]m_1'[\nu_2]m_2'}_{m} |\Phi_0\rangle \\ &= \left(\frac{h_\nu}{h_{\nu_1} h_{\nu_2}}\right)^{1/2} \left(\frac{f_1! f_2!}{f!}\right)^{1/2} \sum_{m_1 m_2 \omega} \left\langle \begin{matrix}[\nu] \\ m\end{matrix} \middle| Q_\omega \middle| [\nu], \begin{matrix}\tau[\nu_1][\nu_2] \\ m_1 m_2\end{matrix} \right\rangle \psi^{[\nu_1]}_{m_1}(\omega_1) \psi^{[\nu_2]}_{m_2}(\omega_2).\end{aligned} \tag{23}$$

Similar expressions were given by Kaplan[2] and Kramer.[4] The label τ' in the lhs of (23) is used to distinguish the different linearly independent functions resulting from using different superscripts in the projection operator (21). We can choose $\tau' = \tau$, since τ' and τ have the same range and the choice of the additional label τ' is arbitrary. Comparing (23) with (18a), we get a new expression for the ORC,

$$C^{[\nu]\tau,m}_{\nu_1 m_1, \nu_2 m_2, \omega} = \left(\frac{h_\nu}{h_{\nu_1} h_{\nu_2}}\right)^{1/2} \left(\frac{f_1! f_2!}{f!}\right)^{1/2} \left\langle \begin{matrix}[\nu] \\ m\end{matrix} \middle| Q_\omega \middle| [\nu], \begin{matrix}\tau[\nu_1][\nu_2] \\ m_1 m_2\end{matrix} \right\rangle \tag{24a}$$

$$= \left(\frac{h_\nu}{h_{\nu_1} h_{\nu_2}}\right)^{1/2} \left(\frac{f_1! f_2!}{f!}\right)^{1/2} \sum_{m'} D^{[\nu]}_{mm'}(Q_\omega) \left\langle \begin{matrix}[\nu] \\ m'\end{matrix} \middle| [\nu], \begin{matrix}\tau[\nu_1][\nu_2] \\ m_1 m_2\end{matrix} \right\rangle. \tag{24b}$$

With the known TC and the Yamanouchi matrix element $D^{[\nu]}_{mm'}$, it is easy to calculate the ORC from (24b).

Setting $Q_\omega = e$ (identity, we have

$$\left\langle \begin{matrix}[\nu] \\ m\end{matrix} \middle| [\nu], \begin{matrix}\tau[\nu_1][\nu_2] \\ m_1 m_2\end{matrix} \right\rangle = \left(\frac{h_{\nu_1} h_{\nu_2}}{h_\nu}\right)^{1/2} \left(\frac{f!}{f_1! f_2!}\right)^{1/2} C^{[\nu]\tau,m}_{\nu_1 m_1, \nu_2 m_2, \omega^0}. \tag{25}$$

C. Calculation of the transformation coefficients from the SU($m+n$) Gel'fand basis to the SU($m+n$)⊃SU(m)⊗SU(n) irreducible basis

The transformation coefficients between the Gel'fand basis $|\begin{smallmatrix}[\nu] \\ W\end{smallmatrix}\rangle$ of SU($m+n$) and the SU($m+n$)⊃SU(m)⊗SU(n) irreducible basis

$$\left| [\nu], \begin{matrix}\tau[\nu_1][\nu_2] \\ W_1 W_2\end{matrix} \right\rangle$$

are defined by the following expansion:

$$\left| [\nu], \begin{matrix}\tau[\nu_1][\nu_2] \\ W_1 W_2\end{matrix} \right\rangle = \sum_W \left| \begin{matrix}[\nu] \\ W\end{matrix} \right\rangle \left\langle \begin{matrix}[\nu] \\ W\end{matrix} \middle| [\nu], \begin{matrix}\tau[\nu_1][\nu_2] \\ W_1 W_2\end{matrix} \right\rangle, \tag{26}$$

where W_1, W_2, and W are the Weyl tableaux corresponding to the groups SU(m), SU(n), and SU($m+n$), respectively, and τ is the multiplicity label. In Ref. 11(c) it was shown that the coefficients

$$\left\langle \begin{matrix}[\nu] \\ W\end{matrix} \middle| [\nu], \begin{matrix}\tau[\nu_1][\nu_2] \\ W_1 W_2\end{matrix} \right\rangle$$

are related to the TC of the permutation group:

$$\left\langle \begin{matrix}[\nu] \\ W\end{matrix} \middle| [\nu], \begin{matrix}\tau[\nu_1][\nu_2] \\ W_1 W_2\end{matrix} \right\rangle = {\sum_m}' \frac{R^{[\nu]m}(\overline{\omega})}{R^{[\nu_1]m_1}(\overline{\omega}_1) R^{[\nu_2]m_2}(\overline{\omega}_2)} \left\langle \begin{matrix}[\nu] \\ m\end{matrix} \middle| [\nu], \begin{matrix}\tau[\nu_1][\nu_2] \\ m_1 m_2\end{matrix} \right\rangle, \tag{27}$$

where R are normalization factors,

$$R^{[\nu]m}(\overline{\omega}) = \left(\left\langle \overline{\omega} \middle| \sum_{p \in S_f} D^{[\nu]}_{mm}(p) p \middle| \overline{\omega} \right\rangle \right)^{1/2},$$

$$R^{[\nu_i]m_i}(\overline{\omega}_i) = \left(\left\langle \overline{\omega}_i \middle| \sum_{p \in S_{f_i}} D^{[\nu_i]}_{m_i m_i}(p) p \middle| \overline{\omega}_i \right\rangle \right)^{1/2}, \quad i = 1,2 \tag{28}$$

and $|\overline{\omega}\rangle$, $|\overline{\omega}_1\rangle$, and $|\overline{\omega}_2\rangle$ are the f-, f_1-, and f_2-particle product states, respectively, for example $|\overline{\omega}\rangle = |i_1 i_2 \cdots i_f\rangle$. The D are the Yamanouchi matrix elements of the permutation group. The prime in the summation symbol means that the sum over m is under the restriction that when the number $1,2,\ldots,f$ in the Young tableaux $Y^{[\nu]}_m$ are replaced by the single particle state labels $i_1, i_2, \ldots, i_f$, the Young tableaux $Y^{[\nu]}_m$ must go over to the Weyl tableaux W.

Using (27), it is easy to calculate the transformation coefficients of unitary groups, which also do not depend on m and n explicitly.

We know that the unitary group approach pioneered by Moshinsky[17] for the nuclear many-body problem turns out to be very powerful for the many-electron problem, as developed mainly by Paldus[18] and Matsen *et al.*[19] in the so-called spin-free quantum chemistry. A considerable progress has been made in both the development and actual implementation of the unitary group approach to the large scale *ab initio* and semiempirical configuration interaction calculation of correlated electronic wavefunctions of molecules.[20] However, in the Matsen or Paldus approach, the point-group symmetry of the molecule has not been incorporated. In Ref. 21 many-electron states with definite spin and point-group symmetry are constructed. The transformation coefficients from this symmetry-adapted basis to the Gel'fand basis can

TABLE I.1. The phase factors Λ^{ν}_{m}.[a]

S_3	S_4			S_5					S_6
[21]	[31]	[22]	[211]	[41]	[32]	[311]	[221]	[21^3]	[21^3]
[51]									
1 2	1 2 3	1 2	1 2 3	1 2 3 4	1 2 3 4 5	1 2 3 4 5 6	1 2 3 4 5	1 2 3 4	1 2 3 4 5
+ −	+ − +	+ −	+ − +	+ − + −	+ − + + −	+ − + + − −	+ − − + −	+ − + −	+ − + − +

S_6																
[42]								[411]								
1 2 3 4 5 6 7 8 9									1 2 3 4 5 6 7 8 9 10							
+ − + − + − + + −									+ − + − + − + + − +							

S_6																					
[31^3]										[2^3]					[2^21^2]						[21^4]
1 2 3 4 5 6 7 8 9 10										1 2 3 4 5					1 2 3 4 5 6 7 8 9					1 2 3 4 5	
+ − + + − + − + − +										+ − − + −					+ − − + − + − + −					+ − + − +	

[a] We use decreasing page order for the Yamanouchi symbols $(r_f r_{f-1} \cdots r_2 r_1)$.

TABLE I.2. The ordering of the irreps.

	S_2	S_3	S_4	S_5
	1 2	1 2 3	1 2 3 4 5	1 2 3 4 5 6 7
[ν]	[2][11]	[3][21][1^3]	[4][31][22][211][1^4]	[5][41][32][311][221][21^3][1^5]

	S_6										
	1 2 3 4 5 6 7 8 9 10 11										
[ν]	[6][51][42][411][33][321][31^3][2^3][2^21^2][21^4][1^6]										

TABLE I.3. The phase $\epsilon(\nu_1\nu_2\nu)$.

$[\nu_1][\nu_2][\nu]$	ϵ	$[\nu_1][\nu_2][\nu]$	ϵ	$[\nu_1][\nu_2][\nu]$	ϵ	$[\nu_1][\nu_2][\nu]$	ϵ
[1][3][31]	+	[1][5][51]	+	[2][22][42]	+	[11][31][411]	−
[1][21]	+	[1][41]	+	[3][3]	+	[3][21]	−
[1][4][41]	−	[2][4]	−	[3][21]	−	[3][1^3]	+
[1][31]	+	[2][31]	+	[21][3]	+	[21][3]	+
[2][3]	+	[3][3]	+	[21][21]	+	[21][21]	+
[2][21]	+	[3][21]	+	[1][41][411]	−	[1][32][33]	+
[1][31][32]	+	[1][41][42]	−	[1][311]	+	[2][31]	−
[1][22]	+	[1][32]	+	[2][31]	+	[21][21]	+
[2][3]	+	[2][4]	+	[2][221]	+		
[2][21]	−	[2][31]	+	[11][4]	+		

be easily evaluated from the Clebsch–Gordan coefficients of the point group and the transformation coefficients of unitary groups. Therefore, the transformation coefficients from the $SU(m+n) \supset SU(m) \otimes SU(n)$ basis to the $SU(m+n)$ Gel'fand basis may have important application in quantum chemistry.

Other applications of the TC of the permutation group can be found in Refs. 2 and 4–9.

IV. THE TABLES OF THE TC

The TC for $f_2 = 1$ are trivial:

$$\left\langle \begin{matrix}[\nu] \\ m\end{matrix} \middle| [\nu] \begin{matrix}[\nu_1][1] \\ m_1\end{matrix} \right\rangle = 1 \quad \text{if } m[\nu_1]m_1$$
$$= 0 \quad \text{otherwise.}$$

The TC for $f_2 = 2$ can be calculated by a simple formula[1,2] and therefore are left out of the tables. Due to the symmetry of the TC under conjugation[4,22]

$$\left\langle \begin{matrix}[\tilde{\nu}] \\ \tilde{m}\end{matrix} \middle| [\tilde{\nu}], \begin{matrix}\tau[\tilde{\nu}_1][\tilde{\nu}_2] \\ \tilde{m}_1\tilde{m}_2\end{matrix} \right\rangle$$
$$= \epsilon(\nu_1\nu_2\nu_\tau) \Lambda^{\ \nu}_{m} \Lambda^{\ \nu_1}_{m_1} \Lambda^{\ \nu_2}_{m_2} \left\langle \begin{matrix}[\nu] \\ m\end{matrix} \middle| [\nu], \begin{matrix}\tau[\nu_1][\nu_2] \\ m_1 m_2\end{matrix} \right\rangle, \quad (29a)$$

when $[\nu_1]$,$[\nu_2]$, and $[\nu]$ are not simultaneously self-conjugate, $[\tilde{\nu}]\tilde{m}$ is the conjugate tableau of $[\nu]m$, Λ 's are phases[16] which we tabulate in Table I.1, and $\epsilon(\nu_1\nu_2\nu_\tau = \epsilon$ is a phase. According to the absolute phase convention, we have

$$\epsilon = \operatorname{sgn}\left(\Lambda^{\ \nu}_{m} \Lambda^{\ \nu_1}_{m_1} \Lambda^{\ \nu_2}_{h_{\nu_2}} \left\langle \begin{matrix}[\nu] \\ m\end{matrix} \middle| [\nu], \begin{matrix}\tau[\nu_1][\nu_2] \\ m_1 m_2\end{matrix} \right\rangle \right)\Big|_{m=\max}. \quad (29b)$$

TABLE II.1. Group S_4, irrep $[31], f_1 = 1, f_2 = 3$.

Standard basis vectors						1	2	3
ν_1	m_1	ν_2	m_2	τ	Com. den.			
1	1	1	1	1	9	1	2	6
1	1	2	1	1	36	32	−1	−3
1	1	2	2	1	4	0	3	−1

TABLE II.2. Group S_4, irrep $[22], f_1 = 1, f_2 = 3$.

Standard basis vectors						1	2
ν_1	m_1	ν_2	m_2	τ	Com. den.		
1	1	2	1	1	4	1	3
1	1	2	2	1	4	3	−1

TABLE II.3. Group S_5, irrep $[41], f_1 = 1, f_2 = 4$.

Standard basis vectors						1	2	3	4
ν_1	m_1	ν_2	m_2	τ	Com. den.				
1	1	1	1	1	48	3	5	10	30
1	1	2	1	1	144	135	−1	−2	−6
1	1	2	2	1	36	0	32	−1	−3
1	1	2	3	1	4	0	0	3	−1

TABLE II.4. Group S_5, irrep $[41], f_1 = 2, f_2 = 3$.

Standard basis vectors						1	2	3	4
ν_1	m_1	ν_2	m_2	τ	Com. den.				
1	1	1	1	1	18	3	5	10	0
1	1	2	1	1	18	15	−1	−2	0
1	1	2	2	1	3	0	2	−1	0
2	1	1	1	1	1	0	0	0	1

TABLE II.5. Group S_5, irrep $[32], f_1 = 1, f_2 = 4$.

Standard basis vectors						1	2	3	4	5
ν_1	m_1	ν_2	m_2	τ	Com. den.					
1	1	2	1	1	9	1	2	6	0	0
1	1	2	2	1	144	32	−1	−3	27	81
1	1	2	3	1	16	0	3	−1	9	−3
1	1	3	1	1	48	32	−1	−3	−3	−9
1	1	3	2	1	16	0	9	−3	−3	1

TABLE II.6. Group S_5, irrep $[32], f_1 = 2, f_2 = 3$.

Standard basis vectors						1	2	3	4	5
ν_1	m_1	ν_2	m_2	τ	Com. den.					
1	1	1	1	1	9	1	2	0	6	0
1	1	2	1	1	9	2	4	0	−3	0
1	1	2	2	1	3	2	−1	0	0	0
2	1	2	1	1	1	0	0	1	0	0
2	1	2	2	1	1	0	0	0	0	1

TABLE II.7. Group S_5, irrep $[311], f_1 = 1, f_2 = 4$.

Standard basis vectors						1	2	3	4	5	6
ν_1	m_1	ν_2	m_2	τ	Com. den.						
1	1	2	1	1	9	1	2	6	0	0	0
1	1	2	2	1	576	−32	1	3	135	405	0
1	1	2	3	1	192	0	−9	3	−15	5	160
1	1	4	1	1	192	160	−5	−15	3	9	0
1	1	4	2	1	576	0	405	−135	−3	1	32
1	1	4	3	1	9	0	0	0	6	−2	1

TABLE II.8. Group S_5, irrep [311], $f_1 = 2, f_2 = 3$.

Standard basis vectors						1	2	3	4	5	6
ν_1	m_1	ν_2	m_2	τ	Com. den.						
1	1	2	1	1	3	1	2	0	0	0	0
1	1	2	2	1	18	−2	1	0	15	0	0
1	1	3	1	1	18	10	−5	0	3	0	0
2	1	1	1	1	18	0	0	3	0	5	10
2	1	2	1	1	18	0	0	15	0	−1	−2
2	1	2	2	1	3	0	0	0	0	2	−1

TABLE II.9. Group S_6, irrep [51], $f_1 = 1, f_2 = 5$.

Standard basis vectors						1	2	3	4	5
ν_1	m_1	ν_2	m_2	τ	Com. den.					
1	1	1	1	1	50	2	3	5	10	30
1	1	2	1	1	1200	1152	−3	−5	−10	−30
1	1	2	2	1	144	0	135	−1	−2	−6
1	1	2	3	1	36	0	0	32	−1	−3
1	1	2	4	1	4	0	0	0	3	−1

TABLE II.10. Group S_6, irrep [51], $f_1 = 2, f_2 = 4$.

Standard basis vectors						1	2	3	4	5
ν_1	m_1	ν_2	m_2	τ	Com. den.					
1	1	1	1	1	20	2	3	5	10	0
1	1	2	1	1	180	162	−3	−5	−10	0
1	1	2	2	1	18	0	15	−1	−2	0
1	1	2	3	1	3	0	0	2	−1	0
2	1	1	1	1	1	0	0	0	0	1

TABLE II.11. Group S_6, irrep [51], $f_1 = 3, f_2 = 3$.

Standard basis vectors						1	2	3	4
ν_1	m_1	ν_2	m_2	τ	Com. den.				
1	1	1	1	1	10	2	3	5	0
1	1	2	1	1	40	32	−3	−5	0
1	1	2	2	1	8	0	5	−3	0
2	1	1	1	1	1	0	0	0	1

TABLE II.12. Group S_6, irrep [42], $f_1 = 1, f_2 = 5$.

Standard basis vectors						1	2	3	4	5	6	7	8	9
ν_1	m_1	ν_2	m_2	τ	Com. den.									
1	1	2	1	1	48	3	5	10	30	0	0	0	0	0
1	1	2	2	1	1296	135	−1	−2	−6	128	256	768	0	0
1	1	2	3	1	324	0	32	−1	−3	64	−2	−6	54	162
1	1	2	4	1	36	0	0	3	−1	0	6	−2	18	−6
1	1	3	1	1	162	135	−1	−2	−6	−2	−4	−12	0	0
1	1	3	2	1	1296	0	1024	−32	−96	−32	1	3	−27	−81
1	1	3	3	1	144	0	0	96	−32	0	−3	1	−9	3
1	1	3	4	1	48	0	0	0	0	32	−1	−3	−3	−9
1	1	3	5	1	16	0	0	0	0	0	9	−3	−3	1

TABLE II.13. Group S_6, irrep [42], $f_1 = 2, f_2 = 4$.

Standard basis vectors						1	2	3	4	5	6	7	8	9
ν_1	m_1	ν_2	m_2	τ	Com. den.									
1	1	1	1	1	108	3	5	10	0	10	20	0	60	0
1	1	2	1	1	108	15	25	50	0	−2	−4	0	−12	0
1	1	2	2	1	54	15	−1	−2	0	8	16	0	−12	0
1	1	2	3	1	9	0	2	−1	0	4	−2	0	0	0
1	1	3	1	1	27	15	−1	−2	0	−2	−4	0	3	0
1	1	3	2	1	9	0	4	−2	0	−2	1	0	0	0
1	1	2	1	1	1	0	0	0	1	0	0	0	0	0
2	1	2	2	1	1	0	0	0	0	0	0	1	0	0
2	1	2	3	1	1	0	0	0	0	0	0	0	0	1

TABLE II.14. Group S_6, irrep [42], $f_1 = 3, f_2 = 3$.

Standard basis vectors						1	2	3	5	6	8
ν_1	m_1	ν_2	m_2	τ	Com. den.						
1	1	1	1	1	18	3	5	0	10	0	0
1	1	2	1	1	72	15	25	0	−32	0	0
1	1	2	2	1	8	5	−3	0	0	0	0
2	1	1	1	1	9	0	0	1	0	2	6
2	1	2	1	1	36	0	0	32	0	−1	−3
2	1	2	2	1	4	0	0	0	0	3	−1

TABLE II.15. Group S_6, irrep [411], $f_1 = 1, f_2 = 5$.

Standard basis vectors						1	2	3	4	5	6	7	8	9	10
ν_1	m_1	ν_2	m_2	τ	Com. den.										
1	1	2	1	1	48	3	5	10	30	0	0	0	0	0	0
1	1	2	2	1	3600	−135	1	2	6	384	768	2304	0	0	0
1	1	2	3	1	1800	0	−64	2	6	−96	3	9	405	1215	0
1	1	2	4	1	200	0	0	−6	2	0	−9	3	−15	5	160
1	1	4	1	1	450	405	−3	−6	−18	2	4	12	0	0	0
1	1	4	2	1	14400	0	12288	−384	−1152	−32	1	3	135	405	0
1	1	4	3	1	4800	0	0	3456	−1152	0	−9	3	−15	5	160
1	1	4	4	1	192	0	0	0	0	160	−5	−15	3	9	0
1	1	4	5	1	576	0	0	0	0	0	405	−135	−3	1	32
1	1	4	6	1	9	0	0	0	0	0	0	0	6	−2	1

TABLE II.16. Group S_6, irrep [411], $f_1 = 2, f_2 = 4$.

Standard basis vectors						1	2	3	4	5	6	7	8	9	10
ν_1	m_1	ν_2	m_2	τ	Com. den.										
1	1	2	1	1	18	3	5	10	0	0	0	0	0	0	0
1	1	2	2	1	180	15	1	2	0	54	108	0	0	0	0
1	1	2	3	1	60	0	−4	2	0	−6	3	0	45	0	0
1	1	4	1	1	60	45	−3	−6	0	2	4	0	0	0	0
1	1	4	2	1	180	0	108	−54	0	−2	1	0	15	0	0
1	1	4	3	1	18	0	0	0	0	10	−5	0	3	0	0
2	1	1	1	1	20	0	0	0	2	0	0	3	0	5	10
2	1	2	1	1	180	0	0	0	162	0	0	−3	0	−5	−10
2	1	2	2	1	18	0	0	0	0	0	0	15	0	−1	−2
2	1	2	3	1	3	0	0	0	0	0	0	0	0	2	−1

TABLE II.17. Group S_6, irrep [411], $f_1 = 3, f_2 = 3$.

Standard basis vectors						1	2	3	5	6	8	10
ν_1	m_1	ν_2	m_2	τ	Com. den.							
1	1	2	1	1	8	3	5	0	0	0	0	0
1	1	2	2	1	40	−5	3	0	32	0	0	0
1	1	3	1	1	10	5	−3	0	2	0	0	0
2	1	1	1	1	10	0	0	2	0	3	0	0
2	1	2	1	1	40	0	0	32	0	−3	−5	0
2	1	2	2	1	8	0	0	0	0	5	−3	0
3	1	1	1	1	1	0	0	0	0	0	0	1

TABLE II.18. Group S_6, irrep [33], $f_1 = 1, f_2 = 5$.

Standard basis vectors						1	2	3	4	5
ν_1	m_1	ν_2	m_2	τ	Com. den.					
1	1	3	1	1	9	1	2	6	0	0
1	1	3	2	1	144	32	−1	−3	27	81
1	1	3	3	1	16	0	3	−1	9	−3
1	1	3	4	1	48	32	−1	−3	−3	−9
1	1	3	5	1	16	0	9	−3	−3	1

TABLE II.19. Group S_6, irrep [33], $f_1 = 2, f_2 = 4$.

Standard basis vectors						1	2	3	4	5
ν_1	m_1	ν_2	m_2	τ	Com. den.					
1	1	2	1	1	9	1	2	0	6	0
1	1	2	2	1	9	2	4	0	−3	0
1	1	2	3	1	3	2	−1	0	0	0
2	1	3	1	1	1	0	0	1	0	0
2	1	3	2	1	1	0	0	0	0	1

TABLE II.20. Group S_6, irrep [33], $f_1 = 3, f_2 = 3$.

Standard basis vectors						1	2	4
ν_1	m_1	ν_2	m_2	τ	Com. den.			
1	1	1	1	1	1	1	0	0
2	1	2	1	1	4	0	1	3
2	1	2	2	1	4	0	3	−1

TABLE II.21. Group S_6, irrep [321], $f_1 = 1, f_2 = 5$.

Standard basis vectors						1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
ν_1	m_1	ν_2	m_2	τ	Com. den.																
1	1	3	1	1	36	1	2	6	0	0	3	6	18	0	0	0	0	0	0	0	0
1	1	3	2	1	2304	128	−4	−12	108	324	−96	3	9	405	1215	0	0	0	0	0	0
1	1	3	3	1	256	0	12	−4	36	−12	0	−9	3	−15	5	160	0	0	0	0	0
1	1	3	4	1	768	−32	1	3	3	9	0	0	0	0	0	0	45	135	135	405	0
1	1	3	5	1	256	0	−9	3	3	−1	0	0	0	0	0	0	45	−15	−15	5	160
1	1	4	1	1	36	3	6	18	0	0	−1	−2	−6	0	0	0	0	0	0	0	0
1	1	4	2	1	2304	384	−12	−36	324	972	32	−1	−3	−135	−405	0	0	0	0	0	0
1	1	4	3	1	768	0	108	−36	324	−108	0	9	−3	15	−5	−160	0	0	0	0	0
1	1	4	4	1	768	0	0	0	0	0	160	−5	−15	3	9	0	108	324	−36	−108	0
1	1	4	5	1	2304	0	0	0	0	0	0	405	−135	−3	1	32	972	−324	36	−12	−384
1	1	4	6	1	36	0	0	0	0	0	0	0	0	6	−2	1	0	0	18	−6	3
1	1	5	1	1	256	160	−5	−15	−15	−45	0	0	0	0	0	0	1	3	3	9	0
1	1	5	2	1	768	0	405	−135	−135	45	0	0	0	0	0	0	9	−3	−3	1	32
1	1	5	3	1	256	0	0	0	0	0	160	−5	−15	3	9	0	−12	−36	4	12	0
1	1	5	4	1	2304	0	0	0	0	0	0	1215	−405	−9	3	96	−324	108	−12	4	128
1	1	5	5	1	36	0	0	0	0	0	0	0	0	18	−6	3	0	0	−6	2	−1

TABLE II.22. Group S_6, irrep [321], $f_1 = 2, f_2 = 4$.

Standard basis vectors						1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
ν_1	m_1	ν_2	m_2	τ	Com. den.																
1	1	2	1	1	9	1	2	0	6	0	0	0	0	0	0	0	0	0	0	0	0
1	1	2	2	1	576	−2	−4	0	3	0	54	108	0	0	0	0	405	0	0	0	0
1	1	2	3	1	192	−2	1	0	0	0	−6	3	0	45	0	0	0	0	135	0	0
1	1	3	1	1	32	2	4	0	−3	0	6	12	0	0	0	0	−5	0	0	0	0
1	1	3	2	1	32	6	−3	0	0	0	−2	1	0	15	0	0	0	0	−5	0	0
1	1	4	1	1	192	30	60	0	−45	0	−10	−20	0	0	0	0	27	0	0	0	0
1	1	4	2	1	576	270	135	0	0	0	10	−5	0	−75	0	0	0	0	81	0	0
1	1	4	3	1	18	0	0	0	0	0	10	−5	0	3	0	0	0	0	0	0	0
2	1	2	1	1	18	0	0	0	0	0	0	0	3	0	5	10	0	0	0	0	0
2	1	2	2	1	576	0	0	81	0	0	0	0	75	0	−5	−10	0	0	0	135	270
2	1	2	3	1	192	0	0	0	0	27	0	0	0	0	20	−10	0	45	0	60	−30
2	1	3	1	1	32	0	0	5	0	0	0	0	15	0	−1	−2	0	0	0	−3	−6
2	1	3	2	1	32	0	0	0	0	5	0	0	0	0	12	−6	0	−3	0	−4	0
2	1	4	1	1	192	0	0	135	0	0	0	0	−45	0	3	6	0	0	0	−1	−2
2	1	4	2	1	576	0	0	0	0	405	0	0	0	0	−108	54	0	−3	0	−4	2
2	1	4	3	1	9	0	0	0	0	0	0	0	0	0	0	0	0	6	0	−2	1

TABLE II.23. Group S_6, irrep [321], $f_1 = 3, f_2 = 3$.

Standard basis vectors						1	2	4	6	7	9	11	12	14	16
ν_1	m_1	ν_2	m_2	τ	Com. den.										
1	1	2	1	1	1	1	0	0	0	0	0	0	0	0	0
1	1	2	2	1	1	0	0	0	1	0	0	0	0	0	0
2	1	1	1	1	32	0	1	3	0	3	5	0	5	15	0
2	1	2	1	1	24	0	4	12	0	−3	−5	0	0	0	0
2	1	2	1	2	96	0	5	15	0	15	25	0	−9	−27	0
2	1	2	2	1	24	0	−3	1	0	0	0	0	15	−5	0
2	1	2	2	2	96	0	15	−5	0	45	−27	0	3	−1	0
2	1	3	1	1	32	0	15	−5	0	−5	3	0	3	−1	0
3	1	2	1	1	1	0	0	0	0	0	0	1	0	0	0
3	1	2	2	1	1	0	0	0	0	0	0	0	0	0	1

The phases ϵ are given in Table I.3.

Owing to (29a), we only tabulate the TC's for partitions whose row lengths are larger than or equal to their column lengths. Furthermore, since the TC's do not vary with m_1, that is, are identical for different values of the m_1 index if all other indices are the same, we omit all TC's other than those for which $m_1 = 1$.

The TC's given in Table II are displayed as rows, each with a common denominator entry followed by the numerators. Each entry is understood to be under a square root sign, and if the TC is negative, the minus sign is given explicitly. The ordering of the irreps is given in Table I.2.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Da Hsuan Feng and Dr. Michel Vallieres of Drexel University and Fan Wang of Nanjing University for helpful discussions and for reading the manuscript.

[1]H. A. Jahn, Phys. Rev. **96**, 989 (1954).
[2]I. G. Kaplan, Zh. Eksp. Teor. Fiz. **41**, 560 (1961) [Sov. Phys. JETP **14**, 401 (1962)].
[3]H. Horie, J. Phys. Soc. Jpn. **19**, 1783 (1964).
[4]P. Kramer, Z. Physik **205**, 181 (1967).
[5]P. Kramer, Z. Physik **216**, 68 (1968).
[6]C. R. Sarma, J. Phys. A **14**, 565 (1981).
[7]Jin-Quan Chen, J. Math. Phys. **22**, 1 (1981).
[8]Jin-Quan Chen, "SU($m + n$)⊃SU(m)×SU(n) Isoscalar Factors," (to be published).
[9]P. Kramer and T. R. Seligman, Nucl. Phys. A **136**, 545 (1969).
[10]V. I. Kukulin, Yu. F. Smirnov, and L. Majling, Nucl. Phys. A **103**, 681 (1967).
[11]Jin-Quan Chen, Fan Wang, and Mei-Juan Gao, (a) Acta Phys. Sinica **26**, 307 (1977) (in Chinese); (b) **26**, 427 (1977) [Chinese Phys. **1**, 533 (1981); **1**, 542 (1981)]; (c) Acta Phys. Sinica **27**, 203 (1978).
[12]Jin-Quan Chen, Fan Wang, and Mei-Juan Gao, Acta Phys. Sinica **27**, 31 (1978).
[13]Jin-Quan Chen and Mei-Juan Gao, "A new approach to permutation group representation," J. Math. Phys. **23**, 928 (1982).
[14]Jin-Quan Chen and Mei-Juan Gao, *Reduction Coefficients of Permutation Groups and Their Applications* (Beijing Science Press, Beijing, 1981).
[15]Michel Vallieres, Da Hsuan Feng, and Jin-Quan Chen, "SU(mn)⊃SU(m)×SU(n) Single Particle Coefficients of Fractional Parentage " (to be published).
[16]M. Hamermesh, *Group Theory and Its Applications to Physical Problems* (Addison-Wesley, Reading, MA, 1962).
[17]M. Moshinsky, *Group Theory and Many-Body Problems* (Gordon and Breach, New York, 1968).
[18]J. Paldus, J. Chem. Phys. **61** 5321 (1974); Phys. Rev. A **14**, 1620 (1976); J. Paldus and M. J. Boyle, Phys. Rev. A **22**, 2299 (1980), and the references therein.
[19]F. A. Matsen, Int. J. Quantum Chem. Symp. **8**, 379 (1974); F. A. Matsen and C. J. Nein, Int. J. Quantum Chem. **15**, 751 (1979); **20**, 861 (1981).
[20]For the most recent advances in this respect see the special volume of Springer's Lecture Notes in Chemistry Series, entitled *The Unitary Group for the Evaluation of Energy Matrix Elements*, Proceedings of the workshop on this topic at the University of Bielefeld, Germany, in 1979 (Springer-Verlag, Berlin) (to be published).
[21]Jin-Quan Chen, Fan Wang, Mei-Juan Gao, and Zu-rong Yu, Scientia Sinica **23**, 1116 (1980).
[22]Jin-Quan Chen, *A New Approach to Group Representation Theory* (Shanghai Science and Technology Press, Shanghai, 1983) (to be published).

Unitary representations of the (4 + 1)-de Sitter group on irreducible representation spaces of the Poincaré group

P. Moylan
Center for Particle Theory and Physics Department, University of Texas, Austin, Texas 78712

(Received 23 March 1982; accepted for publication 12 November 1982)

The construction of the principal continuous series of unitary representations of the simply-connected covering group of the (4 + 1)-de Sitter group on unitary irreducible representation spaces of the Poincaré group is presented. A unitary irreducible representation space of this covering group of the de Sitter group is realized as the direct sum of two irreducible representation spaces of the Poincaré group. Possible physical implications are indicated. In particular, an interpretation of the instantaneous velocity operator in the Dirac theory as the spin part of the de Sitter boosts is given. We obtain a simple method of computing the matrix elements of the generators of the de Sitter group in an SO(4) basis using the matrix elements of the generators of the four-dimensional Euclidean group. Also we obtain explicit expressions for certain matrix elements between the spinor and SO(4) basis of the representation space as functions on the coset space SO(4)/SO(3).

PACS numbers: 02.20.Qs, 11.10.Qr, 11.30.Cp

I. INTRODUCTION

The importance of representation theory to wave equations has long been known. Its role in the solutions of partial differential equations is well known and has been discussed in many books. For example, in his book, *The Theory of Spinors*, Cartan discusses the relationship between groups and wave equations. He must have realized very early in the development of the quantum theory the connection of representations with quantum mechanical equations. In his book, *Linear Representations of the Lorentz Group*, Naimark determined all the Lorentz invariant equations which are linear in the momentum, including Dirac's new relativistic wave equation.[1] Poincaré-invariant finite-dimensional spinor wave equations were discussed by Bargmann and Wigner[2]; they showed that to certain irreducible representation spaces of the Poincaré group corresponds the space of solutions of a certain wave equation, which therefore describes a particle of a definite mass and spin. The results of this paper provide further elucidation of this relationship between representation theory and wave equations. In particular, the larger SO(4,1) symmetry of the Dirac equation is constructed and should be compared with the usual Poincaré symmetry of the theory.

In the following, we will be concerned with two different symmetry groups of the quantum mechanical wave equation for a positive mass particle with arbitrary integer or half-integer spin: the restricted Poincaré group $\mathscr{P}$ and the identity component of the (4 + 1)-de Sitter group, $SO_0(4,1)$. The Poincaré group will be realized in the usual way as described in Ref. 13, the generators of $\mathscr{P}$ being represented as essentially self-adjoint operators defined on an invariant dense domain contained in $\mathscr{H}(m,s,+)$, the positive mass unitary irreducible representation space (UIR) of $\mathscr{P}$ (s denotes the integer of half-integer spin and + means positive energy). The generators, in a UIR of SO(4,1), will be realized as essentially self-adjoint operators defined on an invariant dense domain contained in $\mathscr{H}(m,s,+)\oplus\mathscr{H}(m,s,+)$. [Of course, by irreducible unitary representations we always mean single-valued representations of the simply connected covering groups associated with these groups, which we denote by $\overline{\mathscr{P}}$ and $\overline{SO(4,1)}$.] The groups $\overline{\mathscr{P}}$ and $\overline{SO(4,1)}$ are realized on $\mathscr{H}(m,s,+)$ and $\mathscr{H}(m,s,+)\oplus\mathscr{H}(m,s,+)$, respectively. This result, combined with the fact that $\mathscr{H}(m,s,+)$ has the following decomposition into irreducible representation spaces of its noncompact subgroup $\overline{SO_0(3,1)}$, formally written as[3]

$$\mathscr{H}(m,s,+)=\frac{1}{(2\pi^4)^2}\sum_{l_0}\int_0^\infty \mathscr{H}(l_0,\nu)(l_0^2+\nu^2)\,d\nu$$

gives, in the case of the principal series, the decomposition formula for a UIR of $\overline{SO_0(4,1)}$ with respect to UIR's of $\overline{SO_0(3,1)}$, first obtained by Ström using the quaternion matrix description of $\overline{SO_0(4,1)}$.[4] Our method is, in contrast to Ström's, based on the description of O(4,1) by projective transformations of T_3[5] (T_3 denotes the mass hyperboloid of the momentum space of Minkowski spacetime, i.e., $T_3=\{p_\mu \mid p_\mu p^\mu = m^2\}$). Since we realize the principal series UIR's of $\overline{SO_0(4,1)}$ in terms of the standard UIR's of $\overline{\mathscr{P}}$ used to describe elementary particles, this makes the physical content immediate. In fact, this realization of $\overline{SO_0(4,1)}$ serves as the basis for a description of the relativistic rotator model,[6] which has been most extensively developed by Böhm.[7] In that model the invariant operator, being m^2 in a UIR, is replaced by the first-order Casimir operator of $SO_\lambda(4,1)$ in the above realization—λ is the contraction parameter used in the contraction of $SO_0(4,1)$ into $\mathscr{P}$.[8]

We have organized the material in this paper in the following way. In Sec. II we present a brief résumé of the groups $SO_0(4,1)$, $\mathscr{P}$, $SO_0(3,1)$, and their simply connected covering groups and summarize their pertinent irreducible unitary representations. Section III describes the realization of $SO_0(4,1)$ on the momentum space of Minkowski spacetime. Section IV is devoted to description of the orbital part of the $\overline{SO_0(4,1)}$ principal series representation constructed here. In Sec. V we realize the principal series representations of $\overline{SO_0(4,1)}$ on $\mathscr{H}(m,s,+)\oplus\mathscr{H}(m,s,+)$. A lengthy appendix is devoted to the proof of the irreducibility of these repre-

 0022-2488/83/122706-16$02.50

sentations. Explicit expressions for the $\overline{SO_0(4)}$ basis vectors of the representation together with the matrix elements of the generators in this basis are obtained.

Notation: Lower case Roman indices, in general, range from 1, 2, 3 or 0, 1, 2, 3, 5, 6 and Greek ones from 0, 1, 2, 3. $\eta_{\mu\nu} = \text{diag}\,(1, -1, -1, -1)$ denotes the metric tensor of Minkowski space, $M_{3,1}$, and $g_{ab} = \text{diag}\,(1, -1, -1, -1, -1, +1)$ is the metric tensor of a $(4+2)$-dimensional pseudo-Euclidean space, $M_{4,2}$. $\eta_{ab} = \text{diag}\,(1, -1, -1, -1, -1)$ will be used to denote the metric tensor of a $(4+1)$-dimensional Minkowski space $M_{4,1}$. The translation generators are denoted by a contravariant 4-vector operator $P^\mu = (E^{\text{op}}, \mathbf{P}^{\text{op}})$ and the contravariant 4-momentum vector is, likewise, $p^\mu = (E, \mathbf{p})$. The position operator in the momentum representation is given by $Q^\mu = -i\partial/\partial p_\mu = -(i\partial/\partial E, (1/i)\nabla_p)$ and transforms as a contravariant 4-vector. It follows that

$$[P^\mu, Q^\nu] = i\eta^{\mu\nu} \tag{1}$$

are the relativistic Heisenberg canonical commutation relations. Throughout, both $\hbar$ and c are set equal to 1. When we mean a contraction of two vectors in Euclidean or pseudo-Euclidean space, we denote it by $\xi_\mu \eta^\mu$ (summation convention). Finally, $\dagger$ denotes the complex conjugate transpose and * signifies the complex conjugate.

II. RÉSUMÉ OF $SO_0(4,1)$, $\mathscr{P}$ AND $SO_0(3,1)$ AND THE PERTINENT UNITARY IRREDUCIBLE REPRESENTATIONS OF THEIR SIMPLY CONNECTED COVERING GROUPS

We denote by $O(p,q)$ the group of all homogeneous linear transformations of real $(p+q)$-dimensional space which leave invariant the quadratic form

$$-X_1^2 - X_2^2 - \cdots - X_p^2 + X_{p+1}^2 + \cdots + X_{p+q}^2. \tag{2}$$

By $SO_0(p,q)$ we denote the component of $O(p,q)$ connected to the identity transformation. The Lie algebra of $SO_0(p,q)$ is written as $so(p,q)$.

The isomorphism $B_2 \approx C_2$[9] entails the local isomorphism $SO_0(4,1) \approx Sp(1,1)$.[10] Since $Sp(1,1) = U(2,2) \cap Sp(4,C)$[10] is simply connected, this Lie algebra isomorphism permits a description of $\overline{SO_0(4,1)}$ by 2×2 quaternion matrices, used in Ref. 4. The Hermitian generators A_{ab} of the Lie algebra $so(n,1)$ obey the commutation relations

$$[A_{ab}, A_{cd}] = -i(\eta_{ac}A_{bd} + \eta_{bd}A_{ac} - \eta_{bc}A_{ad} - \eta_{ad}A_{bc}) \tag{3}$$
$$(a,b = 0,1,2,3,5).$$

[For $SO(n+1)$ replace η_{ac} by $-\delta_{ac}$; we need this result in the Appendix.] $A_{\mu\nu}$ generate the homogeneous Lorentz group $SO_0(3,1)$ whose commutation relations are

$$[A_{\mu\nu}, A_{\rho\sigma}] = -i(\eta_{\mu\rho}A_{\nu\sigma} + \eta_{\nu\sigma}A_{\mu\rho} - \eta_{\nu\rho}A_{\mu\sigma} - \eta_{\mu\sigma}A_{\nu\rho}), \tag{4}$$

and $A_{5\mu}$ form a Lorentz vector operator

$$[A_{\mu\nu}, A_{5\rho}] = i(\eta_{\nu\rho}A_{5\nu} - \eta_{\mu\rho}A_{5\nu}), \tag{5}$$

satisfying

$$[A_{5\mu}, A_{5\nu}] = iA_{\mu\nu}. \tag{6}$$

We call $A_{5\mu}$ the generators of the de Sitter boosts.[11]

The covering group of the Lorentz subgroup $SO_0(3,1)$ is isomorphic to the set of all unimodular complex matrices in two dimensions: $SL(2,C)$. The restricted Poincaré group $\mathscr{P}$ is the semidirect product of $SO_0(3,1)$ with the four-dimensional translation group T_4: $\mathscr{P} = SO_0(3,1) \boxplus T_4$. Using the well-known isomorphism of $SO_0(3,1)$ with $SL(2,C)$[12] and the characterization of four-dimensional vectors by Hermitian matrices in two dimensions gives the isomorphism $\overline{\mathscr{P}} \simeq SL(2,C) \boxplus T_4$.[13] The Hermitian generators of $\overline{\mathscr{P}}$ are P_μ and $L_{\mu\nu}$ with the commutation relations

$$[P_\mu, P_\nu] = 0, \tag{7}$$

$$[L_{\mu\nu}, P_\rho] = i(\eta_{\nu\rho}P_\mu - \eta_{\nu\rho}P_\nu), \tag{8}$$

the commutation relations of the generators $L_{\mu\nu}$ of $SO_0(3,1)$ being the same as in (4).

$SO_0(4,1)$ has two independent Casimir operators for which we take

$$Q_1 = -\tfrac{1}{2}A_{ab}A^{ab} \quad (a,b,c,d,e = 0,1,2,3,5), \tag{9}$$

$$Q_2 = -w_a w^a, \quad w_a = \tfrac{1}{8}\epsilon_{abcde}A^{bc}A^{de}. \tag{10}$$

The two independent Casimir operators of the Lorentz subgroup will be denoted by $Q_1^{SO(3,1)}$ and $Q_2^{SO(3,1)}$. $Q_1^{SO(3,1)}$ is the same as its counterpart, Eq. (9) above, except Roman indices are replaced by Latin ones, and $Q_2^{SO(3,1)} = w_5$. Invariant operators of $\mathscr{P}$ are $P_\mu P^\mu$ and $W = -w_\mu w^\mu$, $w_\mu = \tfrac{1}{2}\epsilon_{\mu\nu\rho\sigma}P^\nu L^{\rho\sigma}$.

For $\overline{SO_0(3,1)}$ and $\overline{SO_0(4,1)}$, the values of the two independent Casimir operators can serve to label the UIR's.[8] Besides the values of the two above invariant operators, an additional label is necessary to describe the UIR's of $\overline{\mathscr{P}}$, namely the sign of the energy, p_0. The UIR's of the Lorentz group are labeled by the numbers (k_0, c) with k_0 a nonnegative integer or semi-integer and c purely imaginary, $c = -i\rho$ (the principal series) or $k_0 = 0$, $0 \leqslant c \leqslant 1$ (the supplementary series).[14] The relation between c, k_0 and the two Casimir operators of $\overline{SO_0(3,1)}$ is in an irreducible representation:

$$Q_1^{SO(3,1)} = (1 - c^2 - k_0^2)I, \quad Q_2^{SO(3,1)} = 2ik_0 cI. \tag{11}$$

The finite-dimensional spinor representations of $\overline{SO_0(3,1)}$ have c real with $c^2 = (k_0 + n)^2$ and have dimensions $c^2 - k_0^2$. The UIR's of $\overline{\mathscr{P}}$ were found by Wigner.[13] The ones relevant to our discussion are

$$m_\pm: \quad m^2 > 0, \quad p_0 \gtrless 0, \quad s = 0, \tfrac{1}{2}, 1, \cdots. \tag{12}$$

In a UIR the invariant operators are

$$P_\mu P^\mu = m^2 I, \quad W = m^2 s(s+1)I. \tag{13}$$

The UIR's of $SO(4,1)$ were first determined by Thomas[15] based on the infinitesimal method. The classification was completed by Newton and Dixmier.[16] We are only concerned with the principal series representations (UIR's)

$$(\rho, r): \quad \rho^2 > 0, \quad r = 0, \tfrac{1}{2}, 1, \cdots, \tag{14}$$

for which the Casimir operators take the following values[17]:

$$Q_1 = \rho^2 + \tfrac{9}{4} - r(r+1), \tag{15}$$

$$Q_2 = r(r+1)(\rho^2 + \tfrac{1}{4}). \tag{16}$$

In the following we denote the various UIR's by the corre-

sponding set of numbers labeling them, e.g., $(m,s,+)$ for $\overline{\mathscr{P}}$ or (ρ,r) for $\overline{SO_0(4,1)}$.

III. A REALIZATION OF $\overline{SO_0(4,1)}$ AS PROJECTIVE TRANSFORMATIONS OF THE MASS HYPERBOLOID OF THE MOMENTUM SPACE TO MINKOWSKI SPACE-TIME

In this section we give the geometrical description of $SO_0(4,1)$, which we used to obtain our construction of the principal series irreducible representations. First we present the Klein description of the conformal group O(4,2).[18,19] Imbed $M_{3,1}$ into $M_{4,1}$ as the hyperplane $\overline{\xi}^5 = 0$ (see Fig. 1). Consider the paraboloid

$$\overline{\xi}^5 = +\overline{\xi}^{0^2} - \overline{\xi}^{1^2} - \overline{\xi}^{2^2} - \overline{\xi}^{3^2}. \tag{17}$$

Introduce homogeneous coordinates $(\xi^0,\xi^1,\xi^2,\xi^3,\xi^5,\xi^6)$ by

$$\overline{\xi}^5 = \xi^5/\xi^6, \quad \overline{\xi}^\mu = \xi^\mu/\xi^6 \quad (\mu = 0,1,2,3). \tag{18}$$

Then (17) becomes

$$\xi^5\xi^6 = +\xi^{0^2} - \xi^{1^2} - \xi^{2^2} - \xi^{3^2}. \tag{19}$$

Since

$$\xi^5\xi^6 = \tfrac{1}{4}(\xi^5+\xi^6)^2 - \tfrac{1}{4}(\xi^5-\xi^6)^2, \tag{20}$$

we have that (19) becomes

$$\Omega(\xi) = \tfrac{1}{4}(\xi^5+\xi^6)^2 - \tfrac{1}{4}(\xi^5-\xi^6)^2 - \xi^{0^2} + \xi^{1^2} + \xi^{2^2} + \xi^{3^2} = 0.$$

The linear transformations of $M_{4,2}$ which leave invariant the above form is O(4,2).

The general transformation of the $\overline{\xi}^\mu$'s $(\mu = 0,1,2,3)$ induced by $B \in SO(4,2)$ is

$$\xi^{\mu\prime} = \frac{b^\mu{}_\nu \xi^\nu + b^\mu{}_5 \xi^2 + b^\mu{}_6}{b^6{}_\mu \xi^\mu + b^6{}_5 \xi^2 + b^6{}_6} \quad (\xi^2 = \xi^\mu \xi_\mu), \tag{21}$$

where

$$B = \begin{vmatrix} b^6{}_6 & b^6{}_5 & b^6{}_1 & b^6{}_2 & b^6{}_3 & b^6{}_0 \\ b^5{}_6 & b^5{}_5 & b^5{}_1 & b^5{}_2 & b^5{}_3 & b^5{}_0 \\ b^1{}_6 & b^1{}_5 & b^1{}_1 & b^1{}_2 & b^1{}_3 & b^1{}_0 \\ b^2{}_6 & b^2{}_5 & b^2{}_1 & b^2{}_2 & b^2{}_3 & b^2{}_0 \\ b^3{}_6 & b^3{}_5 & b^3{}_1 & b^3{}_2 & b^3{}_3 & b^3{}_0 \\ b^0{}_6 & b^0{}_5 & b^0{}_1 & b^0{}_2 & b^0{}_3 & b^0{}_0 \end{vmatrix} \in O(4,2).$$

It is easy to show that every transformation of the form (21) can be obtained as a product of an inversion, $\xi^{\mu\prime} = \xi^\mu/|\xi|^2$, a scale transformation, $\xi^{\mu\prime} = \lambda\xi^\mu$, and a Minkowskian motion, $\xi^{\mu\prime} = O^\mu{}_\nu \xi^\nu + a^\mu$, from which it follows by an extension of Liouville's theorem to $M_{3,1}$ that O(4,2) corresponds to the conformal transformations on $M_{3,1}$.[20]

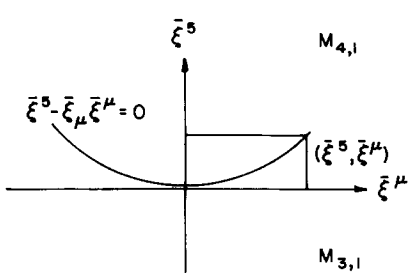


FIG. 1. Projection of paraboloid in $M_{4,1}$ onto $M_{3,1}$.

Next we describe a subgroup of O(4,2) which leaves invariant the hyperboloid $T_3^{(0)} = \{\overline{\xi}^\mu \in M_{3,1} \mid \overline{\xi}_\mu \overline{\xi}^\mu = 1\}$.

We determine it as follows:

$T_3^{(0)}$ invariant means $\overline{\xi}'_\mu \overline{\xi}^{\mu\prime} = \overline{\xi}_\mu \overline{\xi}^\mu$ and because $\overline{\xi}^5 = \overline{\xi}_\mu \overline{\xi}^\mu$ we demand that $\xi^5 - \xi^6$ remain invariant. Since O(4,2) leaves invariant

$$\tfrac{1}{4}(\xi^5+\xi^6)^2 - \tfrac{1}{4}(\xi^5-\xi^6)^2 - \xi^{0^2} + \xi^{1^2} + \xi^{2^2} + \xi^{3^2},$$

we see that the subgroup of the conformal transformations which leaves $T_3^{(0)}$ invariant is the group of linear transformations of $M_{4,1}$, which preserves

$$\Omega'(\xi) = \tfrac{1}{4}(\xi^5+\xi^6)^2 - \xi^{0^2} + \xi^{1^2} + \xi^{2^2} + \xi^{3^2}. \tag{22}$$

The component of this group which is connected to the identity is $SO_0(4,1)$.

Now let A be an element of this $SO_0(4,1)$. Referred to the coordinates $\frac{1}{2}(\xi^5+\xi^6)$, $\frac{1}{2}(\xi^5-\xi^6)$, ξ^μ, its matrix is given by

$$A = \begin{vmatrix} a^6{}_6 & a^6{}_1 & a^6{}_2 & a^6{}_3 & a^6{}_0 \\ a^1{}_6 & a^1{}_1 & a^1{}_2 & a^1{}_3 & a^1{}_0 \\ a^2{}_6 & a^2{}_1 & a^2{}_2 & a^2{}_3 & a^2{}_0 \\ a^3{}_6 & a^3{}_1 & a^3{}_2 & a^3{}_3 & a^3{}_0 \\ a^0{}_6 & a^0{}_1 & a^0{}_2 & a^0{}_3 & a^0{}_0 \end{vmatrix}, \tag{23}$$

and its action on the coordinates is

$$\begin{aligned} \xi^{\bar{5}} &= \tfrac{1}{2}(\xi^{5\prime}+\xi^{6\prime}) = a^6{}_6 \tfrac{1}{2}(\xi^5+\xi^6) + \sum_{\nu=0}^{3} a^6{}_\nu \xi^\nu, \\ \tfrac{1}{2}(\xi^{5\prime}-\xi^{6\prime}) &= \tfrac{1}{2}(\xi^5-\xi^6), \\ \xi^{\mu\prime} &= a^\mu{}_6 \tfrac{1}{2}(\xi^5+\xi^6) + \sum_{j=0}^{3} a^\mu{}_\nu \xi^\nu. \end{aligned} \tag{24}$$

Thus on $V_{3,1}$ (the $\overline{\xi}^\mu$ space), the transformation induced by A is

$$\begin{aligned} \overline{A}:\overline{\xi}^\mu \to \overline{\xi}^{\mu\prime}:\overline{\xi}^{\mu\prime} &= \frac{\xi^{\mu\prime}}{\xi^{6\prime}} \\ &= \frac{a^\mu{}_6 \tfrac{1}{2}(\xi^5+\xi^6) + \Sigma_{\nu=0}^3 a^\mu{}_\nu \xi^\nu}{\tfrac{1}{2}(a^6{}_6-1)\xi^5 + \tfrac{1}{2}(a^6{}_6+1)\xi^6 + \Sigma a^6{}_\nu \xi^\nu} \\ &= \frac{a^\mu{}_6 \tfrac{1}{2}(\overline{\xi}^2+1) + \Sigma a^\mu{}_\nu \overline{\xi}^\nu}{\tfrac{1}{2}(a^6{}_6-1)\overline{\xi}^2 + \tfrac{1}{2}(a^6{}_6+1) + \Sigma a^6{}_\nu \xi^\nu}. \end{aligned}$$

For $\overline{\xi}^\mu \in T_3^{(0)}$ (that is to say, ξ^μ a point on the hyperboloid) we have

$$\overline{A}:\overline{\xi}^\mu \to \overline{\xi}^{\mu\prime}:\overline{\xi}^{\mu\prime} = (a^\mu{}_6 + \textstyle\sum a^\mu{}_\nu \overline{\xi}^\nu)/(a^6{}_6 + \sum a^6{}_\nu \overline{\xi}^\nu). \tag{25}$$

The Jacobian of this transformation (restricted to $T_3^{(0)}$) is[18]

$$J = \left| \frac{\partial \overline{\xi}^{\mu\prime}}{\partial \overline{\xi}^\nu} \right| = + \frac{1}{|a^6{}_6 + \Sigma a^6{}_\rho \overline{\xi}^\rho|^5}. \tag{26}$$

The denominator of this expression may vanish for certain transformations. This means that in order to consider the action of SO(4,1) on $T_3^{(0)}$ we must first compactify it by the adjunction of a surface at infinity. The measure on $T_3^{(0)}$ transforms in the following way under A [21]:

$$d\Omega_{T_3^{(0)}} = \frac{1}{|a^6{}_6 + \Sigma a^6{}_\nu \overline{\xi}^\nu|^3} d\Omega_{T_3^{(0)}}, \tag{27}$$

provided, of course, that the denominator does not vanish. Therefore, we have for $f(\bar{\xi}')$ and $g(\bar{\xi}')$ L^2 functions on $T_3^{(0)}$

$$\int_{T_3^{(0)}} \bar{f}(\bar{\xi}')g(\bar{\xi}')d\Omega_{T_3^{(0)}},$$

$$= \int_{T_3^{(0)}} \frac{\bar{f}(\bar{A}(\bar{\xi}))}{[a^6{}_6 + \Sigma a^6{}_\nu \bar{\xi}^\nu]^{3/2}} \frac{\bar{g}(\bar{A}(\bar{\xi}))}{[a^6{}_5 + \Sigma a^5{}_\nu \bar{\xi}^\nu]^{3/2}} d\Omega_{T_3^{(0)}} \tag{28}$$

with $\bar{A}(\bar{\xi})$ given by (25). Here $d\Omega_{T_3^{(0)}} = \delta(\bar{\xi}^2 - 1)\, d^4\xi$ means

$$d\Omega_{T_3^{(0)}} = \begin{cases} d\Omega_{T_3^{+(0)}} = \delta(\bar{\xi}^2 - 1)\theta(\bar{\xi}^0)\, d^4\bar{\xi}, & \bar{\xi}^0 > 0, \\ d\Omega_{T_3^{-(0)}} = \delta(\bar{\xi}^2 - 1)[1 - \theta(\bar{\xi}^0)]\, d^4\bar{\xi}, & \bar{\xi}^0 < 0. \end{cases} \tag{29}$$

[T_3^+ denotes the upper sheet ($\bar{\xi}^0 > 0$) of the hyperboloid, T_3, and T_3^- denotes the lower sheet ($\bar{\xi}^0 < 0$).]

IV. THE CONSTRUCTION OF THE ORBITAL PART OF THE PRINCIPAL SERIES REPRESENTATIONS OF $\overline{SO_0(4,1)}$ on $\mathscr{H}(m,s,+) \oplus \mathscr{H}(m,s,+)$

After this geometric groundwork we turn to the construction of the (ρ,s) of $SO_0(4,1)$ on unitary irreducible representation spaces of $\overline{\mathscr{P}}$.

First we describe a continuous representation of $(\rho,0)$ on $\mathscr{L}^2(T_3^0)$. Given $\bar{A} \in SO_0(4,1)$ and $f \in \mathscr{L}^2(T_3^0)$, let

$$(T^{\bar{A}}_\rho f)(\bar{\xi}) = \frac{f(\bar{A}^{-1}\bar{\xi})}{|a^6{}_6(\bar{A}^{-1}) + \Sigma_\mu a^6{}_\mu(\bar{A}^{-1})\bar{\xi}^\mu|^{3/2+i\rho}} \tag{30}$$

with

$$(\bar{A}^{-1}\bar{\xi})^\mu = \frac{a^\mu{}_6(\bar{A}^{-1}) + \Sigma_\nu a^\mu{}_\nu(\bar{A}^{-1})\bar{\xi}^\nu}{a^6{}_6(\bar{A}^{-1}) + \Sigma_\nu a^6{}_\nu(\bar{A}^{-1})\bar{\xi}^\nu} \tag{31}$$

extended to include points at infinity.

The representation property

$$T^{\bar{A}_1}\cdot T^{\bar{A}_2} = T^{\bar{A}_1\bar{A}_2} \tag{32}$$

is satisfied because of the multiplicative property of Jacobians[18] and unitarity is assured because of (28).

To obtain representation of $\overline{SO_0(4,1)}$ which acts on $\mathscr{L}^2(T_3)$ we introduce a constant m (which we identify with the mass in a UIR of $\overline{\mathscr{P}}$), and define new variables p^μ by $\xi^\mu = p^\mu/m$. Then (30) and (31) define a unitary representation of $SO_0(4,1)$ on $\mathscr{L}^2(T_3)$, where $T_3 = \{p^\mu \mid p_\mu p^\mu = m^2\}$, by replacing $\bar{\xi}^\mu$ in these equations everywhere by p^μ/m and then redefining f as the original f composed with the function which corresponds to multiplication by m, i.e., $f \to f \circ i_m$, where i_m is the function of multiplication by m.

The proof of the irreducibility of this representation, which is based on the proof presented in Refs. 18 and 19, is given in Appendix A.

It is instructive at this point to compute the action of the infinitesimal generators of $SO_0(4,1)$ on $\mathscr{L}^2(T_3)$. We take the following expressions for the matrices of the generators of the rotations in the various planes of projective space:

(1) The Lorentz rotations:

$$I^{ij} = \begin{vmatrix} \ddots & 0 & \ddots & -1 & \\ & \vdots & & \vdots & \\ & 1 & \cdots & 0 & \\ & & & & \ddots \end{vmatrix}, \quad I^{i0} = -\begin{vmatrix} 0 \cdots & & & \\ \ddots & 0 & \cdots & 1 \\ & \vdots & & \vdots \\ & 1 & \cdots & 0 \end{vmatrix} \tag{33}$$

($i,j = 1,2,3$ and in I^{ij} the 1 is in the ith row, jth column, and the -1 is in the jth row, ith column with $i < j$).

(2) The de Sitter boosts:

$$I^{5i} = \begin{vmatrix} 0 & \cdots & -1 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \\ 1 & \cdots & 0 & & \\ \vdots & & \vdots & \ddots & \\ 0 & & & & 0 \end{vmatrix}, \quad I^{50} = -\begin{vmatrix} 0 & \cdots & \cdot & \cdots & 1 \\ \vdots & & & & \vdots \\ \cdot & & 0 & & \cdot \\ \vdots & & & & \vdots \\ 1 & \cdots & \cdot & \cdots & 0 \end{vmatrix}. \tag{34}$$

Let us define the linear operator $U^0(A)$ in $\mathscr{L}^2(T_3)$ by

$$|\psi') = U^0(A)|\psi), \tag{35}$$

$$\psi'(p) = (p|\psi') = (p|U^0(A)|\psi) = (T^{\bar{A}}_\rho \psi)(p),$$

where $U^0(A)$ has matrix elements

$$(p|U^0(A)|p') = (T^{\bar{A}}_\rho p')(p)$$

$$= \frac{1}{|\mu(\bar{A},p)|^{3/2+i\rho}} |p_0'|\delta^3(\bar{A}^{-1}p - p')$$

[$|\mu(\bar{A},p)|^{3/2+i\rho}$ is the multiplier in (30)]. For an infinitesimal rotation

$$A = I + \omega I^{ab}$$

in the a–b plane of projective space, $U^0(A)$ can be written as:

$$U^{(0)}(A) = I - i\omega M^{ab\,(0)}. \tag{36}$$

Using this equation along with (30) and (35), we explicitly obtain, by expanding $(T^{\bar{A}}_\rho \psi)(p)$ in a Taylor series in ω and keeping only terms of order ω, the following expressions for the infinitesimal generators $M^{(0)}_{ab}$:

$$M^{(0)}_{\mu\nu} = (Q_\mu P_\nu - Q_\nu P_\mu)\cdot I = M_{\mu\nu}\cdot I,$$

$$M^{(0)}_{5\mu} = -\frac{1}{\lambda} B^{(0)}_\mu \cdot I\left(\frac{1}{\lambda} B^{(0)}_\mu = \frac{1}{2m}\{P^\rho, M_{\rho\mu}\} + \frac{\rho}{m} P_\mu\right) \tag{37}$$

$$I=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix},\quad \{A,B\}=AB+BA,$$

where $\bar{5}$ is defined in Eq. (24) and I comes from the following: $\mathscr{L}^2(T_3)\cong\mathscr{L}^2(T_3^+)\oplus\mathscr{L}^2(T_3^-)$ as Hilbert spaces, and a vector $|\psi)\in\mathscr{L}^2(T_3)$ may be written as

$$|\psi)=\begin{pmatrix}|\psi_1)\\ |\psi_2)\end{pmatrix}$$

with

$$|\psi_1)\in\mathscr{L}^2(T_3^+),\ |\psi_2)\in\mathscr{L}^2(T_3^-).$$

The infinitesimal operators M_{ab} may be *formally* written as the direct sum of the operators $Q_\mu P_\nu - Q_\nu P_\mu$ and $-(1/\lambda)B^{(0)}_\mu$, where each of operators acts on dense subspaces of $\mathscr{L}^2(T_3^+)$ and $\mathscr{L}^2(T_3^-)$ in the usual way. We can verify that expressions (37) satisfy (4), (5), and (6), and so they are restrictions of essentially self-adjoint generators of the representation. The general form of the $U^0(A)$ is

$$U^0(A)=e^{-i\omega^{ab}M^{(0)}_{ab}/2}.$$

V. THE CONSTRUCTION OF THE PRINCIPAL SERIES REPRESENTATION $(\sqrt{m^2}/\lambda^2,s)$ OF $\overline{\mathrm{SO}_0(4,1)}$ ON $\mathscr{H}(m,s,+)\oplus\mathscr{H}(m,s,+)$

In order to proceed with the construction, we find it necessary to describe a well-known realization of the UIR's of $\overline{\mathscr{P}}$ in terms of wavefunctions $\psi(p;\xi_1\cdots\xi_{2s})$ with $p\in T_3$ and which are totally symmetric in their $2s$ four-valued variables $\xi_n\cdots\xi_{2s}$.[2] First let $\mathscr{L}^2(T_3)\otimes C^4_{(1)}\otimes C^4_{(2)}\otimes\cdots\otimes C^4_{(2s)}$ be the tensor product of Hilbert spaces, where $C^4_{(i)}$ is the ith copy of C^4, a four-dimensional vector space over C. Next, introduce the generalized basis $|p\rangle\otimes e_{\xi_1}\otimes e_{\xi_2}\otimes\cdots\otimes e_{\xi_{2s}}$. The components of a vector ψ with respect to this basis are denoted by $\psi(p;\xi_1\cdots\xi_{2s})$. Let $\mathscr{S}^{(s)}$ denote the space of all such ψ which are symmetric in their $2s$ variables $\xi_1\cdots\xi_{2s}$. Now for every $k=1,\ldots,2s$ define a set of four matrices $\gamma^\nu_{(k)}$, being four of the 4×4 matrices[22] of the generators of the Clifford algebra S_2[23] which satisfy

$$\{\gamma^\nu_{(k)},\gamma^\mu_{(1)}\}=2\eta^{\nu\mu}_{(k)}\delta_{k1}. \tag{38}$$

[The $\gamma^\mu_{(k)}$'s act on the space $C^4_{(k)}$ in the usual way, so that

$$\psi'(p;\xi_1\cdots\xi_k\cdots\xi_{2s})=\sum_{\xi'_k}(\gamma^\mu_{(k)})_{\xi_k\xi'_k}\psi(p;\xi'_1\cdots\xi'_k\cdots\xi'_{2s})\Big].$$

Next introduce the matrices[24]

$$\Gamma^\mu=\sum_{k=1}^{2s}\Gamma^\mu_{(k)}=\frac{1}{2}\sum_{k=1}^{2s}\gamma^\mu_{(k)}. \tag{39}$$

These lead to the generalized Dirac equation[24]

$$(\Gamma^\mu P_\mu - sm)\psi=0. \tag{40}$$

Consider the set of all $\psi\in\mathscr{S}^{(s)}$ which satisfy the Bargmann–Wigner equations

$$(\gamma^\mu_{(k)}p_\mu - m)\psi=0\quad(k=1,\ldots,2s). \tag{41}$$

Denote the space of all these ψ by "$\mathscr{R}^{(s)}$". We can define in "$\mathscr{R}^{(s)}$" the following inner product: for every $\psi\in$"$\mathscr{R}^{(s)}$"

$$(\psi,\psi)=\int_{T_3}\Big|\sum_\xi\psi^*\gamma^0_1\,\gamma^0_2\cdots\gamma^0_{2s}\psi\Big|\,d\Omega \tag{42}$$

[the inner product is obtained from this norm by polarization; we denote it also by $(\ ,\ )$],

where ψ^* is the Hermitian conjugate of the vector ψ. Let $\mathscr{R}^{(s)}$ be the Hilbert space completion of the set of all $\psi\in$"$\mathscr{R}^{(s)}$" for which the inner product defined by (42) is finite. As shown in Ref. 2, $\mathscr{R}^{(s)}$, for every s, is the carrier space of a unitary representation of the restricted Poincaré group $\overline{\mathscr{P}}$ with invariant inner product defined via (42) and which has the following decomposition into irreducible representations:

$$\mathscr{R}^{(s)}=\mathscr{H}(m,s,+)\oplus\mathscr{H}(m,s,-). \tag{43}$$

A basis in which this decomposition is accomplished is the unitary canonical basis of vectors $|ps_3\epsilon)$ [$\epsilon=\mathrm{sgn}(p_0)$].[13]

The relation of the Bargmann–Wigner wavefunctions $\psi(p;\xi)$ to the canonical basic vectors is obtained through the spinor form of the representation of $\overline{\mathscr{P}}$. Introduce the nonorthogonal spinor basis defined by[25]

$$(-1)^{s-\dot{A}}|p,-\bar{\dot{A}};\epsilon\}=|p,\underline{A};\epsilon\}=\sum_{s_3}|ps_3\epsilon)\mathscr{D}^s_{s_3A}(L(p))$$

and

$$(-1)^{s+\dot{A}}|p,-\underline{\dot{A}};\epsilon\}=|p,\bar{\dot{A}};\epsilon\}=\sum_{s_3}|ps_3\epsilon)\mathscr{D}^{s\dagger}_{s_3\dot{A}}(\epsilon L^{-1}(p))$$

$$(A=s,s-1,\ldots,-s)$$

$$\{p,\underline{A};\epsilon|\phi)^*=(\phi|p,\underline{\dot{A}};\epsilon\};(\phi|p,\bar{A};\epsilon\}=\{p,\bar{\dot{A}};\epsilon|\phi)^*, \tag{44}$$

where $\mathscr{D}^s_{s_3A}(L(p))$ is the representation of the inverse boost $L(p)$, taking p_μ to rest, in the $(2s+1)$-dimensional irreducible representation space of $\overline{\mathrm{SO}_0(3,1)}$ (see Ref. 13).

The transformation law for the components of a vector ϕ in the canonical basis is[2]

$$(ps_3|U(\Lambda,a)\phi)=e^{ia_\mu p^\mu}\sum_{s'_3}\mathscr{D}^s_{s_3s'_3}(R(\Lambda,\Lambda^{-1}p))(\Lambda^{-1}p,s'_3|\phi),$$

where $U(\Lambda,a)$ is the unitary operator corresponding to the Poincaré transformation $p'^\mu=\Lambda^\mu_\nu p^\nu+a^\mu$, and $R(\Lambda,\Lambda^{-1}p)=L(p)\Lambda L^{-1}(\Lambda^{-1}p)$ is the Wigner rotation. We have for the transformation law of the components of ϕ in the spinor basis[25]:

$$\{p,\underline{A};\epsilon|U(\Lambda,a)\phi)=e^{ip_\mu a^\mu}\mathscr{D}^s_{\underline{A}\bar{B}}(\Lambda)\{\Lambda^{-1}p,\underline{B};\epsilon|\phi),$$

$$\{p,\bar{\dot{A}};\epsilon|U(\Lambda,a)\phi)=e^{ip_\mu a^\mu}\mathscr{D}^{s*}_{\dot{A}\underline{B}}(\Lambda)\{\Lambda^{-1}p,\bar{\dot{B}};\epsilon|\phi). \tag{45}$$

We state the following momentum representation form of the Bargmann–Wigner equations in spinor notation[25,26]:

$$\{p,\underline{A};\epsilon|\phi)=\hat{p}^{ss}_{\underline{A}\dot{B}}\{p,\bar{\dot{B}};\epsilon|\phi);\{p,\bar{\dot{B}};\epsilon|\phi)=\hat{p}^{ss}_{\bar{\dot{B}}\underline{A}}\{p,\underline{A};\epsilon|\phi) \tag{46}$$

with

$$\hat{p}^{ss}_{\underline{A}\dot{B}}=\mathscr{D}^s_{AB}(\{p_0+\mathbf{p}\cdot\boldsymbol{\sigma}\}). \tag{47}$$

Up to a possible unitary transformation of the Hilbert space which does not affect the p variable, the components of the vector $\phi\in\mathscr{R}^{(s)}$ in the spinor basis,

$$\frac{1}{\sqrt{2}}\begin{pmatrix}\{p,\underline{A};\epsilon|\phi)\\ \{p,\bar{\dot{B}};\epsilon|\phi)\end{pmatrix},$$[25]

can be identified with the Bargmann–Wigner function $\phi(p;\xi,\ldots,\xi_{2s})$. They have the same transformation laws (45). Also, the infinitesimal generators of the representation of $\mathscr{P}$ acting on the Bargmann–Wigner $\phi(p,\xi,\ldots,\xi_{2s})$ are[2]

$$P_\mu = p_\mu, \quad L_{\mu\nu} = i\left(p_\mu \frac{\partial}{\partial p_\nu} - p_\nu \frac{\partial}{\partial p_\mu}\right) + S_{\mu\nu}$$

with $S_{\mu\nu}$ defined by

$$S_{\mu\nu} = \sum_k S_{\mu\nu(k)} = \frac{i}{4}\sum_k [\gamma_{(k)_\mu}, \gamma_{(k)_\nu}] \quad (\gamma_{(k)_\mu} = \eta_{\mu\nu}\gamma^\nu_{(k)}). \tag{48}$$

These relations are simple consequences of the transformation law of the components of ϕ in the spinor basis [see Eq. (45)].

In the realization of the representation of $\mathscr{P}$ on $\mathscr{H}(m,s,\pm)$, the norm (42) corresponds to the following:

$$\begin{aligned}(\phi,\phi) &= \sum_{s_3}\int_{T_3} d\Omega_{T_3}\,(\phi \mid ps_3,\epsilon)(ps_3,\epsilon\mid\phi) \\ &= \sum_{\epsilon=\pm}\int \frac{d^3p}{2|p_0|}(\phi \mid p\bar{A};\epsilon)\epsilon^{2s}\hat{p}^{*ss}_{\dot{A}\dot{B}}\{p,\bar{B};\epsilon\mid\phi\}. \end{aligned}\tag{49}$$

We readily verify, using (38), (39), and (48), that the Γ_μ and $S_{\mu\nu}$ satisfy the commutation relations of the generators for a representation of $\overline{\mathrm{SO}_0(3,2)}_{\Gamma_\mu,S_{\mu\nu}}$. With the replacement $\Gamma_\mu \to -i\Gamma_\mu$ in (39), we obtain a representation of $\overline{\mathrm{SO}_0(4,1)}_{-i\Gamma_\mu,S_{\mu\nu}}$, which is, in general, defined in $\mathscr{L}_2(T_3)\otimes C^4_{(1)}\otimes C^4_{(2)}\otimes\cdots\otimes C^4_{(2s)}$.

Now we define a mapping U from $\overline{\mathrm{SO}_0(4,1)}$ into the set of all linear operators on $\mathscr{L}^2(T_3)\otimes C^4_{(1)}\otimes\cdots\otimes C^4_{(2s)}$ by

$$\begin{aligned}&\psi'(p;\xi'_1\cdots\xi'_{2s}) \\ &\quad = [U(A)\psi](p;\xi'_1\cdots\xi'_{2s}) \\ &\quad = (p;\xi'_1\cdots\xi'_{2s}\mid\mathscr{D}(A)\tilde{U}^0(A)\mid\psi) \\ &\quad = \sum_{\xi_1\cdots\xi_{2s}=1}^{4}\exp\left[-i\{\tfrac{1}{2}\omega^{\mu\nu}(S_{\mu\nu}) + i\omega^\rho(\Gamma_\rho)\}_{\xi'_i\xi_i}\right] \\ &\qquad\times(T^{\bar{A}}_{(\rho,s)}\psi)(p;\xi_1\cdots\xi_{2s}) \\ &\quad = \sum_{\xi_1\cdots\xi_{2s}=1}^{4}\mathscr{D}_{(1)\xi'_1\xi_1}(A)\mathscr{D}_{(2)\xi'_2\xi_2}(A)\cdots\mathscr{D}_{(2s)\xi'_{2s}\xi_{2s}}(A) \\ &\qquad\times(T^{\bar{A}}_{(\rho,s)}\psi)(p;\xi_1\cdots\xi_{2s}) \quad (\omega^\rho = \omega^{\bar{5}\rho}),\end{aligned}\tag{50}$$

where

$$\begin{aligned}&(T^{\bar{A}}_{(\rho,s)}\psi)(p;\xi_1\cdots\xi_{2s}) \\ &\quad = \frac{1}{|\mu(\bar{A};p/m)|^{3/2-s+ip}}\psi(m\bar{A}^{-1}p/m;\xi_1\cdots\xi_{2s}), \\ &(|\mu(\bar{A};p/m)| = |a^6{}_6(\bar{A}^{-1}) + \sum_\mu a^6{}_\mu(\bar{A}^{-1})p^\mu/m|),\end{aligned}\tag{51}$$

$$\mathscr{D}_{(i)\xi'_i\xi_i}(A) = \exp\left[-i(\tfrac{1}{2}\omega^{\mu\nu}S_{\mu\nu(i)} + i\omega^\rho\Gamma_{\rho(i)})_{\xi'_i\xi_i}\right].$$

We see from (50) that every permutation of the indices $\xi'_1\cdots\xi'_{2s}$ may be obtained by the same permutation of the factors $\mathscr{D}_{(1)\xi'_1\xi_1}(A),\ldots,\mathscr{D}_{(2s)\xi'_{2s}\xi_{2s}}(A)$ of (50) and the indices $\xi_1\cdots\xi_{2s}$ in $\psi(p;\xi_1\cdots\xi_{2s})$. Therefore, if $\psi(p;\xi_1\cdots\xi_{2s})$ is symmetric under the interchange of the ξ_i's then so is $\psi'(p;\xi'_1\cdots\xi'_{2s})$. Thus the representation can be considered as a representation in $\mathscr{S}^{(s)}$.

On $\mathscr{R}^{(1/2)}$ the following operator identity is valid[27]:

$$i\Gamma_\mu = -(1/2m)\{P^\rho, S_{\rho\mu}\} + (i/2m)P_\mu.$$

Using (41), together with (39) and (48), we can obtain a generalization on $\mathscr{R}^{(s)}$

$$\begin{aligned}i\Gamma_\mu &= -(1/2m)\{P^\rho, S_{\rho\mu}\} + i[(2s)/2m]P_\mu \\ &= -(1/\lambda)B^{(s)}_\mu + i[(2s)/2m]P_\mu.\end{aligned}\tag{52}$$

(40) is an immediate consequence of (52). We will use this result shortly.

Now we determine the form of an $\overline{\mathrm{SO}_0(4,1)}$ transformation which leaves (41) invariant: for $A\in\overline{\mathrm{SO}_0(4,1)}$ let $U(A)$ be the operator defined by

$$U(A)\psi(p;\xi_i) = S(p,A)T^{\bar{A}}_{(\rho,s)}\psi(p;\xi'_i), \tag{53}$$

where $S(p,A)$ affects only the ξ_i variables, but may depend upon p, and $T^A_{(\rho,s)}\psi(p,\xi'_i)$ is defined by (51).

For A a de Sitter transformation we have

$$\frac{p^{\mu\prime}}{m} = \left(\bar{A}\left(\frac{p^\nu}{m}\right)\right)^\mu, \tag{54}$$

with $\bar{A}$ as in (31). The invariance of the Bargmann-Wigner equations means[28]

$$p'_\mu\gamma^\mu\psi'(p';\xi) = m\psi'(p';\xi') \tag{55}$$

with

$$\psi'(p';\xi) = U(A)\psi(p;\bar{\xi}) \tag{56}$$

as in (53). Using (54) and (56), we see that (55) is true if S satisfies

$$mS(p,A)\left(A^{-1}\left(\frac{p^\nu}{m}\right)\right)^\mu\gamma_{(k)_\mu}S^{-1}(p,A) = p^\mu\gamma_{(k)_\mu}. \tag{57}$$

For Λ an element of the Lorentz subgroup, the determination of the $S(\Lambda)$ which satisfy (57) is obtained by first noting that a solution of (57) for transformations involving only the p, ξ_k variables (k fixed) is achieved by the usual choice of the spin-$\frac{1}{2}$ Dirac theory case[29]:

$$\begin{aligned}&U_{(k)}(\Lambda)\psi(p;\xi_1\cdots\xi_k\cdots\xi_{2s}) \\ &\quad = \sum_{\xi'_k}\exp\left[-\tfrac{1}{2}i\omega^{\mu\nu}(S_{\mu\nu(k)})_{\xi_k\xi'_k}\right]\psi(\Lambda^{-1}p;\xi'_1\cdots\xi'_k\cdots\xi'_{2s})\end{aligned}\tag{58}$$

so that if we let $U(\Lambda)$ be the operator (50) with

$$\mathscr{D}_{(i)}(\Lambda) = \exp(-\tfrac{1}{2}i\omega^{\mu\nu}S_{\mu\nu(i)}),$$

then invariance of each of the equations of (41) is assured. Using (50) and the fact that the $S_{\mu\nu(k)}$'s for different k's commute with each other, we can rewrite (50) for this Λ as

$$\begin{aligned}&U(\Lambda)\psi(p;\xi_1\cdots\xi_{2s}) \\ &\quad = \sum_{\xi'_1\cdots\xi'_{2s}=1}^{4}\exp(-\tfrac{1}{2}i\{\omega^{\mu\nu}S_{\mu\nu}\}_{\xi_i\xi'_i})\psi(\Lambda^{-1}p;\xi'_1\cdots\xi'_{2s}).\end{aligned}\tag{59}$$

Now consider an infinitesimal de Sitter boost $S(A)$ in the $\bar{5}$–i plane. Let $S_{\bar{5}i}$ denote the infinitesimal generator $S(A)$, so that $S(A)\cong 1 - i\omega^{\bar{5}i}S_{\bar{5}i}$. Using this along with (31), we obtain from (57)

$$[P^\mu\gamma_{(k)_\mu}, S_{\bar{5}i}] = -i\left[(1/m)P_i\gamma^\mu_{(k)_i}P_\mu - m\gamma_{(k)_i}\right]. \tag{60}$$

Because of (52) a solution to this equation is provided by $+i\Gamma_i$ restricted to $\mathscr{R}^{(s)}$. In a similar manner we may establish for an infinitesimal de Sitter boost in the $\bar{5}$–0 plane that $S_{\bar{5}0} = +i\Gamma_0$ on $\mathscr{R}^{(s)}$ is a sufficient condition for invariance of (41). Using (50) and this result, we see that a sufficient

condition for the invariance of the Bargmann–Wigner equation (41) under infinitesimal de Sitter boosts is

$$U(A)\psi(p;\xi_1\cdots\xi_{2s}) = \sum_{\xi'_1\cdots\xi'_{2s}=1}^{4} \exp(-i\{\omega^\mu(+i\Gamma_\mu)\}_{\xi\xi'_i}) T^{\bar{A}}_{(\rho,s)}(p;\xi'_1\cdots\xi'_{2s}) \tag{61}$$

with $\omega^\mu = \omega^{\bar{5}\mu}$. By addition and compounding of transformations, we may establish that the desired form of the $\overline{SO_0(4,1)}$ transformation on $\mathscr{R}^{(s)}$ is identical to the operator (50) in the space $\mathscr{R}^{(s)}$.

Next we make the following definition[30]:

Definition: A unitary irreducible representation U of a Lie group $\mathfrak{G}$ on a Hilbert space $\mathfrak{H}$ is the range of a function U from $\mathfrak{G}$ into the space of all bounded linear operators on $\mathfrak{H}$ satisfying

(i) To every $A\in\mathfrak{G}$ corresponds a unitary operator $U(A)$ on $\mathfrak{H}$;

(ii) for any two elements $A_1, A_2\in\mathfrak{G}$, $U(A_2)U(A_1) = U(A_2A_1)$;

(iii) if the sequence $A_n\in\mathfrak{G}$ converges to A, then $U(A_n)\psi \to U(A)\psi$ for every $\psi\in\mathfrak{H}$;

(iv) no proper closed linear subspace of $\mathfrak{H}$ is invariant with respect to all the $U(A)$'s (irreducibility).

With this definition we can prove that (50) defines an irreducible unitary representation of $\overline{SO_0(4,1)}$ on $\mathscr{R}^{(s)}$. The linearity of $U(A)$ is obvious from (50). The homomorphism condition (ii) follows from the multiplication property of the multipliers

$$\mu(\bar{A}_2^{-1};p)\mu(\bar{A}_1^{-1};(\bar{A}_2^{-1}p)) = \mu((\overline{A_2A_1})^{-1};p) \tag{62}$$

so that from (50) and this equation

$$(U(A_2)(U(A_1)\psi))(p;\xi_1\cdots\xi_{2s}) = \sum_{\xi'_i} \mathscr{D}_{\xi\xi'_i}(A_2\cdot A_1) T^{\overline{A_2\cdot A_1}}_{(\rho,s)}(p;\xi'_1\cdots\xi'_{2s}) = U(A_2\cdot A_1)\psi(p;\xi_1\cdots\xi_{2s}), \tag{63}$$

using the representation property of the $S(A)$'s. Continuity [(iii)] is verified by checking that (iii) is satisfied for arbitrary matrix elements[31]: $A_n\to A$ implies $(\phi,U(A_n)\psi)\to(\phi,U(A)\psi)$ as complex numbers for any $\phi, \psi\in\mathscr{R}^{(s)}$. Using the definition (42) of the inner product, we see that this "matrix element" convergence is clear from the form (50) of the representation along with the fact that $\mathscr{D}(A,p)_{\xi'\xi}(T^{\bar{A}}_{(\rho,s)}\psi)(p,\xi)$ is a continuous function of the group parameters, ω_{ab} (recall that $T^{\bar{A}}_{(\rho,s)}$ is a continuous representation). We consider the notion of the convergence of sequences in $\overline{SO_0(4,1)}$ associated with the topology on this group which is obtained by viewing $\overline{SO_0(4,1)}$ as a surface embedded in R^{n^2}[32]; for the convergence of sequences in C we mean with respect to its usual topology.

We now compute the infinitesimal generators of the representation. Observe that because of (43) ψ can be written as

$$\psi = \begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix} \tag{64}$$

with $\psi_1\in\mathscr{H}(m,s,+)$ and $\psi_2\in\mathscr{H}(m,s,-)$. We have

$$(\psi,\psi) = (\psi_1,\psi_1)_+ + (\psi_2,\psi_2)_-,$$

where $(\ ,\ )_+$ and $(\ ,\ )_-$ are the inner product on $\mathscr{R}^{(s)}$ restricted to $\mathscr{H}(m,s,+)$ and $\mathscr{H}(m,s,-)$, respectively. A generalized basis for $\mathscr{H}(m,s,+)$ consists of the vectors $|ps_3+)$; and $|ps_3-)$ is a generalized orthonormal basis for $\mathscr{H}(m,s,-)$, where the $|ps_3\pm)$ are the canonical basis vectors. Elements of $\mathscr{H}(m,s,-)$ in $\mathscr{R}^{(s)}$ correspond to vectors ψ with $\psi_2 = 0$ and elements of $\mathscr{H}(m,s,-)$ in $\mathscr{R}^{(s)}$ are vectors ψ with $\psi_1 = 0$. Let A be an infinitesimal transformation; then, using (50),

$$U(A) = 1 - i\omega^{ab}L_{ab},$$

with

$$L_{\mu\nu} = M_{\mu\nu}\cdot I + S_{\mu\nu} \tag{65}$$

and

$$L_{\bar{5}\mu} = -\frac{1}{\lambda}B^{(0)}_\mu\cdot I + i\Gamma_\mu - i\frac{(2s)}{2m}P_\mu = -\frac{1}{\lambda}(B^{(0)}_\mu\cdot I + B^{(s)}_\mu) = -\frac{1}{\lambda}B_\mu\cdot I \tag{66}$$

being the generators of the representation.

Next we claim

$$(L_{ab}\phi,\psi) = (\phi,L_{ab}\psi) \tag{67}$$

for any ψ in the domain of L_{ab}. The polarization identity (42) and the unitarity of $T^{\bar{A}}_{\rho}$ as defined in (30) and (52) show that (67) amounts to verifying

$$\begin{aligned}&((1+i\epsilon S_{\mu\nu})\psi,(1+i\epsilon S_{\mu\nu})\psi) = (\psi,\psi) + O(\epsilon^2),\\ &((1+i\epsilon(1/\lambda)B^{(s)}_\mu)\psi,(1+i\epsilon(1/\lambda)B^{(s)}_\mu)\psi) = (\psi,\psi) + O(\epsilon^2).\end{aligned} \tag{68}$$

The first equation can be checked by observing that if both $a,b = 1,2,3$, then $S_{\mu\nu}$ is a Hermitian matrix and commutes with $\Pi_{i=1}^{2s}\gamma_i^0$. If either a or b is 0 and the other 1, 2, 3, S_{ab} is a skew Hermitian, but it anticommutes with $\Pi_{i=1}^{2s}\gamma_i^0$. These same arguments together with the Hermiticity of P_μ suffice to prove the second equation.

Now we show the $U(A)$'s are bounded operators on $\mathscr{R}^{(s)}$. This is easiest to see by introducing the following inner product on $\mathscr{R}^{(s)}$, which is equivalent to (42)[2]:

$$(\phi,\psi) = \int_{T_3} d\Omega_{T_3}\,|m/p_0|^{2s}\sum_{\xi}|(\phi,\psi)|^2 \quad [\,|(\phi,\psi)|^2 = \phi^{+}\phi\,]. \tag{69}$$

Any A can always be written as the product of rotations in the various a–b planes of projective space.[33] Therefore, it suffices to prove boundedness for the operators of rotations in the a–b planes. For $a, b = 0,1,2,3$ the proof is immediate, since they are just representatives of Lorentz rotations which are even unitary. For a de Sitter boost in any of the 5–μ planes we have according to (69) and (50)

$$\|U(A)\psi\| \leqslant \left\{\max_{a,b}\mathscr{D}_{ab}(A)\right\}\|T^{\bar{A}}_{(\rho,s)}\psi\| \quad [\,\|\cdot\| = (\cdot,\cdot)^{1/2}\,] \tag{70}$$

since $\mathscr{D}_{ab}(A)$ is just a finite-dimensional matrix. Therefore, we must show $\|T^{\bar{A}}_{(\rho,s)}\psi\| < k\,\|\psi\|$ for any $\psi\in\mathscr{R}^{(s)}$. This is done in Appendix B. Thus the operators of the representation are bounded ones. Since the representation U consists of bounded operators, we know that

$$\mathfrak{D} = \bigcap_{g\in SO_0(4,1)} \mathfrak{D}(U^*(q))$$

is dense in $\mathscr{R}^{(s)}$. [$U^*(q) = -\frac{1}{2}\omega^{ab}L_{ab}$ and $\mathfrak{D}(A)$ stands for the domain of an operator A.] Therefore, all of the operators of the Lie algebra $U^*(g)$ are symmetric operators in $\mathscr{R}^{(s)}$, and we have a symmetric representation of the Lie algebra of $SO_0(4,1)$ on the Gårding subspace of $\mathfrak{D}$.[34]

We will now show that the representation of the one-parameter groups of rotations in the a–b planes of projective space are uniformly bounded. *A family of operators $U(t)$, $t\in\mathbb{R}$, is said to be uniformly bounded on a Hilbert space $\mathfrak{H}$ if there exists a constant $\kappa > 0$, independent of t, such that $\|U(t)\psi\| \leqslant \kappa\|\psi\|$ for all $t\in\mathbb{R}$.* It is clear that the unitary one-parameter generators of the Lorentz rotations are uniformly bounded. For boosts in the 5–i planes of projective space we know that the $\mathscr{D}(A)$'s are all products of matrices having sine and cosine or hyperbolic sine and cosine functions as entries and their arguments are restricted to finite intervals. Therefore, all of the $\max_{a,b} |\mathscr{D}_{ab}(A)|$ are bounded by some real positive number. Combining this with (70) and (B1) shows that each one-parameter family of boosts in the 5–i plane of projective space is uniformly bounded. For the one-parameter family of boosts in the 5–0 plane of projective space, each $\mathscr{D}(A) = e^{-t\Gamma_0}$ satisfies

$$\max_{a,b} |\mathscr{D}_{ab}(A)| \leqslant e^{st}.$$

[Use (39) and—for simplicity—choose a representation in which Γ_0 is diagonal.] Using (70), (B2), and this result, we obtain

$$\|U(A)\psi\| \leqslant e^{st}e^{-st}\|\psi\| = \|\psi\|$$

for all t. Thus each of the one-parameter groups is represented by a uniformly bounded family of operators.

We have the following theorem whose proof is similar to the proof of Stone's theorem[35]:

Theorem: Let $U(t)$ be a strongly continuous, uniformly bounded one-parameter group of linear transformations on a Hilbert space $\mathfrak{H}$ such that the infinitesimal generator A of $U(t)$ is a symmetric operator on $\mathfrak{H}$. Then $U(t)$ is a one-parameter group of unitary transformations on $\mathfrak{H}$.

Proof: First, A is defined as that operator such that

$$\text{for } \psi \in \mathfrak{D}(A), \quad iA\psi = \lim_{t\to 0} \frac{U(t)\psi - \psi}{t}.$$

In order to establish the theorem, we use the basic criterion for self-adjointness, namely, that $A^+\psi = \pm i\psi$ can have no solutions in $\mathfrak{H}$.[35] Suppose there is a $\psi \in \mathfrak{D}(A^+)$ so that $A^+\psi = i\psi$. Then, for each $\phi \in \mathfrak{D}(A)$,

$$\begin{aligned}\frac{d}{dt}(U(t)\phi,\psi) &= (iAU(t)\phi,\psi)\\ &= -i(U(t)\phi, A^+\psi)\\ &= -i(U(t)\phi, i\psi)\\ &= (U(t)\phi,\psi).\end{aligned}$$

The complex-valued function $f(t) = (U(t)\phi,\psi)$ satisfies the ordinary differential equation $f' = f$. Its solution is $f(t) = f(0)e^t$. Since $U(t)$ is uniformly bounded, $\|U(t)\phi\| < \kappa$, so $|f(t)| < \|U(t)\phi\|\,\|\psi\| \leqslant \kappa\|\psi\|\|\psi\|$ (by Schwartz's inequality), which implies that $f(0) = (\phi,\psi) = 0$. Because $\mathfrak{D}(A)$ is dense in $\mathfrak{H}$, $\psi = 0$. Likewise we may show $A^+\psi = -i\psi$ has no solutions in $\mathfrak{H}$. Therefore, A is essentially self-adjoint on $\mathfrak{D}(A)$. But then by the converse of Stone's theorem,[35] for each $t\in\mathbb{R}$, $\tilde{U}(t) = e^{i\bar{A}t}$ is a strongly continuous unitary operator ($\bar{A}$ is the self-adjoint closure of A). Since $\tilde{U}(t)$ and $U(t)$ agree on a dense domain and are both bounded, they must be equal. QED

For each one-parameter group of rotations in the a–b plane, let $U_k(t)$ denote the representative determined by (51). It satisfies the hypotheses of the theorem, so it is a one-parameter unitary group. Any $U(A)$ can be written as a product of the $U_k(t)$'s, and the reader can easily convince himself that the product of unitary operators is unitary. This proves condition (i). Irreducibility can be proven by demonstrating the unitary equivalence of this representation with another one which is shown to be irreducible. We do this in Appendix A.

If we choose

$$\rho = \sqrt{m^2/\lambda^2} \tag{71}$$

in (37), we obtain for the $(1/\lambda)B_\mu$'s in (66)

$$(1/\lambda)B_\mu = (1/\lambda)\left[P_\mu + (\lambda/2m)\{P^\rho, L_{\rho_\mu}\}\right]. \tag{72}$$

From this we conclude that the infinitesimal generators acting on an arbitrary vector $\psi = \binom{\psi_1}{\psi_2}$ in $\mathscr{R}^{(s)}$, which is in their common invariant domain, may be written as the matrix operators

$$\begin{aligned}&L_{\mu\nu}\times I,\\ &-(1/\lambda)B_\mu \times I,\end{aligned} \tag{73}$$

where $L_{\mu\nu}$ and B_μ are given by (65) and (72), respectively.

From (73) it seems that the subspaces $\mathscr{H}(m,s,+)$ and $\mathscr{H}(m,s,-)$ are invariant under the entire group (exponentiate). This is not so! We show in the appendix that the $e^{-i\omega\{(1/\lambda)B_i\}}$, in fact, do not leave $\mathscr{H}(m,s,+)$ and $\mathscr{H}(m,s,-)$ invariant.[36]

Finally we shall carry this representation of $\overline{SO_0(4,1)}$ over into $\overline{\mathscr{R}}^{(s)} = \mathscr{H}(m,s,+)\oplus\mathscr{H}(m,s,+)$. For this purpose we introduce the linear operator $\theta:\mathscr{R}^{(s)}\to\overline{\mathscr{R}}^{(s)}$ defined by its action on the canonical basis:

$$\theta\begin{pmatrix}|ps_3+)\\ |ps_3-)\end{pmatrix} = \begin{pmatrix}|ps_3+)\\ |-ps_3+)\end{pmatrix}. \tag{74}$$

This mapping is readily demonstrated to be unitary as an operator from $\mathscr{R}^{(s)}$ to $\overline{\mathscr{R}}^{(s)}$. From (74) it follows that

$$\overline{P}_\mu = \theta\begin{pmatrix}P_\mu & 0\\ 0 & P_\mu\end{pmatrix}\theta^{-1} = \begin{pmatrix}P_\mu & 0\\ 0 & -P_\mu\end{pmatrix}. \tag{75}$$

The form of the generators of the Lorentz subgroup acting on wave functions $\phi(ps_3;\pm) = (ps_3;\pm|\phi)$ are given by[37]

$$L_{jk} = -i\left(P_j\frac{\partial}{\partial p_k} - P_k\frac{\partial}{\partial p_j}\right) + S_{jk},$$

$$L_{0i} = N_i = -iP^0\frac{\partial}{\partial p_i} - [(\mathbf{P}\times\mathbf{S})_i/P^0 + \epsilon m]$$

$(S_i = \epsilon_{ijk}S_{jk})$.

From these equations and (74) it follows that

$$\overline{L}_{\mu\nu} = \theta\begin{pmatrix}L_{\mu\nu} & 0\\ 0 & L_{\mu\nu}\end{pmatrix}\theta^{-1} = \begin{pmatrix}L_{\mu\nu} & 0\\ 0 & L_{\mu\nu}\end{pmatrix}. \tag{76}$$

Using (72),

$$\frac{1}{\lambda}\overline{B}_\mu = \theta\frac{1}{\lambda}\begin{pmatrix} B_\mu & 0 \\ 0 & B_\mu \end{pmatrix}\theta^{-1} = \frac{1}{\lambda}\begin{pmatrix} B_\mu & 0 \\ 0 & -B_\mu \end{pmatrix}. \quad (77)$$

Because of the unitarity of θ, the representation $U(A)$ of $SO_0(4,1)$ on $\mathscr{R}^{(s)}$ and $\overline{U}(A) = \theta U(A)\theta^{-1}$ on $\overline{\mathscr{R}}^{(s)}$ are equivalent.

Using (76) and (77) for $\overline{L}_{\mu\nu}$ and $\overline{B}_\mu$, we will find the following expressions for the Casimir operators (9) and (10) of $SO_0(4,1)$ on $\overline{\mathscr{R}}^{(s)}$:

$$\lambda^2 Q_1 = \{m^2 + \tfrac{9}{4}\lambda^2 - \lambda^2 s(s+1)\}\cdot I, \quad (78)$$

$$\lambda^2 Q_2 = \{(m^2 + \tfrac{1}{4}\lambda^2)s(s+1)\}\cdot I. \quad (79)$$

A comparison of these equations with (15) and (16) shows that the irreducible representation which we have constructed is

$$(\sqrt{m^2/\lambda^2}, s).$$

We have thus succeeded in proving the integrability of the representation of the Lie algebra of $\overline{SO_0(4,1)}$ defined on a dense subspace $\mathscr{D}$ of $\overline{\mathscr{R}}^{(s)}$ and which is generated by (76) and (77) with $L_{\mu\nu}$ and B_μ given by (65) and (72), a result first conjectured by Böhm.[8]

VI. CONCLUSIONS

We have constructed on $\overline{\mathscr{R}}^{(s)}$ $= \mathscr{H}(m,s,+)\oplus\mathscr{H}(m,s,+)$ the principal series UIR

$$(\sqrt{m^2/\lambda^2}, s)$$

of $\overline{SO_0(4,1)}$; the advantage of using $\overline{\mathscr{R}}^{(s)}$ instead of $\mathscr{R}^{(s)}$ is that negative energy states are avoided.

Besides providing a very simple derivation of the principal series UIR's of $\overline{SO_0(4,1)}$, the advantage of our method is that, at every stage in the construction, the geometrical and physical meaning is brought out. For example, (52) gives the following interpretation to the Dirac matrices, Γ_μ: $im\Gamma_\mu$ equals $\frac{1}{2}iP_\mu$ minus the "internal" or "spin" part, $(1/\lambda)B_\mu^{(s)}$, of the generators of the de Sitter boosts, on any physical system whose states are described by solutions of the Dirac equation. It also reveals an interesting but little-known $\overline{SO_0(4,1)}$ symmetry of the Dirac equation. The possibility of such a symmetry could have been seen from the fact that the $S_{\mu\nu}$ and $-i\Gamma_\mu$'s form an $\overline{SO_0(4,1)}_{-i\Gamma_\mu, S_{\mu\nu}}$. Supposing this to be the spin part of the symmetry group, the only possible choice for the oribtal part of the generators which are linear differential operators on momentum space are the $M_{\mu\nu}^{(0)}$ and $M_{5\mu}^{(0)}$'s$(\rho = 0)$ up to addition of quantities which commute with the Dirac equation. The fulfillment of the conditions for a principal series UIR $(\sqrt{m^2/\lambda^2}, s)$, as given in the definition, force us to (50) as the only possibility. The physical meaning of the de Sitter boosts is still unclear.

An interesting question is whether or not our method generalizes to an explicit construction for certain UIR's of $SO_0(n,1)$ and $SO(p,q)$. The generalization to $SO_0(n,1)$ would consist in replacing T_3 by T_{n-1} and using higher dimensional spin structures. The generalization to $SO_0(p,q)$ would be in replacing T_3 by a hyperboloid $H_{(p-1,q)}$ in $M_{p-1,q}$: $H_{(p-1,q)} = \{p_\mu | \Sigma_{i=1}^{q}(p_i)^2 - \Sigma_{i=q}^{p+q}(p_i)^2\}$, and again using a different spin structure, that associated with the Clifford algebras, in $M_{p-1,q}$.[23] Also how would our construction compare to other methods of constructing representations of $SO_0(p,q)$, such as those which consider the decomposition of the regular representation acting on $\mathscr{L}^2(SO_0(p,q)/SO_0(p,q-1))$?[38]

Finally we comment on a possible physical consequence of the doubling of states. We may consider a physical system, the relativistic rotator model,[6] which consists in replacing $(P_\mu P^\mu)^{1/2}$ by $(\lambda^2 Q_1)^{1/2}$ as the relativistic Hamiltonian of an elementary particle, with mass m and spin s. A possible complete set of (esa) commuting operators[39] for this system is

$$(\lambda^2 Q_1)^{1/2},\ W,\ \mathbf{L}^2 + \mathbf{B}^2,\ \mathbf{L}\cdot\mathbf{B},\ B_3,\ L_3.$$

The space on which these operators act irreducibly cannot be $\mathscr{H}(m,s,+)$ or $\mathscr{H}(m,s,-)$ because B_3 is not essentially self-adjoint on these spaces. From the results of the preceding sections, we see that the Hilbert space can be chosen to be $\mathscr{H} = \mathscr{H}(m,s,+)\oplus\mathscr{H}(m,s,+)$. As another complete set of commuting observables we may take

$$P_\mu P^\mu,\ W,\ W_3,\ P_\mu,\ \Sigma$$

which correspond to the usual observables of an elementary particle, together with an operator Σ which distinguishes between the two $\mathscr{H}(m,s,+)$'s. The symmetry group or "dynamical group" of this model is $\overline{SO_0(4,1)}$. It is possible, at least for baryons, to identify Σ with the operator of electric charge.[7]

The question of the observability of B_3 is similar to the question of whether a precise description of the *zitterbewegung* in the Dirac theory of the electron in terms of "microscopic" dynamical variables is possible. The general viewpoint is that such an attempt is somewhat irrelevant. For when the Dirac equation is being interpreted at the one-particle level, the only Hermitian operators which can represent observable quantities are those which leave invariant the spaces of positive and negative energy solutions of the equation.[40] The position operator x_i and the B_i's do not satisfy this criterion, and so it would seem, according to the general view, they are only unobservable mathematical curiosities. However, at least for x_i, it can be argued that there are observable effects associated with it.[40] For example, it is $A_\mu(x)$ that appears in the formula of minimal substitution when electromagnetic interactions are brought in. Also it is x_i together with t which form a 4-vector, a fact of great importance in any attempt at a completely relativistic theory.[41] Perhaps the B_μ's are quantities associated with a new "microscopic" geometry of the Dirac theory and can be related to some observable effect.

ACKNOWLEDGMENTS

The author thanks Professor A. Böhm who suggested this problem and provided useful advice in many aspects of the research. He thanks Dr. B. Bourgeois and Professor J. E. Gilbert, both of whom provided much appreciated criticism of some very important parts of the paper. Discussions with Professor M. Gadella, and Professor P. Tomas were very

helpful. I thank Professor E. C. G. Sudarshan for pointing out to me the importance of E. Cartan's work in understanding the connection between representation theory and wave equations. This work was supported in part by the U. S. Department of Energy, Contract #DE-AS05-76ER03992.

APPENDIX A: PROOF OF IRREDUCIBILITY OF THE REPRESENTATION OF $\overline{SO_0(4,1)}$ on $\mathscr{R}^{(s)}$

First we construct a representation of $SO_0(4,1)$ in the Hilbert space of $\mathscr{L}^2$-functions on the unit sphere:

$$S_3 = \{\hat{u}\in\mathbb{R}^4 | \hat{u}^\mu \hat{u}_\mu = \hat{u}_0^2 + \hat{u}_1^2 + \hat{u}_2^2 + \hat{u}_3^2 = 1\} \quad \text{in } \mathbb{R}^4.$$

Analogous to the discussion in Sec. III, we may describe the projective transformations of $\mathbb{R}_4$ which leave invariant S_3, it is again an O(4,1).[18] The formula of the projective transformation is the same as in (25) except that the indefinite metric is no longer used. The representation on $\mathscr{L}^2(S_3)$ is defined just as in (30) and (31) (with $\bar{\xi}$ replaced by $\hat{u}$), except this time the Jacobian of the transformation (26) never vanishes. Introducing $u = m\hat{u}$, the generators of rotations in the various planes of projective space can be calculated similarly as in (36) and (37), but now contraction of two operators is with respect to the four-dimensional Euclidean metric and $Q_\mu = i\partial/\partial u_\mu$. (Note that all of the de Sitter boosts are hyperbolic for this case.) We denote the generators by the same symbols as in (36) and (37) except with carets placed over them.

Next we define a representation of $\overline{SO_0(4,1)}$ on $\mathscr{L}_2(S_3)\otimes C^4_{\hat{1}}\otimes C^4_{\hat{2}}\otimes\cdots\otimes C^4_{\hat{2s}}$. We introduce the quantities $\hat{\gamma}_{\mu\hat{k}}$ which satisfy

$$\{\hat{\gamma}^a_{\hat{k}},\hat{\gamma}^b_{\hat{l}}\} = 2\delta^{ab}_{\hat{k}}\delta_{kl}, \tag{38'}$$

i.e., the generators of the Clifford algebra corresponding to the Riemannian space, $\mathbb{R}^4$.[23] We define $\hat{\Gamma}^\mu$ as in (39) except the $\gamma_{\mu\hat{k}}$'s in these equations are replaced by $\hat{\gamma}_{\mu\hat{k}}$'s. The $\hat{\Gamma}_\mu$'s and

$$S_{\mu\nu} = -\frac{i}{4}\sum_k [\hat{\gamma}_{\hat{k}_\mu},\hat{\gamma}_{\hat{k}_\nu}]$$

can all be chosen Hermitian (consider the representation $\hat{\gamma}_0 = \gamma_0$, $\hat{\gamma}_i = \alpha_i$). The $\hat{S}_{\mu\nu}$ satisfy the commutation relations of the Lie algebra of $\overline{SO_0(4)}_{\hat{S}_{\mu\nu}}$ and the $\hat{S}_{\mu\nu}$ along with $-i\hat{\Gamma}_\mu$ give an $\overline{SO_0(4,1)}_{\hat{S}_{\mu\nu},-i\hat{\Gamma}_\mu}$. We have

$$(1/\lambda)\hat{B}^{(s)}_\mu = (1/2m)\{\hat{P}^\rho,\hat{S}_{\rho\mu}\} = i\hat{\Gamma}_\mu - i[(2s)/2m]\hat{P}_\mu \tag{52'}$$

as the analog of (52). It is needed in order to prove invariance of the representation defined below.

Now we consider the Hilbert space completion of the set [inner product is given by (42')] of all completely symmetric wavefunctions $\psi(u;\xi_1\cdots\xi_{2s})$ $(\xi_k\in C^4_{\hat{k}})$ which satisfy the Bargmann–Wigner equations $(\hat{\gamma}^\mu_{\hat{k}}\hat{p}_\mu - m)\psi = 0$ $(k = 1,\ldots,2s)$. Denote this space by $\hat{\mathscr{R}}^{(s)}$. Define the representation $\hat{U}$ of $\overline{SO_0(4,1)}$ by (50) except replace everywhere p by u and ξ by ζ,

$$[\hat{U}(\hat{A})\psi](u;\zeta'_1\cdots\zeta'_{2s}) = \sum \exp\left[-i(\tfrac{1}{2}\hat{\omega}^{\mu\nu}\hat{S}_{\mu\nu} + i\hat{\omega}^\rho\hat{\Gamma}_\rho)_{\zeta'\zeta}\right] \times T^{\bar{A}}_{(\rho,s)}\psi(u;\zeta_1\cdots\zeta_{2s}) \quad (\hat{\omega}^\mu = \hat{\omega}^{5\mu}). \tag{A1}$$

We verify, in the same way as for $\mathscr{R}^{(s)}$, that $\hat{\mathscr{R}}^{(s)}$ is invariant under the action of this representation. Since, for this case, the multiplier is a continuous never-vanishing function on the compact set S_3,[18] it obtains a nonzero minimum and the representation is easily shown to be bounded with respect to the inner product

$$\langle\phi,\psi\rangle = \int_{S_3}\left(\sum_\zeta \phi^*\psi\right)d\Omega_{S_3}. \tag{42'}$$

Unitarity is proved by using the theorem and also Eq. (52'). We can easily verify the $\hat{U}(\hat{A})$'s satisfy the homomorphism property of a representation. Therefore, this defines a representation on $\hat{\mathscr{R}}^{(s)}$ which satisfies conditions (i)–(iii) of the definition.

At this point we must construct the positive mass UIR's of the four-dimensional Euclidean group (actually its covering group), $\overline{E}(4)$, analogous to Wigner's construction of the UIR's of $\overline{\mathscr{P}}$ by Frobenius' method of induced representations. We introduce the canonical basis of vectors $|u, s_3\, s\rangle$ defined as[13]

$$|u_3,s_3s\rangle = \hat{U}^{-1}(\hat{L}(u))|(m,\mathbf{0})s_3s\rangle = \hat{U}^{-1}(\hat{L}(u))||(m,\mathbf{0})\rangle\otimes|s_3s\rangle\rangle,$$

where $\hat{L}^{-1}(u)$ is the compact analog of the Lorentz boost, i.e., $\hat{L}^{-1}(u)$:$(m,\mathbf{0})\to(u_0,\mathbf{u})\in S_3$. Proceeding as in Ref. 13, we obtain the desired form of the representation of $\overline{E}(4)$: Let $(\hat{\Lambda},\hat{a})\in\overline{E}(4)$; then

$$\hat{U}(\hat{\Lambda},\hat{a})|us_3s\rangle = e^{iu\cdot\hat{a}}\sum_{s_3}|(\hat{\Lambda}u)s'_3s\rangle\mathscr{D}_{s'_3s_3}(\hat{\mathscr{R}}),$$

where

$$\hat{\mathscr{R}} = \hat{L}(\hat{\Lambda}u)\hat{\Lambda}\hat{L}^{-1}(u) = \hat{R}(\hat{\Lambda},u)$$

is a rotation in the three-dimensional spaces of $u_0 = \text{const}$ and

$$\mathscr{D}_{s'_3s_3}(\hat{R}) = \exp\left[-i(\tfrac{1}{2}\omega^{ij}\hat{S}_{ij})_{s'_3s_3}\right],$$

ω_{ij} being the parameters of the rotation $\hat{\mathscr{R}}$. This defines a unitary representation on $\hat{\mathscr{H}}(m,s)$, the Hilbert space completion of the set of all vectors, ψ, in the linear span of the $|us_3s\rangle$'s whose components with respect to this generalized basis satisfy

$$\langle\psi,\psi\rangle = \sum_{s_3}\int_{S_3} d\Omega\ \psi^*(u;s_3)\psi(u,s_3) < \infty.$$

[$\langle\ ,\ \rangle$ denotes the inner product on $\hat{\mathscr{H}}(m,s)$ which makes the $|us_3s\rangle$'s "orthonormal."] The method of proof used in Ref. 22 to demonstrate the equivalence of the representations of $\overline{\mathscr{P}}$ on $\mathscr{H}(m,s,+)\oplus\mathscr{H}(m,s,-)$ and $\mathscr{R}^{(s)}$ can be applied in essentially the same way to demonstrate the unitary equivalence of this representation of $\overline{E}(4)$ on $\hat{\mathscr{H}}(m,s)$ and the one defined on $\hat{\mathscr{R}}^{(s)}$ as follows:

$$\hat{U}(\hat{\Lambda},\hat{a})\psi(u;\zeta_i')$$

$$= e^{-iu\cdot\hat{a}} \sum_{\zeta_i} \exp\left[-\tfrac{1}{2}i(\hat{\omega}^{\mu\nu}\hat{S}_{\mu\nu})_{\zeta_i'\zeta_i}\right]\psi(\hat{\Lambda}^{-1}u;\zeta_i),$$

where $\omega^{\mu\nu}$ are the parameters of the $\overline{\mathrm{SO}_0(4)}$ transformation $\hat{\Lambda}$.

Next we prove irreducibility of the representation of $\overline{\mathrm{SO}_0(4,1)}$ on $\hat{\mathscr{H}}(m,s)$. For this we must know the UIR's of $\overline{\mathrm{SO}_0(4)}$; they are characterized by two numbers $\pm j_0$ ($j_0>0$) and c, where $j_0=\frac{1}{2},\frac{3}{2},\cdots$ or, 0, 1, 2,$\cdots$ and $c=j_0+n$, $n=1,2,\cdots$.[42] We denote the characters of the UIR's by $X^{[c-1,j_0]}$.[43] We can introduce a basis in a UIR space $\mathscr{H}(\pm j_0,c)$ of $\overline{\mathrm{SO}_0(4)}$ which diagonalizes $\overline{\mathrm{SO}_0(3)}$; it is denoted by

$$|j_3 j;(\pm j_0,c)\rangle, \quad j_3=j,j-1,\ldots,-j, \quad j=j_0,j_0+1,\ldots,c-1. \tag{A2}$$

The values of the Casimir operators in $\mathscr{H}(\pm j_0,c)$ are

$$Q_1^{\mathrm{SO}(4)}|j_3 j;(\pm j_0,c)\rangle = (j_0^2+c^2-1)|j_3 j;(\pm j_0,c)\rangle, \tag{A3}$$

$$Q_2^{\mathrm{SO}(4)}|j_3 j;(\pm j_0,c)\rangle = \pm j_0 c|j_3 j;(\pm j_0,c)\rangle. \tag{A4}$$

To prove irreducibility, we will show that the Lie algebra of $\hat{U}(\overline{\mathrm{SO}_0(4,1)})$ satisfies the following conditions:

(i) There exists a dense subspace $\hat{\mathscr{D}}(m,s)$ of $\hat{\mathscr{H}}(m,s)$ which is invariant with respect to the Lie algebra of $\hat{U}(\overline{\mathrm{SO}_0(4,1)})$; and, furthermore, all of the generators $\hat{B}_\mu$, $\hat{L}_{\mu\nu}$ along with the Nelson operator

$$\mathfrak{N}=\sum_\mu \hat{B}_\mu\hat{B}_\mu+\frac{1}{2}\sum_{\mu,\nu}\hat{L}_{\mu\nu}\hat{L}_{\mu\nu}$$

are essentially self-adjoint on $\hat{\mathscr{D}}(m,s)$.

(ii) Every irreducible representation $\mathscr{H}(\pm j_0,c)$ with $j_0=s,s-1,\cdots\geqslant 0$ and $s+1\leqslant c$ of the Lie algebra of $\overline{\mathrm{SO}_0(4,1)}$ lies completely in $\hat{\mathscr{D}}(m,s)$.

(iii) Every irreducible representation space $\mathscr{H}(\pm j_0,c)$, if it occurs in $\hat{\mathscr{D}}(m,s)$, occurs at most once.

(iv) The vectors $|jj_3;(\pm j_0,c)\rangle$ form a complete orthonormal basis in $\hat{\mathscr{H}}(m,s)$.

(v) The generators $\hat{L}_{\mu\nu}$ acting on the basis vectors $|j_3 j;(\pm j_0,c)\rangle$ are given by

$$\hat{J}_q|j_3 j;(\pm j_0,c)\rangle$$

$$= -[j(j+1)]^{1/2}\langle 1qjj_3|1jj_3+q\rangle|j_3+q\,j(\pm j_0,c)\rangle,$$

$$i\hat{N}_q|j_3 j;(\pm j_0 c)\rangle$$

$$= \{[(j+1)^2-j_0^2][(j+1)^2-c^2]/(2j+3)(j+1)\}^{1/2}$$
$$\times\langle 1qjj_3|1jj+1j_3+q\rangle|j_3+q,j+1;(\pm j_0\,c_0\rangle$$
$$\pm i\{j_0 c/[j(j+1)]^{1/2}\}\langle 1qjj_3|1jjj_3+q\rangle|j_3+q,j;(\pm j_0 c)\rangle$$
$$-[(j^2-j_0^2)(j^2-c^2)/(2j-1)j]^{1/2}$$
$$\times\langle 1qjj_3|1jj-1j_3+q\rangle|j_3+q,j;(\pm j_0,c)\rangle.$$

[$q=0,\ \pm 1$ with $\hat{J}_0=\hat{L}_{12}$, $\hat{J}_\pm=+(1/\sqrt{2})(\hat{L}_{23}\pm i\hat{L}_{31})$ and $\hat{N}_0=\hat{L}_{30}$, $\hat{N}_\pm=\mp(1/\sqrt{2})(\hat{L}_{10}\pm i\hat{L}_{20})$]; and the generators $\hat{B}_\mu$ acting on these basis vectors are given as in Table II of Ref. 8. ($\langle|qjj_3|\ |jjj+q\rangle$ etc. are Clebsch–Gordan or vector coupling coefficients as given in Edmonds.[51]

That these conditions determine the UIR

$$(\sqrt{m^2/\lambda^2},s)$$

of $\overline{\mathrm{SO}_0(4,1)}$ can be seen from the decomposition of this UIR into UIR's of its maximal compact subgroup $\overline{\mathrm{SO}_0(4)}$ written formally as

$$\mathscr{H}(\sqrt{m^2/\lambda^2},s)$$

$$=\sum_{c>s+1}\ \sum_{\substack{j_0=0,1,\cdots\\ j_0=1/2,3/2,\cdots}}^{s}\oplus(\mathscr{H}(+j_0,c)\oplus\mathscr{H}(-j_0,c)), \tag{A5}$$

which we infer from the results of Ref. 8 or 16. Conditions (ii)–(iv) show that $\hat{\mathscr{R}}^{(s)}$ and $\mathscr{H}(\sqrt{m^2/\lambda^2},s)$ are equivalent as Hilbert spaces, and (i) and (v) show that the representation of $\overline{\mathrm{SO}_0(4)}$ on $\hat{\mathscr{R}}^{(s)}$ is in completely reduced form with the expression for the generators of $\overline{\mathrm{SO}_0(4,1)}$ being the same as in a UIR (cf. Ref. 8). Therefore, they determine equivalent representations.

Condition (i) is true because of Nelson's theorem.[44] In order to prove (ii)–(iv) we describe the $\overline{\mathrm{E}}(4)$ analog of Joos' results on the decomposition of a UIR of $\overline{\mathscr{P}}$ with respect to $\overline{\mathrm{SO}_0(3,1)}$,[45] i.e., the decomposition of a positive mass UIR of $\overline{\mathrm{E}}(4)$ with respect to $\overline{\mathrm{SO}_0(4)}$.

We consider the set of vectors $|j_3,j;(\pm j_0,c),\eta\rangle$ which reduce $\overline{\mathrm{SO}_0(4)}$:

$$|j_3,j;(\pm j_0,c),\eta\rangle$$

$$=\sum_{s_3}\int_{s_3} d\Omega\,|u,s_3;(M,s)\rangle\langle u,s_3;(M,s)|j_3,j;(\pm j_0,c)\eta\rangle \tag{A6}$$

(η is a parameter labeling the possible multiplicity).

For $s=0$, (A3) and (A4) give the following simultaneous eigenvalue equations for the transformation coefficients in (A6) ($j_0=0$)[46]:

$$\hat{J}_0(u|j_3,j;(0,c),\eta\rangle = j_3<u|j_3,j;(0,c),\eta\rangle, \tag{A7a}$$

$$\hat{J}^2\langle u|j_3,j;(0,c),\eta\rangle = j(j+1)\langle u|j_3,j;(0,c),\eta\rangle, \tag{A7b}$$

$$Q_1^{\mathrm{SO}(4)}\langle u|j_3,j;(0,c),\eta\rangle = (c^2-1)\langle u|j_3,j;(0,c),\eta\rangle, \tag{A7c}$$

where $Q_1^{\mathrm{SO}(4)}=\hat{J}^2+\hat{N}^2$.

Normalized solutions in $\mathscr{L}^2(S_3)$ ($=\hat{\mathscr{R}}^{(0)}$) of these equations are the four-dimensional spherical harmonics, $\langle u|j_3,j;(0,c)\rangle=\overline{Y}^{(\hat{u})}_{cjj_3}$, given by[47]

$$\overline{Y}^{(\hat{u})}_{cjj_3}=Y_{cjj_3}(u)=\frac{1}{M^{3/2}}\left[\frac{c(c+j)!}{(c-1-j)!}\right]^{1/2}\left(\frac{M}{|\mathbf{u}|}\right)^{1/2}$$
$$\times Y_j^{j_3}\left(\frac{\mathbf{u}}{|\mathbf{u}|}\right)P_{c-1/2}^{-(j+1/2)}\left(\frac{|u_0|}{M}\right),$$

$$c=1,2,\cdots,\quad j=0,1,\ldots,c-1,\quad j_3=-j,-j+1,\ldots,j, \tag{A8}$$

with $M^2=u_0+\mathbf{u}^2$ the four-dimensional sphere, S_3, $Y_j^{j_3}(\mathbf{u}/|\mathbf{u}|)$ denoting the three-dimensional spherical harmonic, and $P_{c-1/2}^{-(j+1/2)}(x)$ being a Legendre function having real parameters $c-\frac{1}{2}$ and $-(j+\frac{1}{2})$.[48] The normalization factor, necessary to ensure orthonormality of the Y_{cjj_3}'s,

$$N(c,j)=[c(c+j)!/M^3(c-1-j)!]^{1/2},$$

follows from the orthonormality properties of the three-dimensional spherical harmonics and the Legendre functions.

The general case ($s \neq 0$) is considerably more complicated. From the definition of the spinor base of $\overline{E}(4)$ [replace p by u in (44) and neglect the factor of ϵ], we obtain the following transformation coefficients on which we base our discussion[49]:

$$\{u, \underline{A};(M,s)|\, j_3, j;(\pm j_0, c), \eta\rangle$$
$$= \sum_{s_3} \mathscr{D}^{s}_{\Lambda s_3}(\hat{L}^{-1}(u))\langle u, s_3;(M,s)|\, j_3, j;(\pm j_0, c)\eta\rangle,$$
$$\{u, \overline{B};(M,s)|\, j_3, j;(\pm j_0, c), \eta\rangle$$
$$= \sum_{s_3} \mathscr{D}^{s^*}_{s_3 B}(\hat{L}(u))\langle u, s_3;(M,s)|\, j_3, j;(\pm j_0, c), \eta\rangle, \tag{A9}$$
$$[\,(\phi|u, A\,\}^* = \{u, \underline{A}\,|\phi\,) \quad \text{and} \quad \mathscr{D}^{s\dagger}_{\Lambda s_3}(\Lambda)$$
$$= \mathscr{D}^{s}_{\Lambda s_3}(\Lambda^{-1}) \text{ for } \Lambda \in \overline{SO_0(4)}\,].$$

In order to determine the coefficients

$$\{u, \underline{A};(M,s)|\, j_3, j;(\pm j_0, c), \eta\rangle,$$

we seek the matrix which reduces the Kronecker product of two UIR's of SO(4), i.e., the product of a self-conjugate one and $X^{[s,s]}$

$$X^{[\bar{c}-1,0]} \otimes X^{[s,s]}. \tag{A10}$$

The result obtained by Biedenharn is[50]

$$|j_3, j;(\pm j_0, c)\rangle$$
$$= \sum_{\substack{m_r, r, \bar{c} \\ \underline{A}, s}} \mathbf{Y}_{\bar{c}, r, m_r} \otimes |s, \underline{A}\,\rangle [(c \pm j_0)(c \mp j_0)(2r+1)(2s+1)]^{1/2}$$
$$\times \langle s\underline{A} r m_r | s r j j_3\rangle \times \begin{Bmatrix} s & \frac{1}{2}(\bar{c}-1) & \frac{1}{2}(c \pm j_0 - 1) \\ 0 & \frac{1}{2}(\bar{c}-1) & \frac{1}{2}(c \mp j_0 - 1) \\ s & r & j \end{Bmatrix}, \tag{A11}$$

where $\{\cdots\}$ represents Wigner's (9-j) coefficients—or Fano's X coefficient[51]—, $\mathbf{Y}_{\bar{c}, r, m_r}$ is the vector in $\mathscr{L}^2(S_3)$ corresponding to the spherical harmonic, $Y_{\bar{c} r m_r}(u)$, and $|s, \underline{A}\,\rangle$ represents the standard basis vectors in the $(2s+1)$-dimensional UIR of $\overline{SO_0(4)}$, $X^{[s,s]}$. We have no degeneracy and the multiplicity parameter η has been dropped. $\{u, \underline{A};(M,s)|$ $= \langle u| \otimes \langle s, \underline{A}\,|$,[52] so that using (A11), we obtain for the transformation coefficients

$$\{u, \underline{A};(M,s)|\, j_3, j;(\pm j_0, c)\rangle$$
$$= \sum_{\bar{c}, r, m_r} [(c \pm j_0)(c \mp j_0)(2r+1)(2s+1)]^{1/2} Y_{\bar{c} r m_r}(u)$$
$$\times \langle s\underline{A} r m_r | s r j j_3\rangle \begin{Bmatrix} s & \frac{1}{2}(\bar{c}-1) & \frac{1}{2}(c \pm j_0 - 1) \\ 0 & \frac{1}{2}(\bar{c}-1) & \frac{1}{2}(c \mp j_0 - 1) \\ s & r & j \end{Bmatrix}. \tag{A12}$$

[In commuting this we have used

$$\mathbf{Y}_{\bar{c} r m_r} \otimes |s, \underline{A}\,\rangle = \int_{S_3} d\Omega'_{S_3}\, Y_{\bar{c} r m_r}(u')|u'\rangle \otimes |s\underline{A}\,\rangle$$

together with $|u', \underline{A};(M,s)\} = |u'\rangle \otimes |s\underline{A}\,\rangle$ and also $\{u, \underline{A}\,';$ $(M,s)|u', \underline{A};(M,s)\} = \delta_{S_3}(u,u')\,\delta_{\underline{A}'\underline{A}}$.[53]]

Next, from the defining properties of the $|j_3, j;(\pm j_0, c)\rangle$ and the spinor basis we obtain the following equations:

$$\hat{J}_q\{u|\, j_3, j\rangle = -[j(j+1)]^{1/2}\langle 1q\, j j_3 | 1 j j_3 + q\rangle\{u|\, j_3 + q, j\rangle, \tag{A13a}$$

$$i\hat{N}_q\{u|\, j_3, j\rangle$$
$$= \{[(j+1)^2 - j_0^2][(j+1)^2 - c^2]/(2j+3)(j+1)\}^{1/2}$$
$$\times \langle 1q\, j j_3 | 1 j\, j+1\, j_3 + q\rangle\{u|\, j_3 + q, j+1\rangle$$
$$\pm i\{j_0 c/[j(j+1)]^{1/2}\}\langle 1q\, j j_3 | 1 j j\, j_3 + q\rangle\{u|\, j_3 + q, j\rangle$$
$$- [(j^2 - j_0^2)(j^2 - c^2)/(2j-1)j]^{1/2}$$
$$\times \langle 1q\, j j_3 | 1 j j - 1\, j_3 + q\rangle\{u|\, j_3 + q, j\rangle, \tag{A13b}$$

$$(\hat{N}^2 + \hat{J}^2)\{u|\, j_3, j\rangle = (c^2 + j_0^2 - 1)\{u|\, j_3, j\rangle, \tag{A13c}$$

$$\mathbf{J}\cdot\hat{\mathbf{N}}\{u|\, j_3, j\rangle = \pm c j_0\{u|\, j_3, j\rangle, \tag{A13d}$$

$$\int_{S_3} d\Omega^{(u)}_{S_3}\{u, \underline{A}\,|\, j_3, j;(\pm j_0 c)\rangle^* \delta_{AB}\{u, \underline{B}\,|\, j_3', j';(\pm j_0 c)\rangle$$
$$= \delta_{j_3 j_3'}\delta_{j j'}\delta_{\pm j_0, \pm j_0'}\delta_{c c'}. \tag{A13e}$$

(For simplicity we have omitted the fixed parameters $M, s, c,$ j_0 and $\underline{A}$, s_3, except where necessary.)

From (A13c) and (A13d) and (A12) we obtain with $\hat{\mathbf{J}} = \hat{\mathbf{J}}_0 + \hat{\mathbf{S}}$, $\hat{\mathbf{N}} = \hat{\mathbf{N}}_0 + \hat{\mathbf{S}}$ or $\hat{\mathbf{J}} = \hat{\mathbf{J}}_0 + \hat{\mathbf{S}}$, $\hat{\mathbf{N}} = \hat{\mathbf{N}}_0 - \hat{\mathbf{S}}$[54]:

$$\{+\hat{\mathbf{N}}_0^2 + \hat{\mathbf{J}}_0^2 \pm 2(\hat{\mathbf{S}}^2 + \hat{\mathbf{S}}\cdot\hat{\mathbf{J}}_0 + \hat{\mathbf{S}}\cdot\hat{\mathbf{N}})\}\{u|\, j_3, j;(\pm j_0 c)\rangle$$
$$= (c^2 + j_0^2 - 1)\{u|\, j_3, j;(\pm j_0 c)\rangle,$$
$$\pm(\hat{\mathbf{S}}^2 + \hat{\mathbf{S}}\cdot\hat{\mathbf{J}}_0 + \hat{\mathbf{S}}\cdot\hat{\mathbf{N}}_0)\{u|\, j_3, j;(\pm j_0 c)\rangle$$
$$= \pm c j_0\{u|\, j_3, j;(\pm j_0, c)\rangle$$

(the $+$ sign comes from $\hat{\mathbf{N}} = \hat{\mathbf{N}}_0 + \hat{\mathbf{S}}$ and the $-$ sign comes from $\hat{\mathbf{N}} = \hat{\mathbf{N}}_0 - \hat{\mathbf{S}}$). Using these equations with (A7) gives

$$\bar{c} = +(c \mp j_0). \tag{A14}$$

(Since $c = j_0 + \eta$ and $\bar{c}$ must be positive, the solution with the minus sign has to be discarded.) Substitution of (A14) into (A12) gives[55]

$$\{u, \underline{A};(M,s)|\, j_3, j;(\pm j_0, c)\rangle$$
$$= (-1)^{2j_2 + j + s}\sum_{r m_r} [(c \pm j_0)(2r+1)]^{1/2} Y_{c \mp j_0 r m_r}(u)$$
$$\times \langle s, \underline{A} r m_r | s r j j_3\rangle \begin{Bmatrix} s & j_2 & j_1 \\ j_2 & j & r \end{Bmatrix} \tag{A15}$$

with $j_1 = \frac{1}{2}(c \pm j_0 - 1)$, $j_2 = -\frac{1}{2}(\pm j_0 - c + 1)$, and $\{\cdots\}$ is the 6-j symbol as defined in Edmonds.[51] The triple s, j_2, j_1 must satisfy the triangle inequality[56]:

$$|s - j_2| \leqslant j_1 \leqslant s + j_2 \quad \text{and} \quad s + j_2 + j_2 \text{ an integer},$$

from which we obtain the desired restrictions on j_0 and c:

$$s + 1 \leqslant c \quad \text{and} \quad j_0 \leqslant s. \tag{A16}$$

To demonstrate the completeness condition, we first observe, on the basis of unitarity of the representation $\hat{U}^s(\hat{\Lambda})$, the validity of the following equation[57]:

$$\sum_{j j_3} \{u, \underline{A}\,|\, j_3, j;(\pm j_0, c)\rangle\{u', \underline{A}\,'|\, j_3, j;(\pm j_0, c)\rangle^*$$
$$= \sum_{j j_3} \mathscr{D}^{s}_{AB}(\hat{L}_{u'})\{\hat{L}_{u'}^{-1}u, \underline{B}\,|\, j_3, j;(\pm j_0, c)\rangle$$
$$\times \{(M, \mathbf{0}), \underline{B}\,'|\, j_3, j;(\pm j_0, c)\rangle^* \mathscr{D}^{s+}_{B'A'}(\hat{L}_{\mu'}), \tag{A17}$$

where $\hat{L}_{u'}^{-1}$ is the inverse boost taking u' to rest. We have[58]

$$(\hat{L}_{u'}^{-1})_0^k = \hat{u}^{k'} \quad \text{and} \quad (\hat{L}_{u'}^{-1})_0^0 = \hat{u}_0', \tag{A18}$$

and we observe that[59]

$$\left(\frac{M}{|u|}\right)^{1/2} P_{c-1/2}^{-(r+1/2)}\left(\frac{|u_0'|}{M}\right) \approx \frac{2^{-(r+1)/2}}{(r+\frac{1}{2})!}\left(\frac{|u_0'|}{M}-1\right)^{r/2} \quad \text{for} \quad \frac{|u_0'|}{M}\to 1 \tag{A19}$$

so that, using (A15), (A8), (A18), and (A19), we obtain

$$\begin{aligned}
&\sum_{c,j_0}\sum_{\underline{j}j_3}\{u,\underline{B}\,|\,j_3,j;(\pm j_0,c)\rangle\langle(M,\mathbf{0}),\underline{B}'\,|\,j_3,j;(\pm j_0,c)\rangle^* \\
&\quad=\sum_{r=0}^{2s}(2r+1)^{-1/2}\Bigg(\sum_{c,\pm j_0}\frac{(c\pm j_0)(c\mp j_0)}{\sqrt{2}\pi M^3} \\
&\qquad\times\left[\frac{(c\mp j_0)(c\mp j_0+r)!}{(c\mp j_0-1-r)!}\right]^{1/2}(1-\hat{u}_0)^{-1/4} \\
&\qquad\times P_{c\mp j_0-1/2}^{-(r+1/2)}(\hat{u}_0)(2r+1)\begin{Bmatrix} s & j_2 & j_1\\ j_2 & s & r\end{Bmatrix}\begin{Bmatrix} s & j_2 & j_1\\ j_2 & s & 0\end{Bmatrix}\Bigg) \\
&\qquad\times\langle sBrm_r\,|srsB'\rangle\, Y_r^{m_r}(\{\hat{L}_{u'}^{-1}u\}_i) \\
&\quad=\sum_{r=0}^{2s}(\mathscr{F}(s,r;\hat{u}_0))\langle sBrm_r\,|srsB'\rangle\, Y_r^{m_r}(\{\hat{L}_{u'}^{-1}u\}_i),
\end{aligned} \tag{A20}$$

where $\mathscr{F}(s,r;\hat{u}_0)$ is the quantity in the parentheses. Substituting for the Legendre polynomial the following,[60]

$$P_{c\mp j_0-1/2}^{-(r+1/2)}(\hat{u}_0) = \frac{2^{r+1/2}r!(c\mp j_0-r-1)!}{\sqrt{\pi}(c\mp j_0+r)!}\times(1-u_0^2)^{1/4+r/2}C_{c\mp j_0-r-1}^{r+1}(\hat{u}_0),$$

where $C_{c\mp j_0-r-1}^{r+1}(\hat{u}_0)$ is a Gegenbauer polynomial, we obtain for the "radial part"

$$\begin{aligned}
F(s,r;\hat{u}_0) &= \sum_{j_2}\frac{2^r r!(2j_2+1)^{3/2}}{\sqrt{2\pi}\pi M^3}\times\left[\frac{(2j_2-r)!}{(2j_2+1+r)!}\right]^{1/2}(1-u_0^2)^{r/2}C_{2j_2}^{r+1}{}_r(\hat{u}_0) \\
&\quad\times\sum_{j_1}(2j+1)(2r+1)\begin{Bmatrix} j_2 & j_2 & r\\ s & s & j_1\end{Bmatrix}\begin{Bmatrix} s & j_2 & j_1\\ j_2 & s & 0\end{Bmatrix} \\
&= \frac{1}{\sqrt{2\pi}\pi M^3}\sum_{n=1}^{\infty} nC_{n-1}^1(\hat{u}_0)\delta_{r0}.
\end{aligned}$$

Using this result, (A20) becomes equal to the sum

$$\frac{1}{2\pi^2 M^3}\sum_{n=1}^{\infty} nC_{n-1}^1(\hat{u}_0), \tag{A21}$$

which must be interpreted in the sense of a distribution on $\hat{\mathscr{R}}^{(s)}$.

Denote spherical polar coordinates on S_3 by χ,θ,ϕ with $u_0 = M\cos\chi$, $u_1 = M\sin\chi\sin\theta\cos\phi$, $u_2 = M\sin\chi\sin\theta\sin\phi$, $u_3 = M\sin\chi\cos\theta$ and $d\Omega_{S_3} = M^3\sin^2\chi\times\sin\omega\, d\chi d\theta d\phi$. Let $N = (1,0,0,0)$ be the north pole on S_3. We wish to show (A21) is equal to $\delta_{S_3}(u,N)$, where $\delta_{S_3}(u_1,u_2)$ is the δ function on S_3 defined by

$$\int_{S_3} d\Omega_{S_3}\,\psi(u_1,\zeta_i)\delta_{S_3}(u_1,u_2) = \psi(u_2,\zeta_i).$$

Let $\psi(\chi,\mathbf{n},\zeta_i)$ [$\mathbf{n} = (\theta,\phi)$] be an element of $\hat{\mathscr{R}}^{(s)}$. We have

$$\begin{aligned}
&\int_{S_3} d\Omega_{S_3}\,\psi(\chi,\mathbf{n},\zeta_i)\frac{1}{2M^3\pi^2}\sum_{c=1}^{\infty} cC_{c-1}^1(\hat{u}_0) \\
&\quad=\frac{1}{2\pi}\sum_{c=1}^{\infty} c\int_{-1}^{1} d\hat{u}_0\int_{S_2} d\Omega_{S_2}\sqrt{1-\hat{u}_0^2}\,\psi(\hat{u}_0,\mathbf{n},\zeta_i)C_{c-1}^1(\hat{u}_0).
\end{aligned} \tag{A22}$$

Now let $\mathscr{L}_{\sigma_1}^2([-1,1])$ denote the space of all complex-valued functions, $f(\hat{u}_0)$, on the interval $[-1,1]$ such that

$$\int_{-1}^{1}|f(\hat{u}_0)|^2(1-\hat{u}_0^2)^{1/2}\,d\hat{u}_0 < \infty.$$

The Gegenbauer polynomials $C_{c-1}^1(\hat{u}_0)$ are complete in $\mathscr{L}_{\sigma_1}^2([-1,1])$, and we have the following expansion[61]: For $f(x)\in\mathscr{L}_{\sigma_1}^2([-1,1])$

$$f(x) = \sum_{l=0}^{\infty} b_l C_l^1(x) \tag{A23}$$

with

$$b_l = \frac{2}{\pi}\int_{-1}^{1} f(x)C_l^1(x)(1-x^2)^{1/2}\,dx. \tag{A24}$$

Also[62]

$$C_l^1(1) = l+1. \tag{A25}$$

For fixed $\mathbf{n}$ and ζ_i, $\psi(x,\mathbf{n},\zeta_i) = \psi_{\mathbf{n},\zeta_i}(\hat{u}_0)\in\mathscr{L}_{\sigma_1}^2[-1,1]$, so using (A24), (A25), and then (A23), we see that (A22) is $\psi(N,\zeta_i)$ [which stands for $\psi(N,\mathbf{n},\zeta_i)$, $\mathbf{n}$ arbitrary—the value $\psi(N,\mathbf{n},\zeta_i)$being the same for all $(\mathbf{n})$]. Thus (A21) is, in fact, equal to $\delta_{S_3}(u,N)$. On the grounds of rotational invariance [Eq. (A17)] we infer for arbitrary u' with $u'^2 = M^2$ the desired completeness relation

$$\begin{aligned}
&\sum_{c,\pm j_0}\sum_{\underline{j}j_3}\{u\underline{A}\,|\,j_3,j;(\pm j_0,c)\rangle\{u',\underline{A}'\,|\,j_3,j;(\pm j_1,c)\rangle^* \\
&\quad=\delta_{S_3}(u,u')\delta_{AA'}
\end{aligned} \tag{A26}$$

since

$$\mathscr{D}_{AB}^{s}(\hat{L}_{u'})\mathscr{D}_{BA'}^{s'}(\hat{L}_{u'}) = \delta_{AA'}$$

because of the unitarity of the $\mathscr{D}(\hat{A})$'s for $\hat{A}$ an element of the $\overline{\mathrm{SO}_0(4)}$ subgroup.

Finally, in order to establish irreducibility of $\hat{\mathscr{R}}^{(s)}$, condition (v) needs to be verified. The expressions for the generators $\hat{J}_q$ and $\hat{N}_q$ in (A13) are the same as in condition (v). To compute matrix elements of the generators $\hat{B}_\mu$ of the representation $\hat{\mathscr{R}}^{(s)}$ in the $|j,j_3;(c,\pm j_0)\rangle$ basis, we may use the commutation relations directly as is done in Ref. 15 or calculate first the matrix elements of $\hat{P}_\mu$ in the $|j,j_3;(c,\pm j_0)\rangle$ basis and then determine the matrix elements of the $\hat{B}_\mu$'s, using the matrix elements of $\hat{P}_\mu$ and $\hat{L}_{\mu\nu}$ and the equation[63,64]

$$\begin{aligned}
\hat{L}_{\mu\bar{5}} &= (1/\lambda)\hat{B}_\mu \\
&= (1/\lambda)\{\hat{P}_\mu + (i\lambda/2m)[\hat{P}_\mu, Q_1^{\mathrm{SO}(4)}]\} \quad [\lambda = \sqrt{(m/\rho)^2}].
\end{aligned}$$

For example, using Eq. (A7c) and this equation along with the P_0 of Ref. 64, we obtain the following upon replacing the

λ in Eq. (2.19) of Ref. 64 by $\pm ic$ and choosing the phases in their paper to be $\phi = \arctan[(c+\frac{1}{2})/i\rho]$ and $\phi' = \arctan\{-[(j_0-\frac{1}{2})/i\rho]$:

$$\begin{aligned}\hat{B}_0|j,j_3;(\pm j_0,c)\rangle &= \tfrac{1}{2}a(\pm j_0,c)\,[(j+c+1)(c-j)]^{1/2}|j,j_3;(\pm j_0,c+1)\rangle \\ &+ \tfrac{1}{2}b(\pm j_0,c)\,[(j\mp j_0)(j\pm j_0+1)]^{1/2}|j,j_3;(\pm j_0+1,c)\rangle \\ &- \tfrac{1}{2}c(\pm j_0,c)\,[(j\pm j_0)(j\mp j_0+1)]^{1/2}|j,j_3;(\pm j_0-1,c)\rangle \\ &+ \tfrac{1}{2}d(\pm j_0,c)\,[(c+j)(c-j-1)]^{1/2}|j,j_3;(\pm j_0,c-1)\rangle,\end{aligned}$$

where

$$a(\pm j_0,c) = -i\left[\frac{\{(s+\frac{1}{2})^2-(c+\frac{1}{2})^2\}\{-\rho^2-(c+\frac{1}{2})^2\}}{(j_0^2-c^2)[j_0^2-(c+1)^2]}\right]^{1/2}$$

$$b(\pm j_0,c) = +i\left[\frac{\{(s+\frac{1}{2})^2-(\pm j_0+\frac{1}{2})^2\}\{(\pm j_0+\frac{1}{2})^2+\rho^2\}}{(c^2-j_0^2)[c^2-(j_0+1)^2]}\right]^{1/2},$$

$$d(\pm j_0,c) = -a(\pm j_0,c-1),$$

$$c(\pm j_0,c) = b(\pm j_0-1,c).$$

Similarly, we may write down the matrix elements of $\hat{B}_i$ by the use of the above expression for $\hat{B}_i$ and also Eq. (2.20) of Ref. 64, which determines the matrix elements of $\hat{P}_i$ in the $\overline{SO_0(4)}$ basis.

The remainder of the Appendix will be devoted to the proof of the unitary equivalence of $\hat{\mathscr{R}}^{(s)}$ and $\mathscr{R}^{(s)}$. For this purpose we define a transformation τ which maps T_3 into S_3:

$$u\in S_3 \to p = \tau u \in T_3 : p = [m^2u_0^{-1},(-mu_i/u_0)], \tag{A27}$$

$$p\in T_3 \to u = \tau^{-1}p \in S_3 : u = [m^2p_0^{-1},(-mp_i/p_0)].$$

Let $u\to u' = \hat{\hat{A}}u$ be a conformal transformation of S_3: $\hat{\hat{A}}\in O(4,1)$ and

$$u_0' = (a_{00}u_0 + \Sigma a_{0j}u_j + a_{0t})/(a_{t0}u_0 + \Sigma a_{tj}u_j + a_{tt}), \tag{A28}$$

$$u_i' = (a_{i0}u_0 + \Sigma a_{ij}u_j + a_{it})/(a_{t0}u_0 + \Sigma a_{tj}u_j + a_{tt}),$$

with

$$\hat{\hat{A}} = \begin{vmatrix} a_{tt} & a_{t\nu} \\ a_{\mu t} & a_{\mu\nu}\end{vmatrix}. \tag{A29}$$

Under the mapping τ, we find that the corresponding point on T_3 undergoes the following transformation:

$$p_0' = (a_{tt}p_0 - \Sigma a_{tj}p_j + a_{t0})/(a_{0t}p_0 - \Sigma a_{0j}p_j + a_{00}), \tag{A30}$$

$$p_i' = (-a_{it}p_0 + \Sigma a_{ij}p_j - a_{i0})/(a_{0t}p_0 - \Sigma a_{0j}p_j - a_{00}).$$

We can write this projection as $\hat{\hat{A}}\to\bar{A}$ with

$$\bar{A} = \Lambda_0\hat{\hat{A}}\Lambda_0^{-1} = \begin{vmatrix} a_{tt} & -a_{tj} & a_{t0} \\ -a_{it} & a_{ij} & -a_{i0} \\ a_{0t} & -a_{0j} & a_{00}\end{vmatrix}, \tag{A31}$$

$$\Lambda_0 = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 0 & -I_3 \\ 0 & 1 & 0\end{vmatrix}.$$

We see that Λ_0 is itself an element of $SO_0(4,1)$.[65]

If $p\in\tau u$, using (A27), we find

$$d\Omega_{S_3} = |m/p_0|^3\,d\Omega_{T_3^+}. \tag{A32}$$

An alternative form for (42) is given by Eq. (69):

$$(\phi,\psi) = \int_{T_3'}|m/p_0|^{2s}\sum_{\xi}|(\phi,\psi)|^2\,d\Omega_{T_3^+}. \tag{A33}$$

Hence, if $\psi\in\hat{\mathscr{R}}^{(s)}$

$$\int_{S_3}\sum_{\xi}|\psi(u,\xi)|^2\,d\Omega_{S_3} = \int_{T_3}|m/p_0|^{2s}|p_0/m|^{2s-3}\sum_{\xi}|\psi(\tau^{-1}p,\xi)|^2\,d\Omega_{T_3^+} \tag{A34}$$

(for simplicity we assume the labels ζ and ξ refer to identical bases in their respective internal spaces, and we will use these labels interchangeably.) Using this equation, we define a unitary mapping $\Pi_{(\rho,s)}$ from $\hat{\mathscr{R}}^{(s)}$ to $\mathscr{R}^{(s)}$ by

$$\hat{\mathscr{R}}^{(s)}\ni\psi(u,\zeta)\to(\Pi_{(\rho,s)}\psi)(p,\xi) = \frac{\psi(\tau^{-1}p,\xi)}{|p_0/m|^{3/2-s+i\rho}}\in\mathscr{R}^{(s)}, \tag{A35}$$

$$\mathscr{R}^{(s)}\ni\phi(p,\xi)\to(\Pi_{(\rho,s)}^{-1}\phi)(u,\xi) = \frac{\phi(\tau u,\zeta)}{|u_0/m|^{3/2-s+i\rho}}\in\hat{\mathscr{R}}^{(s)}.$$

Next we choose the $\hat{\Gamma}_\mu$'s in terms of the Γ_μ's

$$\hat{\Gamma}^0{}_\kappa = \Gamma^0{}_\kappa, \quad \hat{\Gamma}^i{}_{(\kappa)} = -\gamma^0_{(\kappa)}\Gamma^i{}_{(\kappa)} = -iS^{0i} \tag{A36}$$

(other possible choices for the $\hat{\Gamma}_\mu$'s would be connected to this choice by a similarity transformation of the internal spaces of the ζ's and ξ's). We have then

$$\hat{S}_{ij} = S_{ij} \quad\text{and}\quad \hat{S}_{0i} = i\Gamma_i. \tag{A37}$$

We now prove that

$$\Pi_{(\rho,s)}\hat{U}(\hat{A}) = U(A)\Pi_{(\rho,s)}. \tag{A38}$$

First, because of (A31)

$$U(A) = U(A_0)U(\hat{A})U(A_0)^{-1}. \tag{A39}$$

Since all operators in (A38) and (A39) are unitary, substitution of (A39) into (A38) reveals that (A38) is a statement of unitary equivalence. To prove (A38), we consider

$$[\Pi_{(\rho,s)}\hat{U}(\hat{A})\psi](p,\xi) = \exp[-i(\tfrac{1}{2}\hat{\omega}^{\mu\nu}\hat{S}_{\mu\nu} - i\hat{\omega}^\rho\hat{\Gamma}_\rho)] \times\frac{\psi(\hat{\hat{A}}^{-1}\tau^{-1}p/m)}{[a_{tt}(\hat{\hat{A}}^{-1})p_0/m - \Sigma a_{ti}(\hat{\hat{A}}^{-1})p_i/m + a_{t0}(\hat{\hat{A}}^{-1})]^{3/2-s+i\rho}}. \tag{A40}$$

We also have

$$[U(A)\Pi_{(\rho,s)}\psi](p,\xi) = \exp[-i(\tfrac{1}{2}\omega^{\mu\nu}S_{\mu\nu} - i\omega^{\rho}\Gamma_{\rho})] \times \frac{\psi(\tau^{-1}\hat{A}^{-1}p/m)}{[a_{0t}(\bar{A}^{-1})p_0/m - \Sigma a_{0j}(\bar{A}^{-1})p_j/m + a_{00}(\bar{A}^{-1})]^{3/2-s+i\rho}} \times \left|\frac{a_{0t}(\bar{A}^{-1})p_0/m - \Sigma a_{0j}(\bar{A}^{-1})p_j/m + a_{00}(\bar{A}^{-1})}{a_{tt}(\bar{A}^{-1})p_0/m - \Sigma a_{tj}(\bar{A}^{-1})p_j/m + a_{t0}(\bar{A}^{-1})}\right|^{3/2-s+i\rho} \tag{A41}$$

By the definition of the automorphism τ, $\tau^{-1}\bar{A}^{-1} = \hat{\bar{A}}^{-1}\tau^{-1}$, so that Eq. (A38) will be established provided we can show

$$\exp[-i(\tfrac{1}{2}\hat{\omega}^{\mu\nu}\hat{S}_{\mu\nu} - i\hat{\omega}^{\rho}\hat{\Gamma}_{\rho})] = \exp[-i(\tfrac{1}{2}\omega^{\mu\nu}S_{\mu\nu} - i\omega^{\rho}\Gamma_{\rho})]\ . \tag{A42}$$

To prove Eq. (A42), it suffices to show that it is true for any one-parameter group of rotations in the a–b planes of projective space. First consider a rotation in the i–j plane. We must show

$$\hat{S}(e^{\hat{\omega}I^{ij}}) = S(e^{\omega I^{ij}}).$$

From (A31), $\omega = \hat{\omega}$, and, using (50) and (A1), we obtain

$$\hat{S}(e^{\hat{\omega}I^{ij}}) = e^{-i\hat{\omega}\hat{S}^{ij}}, \quad S(e^{\omega I^{ij}}) = e^{-i\hat{\omega}S^{ij}} = e^{-i\hat{\omega}\hat{S}^{ij}},$$

where the last equality follows from (A37). For a rotation in the 5–i plane of projective space we have

$$S(e^{\hat{\omega}I^{5i}}) = e^{-i\hat{\omega}\hat{S}^{0i}}$$

by (A1), and

$$S(e^{\omega I^{5i}}) = S(e^{-\hat{\omega}I^{5i}}) = e^{+i|\hat{\omega}|(-i\Gamma^{i})|}$$

using (A31) and (50) [$\omega^{i}(i\Gamma_{i}) = \omega^{5i}S_{5i}$ so $S_{5i} = i\Gamma_{i}$ and $S^{5i} = -i\Gamma^{i}$]. By (A37), we conclude from these two equations

$$\hat{S}(e^{\hat{\omega}I^{5i}}) = S(e^{\omega I^{5i}}).$$

For a rotation in the 0–i plane of projective space,

$$\hat{S}(e^{\hat{\omega}I^{0i}}) = e^{-i|\hat{\omega}|(+i\hat{\Gamma}^{i})|}$$

using (A1) and

$$S(e^{\omega I^{0i}}) = e^{-i|-\hat{\omega}|(S^{0i})|}.$$

By (A36), $i\hat{\Gamma}^{i} = S^{0i}$, so

$$\hat{S}(e^{\hat{\omega}I^{0i}}) = S(e^{\omega I^{0i}}).$$

Finally, for a rotation in the 5–0 plane of projective space

$$\hat{S}(e^{\hat{\omega}I^{50}}) = e^{-i|\hat{\omega}\hat{S}^{05}|} = e^{-i|\hat{\omega}(-i\hat{\Gamma}_{0})|} = e^{-i|\hat{\omega}(-i\Gamma_{0})|} = S(e^{\omega I^{50}})$$

using (A31), (50), (A1), and (A36), which completes the proof.

Now we can understand why the $\exp[-i\omega(1/\lambda)B_{i}$'s] defined in (50) do not leave $\mathscr{H}(m,s,+)$ and $\mathscr{H}(m,s,-)$ invariant. Consider $U(\exp(\overline{A}_{5i}))$, the transformation in $\mathscr{R}^{(s)}$ corresponding to a rotation in the i–5 plane of the projective space defined in Sec. III. Comparing (A28) and (A30) we see this is just an ordinary rotation of S_3 in the i–0 plane of u-space. But depending upon ω, points in the upper hemisphere may move to points in the lower hemisphere and the other way around. With the aid of the transformation τ, the claim follows from this remark.[66]

APPENDIX B: PROOF OF $\|T^{\bar{A}}_{(\rho,s)}\psi\| < k\|\psi\|$ FOR A A DE SITTER BOOST

For a de Sitter boost in the 5–i ($i = 1,2,3$) plane:

$$\|T^{\bar{A}}_{(\rho,s)}\psi\|^2 = (T^{\bar{A}}_{(\rho,s)}\psi, T^{\bar{A}}_{(\rho,s)}\psi) = \sum_{\xi}\int d\Omega_{T_3}\,|m/p_0|^{2s}\,[T^{\bar{A}}_{(\rho,s)}\psi]^{*}(p,\xi)[T^{\bar{A}}_{(\rho,s)}\psi](p,\xi)$$

and

$$\sum_{\xi}\int d\Omega_{T_3}\,|m/p_0|^{2s}\,[T^{\bar{A}}_{(\rho,s)}\psi]^{*}(p;\xi)[T^{\bar{A}}_{(\rho,s)}\psi](p;\xi)$$

$$= \int \sum_{\xi} d\Omega_{T_3}\,|m/p_0|^{2s} \times \frac{\psi^{*}(p';\xi)\psi(p';\xi)}{|\mu(\bar{A}^{-1};p/m)|^{3-2s}} \quad (p' = \bar{A}^{-1}p)$$

$$= \int \sum_{\xi} d\Omega'_{T_3}\,|m/p'_0|^{2s}\left|\frac{p'_0}{p_0}\right|^{2s} \times |\mu(\bar{A}^{-1};p/m)|^{2s}\psi^{*}(p';\xi)\psi(p';\xi) \quad [\text{by (27)}]$$

$$= \int \sum_{\xi} d\Omega'_{T_3}\,|m/p'_0|^{2s}\psi^{*}(p';\xi)\psi(p';\xi) = \|\psi\|^2. \tag{B1}$$

[In the second to last step we used (27) and in the last step we used (31).

For A a de Sitter boost in the 5–0 plane, we have

$$\|T^{\bar{A}}_{(\rho,s)}\psi\|^2 = \int d\Omega'_{T_3}\left|\frac{m}{p'_0}\right|^{2s}\left|\frac{a_{06}(\bar{A}^{-1}) + a_{00}(\bar{A}^{-1})p_0}{p_0}\right| \times \sum_{\xi}^{2s}\psi^{*}(p';\xi)\psi(p';\xi)$$

$$= \int d\Omega'_{T_3}\left|\frac{m}{p'_0}\right|^{2s}\left|\frac{a_{06}}{p_0} + a_{00}\right|^{2s} \times \sum_{\xi}\psi^{*}(p';\xi)\psi(p';\xi)$$

(since $p'_0 \neq 0$). Let $u_0 = m/p_0$, then $-1 < u_0 < 1$ and

$$[a_{06}u_0]^2 \leqslant a_{06}{}^2 = a_{00}{}^2 - 1.$$

Hence we have $|a_{00} + a_{06}u_0| \leqslant |a_{00}| + (a_{00}{}^2 - 1)^{1/2} = e^{-t}$ since $|a_{00}| = |\cosh(-t)|$; so

$$\|T^{\bar{A}}_{(\rho,s)}\psi\|^2 \leqslant e^{-2st}\|\psi\|^2. \tag{B2}$$

The proof follows from (B1) and (B2).

[1]P. A. M. Dirac, Proc. R. Soc. London A **322**, 435 (1971); **328**, 1 (1972).
[2]V. Bargmann and E. P. Wigner, Proc. Nat. Acad. Sci. USA **34**(5), 211–233 (1946).
[3]H. Joos, Fortschr. Phys. **10**, 65 (1962); Chou Kuang-Chao and L. G. Zastavenko, Zh. Eksp. Teor. Fiz. **35**, 1417 (1958) [Sov. Phys. JETP **8**, 990 (1959)].
[4]S. Ström, Ark. Fys. **40**, 1 (1969).
[5]M. Bander and C. Itzykson, Rev. Mod. Phys. **38**, 354 (1966).
[6]A. Böhm, in *Lectures in Theoretical Physics*, edited by W. E. Britton, A. O. Barut, and M. Guenin (Colorado U.P., Boulder, 1967), Vol. 9 B.
[7]A. Böhm, Phys. Rev. **175**(5), 1767–73 (1968); Phys. Rev. Lett. **23**(8), 436 (1969).
[8]A. Böhm, "Generalized Eigenvectors and Group Representations—The Connection between Representations of SO(4,1) and the Poincaré Group," in *Studies in Mathematical Physics*, edited by A. O. Barut (Reidel, New York, 1973).
[9]J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory* (Springer-Verlag, New York, 1972), p. 58.
[10]S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962), pp. 339–55.
[11]If steriographic projection coordinates are introduced in the de Sitter space-time V_4' of general relativity embedded in $M_{4,1}$, then the $A_{\mu\nu}$ generate Lorentz transformations which are physically the same as in the flat Minkowski space, whereas the rotations $A_{5\mu}$ generate the de Sitter boosts which go over into the translations in the contraction of SO(4,1) into the Poincaré group. (This identification of the $A_{\mu\nu}$ and $L_{\mu\nu}$—generators of Lorentz transformations—is possible through a "soldering" of $T_* V_4'$ with $T_* M_{3,1}$.) See W. Drechsler, Fortschr. Phys. **23**, 641 (1975).
[12]V. Bargmann, Ann. Math. **48**(3), 590 (1946).
[13]E. P. Wigner, *Istanbul Lectures* edited by F. Gürsey (Gordon and Breach, New York, 1964).
[14]M. A. Naimark, *Linear Representations of the Lorentz Group* (Pergamon, New York, 1964); A Böhm, *Quantum Mechanics* (Springer-Verlag, New York, 1979), pp. 145–56.
[15]L. H. Thomas, Ann. Math. **42**, 113 (1941).
[16]T. D. Newton, Ann. Math. **51**, 730 (1950); J. Dixmier, Bull. Soc. Math. France **89**, 9 (1961). [The clearest and most concise derivation known to me is given in Ref. 8; it uses the Gel'fand pattern. The representations of SO(n), $n = 2,3,4,5$, are assumed.]
[17]Ref. 8, p. 51.
[18]M. Bander and C. Itzykson, Rev. Mod. Phys. **38**, 330–45 (1966).
[19]M. Bander and C. Itzykson, Rev. Mod. Phys. **38**, 346–58 (1966).
[20]J. Haantjes, K. Ned. Akad. Wet. Proc. **40**, 700 (1937).
[21]Ref. 18, p. 341, and Ref. 19, p. 354.
[22]For various "representations" of the γ matrices, see A. Böhm and G. B. Mainland, Fortschr. Phys. **18**, 290–93 (1970).
[23]Nikos Salingaros, J. Math. Phys. **23**(1), 1 (1982).
[24]H. A. Kramers, F. J. Belifante, and J. K. Lubański, Physica **VIII**, 597–627 (1941).
[25]H. Joos, Fortschr. Phys. **10**, 75–83 (1962).
[26]P. A. M. Dirac, Proc. Roy. Soc. London A **155**, 447–59 (1936).
[27]R. P. Feynman, *Quantum Electrodynamics* (Benjamin, New York, 1962), p. 48; H. C. Corben, *Classical and Quantum Theories of Spinning Particles* (Holden-Day, San Francisco, 1968), p. 173.
[28]J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics* (McGraw-Hill, New York, 1964), p. 18.
[29]Ref. 28, p. 22.
[30]Ref. 12, p. 599.
[31]G. Warner, *Harmonic Analysis on Semi-Simple Lie Groups I* (Springer-Verlag, New York, 1972), p. 237. [We must assume the representation is bounded in order to establish (iii). We prove boundedness independent of (iii) later on.]
[32]J. E. Gilbert, Theory and Applications of Representations of Lie Groups, lecture notes (unpublished), Chap. I.
[33]N. J. Villenkin, *Special Functions and the Theory of Group Representations* (American Mathematical Society, Providence, RI, 1968), p. 438.
[34]L. Garding, Proc. Nat. Acad. Sci. USA **33**, 331–2 (1947).
[35]M. Reed and B. Simon, *Functional Analysis* (Academic, New York, 1972), Vol. 1, p. 266.
[36]Harish Chandra first pointed out this possibility of invariance of the Lie algebra under a subspace which is not dense in the representation space: Harish Chandra, Trans. Am. Math. Soc. **75**, 185 (1953).
[37]A. Chakrabarti, J. Math. Phys. **7**(3), 426 (1966); L. Fonda and G. C. Ghirardi, *Symmetry Principles in Quantum Physics* (Marcel Dekker, New York, 1970), Vol. 1, p.301.
[38]R. Strichartz, J. Funct. Anal. **12**(4), 341 (1973).
[39]A. Böhm, *Quantum Mechanics*, p. 125, Springer (1980); and Fonda and Ghirardi, Ref. 37, Chap. I.
[40]A. O. Barut and A. J. Bracken, Phys. Rev. D **23**(10), 2454 (1981).
[41]J. R. Fanchi, Phys. Rev. D **20**(12), 3108 (1979); L. P. Horowitz, C. Piron, and F. Reuse, Helv. Phys. Acta **48**, 546 (1975); L. P. Horowitz and Y. Rabin, UT CPT Preprint ORO 298, University of Texas, Center for Particle Theory, 1977; R. P. Feynman, Phys. Rev. **80**, 440 (1950).
[42]Ref. 8, p. 3, or Böhm, Ref. 39, p. 153.
[43]E. P. Wigner, *Group Theory* (Academic, New York, 1959), p. 83.
[44]E. Nelson, Ann. Math. **70**, 572–615 (1959).
[45]Ref. 25, pp. 92–104.
[46]Ref. 25, p. 92.
[47]Ref. 25, p. 94.
[48]A. Erdelyi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Chap. 3.
[49]Compare Ref 25, p. 99.
[50]L. C. Biedenharn, J. Math. Phys. **2**, 433 (1961).
[51]M. E. Rose, *Elementary Theory of Angular Momentum* (Wiley, New York, 1957); G. Racah and U. Fano, *Irreducible Tensorial Sets* (Academic, New York, 1959); A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U. P., Princeton, NJ, 1960).
[52]This form of the spinor basis could actually have served as its definition in (44). For the proof that this form and (44) are equivalent, see Ref. 22, pp. 285–304—there they consider only spin $\frac{1}{2}$, but the generalization to higher spins is immediately clear.
[53]For the analogous formula for nonunitary representations of $\overline{SO_0(3,1)}$ see Ref. 25, p. 81, Eq. (3.18″), or Ref. 22, p. 298, Eq. (88).
[54]Compare Ref. 25, p. 80, and Ref. 22, Eqs. (19) and (41).
[55]Edmonds, Ref. 51, p. 105–6, and also pp. 94–95 for symmetries of the 6-j symbol.
[56]J. D. Talman, *Special Functions: A Group Theoretical Approach* (Benjamin, New York, 1968), p. 122.
[57]Ref. 25, p. 103.
[58]Fonda and Ghirardi, Ref. 37, p. 299.
[59]Ref. 48, Chap. 3.
[60]N. J. Villenkin, *Special Functions and the Theory of Group Representations* (American Mathematical Society, Providence, RI, 1968), p. 486, Eq. (10).
[61]Villenkin, Ref. 60, p. 463.
[62]Villenkin, Ref.60, p. 461, Eq. (9).
[63]Ref. 8, p. 49.
[64]A. Chakrabarti, M. Levy-Nehas, and R. Seneor, J. Math. Phys. **9**(8), 1274 (1968).
[65]Ref. 18, p. 355.
[66]N. Mukunda, J. Math. Phys. **1**(9), 50 (1968) (in this paper the six-dimensional analog of our de Sitter group is examined, namely, the homogeneous Lorentz group).

General indices of simple Lie algebras and symmetrized product representations

S. Okubo
Department of Physics, University of Rochester, Rochester, New York 14627

J. Patera
Centre de recherche de mathématiques appliquées, Université de Montréal, Montréal, Québec, H3C 3J7, Canada

(Received 25 February 1983; accepted for publication 29 June 1983)

In many branches of physics, it is important to know the decomposition of a product representation $\rho \otimes \rho \otimes \cdots \otimes \rho$ (n times) of identical representations ρ of a simple Lie algebra into irreducible components with a given Young tableau symmetry. We show that the notion of representation indices introduced elsewhere is a very useful tool in dealing with this problem. We calculate explicit formula for general pth order indices $D^{(p)}(\rho)$ for all classical simple Lie algebras. Sixth-order indices for exceptional Lie algebras are also discussed.

PACS numbers: 02.20.Sv

1. INTRODUCTION

Let ψ be a one-particle wave function in any quantum-mechanical system. The wave function corresponding to n identical particles in the same state ψ will be described in terms of the tensor product

$$\psi \otimes \psi \otimes \cdots \otimes \psi \quad (n \text{ times}).$$

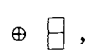
Because of the Pauli principle, we have to symmetrize or antisymmetrize these wave functions. In reality, the situation is more complicated in view of many internal degrees of freedom associated with some symmetry group inherent to the system. Let L be a Lie algebra and let ρ be a representation of L, which is not necessarily irreducible. We often encounter the following problem in many branches of physics. We consider the nth tensor power

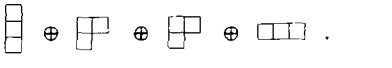
$$\rho \otimes \rho \otimes \cdots \otimes \rho \quad (n \text{ times}) \tag{1.1}$$

of a representation ρ, which will be simply denoted hereafter as ρ^n. Then we first decompose ρ^n into a sum of representations $\rho^n(\Gamma)$ possessing general permutation symmetry associated with Young tableau[1] Γ. We write this fact as

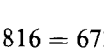
$$\rho^n = \sum_{\Gamma} \oplus \rho^n(\Gamma). \tag{1.2}$$

It is also convenient to denote the representation ρ by a single box $\square$. Then, for case $n = 2$ and 3, Eq. (1.2) is graphically depicted as

$\square \otimes \square =$  ,

$\square \otimes \square \otimes \square =$  .

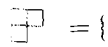
However, since this procedure is well known,[2] we will not dwell upon it. Moreover, if the Lie algebra L is of type A_{N-1} corresponding to the SU(N) group and if the representation ρ is the N-dimensional defining (or basic) representation of A_{N-1}, then $\rho^n(\Gamma)$ is irreducible.[1] Nevertheless, this is not generally true if L is not of type A_{N-1} and/or if ρ is not the defining representation of A_{N-1}. For example, let us consider the case of $L = D_5$ corresponding to the SO(10) group and of ρ being the 16-dimensional spinor representation with highest weight Λ_5. Then, we find

 $= \{3\Lambda_5\} \oplus \{\Lambda_1 + \Lambda_5\}$, (1.3a)

$816 = 672 + 144$,

 $= \{\Lambda_2 + \Lambda_4\}$, (1.3b)

$560 = 560$,

 $= \{\Lambda_3 + \Lambda_5\} \oplus \{\Lambda_1 + \Lambda_5\} \oplus \{\Lambda_4\}$, (1.3c)

$1360 = 1200 + 144 + \overline{16}$,

where $\{\Lambda\}$ stands for the irreducible representation of L with the highest weight Λ and where the dimension is shown under each representation. Hereafter we adopt the numbering of simple roots of L as in Ref. 3. Let us point out that we have chosen the example $L = D_5$ because of its relevance to the grand unified models based on the SO(10) group.[4] We return to this problem in Sec. 3.

The decomposition of $\rho^n(\Gamma)$ becomes rapidly more complicated if the rank of L gets larger and/or if the dimension of ρ is larger. In special cases efficient particular decomposition prescriptions can be found such as generating functions (cf. for instance Ref. 5). However, for a general L and ρ there are no efficient general methods. An ideal method would have to be general as to the type of the Lie algebra L, its representation ρ, and the symmetry Γ. Let us recall the known methods to the problem which has some degree of generality at least in one of the three quantities L, ρ, and Γ involved.

The method of Ref. 5 provides the decomposition of $\rho^n(\Gamma)$ for all Γ and G compatible with the requirement that $\dim \rho \leqslant 5$. In each particular case the decomposition is given by the coefficient of the corresponding term in the power series expansion of a generating function.

The method of Ref. 6 is general as to the choice of ρ and relatively easily extendable by recursion calculations to any Γ of interest. However, the Lie algebra L is only that of SU(2). It appears that it can be extended to other L only with extreme efforts.

In Ref. 7, besides computations of some particular cases motivated by physics, one finds decompositions of $\rho^n(\Gamma)$ for

 0022-2488/83/122722-12$02.50

a few lowest choices of ρ and Γ, and for the Lie algebras of type A_N with a general N.

In the absence of an ideal method, it is important to develop methods which would allow finding the decomposition of $\rho^n(\Gamma)$ at least in most cases of interest. The only method of this kind which is general as to the choice of L, ρ, and recursively extendable to rather large Γ, was invented in Ref. 8 and further extended in Refs. 7 and 9. For each decomposition it provides equalities similar to the equality of dimensions of representations. A decomposition is then found by a systematic (computer) search through all possible candidates for the decompositions which would satisfy those equalities with the hope that the solution will be unique. It often turns out to be the case as demonstrated in Refs. 7 and 9 on rather large examples. The role of the dimension of representations of L is played in these equalities by a more general (integer-valued) quantities: anomaly numbers[8] and indices of representations.[10]

The recent generalization[11] of the indices and anomaly numbers allows writing even more such independent equalities and thus providing a more powerful method, i.e., more restrictions on the possible candidates for the decomposition. However, in order to make use of it, the results of Refs. 8 and 9 have to be rederived for the new indices of Ref. 11. That is the first of the three aims in this paper. The second is a derivation of a closed algebraic expression for the new pth-order index for all classical simple Lie algebras. Corresponding expressions for the five exceptional simple Lie algebras cannot be obtained the same way because some auxiliary information for that is not available in a suitable form. Therefore as the last result of the paper we find explicit expressions for the sixth-order index for G_2, F_4, E_6, and E_7. This index is trivial for E_8. As in I, let $D^{(p)}(\omega) = Tr^{(\omega)} J_p$ be the pth-order fundamental indices for a generic finite-dimensional representation ω which may not necessarily be irreducible. Here, J_p is the pth-order fundamental Casimir invariant of L [see Eq. (1.8)].

If a given representation $\rho^n(\Gamma)$ for a given Young tableau Γ is written as a direct sum of irreducible representations ρ_j of L as

$$\rho^n(\Gamma) = \sum_j \oplus \rho_j, \tag{1.4}$$

then we can easily show the validity of

$$D^{(p)}(\Gamma) = \sum_j D^{(p)}(\rho_j), \tag{1.5}$$

where we have written for simplicity

$$D^{(p)}(\Gamma) \equiv D^{(p)}[\rho^n(\Gamma)] \tag{1.6}$$

with $\Gamma \equiv \rho^n(\Gamma)$ hereafter.

Note that the validity of Eq. (1.5) is nontrivial for all values of integers p such that the pth-order fundamental Casimir invariants J_p exist nontrivially, as we see from the result of I. Since the value of $D^{(p)}(\rho_j)$ can be readily calculated, Eq. (1.6) would furnish a strong check for the correct decomposition equation (1.4), provided the numerical value of $D^{(p)}(\Gamma)$ is calculable. We give an explicit formula for $D^{(p)}(\Gamma)$ in Sec. 2. A formula for $D^{(p)}(\rho)$ is given in Sec. 4 for any classical simple L and its irreducible representation ρ. Also, $l_4(\Gamma)$ has been computed in Sec. 2 for any simple Lie algebra. Moreover, some practical illustrative examples such as the decompositions (1.3) are shown in Sec. 3.

For later use, we recapitulate here the definition[11] of the pth-order index $D^{(p)}(\rho)$. Let L be a simple Lie algebra with basis $t_1, t_2, \ldots, t_d$. The Lie multiplication table is written then as

$$[t_\mu, t_\nu] = c_{\mu\nu}^{\lambda} t_\lambda, \tag{1.7}$$

where the usual summation convention on the repeated index is understood. If the rank of L is r, then L has[12] precisely r fundamental Casimir invariants. We construct their basis J_p for a pth-order Casimir invariant according to the prescription given in I. Writing J_p as

$$J_p = g^{\mu_1\mu_2\cdots\mu_p} t_{\mu_1} t_{\mu_2} \cdots t_{\mu_p} \tag{1.8}$$

for some totally symmetric coefficients $g^{\mu_1\cdots\mu_p}$, they satisfy orthogonality conditions such as

$$g^{\mu\nu\alpha\beta} g_{\mu\nu} g_{\alpha\beta} = 0 \quad (p = 4), \tag{1.9a}$$

$$g^{\mu\nu\lambda\alpha\beta} g_{\mu\nu\lambda} g_{\alpha\beta} = 0 \quad (p = 5), \tag{1.9b}$$

$$g^{\mu\nu\lambda\alpha\beta\gamma} g_{\mu\nu} g_{\lambda\alpha} g_{\beta\gamma} = g^{\mu\nu\lambda\alpha\beta\gamma} g_{\mu\nu\lambda\alpha} g_{\beta\gamma} = g^{\mu\nu\lambda\alpha\beta\gamma} g_{\mu\nu\lambda} g_{\alpha\beta\gamma} = 0 \quad (p = 6). \tag{1.9c}$$

Here, we lowered suffixes by means of the Killing metric tensor $g_{\mu\nu}$ defined by

$$g_{\mu\nu} = c\,\mathrm{Tr}(\mathrm{ad}\, t_\mu\ \mathrm{ad}\, t_\nu), \quad c \neq 0 \tag{1.10}$$

for unspecified normalization constant c. The explicit forms of J_p for $p \leqslant 5$ are found in I. Let ρ be a representation of L, which is not necessarily irreducible. Then, the pth-order index $D^{(p)}(\rho)$ is now defined by

$$D^{(p)}(\rho) = \mathrm{Tr}\, J_p = g^{\mu_1\cdots\mu_p}\, \mathrm{Tr}(X_{\mu_1} X_{\mu_2} \cdots X_{\mu_p}), \tag{1.11}$$

where X_μ is the representation matrix of t_μ in the representation ρ. When ρ is irreducible, then we have

$$D^{(p)}(\rho) = J_p(\rho) d(\rho), \tag{1.12}$$

where $d(\rho)$ is the dimension of ρ and $J_p(\rho)$ is the eigenvalue of J_p in ρ. If ρ is reducible and is decomposed as a direct sum of irreducible components ρ_j by

$$\rho = \sum_j \oplus \rho_j,$$

then we have

$$D^{(p)}(\rho) = \sum_j D^{(p)}(\rho_j). \tag{1.13}$$

The explicit formula for $J_p(\rho)$ (ρ being irreducible) is computed in Ref. 13 for cases $p \leqslant 4$ and in I for $p = 5$. More general formulas for $D^{(p)}(\rho)$ will be presented in Sec. 4, and the case of $p = 6$ for exceptional Lie algebras is also dicussed in Sec. 5.

Also, for many practical problems, the second- and fourth-order indices $l_2(\rho)$ and $l_4(\rho)$ of Ref. 9 are very useful in view of extensive tabulation[3] of their numerical values, which are defined by

$$l_{2p}(\rho) = \sum_M (M, M)^p = \mathrm{Tr}\,(g^{ij} H_i H_j)^p \tag{1.14}$$

for nonnegative integer values of p. Here, H_j $(j = 1, 2,...,r)$ refer to the Cartan subalgebra elements of L, and the summation extends over all weights M of ρ. Actually, $l_2(\rho)$ is simply proportional to $D^{(2)}(\rho)$, i.e.,

$$l_2(\rho) = c_0 D^{(2)}(\rho) \tag{1.15}$$

for some constant c_0. Since the presence of the normalization constant c_0 is often irrelevant, we may identify $l_2(\rho)$ with $D^{(2)}(\rho)$. The relation between $l_4(\rho)$ and $D^{(4)}(\rho)$ is more complicated and is discussed in Ref. 13.

Finally we simply remark that $D^{(p)}(\rho)$ are also useful[11] for finding the Clebsch–Gordan decomposition of the product of two irreducible representations as well as for the branching rules.

2. EVALUATION OF $D^{(\rho)}(\Gamma)$

Let L be a simple Lie algebra and let ρ be a representation of L, which is not necessarily irreducible. Any generic element t of L is written as

$$t = \xi^{\mu} t_{\mu}, \tag{2.1}$$

where $t_1, t_2,...,t_d$ is a basis of L and where $\xi^1, \xi^2,...,\xi^d$ are some real or complex numbers. Then, the matrix X representing t in the representation ρ is given by

$$X = \xi^{\mu} X_{\mu} = \rho(t). \tag{2.2}$$

Here X_{μ} $(= \rho(t_{\mu}))$ is the representation matrix of t_{μ}. Let G be a connected component of the Lie group generated by the Lie algebra L. Then, any $g \in G$ can be written as

$$g = e^{t} \tag{2.3}$$

for some $t \in L$. Correspondingly, its representation matrix U is expressed as

$$U \equiv \rho(g) = e^{X} = \exp(\xi^{\mu} X_{\mu}). \tag{2.4}$$

Since we have

$$\mathrm{Tr}\, X = 0 \tag{2.5}$$

for any semisimple Lie algebra, we find

$$\det U = 1. \tag{2.6}$$

Moreover, assuming L to be a compact Lie algebra, U is a unitary matrix, i.e.,

$$U^{\dagger} U = E, \tag{2.7}$$

where E is the identity matrix in ρ.

If ψ is a vector in the representation space of ρ, then by action of $g \in G$, ψ will transform into

$$\psi \overset{g}{\longrightarrow} \psi' = U\psi. \tag{2.8}$$

Let $d(\rho)$ be the dimension of ρ and set for simplicity

$$N = d(\rho). \tag{2.9}$$

Then, in view of Eqs. (2.6) and (2.7), we may regard $N \times N$ matrix U to be an element of the defining (or basic) representation of the unitary unimodular group SU(N). Let Γ be a given Young tableau with k rows specified by Young's symbol $(f_1, f_2,...,f_k)$ satisfying

$$f_1 \geqslant f_2 \geqslant f_3 \geqslant \cdots \geqslant f_k \geqslant 0. \tag{2.10}$$

If $f_k = 0$, then we may simply omit it and replace k by $k - 1$. Correspondingly, we often write

$$\Gamma = (f_2, f_2,...,f_k). \tag{2.11}$$

We emphasize the fact that the notion of this Young tableau is independent of any specific Lie algebra L or of the SU(N) group under consideration. Next, consider the tensor product $\psi^n = \psi \otimes \psi \otimes \cdots \otimes \psi$ (n times) and construct the irreducible representation space $\psi^n(\Gamma)$ of the SU(N) group with a given permutation symmetry associated with the tableau Γ out of ψ^n. Under the transformation equation (2.8), we will then have

$$\psi^n(\Gamma) \overset{g}{\longrightarrow} \psi^n(\Gamma) = U^n(\Gamma)\psi^n(\Gamma) \tag{2.12}$$

for some matrix $U^n(\Gamma)$, whose trace defines the character $\chi_{\Gamma}(U)$ by

$$\chi_{\Gamma}(U) = \mathrm{Tr}\, U^n(\Gamma). \tag{2.13}$$

Note that the total number of boxes in the tableau Γ is given by

$$n = f_1 + f_2 + \cdots + f_k. \tag{2.14}$$

Next, we restrict SU(N) to its subgroup G. Then, its restricted representation $\rho^n(\Gamma)$ will generally be reducible under G. Our task is to compute $D^{(\rho)}[\rho^n(\Gamma)]$. As we will see below, this can be readily accomplished once the explicit formula for the character $\chi_{\Gamma}(U)$ of the SU(N) is known. However, since its derivation is a little complicated, we will explain it first by a simpler example of $n = 3$ with $\Gamma =$. In that case, the character is easily computed to be

$$\chi_{\square\square\square}(U) = \tfrac{1}{6}\{(\mathrm{Tr}\, U)^3 + 2\,\mathrm{Tr}(U^3) + 3\,\mathrm{Tr}\, U\,\mathrm{Tr}(U^2)\}, \tag{2.15}$$

where $\mathrm{Tr}\, U^l$ $(l = 1, 2, 3)$ refer to the trace with respect to the N-dimensional space. Similarly, for the antisymmetric representation, we have

$$\chi_{\begin{smallmatrix}\square\\\square\\\square\end{smallmatrix}}(U) = \tfrac{1}{6}\{(\mathrm{Tr}\, U)^3 + 2\,\mathrm{Tr}(U^3) - 3\,\mathrm{Tr}\, U\,\mathrm{Tr}(U^2)\}. \tag{2.16}$$

Since U is defined by Eq. (2.4), the right side of Eq. (2.15) can be evaluated as

$$\begin{aligned}\chi_{\square\square\square}(U) &= \tfrac{1}{6}(N+1)(N+2) + \tfrac{1}{4}(N+2)(N+3)\mathrm{Tr}\, X^2 \\ &\quad + \tfrac{1}{12}(N+3)(N+6)\mathrm{Tr}\, X^3 + \tfrac{1}{8}(N+4)(\mathrm{Tr}\, X^2)^2 \\ &\quad + \tfrac{1}{48}[N^2 + 17N + 54]\mathrm{Tr}\, X^4 + \cdots,\end{aligned} \tag{2.17}$$

where $\cdots$ in Eq. (2.17) implies terms involving traces of higher powers of X such as $\mathrm{Tr}\, X^5$, $\mathrm{Tr}\, X^2\, \mathrm{Tr}\, X^3$, etc.

Let $\tilde{X}$ be the representation matrix of t in the general representation $\rho^n(\Gamma)$ of L. Then,

$$\begin{aligned}\chi_{\Gamma}(U) = \mathrm{Tr}^{(\Gamma)}\exp\tilde{X} &= \sum_{n=0}^{\infty}\frac{1}{n!}\mathrm{Tr}^{(\Gamma)}(\tilde{X})^n \\ &= d(\Gamma) + \frac{1}{2!}\mathrm{Tr}^{(\Gamma)}(\tilde{X})^2 + \frac{1}{3!}\mathrm{Tr}^{(\Gamma)}(\tilde{X})^3 + \cdots,\end{aligned} \tag{2.18}$$

where $d(\Gamma)$ is the dimension of $\rho^n(\Gamma)$ and $\mathrm{Tr}^{(\Gamma)}$ indicates the trace in the representation space $\rho^n(\Gamma)$. Now we specialize Γ

to correspond to the tableau $\square\square\square$ and note

$$X = \xi^{\mu} X_{\mu}, \quad \tilde{X} = \xi^{\mu} \tilde{X}_{\mu}. \tag{2.19}$$

Since ξ^{μ}'s are arbitrary real or complex numbers, we compare coefficients of $\xi^{\mu_1}\xi^{\mu_2}\cdots\xi^{\mu_p}$ of both sides of Eqs. (2.17) and (2.18). In this way, we find

$$d(\Gamma) = \tfrac{1}{6} N(N+1)(N+2), \tag{2.20a}$$

$$\frac{1}{2!}\sum_P \mathrm{Tr}^{(\Gamma)}(\tilde{X}_\mu \tilde{X}_\nu) = \frac{1}{2}(N+2)(N+3)\frac{1}{2!}\sum_P \mathrm{Tr}(X_\mu X_\nu), \tag{2.20b}$$

$$\frac{1}{3!}\sum_P \mathrm{Tr}^{(\Gamma)}(\tilde{X}_\mu \tilde{X}_\nu \tilde{X}_\lambda) = \frac{1}{2}(N+3)(N+6)\frac{1}{3!}\sum_P \mathrm{Tr}(X_\mu X_\nu X_\lambda), \tag{2.20c}$$

for $p = 0$, 2, and 3, and

$$\frac{1}{4!}\sum_P \mathrm{Tr}^{(\Gamma)}(\tilde{X}_\mu \tilde{X}_\nu \tilde{X}_\alpha \tilde{X}_\beta) = \frac{1}{2}[N^2 + 17N + 54]\frac{1}{4!}\sum_P \mathrm{Tr}(X_\mu X_\nu X_\alpha X_\beta) + 3(N+4)\frac{1}{4!}\sum_P \mathrm{Tr}(X_\mu X_\nu)\mathrm{Tr}(X_\alpha X_\beta), \tag{2.21}$$

for $p = 4$. Here, the summation over P stands for $p!$ permutations of indices μ, ν, α, β, etc. Multiplying $g^{\mu\nu}$ and $g^{\mu\nu\lambda}$ to both sides of Eqs. (2.20b) and (2.20c), respectively, and noting Eq. (1.11), we find

$$D^{(2)}(\square\square\square) = \tfrac{1}{2}(N+2)(N+3)D^{(2)}(\square), \tag{2.22a}$$

$$D^{(3)}(\square\square\square) = \tfrac{1}{2}(N+3)(N+6)D^{(3)}(\square), \tag{2.22b}$$

where we identified ρ with the single box $\square$ and set

$$D^{(p)}(\square) \equiv D^{(p)}(\rho). \tag{2.23}$$

For $p = 4$, the situation is slightly more complicated. We first note

$$\mathrm{Tr}(X_\mu X_\nu) = \frac{D^{(2)}(\square)}{d(\rho_0)} g_{\mu\nu}, \tag{2.24}$$

where ρ_0 refers hereafter to the adjoint representation of L. Multiplying $g^{\mu\nu\alpha\beta}$ to both sides of Eq. (2.21) and noting the orthogonality relation $g^{\mu\nu\alpha\beta}g_{\mu\nu}g_{\alpha\beta} = 0$ [see Eq. (1.9a)], we find then

$$D^{(4)}(\square\square\square) = \tfrac{1}{2}(N^2 + 17N + 54)D^{(4)}(\square). \tag{2.25}$$

We remark that if we had started from the antisymmetric representation $\Gamma = \begin{smallmatrix}\square\\\square\\\square\end{smallmatrix}$ [as in Eq. (2.16)], all these formulas for $D^{(p)}(\Gamma)$ and $d(\Gamma)$ will remain the same if we make the formal change $N \to -N$. As we shall prove shortly, this property is always valid for any tableau Γ and its conjugate tableau Γ^*, except for the sign change factor $(-1)^{n-1}$ where n is defined by Eq. (2.14), i.e., n is the total number of boxes contained in Γ.

If we multiply $g^{\mu\nu}g^{\alpha\beta}$ to both sides of Eq. (2.21), then we will obtain a formula for the fourth-order index[11] $\bar{D}^{(4)}(\Gamma)$ corresponding to the Casimir invariant $(I_2)^2$. This will be discussed later. However, we will consider here the index $l_4(\Gamma)$. We restrict Greek indices μ, ν, α, β in Eq. (2.21) to Cartan subalgebra indices i, j, k, l with $X_\mu = H_i$, $X_\nu = H_j$, $X_\alpha = H_k$, and $X_\beta = H_l$, and multiply $g^{ij}g^{kl}$. Then, we obtain

$$l_4(\square\square\square) = \tfrac{1}{2}(N^2 + 17N + 54)l_4(\square) + ((r+2)/r)(N+4)(l_2(\square))^2, \tag{2.26}$$

where r is the rank of L. Note that the extra term in the right-hand side of Eq. (2.26) comes from the second contribution on Eq. (2.21), which gives however no contribution to $D^{(4)}(\square\square\square)$.

The procedure explained above is applicable to general cases, once the character formula is given. First let

$$h(f) \equiv \chi_f(U) \tag{2.27}$$

be the character of the SU(N) group corresponding to a completely symmetric representation corresponding to $\Gamma = \square\cdots\square$ (f boxes). If z is an arbitrary real or complex variable, then the generating function for $\chi_f(U)$ is well known[1] and is shown as

$$\sum_{f=0}^{\infty} z^f \chi_f(U) = \frac{1}{\det(E - zU)} = \exp\left(\sum_{l=1}^{\infty} \frac{z^l}{l}\mathrm{Tr}(U^l)\right). \tag{2.28}$$

For a later purpose we introduce $a_f(U)$ which is the character of the completely antisymmetric representation of the SU(N) group with f boxes in the single column. Then, we have analogously

$$\sum_{f=0}^{N} (-1)^f z^f a_f(U) = \det(E - zU) = \exp\left(-\sum_{l=1}^{\infty} \frac{z^l}{l}\mathrm{Tr}(U^l)\right). \tag{2.29}$$

Now, the general character $\chi_\Gamma(U)$ for a given Young tableau Γ with $\Gamma = (f_1, f_2, \ldots, f_k)$ is determined from the Weyl's formula[1] to be

$$\chi_\Gamma(U) = \det \alpha_{ij}, \tag{2.30}$$

where $k \times k$ matrix α_{ij} ($i, j = 1, 2, \ldots, k$) is defined by

$$\alpha_{ij} = h(f_j + i - j) \tag{2.31}$$

with the understanding that $h(0) = 1$ and $h(l) = 0$ for $l < 0$. Next, we restrict SU(N) to its subgroup G and we insert $U = \exp X$ [see Eq. (2.4)] into the right side of (2.30). However, for the purpose of calculating $D^{(p)}(\Gamma)$ ($p \geqslant 2$), only terms containing $\mathrm{Tr}\, X^p$ in the expansion are relevant by the orthogonality conditions, Eqs. (1.9), just as the second term in Eq. (2.21) did not contribute to the evaluation of $D^{(4)}(\square\square\square)$ in Eq. (2.25). Thus, we expand for $p \geqslant 2$,

$$h(f) = \frac{(N+f-1)!}{f!(N-1)!} + \frac{1}{p!}\beta^{(p)}(f)\mathrm{Tr}\, X^p + \cdots. \tag{2.32}$$

Here, we have set

$$\beta^{(p)}(f) = \sum_{l=0}^{f} \frac{(N+l-1)!}{l!(N-1)!}(f-l)^{p-1}, \tag{2.33}$$

and other terms in the right side of Eq. (2.32) are either those involving traces of different powers of X such as $\mathrm{Tr}\, X^{p+1}$ or terms involving products of traces such as $\mathrm{Tr}\, X^2\, \mathrm{Tr}\, X^{p-2}$, etc. As we shall see shortly, all these neglected terms are not necessary for our consideration. Now, we insert Eq. (2.32) into Eq. (2.30). The left side of $\chi_\Gamma(U)$ is again expanded by Eq. (2.18). Comparing both sides of coefficients of

$\xi^{\mu_1}\xi^{\mu_2}\cdots\xi^{\mu_p}$, we find

$$\frac{1}{p!}\sum_P \mathrm{Tr}^{(\Gamma)}(\tilde{X}_{\mu_1}\tilde{X}_{\mu_2}\cdots\tilde{X}_{\mu_p}) = Q_p(\Gamma)\frac{1}{p!}\sum_P \mathrm{Tr}(X_{\mu_1}X_{\mu_2}\cdots X_{\mu_p}) + \cdots, \tag{2.34}$$

where the second term of the right side of Eq. (2.34) contains products of traces such as

$$\mathrm{Tr}(X_{\mu_1}X_{\mu_2})\mathrm{Tr}(X_{\mu_3}\cdots X_{\mu_p}),$$

etc. Multiplying $g^{\mu_1\cdots\mu_p}$ to both sides of Eq. (2.34) and noting the orthogonality condition, Eqs. (1.9), together with the reasoning explained in I, only the first term in the right side on Eq. (2.34) survives, and we obtain

$$D^{(p)}(\Gamma) = Q_p(\Gamma)D^{(p)}(\square). \tag{2.35}$$

The coefficient $Q_p(\Gamma)$ can be readily computed as follows. We first define $k\times k$ matrix A by

$$A_{ij} = \frac{(N+i-j+f_j-1)!}{(N-1)!(i-j+f_j)!} = \binom{N+i-j+f_j-1}{N-1}. \tag{2.36}$$

Let B_{ij} be the (i,j) cofactor of the matrix A so that the $k\times k$ matrix B is given by

$$B = (A^T)^{-1}\det A, \tag{2.37}$$

where A^T is the transpose matrix of A. Then, $Q_p(\Gamma)$ is now calculated by the formula

$$Q_p(\Gamma) = \sum_{i,j=1}^{k}\beta^{(p)}(f_j+i-j)B_{ij}, \tag{2.38}$$

where $\beta^{(p)}(f)$ is defined by Eq. (2.33) with $\beta^{(p)}(f)\equiv 0$ for $f\leqslant 0$. We also note that the dimension $d(\Gamma)$ is given by

$$d(\Gamma) = \det A. \tag{2.39}$$

When the Young tableau Γ is fixed, then $Q_p(\Gamma)$ as a function of N is easily seen from Eqs. (2.33)–(2.38) to be a polynomial of N of degree $n-1$ with n being given by Eq. (2.14). We also remark that we have $D^{(p)}(\Gamma) = D^{(p)}(\rho^n(\Gamma)) = 0$ identically whenever $D^{(p)}(\rho) = 0$. Of course, this fact is trivially satisfied, if L does not possess pth-order fundamental Casimir invariant J_p. At any rate, we rewrite often Eq. (2.35) simply as

$$Q_p(\Gamma) = D^{(p)}(\Gamma)/D^{(p)}(\square) \tag{2.40}$$

without cautioning that Eq. (2.40) is really meaningless when we have $D^{(p)}(\square) = 0$.

Next, let Γ^* be the Young tableau conjugate to Γ, i.e., the tableau which can be obtained from Γ by its mirror reflection with respect to its diagonal blocks. Then, by Jacobi–Trudi, as well as Naegelsbach–Kostka identities,[14] we can rewrite the character $\chi_{\Gamma^*}(U)$ of Γ^* in Eq. (2.30) to be

$$\chi_{\Gamma^*}(U) = \det\tilde{\alpha}_{ij}, \tag{2.41}$$

where $k\times k$ matrix $\tilde{\alpha}_{ij}$ is defined by

$$\tilde{\alpha}_{ij} = a(f_j+i-j), \quad a(f)\equiv a_f(U) \tag{2.42}$$

in terms of $a_f(U)$ defined by Eq. (2.29). Then, repeating the same procedure as before, we calculate

$$Q_p(\Gamma^*) = \frac{D^{(p)}(\Gamma^*)}{D^{(p)}(\square)} = \sum_{i,j=1}^{k}\tilde{B}_{ij}\tilde{\beta}^{(p)}_{ij}, \tag{2.43}$$

$$d(\Gamma^*) = \det\tilde{A}. \tag{2.44}$$

Here, $k\times k$ matrix $\tilde{A}$ is defined by

$$\tilde{A}_{ij} = \frac{N!}{(f_j-j+i)!(N-f_j+j-i)!} = \binom{N}{f_j-j+i} \tag{2.45}$$

and $\tilde{B}_{ij}$ is (i,j) cofactor of the matrix $\tilde{A}$. Finally,

$$\tilde{\beta}^{(p)}_{ij} = \tilde{\beta}^{(p)}(f_j+i-j), \tag{2.46a}$$

$$\tilde{\beta}^{(p)}(f) = -\sum_{l=0}^{f}\frac{N!}{l!(N-l)!}(-1)^{f-l}(f-l)^{p-1}. \tag{2.46b}$$

Comparing Eqs. (2.38) and (2.43) and noting

$$\binom{-N}{l} = (-1)^l\binom{N+l-1}{l},$$

we obtain an identity

$$Q_p(\Gamma^*,N) = (-1)^{n-1}Q_P(\Gamma,-N), \tag{2.47a}$$

$$d(\Gamma^*,N) = (-1)^n d(\Gamma,-N), \tag{2.47b}$$

where n is again given by Eq. (2.14) and we wrote $Q_p(\Gamma)$, etc. as $Q_p(\Gamma,N)$ in order to emphasize its N dependence for a given Young tableau Γ. The validity of Eq. (2.47a) for the case of $p=2$ and 3 has been noted in Refs. 5 and 7. These are also consistent with a theorem of Cvitanović and Kennedy.[15]

The advantage of using $D^{(p)}(\Gamma)$ instead of $l_p(\Gamma)$ is that its calculation is simpler. However, since an extensive table of $l_2(\rho)$ and $l_4(\rho)$ is available,[3] we have computed also the expression for $l_4(\Gamma)$ to be

$$l_4(\Gamma) = Q_4(\Gamma)l_4(\square) + ((r+2)/r)R_4(\Gamma)(l_2(\square))^2, \tag{2.48}$$

where $Q_4(\Gamma)$ is defined by Eq. (2.38) for $p=4$, and $R_4(\Gamma)$ is given by

$$R_4(\Gamma) = \sum_{i,j=1}^{k}\binom{N+f_j+i-j+1}{N+3}B_{ij} + \sum_{\substack{i_1\neq i_2\\ j_1\neq j_2}}^{k}\binom{N+f_{j_1}+i_1-j_1}{N+1}\binom{N+f_{j_2}+i_2-j_2}{N+1} \times B_{i_1i_2j_1j_2}. \tag{2.49}$$

Note that in order to derive this formula, we had to include terms proportional to $\mathrm{Tr}(X_\mu X_\nu)\,\mathrm{Tr}\,(X_\alpha X_\beta)$ which gives no contribution to $D^{(p)}(\Gamma)$. Here, $B_{i_1,i_2j_1j_2}$ is cofactor of the matrix A corresponding to two elements $A_{i_1j_1}$ and $A_{i_2j_2}$ $(i_1\neq i_2, j_1\neq j_2)$, and we have set

$$\binom{l}{j} = \frac{l!}{j!(l-j)!}$$

for simplicity. As a function of N for a given Γ, $Q_4(\Gamma)$ and $R_4(\Gamma)$ are readily seen to be polynomials of N of degree, respectively, of $n-1$ and $n-2$ in agreement with the conclusion of Refs. 6–9. Writing $R_4(\Gamma)$ as $R_4(\Gamma,N)$, we find also

$$R_4(\Gamma^*,N) = (-1)^n R_4(\Gamma,-N) \tag{2.50}$$

as an analog of Eq. (2.47). If we decompose $\rho^n(\Gamma)$ as a direct sum of irreducible components ρ_j as in Eq. (1.4), then $l_4(\Gamma)$ satisfies

$$l_4(\Gamma) = \sum_j l_4(\rho_j), \tag{2.51}$$

which can be used as a check of Eq. (1.4).

In I, we have introduced another fourth-order index $\overline{D}^{(4)}(\rho)$ by

$$\overline{D}^{(4)}(\rho) = b^{\mu\nu\alpha\beta}\mathrm{Tr}(X_\mu X_\nu X_\alpha X_\beta), \tag{2.52a}$$

$$b^{\mu\nu\alpha\beta} = \tfrac{1}{3}(g^{\mu\nu}g^{\alpha\beta} + g^{\mu\alpha}g^{\nu\beta} + g^{\mu\beta}g^{\nu\alpha}). \tag{2.52b}$$

If Γ is a direct sum of irreducible components ρ_j as $\Gamma = \sum_j \oplus \rho_j$, then we have

$$\overline{D}^{(4)}(\Gamma) = \sum_j D^{(2)}(\rho_j)\left\{\frac{D^{(2)}(\rho_j)}{d(\rho_j)} - \frac{1}{6}\frac{D^{(2)}(\rho_0)}{d(\rho_0)}\right\} = \sum_j \overline{D}^{(4)}(\rho_j), \tag{2.53}$$

where $\overline{D}^{(4)}(\Gamma)$ is calculated similarly also as

$$\overline{D}^{(4)}(\Gamma) = Q_4(\Gamma)\overline{D}^{(4)}(\square) + \frac{d(\rho_0)+2}{d(\rho_0)} R_4(\Gamma)(D^{(2)}(\square))^2. \tag{2.54}$$

Concluding this section, we may note the following. We have already remarked that we have $D^{(p)}(\Gamma) = 0$ identically for any Γ if $D^{(p)}(\rho) = D^{(p)}(\square) = 0$. This statement holds valid even if ρ ($= \square$) is reducible. For $p = 3$, this fact may have the following physical implication, since $D^{(3)}(\rho)$ is the anomaly coefficient[16] of the gauge field theory. Suppose that $\square$ ($= \rho$) corresponds to some fundamental constituent representation of, say, preon, or quarks and leptons in any grand unified theory such as SU(5) or SO(10). Then, the renormalizability of theory[17] requires $D^{(3)}(\square) = 0$. Therefore the consequence $D^{(3)}[\rho^n(\Gamma)] = 0$ implies also the fact that the effective Lagrangian field theory for any type of bound states of these preons or quark-leptons with the same gauge fields must satisfy also the same anomaly-free condition.

3. SOME EXAMPLES

In order to make our formulation useful for practical problems, we compute here some explicit formulas for $Q^p(\Gamma)$ and $R_4(\Gamma)$. As we have emphasized already, they depend only upon a given Young tableau Γ and N ($= d(\square)$) but not directly upon the original Lie algebra L. The simplest case is for a totally symmetric or totally antisymmetric Young tableau. We adopt here the standard notation

$$\underbrace{\square\cdots\square}_{f \text{ times}} = \{f\} = \{f\Lambda_1\}, \tag{3.1}$$

$$\begin{array}{l}\square\\ \vdots\\ \square\end{array}\ (f \text{ times}) = \{1^f\} = \{\Lambda_f\}. \tag{3.2}$$

Then, we use Eq. (2.38) with $k = 1$ for Eq. (3.1) and Eq. (2.43) with $k = 1$ for Eq. (3.2). The results are

$$Q_p(\Lambda_f) = \widetilde{\beta}^{(p)}(f) = -\sum_{l=0}^{f-1}(-1)^{f-l}(f-l)^{p-1}\frac{N!}{l!(N-l)!}, \tag{3.3a}$$

$$Q_p(f\Lambda_1) = \beta^{(p)}(f) = \sum_{l=0}^{f-1}(f-l)^{p-1}\frac{(N+l-1)!}{l!(N-1)!}. \tag{3.3b}$$

Similarly, we calculate

$$R_4(\Lambda_f) = \frac{(N-4)!}{(f-2)!(N-f-2)!}, \tag{3.4a}$$

$$R_4(f\Lambda_1) = \frac{(N+f+1)!}{(f-2)!(N+3)!}. \tag{3.4b}$$

The expression for $Q_p(\Lambda_f)$ may be identified with the pth-order generalized anomaly coefficient $A_p(N,f)$. Equation (3.3a) has been applied to grand unified theory by Frampton and Kephart[18] for a study of the fermion family problem in particle physics.

For special cases of $p \leqslant 5$, we can sum up the expression for $\beta^{(p)}(f)$ to be

$$\begin{aligned}
\beta^{(2)}(f) &= \frac{(N+f)!}{(N+1)!(f-1)!},\\
\beta^{(3)}(f) &= (N+2f)\frac{(N+f)!}{(N+2)!(f-1)!},\\
\beta^{(4)}(f) &= \{N(N-1)+6f(N+f)\}\frac{(N+f)!}{(N+3)!(f-1)!},\\
\beta^{(5)}(f) &= (N+2f)\{N(N-5)+12f(N+f)\}\frac{(N+f)!}{(N+4)!(f-1)!}.
\end{aligned} \tag{3.5}$$

For more detail see the Appendix. Then, the value of $\widetilde{\beta}^{(p)}(f)$ can be evaluated from $\beta^{(p)}(f)$ by letting $N \to -N$ as before. The values of $Q^p(\Lambda_f)$ for the antisymmetric representations are evaluated from these to be

$$\begin{aligned}
Q_2(\Lambda_f) &= \frac{(N-2)!}{(f-1)!(N-f-1)!},\\
Q_3(\Lambda_f) &= (N-2f)\frac{(N-3)!}{(f-1)!(N-f-1)!},\\
Q_4(\Lambda_f) &= \{N(N+1)-6(N-f)\}\frac{(N-4)!}{(f-1)!(N-f-1)!},\\
Q_5(\Lambda_f) &= (N-2f)\{N(N+5)-12f(N-f)\}\frac{(N-5)!}{(f-1)!(N-f-1)!},
\end{aligned} \tag{3.6}$$

which reproduces, of course, the same results as those directly computed[11,13] by the formula $D^{(p)}(\rho) = d(\rho)J_p(\rho)$ for the SU(N) group. The results for $Q_p(\Lambda_f)$ ($p = 2$ and 3), are previously given in Refs. 8, 9, and 16.

Here, we present also explicit formulas for $Q_p(\Gamma)$ ($p \leqslant 4$), and $R_4(\Gamma)$ for cases of $n = 2$, 3, or 4.

(1) $n = 2$

(1a) $\Gamma = \square\square$,

$$d(\Gamma) = \tfrac{1}{2}N(N+1), \quad R_4(\Gamma) = 1,$$
$$Q_p(\Gamma) = N + 2^{p-1} \quad (p \geqslant 2).$$

(1b) $\Gamma = \begin{array}{l}\square\\ \square\end{array}$,

$$d(\Gamma) = \tfrac{1}{2}N(N-1), \quad R_4(\Gamma) = 1,$$
$$Q_p(\Gamma) = N - 2^{p-1} \quad (p \geqslant 2).$$

(2) $n=3$

(2a) $\Gamma =$  ,

$$d(\Gamma)=\tfrac{1}{6}N(N+1)(N+2),$$
$$Q_2(\Gamma)=\tfrac{1}{2}(N+2)(N+3),$$
$$Q_3(\Gamma)=\tfrac{1}{2}(N+3)(N+6),$$
$$Q_4(\Gamma)=\tfrac{1}{2}(N^2+17N+54),$$
$$R_4(\Gamma)=N+4,$$
$$Q_p(\Gamma)=\tfrac{1}{2}N^2+\tfrac{1}{2}[1+2^p]N+3^{p-1}\quad (p\geqslant 2).$$

(2b) $\Gamma =$  ,

$$d(\Gamma)=\tfrac{1}{6}N(N-1)(N-2),$$
$$Q_2(\Gamma)=\tfrac{1}{2}(N-2)(N-3),$$
$$Q_3(\Gamma)=\tfrac{1}{2}(N-3)(N-6),$$
$$Q_4(\Gamma)=\tfrac{1}{2}[N^2-17N+54],$$
$$R_4(\Gamma)=N-4,$$
$$Q_p(\Gamma)=\tfrac{1}{2}N^2-\tfrac{1}{2}[1+2^p]N+3^{p-1}\quad (p\geqslant 2).$$

(2c) $\Gamma =$  ,

$$d(\Gamma)=\tfrac{1}{3}N(N^2-1),$$
$$Q_2(\Gamma)=N^2-3,$$
$$Q_3(\Gamma)=N^2-9,$$
$$Q_4(\Gamma)=N^2-27,$$
$$R_4(\Gamma)=2N,$$
$$Q_p(\Gamma)=N^2-3^{p-1}\quad (p\geqslant 2).$$

(3) $n=4$

(3a) $\Gamma =$  ,

$$d(\Gamma)=\tfrac{1}{24}N(N+1)(N+2)(N+3),$$
$$Q_2(\Gamma)=\tfrac{1}{6}(N+2)(N+3)(N+4),$$
$$Q_3(\Gamma)=\tfrac{1}{6}(N+3)(N+4)(N+8),$$
$$Q_4(\Gamma)=\tfrac{1}{6}(N+4)(N^2+23N+96),$$
$$R_4(\Gamma)=\tfrac{1}{2}(N+4)(N+5).$$

(3b) $\Gamma =$  ,

$$d(\Gamma)=\tfrac{1}{24}N(N-1)(N-2)(N-3),$$
$$Q_2(\Gamma)=\tfrac{1}{6}(N-2)(N-3)(N-4),$$
$$Q_3(\Gamma)=\tfrac{1}{6}(N-3)(N-4)(N-8),$$
$$Q_4(\Gamma)=\tfrac{1}{6}(N-4)(N^2-23N+96),$$
$$R_4(\Gamma)=\tfrac{1}{2}(N-4)(N-5).$$

(3c) $\Gamma =$  ,

$$d(\Gamma)=\tfrac{1}{12}N^2(N^2-1),$$
$$Q_2(\Gamma)=\tfrac{1}{3}N(N^2-4),$$
$$Q_3(\Gamma)=\tfrac{1}{3}N(N^2-16),$$
$$Q_4(\Gamma)=\tfrac{1}{3}N(N^2-58),$$
$$R_4(\Gamma)=N^2+2.$$

(3d) $\Gamma =$,

$$d(\Gamma)=\tfrac{1}{8}N(N-1)(N+1)(N+2),$$
$$Q_2(\Gamma)=\tfrac{1}{2}(N+2)(N^2+N-4),$$
$$Q_3(\Gamma)=\tfrac{1}{2}(N+4)(N^2+N-8),$$
$$Q_4(\Gamma)=\tfrac{1}{2}(N+8)(N^2+N-16),$$
$$R_4(\Gamma)=\tfrac{1}{2}(3N^2+9N-8).$$

(3e) $\Gamma =$,

$$d(\Gamma)=\tfrac{1}{8}N(N-1)(N+1)(N-2),$$
$$Q_2(\Gamma)=\tfrac{1}{2}(N-2)(N^2-N-4),$$
$$Q_3(\Gamma)=\tfrac{1}{2}(N-4)(N^2-N-8),$$
$$Q_4(\Gamma)=\tfrac{1}{2}(N-8)(N^2-N-16),$$
$$R_4(\Gamma)=\tfrac{1}{2}(3N^2-9N-8).$$

These reproduce of course the results of Refs. 6 and 7. More general formulas will be given in the Appendix.

We should note that for the $G=\mathrm{SU}(N)$ group, the numerical values of $l_4(\square)$ and $l_2(\square)$ with the normalization specified in Ref. 3 are given by

$$l_2(\square)=N-1,\quad l_4(\square)=(N-1)^2/N. \tag{3.7}$$

Therefore, $l_4(\Gamma^*)$ for the $\mathrm{SU}(N)$ group cannot be obtained from $l_4(\Gamma)$ by simply substituting $N\to -N$ from $l_4(\Gamma)$, although $Q_4(\Gamma^*)$ and $R_4(\Gamma^*)$ can be, respectively, obtained from $Q_4(\Gamma)$ and $R_4(\Gamma)$ by $N\to -N$. This is due to normalization condition (3.7), and $r=N-1$.

For a practical application of these results we note the following. If $\Gamma\equiv\rho^n(\Gamma)$ is a direct sum of irreducible representations ρ_j as

$$\Gamma=\sum_j \oplus \rho_j, \tag{3.8}$$

then, we should have

$$d(\Gamma)=\sum_j d(\rho_j), \tag{3.9a}$$

$$l_p(\Gamma)=\sum_j l_p(\rho_j), \tag{3.9b}$$

$$D^{(p)}(\Gamma)=\sum_j D^{(p)}(\rho_j), \tag{3.9c}$$

$$\overline{D}^{(4)}(\Gamma)=\sum_j \overline{D}^{(4)}(\rho_j). \tag{3.9d}$$

Also, we remark that $l_2(\rho)$ and $l_2(\Gamma)$ are essentially equivalent to $D^{(2)}(\rho)$ and $D^{(2)}(\Gamma)$, respectively, apart from some common normalization constant. Therefore we have

$$l_2(\Gamma)=Q_2(\Gamma)l_2(\square). \tag{3.10}$$

Also, the $D^{(3)}(\Gamma)$ sum rule is useful only for the case of the $G=\mathrm{SU}(m)$ group, since $D^{(3)}(\rho)=0$ identically[16] for all other simple Lie groups.

For example, consider the case of $G=\mathrm{SO}(10)$ with $\rho=\square$ being the 16-dimensional spinor representation. To be definite, we choose the highest weight of ρ to be Λ_5, in the lexicographical ordering convention of simple root system of

Ref. 3. We now want to show the decomposition

$$\square\!\square = \{2\Lambda_3\} \oplus \{\Lambda_1 + 2\Lambda_5\} \oplus \{\Lambda_4 + \Lambda_5\} \oplus \{2\Lambda_1\} \oplus \{0\}, \tag{3.11}$$

where $\{\Lambda\}$ implies the irreducible representation of SO(10) with the highest weight Λ. As we shall see, the sum rules for $d(\Gamma), l_2(\Gamma)$, and $l_4(\Gamma)$ essentially suffice for the establishment of Eq. (3.11). For a later purpose, we also note

$$Q_p(\square\!\square) = \tfrac{1}{3} N\{N^2 - 1 - 3(3^{p-1} - 2^{p-1})\}. \tag{3.12}$$

In the present problem we have $N = 16$ and $r = 5$. Then, our formulas lead to numerical values of

$$d(\square\!\square) = 5440,$$

$$l_2(\square\!\square) = 26\,880,$$

$$l_4(\square\!\square) = 170\,880,$$

where we used $l_2(\square) = l_2(\Lambda_5) = 20$ and $l_2(\square) = l_4(\Lambda_5) = 25$ for D_5 from the table of Ref. 3. Then, the validity of Eq. (3.11) can be established by noting

(i) $d(\rho)$ *sum rule* (3.9a),

$$5440 = 4125 + 1050 + 210 + 54 + 1;$$

(ii) $l_2(\rho)$ *sum rule* (3.9b) *for* $p = 2$,

$$26\,880 = 22\,000 + 4\,200 + 560 + 120 + 0;$$

(iii) $l_4(\rho)$ *sum rule* (3.9b) *for* $p = 4$,

$$170\,880 = 148\,000 + 20\,800 + 1760 + 320 + 0,$$

again from the table of Ref. 3. However, the solution (3.11) is actually not unique, since we could have used the representation $\{\Lambda_1 + 2\Lambda_4\}$ instead of $\{\Lambda_1 + 2\Lambda_5\}$, both of which have the same values for $d(\rho)$, $l_2(\rho)$, and $l_4(\rho)$. Even the congruence selection rule[19] does not help us to choose one of them since both belong to the same congruence class. However, we could resolve this issue by utilizing the fifth-order indices $D^{(5)}(\rho)$ whose eigenvalues are given by [see Sec. 4, Eq. (4.22)]

$$D^{(5)}(\rho) = l_1 l_2 l_3 l_4 l_5 d(\rho) \tag{3.13}$$

for any irreducible representation ρ, where l_j $(1 \leqslant j \leqslant 5)$ are defined by

$$l_j = f_j + 5 - j \quad (1 \leqslant j \leqslant 5), \text{ when } \rho = (f_1, f_2, f_3, f_4, f_5).$$

Then we calculate

$$D^{(5)}(\square\!\square) = 320 D^{(5)}(\square)$$

from Eq. (3.12) and

$$D^{(5)}(\Lambda_1 + 2\Lambda_5) = -D^{(5)}(\Lambda_1 + 2\Lambda_4) = 320 D^{(5)}(\square),$$

$$D^{(5)}(\square) = D^{(5)}(\Lambda_5) = \tfrac{1}{2} \times 9!!$$

from Eq. (3.13), while we find

$$D^{(5)}(2\Lambda_3) = D^{(5)}(\Lambda_4 + \Lambda_5) = D^{(5)}(2\Lambda_1) = D^{(5)}(0) = 0$$

when we note $l_5 = f_5 = 0$ for these irreducible representations. Then, the sum rule (3.9c) for $p = 5$ can be satisfied for the present problem only by the solution (3.11). Similarly, we can verify the validity of Eqs. (1.3) by the same technique. Our results agree with those given by Schellekens *et al.*[7] who have also studied many decompositions of this kind up to $n = 5$ by using both $l_2(\rho)$ and $l_4(\rho)$ sum rules but not the $D^{(5)}(\rho)$ sum rule. For more complicated Young tableaux, the use of higher-order indices is expected to be more useful.

As we emphasized in the Introduction, the reason for our special choice of $L = D_5$ with $\rho = \{\Lambda_5\}$ here is motivated for its potential relevance to the grand unified theory.[4] However, many other Lie algebras such as E_6 could also be potential candidates for GUT. Schellekens *et al.*[20] as well as Tosa and Marshak[21] have studied such problems for possible applications to preon models in elementary particle physics.

4. FORMULA FOR GENERAL INDICES

In order to apply our formalism to practical problems, it is necessary to know the explicit form of $D^{(p)}(\rho)$. Formulas for $J_p(\rho)$ with $p \leqslant 4$ have been given in Ref. 13, while $J_5(\rho)$ has been evaluated in Ref. 11. From these, we can calculate $D^{(p)}(\rho) = J_p(\rho) d(\rho)$ for $p \leqslant 5$. However, the method utilized in these papers is not suitable for general evaluation of $D^{(p)}(\rho)$. We shall show that we can find a general formula of $D^{(p)}(\rho)$ for any simple classical Lie algebra L.

For the $G = \mathrm{SU}(N)$ group corresponding to $L = A_{N-1}$, we observe first that $\rho^n(\Gamma)$ will be irreducible if we choose ρ to be the N-dimensional basic (or defining) representation of SU(N). Therefore, the general formula for $D^{(p)}(\Gamma)$ corresponding to the irreducible representation $\{\Gamma\}$ of the SU(N) group is simply given by Eqs. (2.35)–(2.38), by choosing $\square$ to be the basic representation with $d(\square) = N$.

The same reasoning is, however, not directly applicable to other classical groups Sp($2N$), SO($2N$), and SO($2N+1$), since $\rho^n(\Gamma)$ will not be in general irreducible even for ρ being the defining representation of these groups. Nevertheless, we can circumvent this and proceed as follows. In order to avoid possible confusion, we simply ignore the construction of $\rho^n(\Gamma)$, and assume hereafter ρ to be a generic irreducible tensor representation of these classical Lie groups, with Young tableau specified by the integers satisfying

$$f_1 \geqslant f_2 \geqslant \cdots \geqslant f_k \geqslant 0. \tag{4.1}$$

We may omit f_k and replace k by $k-1$, if $f_k = 0$. We also change the meaning and notation of Γ so that Γ now signifies the Young tableau $(f_1, f_2, \ldots, f_k)$ associated with the irreducible representation ρ. Let us first consider the case of the Sp($2N$) group. Since the basic (or defining) representation of Sp($2N$) is $2N$ dimensional, the matrix U discussed in Sec. 2 is now $2N$ dimensional. We define $h(f) = \chi_f(U)$ again by Eq. (2.28). Then, the character formulas for the irreducible representation ρ is now given by the Weyl's formula[1]

$$\chi_\rho(U) = \tfrac{1}{2} \det \alpha_{ij}, \tag{4.2}$$

where $k \times k$ matrix α_{ij} is defined by

$$\alpha_{ij} = h(f_j + i - j) + h(f_j - i - j + 2). \tag{4.3}$$

Note that in order to avoid possible confusion, we used the notation $\chi_\rho(U)$ in Eq. (4.2) instead of $\chi_\Gamma(U)$ of Sec. 2. Nevertheless, we can utilize exactly the same procedure of Sec. 2 for evaluation of $D^{(p)}(\rho)$, once the character formula Eq. (4.2) is known. We define the $k \times k$ matrix A now by

$$A_{ij} = \binom{2N + f_j + i - j}{2N} + \binom{2N + f_j - i - j + 2}{2N}. \tag{4.4}$$

Let B_{ij} be similarly (i,j) cofactor of the matrix A so that

$$B=(A^{T})^{-1}\det A, \tag{4.5}$$

again in the matrix notation. We now set

$$\beta^{(p)}(f)=\sum_{l=0}^{f}\binom{2N+l-1}{l}(f-l)^{p-1} \tag{4.6}$$

and

$$\gamma_{ij}^{(p)}=\beta^{(p)}(f_j+i-j)+\beta^{(p)}(f_j-i-j+2). \tag{4.7}$$

Note that $\beta^{(p)}(f)$ given by Eq. (4.6) is the same as in Eq. (2.33) if we let $N\to 2N$. We now find

$$D^{(p)}(\rho)=Q_p(\rho)D^{(p)}(\square), \tag{4.8}$$

$$Q_p(\rho)=\frac{1}{2}\sum_{i,j=1}^{k}B_{ij}\gamma_{ij}^{(p)}, \tag{4.9}$$

where $\square$ is now the $2N$-dimensional basic (or defining) representation of Sp$(2N)$. We also have

$$d(\rho)=\tfrac{1}{2}\det A. \tag{4.10}$$

We next consider the case of the SO$(2N+1)$ group, corresponding to the Lie algebra B_N. We first restrict ourselves to the irreducible tensor representation ρ with the Young tableau $\Gamma=(f_1,\ldots,f_k)$ satisfying Eq. (4.1). The matrix U of Sec. 2 is now chosen to be a $(2N+1)$-dimensional basic representation. The character formula for this case is given by[1,22]

$$\chi_\rho(U)=\tfrac{1}{2}\det\alpha_{ij}, \tag{4.11}$$

$$\alpha_{ij}=h(f_j+i-j)+h(f_j-i-j+2)-h(f_j+i-j-2)-h(f_j-i-j), \tag{4.12}$$

The corresponding $k\times k$ matrix A is expressed now as

$$A_{ij}=\binom{2N+f_j+i-j}{2N}-\binom{2N+f_j-i-j}{2N}+\binom{2N+f_j-i-j+2}{2N}-\binom{2N+f_j+i-j-2}{2N}, \tag{4.13}$$

while B_{ij} is again defined as (i,j) cofactor of the matrix A. Also, setting

$$\gamma_{ij}^{(p)}=\beta^{(p)}(f_j+i-j)+\beta^{(p)}(f_j-i-j+2)-\beta^{(p)}(f_j-i-j)-\beta^{(p)}(f_j+i-j-2), \tag{4.14}$$

$$\beta^{(p)}(f)=\sum_{l=0}^{f}\binom{2N+l}{l}(f-l)^{p-1}, \tag{4.15}$$

we find the validity of Eqs. (4.8) and (4.9) again for the case of SO$(2N+1)$. We can generalize our formulas for spinor representations as follows. The most general irreducible representation of the Lie algebra B_N is well known to be characterized by N nonnegative numbers f_j satisfying

$$f_1\geqslant f_2\geqslant\cdots\geqslant f_N\geqslant 0. \tag{4.16}$$

The tensor representations correspond to the case when all f_j are nonnegative integers. Also, if we have $f_{k+1}=f_{k+2}=\cdots=f_N=0$, then we may simply drop all these extra symbols f_j $(k+1\leqslant j\leqslant N)$ as in Eq. (4.1). However, we have to use positive half-integer values for f_j for spinor representations, so that we must always set $k=N$ for such cases. Regardless, we generally define l_j by

$$l_j=f_j+N+\tfrac{1}{2}-j \tag{4.17}$$

for $1\leqslant j\leqslant N$. It is known that the dimensional formula for $d(\rho)$ is a completely antisymmetric polynomial of $l_1,l_2,\ldots,l_N$. Similarly, from explicit formulas of nonsymmetrized Casimir invariants $I_p^{(N,S)}(\rho)$ calculated by many authors,[23–26] we can show[27] that the $2p$th-order Casimir invariant $J_{2p}(\rho)$ is a totally symmetric pth-order polynomial of $l_1^2,l_2^2,\ldots,l_N^2$. Since this fact holds valid irrespective of whether ρ is a tensor or spinor representation of B_N, and since $Q_p(\rho)$ in Eq. (4.9) can be rewritten as antisymmetric polynomials of $l_1,l_2,\ldots,l_N$ for $k=N$, we conclude that formulas (4.8)–(4.10) must also be valid for all these cases if we set $k=N$.

The case of the SO$(2N)$ group corresponding to the Lie algebra D_N is slightly more involved. Any irreducible representation ρ of D_N is characterized by N real numbers $f_1,f_2,\ldots,f_N$ satisfying now

$$f_1\geqslant f_2\geqslant\cdots\geqslant f_{N-1}\geqslant|f_N|\geqslant 0, \tag{4.18}$$

where the f_j are simultaneously either all integers or all half-integers, depending upon the tensor or spinor representations. Also, if $f_N\neq 0$, we have to distinguish two distinct irreducible representations corresponding to

$$\rho=(f_1,f_2,\ldots f_{N-1},f_N),\qquad \rho'=(f_1,f_2,\ldots f_{N-1},-f_N). \tag{4.19}$$

Now, the Lie algebra D_N possesses N fundamental Casimir invariants[26] which can be classified into two classes. The first class may be called regular ones with Casimir invariants $J_2,J_4,\ldots,J_{2N-2}$, while the second one consists of one N th-order invariant we specify by $\hat{J}_N$. When we set

$$l_j=f_j+N-j \tag{4.20}$$

for $1\leqslant j\leqslant N$, then eigenvalues $J_{2p}(\rho)$ $(1\leqslant p\leqslant N-1)$ of the regular class Casimir invariants are pth-order symmetric polynomials[27] of $l_1^2,l_2^2,\ldots,l_N^2$, while we have[13,23,24,26]

$$\hat{J}_N(\rho)=l_1l_2\cdots l_N. \tag{4.21}$$

Therefore, the N th-order index $\hat{D}^{(N)}(\rho)$ corresponding to $\hat{J}_N$ is simply given[11] by

$$\hat{D}^{(N)}(\rho)=l_1l_2\cdots l_N d(\rho). \tag{4.22}$$

The special case of $N=5$ reproduces the result of Eq. (3.13), although we changed the notation there simply as $D^{(5)}(\rho)$. Noting $l_N=f_N$, we also see that we have

$$\hat{D}^{(N)}(\rho')=-\hat{D}^{(N)}(\rho) \tag{4.23}$$

for two irreducible representations ρ and ρ' given by Eq. (4.19). Contrarily, we have

$$D^{(p)}(\rho')=D^{(p)}(\rho) \tag{4.24}$$

for the regular class indices.

In view of Eqs. (4.23) and (4.24) it suffices to consider only the cases of $f_N\geqslant 0$ for the regular indices $D^{(p)}(\rho)$. Again, first consider the case of tensor representations with Young tableau (4.1). The character formula for the SO$(2N)$ group for this case is then exactly the same[1,22] as for the SO$(2N+1)$ except for the face that U is now a $2N\times 2N$ defining representation of the SO$(2N)$. Therefore, for a fixed Young tableau $\Gamma=(f_1,f_2,\ldots,f_k)$ with a fixed value of k, the formula for $D^{(p)}(\rho)$ remains exactly the same and is given by Eqs. (4.8) and (4.9) except that we have to replace $2N$ by $2N-1$ in Eqs.

(4.13)–(4.15). Note that this replacement is consistent also from the transition of l_j defined by Eq. (4.17) to that of Eq. (4.20). Once we have shown this, the general case involving the spinor representations can be dealt with just as in the SO($2N+1$) case, simply by setting $k=N$ in the formula thus obtained. We should remark that the replacement of N by $N-\frac{1}{2}$ from the SO($2N+1$) to SO($2N$) group should be made before we set $k=N$. Otherwise, we will obtain incorrect results. We can verify these facts more directly for dimensional formula of $d(\rho)$ as well as for explicit formula of $J_2(\rho)$ and $J_4(\rho)$ given in Ref. 13.

For the SO($2N$) group, we note $\hat{D}^{(N)}(\square)=0$ for the basic representation $\square$. Also, we shall give here expressions of $D^{(p)}(\Lambda_f)$ and $D^{(p)}(f\Lambda_1)$ of SO($2N$), corresponding to totally antisymmetric and totally symmetric irreducible representations to be

$$Q_p(f\Lambda_1)=\sum_{l=0}^{f}\left\{\binom{2N+l-1}{l}-\binom{2N+l-3}{l-2}\right\}(f-l)^{p-1}, \tag{4.25a}$$

$$Q_p(\Lambda_f)=-\sum_{l=0}^{f}(-1)^{f-l}\binom{2N}{l}(f-l)^{p-1}, \tag{4.25b}$$

where $D^{(p)}(\rho)=Q_p(\rho)D^{(p)}(\square)$, again.

5. SIXTH-ORDER INDEX

We have computed $D^{(p)}(\rho)$ for all simple classical Lie algebras A_N, B_N, C_N, and D_N in the previous section. Unfortunately, the same technique used there is not applicable to five exceptional Lie algebras G_2, F_4, E_6, E_7, and E_8, since analogous simple characteristic formulas for these algebras are not available. We know that all these Lie algebras possess nontrivial second-order index $D^{(2)}(\rho)$, but not the fourth-order one $D^{(4)}(\rho)$. E_6 alone has nonzero fifth-order index $D^{(5)}(\rho)$ which can be computed as in I. The next higher sixth-order index $D^{(6)}(\rho)$ is however nontrivial for G_2, F_4, E_6, and E_7 but not for E_8. It is of some interest both theoretically and practically to evaluate $D^{(6)}(\rho)$ for these algebras. For this purpose, we have to utilize the method given in Refs. 13 and I. Below, we shall sketch a procedure which will enable us to compute $D^{(6)}(\rho)$. As a by-product, we find a sixth-order trace identity for those algebras.

Let ρ be a generic irreducible representation of a simple Lie algebra L.

Let λ be the reference representation[13,26] which we may identify with the lowest-dimensional basic representation of L. Let x_μ be the representation matrix of the basis t_μ of L in the representation λ, and set

$$h_{\mu_1\cdots\mu_p}=\frac{1}{p!}\sum_P \mathrm{Tr}(x_{\mu_1}x_{\mu_2}\cdots x_{\mu_p}) \tag{5.1}$$

as in I with normalization condition[13]

$$g_{\mu\nu}=h_{\mu\nu}=\mathrm{Tr}(x_\mu x_\nu). \tag{5.2}$$

Then, a pth-order Casimir invariant $I_p(\rho)$ can be defined by

$$I_p(\rho)=h^{\mu_1\mu_2\cdots\mu_p}X_{\mu_1}X_{\mu_2}\cdots X_{\mu_p}, \tag{5.3}$$

where X_μ is the representation matrix of t_μ in ρ. Hereafter, we assume for simplicity that the algebra L is *not* of type A_N ($N\geqslant 2$). We can construct the sixth-order fundamental Casimir invariant $J_6(\rho)$ as follows. First we set

$$S(\rho)=\frac{1}{8+d(\rho_0)}\left\{\frac{d(\rho_0)}{d(\rho)}-\frac{1}{3}\frac{D^{(2)}(\rho_0)}{D^{(2)}(\rho)}+\frac{1}{30}\frac{D^{(4)}(\rho_0)}{D^{(4)}(\rho)}\right\}, \tag{5.4}$$

$$T(\rho)=\frac{1}{4+d(\rho_0)}\left\{\left[\frac{d(\rho_0)}{d(\rho)}\right]^2-\frac{1}{2}\frac{d(\rho_0)}{d(\rho)}\frac{D^{(2)}(\rho_0)}{D^{(2)}(\rho)}+\frac{1}{12}\left[\frac{D^{(2)}(\rho_0)}{D^{(2)}(\rho)}\right]^2\right\}, \tag{5.5}$$

and moreover, define

$$a_{\mu\nu\lambda\alpha\beta\gamma}=\frac{1}{2!4!}\sum_P g_{\mu\nu}g_{\lambda\alpha\beta\gamma}, \tag{5.6}$$

$$b_{\mu\nu\lambda\alpha\beta\gamma}=\frac{1}{(2!)^3 3!}\sum_P g_{\mu\nu}g_{\lambda\alpha}g_{\beta\gamma}, \tag{5.7}$$

where ρ_0 is again the adjoint representation of L.

Here, the summation is over all 6! permutations P of six indices $\mu,\nu,\lambda,\alpha,\beta$, and γ so that the right sides of both Eqs. (5.6) and (5.7) consist of 15 different terms. The explicit form of $g_{\lambda\alpha\beta\gamma}$ is found in Ref. 13. We now set for any simple Lie algebra other than the type $A_N(N\geqslant 2)$,

$$g_{\mu\nu\lambda\alpha\beta\gamma}=[2+d(\rho_0)]h_{\mu\nu\lambda\alpha\beta\gamma}-S(\lambda)a_{\mu\nu\lambda\alpha\beta\gamma}-T(\lambda)b_{\mu\nu\lambda\alpha\beta\gamma}, \tag{5.8}$$

which can be shown to satisfy the orthogonality condition (1.9c). Then, the sixth-order fundamental Casimir invariant $J_6(\rho)$ is now given by

$$J_6(\rho)=g^{\mu\nu\lambda\alpha\beta\gamma}X_\mu X_\nu X_\lambda X_\alpha X_\beta X_\gamma. \tag{5.9}$$

Let X be as in Eq. (2.2). Then, following the procedure of Ref. 13, we can prove the sixth-order trace identity

$$[2+d(\rho_0)]\mathrm{Tr}\,X^6-15[2+d(\rho_0)]S(\rho)\{\mathrm{Tr}\,X^4-K(\rho)(\mathrm{Tr}\,X^2)^2\}\mathrm{Tr}\,X^2-15T(\rho)(\mathrm{Tr}\,X^2)^3=C_6(t)D^{(6)}(\rho) \tag{5.10}$$

for any simple Lie algebra other than A_N ($N\geqslant 2$), D_4 and D_6, where $C_6(t)$ depends upon the generic element $t\in L$ but not upon ρ. Also, we have set[13,28]

$$K(\rho)=\frac{1}{2(2+d(\rho_0))}\left\{6\frac{d(\rho_0)}{d(\rho)}-\frac{D^{(2)}(\rho_0)}{D^{(2)}(\rho)}\right\}. \tag{5.11}$$

We now restrict ourselves to Lie algebras A_1, G_2, F_4, E_6, E_7, and E_8, where we have identity[28]

$$\mathrm{Tr}\,X^4-K(\rho)(\mathrm{Tr}\,X^2)^2=0. \tag{5.12}$$

Of course this fact is related[13] to identical vanishing of $D^{(4)}(\rho)$ for these algebras. At any rate we then have

$$[2+d(\rho_0)]\mathrm{Tr}\,X^6-15T(\rho)(\mathrm{Tr}\,X^2)^3=C_6(t)D^{(6)}(\rho) \tag{5.13}$$

for all these Lie algebras. Correspondingly,[13] we define $\bar{I}_6(\rho)$ by

$$\bar{I}_6(\rho)=I_6(\rho)-\frac{T(\rho)}{2+d(\rho_0)}\frac{(r+2)(r+4)}{r^2}[I_2(\rho)]^3, \tag{5.14}$$

where r is the rank of L and $I_{2p}(\rho)$ is defined by Eq. (1.14), i.e.,

$$I_{2p}(\rho)=\sum_M (M,M)^p. \tag{5.15}$$

Following the same reasoning given in Ref. 13 to establish the relationship between $\bar{I}_4(\rho)$ and $D^{(4)}(\rho)$, we can now find

$$\bar{I}_6(\rho) = \text{const} \times D_6(\rho). \tag{5.16}$$

Expressing $T(\rho)$ in terms of $I_2(\rho)$, we can rewrite $\bar{I}_6(\rho)$ as

$$\begin{aligned}\bar{I}_6(\rho) = I_6(\rho) &- \frac{1}{12}\,\frac{(r+2)(r+4)}{r^2}\,\frac{1}{[2+d(\rho_0)][4+d(\rho_0)]} \\ &\times\left\{12\left[\frac{d(\rho_0)I_2(\rho)}{d(\rho)}\right]^2\right. \\ &\left. - 6\,\frac{d(\rho_0)I_2(\rho_0)}{d(\rho)}\,I_2(\rho) + [I_2(\rho_0)]^2\right\} I_2(\rho).\end{aligned} \tag{5.17}$$

Therefore, if we compute $\bar{I}_6(\rho)$ by Eq. (5.17), we can then numerically evaluate $D_6(\rho)$ apart from the normalization constant for exceptional Lie algebras G_2, F_4, E_6, and E_7. Similarly, let H_j $(j = 1, 2,..., r)$ be elements of a Cartan subalgebra of L, and let v^j $(j = 1, 2,..., r)$ be an arbitrary but fixed vector in the root space. If we choose

$$X = \sum_{j=1}^{r} v^j H_j$$

in Eq. (5.13), it gives

$$[2 + d(\rho_0)]I'_6(\rho) - 15T(\rho)[I'_2(\rho)]^3 = \text{const} \times D^{(6)}(\rho), \tag{5.18}$$

where we have set[6]

$$I'_p(\rho) = \sum_M (v, M)^p \tag{5.19}$$

as in I. From Eqs. (5.18) and (5.19) we may numerically compute $D^{(6)}(\rho)$ again.

For Lie algebras A_1 and E_8, both $\bar{I}_6(\rho)$ and $D^{(6)}(\rho)$ should be identically zero, since these possess no fundamental sixth-order Casimir invariants.[12] Especially Eq. (5.13) requires the validity of

$$[2 + d(\rho_0)]\text{Tr}\,X^6 - 15T(\rho)(\text{Tr}\,X^2)^3 = 0 \tag{5.20}$$

for any generic element X of A_1 and E_8. For A_1, the relation Eq. (5.20) or $\bar{I}_6(\rho) = 0$ gives an identity

$$\begin{aligned}\sum_{m=-j}^{j} m^6 = \frac{1}{21} j(j+1)(2j+1)\{3[j(j+1)]^2 \\ - 3j(j+1) + 1\}.\end{aligned} \tag{5.21}$$

for all positive integer and half-integer values of j, i.e., for $j = \frac{1}{2}, 1, \frac{3}{2}, 2,....$ Equation (5.21) can be more directly verified. We remark that the validity of Eq. (5.20) for E_8 has been previously noted elsewhere.[29]

The numerical evaluation of $\bar{I}_6(\rho)$ and hence of $D^{(6)}(\rho)$ for G_2, F_4, E_6, and E_7 will be reported elsewhere.

Note added in proof: After this paper had been completed, it came to our attention that there are additional references which are also of some relevance to the subject matter of this paper.[30–35] The present authors would like to express our gratitude to Professor R. C. King for calling our attention to these references.

ACKNOWLEDGMENTS

One of the authors (S. O.) is supported in part by the U.S. Department of Energy under contract No. DE-AC02-76ER-13605. The other (J. P.) is supported by the Natural Science and Engineering Research Council of Canada and by the Ministère de l'Education du Québec.

APPENDIX

Here, we shall give formulas for $d(\Gamma)$, $Q_p(\Gamma)$, and $R_4(\Gamma)$ of Sec. 2, when we have $k = 2$ with Γ being specified by two integers f_1 and f_2 satisfying

$$f_1 \geqslant f_2 \geqslant 0. \tag{A1}$$

It is convenient to set

$$\theta_1 = f_1 + 1, \quad \theta_2 = f_2. \tag{A2}$$

Then, from Eqs. (2.36)–(2.39), we calculate

$$d(\Gamma) = (\theta_1 - \theta_2)\,\frac{(N+\theta_1-2)!(N+\theta_2-2)!}{(N-1)!(N-2)!\theta_1!\theta_2!}, \tag{A3}$$

$$\begin{aligned}Q_p(\Gamma) = \frac{1}{(N-1)!}\Bigg\{&\frac{(N+\theta_1-2)!}{(\theta_1-1)!}\beta^{(p)}(\theta_2) \\ &- \frac{(N+\theta_1-1)!}{\theta_1!}\beta^{(p)}(\theta_2-1) \\ &+ \frac{(N+\theta_2-1)!}{\theta_2!}\beta^{(p)}(\theta_1-1) \\ &- \frac{(N+\theta_2-2)!}{(\theta_2-1)!}\beta^{(p)}(\theta_1)\Bigg\},\end{aligned} \tag{A4}$$

$$\begin{aligned}R_4(\Gamma) = &\frac{1}{(N+3)!(N-1)!}\,\frac{(N+\theta_1)!(N+\theta_2-2)!}{(\theta_1-2)!\theta_2!} \\ &\times\{(N-1)(\theta_1-2) - (N+3)\theta_2\} \\ &+ \frac{1}{(N+3)!(N-1)!}\,\frac{(N+\theta_1-2)!(N+\theta_2)!}{\theta_1!(\theta_2-2)!} \\ &\times\{(N+3)\theta_1 - (N-1)(\theta_2-2)\} \\ &+ 2(\theta_1-\theta_2)\,\frac{(N+\theta_1-1)!(N+\theta_2-1)!}{N!(N+1)!(\theta_1-1)!(\theta_2-1)!},\end{aligned} \tag{A5}$$

which are odd functions of θ_1 and θ_2 when we interchange them. Then, all formulas of Sec. 2 can be obtained as special cases of these formulas when we also utilize Eqs. (2.47) and (2.50). Especially for $f_1 = f$ and $f_2 = 1$, we calculate

$$\begin{aligned}R_4(f,1) = &\frac{(N+f)!}{(N+3)!(f-1)!}\{fN^2 \\ &+ (f^2 - 2f + 6)N - (f+5)(f-2)\},\end{aligned} \tag{A6}$$

$$\begin{aligned}Q_p(f,1) &= \frac{(N+f-1)!}{(N-1)!f!} + N\beta^{(p)}(f) - \beta^{(p)}(f+1) \\ &= N\frac{(N+f-2)!}{(f-1)!(N-1)!} - (f+1)^{p-1} \\ &\quad + \sum_{l=1}^{f-2}\frac{(N+l-1)!}{(l+1)!(N-2)!}\,l(f-l)^{p-1},\end{aligned} \tag{A7}$$

which for example gives for $f = 3$,

$$Q_p(3,1) = \tfrac{1}{2}(N + 2^{p-1})(N^2 + N - 2^p).$$

The function $\beta^{(p)}(f)$ defined by Eq. (2.33) may be rewritten as

$$\beta^{(p)}(f) = \sum_{l=1}^{p-1}\alpha_l^{(p)}\binom{N+p+f-l-1}{f-l}, \tag{A8}$$

$$\tilde{\beta}^{(p)}(f) = -\sum_{l=1}^{p-1}(-1)^l\alpha_l^{(p)}\binom{N-p}{f-l}, \tag{A9}$$

where $\alpha_l^{(p)}$ ($p \geqslant 2$) are determined by the generating function of the form

$$\sum_{l=1}^{p-1} \alpha_l^{(p)} z^l = (1-z)^p \left(z \frac{d}{dz}\right)^{p-1} \frac{1}{1-z} = (1-z)^p \sum_{l=0}^{\infty} l^{p-1} z^l. \tag{A10}$$

For $p \leqslant 5$, Eqs. (A8) and (A10) lead to Eq. (3.5), while for $p = 6$ and 7, we compute

$$\begin{aligned} p=6: \quad & \alpha_1^{(6)} = \alpha_5^{(6)} = 1, \\ & \alpha_2^{(6)} = \alpha_4^{(6)} = 26, \\ & \alpha_3^{(6)} = 66, \end{aligned} \tag{A11}$$

$$\begin{aligned} p=7: \quad & \alpha_1^{(7)} = \alpha_6^{(7)} = 1, \\ & \alpha_2^{(7)} = \alpha_5^{(7)} = 57, \\ & \alpha_3^{(7)} = \alpha_4^{(7)} = 302. \end{aligned} \tag{A12}$$

Using the formulas given here together with Eqs. (2.47) and (2.50), we can compute all $Q_p[\rho^5(\Gamma)]$ and $R_4[\rho^5(\Gamma)]$ for $n = 5$ except for the Young tableau Γ_0 specified by

$$\Gamma_0 = \begin{array}{|c|c|c|}\hline \multicolumn{1}{|c}{} & \multicolumn{1}{|c}{} & \multicolumn{1}{|c|}{} \\ \hline \multicolumn{1}{|c}{} & \multicolumn{1}{|c|}{} \\ \cline{1-2} \end{array}, \quad f_1 = 3, \quad f_2 = f_3 = 1. \tag{A13}$$

Because of this, we also calculate here the corresponding formulas for this case to be

$$d(\Gamma_0) = \tfrac{1}{20} N(N^2-1)(N^2-4), \tag{A14}$$

$$Q_2(\Gamma_0) = \tfrac{1}{4}(N^2-4)(N^2-5), \tag{A15}$$

$$Q_3(\Gamma_0) = \tfrac{1}{4}(N^4 - 17N^2 + 100), \tag{A16}$$

$$Q_4(\Gamma_0) = \tfrac{1}{4}(N^4 - 33N^2 + 500), \tag{A17}$$

$$R_4(\Gamma_0) = N(N^2 - 12). \tag{A18}$$

Finally, we shall make a comment on the Clebsch–Gordan decomposition. Let ρ_A and ρ_B be two representations of L and suppose that the product representation $\rho_A \otimes \rho_B$ will decompose as

$$\rho_A \otimes \rho_B = \sum_j \oplus \rho_j \tag{A19}$$

of a sum of representations ρ_j. We have shown in I that we have the sum rules

$$d(\rho_A) D^{(p)}(\rho_B) + d(\rho_B) D^{(p)}(\rho_A) = \sum_j D^{(p)}(\rho_j) \tag{A20}$$

as well as

$$d(\rho_A)\overline{D}^{(4)}(\rho_B) + d(\rho_B)\overline{D}^{(4)}(\rho_A) + 2\frac{d(\rho_0)+2}{d(\rho_0)} D^{(2)}(\rho_A) D^{(2)}(\rho_B) = \sum_j \overline{D}^{(4)}(\rho_j), \tag{A21}$$

where $\overline{D}^{(4)}(\rho)$ is defined by Eq. (2.53). We now consider the case of the $G = \mathrm{SU}(N)$ group, and apply the results of Sec. 2 to Eqs. (A20) (with $p = 4$) and (A21). Also, in conformity with the notation of Sec. 2, we use Young tableau symbols Γ_A, Γ_B, and Γ_j for corresponding irreducible representations ρ_A, ρ_B, and ρ_j, respectively, of the SU(N) group. In this way, we find the sum rule

$$d(\Gamma_A) R_4(\Gamma_B) + d(\Gamma_B) R_4(\Gamma_A) + 2 Q_2(\Gamma_A) Q_2(\Gamma_B) = \sum_j R_4(\Gamma_j) \tag{A22}$$

for the Clebsch–Gordan decomposition of the product of the Young tableaux Γ_A and Γ_B,

$$\Gamma_A \otimes \Gamma_B = \sum_j \oplus \Gamma_j \tag{A23}$$

of the SU(N) group. Noting $R_4(\square) = 0$ and $Q_2(\square) = 1$, we can calculate $R_4(\Gamma)$ for many Γ by repeated uses of Eq. (A22) from some known values of $R_4(\Gamma_A)$ and $R_4(\Gamma_B)$. Also, Eq. (A22) may be used as a check for the validity of the correct Clebsch–Gordan decomposition, Eq. (A23). Similarly, for any p satisfying $N \geqslant p$, we must have

$$d(\Gamma_A) Q_p(\Gamma_B) + d(\Gamma_B) Q_p(\Gamma_A) = \sum_j Q_p(\Gamma_j). \tag{A24}$$

[1] H. Weyl, *Classical Groups* (Princeton U. P., Princeton, NJ, 1939).
[2] M. Hammermesh, *Group Theory and its Applications to Physical Problems* (Addison-Wesley, Reading, MA, 1962).
[3] W. G. McKay and J. Patera, *Tables of Dimensions, Indices, and Branching Rules for Representations of Simple Lie Algebras* (Dekker, New York, 1981).
[4] H. Fritzsch and P. Minkowski, Ann. Phys. **93**, 193 (1975); H. Georgi, in *Particles and Fields*, edited by C. E. Carlson (AIP, New York, 1975).
[5] J. Patera and R. T. Sharp, J. Phys. A: Math. Gen. **13**, 347 (1980).
[6] J. Patera and R. T. Sharp, J. Math. Phys. **22**, 261 (1980).
[7] A. N. Schellekens, I.-G. Koh, and K. Kang, J. Math. Phys. **23**, 2244 (1982).
[8] J. Patera and R. T. Sharp, J. Math. Phys. **22**, 2352 (1981).
[9] J. McKay, J. Patera, and R. T. Sharp, J. Math. Phys. **22**, 2770 (1981).
[10] J. Patera, R. T. Sharp, and P. Winternitz, J. Math. Phys. **17**, 1972 (1976); erratum **18**, 1519 (1977). We have replaced the symbol $I^{(2p)}$ to $l_{2p}(\rho)$ here, however.
[11] S. Okubo and J. Patera, "General indices of representations and Casimir invariants," (to appear in J. Math. Phys.) This paper will hereafter be referred to as I.
[12] J. Dixmier, "*Enveloping Algebras*" (North-Holland, Amsterdam, 1977).
[13] S. Okubo, J. Math. Phys. **23**, 8 (1982).
[14] E. g., see D. E. Littlewood, *The Theory of Group Character* (Clarendon, Oxford, 1940), pp. 88 and 89.
[15] P. Cvitanović and A. D. Kennedy, Physica Scripta **26**, 5 (1982).
[16] J. Banks and H. Georgi, Phys. Rev. D **14**, 1159 (1976); S. Okubo, Phys. Rev. D **16**, 3528 (1977).
[17] D. Gross and R. Jackiw, Phys. Rev. D **6**, 477 (1972); C. Bouchiat, J. Illiopoulos, and P. Meyer, Phys. Lett. **35**, 519 (1972); H. Georgi and S. L. Glashow, Phys. Rev. D **6**, 429 (1972).
[18] P. H. Frampton and T. W. Kephart, University of North Carolina Reports IFP-184-UNC (1982) and IFP-191/192-UNC, 1983.
[19] F. W. Lemire and J. Patera, J. Math. Phys. **21**, 2026 (1980).
[20] A. N. Schellekens, K. Kang, and I.-G. Koh, Phys. Rev. D **26**, 658 (1982).
[21] Y. Tosa and R. E. Marshak, Phys. Rev. D **27**, 616 (1983).
[22] F. D. Murnaghan, *The Theory of Group Representations* (John Hopkins, Baltimore, 1938).
[23] J. D. Louck, "Theory of angular momentum in N-dimensional space," Los Alamos Scientific Lab, 1960 (unpublished).
[24] V. S. Popov and A. M. Perelomov, Yad. Fiz. **5**, 693 (1964) [Sov. J. Nucl. Phys. **5**, 489 (1964)].
[25] C. O. Nwachuku and M. A. Rashid, J. Math. Phys. **17**, 1611 (1976).
[26] S. Okubo, J. Math. Phys. **18**, 2382 (1977).
[27] This fact has been shown in Ref. 26 for the nonsymmetrized Casimir invariants $I_{2p}^{(N,S)}(\rho)$. However, the transition from $I_{2p}^{(N,S)}(\rho)$ to $J_{2p}(\rho)$ does not spoil this property as we see from the discussions of I and of Ref. 13.
[28] S. Okubo, J. Math. Phys. **20**, 586 (1979).
[29] Y. Tosa, R. E. Marshak, and S. Okubo, Phys. Rev. D **27**, 444 (1983).
[30] B. G. Wyburne and P. H. Butler, J. Phys. **30**, 181 (1969).
[31] P. H. Butler and B. G. Wyburne, At. Data **3**, 133 (1971).
[32] P. H. Butler and R. C. King, J. Math. Phys. **14**, 741 (1973).
[33] M. J. Englefield and R. C. King, J. Phys. A **13**, 2299 (1980).
[34] B. G. Wyburne and M. J. Bowick, Austr. J. Phys. **30**, 259 (1977).
[35] B. G. Wyburne, Auster. J. Phys. **34**, 417 (1979).

Computation of nonlinear behavior of Hamiltonian systems using Lie algebraic methods [a]

Alex J. Dragt[b] and Etienne Forest[b,c]
Center for Theoretical Physics, Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742

(Received 15 March 1983; accepted for publication 18 July 1983)

Lie algebraic methods are developed to describe the behavior of trajectories near a given trajectory for general Hamiltonian systems. A procedure is presented for the computation of nonlinear effects of arbitrarily high degree, and explicit formulas are given through effects of degree 5. Expected applications include accelerator design, charged particle beam and light optics, other problems in the general area of nonlinear dynamics, and, perhaps, with suitable modification, the area of S-matrix expansions in quantum field theory.

PACS numbers: 03.20. + i, 02.20.Sv, 02.30.Hq, 11.20. − e

1. INTRODUCTION

In many cases one knows a particular trajectory for a Hamiltonian system, and wishes to find the behavior, both linear and nonlinear, of trajectories near this known trajectory. For example, one knows the central design trajectory for an accelerator, and wishes to obtain information about the trajectories near the design trajectory.[1] Similarly, in the field of light optics (as well as charged particle beam optics) one knows the path of the axial ray, and wishes to characterize the behavior of rays near the axial ray.[2] Our purpose is to indicate how this problem can be treated with the aid of Lie algebraic methods. We expect that our results will have important applications both in accelerator design and optics, and in the general area of nonlinear dynamics. With suitable modification, there may also be useful applications in the area of S-matrix expansions in quantum field theory.

2. PRELIMINARY CALCULATIONS

To set the stage for further discussion, consider Hamiltonian motion in a $2n$-dimensional phase space with generalized coordinates $q_1 \cdots q_n$ and generalized momenta $p_1 \cdots p_n$. It is convenient to treat these quantities together by the introduction of a $2n$-vector z having the q's and p's as components,

$$(z_1 \cdots z_{2n}) = (q_1 \cdots q_n, p_1 \cdots p_n). \tag{2.1}$$

In this notation, we wish to deal with motion described by some Hamiltonian $H(z, t)$.

Now suppose that $z^g(t)$ is some *given* trajectory, which is assumed to be known, and that our task is to characterize all trajectories near z^g. Introduce $2n$ new variables ζ by the rule

$$z = z^g + \zeta. \tag{2.2}$$

The transformation (2.2) is canonical. Consequently, the time evolution of the new variables ζ will also be described by some Hamiltonian. Call this Hamiltonian $H^{\text{new}}(\zeta, t)$. Evidently, the problem of studying trajectories near z^g is equivalent to studying the trajectories governed by $H^{\text{new}}(\zeta, t)$ in the case where ζ is small.

What is the relation between $H(z, t)$ and $H^{\text{new}}(\zeta, t)$? Suppose that the quantity $H(z^g(t) + \zeta, t)$ is expressed as a power series in ζ by writing the expansion

$$H(z^g(t) + \zeta, t) = \sum_{m=0}^{\infty} H_m(\zeta, t). \tag{2.3}$$

Here each quantity $H_m(\zeta, t)$ is a homogeneous polynomial of degree m in the components of ζ. Then it is easily verified that $H^{\text{new}}(\zeta, t)$ is given by the expression

$$H^{\text{new}}(\zeta, t) = \sum_{m=2}^{\infty} H_m(\zeta, t). \tag{2.4}$$

3. LIE ALGEBRAIC TOOLS

The purpose of this section is to present a brief summary of the Lie algebraic tools and concepts required for our purpose. A more complete discussion may be found elsewhere.[1,3]

To begin, let f be a specified function on phase space, and let g be any function. Associated with each f is a *Lie operator* that acts on general functions g. The Lie operator associated with the function f will be denoted by the symbols $:f:$, and is defined by the rule

$$:f: g = [f, g]. \tag{3.1}$$

Here the square bracket [,] denotes the familiar Poisson bracket operation of classical mechanics.

Let the symbols $\{:f:,: g:\}$ denote the commutator of two Lie operators $:f:$ and $:g:$,

$$\{:f:,: g:\} = :f: :g: - :g: :f:. \tag{3.2}$$

Then it can be shown from the Jacobi identity for Poisson brackets that one has the relation

$$\{:f:,: g:\} = :[f, g]:. \tag{3.3}$$

That is, the commutator of two Lie operators is again a Lie operator, and this Lie operator can be calculated in terms of a Poisson bracket. This fact will be essential for later discussion.

Next consider the object $\exp(:f:)$, called a *Lie transfor-*

---

[a] From a dissertation to be submitted to the Graduate School, University of Maryland, by Etienne Forest in partial fulfillment of the requirements for the Ph.D. degree in Physics.

[b] Work supported in part by Department of Energy Contract AS05-80ER10666.

[c] Work also supported in part by National Sciences and Engineering Research Council of Canada.

 0022-2488/83/122734-11$02.50

mation, and defined by the exponential series

$$\exp(:f:) = \sum_{m=0}^{\infty} \frac{:f:^m}{m!}. \tag{3.4}$$

More explicitly, the action of $\exp(:f:)$ on any function g is given by the expression

$$\exp(:f:)\, g = g + [f, g] + [f,[f, g]]/2! + \cdots. \tag{3.5}$$

Now consider all phase-space trajectories generated by the Hamiltonian (2.4) that start at some *initial* time t^{in} and end at some *final* time t^{fin}. Let ζ^{in} denote an initial condition at the initial time t^{in}, and suppose the trajectory with this initial condition arrives at the point ζ^{fin} at the final time t^{fin}. Evidently, this following of trajectories assigns to each ζ^{in} a unique ζ^{fin}. Equivalently, we say that the Hamiltonian (2.4) gives rise to a mapping $\mathcal{M}$ (in general nonlinear) with the property

$$\zeta^{\text{fin}} = \mathcal{M}\zeta^{\text{in}}. \tag{3.6}$$

Observe that, by construction, $\zeta = 0$ is a trajectory. It follows that $\mathcal{M}$ maps the origin of phase space into itself. In this circumstance, there is a *factorization theorem* which shows that $\mathcal{M}$ can be written as an infinite product of Lie transformations in the form[4]

$$\mathcal{M} = \cdots \exp(:f_5:) \exp(:f_4:) \exp(:f_3:) \exp(:f_2:). \tag{3.7}$$

Here each function f_m is a homogeneous polynomial of degree m in the variables ζ^{in}.

Evidently, a knowledge of $\mathcal{M}$ is equivalent to a knowledge of the trajectories generated by the Hamiltonian (2.4). And, according to (3.7), a knowledge of $\mathcal{M}$ amounts to determining certain homogeneous polynomials f_2, f_3, f_4, etc.

With this brief background, we can be more specific about the contents of the remaining sections of this paper. Our aim is to derive explicit formulas for the polynomials f_2, f_3, etc., that characterize $\mathcal{M}$ in terms of the polynomials H_2, H_3, etc., that characterize H^{new}.[5] Sections 4–9 and an appendix develop various mathematical tools. Formulas for f_3 and f_4 are given in Sec. 10, and formulas for f_5 and f_6, as well as additional tools of use for the general case, are given in Secs. 11 and 12.

4. EQUATION OF MOTION FOR $\mathcal{M}$

For ease of notation, we will henceforth drop the superscript "new" and simply write H for $H^{\text{new}}(\zeta, t)$. Also, we will view t^{fin} as a variable time, and simply refer to it as the time t. Thus, we write (3.6) in the form

$$\zeta(t) = \mathcal{M}\zeta^{\text{in}}. \tag{4.1}$$

Suppose g is any function on phase space. Then, because $\mathcal{M}$ is a product of Lie transformations, it can be shown that $\mathcal{M}$ has the property[1,3]

$$g(\zeta) = g(\mathcal{M}\zeta^{\text{in}}) = \mathcal{M} g(\zeta^{\text{in}}). \tag{4.2}$$

Now take the time derivative (along a trajectory) of both sides of (4.2) to obtain the relation

$$\dot{g}(\zeta) = \dot{\mathcal{M}} g(\zeta^{\text{in}}). \tag{4.3}$$

From the equations of motion it follows that $\dot{g}$ is given also by the Poisson bracket equation

$$\dot{g}(\zeta) = [g(\zeta), H(\zeta, t)]. \tag{4.4}$$

Manipulate the right-hand side of (4.4) to obtain the relation

$$\begin{aligned}[g(\zeta), H(\zeta, t)] &= [g(\mathcal{M}\zeta^{\text{in}}), H(\mathcal{M}\zeta^{\text{in}}, t)] \\ &= [\mathcal{M} g(\zeta^{\text{in}}), \mathcal{M} H(\zeta^{\text{in}}, t)] \\ &= \mathcal{M}[g(\zeta^{\text{in}}), H(\zeta^{\text{in}}, t)] \\ &= \mathcal{M}[-H(\zeta^{\text{in}}, t), g(\zeta^{\text{in}})] \\ &= \mathcal{M}:-H(\zeta^{\text{in}}, t): g(\zeta^{\text{in}}).\end{aligned} \tag{4.5}$$

Here use has been made of the further property of $\mathcal{M}$ that it can be taken outside a Poisson bracket.[1,3]

Now compare Eqs. (4.3)–(4.5). One finds the result

$$\dot{\mathcal{M}} g(\zeta^{\text{in}}) = \mathcal{M}:-H(\zeta^{\text{in}}, t): g(\zeta^{\text{in}}). \tag{4.6}$$

Since the function g is arbitrary, it follows that $\mathcal{M}$ obeys the equation of motion

$$\dot{\mathcal{M}} = \mathcal{M}:-H(\zeta^{\text{in}}, t):. \tag{4.7}$$

5. SOLUTION IN THE COMMUTING CASE

Suppose the time interval $(t^{\text{in}}, t^{\text{fin}})$ is divided into N equal subintervals of duration Δt. Introduce intermediate times $t^{(m)}$ by the rules

$$t^{(m)} = t^{\text{in}} + m\Delta t, \quad m = 0, 1, \ldots N, \tag{5.1a}$$

$$t^{(0)} = t^{\text{in}}, \quad t^{(N)} = t^{\text{fin}}. \tag{5.1b}$$

Also, introduce the shorthand notation

$$H^{(m)} = H(\zeta^{\text{in}}, t^{(m)}). \tag{5.2}$$

Then, to lowest order in Δt, a Taylor expansion gives the result

$$\begin{aligned}\mathcal{M}(t^{(m+1)}) &= \mathcal{M}(t^{(m)} + \Delta t) = \mathcal{M}(t^{(m)}) + \dot{\mathcal{M}}(t^{(m)})\,\Delta t \\ &= \mathcal{M}(t^{(m)}) + \mathcal{M}(t^{(m)}):-H^{(m)}:\Delta t \\ &= \mathcal{M}(t^{(m)})(\mathcal{I} + :-H^{(m)}:\Delta t) \\ &= \mathcal{M}(t^{(m)}) \exp(:-H^{(m)}:\Delta t).\end{aligned} \tag{5.3}$$

Here $\mathcal{I}$ denotes the identity operator, and use has been made of the equation of motion (4.7).

Equation (5.3) can be solved sequentially to give the result

$$\begin{aligned}\mathcal{M}(t^{(N)}) = \mathcal{M}(t^{(0)}) &\exp(:-H^{(0)}:\Delta t) \exp(:-H^{(1)}:\Delta t) \\ &\times \cdots \exp(:-H^{(N-1)}:\Delta t).\end{aligned} \tag{5.4}$$

Thus, in view of (5.1b) and the fact that $\mathcal{M}(t^{\text{in}})$ must be the identity operator $\mathcal{I}$, we have to lowest order in Δt the formal solution

$$\begin{aligned}\mathcal{M} = &\exp(:-H^{(0)}:\Delta t) \exp(:-H^{(1)}:\Delta t) \\ &\cdots \exp(:-H^{(N-1)}:\Delta t).\end{aligned} \tag{5.5}$$

At this point it is possible to make two observations. First, suppose that for any two times t' and t'' the Hamiltonian H has the commuting property

$$\{:H(\zeta^{\text{in}}, t'):,:H(\zeta^{\text{in}}, t''):\} = 0, \tag{5.6}$$

or, equivalently, in view of (3.3), the property

$$[H(\zeta^{\text{in}}, t'), H(\zeta^{\text{in}}, t'')] = 0. \tag{5.7}$$

[Note that (5.7) will certainly be satisfied if H is time-independent.] Then the various exponents in (5.5) all commute, and therefore can be combined into one grand exponent to give, to lowest order in Δt, the result

$$\mathcal{M} = \exp(:-H^{(0)}:\Delta t + :-H^{(1)}:\Delta t + \cdots + :-H^{(N-1)}:\Delta t). \tag{5.8}$$

Upon taking the limits $N \to \infty, \Delta t \to 0$, we obtain the exact result

$$\mathcal{M} = \exp\left(-:\int_{t^{\text{in}}}^{t} H(\zeta^{\text{in}}, t')\,dt':\right). \tag{5.9}$$

Here we have again followed our convention of viewing t^{fin} as a variable time, and simply referring to it as the time t. Note that the correctness of (5.9) can also be checked directly, by simple differentiation with the aid of (5.6), to reproduce the equation of motion (4.7).

The second observation concerns the general noncommuting case. In this case too, by means of the Campbell–Baker–Hausdorff formula for manipulating noncommuting exponents, it is in principle possible to combine the various exponents in (5.4).[3] If this were done, the result would in general involve the various Lie operators $:-H^{(m)}:$ and all their various (multiple) commutators. That is, according to the Campbell–Baker–Hausdorff formula, products of Lie operators would occur *only* in the form of commutators. A direct calculation at this point is too awkward to carry out with our present tools. Subsequent sections will be devoted to the development of further tools, and the eventual treatment of the general noncommuting case. It is sufficient to observe here that the general case will involve (and *only* involve) exponentials of operators which are linear combinations of the operators $:-H(\zeta^{\text{in}}, t'):$ at various times t' and their multiple commutators.

Let us return, for the moment, to the commuting case. Suppose, as assumed, that H has the form (2.4). Then its integral can be written in the form

$$-\int_{t^{\text{in}}}^{t} H(\zeta^{\text{in}}, t')\,dt' = \sum_{m=2}^{\infty} h_m(\zeta^{\text{in}}, t), \tag{5.10}$$

where the quantities h_m are homogeneous polynomials of degree m in the variables ζ^{in}. Consequently, $\mathcal{M}$ in the commuting case has the form

$$\mathcal{M} = \exp\left(:\sum_{m=2}^{\infty} h_m:\right). \tag{5.11}$$

According to Sec. 3, our problem is to write $\mathcal{M}$ in the factored product form (3.7). That is, we need to reexpress (5.11) in the form (3.7). This too can be carried out with the aid of the Campbell–Baker–Hausdorff formula. A complete explicit solution for the f's in terms of the h's has been given, and it has been used extensively in the construction of the Lie algebraic computer code MARYLIE, designed for the computation of charged particle beam transport.[6,7] We will not elaborate further on specific results here because they will turn out to be a special instance of the general case to be treated in subsequent sections.

6. THE INTERACTION PICTURE

The discussion of the noncommutative case is facilitated by the introduction of an "interaction picture" similar to that employed in quantum mechanical calculations.[8]

According to (3.7), the map $\mathcal{M}$ can be written in the factored product form

$$\mathcal{M} = \cdots \mathcal{M}_5 \mathcal{M}_4 \mathcal{M}_3 \mathcal{M}_2 = \mathcal{M}_R \mathcal{M}_2, \tag{6.1}$$

where each factor $\mathcal{M}_m$ is given by the expression

$$\mathcal{M}_m = \exp(:f_m:). \tag{6.2}$$

Also, (2.4) states that H decomposes into a sum of homogeneous polynomials,

$$H = H_2 + H_3 + H_4 + \cdots = H_2 + H_R. \tag{6.3}$$

Here, as in (6.1), we use the subscript "R" to denote "remaining" terms.

Differentiating (6.1) gives the result

$$\dot{\mathcal{M}} = \dot{\mathcal{M}}_R \mathcal{M}_2 + \mathcal{M}_R \dot{\mathcal{M}}_2. \tag{6.4}$$

Now combine the equation of motion (4.7), the decompositions (6.1) and (6.3), and the relation (6.4) to obtain the result

$$\dot{\mathcal{M}}_R \mathcal{M}_2 + \mathcal{M}_R \dot{\mathcal{M}}_2 = \mathcal{M}_R \mathcal{M}_2 :-H_2 - H_R: = \mathcal{M}_R \mathcal{M}_2 :-H_2: + \mathcal{M}_R \mathcal{M}_2 :-H_R:. \tag{6.5}$$

Suppose, as will later be shown to be consistent, that $\mathcal{M}_2$ is required to satisfy the equation

$$\dot{\mathcal{M}}_2 = \mathcal{M}_2 :-H_2:. \tag{6.6}$$

It then follows from (6.5) and (6.6) that $\mathcal{M}_R$ obeys the equation

$$\dot{\mathcal{M}}_R \mathcal{M}_2 = \mathcal{M}_R \mathcal{M}_2 :-H_R:, \tag{6.7}$$

or, equivalently,

$$\dot{\mathcal{M}}_R = \mathcal{M}_R \mathcal{M}_2 :-H_R: \mathcal{M}_2^{-1}. \tag{6.8}$$

The quantity $\mathcal{M}_2 :-H_R: \mathcal{M}_2^{-1}$ occurring on the right-hand side of (6.8) can be simplified. We claim that

$$\mathcal{M}_2 :-H_R: \mathcal{M}_2^{-1} = :-H_R^{\text{int}}:, \tag{6.9}$$

where the "interaction" Hamiltonian H_R^{int} is given by the expression

$$H_R^{\text{int}}(\zeta^{\text{in}}, t) = \mathcal{M}_2 H_R(\zeta^{\text{in}}, t) = H_R(\mathcal{M}_2 \zeta^{\text{in}}, t). \tag{6.10}$$

If Eqs. (6.9) and (6.10) are accepted, then the equation of motion for $\mathcal{M}_R$ takes the final form

$$\dot{\mathcal{M}}_R = \mathcal{M}_R :-H_R^{\text{int}}:. \tag{6.11}$$

The verification of Eqs. (6.9) and (6.10) is straightforward, based on previously mentioned properties of Lie transformations. Suppose $\mathcal{M}$ is a Lie transformation, and let f and g be any two functions. Then one has the relation

$$\mathcal{M}:f:\mathcal{M}^{-1}g = \mathcal{M}[f, \mathcal{M}^{-1}g] = [\mathcal{M}f, \mathcal{M}\mathcal{M}^{-1}g] = [\mathcal{M}f, g] = :\mathcal{M}f:g. \tag{6.12}$$

Since the function g is arbitrary, (6.12) is equivalent to the operator identity

$$\mathcal{M}:f:\mathcal{M}^{-1} = :\mathcal{M}f:. \tag{6.13}$$

Now replace $\mathscr{M}$ and f in (6.13) by $\mathscr{M}_2$ and $(-H_R)$, respectively, in order to obtain the desired result.

Note that by construction $\mathscr{M}_R$ and H_R involve polynomials only of degree 3 and higher. It will be shown in a later section that the same is true for H_R^{int}. Thus, the equation of motion (6.11) for $\mathscr{M}_R$ involves polynomials only of degree 3 and higher.

7. COMPUTATION OF $\mathscr{M}_2$

Let $\bar{\zeta}(t)$ denote the result of applying $\mathscr{M}_2$ to ζ^{in}. Then, by definition, we have the relation

$$\bar{\zeta}(t) = \mathscr{M}_2 \zeta^{\text{in}} = \exp(:f_2:)\zeta^{\text{in}}, \tag{7.1}$$

or, equivalently,

$$\bar{\zeta}(t) = \zeta^{\text{in}} + :f_2:\zeta^{\text{in}} + (:f_2:^2/2!)\zeta^{\text{in}} + \cdots. \tag{7.2}$$

At this point it is useful to pause a moment in order to examine the degrees of various combinations of Lie operators and polynomials. Suppose f_l and f_m are any two homogeneous polynomials of degrees l and m, respectively. We shall let $d(f)$ denote the degree of f. Thus, with this notation, we write the relation

$$d(f_l) = l. \tag{7.3}$$

Next consider the operation of Poisson bracketing. Since this operation involves multiplication and two differentiations, we have the relation

$$d([f_l, f_m]) = l + m - 2 = d(f_l) + d(f_m) - 2, \tag{7.4}$$

or, equivalently,

$$d(:f_l:f_m) = d(f_l) + d(f_m) - 2. \tag{7.5}$$

Finally, consider a set of k homogeneous polynomials $f^1, f^2, \ldots, f^k$. Here the superscript merely labels the polynomial, and has nothing to do with its degree. Also, let g_m be a homogeneous polynomial of degree m. Then, trivial induction on (7.5) gives the result

$$d(:f^1::f^2:\cdots:f^k:g_m) = m - 2k + \sum_{i=1}^{k} d(f^i). \tag{7.6}$$

Now apply the results of the previous paragraph to the right-hand side of (7.2). Evidently, all terms are of degree 1. Consequently, (7.1) is a *linear* transformation which can be written in the component form

$$\bar{\zeta}_a(t) = \mathscr{M}_2 \zeta_a^{\text{in}} = \sum_b M_{ab}(t)\zeta_b^{\text{in}}, \tag{7.7}$$

or, more compactly, in the matrix form

$$\bar{\zeta} = \mathscr{M}_2 \zeta^{\text{in}} = M\zeta^{\text{in}}. \tag{7.8}$$

Thus, the computation of $\mathscr{M}_2$ is equivalent to finding the matrix M.

According to (6.6), the time evolution of $\mathscr{M}_2$, and consequently of M, is governed by H_2. Suppose that H_2 is written in the form

$$H_2(\zeta^{\text{in}}, t) = \frac{1}{2}\sum_{a,b} S_{ab}(t)\zeta_a^{\text{in}}\zeta_b^{\text{in}}. \tag{7.9}$$

Evidently S is a *symmetric* matrix. Next compute the quantity $:-H_2:\zeta^{\text{in}}$. One finds the result

$$:-H_2:\zeta_c^{\text{in}} = [\zeta_c^{\text{in}}, H_2] = \sum_{a,b} J_{ca} S_{ab} \zeta_b^{\text{in}}, \tag{7.10}$$

where the quantities J_{ca} are defined in terms of the fundamental Poisson brackets by the relation

$$J_{ca} \overset{\text{def}}{=} [\zeta_c^{\text{in}}, \zeta_a^{\text{in}}]. \tag{7.11}$$

Specifically, in matrix form, J is $2n \times 2n$ and is given by the equation

$$J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}. \tag{7.12}$$

Here each entry in J is an $n \times n$ matrix, I denotes the $n \times n$ identity matrix, and all other entries are zero. In computing (7.10), use has been made also of the symmetric property of S. Finally, using matrix notation, Eq. (7.10) can be written more compactly in the form

$$:-H_2:\zeta^{\text{in}} = JS\zeta^{\text{in}}. \tag{7.13}$$

Suppose both sides of (6.6) are applied to the quantity ζ^{in}. On the left-hand side one obtains the result

$$\dot{\mathscr{M}}_2 \zeta^{\text{in}} = \dot{\bar{\zeta}} = \dot{M}\zeta^{\text{in}}. \tag{7.14}$$

The right-hand side gives the result

$$\begin{aligned} \mathscr{M}_2 :-H_2:\zeta^{\text{in}} &= \mathscr{M}_2 JS\zeta^{\text{in}} \\ &= JS\mathscr{M}_2\zeta^{\text{in}} = JSM\zeta^{\text{in}}. \end{aligned} \tag{7.15}$$

This last step may require some elaboration. In terms of components one finds the relations

$$\begin{aligned} \mathscr{M}_2(JS\zeta^{\text{in}})_c &= \mathscr{M}_2 \sum_{a,b} J_{ca} S_{ab} \zeta_b^{\text{in}} \\ &= \sum_{a,b} J_{ca} S_{ab} \mathscr{M}_2 \zeta_b^{\text{in}} = \sum_{a,b} J_{ca} S_{ab} \bar{\zeta}_b \\ &= \sum_{a,b,d} J_{ca} S_{ab} M_{bd} \zeta_d^{\text{in}} = (JSM\zeta^{\text{in}})_c, \end{aligned} \tag{7.16}$$

where use has been made of (7.7).

Now compare the right-hand sides of Eqs. (7.14) and (7.15). They are both of degree 1 in ζ^{in}, and therefore (6.6) is consistent as advertised. Also, the matrix M must evidently obey the differential equation

$$\dot{M} = JSM. \tag{7.17}$$

Finally, the stipulation that $\mathscr{M}$ be the identity map $\mathscr{I}$ when $t = t^{\text{in}}$ requires, for consistency, that $\mathscr{M}_R$ and $\mathscr{M}_2$ also be the identity map when $t = t^{\text{in}}$. See Eq. (6.1). Consequently, the matrix M is subject to the initial condition

$$M(t^{\text{in}}) = I. \tag{7.18}$$

The differential equation (7.17) with the initial condition (7.18) has a unique solution whose computation, in most cases, requires numerical integration. In the special case when the matrices $JS(t')$ and $JS(t'')$ commute for all times t' and t'', one has, in analogy to (5.10) and (5.11), the explicit solution

$$M = \exp\left[\int_{t^{\text{in}}}^{t} JS(t')\,dt'\right]. \tag{7.19}$$

Indeed, it can be shown that the solution of (7.17) depends entirely upon the Lie algebra generated by the matrices

$JS(t)$.[9] In the even more special case that S (and therefore H_2) is time-independent, the integration required in (7.19) is immediate, and one obtains the result

$$M = \exp[(t - t^{in})\, JS]. \tag{7.20}$$

In the case (7.20) one may immediately write

$$\mathcal{M}_2 = \exp(:f_2:) \tag{7.21}$$

with f_2 given by the relation

$$f_2 = -(t - t^{in})\, H_2. \tag{7.22}$$

The determination of f_2 can also be carried out in the general case. However, since the explicit form of f_2 is not required for the work of this paper, we shall not pursue the matter further here.

8. COMPUTATION OF H_R^{int}

By definition, H_R consists of terms of degree 3 and higher,

$$H_R = H_3 + H_4 + \cdots. \tag{8.1}$$

Also, in view of (6.10) and the fact that $\mathcal{M}_2$ produces a linear transformation when acting on ζ^{in} [see (7.7)], it follows that H_R^{int} has the decomposition

$$H_R^{int} = H_3^{int} + H_4^{int} + \cdots, \tag{8.2}$$

where each term H_m^{int} is a homogeneous polynomial of degree m given by the relation

$$H_m^{int}(\zeta^{in}, t) = H_m(\mathcal{M}_2 \zeta^{in}, t). \tag{8.3}$$

To see how this works out in a specific case, consider the computation of H_3^{int}. The terms of still higher degree are handled analogously. Suppose that H_3 is written in the explicit form

$$H_3(\zeta^{in}, t) = \sum_{abc} T_{abc}(t)\, \zeta_a^{in}\, \zeta_b^{in}\, \zeta_c^{in}, \tag{8.4}$$

where T_{abc} is a set of (possibly time-dependent) coefficients. Then use of (8.3) gives the relation

$$H_3^{int}(\zeta^{in}, t) = \sum_{abc} T_{abc}(\mathcal{M}_2 \zeta^{in})_a (\mathcal{M}_2 \zeta^{in})_b (\mathcal{M}_2 \zeta^{in})_c. \tag{8.5}$$

However, thanks to (7.7), the terms on the right-hand side of (8.5) may be evaluated explicitly so that H_3^{int} can be expressed in the form

$$H_3^{int}(\zeta^{in}, t) = \sum_{\substack{abc \\ a'b'c'}} T_{abc}\, M_{aa'} M_{bb'} M_{cc'}\, \zeta_{a'}^{in}\, \zeta_{b'}^{in}\, \zeta_{c'}^{in}. \tag{8.6}$$

Finally, the sums in (8.6) can be grouped so that H_3^{int} can be written in the final form

$$H_3^{int}(\zeta^{in}, t) = \sum_{a'b'c'} T^{int}_{a'b'c'}(t)\, \zeta_{a'}^{in}\, \zeta_{b'}^{in}\, \zeta_{c'}^{in}, \tag{8.7}$$

where T^{int} is defined by the equation

$$T^{int}_{a'b'c'}(t) = \sum_{abc} T_{abc}(t)\, M_{aa'}(t)\, M_{bb'}(t)\, M_{cc'}(t). \tag{8.8}$$

Note that because of the time dependence of M, H_3^{int} is in general *time-dependent* even if H_3 is not.

9. COMPUTATION OF $\mathcal{M}_R$ BY ITERATION

Suppose that both sides of Eq. (6.11) are integrated with respect to the time from the initial time t^{in} to some final time t. Then one finds the result

$$\mathcal{M}_R(t) - \mathcal{M}_R(t^{in}) = \int_{t^{in}}^{t} dt'\, \mathcal{M}_R(t'):-H_R^{int}(t'):, \tag{9.1}$$

or, equivalently,

$$\mathcal{M}_R(t) = \mathcal{I} + \int_{t^{in}}^{t} dt'\, \mathcal{M}_R(t'):-H_R^{int}(t'):. \tag{9.2}$$

Here use has been made of the earlier result that $\mathcal{M}_R(t^{in})$ is the identity map $\mathcal{I}$.

Now iterate (9.2) by substituting the right-hand side back into the integral. If this is done once, we obtain the result

$$\mathcal{M}_R = \mathcal{I} + \int_{t^{in}}^{t} dt':-H_R^{int}(t'): + \int_{t^{in}}^{t} dt' \int_{t^{in}}^{t'} dt''\, \mathcal{M}_R(t''):-H_R^{int}(t''):::-H_R^{int}(t'):. \tag{9.3}$$

Evidently, repeated iteration gives the result

$$\mathcal{M}_R = \mathcal{I} + \int_{t^{in}}^{t} dt':-H_R^{int}(t'): + \int_{t^{in}}^{t} dt' \int_{t^{in}}^{t'} dt'':-H_R^{int}(t''):::-H_R^{int}(t'):+\cdots. \tag{9.4}$$

Note that in Eq. (9.4) the terms in the integrals occur in chronological order with earlier times preceding later times. We conclude that $\mathcal{M}_R$ can be expressed as an infinite sum of multiple time ordered integrals over the Lie operators $:-H_R^{int}(t'):$.

10. COMPUTATION OF f_3 AND f_4

By definition $\mathcal{M}_R$ has the factorization

$$\mathcal{M}_R = \cdots \exp(:f_5:) \exp(:f_4:) \exp(:f_3:). \tag{10.1}$$

The purpose of this section is to compare the two expressions (9.4) and (10.1) for $\mathcal{M}_R$ in order to obtain explicit formulas for f_3 and f_4. After the pattern of computation has been established, the determination of f_5, f_6, etc., will be treated in a subsequent section.

Suppose $\mathcal{M}_R$ as given by (10.1) is applied to some homogeneous polynomial g_m. Then, using the exponential expansion (3.4), we obtain the result

$$\begin{aligned}\mathcal{M}_R\, g_m &= \cdots(1 + :f_5: + \cdots)(1 + :f_4: + :f_4:^2/2! + \cdots) \\ &\quad \times (1 + :f_3: + :f_3:^2/2! + :f_3:^3/3! + \cdots)\, g_m \\ &= g_m + (:f_3: g_m) + (:f_3:^2/2!\, g_m + :f_4: g_m) + \cdots.\end{aligned} \tag{10.2}$$

Here, with the aid of (7.6), the various terms appearing in (10.2) have been grouped according to degree. Specifically, the degrees displayed are m, $(m+1)$, and $(m+2)$, respectively.

Next, suppose $\mathcal{M}_R$ as given by (9.4) is applied to the

same homogeneous polynomial g_m. Then, using the decomposition (8.2), we obtain the result

$$\mathscr{M}_R\, g_m = g_m + \left(\int_{t^{\text{in}}}^{t} dt' : -H_3^{\text{int}}: g_m \right) + \left(\int_{t^{\text{in}}}^{t} dt' : -H_4^{\text{int}}: g_m + \int_{t^{\text{in}}}^{t} dt' \times \int_{t^{\text{in}}}^{t'} dt'' : -H_3^{\text{int}}(t'') :: -H_3^{\text{int}}(t') : g_m \right) + \cdots . \tag{10.3}$$

Here again the terms appearing in (10.3) have been grouped according to degree.

Now compare terms in (10.2) and (10.3) of like degree. Also observe that g_m is arbitrary. It follows that we must have the relations

$$:f_3: = \int_{t^{\text{in}}}^{t} dt' : -H_3^{\text{int}}(t') :, \tag{10.4}$$

$$\frac{:f_3:^2}{2!} + :f_4: = \int_{t^{\text{in}}}^{t} dt' : -H_4^{\text{int}}: + \int_{t^{\text{in}}}^{t} dt' \times \int_{t^{\text{in}}}^{t'} dt'' : -H_3^{\text{int}}(t'') :: -H_3^{\text{int}}(t') :. \tag{10.5}$$

Both sides of Eq. (10.4) are manifestly Lie operators. Consequently, the colons can be removed from both sides of the equation to give for f_3 the explicit formula

$$f_3 = -\int_{t^{\text{in}}}^{t} dt'\, H_3^{\text{int}}(t'). \tag{10.6}$$

Note that both sides of (10.6) are of the same degree, as is required for consistency.

The determination of f_4 is somewhat more difficult. Upon solving (10.5) for $:f_4:$, we find the result

$$:f_4: = \int_{t^{\text{in}}}^{t} dt' : -H_4^{\text{int}}: - \frac{:f_3:^2}{2!} + \int_{t^{\text{in}}}^{t} dt' \int_{t^{\text{in}}}^{t'} dt'' : -H_3^{\text{int}}(t'') :: -H_3^{\text{int}}(t') :. \tag{10.7}$$

The first term appearing on the right-hand side of (10.7) is evidently a Lie operator. However, it is not so clear that the remaining terms also produce a Lie operator, although we know from the left-hand side of (10.7) that they must. Our task, therefore, is to manipulate the terms in question to see if they can be brought into the form of a Lie operator.

By using the explicit result (10.4) for $:f_3:$, the quantity $:f_3:^2/2!$ can be expressed in the form

$$\frac{:f_3:^2}{2!} = \frac{1}{2} \int_{t^{\text{in}}}^{t} dt' \times \int_{t^{\text{in}}}^{t} dt'' : -H_3^{\text{int}}(t'') :: -H_3^{\text{int}}(t') :. \tag{10.8}$$

Observe that the domain of integration in (10.8) consists of the two regions $t'' < t'$ and $t'' > t'$. If we split the integral into separate integrals over these regions and exchange integration variables in the region $t'' > t'$, then (10.8) can be rewritten in the form

$$\frac{:f_3:^2}{2!} = \frac{1}{2} \int_{t^{\text{in}}}^{t} dt' \int_{t^{\text{in}}}^{t'} dt'' (: -H_3^{\text{int}}(t'') :: -H_3^{\text{int}}(t') : + : -H_3^{\text{int}}(t') :: -H_3^{\text{int}}(t'') :). \tag{10.9}$$

Now substitute (10.9) into (10.7). We obtain the result

$$:f_4: = \int_{t^{\text{in}}}^{t} dt' : -H_4^{\text{int}}(t') : + \frac{1}{2} \int_{t^{\text{in}}}^{t} dt' \int_{t^{\text{in}}}^{t'} dt'' \times \{ : -H_3^{\text{int}}(t'') :, : -H_3^{\text{int}}(t') : \}. \tag{10.10}$$

In accord with our expectations from Sec. 5, the terms (10.7) involving H_3^{int} have been reexpressed as a commutator!

We are now ready for the last step. According to (3.3), the commutator of two Lie operators is again a Lie operator. Specifically, for the case in question, one has the relation

$$\{ : -H_3^{\text{int}}(t'') :, : -H_3^{\text{int}}(t') : \} = : [-H_3^{\text{int}}(t''), -H_3^{\text{int}}(t')] :. \tag{10.11}$$

It is now obvious how to remove the colons from both sides of (10.10). Thanks to (10.11), we find the final result

$$f_4 = -\int_{t^{\text{in}}}^{t} dt'\, H_4^{\text{int}}(t') + \frac{1}{2} \int_{t^{\text{in}}}^{t} dt' \int_{t^{\text{in}}}^{t'} dt'' \, [-H_3^{\text{int}}(t''), -H_3^{\text{int}}(t')]. \tag{10.12}$$

11. COMPUTATION OF TERMS BEYOND f_4

The purpose of this section is to extend the results of the previous section to compute the f_m of higher degree. The general problem is to reexpress various integrals over products of Lie operators as integrals involving only commutators so that the identity (3.3) can be used to remove colons. In order to minimize algebraic complexity in this process of "decolonization," it is necessary to extract the important features of the calculations we have just presented and to omit all irrelevant aspects. This will be done by showing that the entire calculation reduces to an exercise in the permutation of labels and involves only the commutators of symbols.

We begin by defining a generalized "bracket operator," denoted by the symbol $\langle {}^{i\ j\ \cdots\ m}_{\alpha\ \beta\ \cdots\ \epsilon} \rangle$, by the rule

$$\langle {}^{i\ j\ \cdots\ m}_{\alpha\ \beta\ \cdots\ \epsilon} \rangle \overset{\text{def}}{=} \int_{t^{\text{in}}}^{t} dt_1 \int_{t^{\text{in}}}^{t_1} dt_2 \cdots : -H_\alpha^{\text{int}}(t_i) :: -H_\beta^{\text{int}}(t_j) : \times \cdots : -H_\epsilon^{\text{int}}(t_m) :. \tag{11.1}$$

Here $t_i, t_j, \ldots, t_m$ is some permutation of the variables $t_1, t_2, \cdots$. In this notation, Eqs. (10.6) and (10.7) take the more compact forms

$$:f_3: = \langle {}^{1}_{3} \rangle, \tag{11.2}$$

$$:f_4: = \langle {}^{1}_{4} \rangle - \langle {}^{1}_{3} \rangle \langle {}^{1}_{3} \rangle / 2 + \langle {}^{2\ 1}_{3\ 3} \rangle. \tag{11.3}$$

The next step is to observe that products of bracket operators can be reexpressed in terms of sums of single bracket operators. For example, the second term in (11.3) can be rewritten in the form

$$\langle {}^{1}_{3} \rangle \langle {}^{1}_{3} \rangle = \langle {}^{1\ 2}_{3\ 3} \rangle + \langle {}^{2\ 1}_{3\ 3} \rangle. \tag{11.4}$$

This decomposition is simply a restatement that the region of integration can be split into the two regions $t_1 > t_2$ and $t_1 < t_2$. If we substitute (11.4) into (11.3), we immediately obtain the result

$$:f_4: = \langle {}^{1}_{4} \rangle + \langle {}^{2\ 1}_{3\ 3} \rangle/2 - \langle {}^{1\ 2}_{3\ 3} \rangle/2, \tag{11.5}$$

or, equivalently,

$$:f_4: = \langle {}^{1}_{4} \rangle + \{ {}^{2}_{3}, {}^{1}_{3} \}/2. \tag{11.6}$$

Here we have introduced the obvious notation $\{ {}^{2}_{3}, {}^{1}_{3} \}$ to denote the time ordered integral of a commutator. Equation (11.6) is identical to (10.10), and the colons can be removed immediately to give (10.12).

We can state the general rule as follows: Any product of bracket operators can be replaced by a sum of single operators. The lower indices in the single operators are identical to and in the same order as they were in the original expression. The upper indices consist of all possible permutations of time ordering consistent with whatever time ordering existed *within* the factors of the original expression. Suppose, for example, we are given the product

$$\langle {}^{j\ k}_{\alpha\ \beta} \rangle \langle {}^{l}_{\gamma} \rangle = \langle {}^{2\ 1}_{\alpha\ \beta} \rangle \langle {}^{1}_{\gamma} \rangle. \tag{11.7}$$

In accord with the previous discussion, this product can be decomposed into the sum of three single operators,

$$\langle {}^{2\ 1}_{\alpha\ \beta} \rangle \langle {}^{1}_{\gamma} \rangle = \langle {}^{2\ 1\ 3}_{\alpha\ \beta\ \gamma} \rangle + \langle {}^{3\ 1\ 2}_{\alpha\ \beta\ \gamma} \rangle + \langle {}^{3\ 2\ 1}_{\alpha\ \beta\ \gamma} \rangle. \tag{11.8}$$

Note that the relative time ordering of the α and β factors never violates the $j > k$ ordering of the original expression (11.7). In the simple case of (11.4), the original expression had no special time ordering, with the result that there was no preferential order for the decomposition.

Let us apply the tools developed so far to the computation of $:f_5:$. Consider first Eq. (10.2). When acting on g_m, the Lie operators giving rise to terms of degree $(m + 3)$ are given by the expression

$$:f_5: + :f_4: :f_3: + :f_3:^3/3!. \tag{11.9}$$

Next consider (10.3). Using the generalized bracket notation, when acting on g_m the factors giving rise to terms of degree $(m + 3)$ are given by the expression

$$\langle {}^{1}_{5} \rangle + \langle {}^{2\ 1}_{3\ 4} \rangle + \langle {}^{2\ 1}_{4\ 3} \rangle + \langle {}^{3\ 2\ 1}_{3\ 3\ 3} \rangle. \tag{11.10}$$

Comparison of (11.9) and (11.10) shows that $:f_5:$ must be given by the relation

$$:f_5: = \langle {}^{1}_{5} \rangle + \langle {}^{2\ 1}_{3\ 4} \rangle + \langle {}^{2\ 1}_{4\ 3} \rangle + \langle {}^{3\ 2\ 1}_{3\ 3\ 3} \rangle - :f_4: :f_3: - :f_3:^3/6. \tag{11.11}$$

Now use the previous results for $:f_3:$ and $:f_4:$ as given by (11.2) and (11.6). Substituting them into (11.11) gives the relation

$$:f_5: = (\langle {}^{1}_{5} \rangle) + (\langle {}^{2\ 1}_{3\ 4} \rangle + \langle {}^{2\ 1}_{4\ 3} \rangle - \langle {}^{1}_{4} \rangle \langle {}^{1}_{3} \rangle) + (\langle {}^{3\ 2\ 1}_{3\ 3\ 3} \rangle - \{ {}^{2}_{3}, {}^{1}_{3} \} \langle {}^{1}_{3} \rangle/2 - \langle {}^{1}_{3} \rangle \langle {}^{1}_{3} \rangle \langle {}^{1}_{3} \rangle/6). \tag{11.12}$$

Here we have collected the various terms into three groups in anticipation of the fact that each group can be decolonized separately.

The first group in (11.12) consists of a single term and is therefore already manifestly a Lie operator. The terms in the second group can be manipulated using the decomposition

$$\langle {}^{1}_{4} \rangle \langle {}^{1}_{3} \rangle = \langle {}^{1\ 2}_{4\ 3} \rangle + \langle {}^{2\ 1}_{4\ 3} \rangle. \tag{11.13}$$

It follows that the terms in the second group can be written as a commutator as anticipated:

$$\langle {}^{2\ 1}_{3\ 4} \rangle + \langle {}^{2\ 1}_{4\ 3} \rangle - \langle {}^{1}_{4} \rangle \langle {}^{1}_{3} \rangle = \langle {}^{2\ 1}_{3\ 4} \rangle - \langle {}^{1\ 2}_{4\ 3} \rangle = \{ {}^{2}_{3}, {}^{1}_{4} \}. \tag{11.14}$$

The conversion of the terms in the third group into commutators requires more work. One procedure is simply to combine products using the analog of (11.8), and then collect the resulting terms into multiple commutators. However, there is another procedure which is at once simpler and more profound.

Consider the set of all $(N - 1)$-fold commutators of N different abstract linear operators $O_1, O_2, \ldots, O_N$. It is shown in an appendix that this set has the remarkable properties listed below:

(1) The set forms a vector space of dimension $(N - 1)!$.

(2) The basis vectors for this vector space may all be taken to be *nests*. A nest is defined to be a multiple commutator of N objects that ends with $(N - 1)$ right braces. For example, $\{\alpha,\{\beta,\{\gamma,\delta\}\}\}$ is a nest, and $\{\{\alpha,\beta\},\{\gamma,\delta\}\}$ is not.

(3) In forming a basis, it is sufficient to use only those nests that end with a particular but arbitrary operator selected from the collection $O_1, O_2, \ldots, O_N$. Suppose, for example, that this operator is selected to be O_1. Then a basis is formed by the set of nests given by

$$\{\ldots, \{O_k, \{O_j, \{O_i, O_1\}\cdots\},$$

where the indices $\ldots, k, j, i$ are all possible permutations of the numbers $N, \ldots, 4, 3, 2$. Note that there are $(N - 1)!$ such nests in accord with property 1 above.

(4) The N-tuple $(\ldots, k, j, i, 1)$ will be referred to as the *defining string* of a nest. Defining strings play a special role. Consider an expression which is known to be decomposable into a set of $(N - 1)$-fold commutators of N objects. Suppose all nests ending with O_1 are used as a basis for the decomposition. Then, in this decomposition, the coefficient of the nest with the defining string $(\ldots, k, j, i, 1)$ will be the sum of all the coefficients of the factor $\cdots O_k O_j O_i O_1$ in the original expression.

Let us apply the results of the previous paragraph to the multiple commutator decomposition of the third term in (11.12). Since there are three operators in question, we must deal with the set of all twofold commutators of three operators. This set is of dimension 2, and a convenient set of basis vectors is given by the nests $\{ {}^{3}_{3}, \{ {}^{2}_{3}, {}^{1}_{3} \} \}$ and $\{ {}^{2}_{3}, \{ {}^{3}_{3}, {}^{1}_{3} \} \}$. Consider the defining string (3,2,1) for the first nest. Examine the third term in (11.12) to find the coefficients of the corresponding factor. Evidently $\langle {}^{3\ 2\ 1}_{3\ 3\ 3} \rangle$ contains the factor with a coefficient $(+1)$, $\{ {}^{2}_{3}, {}^{1}_{3} \} \langle {}^{1}_{3} \rangle$ when expanded and decomposed contains the factor with a coefficient of $(+1)$, and $\langle {}^{1}_{3} \rangle \langle {}^{1}_{3} \rangle \langle {}^{1}_{3} \rangle$ when decomposed contains the factor with a coefficient of $(+1)$. Thus the coefficient of the nest $\{ {}^{3}_{3}, \{ {}^{2}_{3}, {}^{1}_{3} \} \}$ in the decomposition is $1 + (1)(-1/2) + (1)(-1/6) = (1/3)$. Similarly, one finds the coefficient of the nest $\{ {}^{2}_{3}, \{ {}^{3}_{3}, {}^{1}_{3} \} \}$ is also $(1/3)$. Consequently, the third term in (11.12) can be rewritten in the form

$$\langle {}^{3}_{3}\,{}^{2}_{3}\,{}^{1}_{3}\rangle - \{{}^{2}_{3},{}^{1}_{3}\}\langle {}^{1}_{3}\rangle/2 - \langle {}^{1}_{3}\rangle\langle {}^{1}_{3}\rangle\langle {}^{1}_{3}\rangle/6$$
$$= \{{}^{3}_{3},\{{}^{2}_{3},{}^{1}_{3}\}\}/3 + \{{}^{2}_{3},\{{}^{3}_{3},{}^{1}_{3}\}\}/3. \tag{11.15}$$

Now substitute (11.14) and (11.15) into (11.12). We find the result

$$:f_5: = \langle {}^{1}_{5}\rangle + \{{}^{2}_{3},{}^{1}_{4}\} + \{{}^{3}_{3},\{{}^{2}_{3},{}^{1}_{3}\}\}/3 + \{{}^{2}_{3},\{{}^{3}_{3},{}^{1}_{3}\}\}/3, \tag{11.16}$$

or, after decolonization, the final result

$$f_5 = -\int_{t^{\text{in}}}^{t} dt_1\, H_5^{\text{int}}(t_1) + \int_{t^{\text{in}}}^{t} dt_1 \times \int_{t^{\text{in}}}^{t_1} dt_2\, [-H_3^{\text{int}}(t_2), -H_4^{\text{int}}(t_1)] + \frac{1}{3}\int_{t^{\text{in}}}^{t} dt_1 \int_{t^{\text{in}}}^{t_1} dt_2 \int_{t^{\text{in}}}^{t_2} dt_3 \times ([-H_3^{\text{int}}(t_3),[-H_3^{\text{int}}(t_2), -H_3^{\text{int}}(t_1)]] + [-H_3^{\text{int}}(t_2),[-H_3^{\text{int}}(t_3), -H_3^{\text{int}}(t_1)]]). \tag{11.17}$$

Note that in decolonizing the triple commutator we have used the relation

$$\{:f:,\{:g:,:h:\}\} = \{:f:,:[g,h]:\} = :[f,[g,h]]:, \tag{11.18}$$

which follows directly from (3.3).

Let us, in analogy with (11.1), introduce the shorthand notation

$$[{}^{i}_{\alpha}] \overset{\text{def}}{=} \int_{t^{\text{in}}}^{t} dt_1(-H_\alpha^{\text{int}}(t_1)), \tag{11.19a}$$

$$[{}^{i}_{\alpha},{}^{j}_{\beta}] = \int_{t^{\text{in}}}^{t} dt_1 \int_{t^{\text{in}}}^{t_1} dt_2[-H_\alpha^{\text{int}}(t_i), -H_\beta^{\text{int}}(t_j)], \quad \text{etc.} \tag{11.19b}$$

Then, Eqs. (10.6), (10.12), and (11.17) can be written more compactly in the form

$$f_3 = [{}^{1}_{3}], \tag{11.20}$$

$$f_4 = [{}^{1}_{4}] + [{}^{2}_{3},{}^{1}_{3}]/2, \tag{11.21}$$

$$f_5 = [{}^{1}_{5}] + [{}^{2}_{3},{}^{1}_{4}] + [{}^{3}_{3},[{}^{2}_{3},{}^{1}_{3}]]/3 + [{}^{2}_{3},[{}^{3}_{3},{}^{1}_{3}]]/3. \tag{11.22}$$

The reader should compare these results with their colonized versions given in (11.2), (11.6), and (11.16), respectively.

The calculation of $:f_6:$ may be carried out in similar fashion. One finds that it has the multiple commutator decomposition given by the formula

$$:f_6: = \langle {}^{1}_{6}\rangle + \{{}^{2}_{3},{}^{1}_{5}\} + \{{}^{2}_{4},{}^{1}_{4}\}/2 + (\{{}^{3}_{4},\{{}^{2}_{3},{}^{1}_{3}\}\} + \{{}^{2}_{4},\{{}^{3}_{3},{}^{1}_{3}\}\} + 3\{{}^{3}_{3},\{{}^{2}_{3},{}^{1}_{4}\}\} + \{{}^{2}_{3},\{{}^{3}_{3},{}^{1}_{4}\}\})/4 + (\{{}^{3}_{3},\{{}^{2}_{3},\{{}^{4}_{3},{}^{1}_{3}\}\}\} + \{{}^{4}_{3},\{{}^{2}_{3},\{{}^{3}_{3},{}^{1}_{3}\}\}\} + \{{}^{4}_{3},\{{}^{3}_{3},\{{}^{2}_{3},{}^{1}_{3}\}\}\})/4. \tag{11.23}$$

Decolonization gives the final result

$$f_6 = [{}^{1}_{6}] + [{}^{2}_{3},{}^{1}_{5}] + [{}^{2}_{4},{}^{1}_{4}]/2 + ([{}^{3}_{4},[{}^{2}_{3},{}^{1}_{3}]] + [{}^{2}_{4},[{}^{3}_{3},{}^{1}_{3}]] + 3[{}^{3}_{3},[{}^{2}_{3},{}^{1}_{4}]] + [{}^{2}_{3},[{}^{3}_{3},{}^{1}_{4}]])/4 + ([{}^{3}_{3},[{}^{2}_{3},[{}^{4}_{3},{}^{1}_{3}]]] + [{}^{4}_{3},[{}^{2}_{3},[{}^{3}_{3},{}^{1}_{3}]]] + [{}^{4}_{3},[{}^{3}_{3},[{}^{2}_{3},{}^{1}_{3}]]])/4. \tag{11.24}$$

12. IMMEDIATE DECOLONIZATION

There is an alternate complementary approach to the computation of the f_m that is worthy of mention. It has the feature that decolonization can be performed explicitly in the general case prior to subsequent calculation.

From the factorization (10.1) it follows that $\dot{\mathcal{M}}_R$ can be written in the form

$$\dot{\mathcal{M}}_R = \cdots + \cdots \dot{\mathcal{M}}_5 \mathcal{M}_4 \mathcal{M}_3 + \cdots \mathcal{M}_5 \dot{\mathcal{M}}_4 \mathcal{M}_3 + \cdots \mathcal{M}_5 \mathcal{M}_4 \dot{\mathcal{M}}_3. \tag{12.1}$$

Suppose (12.1) is substituted into the equation of motion (6.11) and both sides of the resulting relation are multiplied by $\mathcal{M}_R^{-1}$. Then we obtain the result

$$\cdots + \mathcal{M}_3^{-1}\mathcal{M}_4^{-1}\mathcal{M}_5^{-1}\dot{\mathcal{M}}_5\mathcal{M}_4\mathcal{M}_3 + \mathcal{M}_3^{-1}\mathcal{M}_4^{-1}\dot{\mathcal{M}}_4\mathcal{M}_3 + \mathcal{M}_3^{-1}\dot{\mathcal{M}}_3 = :-H_R^{\text{int}}:. \tag{12.2}$$

Next, let $\#f\#$ denote the *adjoint* of the Lie operator $:f:$. It is a kind of superoperator which acts on the general Lie operator $:g:$ according to the rule

$$\#f\#:g: = \{:f:,:g:\}. \tag{12.3}$$

Then, using (6.2), it can be shown that[1]

$$\mathcal{M}_3^{-1}\mathcal{M}_4^{-1}\mathcal{M}_5^{-1}\dot{\mathcal{M}}_5\mathcal{M}_4\mathcal{M}_3 = \exp(-\#f_3\#)\exp(-\#f_4\#)\mathcal{M}_5^{-1}\dot{\mathcal{M}}_5, \quad \text{etc.} \tag{12.4}$$

We now need to compute $\dot{\mathcal{M}}_m$. Again with the aid of adjoint operators, it can be shown that the formula for differentiating an exponential is given by the relation[3]

$$\mathcal{M}_m^{-1}\dot{\mathcal{M}}_m = \text{iex}(-\#f_m\#):\dot{f}_m:. \tag{12.5}$$

Here the *integrated exponential* function $\text{iex}(w)$ for general argument w is defined by the equations

$$\text{iex}(w) \overset{\text{def}}{=} \int_0^1 d\tau \exp(\tau w) = \sum_{m=0}^{\infty} \frac{w^m}{(m+1)!}. \tag{12.6}$$

Upon combining the fruits of our labor, we find that (12.2) can be rewritten in the form

$$\cdots + \exp(-\#f_3\#)\exp(-\#f_4\#)\,\text{iex}(-\#f_5\#):\dot{f}_5: + \exp(-\#f_3\#)\,\text{iex}(-\#f_4\#):\dot{f}_4: + \text{iex}(-\#f_3\#):\dot{f}_3: = :-H_R^{\text{int}}:. \tag{12.7}$$

The colons can now be removed immediately from both sides of (12.7) to give the result

$$\cdots + \exp(-:f_3:)\exp(-:f_4:)\,\text{iex}(-:f_5:)\dot{f}_5 + \exp(-:f_3:)\,\text{iex}(-:f_4:)\dot{f}_4 + \text{iex}(-:f_3:)\dot{f}_3 = -H_R^{\text{int}}. \tag{12.8}$$

Essentially all that is needed to pass from (12.7) to (12.8) is repeated use of relations of the form (3.3), (11.18), etc.

Suppose we examine (12.8) with the aim of equating terms of like degree. Using the expansion (12.6), we find the result

$$\text{iex}(-:f_m:)\dot{f}_m = (1 - :f_m:/2! + :f_m:^2/3! + \cdots)\dot{f}_m. \tag{12.9}$$

According to (7.6), the terms on the right-hand side of (12.9) have degree m, $2m-2$, $3m-4$, etc. Consequently, upon using (8.2), and equating terms of like degree in (12.8), we find the explicit formulas

$$\begin{aligned}
\dot{f}_3 &= -H_3^{\text{int}}, \\
\dot{f}_4 &= -H_4^{\text{int}} - P_4\,\text{iex}(-:f_3:)\dot{f}_3, \\
\dot{f}_5 &= -H_5^{\text{int}} - P_5\,\text{iex}(-:f_3:)\dot{f}_3 \\
&\quad - P_5\exp(-:f_3:)\,\text{iex}(-:f_4:)\dot{f}_4, \\
\dot{f}_6 &= -H_6^{\text{int}} - P_6\,\text{iex}(-:f_3:)\dot{f}_3 - P_6\exp(-:f_3:) \\
&\quad \times\text{iex}(-:f_4:)\dot{f}_4 - P_6\exp(-:f_3:) \\
&\quad \times\exp(-:f_4:)\,\text{iex}(-:f_5:)\dot{f}_5, \quad \text{etc.}
\end{aligned} \tag{12.10}$$

Here P_m denotes a projection operator which projects out terms of degree m.

Equations (12.10) may be solved and integrated successively to find the desired functions f_m. For example, upon integrating the first of equations (12.10), we find the expected result (11.20). Here we have used the condition that $f_m = 0$ when $t = t^{\text{in}}$, which corresponds to the condition $\mathcal{M}_R(t^{\text{in}}) = \mathcal{I}$.

Let us examine the second of Eqs. (12.10). Use of (7.6) and (12.9) gives the result

$$P_4\,\text{iex}(-:f_3:)\dot{f}_3 = -:f_3:\dot{f}_3/2. \tag{12.11}$$

Consequently, f_4 obeys the differential equation

$$\dot{f}_4 = -H_4^{\text{int}} + :f_3:(-H_3^{\text{int}})/2. \tag{12.12}$$

By making use of (11.2), or (11.20), we obtain the equivalent result

$$\dot{f}_4(t) = -H_4^{\text{int}}(t) + \langle{}_3^1\rangle(-H_3^{\text{int}}(t))/2. \tag{12.13}$$

Equation (12.13) may now be integrated directly. In doing so, it is convenient to let t_1 replace t as the variable of integration and, correspondingly, to increase the subscripts on all other variables and limits of integration by 1. When this is done, integration of (12.13) gives the previous result (11.21).

Similarly, Eqs. (11.22) and (11.24) can, after suitable manipulation, be obtained by integration, respectively, of the third and fourth of Eqs. (12.10). As a check of her or his understanding, the reader is encouraged to reproduce (11.22) by this method. In this connection identities such as (11.4) and (11.8) are again of use.

13. CONCLUDING REMARKS

We have shown how to compute the polynomials f_2, f_3, etc. that characterize the mapping $\mathcal{M}$ in terms of the polynomials H_2, H_3, etc. that characterize the Hamiltonian.[10] Explicit formulas have been given for the polynomials f_3–f_6, and general machinery has been developed for the computation of the polynomials of still higher degree should the need arise.

These results are expected to be of value in the fields of accelerator design and charged particle beam and light optics. They may also be of value in other areas of nonlinear dynamics. For example, suppose $z^g(t)$ is a periodic orbit. Then it can be shown that the stability of this orbit, when one goes beyond the linear approximation, is governed by the polynomials f_3, f_4, etc.[1,11]

It may also be remarked that expressions such as (11.17) are not as formidable as they may appear. Once the indicated Poisson brackets have been performed, the various monomials in the variables ζ^{in} may be taken outside the integral sign since they are, in fact, time independent. The integrals then involve only products of matrix elements of $M(t)$ and various coefficients such as those appearing, for example, in (8.4).

ACKNOWLEDGMENT

We are grateful to Richard K. Cooper for many helpful comments.

APPENDIX

The purpose of this appendix is to prove the assertions made in Sec. 11 about $(N-1)$-fold commutators of N different linear operators. We begin by showing that all such multiple commutators can be written as linear combinations of nests. The proof, which is by induction, will be presented in a somewhat discursive fashion in order to illustrate along the way various properties of multiple commutators.

Consider first the case $N = 2$ where there are only two operators O_1 and O_2. The possible commutators are $\{O_1, O_2\}$ and $\{O_2, O_1\}$. Each is obviously a nest. Also, from the antisymmetry condition we have the relation

$$\{O_1, O_2\} = -\{O_2, O_1\}. \tag{A1}$$

Thus, the two commutators are linearly dependent, the vector space they span is one-dimensional, and either nest may serve as a basis.

Next consider the case $N = 3$ where there are three operators O_1, O_2, and O_3. From these operators one may form 12 commutators which superficially appear to be distinct. They are the commutators of the form $\{X, \{Y, Z\}\}$ and $\{\{X, Y\}, Z\}$, where the roles of X, Y, and Z are filled by the six possible permutations of O_1, O_2, and O_3. However, the commutators of the form $\{\{X, Y\}, Z\}$ are related to those which are nests through the antisymmetry condition

$$\{\{X, Y\}, Z\} = -\{Z, \{X, Y\}\}. \tag{A2}$$

Thus in the case $N = 3$, all twofold commutators are again expressible in terms of nests.

How many nests are required to form a basis? All six possible nests for the case $N = 3$ are listed below:

$$\begin{aligned}
u_1 &= \{O_3, \{O_2, O_1\}\}, \quad u_4 = \{O_1, \{O_2, O_3\}\}, \\
u_2 &= \{O_1, \{O_3, O_2\}\}, \quad u_5 = \{O_2, \{O_3, O_1\}\}, \\
u_3 &= \{O_2, \{O_1, O_3\}\}, \quad u_6 = \{O_3, \{O_1, O_2\}\}.
\end{aligned} \tag{A3}$$

However, the antisymmetry condition gives the relations

$$u_4 = -u_2, \quad u_5 = -u_3, \quad u_6 = -u_1. \tag{A4}$$

Thus, at most only three nests are needed to form a basis. Moreover, the Jacobi identity gives the relation

$$u_1 + u_2 + u_3 = 0. \tag{A5}$$

Thus, the vector space of twofold commutators of three operators is two dimensional. A convenient basis is given by the nests u_1 and u_5. Note that both these nests end with the operator O_1.

We now proceed with induction. Let C_N denote the space of all $(N-1)$-fold commutators of N different linear operators. We assume that it is known that these spaces can be spanned by nests for all values of $N \leqslant M$. Now consider the space C_{M+1}. Let u be some element in C_{M+1}. Then since u is a multiple commutator, it can be written in the form

$$u = \{X, Y\}, \tag{A6}$$

where X and Y are members of C_j and C_k, respectively, and j and k are related by the equation

$$j + k = M + 1. \tag{A7}$$

Also observe that X and Y have no operators in common.

According to (A7), the quantities j and k are less than or equal to M. It follows by assumption that X and Y can be expressed in terms of nests. We denote these nests by $\mathcal{N}_l^a$. The upper index labels the nest, and the lower index labels the space C_l to which it belongs. Thus, we may rewrite (A6) in the form

$$u = \{X, Y\} = \left\{\sum_a \alpha_a \mathcal{N}_j^a, \sum_b \beta_b \mathcal{N}_k^b\right\}$$

$$= \sum_{a,b} \alpha_a \beta_b \{\mathcal{N}_j^a, \mathcal{N}_k^b\}. \tag{A8}$$

Since $\mathcal{N}_j^a$ is a nest, it can be written in the form

$$\mathcal{N}_j^a = \{O_d, \mathcal{N}_{j-1}^c\}, \tag{A9}$$

where O_d is a single operator in the $(M+1)$-fold collection $O_1, O_2, \ldots, O_{M+1}$. Apply the Jacobi identity to each term in (A8) using the representation (A9). This gives the result

$$\{\mathcal{N}_j^a, \mathcal{N}_k^b\} = \{\{O_d, \mathcal{N}_{j-1}^c\}, \mathcal{N}_k^b\}$$
$$= \{O_d, \{\mathcal{N}_{j-1}^c, \mathcal{N}_k^b\}\} - \{\mathcal{N}_{j-1}^c, \{O_d, \mathcal{N}_k^b\}\}. \tag{A10}$$

Consider the first term on the right-hand side of (A10). The commutator $\{\mathcal{N}_{j-1}^c, \mathcal{N}_k^b\}$ belongs to C_l with $l = j + k - 1 = M$. Thus, by the induction hypothesis, this commutator can be written as a sum of nests in the form

$$\{\mathcal{N}_{j-1}^c, \mathcal{N}_k^b\} = \sum_e \gamma_e \mathcal{N}_M^e. \tag{A11}$$

Correspondingly, the first term on the right-hand side of (A10) can be written in the form

$$\{O_d, \{\mathcal{N}_{j-1}^c, \mathcal{N}_k^b\}\} = \sum_e \gamma_e \{O_d, \mathcal{N}_M^e\}. \tag{A12}$$

But each term $\{O_d, \mathcal{N}_M^e\}$ is some nest $\mathcal{N}_{M+1}^f$. It follows that the first term on the right-hand side of (A10) can be expressed as a sum of nests in C_{M+1}.

What can be said about the second term on the right-hand side of (A10)? Evidently each term $\{O_d, \mathcal{N}_k^b\}$ is some nest $\mathcal{N}_{k+1}^g$. Thus, the terms on the right-hand side of (A10) which are not manifestly composed of nests are of the form $\{\mathcal{N}_{j-1}^c, \mathcal{N}_{k+1}^g\}$. Consequently, Eq. (A10) can be rewritten in the form

$$\{\mathcal{N}_j^a, \mathcal{N}_k^b\} = \{\mathcal{N}_{j-1}^c, \mathcal{N}_{k+1}^g\} + \sum_f \delta_f \mathcal{N}_{M+1}^f. \tag{A13}$$

Comparison of the left- and right-hand sides of (A13) shows that the value of j has been decreased by 1 and the value of k has been increased by 1. Evidently the operations which led to (A13) can be repeated at will until the only terms which are not manifestly nests are of the form $\{\mathcal{N}_1^h, \mathcal{N}_{j+k-1}^i\}$. But $\mathcal{N}_1^h$ must be one of the operators $O_1 \cdots O_{M+1}$, and, consequently, $\{\mathcal{N}_1^h, \mathcal{N}_{j+k-1}^i\}$ is also a nest. We have seen that all the terms on the right-hand side of (A8) can be written in terms of nests. Therefore, u itself can be written in terms of nests, and the induction process is complete.

At this point we suggest that the reader consider the case $N = 4$ and explicitly work out a decomposition into nests for the multiple commutator $\{\{O_1, O_2\}, \{O_3, O_4\}\}$.

We next show that only those nests ending with a particular operator O_l are needed to form a basis. Without loss of generality we choose $O_l = O_1$. Again we proceed by induction. As shown in the previous discussion, the assertion is already known to be true for $N = 2$ and $N = 3$.

Suppose the assertion is assumed to be true for all $N \leqslant M$. Let $\mathcal{N}_{M+1}$ be a nest formed from the operators $O_1 \cdots O_{M+1}$. Suppose the first entry in this nest is the operator O_i. That is, $\mathcal{N}_{M+1}$ can be written in the form

$$\mathcal{N}_{M+1} = \{O_i, \mathcal{N}_M\}, \tag{A14}$$

where, by construction, $\mathcal{N}_M$ is a nest composed of M operators different from O_i and each other. There are now only two possibilities. Either $O_i \neq O_1$ or $O_i = O_1$.

In the first possibility, $O_i \neq O_1$, $\mathcal{N}_M$ is a nest of M elements including the operator O_1. Since the assertion in question is assumed to be true for $N \leqslant M$, it follows that $\mathcal{N}_M$ can be written as a linear combination of nests all having O_1 as their last entry. Then, according to (A14), $\mathcal{N}_{M+1}$ can also be written as a linear combination of nests all having O_1 as their last entry, and the induction process is complete.

Suppose, according to the second possibility, that $O_i = O_1$. Let O_j and O_k be the last two entries of $\mathcal{N}_{M+1}$, respectively. That is, $\mathcal{N}_{M+1}$ is taken to have the form

$$\mathcal{N}_{M+1} = \{O_1, \{\ldots, \{O_j, O_k\}\} \cdots\}. \tag{A15}$$

Let V denote the commutator of O_j and O_k,

$$V = \{O_j, O_k\}. \tag{A16}$$

If V is regarded as a single operator, then (A15) shows that $\mathcal{N}_{M+1}$ may be viewed as a nest consisting of the $(M-1)$ operators O_l, with $O_l \neq O_j$ and $O_l \neq O_k$, and the operator V. From this perspective, $\mathcal{N}_{M+1}$ is a nest of M operators including the operator O_1. Therefore, according to the induction hypothesis, $\mathcal{N}_{M+1}$ may be written in the form

$$\mathcal{N}_{M+1} = \sum_a \alpha_a \mathcal{L}_M^a, \tag{A17}$$

where each $\mathscr{L}_M^a$ is a nest containing V as a single operator and having O_1 as its last entry. That is, each $\mathscr{L}_M^a$ is of the form

$$\mathscr{L}_M^a = \{O_m, \{..., \{V, \{..., \{O_n, O_1\}\} \cdots\}\}. \tag{A18}$$

Evidently, each $\mathscr{L}_M^a$ can also be written in the form

$$\mathscr{L}_M^a = \{O_m, \{..., \{V, \mathscr{N}\} \cdots\}, \tag{A19}$$

where $\mathscr{N}$ is either a nest ending with O_1, or is the single operator O_1 itself.

We are ready again to use the Jacobi identity, this time on the quantity $\{V, \mathscr{N}\}$. We find the result

$$\{V, \mathscr{N}\} = \{\{O_j, O_k\}, \mathscr{N}\}$$
$$= \{O_j, \{O_k, \mathscr{N}\}\} - \{O_k, \{O_j, \mathscr{N}\}\}. \tag{A20}$$

Observe that both terms on the right-hand side of (A20) are nests having O_1 as the last entry. Upon inserting (A20) into (A19), we see that each $\mathscr{L}_M^a$ can also be expressed as a linear combination of nests composed of the operators $O_1 \cdots O_{M+1}$ and all having O_1 as the last entry. It follows from (A17) that $\mathscr{N}_{M+1}$ also has this decomposition, and the induction process is again complete.

We have seen that any element of C_N can be expanded in terms of nests and that it is sufficient to use only those nests ending with O_1. The last assertion to verify is the linear independence of the nests with the defining strings $(...,k,j,i,1)$ when the quantities $...,k,j,i$ have as values all possible permutations of the indices $N,...,4,3,2$. The simplest proof follows from direct inspection.

Consider the nest with the defining string $(...,k,j,i,1)$. When written out in full, this nest has the expansion

$$\{..., \{O_k, \{O_j, \{O_i, O_1\}\} \cdots\}\}$$
$$= \text{other terms} + \cdots O_k O_j O_i O_1. \tag{A21}$$

The last term displayed on the right-hand side of (A21) will be called the *trailing term*. It has the property that it ends with O_1. Observe that there is only *one* term which ends with O_1, and therefore the trailing term is *uniquely* defined. That is, all other terms in the expansion (A21) have O_1 either at the beginning or buried somewhere in the middle.

Now suppose $(...,k',j',i',1)$ is some other defining string. Its corresponding nest has the expansion

$$\{..., \{O_{k'}, \{O_{j'} \{O_{i'}, O_1\}\} \cdots\}\}$$
$$= \text{other terms} + \cdots O_{k'} O_{j'} O_{i'} O_1. \tag{A22}$$

Evidently, the trailing terms of (A21) and (A22) are distinct unless $i' = i, j' = j, k' = k$, etc. It follows that the nests in question are linearly independent.

Even more can be said. Suppose u is some combination of products of operators, and suppose u is known to belong to C_N. Let us expand u in the basis set of nests corresponding to the defining strings $(...,k,j,i,1)$. Then the expansion coefficient for a given nest is the same as the coefficient of the trailing term of that nest as it appears in u.

[1] A. J. Dragt, "Lectures on nonlinear orbit dynamics," *Physics of High Energy Particle Accelerators*, AIP Conference Proceedings No. 87, edited by R. A. Carrigan *et al.* (Am. Inst. Phys., New York, 1982).

[2] A. J. Dragt, "Lie algebraic theory of geometrical optics and optical aberrations," J. Opt. Soc. Am. **72**, 372 (1982).

[3] A. J. Dragt and J. M. Finn, "Lie series and invariant functions for analytic symplectic maps," J. Math. Phys. **17**, 2215 (1976).

[4] See Refs. 1 and 3 above. Strictly speaking in the general case, as shown in Ref. 1, the last factor in (3.7) should be written as a product of two factors in the form $\exp(:f_2^a:)\exp(:f_2^c:)$. Here, f_2^a and f_2^c are polynomials of degree 2 having certain definite symmetry properties. However, this complication need not concern us here. Also, if desired, the factorization (3.7) can be performed (generally with different f's) in the reverse order.

[5] The expansions in (2.4) and (3.7) are organized according to degree. One could also consider the case where H is expanded in powers of some coupling constant ϵ. Then the analog of (3.7) would be an infinite product of factors whose exponents are proportional to successively higher powers of ϵ. Expansions organized in this manner may be of use for expressing the S matrix in quantum field theory. The formulas for these expansions will be similar to those obtained here.

[6] S. Steinberg, "Lie series, perturbation theory, and factored products," preprint, Department of Mathematics, University of New Mexico, Albuquerque, 1981.

[7] D. R. Douglas and A. J. Dragt, "Charged particle beam transport using Lie algebraic methods," IEEE Trans. Nucl. Sci. **NS-28**, 2522 (1981); D. R. Douglas, "Lie algebraic methods for particle accelerator theory," University of Maryland Ph.D. thesis, 1982, unpublished; D. R. Douglas and A. J. Dragt, "MARYLIE, the Maryland Lie algebraic beam transport and particle tracking program," IEEE Trans. Nucl. Sci. **NS-30**, 2442 (1983).

[8] A. L. Fetter and J. D. Walecka, *Quantum Theory of Many-Particle Systems* (McGraw-Hill, New York, 1971), p. 54.

[9] E. H. Wichmann, "A note on the algebraic aspect of the integration of a system of ordinary linear differential equations," J. Math. Phys. **2**, 876 (1961).

[10] This paper has given formulas for the f's assuming a factorization in the order (3.7). Similar formulas may be obtained for a factorization in the reverse order.

[11] A. J. Dragt, "A method of transfer maps for linear and nonlinear beam elements," IEEE Trans. Nucl. Sci. **NS-26**, 3601 (1979).

Generalized canonical transformations for time-dependent systems

Manuel Asorey, José F. Cariñena, and Luis A. Ibort
Departamento de Física Teórica, Facultad de Ciencias, Universidad de Zaragoza, Zaragoza, Spain

(Received 5 April 1983; accepted for publication 29 June 1983)

We introduce the concept of generalized canonical transformations as symplectomorphisms of the extended phase space. We prove that any such transformation factorizes in a standard canonical transformation times another one that changes only the time variable. The theory of generating functions as well as that of Hamilton–Jacobi is developed. Some further applications are developed.

PACS numbers: 03.20. + i, 02.40. + m

1. INTRODUCTION

Over the recent years there has been a considerable amount of literature devoted to the study of the applications in physics of differential geometry tools and in particular in classical mechanics. In the case of time-independent systems there are two alternative descriptions: both the Lagrangian and the Hamiltonian formalism are particular cases of the theory of locally Hamiltonian dynamical systems,[1] but in most cases the Hamiltonian formalism is easier to deal with because geometry—the symplectic structure—and dynamics—the Hamiltonian 1-form dH—are independent ingredients; on the other side, the Lagrangian function contains information on the symplectic structure as well as on the dynamical vector field. It makes it more difficult, for instance, to study symmetries of the geometry which are not symmetries of the dynamics and conversely.[2]

The geometrical framework for describing time-dependent systems is not so well established. The usual Hamiltonian description makes use of a contact structure which is built up from the Hamiltonian[1,3]; dynamics and geometry are coupled again, and for this reason the geometrical meaning of a canonical transformation is not clear. On the other hand, some transformations changing only the time have succeeded to deal with certain dynamical systems.[4–8] However, this kind of transformations has no relation to the canonical transformations defined in Refs. 1 and 3. The aim of this paper is to develop the concept of generalized canonical transformation for a time-dependent system in the framework of the extended phase space formalism. These generalized canonical transformations are but particular cases of symplectomorphisms of the extended phase space when it is endowed with the natural symplectic structure; this generalization covers the second kind of transformations considered above. Moreover, it will be shown that any generalized canonical transformation can be factorized as a product of a transformation of this kind times a standard canonical transformation.

The extended phase space formalism, which appears in a natural way in general relativity, reduces the problem of a time-dependent Hamiltonian to that of an autonomous dynamical system, and the dynamical evolution can be considered as a one-parameter group of symplectomorphisms. The transformation carrying the system to equilibrium will then be a symplectomorphism which allows development of the Hamilton–Jacobi theory.

The organization of the paper is as follows. In Sec. 2 a short description of the usual and the extended phase space formalisms, as well as a relation between them, is given. In particular, the reason for the H dependence of the usual contact form is shown to correspond to the different ways of identifying the contact manifold as a submanifold of the extended phase space. The concept and main properties of generalized canonical transformations are introduced in Sec. 3, as well as the standard case. Section 4 is devoted to the study of the generating functions of canonical transformations and to establishing the generalized Hamilton–Jacobi equation. Some applications are discussed in Sec. 5.

2. TIME-DEPENDENT SYSTEMS

The standard geometric description of time-dependent systems is carried out by means of a $2n + 1$ differentiable manifold $T^*Q \times \mathbb{R}$ as well as a Hamiltonian function $H \in \Lambda^0(T^*Q \times \mathbb{R})$, which enables us to define a contact structure on $T^*Q \times \mathbb{R}$[1,3]: the closed 2-form ω_H of rank $2n$ is defined by $\omega_H = \tau^*\omega_0 - dH \wedge dt$, where ω_0 is the canonical 2-form on T^*Q, t is the natural coordinate function on $\mathbb{R}$, and τ denotes the projection $\tau: T^*Q \times \mathbb{R} \to T^*Q$. The time-dependent vector field X_H giving the time evolution is then defined by

$$i(X_H)\omega_H = 0 \quad \text{and} \quad i(X_H)\,dt = 1, \tag{2.1}$$

i.e., if (q^i, p_i) are local canonical coordinates in T^*Q the expression of X_H in the corresponding coordinates in $T^*Q \times \mathbb{R}$ will be

$$X_H = \frac{\partial H}{\partial p_i}\frac{\partial}{\partial q^i} - \frac{\partial H}{\partial q^i}\frac{\partial}{\partial p_i} + \frac{\partial}{\partial t}. \tag{2.2}$$

In this section we will develop an extended Hamiltonian formalism where the time does not play any distinguished role. The advantage of such a formalism rests on the existence of a canonical symplectic form which allows us to relate functions and Hamiltonian vector fields, to introduce Poisson brackets, and to clarify the meaning of the traditional concept of canonical transformations such as will be shown in next sections.

In this approach the configuration space Q is replaced from the beginning by $Q \times \mathbb{R}$, the space of events, and therefore the cotangent bundle T^*Q will change to the extended phase space,[9,10] the cotangent bundle $T^*(Q \times \mathbb{R})$. In such manifold there is a canonical 2-form Ω_0 that in local canoni-

 0022-2488/83/122745-06$02.50

cal coordinates will be given by $\Omega_0 = dp_i \wedge dq^i + du \wedge dt$. We recall the natural diffeomorphisms $T^*(Q \times \mathbb{R}) \simeq T^*Q \times T^*\mathbb{R} \simeq T^*Q \times \mathbb{R} \times \mathbb{R}^*$. The coordinate function u is given by $u(\alpha) = \alpha(\partial/\partial t) \,|\, \forall \alpha \in T^*(Q \times \mathbb{R})$ and the projection of $T^*(Q \times \mathbb{R})$ on $T^*Q \times \mathbb{R}$ will be denoted by μ. The dynamics in this formalism can be defined by making use of an extended Hamiltonian $\Phi \in \Lambda^0(T^*(Q \times \mathbb{R}))$ as indicated in the following theorem.

Theorem 1: Let $H \in \Lambda^0(T^*Q \times \mathbb{R})$ be a time-dependent Hamiltonian and define the extended Hamiltonian $\Phi \in \Lambda^0(T^*(Q \times \mathbb{R}))$ by $\Phi = H \circ \mu + u$. The dynamical vector field X_Φ on $T^*(Q \times \mathbb{R})$ of the globally Hamiltonian dynamical system $(T^*(Q \times \mathbb{R}), \Omega_0, \Phi)$ is μ-related to X_H.

Proof: The vector field X_Φ is expressed in local canonical coordinates as

$$X_\Phi = -\hat{\Omega}_0^{-1}(d\Phi) = -\hat{\Omega}_0^{-1}(d(H \circ \mu) + du),$$

i.e.,

$$X_\Phi = \frac{\partial(H \circ \mu)}{\partial p_i}\frac{\partial}{\partial q^i} - \frac{\partial(H \circ \mu)}{\partial q^i}\frac{\partial}{\partial p_i} + \frac{\partial}{\partial t} - \frac{\partial(H \circ \mu)}{\partial t}\frac{\partial}{\partial u}$$

and therefore

$$\mu_*(X_\Phi) = \frac{\partial(H \circ \mu)}{\partial p_i}\frac{\partial}{\partial q^i} - \frac{\partial(H \circ \mu)}{\partial q^i}\frac{\partial}{\partial p_i} + \frac{\partial}{\partial t} = X_H.$$

Notice that the preceding result means that the integral curves of X_Φ project on integral curves of X_H, which are usually considered the trajectories of the time-dependent mechanical system defined by H. But in this formulation two new equations for the integral curves of X_Φ arise, namely, $(dt \circ \mu)/ds = 1$ and $du/ds = -\partial(H \circ \mu)/\partial t$; while the first one identifies the parameter s as the time, the second suggests the identification of $-u$ with the "energy."

The constant value hypersurfaces for Φ are invariant under time evolution, and the restriction of Ω_0 endows them with a contact structure in such way that they are isomorphic as contact manifolds to $(T^*Q \times \mathbb{R}, \omega_H)$:

Theorem 2: Let S_r be $S_r = \{m \in T^*(Q \times \mathbb{R}) \,|\, \Phi(m) = r\}$ and denote by i_r the natural injection $i_r : S_r \to T^*(Q \times \mathbb{R})$. The pair $(S_r, i_r^*\Omega_0)$ is a contact manifold and the restriction μ_r of μ to S_r is an isomorphism between the contact manifolds $(S_r, i_r^*\Omega_0)$ and $(T^*Q \times \mathbb{R}, \omega_H)$.

Proof: S_r is a regular hypersurface and therefore $(S_r, i_r^*\Omega_0)$ is a contact manifold (see, e.g., Proposition 5.1.7 in Ref. 1). Furthermore, if $m \in S_r$, then $\ker \mu_*(m) = \{\lambda \partial/\partial u|_m \,;\, \lambda \in \mathbb{R}\}$. The tangent space $T_m S_r$ is the kernel of $d\Phi(m)$ and therefore $\ker \mu_*(m) \cap T_m S_r = \varnothing$. Consequently, the rank of the restriction of $\mu_*(m)$ to $T_m S_r$ is $2n + 1$, i.e., $\mu_{r*}(m)$ is an isomorphism and μ_r is a local diffeomorphism. On the other hand, μ_r is an injective map of S_r onto $T^*Q \times \mathbb{R}$, and, consequently, μ_r is a diffeomorphism. Finally it is easy to check that $\mu_r^*\omega_H = i_r^*\Omega_0$. In fact, $\mu_r = \mu \circ i_r$, $H \circ \mu_r = r - u \circ i_r$, and therefore

$$\mu_r^*\omega_H = \mu_r^*(\tau^*\omega_0 - dH \wedge dt) = (\tau \circ u_r)^*\omega_0 - d(H \circ \mu_r) \wedge d(t \circ \mu_r)$$
$$= i_r^*[(\tau \circ \mu)^*\omega_0 + du \wedge d(t \circ \mu)],$$

i.e.,

$$\mu_r^*\omega_H = i_r^*\Omega_0.$$

Noteworthy is the fact that the extended phase space is foliated by the family of hypersurfaces S_r, $r \in \mathbb{R}$. This foliation is regular and any two leaves S_{r_1} and S_{r_2} are diffeomorphic; the diffeomorphism is given by $\mu_{r_2}^{-1} \circ \mu_{r_1}$.

3. THE CANONICAL TRANSFORMATIONS IN TIME-DEPENDENT SYSTEMS

In the case of a (time-independent) Hamiltonian dynamical system, canonical transformations are symplectomorphisms of the symplectic manifold; but in the standard formulation of time-dependent mechanical systems the geometric structure in $T^*Q \times \mathbb{R}$ is not a symplectic but a contact structure, and there is no way to give a similar geometric definition of the concept of canonical transformation. The extended formalism developed in Sec. 2 suggests to define such a concept as that of a fiber-preserving symplectomorphism in $(T^*(Q \times \mathbb{R}), \Omega_0)$. We will show later that this concept generalizes the one usually found in Refs. 1 and 3.

Definition 1: A canonical transformation of a time-dependent system $(T^*Q \times \mathbb{R}, \omega_H)$ is a pair (ψ, ϕ) of diffeomorphisms, $\psi \in \mathrm{Diff}(T^*(Q \times \mathbb{R}))$, $\phi \in \mathrm{Diff}(T^*Q \times \mathbb{R})$ such that:

(i) $\mu \circ \psi = \phi \circ \mu$;

(ii) ψ is a symplectomorphism of $(T^*(Q \times \mathbb{R}), \Omega_0)$.

In an appropriate coordinate system the condition (i) means that $\psi(q, p, t, u) = (\phi(q, p, t), \bar{\phi}(q, p, t, u))$, where $\bar{\phi} \in \Lambda^0(T^*(Q \times \mathbb{R}))$.

It is to be remarked that this concept of canonical transformation does not depend on the Hamiltonian function H such as was in the case of an autonomous system.

The set of these generalized canonical transformations, $\mathrm{Can}(T^*Q \times \mathbb{R})$ is a group with respect to the natural composition law.

In order to analyze this concept of generalized canonical transformation in terms of the standard framework, we recall that if (M, ω, Φ) is a Hamiltonian dynamical system and $S_r(\Phi)$ is a constant value hypersurface for Φ, then, for any $\psi \in \mathrm{Diff}(M)$, $\psi(S_r(\Phi)) = S_r(\Phi \circ \psi^{-1}) = S_r(\psi^{-1*}\Phi)$. Moreover, the diffeomorphism ψ defines a new Hamiltonian dynamical system $\psi(M, \omega, \Phi) = (M, \psi^{-1*}\omega, \psi^{-1*}\Phi)$. The dynamical vector fields are related by $\psi_*(X_\Phi) = X_{\Phi \circ \psi} = (X_{\psi^{-1*}\Phi})$. In a similar way, if (M, ω) is a contact structure, $(M, \psi^{-1*}\omega)$ is also a contact structure to be denoted $\psi(M, \omega)$. Hereafter we will use ψ_* as a shorthand for ψ^{-1*}, and the former expressions become $\psi(M, \omega, \Phi) = (M, \psi_*\omega, \psi_*\Phi)$ and $\psi_* X_\Phi = X_{\psi_*\Phi}$.

In the particular case we are considering, $S_r(\phi_*\Phi)$ is also a regular submanifold of $T^*(Q \times \mathbb{R})$. Moreover, since (ψ, ϕ) is a fiber-preserving map, the restriction of μ to $S_r(\psi_*\Phi)$ establishes a diffeomorphism between $S_r(\psi_*\Phi)$ and $T^*Q \times \mathbb{R}$. Thus, the new Hamiltonian function $\psi_*\Phi$ defines a regular projectable foliation of $T^*(Q \times \mathbb{R})$; its leaves

are in one-to-one correspondence with those of the foliation defined by Φ.

Theorem 3: Let (ψ, ϕ) be a canonical transformation of the time-dependent mechanical system $(T^*Q \times \mathbb{R}, \omega_H)$. The contact manifold $\phi(T^*Q \times \mathbb{R}, \omega_H) \equiv (T^*Q \times \mathbb{R}, \phi_* \omega_H)$ is isomorphic to the contact manifold $(S_r(\psi_* \Phi), i_r'^* \Omega_0)$, where i_r' is the natural inclusion of $S_r(\psi_* \Phi)$ in $T^*(Q \times \mathbb{R})$.

Proof: The diffeomorphism $\mu_r' = \mu \circ i_r'$ is such that $\mu_r' \phi_* \omega_H = (\mu \circ \psi^{-1} \circ i_r')^* \omega_H = (i_r'^{-1} \circ \psi^{-1} \circ i_r')^* \mu_r^* \omega_H$. From Theorem 2 it follows that $\mu_r^* \omega_H = i_r'^* \Omega_0$, and, consequently,

$$\mu_r'^* \phi_* \omega_H = (\psi^{-1} \circ i_r')^* \Omega_0 = i_r'^* \Omega_0.$$

Theorem 4: If (ψ, ϕ) is a canonical transformation of the time-dependent mechanical system $(T^*Q \times \mathbb{R}, \omega_H)$, then

$$\psi_* \omega_H = \omega_{\hat{K}}, \tag{3.1}$$

where $\hat{K} = K \circ i_r' \circ \mu_r'^{-1}$ and

$$K = \psi_*(H \circ \mu) + \psi_* u - u \tag{3.2}$$

is a function $K \in \Lambda^0(T^*(Q \times \mathbb{R}))$.

Proof: Let K be defined by $K = \psi_* \Phi - u$. Then, the identity $K \circ i_r' + u \circ i_r' = r$ permits us to write

$$\begin{aligned} i_r'^* \Omega_0 &= i_r'^* \mu^* \tau^* \omega_0 + i_r'^* [du \wedge d(t \circ \mu)] \\ &= i_r'^* [\mu^* \tau^* \omega_0 - dK \wedge d(t \circ \mu)]. \end{aligned}$$

On the other hand,

$$\begin{aligned} \mu_r'^* \omega_{\hat{K}} &= \mu_r'^* [\tau^* \omega_0 - d(K \circ i_r' \circ \mu_r'^{-1}) \wedge dt] \\ &= i_r'^* [\mu^* \tau^* \omega_0 - dK \wedge d(t \circ \mu)] \end{aligned}$$

and the relation $i_r'^* \Omega_0 = \mu_r'^* \omega_K$ follows.

It is noteworthy that the new Hamiltonian K is not defined on $T^*Q \times \mathbb{R}$ but on $T^*(Q \times \mathbb{R})$, i.e., K does generally depend on the variable u because of the difference $\psi_* u - u \equiv W_{\psi^{-1}}$ may depend on it. Consequently, the function $\hat{K}$ depends on the choice of the constant value r.

On the other hand, the existence of the function K is assured for any H as a consequence of the canonical character of the transformation.

Theorem 5: Let (ψ, ϕ) be a canonical transformation of a mechanical system $(T^*Q \times \mathbb{R}, \omega_H)$. The closed 2-form $\tilde{\omega}_0 = \tau^* \omega_0$ transforms as follows:

$$\phi_* \tilde{\omega}_0 = \tilde{\omega}_0 - d\alpha$$

with

$$\alpha = \hat{W}_{\psi^{-1}}\, dt - (\phi_* H)\, d(\phi_* t - t). \tag{3.3}$$

Here $\hat{W}_{\psi^{-1}}$ denotes the restriction of $W_{\psi^{-1}}$ to $T^*Q \times \mathbb{R}$ by means of $S_r(\psi_* \Phi)$, i.e., $\hat{W}_{\psi^{-1}} = W_{\psi^{-1}} \circ \mu_r'^{-1}$.

Proof: The 2-form $\psi_* \omega_H$ is given by $\phi_* \tilde{\omega}_0 - d(\phi_* H) \wedge d(\phi_* t)$. On the other hand, $\hat{K} = \phi_* H + \hat{W}_{\psi^{-1}}$; hence $\omega_{\hat{K}} = \tilde{\omega}_0 - d(\phi_* H) \wedge dt - d\hat{W}_{\psi^{-1}} \wedge dt$, and the result of the theorem follows.

Only time-preserving canonical transformations are usually considered, and in this particular case the theory we are proposing reduces to the standard one.

Definition 2: A standard canonical tranformation of $(T^*Q \times \mathbb{R}, \omega_H)$ is a canonical transformation (ψ, ϕ) such that $\phi_* t = t$.

As a consequence of Theorem 5, a standard canonical transformation changes $\tilde{\omega}_0$ to $\tilde{\omega}_0 - d\hat{W}_{\psi^{-1}} \wedge dt$. Moreover, the next theorem shows that the new Hamiltonian K does not depend on u.

Theorem 6: If (ψ, ϕ) is a standard canonical transformation of $(T^*Q \times \mathbb{R}, \omega_H)$, the new Hamiltonian K factorizes as $K = \hat{\hat{K}} \circ \mu$, where $\hat{\hat{K}}$ is a function of $\Lambda^0(T^*Q \times \mathbb{R})$.

Proof: Let $\tilde{\tilde{\omega}}_0$ be the pullback of $\tilde{\omega}_0$ by μ, i.e., $\omega_0 = \mu^* \tilde{\tilde{\omega}}_0$. Then $\psi_* \tilde{\tilde{\omega}}_0 = \mu^* \phi_* \tilde{\omega}_0$ and therefore $\psi_* \tilde{\tilde{\omega}}_0 = \tilde{\tilde{\omega}}_0 - d(\hat{W}_{\psi^{-1}} \circ \mu) \wedge d(t \circ \mu)$. Neither $\psi_* \tilde{\tilde{\omega}}_0$ nor $\tilde{\omega}_0$ depend on the variable u and the corresponding u independence of the function $\hat{W}_{\psi^{-1}}$ implies that $\hat{W}_{\psi^{-1}}$ does not depend on the choice of the constant value r. Finally K can then be expressed as $K = \hat{\hat{K}} \circ \mu$ with $\hat{\hat{K}} = (\phi_* H + \hat{W}_{\psi^{-1}})$.

As a corollary of this theorem we find that a standard canonical transformation can also be considered as a diffeomorphism ϕ of $T^*Q \times \mathbb{R}$ such that

$$\phi_* \omega_H = \omega_{\hat{\hat{K}}} \quad \text{with} \quad \hat{\hat{K}} = \phi_* H + \hat{W}_{\psi^{-1}}, \tag{3.4}$$

$$\hat{W}_{\psi^{-1}} \in \Lambda^0(T^*Q \times \mathbb{R}),$$

$$\phi_* t = t, \tag{3.5}$$

which is the usual definition of canonical transformation.[1,3] In fact, the comparison of the theory in the extended phase space formalism we are developing to the standard one proceeds via the group homomorphism Δ: $\mathrm{Can}(T^*Q \times \mathbb{R}) \to \mathrm{Diff}(T^*Q \times \mathbb{R})$, given by $\Delta(\psi, \phi) = \phi$. Two canonical transformations (ψ_1, ϕ) and (ψ_2, ϕ) are physically equivalent, and they are related by $\psi_2^* u = \psi_1^* u + f \circ \phi^*(t) \circ \mu$ with $f \in \Lambda^0(\mathbb{R})$. We remark that (ψ, id) is a symplectomorphism if and only if $\psi^* u = u + f \circ t \circ u$.

The main question is to elucidate whether a given diffeomorphism ϕ of $T^*Q \times \mathbb{R}$ can be lifted to a canonical transformation.

Theorem 7: Let (ψ, ϕ) be a canonical transformation. Then,

(i) There exists $\hat{W}_{\phi^{-1}} \in \Lambda^0(T^*Q \times \mathbb{R})$ such that

$$\phi_* \tilde{\omega}_0 = \tilde{\omega}_0 - d\hat{W}_{\phi^{-1}} \wedge dt + d(\phi_* H) \wedge d(\phi_* t - t) \tag{3.6}$$

(ii) $\phi_* t = f \circ t$ with $f \in \Lambda_0(\mathbb{R})$ (3.7)

Proof: The first claim follows from Theorem 5 with $\hat{W}_{\phi^{-1}} = \hat{W}_{\psi^{-1}}$. Furthermore, we can compute $\psi_* \Omega_0$, and we obtain $\psi_* \Omega_0 = \Omega_0 + \Gamma$ with Γ defined by $\Gamma = d(\phi_* H) \circ \mu \wedge d(\phi_* t - t) \circ \mu + d(\psi_* u) \wedge d(\phi_* t - t) \circ \mu$. The map ψ being symplectic, the 2-form Γ vanishes, and this implies the relation $d(\psi_* u) \wedge d(\psi_* t) \circ \mu = du \wedge d(t \circ \mu) + d\hat{W}_{\phi^{-1}} \circ \mu \wedge d(t \circ \mu) - d(\phi_* H) \circ \mu \wedge d(\phi_* t - t) \circ \mu$. As a consequence, the following equations hold:

$$\frac{\partial \psi_* u}{\partial u} \cdot \frac{\partial \phi_* t}{\partial t} = 1, \quad \frac{\partial \psi_* u}{\partial u} \cdot \frac{\partial \phi_* t}{\partial q} = 0,$$

$$\frac{\partial \psi_* u}{\partial u} \cdot \frac{\partial \phi_* t}{\partial p} = 0,$$

and therefore the second statement follows.

Theorem 8: A diffeomorphism ϕ of $T^*Q \times \mathbb{R}$ such that it satisfies both conditions (3.6) and (3.7) can be lifted to a canonical transformation (ψ, ϕ). An explicit solution is given by

$$\psi_* u = \left(\frac{df}{dt}\right)^{-1} [u + \hat{W}_{\phi^{-1}} \circ \mu + (\phi_* H) \circ \mu] - (\phi_* H) \circ \mu. \tag{3.8}$$

Proof: It suffices to show that ψ defined by (3.8) is a symplectomorphism. If we use the condition (ii), we will obtain

$$(\psi_* u)\, d(\phi_* t) \circ \mu = [u + \hat{W}_{\phi^{-1}} \circ \mu + (\phi_* H) \circ \mu]\, d(t \circ \mu) - (\phi_* H) \circ \mu\, d(\phi_* t) \circ \mu,$$

and, if we substitute $d(\psi_* u) \wedge d(\phi_* t) \circ \mu$ in the expression for $\psi_* \Omega_0$, we will find $\psi_* \Omega_0 = \Omega_0$.

The particular case of a nonstandard canonical transformation where ϕ is such that it only changes the time variable has been used by some authors,[4-8] but the fact that it can be considered as a symplectomorphism has not yet been claimed, as far as we know. Theorem 8 shows this possibility because such a diffeomorphism in $T^*Q \times \mathbb{R}$ can be lifted to a canonical transformation. Moreover, the set of all diffeomorphisms ϕ of $T^*Q \times \mathbb{R}$ that can be lifted to a canonical transformation, i.e., Δ (Can $T^*Q \times \mathbb{R}$), is a semidirect product group of the subgroup of diffeomorphisms that only change the time to a time function by the normal subgroup of diffeomorphisms corresponding to standard canonical transformations. This means that any diffeomorphisms ϕ that can be lifted factorizes in a unique way as a product of two factors, one in each subgroup.

4. GENERATING FUNCTIONS OF CANONICAL TRANSFORMATIONS

We are now able to apply the theory of generating functions of symplectomorphisms to study canonical transformations of time-dependent systems. Before going deeper in such application we start by recalling some basic concepts in the theory of generating functions.

Given a symplectic manifold, namely (P, Ω) a submanifold $L \subset P$ is Lagrangian if $\dim L = n = \frac{1}{2} \cdot \dim P$ and $i^*\Omega = 0$, where i is the canonical injection of L in P. There is then a function G_L locally defined such that $i^*\Theta = dG_L$, where Θ is such that $\Omega = d\Theta$; such a function G_L is said to be the Weinstein generating function for L.[1,11] A rather interesting case is that of a symplectic manifold $(T^*Q_1 \times T^*Q_2, \Omega_1 \ominus \Omega_2)$ arising from the pair of symplectic manifolds (T^*Q_1, Ω_1) and (T^*Q_2, Ω_2). The 2-form $\Omega_1 \ominus \Omega_2$ is defined by $\Omega_1 \ominus \Omega_2 = \pi_1^* \Omega_1 - \pi_2^* \Omega_2$ with π_i the projection of $T^*Q_1 \times T^*Q_2$ on T^*Q_i $(i = 1, 2)$. A diffeomorphism ψ: $T^*Q_1 \to T^*Q_2$ is a symplectomorphism if and only if the graph of ψ is a Lagrangian submanifold. On the other hand, the restriction of π_1 to the graph of any function f: $T^*Q_1 \to T^*Q_2$ is a global diffeomorphism. The Weinstein generating function G_ψ of the symplectomorphism ψ: $T^*Q_1 \to T^*Q_2$ can be related to the corresponding Poincaré generating function as follows:

Proposition 1: Let ψ be a symplectomorphism ψ: $T^*Q_1 \to T^*Q_2$. The locally defined Weinstein and Poincaré generating functions are related (up to an additive constant) by

$$G_\psi \circ \pi_1^{-1} = -S_\psi. \tag{4.1}$$

Proof: Let θ_i be the canonical 1-form on T^*Q_i $(i = 1, 2)$. The function S_ψ is locally defined by $\psi^*\theta_2 = \theta_1 + dS_\psi$ while G_ψ is locally defined by $i^*(\theta_1 \ominus \theta_2) = dG_\psi$. Then, $d(G_\psi \circ \pi_1^{-1}) = (i \circ \pi_1^{-1})^*(\theta_1 \ominus \theta_2)$, and, therefore, if we recall the definition of $\theta_1 \ominus \theta_2$, it is easy to check that $d(G_\psi \circ \pi_1^{-1}) = -dS_\psi$ because of the relation $d(G_\psi \circ \pi_1^{-1}) = (\pi_1 \circ i \circ \pi_1^{-1})^* \theta_1 - (\pi_2 \circ i \circ \pi_1^{-1})^* \theta_2 = \theta_1 - \psi^* \theta_2$.

Hereafter only Poincaré generating functions will be considered. It is worthy of remark the identity $d(S_\psi + \psi^* S_{\psi^{-1}}) = 0$, which is easily obtained by application of ψ^* on the relation defining $dS_{\psi^{-1}}$. In general,

$$S_{\psi_1 \circ \psi_2} = S_{\psi_2} + \psi_2^* S_{\psi_1}. \tag{4.2}$$

Definition 3: A generating function of a canonical transformation (ψ, ϕ) of the time-dependent system $(T^*Q \times \mathbb{R}, \omega_H)$ is a generating function of the symplectomorphism ψ of $(T^*(Q \times \mathbb{R}), \Omega_0)$, i.e., it is defined by $\psi_* \Theta_0 = \Theta_0 + dS_{\psi^{-1}}$. It is only locally defined.

In the particular case of a standard canonical transformation we will see that the generating function above defined reduces to the usual concept. In fact, if $\tilde{\theta}_0$, θ_H, $\theta_{\hat{W}_\phi}$, and $\theta_{\hat{K}}$ are the 1-forms of $T^*Q \times \mathbb{R}$ defined by $\tilde{\theta}_0 = \tau^* \theta_0$, $\theta_H = \tilde{\theta}_0 - H\, dt$, $\theta_{\hat{K}} = \tilde{\theta}_0 - \hat{K}\, dt$, and $\theta_{\hat{W}_\phi} = \tilde{\theta}_0 - \hat{W}_{\phi^{-1}}\, dt$; then, if (ψ, ϕ) is a canonical transformation, $\phi_* \theta_H - \theta_{\hat{K}}$ is a closed 1-form and there is a (locally defined) function $\hat{S}_{\phi^{-1}}$, such that $\phi_* \theta_H - \theta_{\hat{K}} = d\hat{S}_{\phi^{-1}}$.

Lemma 1: With the above notations, the 1-form $\tilde{\theta}_0$ transforms as follows:

$$\phi_* \tilde{\theta}_0 = \theta_{\hat{W}_{\phi^{-1}}} + d\hat{S}_{\phi^{-1}} + (\phi_* H)\, d(\phi_* t - t). \tag{4.3}$$

Proof: The relation $\phi_* \tilde{\theta}_0 = \phi_* \theta_H + (\phi_* H)\, d(\phi_* t)$, is written in terms of the function $S_{\phi^{-1}}$ as follows: $\phi_* \tilde{\theta}_0 = \theta_{\hat{K}} + d\hat{S}_{\phi^{-1}} + (\phi_* H)\, d(\phi_* t)$. We can replace $\hat{K}$ by $\phi_* H + \hat{W}_{\phi^{-1}}$, and we will find the relation we were looking for.

As a corollary of this proposition, in the particular case of a standard canonical tranformation, $\phi_* \tilde{\theta}_0 = \theta_{\hat{W}_{\phi^{-1}}} + d\hat{S}_{\phi^{-1}}$, which is the usual expression for the generating function.[1]

We are now interested in looking for a relation between the generating functions $S_{\psi^{-1}}$ and $\hat{S}_{\phi^{-1}}$. This is given by the next theorem, where it will be also shown that $\hat{S}_{\phi^{-1}}$ is a good generating function.

Theorem 9: If (ψ, ϕ) is a canonical transformation of $(T^*Q \times \mathbb{R}, \omega_H)$, the (locally defined) generating function $S_{\psi^{-1}}$ is related to $\hat{S}_{\phi^{-1}}$ by

$$dS_{\psi^{-1}} = d\hat{S}_{\phi^{-1}} \circ \mu + d(h \circ t \circ \mu), \tag{4.4}$$

where $h \in \Lambda^0(\mathbb{R})$.

Proof: We recall the definition of $S_{\psi^{-1}}$, $\psi_* \Theta_0 = \Theta_0 + dS_{\psi^{-1}}$ as well as the identity $\Theta_0 = \mu^* \tilde{\theta}_0 + u\, d(t \circ \mu)$. Therefore, $\psi_* \Theta_0 = (\mu \circ \psi^{-1})^* \tilde{\theta}_0 + (\psi_* u)\, d(t \circ \mu \circ \psi^{-1}) = \mu^* \phi_* \tilde{\theta}_0 + (\psi_* u)\, d(\phi_* t) \circ \mu$.

On the other side, if we make use of Theorem 8 and define a symplectomorphism $\bar{\psi}$ in $(T^*Q \times \mathbb{R})$ by $\bar{\psi}_* u = (df/dt)^{-1} [u + \hat{W}_{\phi^{-1}} \circ \mu + (\phi_* H) \circ \mu] - (\phi_* H) \circ \mu$, both diffeomorphisms ψ and $\bar{\psi}$ are related by $\psi_* u = \bar{\psi}_* u + g \circ \phi_* (t) \circ \mu = \psi_* u + g \circ f \circ t \circ \mu$, where g is a real differentiable function. A comparison with $\psi_* \Theta_0$

$= \mu^* \tilde{\theta}_0 + u\, d(t \circ \mu) + d\hat{S}_{\psi^{-1}}$ leads to the relation $dS_{\psi^{-1}} = d\hat{S}_{\phi^{-1}} \circ \mu + g \circ f \circ (t \circ \mu)\, d(f \circ t \circ \mu)$ with $\phi_* t = f \circ t$. Finally, if h is a primitive for $g \circ f$, we will find $dS_{\psi^{-1}} = d\hat{S}_{\phi^{-1}} \circ \mu + d(h \circ t \circ \mu)$.

The structure of semidirect product of the set of diffeomorphisms of $T^*Q \times \mathbb{R}$ that can be lifted to a canonical transformation as identified with the corresponding equivalence classes suggests study of the generating functions associated with elements of each factor.

Lemma 2: Let (ψ, ϕ) be a canonical transformation such that changes only the time t to $\phi_* t = f \circ t$. Then $\hat{W}_{\phi^{-1}}$ is given by

$$\hat{W}_{\phi^{-1}} = (\phi_* H)\left(\frac{df}{dt} - 1\right) + \frac{d\hat{S}_{\phi^{-1}}}{\partial t}. \tag{4.5}$$

Proof: The invariance of $\tilde{\theta}_0$ under ϕ_* implies that $-\hat{W}_{\phi^{-1}}\, dt + (\phi_* H)\, d(\phi_* t - t) + d\hat{S}_{\phi^{-1}} = 0$, and, therefore, $\hat{W}_{\phi^{-1}}$ will be given by the expression (4.5). The transformation law of $\tilde{\theta}_0$ given by Lemma 1 shows that $d\hat{S}_{\phi^{-1}} = d\hat{S}_{\phi^{-1}}/\partial t\, dt$.

Proposition 2: Let (ψ, ϕ) be a canonical transformation that factorizes as a product $\psi = f \circ \bar{\phi}$. The generating function is $S_{\psi^{-1}} = S_{f^{-1}} + f_* S_{\bar{\phi}^{-1}}$, i.e., $\hat{S}_{\psi^{-1}} = f_* \hat{S}_{\bar{\phi}^{-1}} + \hat{h}$ (up to a constant), where $\bar{\phi}$ is the corresponding standard canonical transformation in the factorization.

Proof: Since $\psi = f \circ \bar{\phi}$, we obtain $S_{\psi^{-1}} = S_{\bar{\phi}^{-1} \circ f^{-1}} = S_{f^{-1}} + f_* S_{\bar{\phi}^{-1}}$ (up to an additive constant). The relations $S_{\bar{\phi}^{-1}} = \hat{S}_{\bar{\phi}^{-1}} \circ \mu$ and $S_{\psi^{-1}} = \hat{S}_{\psi^{-1}} \circ \mu + h \circ t \circ \mu$ lead to the second expression when $\hat{h}$ is a real function defined by $\hat{h} = \hat{S}_{f^{-1}} - h \circ t$.

A particularly important application of canonical transformations is that of the Hamilton–Jacobi equation for the determination of Hamilton's principal function.[1,3,10–12] We look for a generating function of the canonical transformation ψ^{-1} carrying H to equilibrium. It is given by $G_{\psi^{-1}} = -S_{\psi^{-1}} \circ \pi_1$, where π_1 is the projection π_1: $T^*(Q \times \mathbb{R}) \times T^*(Q \times \mathbb{R}) \to T^*(Q \times \mathbb{R})$ on the first factor. Let π denote the natural projection π: $T^*(Q \times \mathbb{R}) \times T^*(Q \times \mathbb{R}) \to (Q \times \mathbb{R}) \times (Q \times \mathbb{R})$. Let assume that $G_{\psi^{-1}}$ may be written in terms of the local coordinates $(q, t; q', t')$ by means of the identification of the graph of ψ^{-1} to $(Q \times \mathbb{R}) \times (Q \times \mathbb{R})$, i.e., $\det \partial^2 G / \partial q^i \partial p'_j \neq 0$. Hereafter by $S_{\psi^{-1}}$ we will understand $S_{\psi^{-1}} \circ \pi_1 \circ \pi^{-1}$.

The symplectic structure given on $T^*(Q \times \mathbb{R})$ reduces the time-dependent case to the simpler one of time-independent Hamilton–Jacobi equation and therefore the symplectomorphism ψ^{-1} transforming the system to equilibrium will be given by a generating function $S_{\psi^{-1}}$ satisfying the Hamilton–Jacobi equation $\Phi \circ d_1 S_{\psi^{-1}} = r$, i.e., in local coordinates $\Phi(q, \partial S_{\psi^{-1}}/\partial q, t, \partial S_{\psi^{-1}}/\partial t) = r$. The subindex 1 in d_1 indicates that the differential is only with respect to the first coordinates. If we recall $\Phi = H + u$, the Hamilton–Jacobi equation reads

$$H\left(q, \frac{\partial S_{\psi^{-1}}}{\partial q}, t\right) = r - \frac{\partial S_{\psi^{-1}}}{\partial t}. \tag{4.6}$$

The form of the generating function $S_{\psi^{-1}}$ established in the last proposition give us a more explicit form of the H–J equation. That is, the generating function $S_{\psi^{-1}}$ is $S_{\psi^{-1}} = f_* S_{\bar{\phi}^{-1}} + S_{f^{-1}}$, where f is denoting here the part of the transformation ψ^{-1} that changes the time only. Hence

$$r = \Phi(d_1 S_{\psi^{-1}}) = H(d_1 S_{\psi^{-1}}) + u(d_1 S_{\psi^{-1}}) = H(d_1 S_{\psi^{-1}}) + \left(\frac{df}{dt}\right)^{-1} \frac{\partial S_{\bar{\phi}^{-1}}}{\partial t}$$

and therefore

$$r = H(d_1 \hat{S}_{\bar{\phi}^{-1}} \circ \mu) + \left(\frac{df}{dt}\right)^{-1} \frac{d\hat{S}_{\bar{\phi}^{-1}} \circ \mu}{\partial t} \tag{4.7}$$

and $S_{\bar{\phi}^{-1}} = \hat{S}_{\phi^{-1}} + \text{const}$, because $\bar{\phi}$ is a standard canonical transformation. In the easier case of ψ being a standard canonical transformation, the Hamilton–Jacobi equation reduces to the usual form

$$H(d_1 \hat{S}_{\phi^{-1}} \circ \mu) = r - \frac{\partial(\hat{S}_{\phi^{-1}} \circ \mu)}{\partial t}.$$

5. SOME APPLICATIONS AND COMMENTS

We will show now that the generalized concept of time-dependent canonical transformation we have developed in preceding sections is the appropriate one to study some transformations which have been used in the literature, but which have not been identified as canonical transformations; in particular, the Kustaanheimo and Stiefel transformation[4,5] regularizing the problem of the Kepler motion and that used by Leach[6–8] when studying a time-dependent harmonic oscillator are nonstandard canonical transformations which can be factorized as indicated in Sec. 4. Both of them are particular examples of a broader class of problems where a canonical transformation arises in which the time changes as

$$s = \int_0^t \rho^{-2}(t')\, dt' \tag{5.1}$$

with ρ a solution of some differential equation. Leach[6–8] developed a method of finding a standard canonical transformation leading the old Hamiltonian $H = p^2/2 + \frac{1}{2}\omega^2(t) q^2$ to a new Hamiltonian $H' = (P^2 + Q^2)/\rho^2$, where ρ is a time function solution of $\ddot{\rho} + \omega^2 \rho = \rho^{-3}$. After this first step has been carried out, a second nonstandard canonical transformation in which only the time changes, given by (5.1), is to be done.

In a similar way in the two-dimensional Kepler problem a standard canonical transformation given by $x = q_1^2 - q_2^2$ and $y = 2q_1 q_2$ (it is remarkable that it will be no one-to-one if no restriction on the domain is imposed, see, e.g., Ref. 13) transforms the old Hamiltonian $H = (p_1^2 + p_2^2)/2m - k/r$ to the new one $H' = (P_1^2 + P_2^2)/8m\rho^2 - k/\rho^2 = (1/\rho^2)[(P_1^2 + P_2^2)/8m - k]$. A nonstandard canonical transformation given by (5.1) is also to be done but the function ρ^2 is now defined by the norm of the position vector running an actual trajectory.

The fundamental point is that this transformation when considered as a canonical one tells us how to find a new Hamiltonian while so far the new Hamiltonian was an ansatz.

We develop more explicitly the prescription of our theory for finding the new Hamiltonian: The generating func-

tion $\hat{S}_{\phi^{-1}}$ is constant while $\hat{W}_{\phi^{-1}}$ is given by (4.5) and hence the new Hamiltonian $\hat{K}$ is $\hat{K} = (\phi_* H) + \hat{W}_{\phi^{-1}}$ $= (df/dt)(\phi_* H) + \partial \hat{S}_{\phi^{-1}}/\partial t$. Furthermore, $\psi_* u$ $= (df/dt)^{-1}[u + \hat{W}_{\phi^{-1}} + \phi_* H] - \phi_* H$ (see Theorem 8), and, consequently, one also finds $\psi_* \Phi$ $= \hat{K} \circ \mu + u = (df/dt)(\phi_* H) + u$. The choice $df/dt = \rho^2$ for f reduces both cases, the harmonic oscillator and the Kepler problem, to $\hat{K} = P^2 + Q^2$ and $\hat{K} = (P_1^2 + P_2^2)/8m - k + \rho^2(\partial \hat{S}_{\phi^{-1}}/\partial t)$, respectively.

Finally, some remarks are in order to show that our formalism could be useful for the study of some constrained autonomous dynamical systems. For instance, the system given by a test particle in a gravitational field is usually formulated in an extended formalism, but, because of the Hamiltonian constraint, it is reduced to an effective time-dependent system. The relation between both systems is similar to that considered in this paper. A more interesting case is general relativity itself, because the canonical formulation leads to a time-independent Hamiltonian system with the constraint $H = 0$,[14,15] which is just the condition relating an autonomous and the corresponding nonautonomous system in our framework. This problem will be considered elsewhere.

ACKNOWLEDGMENTS

The authors are grateful to J. Gracia-Bondía for stimulating discussions. We acknowledge partial financial support by Instituto de Estudios Nucleares and Comisión Asesora de Investigación Científica y Técnica.

[1] R. Abraham and J. Marsden, *Foundations of Mechanics* (Benjamin, New York, 1978), 2nd ed.
[2] J. F. Cariñena and L. A. Ibort, J. Phys. A **16**, 1 (1983).
[3] L. H. Loomis and S. Sternberg, *Advanced Calculus* (Addison-Wesley, Reading, MA, 1968).
[4] P. Kustaanheimo and E. Stiefel, J. Reine Angew. **218**, 204 (1965).
[5] E. Stiefel and G. Scheifele, C. R. Acad. Sci. Paris A **267**, 950 (1968).
[6] P. G. L. Leach, J. Math. Phys. **18**, 1902 (1977).
[7] P. G. L. Leach, J. Austr. Math. Soc. B **22**, 12 (1980).
[8] H. R. Lewis and P. G. L. Leach, J. Math. Phys. **23**, 2371 (1982).
[9] W. Thirring, *Classical Dynamical Systems* (Springer-Verlag, Berlin, 1978).
[10] N. Woodhouse, *Geometric Quantization* (Clarendon, Oxford, 1980).
[11] A. M. Vinogradov and B. A. Kupershmidt, "The Structures of Hamiltonian Mechanics," in *Integrable Systems*, edited by S. P. Novikov, London Math. Soc. Lecture Note Series 60 (Cambridge U. P., London, 1981).
[12] V. I. Arnold, *Les méthodes mathématiques de la méchanique classique* (Mir, Moscow, 1976).
[13] M. Boiteux, J. Math. Phys. **23**, 1311 (1982).
[14] R. Arnowit, S. Deser, and C. W. Misner, *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962).
[15] K. Kuchař, *Canonical Quantization of Gravity*, Proceedings of the 1972 Banff Summer School (Reidel, Dordrecht, 1974).

Stochastic electrodynamics for the free particle

L. de la Peña
Instituto de Física, Universidad Nacional Autónoma de México, Apdo. Postal 20-364, 01000 México, D. F., Mexico

A. Jáuregui[a)]
Escuela de Altos Estudios, Universidad de Sonora, Hermosillo, Sonora, Mexico

(Received 22 December 1980; accepted for publication 18 November 1982)

The theory of stochastic electrodynamics is applied to the free particle and to the particle moving in a homogeneous field, leading to a complete temperature- and time-dependent description in phase space. After a transient time, the marginal description in configuration space coincides entirely with quantum mechanics, while the phase-space description is only mathematically related to the Wigner distribution. The Schrödinger equation appears as a natural—though incomplete—means of describing the statistical behavior of the electron under these conditions.

PACS numbers: 03.65.Bz, 05.40. + j

I. INTRODUCTION

In the last two decades or so the theory of stochastic electrodynamics (SED) has been the subject of intensive attention by several workers.[1] The theory starts from the recognition—that goes back to Planck[2] and Nernst[3]—of the physical reality of the zero-point radiation field with energy per mode $\frac{1}{2}\hbar\omega$. This field is introduced into the Abraham–Lorentz equation of motion for the electron, thus transforming it into a stochastic equation that has been named after Braffort and Marshall.[4] It is believed that this theory may help in laying new foundations for nonrelativistic quantum mechanics, and in fact a certain progress has been achieved in this direction. With its help and without any quantum assumption, the Planck distribution has been derived,[5–7] as well as the Van der Waals forces[8,9]; the detailed behavior of the quantum harmonic oscillator has been explained, including the radiative corrections[1–16]; the connection with Schrödinger theory and the radiative corrections from quantum electrodynamics have also been explored in the general case,[13,16–19] etc. However, in more elaborated problems the theory has met severe difficulties, which are revealed by the fact that it predicts a behavior at variance with the one predicted by quantum theory. In the list of these defiant problems we must include not only nonlinear systems as the hydrogen atom[20] or the anharmonic oscillator,[21,22] but even some linear problems as a system of degenerate coupled harmonic oscillators.[23] Even the free particle presents some problems, since only qualitative agreement has been obtained up to now with quantum mechanics.[15] The situation is certainly confused, but if one is still confident about the potentialities of SED, one is led to the obvious conclusion that the theory requires some revisions.

It is not difficult to trace back a possible—although not necessarily the single one—source of the difficulties in the Abraham–Lorentz (AL) equation of motion used in SED, since, as is well known, this equation predicts a noncausal behavior of the self-interacting point charge. Fortunately, it is possible to construct an alternative to the AL equation for an *extended* particle that is free of noncausal behavior. In this equation, the usual radiation reaction term $(2e^2/3c^3)\dddot{x}$ of AL theory is substituted by an integral expression of the form $\int_{-\infty}^{t} g(ct - ct')[\ddot{x}(t') - \ddot{x}(t)]\,dt'$, this implying that the motion at a given time depends on all past accelerations, i.e., on the whole trajectory. The kernel $g(ct)$ is related to the spatial structure of the extended particle, as explained in the literature.[24,25] Elsewhere we make an attempt to explore the theory to the simplest possible problem, namely, the free particle. In spite of the simplicity of the problem, the theory is highly nontrivial and demands the use of sophisticated techniques and long calculations, several of which are still in progress. Therefore, it seems preferable to try to explore this problem with simple even if approximate means, as those afforded by usual SED, in spite of the shortcomings of this theory, a task to which we devote the present paper.

To partially overcome the difficulties associated with the AL equation we will pay attention only to the asymptotic behavior of the system. It is clear that the structure of the particle may play an important role—even an essential one—on the motion of a free particle only for short times, but once the transient has disappeared, the dynamics will be practically independent of the structure, if we are allowed to neglect small radiative corrections to the self-energy. This is confirmed by the explicit calculations (see, e.g., Ref. 25). Hence by considering only the predictions of the theory for times long enough we are able to apply the AL equation to the free particle as an approximate representation of the more elaborate theory. Since adding a constant force hardly makes any important calculational difference, we here investigate not the free particle, but the particle acted on by a homogeneous field, thus extending somewhat the range of applicability of the theory. On the other hand, from the point of view of the principles, this last problem has the great advantage of allowing us to make legitimate use of the AL equation, by assuming an external homogeneous force that is everywhere much stronger than the radiation reaction force.[27]

The asymptotic phase-space distribution obtained may be compared with the one predicted by quantum mechanics:

[a)] With partial support by DGICSA, Secretaría de Educación Pública, Mexico.

 0022-2488/83/122751-11$02.50

Both distributions are different but related by a Gaussian integral transform, a result that has been suggested in the literature on repeated occasions,[28] though never convincingly demonstrated, and that seems to be valid only in this case. On the other hand, the marginal distribution in configuration space predicted coincides with the usual quantum mechanical one at any temperature; thus the simplest possible way of studying the statistical behavior of the free particle in configuration space is just by solving Schrödinger's equation. In this way a natural link is established between usual quantum theory and SED for the free particle and for the particle in a homogeneous field, in spite of the fact that there exist important differences between the predictions of both theories.

A point which is worth mentioning is that SED predicts for the free particle a fixed minimum asymptotic variance for the velocity; unfortunately, the numerical value of this quantity is strongly cutoff dependent, and hence we may offer, for the time being, only a rough preliminary estimation of its value. Nevertheless, this is an important prediction, since quantum mechanics affords no means to evaluate it. Due to the fact that the dispersion of the velocity of the free particle attains a constant finite value, the final average energy is also a constant, thus suggesting that in the long run an energetic equilibrium between the particle and the zero-point radiation field is reached. However, at the end of the paper a spectral analysis of the power exchanged between particle and field shows that an intensive instantaneous exchange of energy between the different modes of the field takes place, even under the condition of constancy of the average kinetic energy, thus extending for the present linear case earlier results of Boyer.[29]

II. STOCHASTIC ELECTRODYNAMICS FOR THE PARTICLE IN A HOMOGENEOUS FIELD

The Braffort–Marshall equations for a point charge acted on by a constant external force $\mathbf{F}$ is the AL equation

$$m\ddot{\mathbf{x}} = \mathbf{F} + e\mathbf{E}(t) + m\tau\dddot{\mathbf{x}}, \tag{1}$$

where

$$\tau = 2e^2/3mc^3. \tag{2}$$

Here, as has become customary in the study of nonrelativistic problems in SED,[1,4] we approximate the Lorentz force by its electric part $e\mathbf{E}(t)$, written in the long-wavelength (or dipole) approximation. The random field $\mathbf{E}(t)$ is assumed to have a Gaussian distribution with zero average and autocorrelation given by the Wiener–Khinchin theorem

$$\langle E_i(t)E_j(t')\rangle = \delta_{ij}\int_0^\infty S_E(\omega, T)\cos\omega(t-t')\,d\omega, \tag{3}$$

with a power spectrum $S_E(\omega, T)$ given by Planck's law at temperature T:

$$S_E(\omega, T) = (2\hbar\omega^3/3\pi c^3)[1 + 2/(e^{\beta\hbar\omega} - 1)], \tag{4}$$

where, as usual, β represents the inverse absolute temperature in units of the Boltzmann constant:

$$\beta = 1/k_B T. \tag{5}$$

Here a difficulty appears, namely, that Planck's distribution includes the zero-point contribution

$$S_E(\omega, 0) = (2\hbar/3\pi c^3)\omega^3,$$

which is not integrable over the whole frequency spectrum. Since a deeper theory of the vacuum that may solve this fundamental problem is still lacking, we will go around this problem when necessary by introducing an appropriate cutoff, as has been done in previous work.[4,7,16,18]

As is well known, the solutions to Eq. (1) may exhibit "runaway" behavior, which can be eliminated by demanding that $\ddot{\mathbf{x}}\to 0$ as $t\to\infty$, as proposed by Ivanenko and Sokolov.[30] By an integration by parts and under this condition, Eq. (1) takes the form

$$m\ddot{\mathbf{x}} = \mathbf{F} + e\mathbf{E}_m(t), \tag{6}$$

where

$$\mathbf{E}_m(t) = \frac{1}{\tau}\int_t^\infty \mathbf{E}(t')e^{(t-t')/\tau}dt' \tag{7}$$

has the power spectrum

$$S_m = S_E(\omega)/(1 + \tau^2\omega^2), \tag{8}$$

as follows immediately from Eq. (3). Equation (6) shows preacceleration with respect to the stochastic force $\mathbf{E}_m(t)$ only [see Eq. (7)], being free of this problem with respect to the external force as this external force has a constant value.

We assume well-defined (nonstochastic) initial conditions $\mathbf{x}(0) = \mathbf{x}_0$, $\mathbf{p}(0) = \mathbf{p}_0$; then Eq. (6) has the solutions

$$\mathbf{p}(t) \equiv m\dot{\mathbf{x}}(t) = \mathbf{p}_0 + \mathbf{F}t + e\int_0^t \mathbf{E}_m(t')\,dt', \tag{9a}$$

$$\mathbf{x}(t) = \mathbf{x}_0 + (1/m)\int_0^t \mathbf{p}(t')\,dt'. \tag{9b}$$

For the average of these equations over the ensemble of realizations of the stochastic field $\mathbf{E}_m(t)$, recalling that $\langle \mathbf{E}_m\rangle = 0$, we get

$$\langle \mathbf{p}(t)\rangle = \mathbf{p}_0 + \mathbf{F}t, \tag{10a}$$

$$\langle \mathbf{x}(t)\rangle = \mathbf{x}_0 + (\mathbf{p}_0/m)t + (\mathbf{F}/2m)t^2, \tag{10b}$$

whereas the correlations between the different components of $\mathbf{p}(t)$ and $\mathbf{x}(t)$ become

$$\langle p_i(t)p_j(t')\rangle = \langle p_i(t)\rangle\langle p_j(t')\rangle + C_{ij}(t, t'), \tag{11a}$$

$$\langle x_i(t)p_j(t')\rangle = \langle x_i(t)\rangle\langle p_j(t')\rangle + (1/m)\int_0^t dt_1\, C_{ij}(t_1, t'), \tag{11b}$$

$$\langle x_i(t)x_j(t')\rangle = \langle x_i(t)\rangle\langle x_j(t')\rangle + (1/m^2)\int_0^t dt_1\int_0^{t'} dt_2\, C_{ij}(t_1, t_2), \tag{11c}$$

where the tensor $C_{ij}(t, t')$ is given by

$$C_{ij}(t, t') = e^2\int_0^t dt_1\int_0^{t'} dt_2\langle E_{mi}(t_1)E_{mj}(t_2)\rangle. \tag{12}$$

With the help of Eqs. (3) and (7), Eq. (12) may be given the form

$$C_{ij}(t, t') = e^2\delta_{ij}\int_0^\infty d\omega\, S_m(\omega)\int_0^t dt_1\int_0^{t'} dt_2\cos\omega(t_1 - t_2)$$

or, performing the time integrations,

$$C_{ij}(t,t')$$
$$= e^2\delta_{ij}\int_0^\infty d\omega\,[S_m(\omega)/\omega^2]$$
$$\times[1-\cos\omega t-\cos\omega t'+\cos\omega(t-t')] \tag{13a}$$

and also

$$\int_0^t C_{ij}(s,t')\,ds$$
$$= e^2\delta_{ij}\int_0^\infty d\omega\,[S_m(\omega)/\omega^3]$$
$$\times[\omega t-\omega t\cos\omega t'-\sin\omega t+\sin\omega t'$$
$$+\sin\omega(t-t')], \tag{13b}$$

$$\int_0^t ds\int_0^{t'} ds'\,C_{ij}(s,s')$$
$$= e^2\delta_{ij}\int_0^\infty d\omega\,[S_m(\omega)/\omega^4]$$
$$\times[\omega^2 tt'-\omega t'\sin\omega t-\omega t\sin\omega t$$
$$+1-\cos\omega t'+\cos\omega(t-t')-\cos\omega t]. \tag{13c}$$

From these results and Eqs. (11) we see that all covariance matrices are diagonal. In particular, we get for the variance of p_i

$$\sigma_p^2(t)=2e^2\int_0^\infty[S_m(\omega)/\omega^2](1-\cos\omega t)\,d\omega, \tag{14}$$

the covariance at equal times of x_i and p_i,

$$\Gamma_{xp}(t)=(1/2m)t\sigma_p^2(t), \tag{15}$$

and the variance of x_i,

$$\sigma_x^2(t)=\frac{2e^2}{m^2}\int_0^\infty d\omega\,\frac{S_m(\omega)}{\omega^2}\left(\frac{1}{2}t^2-\frac{t}{\omega}\sin\omega t-\frac{1}{\omega^2}\cos\omega t+\frac{1}{\omega^2}\right).$$

Noticing that the factor within parentheses in the integrand is just the value of the integral $\int_0^t dt'\,t'(1-\cos\omega t')$, we may rewrite the last result in the form

$$\sigma_x^2(t)=(1/m^2)\int_0^t t'\sigma_p^2(t')\,dt'. \tag{16}$$

Deriving this expression with respect to time and noticing Eq. (15), we see that

$$\frac{d\sigma_x^2}{dt}=\frac{2}{m}\Gamma_{xp}(t). \tag{17}$$

This is a general result that may be derived directly from the definitions of σ_x^2 and Γ_{xp}.

To get an expression for $\sigma_p^2(t)$, we substitute Eqs. (4) and (8) into Eq. (14):

$$\sigma_p^2(t)=\frac{2\hbar m\tau}{\pi}\int_0^\infty\frac{\omega}{1+\tau^2\omega^2}\left(1+\frac{2}{e^{\beta\hbar\omega}-1}\right)\times(1-\cos\omega t)\,d\omega. \tag{18}$$

This expression contains a divergent term, that which comes from the product of the first terms within the parentheses; we regularize it by introducing a cutoff at frequency ω_c, as said before, which we select as the pair-production frequency[4,7,16,18]

$$\omega_c=2mc^2/\hbar. \tag{19}$$

Since $\tau\omega_c\sim\alpha$ (α is the fine structure constant), we can now approximate the factor $(1+\tau^2\omega^2)^{-1}$ in Eq. (18) by 1, the corrections being of order α^2 and hence negligible. Developing the thermal contribution to the spectrum in a power series of $e^{-\beta\hbar\omega}$, we get from Eq. (18)

$$\sigma_p^2(t)=\frac{2\hbar m}{\pi}\left[\int_0^{\omega_c}\omega(1-\cos\omega t)\,d\omega+2\sum_{n=1}^\infty\int_0^\infty\omega e^{-n\beta\hbar\omega}(1-\cos\omega t)\,d\omega\right]$$

or

$$\sigma_p^2(t)=\sigma_0^2\left[1-\frac{2}{\omega_c t}\sin\omega_c t+\frac{2}{\omega_c^2t^2}-\frac{2}{\omega_c^2t^2}\cos\omega_c t\right]+\frac{4\sigma_0^2}{\omega_c^2}\sum_{n=1}^\infty\left[\frac{1}{n^2\beta^2\hbar^2}-\frac{n^2\beta^2\hbar^2-t^2}{(n^2\beta^2\hbar^2+t^2)^2}\right], \tag{20}$$

where

$$\sigma_0^2=(\hbar m\tau/\pi)\omega_c^2=(8\alpha/3\pi)m^2c^2. \tag{21}$$

When $t\gg\beta\hbar\gg\omega_c^{-1}$, this function reaches the stationary value

$$\sigma_p^2=\sigma_0^2\left[1+\tfrac{1}{6}(2\pi/\beta\hbar\omega_c)^2\right]. \tag{22}$$

At room temperatures, $\beta\hbar\sim10^{-14}$ s, so that the particle will reach this stationary dispersion very fast. However, as $T\to0$ the time needed to reach the stationary value for σ_p^2 becomes arbitrarily large. Also, since at room temperatures $\beta^{-1}\sim10^{-2}$ eV whereas $\hbar\omega_c\sim1$ MeV, for all usual temperatures the thermal corrections to σ_p^2 are entirely negligible, and we may approximate σ_p^2 by σ_0^2. This is quite understandable, since the energy $\hbar\omega/2$ of the mode ω of the zero-point field greatly exceeds the thermal contribution to the spectrum at the same frequency ω, unless the temperature reaches extremely high values. Therefore, for $t\gg\omega_c^{-1}$ and not unusually high temperatures we may write the variance of the momentum as

$$\sigma_{p_i}^2=\sigma_0^2. \tag{23}$$

Obviously, if we assumed that the initial momentum is distributed with variance $\sigma_{p_0}^2$, but uncorrelated with the radiation field, instead of Eq. (23) we would get

$$\sigma_{p_i}^2=\sigma_{p0i}^2+\sigma_0^2. \tag{24}$$

These results assign to the parameter σ_0^2 the meaning of the minimum (or intrinsic) asymptotic momentum variance for the free particle at temperature $T=0$ (the presence of the homogeneous field does not affect this result). According to Eq. (21), σ_0^2 is strongly cutoff-dependent; in particular, for the selected cutoff [see Eq. (1)] it takes a fairly high value that corresponds to an energy associated to these fluctuations of order αmc^2, and which will usually be treated as part of the self-energy of the free particle, by means of a renormalization procedure.

These results, whose numerical value depends on the value adopted for ω_c and hence must be considered merely a

crude evaluation of order of magnitude, are related to an effect discussed several years ago by Boyer[5] and lately also by Rueda.[7] The Boyer effect may easily be shown by noticing that for $\omega_c \to \infty$, the variance of the momentum of the particle grows without limit, which means that the zero-point radiation field may accelerate free particles to any arbitrarily high velocity. In his penetrating analysis on the origin of the Planck spectrum,[5] Boyer restores equilibrium by confining the particles in a box and assuming that the radiation energy lost during the collisions with the walls accounts for the balance. In his turn, Rueda[7] sees in Boyer's effect a possible mechanism for explaining the high velocities attained by cosmic ray particles.

It is possible to write Eq. (20) in closed form in terms of the function

$$\phi(z) = \sum_{n=1}^{\infty} \frac{1}{n^2 + z^2}, \tag{25a}$$

whose numerical value for real z may be calculated with the formula (Ref. 31, p. 36)

$$\sum_{n=1}^{\infty} \frac{1}{n^2 + z^2} = \frac{\pi}{2z}\left(\coth \pi z - \frac{1}{\pi z}\right). \tag{25b}$$

We get

$$\sigma_p^2(t) = \sigma_0^2\left[1 - \frac{2}{\omega_c t}\sin \omega_c t + \frac{2}{\omega_c^2 t^2}(1 - \cos \omega_c t)\right] + \sigma_0^2\left(\frac{2}{\beta\hbar\omega_c}\right)^2\left[\frac{\pi^2}{6} - \phi\left(\frac{t}{\beta\hbar}\right) - t\frac{d}{dt}\phi\left(\frac{t}{\beta\hbar}\right)\right]. \tag{26}$$

With this result and Eq. (15) we get a closed expression for $\Gamma_{xp}(t)$. To calculate $\sigma_x^2(t)$, we use Eq. (16):

$$\sigma_x^2(t) = \frac{\sigma_0^2}{2m^2} t^2\left\{1 + \left(\frac{2}{\beta\hbar\omega_c}\right)^2\left[\frac{\pi^2}{6} - 2\phi\left(\frac{t}{\beta\hbar}\right)\right] + \left(\frac{2}{\omega_c t}\right)^2\left[\gamma - 1 + \cos \omega_c t + \ln \omega_c t - \ln \frac{\pi t}{\beta\hbar} + \ln \sinh \frac{\pi t}{\beta\hbar} - \operatorname{ci}(\omega_c t)\right]\right\}, \tag{27}$$

where γ is Euler's constant. Since for $t \gg \beta\hbar \gg \omega_c^{-1}$ the cosine integral $\operatorname{ci}(\omega_c t) = -\int_{\omega_c t}^{\infty}[(\cos x)/x]\,dx$ tends to zero and the sinh may be approximated by one-half the exponential, $\sigma_x^2(t)$ reduces to

$$\sigma_x^2(t) = \frac{\sigma_0^2}{2m^2} t^2\left\{1 + \left(\frac{2}{\beta\hbar\omega_c}\right)^2\left[\frac{\pi^2}{6} + \frac{\pi\beta\hbar}{2t} - 2\phi\left(\frac{t}{\beta\hbar}\right)\right]\right\}, \tag{28}$$

which means a quadratic asymptotic dependence on t for the variance of x_i,

$$\sigma_x^2(t) = (\sigma_p^2/2m^2)t^2, \tag{29}$$

where σ_p^2 is given by Eq. (22).

Summing up, for $t \gg \beta\hbar$ the components of the correlation tensors between the components of $\mathbf{x}$ and $\mathbf{p}$ at equal times are

$$\langle p_i p_j \rangle = \langle p_i \rangle \langle p_j \rangle + \delta_{ij}\sigma_p^2, \tag{30a}$$

$$\langle x_i p_j \rangle = \langle x_i \rangle \langle p_j \rangle + \delta_{ij}(\sigma_p^2/2m)t, \tag{30b}$$

$$\langle x_i x_j \rangle = \langle x_i \rangle \langle x_j \rangle + \delta_{ij}(\sigma_p^2/2m^2)t^2. \tag{30c}$$

III. PHASE-SPACE DENSITY

A. The phase-space density

Since $\mathbf{x}(t)$ and $\mathbf{p}(t)$ are both linear functionals of $\mathbf{E}(t)$ when the system is linear, and $\mathbf{E}(t)$ has a Gaussian distribution, for well-defined initial values of $\mathbf{x}$ and $\mathbf{p}$ the phase-space distribution is a bivariate Gaussian centered around the values $\langle \mathbf{x} \rangle$ and $\langle \mathbf{p} \rangle$ given by Eqs. (10) and with covariance matrices given by

$$\sigma_{p_{ij}}^2 = \delta_{ij}\sigma_p^2(t), \tag{31a}$$

$$\Gamma_{xp_{ij}} = \delta_{ij}\frac{t}{2m}\sigma_p^2(t) = \delta_{ij}\Gamma_{xp}, \tag{31b}$$

$$\sigma_{x_{ij}}^2 = \delta_{ij}\sigma_x^2(t), \tag{31c}$$

where σ_x^2 is given by Eq. (27). Explicitly,

$$P(\mathbf{x}, \mathbf{p}, t) = \frac{1}{(2\pi)^3(\sigma_x^2\sigma_p^2 - \Gamma_{xp}^2)^{3/2}} \exp\left\{-\frac{1}{2(\sigma_x^2\sigma_p^2 - \Gamma_{xp}^2)} \times \left[\sigma_p^2(\mathbf{x} - \langle \mathbf{x} \rangle)^2 - 2\Gamma_{xp}(\mathbf{x} - \langle \mathbf{x} \rangle)\cdot(\mathbf{p} - \langle \mathbf{p} \rangle) + \sigma_x^2(\mathbf{p} - \langle \mathbf{p} \rangle)^2\right]\right\}. \tag{32}$$

For $t \gg \beta\hbar \gg \omega_c^{-1}$ Eqs. (30) hold and hence

$$\sigma_x^2\sigma_p^2 - \Gamma_{xp}^2 \to \frac{\sigma_p^4}{4m^2}t^2.$$

The phase-space density then takes the simple form

$$P(\mathbf{x}, \mathbf{p}, t) = \left(\frac{m}{\pi\sigma_p^2 t}\right)^3 \exp\left[-\frac{2m^2}{\sigma_p^2 t^2}(\mathbf{x} - \langle \mathbf{x} \rangle)^2 + \frac{2m}{\sigma_p^2 t}(\mathbf{x} - \langle \mathbf{x} \rangle)\cdot(\mathbf{p} - \langle \mathbf{p} \rangle) - \frac{1}{\sigma_p^2}(\mathbf{p} - \langle \mathbf{p} \rangle)^2\right], \tag{33}$$

where σ_p^2 is given by Eq. (22).

B. Comments on the validity of the formula for the phase-space density

Since in the following we will make extensive use of Eq. (33) it seems opportune to make some comments on it. First of all, we notice that this density depends on the temperature through σ_p^2 only, which in its turn has a very weak temperature dependence, as we say in Sec. II; hence, $P(\mathbf{x}, \mathbf{p})$ is very insensitive to temperature changes.

A more interesting comment refers to the validity of Eq. (33). In the case of a constant external force, the AL equation is applicable and only the term $\mathbf{E}_m(t)$ [see Eq. (7)] may show preacceleration effects; but these effects are noticeable only for times $t \lesssim \tau$, i.e., for frequencies $\omega \gg 2\pi/\tau \sim \omega_c/\alpha$, and hence can do no harm since they fall outside the physical spectrum considered. In other words, the acausal effects due to preacceleration are entirely negligible in this case and Eq. (33) is fully legitimate. A little more caution is needed when considering the free particle, since in this case it may be argued that the AL equation itself does not apply. In fact, for an absolutely *free* particle it does not apply, as is well

known[27,32]; however, the particle we are dealing with is not free, but subject to the stochastic force $\mathbf{E}(t)$. As stated in the Introduction, there are arguments to support the idea that the presence of this field—which at $T=0$ subjects the particle to such an intense interaction that it becomes essentially insensitive to thermal perturbations, as we have seen—is enough to guarantee the applicability of the AL equation. This belief is reinforced by observing that the distribution $P(\mathbf{x}, \mathbf{p})$ given by Eq. (33) goes smoothly to a well-defined limit when $\mathbf{F}\to 0$, since $\mathbf{F}$ enters linearly in $\langle \mathbf{x} \rangle$ and $\langle \mathbf{p} \rangle$.

It is possible to confirm rigorously the validity of this conclusion by performing a more careful calculation, which we omit entirely in order not to complicate and obscure the present calculation, but which we may explain as follows. As discussed in the Introduction, it is possible to construct an entirely causal theory of the radiating particle, by taking its structure into account[24,25]; starting from this theory, we may calculate $\mathbf{p}(t)$ and from it the covariance $\langle p_i p_j \rangle$ and its limit when $t\to\infty$. For example, a concrete calculation of this kind can be done on the basis of the results presented in §§IV of Ref. 25. The point is that for times long enough we recover in this way Eq. (14), and hence the whole of the above theory, thus confirming the applicability of Eq. (33) to the free particle. The advantage of the present procedure is its relative simplicity and clarity.

IV. RELATIONSHIP WITH WIGNER'S DISTRIBUTION

A. Quantum mechanics and SED are different theories

A straightforward application of stochastic theory has led us to the phase-space density $P(\mathbf{x}, \mathbf{p}, t)$ as the basis for a complete statistical description of the behavior of the mechanical system. Our present purpose is to compare the above results with those predicted by quantum mechanics for the same system under the same conditions and investigate if any reasonable connection may be established between both. First of all, we must stress that we are dealing with two theories that give different predictions for many important quantities. To show this in a simple way, we start from Eq. (29), which we rewrite for convenience as

$$\sigma_x^2(t) = (\sigma_p^2/2m^2)t^2, \tag{29}$$

with σ_p^2 given by Eqs. (21) and (22), and compare with the corresponding quantum mechanical result (which we will derive below),

$$\sigma_x^2(t) = (\sigma_p^2/m^2)t^2. \tag{34}$$

The word "corresponding" seems to require some explanation. We are here working with a particular solution to our problem, namely the one that corresponds to a Gaussian distribution for all times. We could as well consider more complicated—even non-Gaussian—initial conditions, but this would unnecessarily complicate the problem. In the quantum-mechanical case we also have an infinity of possible solutions, depending on the initial conditions. We should therefore compare Gaussian solutions of both theories; this is the meaning of the word "corresponding." A word of caution is, however, necessary: In the quantum-mechanical case we cannot go to the dispersionless limit ($t\to 0$) because in this limit the density would violate the Heisenberg inequalities. The comparison can only be made for times such that the Heisenberg inequalities hold, as discussed in the next section.

As seen from Eqs. (29) and (34), both theories predict the same time dependence (for large times; SED gives a more detailed description for smaller times); nevertheless, there exists an essential difference, associated to the presence of the factor $\frac{1}{2}$ in Eq. (29). This factor is responsible for several important differences in the behavior of the SED particle with respect to that of quantum mechanics, as we will see below. We will have an opportunity below to discuss several related points.

Before making a deeper comparison with quantum mechanics, it seems worthwhile to make the following consideration. Since we have assumed well-defined initial conditions, the initial dispersion of both x_i and p_i are zero, and so also is their product; for $\omega_c t \gg \beta\hbar\omega_c \gg 1$ we have, however, that

$$\sigma_x^2\sigma_p^2 = (\sigma_0^4/2m^2)t^2.$$

The onset of the quantum regime[18] is reached when this quantity attains a value not smaller than $\hbar^2/4$. This consideration offers us a heuristic interpretation of the Heisenberg inequalities, namely, that there exists a minimum (Heisenberg) time t_H needed for the system to relax into the quantum-mechanical regime, when starting from well-defined initial conditions in phase space. With the aid of Eqs. (19) and (21), we may estimate this t_H to be such that

$$\omega_c t_H \geqslant 3\pi/4\sqrt{2}\alpha,$$

a result that is stronger than the simple requirement $\omega_c t \gg 1$. This is easily understood by recalling that Eq. (29) is legitimate only for $\omega_c t \gg \beta\hbar\omega_c = 2mc^2/k_B T$, which tends to infinity when $T\to 0$ and is much greater than unity even for room temperatures. Therefore, we conclude that when Eqs. (30) apply, we are surely well within the quantum regime, but a long time is needed to reach this regime at normal or small temperatures.

B. The Wigner distribution

Now we go over to a more quantitative comparison with quantum mechanics. We start by constructing the quantum-mechanical phase-space distribution, i.e., the Wigner distribution of the problem, which we denote by $P_W(\mathbf{x}, \mathbf{p}, t)$. Let us recall that for a general potential $V(\mathbf{x})$ the equation of evolution of P_W is[33]

$$\frac{\partial P_W}{\partial t} + \frac{\mathbf{p}}{m}\cdot\nabla P_W - \frac{2}{\pi^3\hbar^4}\int d^3x' \int d^3p'\, P_W(\mathbf{x}, \mathbf{p}', t) V(\mathbf{x}') \times \sin\frac{2}{\hbar}(\mathbf{x}-\mathbf{x}')\cdot(\mathbf{p}-\mathbf{p}') = 0.$$

For $V = -\mathbf{F}\cdot\mathbf{x}$, as is our case, this reduces to

$$\frac{\partial P_W}{\partial t} + \frac{\mathbf{p}}{m}\cdot\nabla P_W + \mathbf{F}\cdot\nabla_p P_W = 0. \tag{35}$$

We see that the equation of evolution of the Wigner distribution coincides with the *classical* Liouville equation of the problem. It is easy to check that Eq. (35) does not accept a

solution of the form of Eq. (33); therefore, to construct the P_W of interest, we proceed as follows. As a first step, we notice that Eq. (35) accepts solutions of the form

$$P_W(\mathbf{x}, \mathbf{p}, t) = \Phi_1(\mathbf{p} - \langle \mathbf{p} \rangle)\Phi_2(\mathbf{x} - \langle \mathbf{x} \rangle - (t/m)(\mathbf{p} - \langle \mathbf{p} \rangle))$$

with $\langle \dot{\mathbf{p}} \rangle = \mathbf{F}$ and $\langle \dot{\mathbf{x}} \rangle = \langle \mathbf{p} \rangle / m$, which means that $\langle \mathbf{p} \rangle$ and $\langle \mathbf{x} \rangle$ are given by Eqs. (10). For a classical (deterministic) problem we would demand that $\mathbf{p} = \langle \mathbf{p} \rangle$ and $\mathbf{x} = \langle \mathbf{x} \rangle$, and, to guarantee that P_W differs from zero only if these conditions are met, we would select both functions Φ_1 and Φ_2 as Dirac deltas. In the present problem, as discussed previously, we are interested in a distribution which contains at least a Gaussian factor in the momentum with a constant variance. Therefore, we take the particular solution

$$P_W(\mathbf{x}, \mathbf{p}, t) = \frac{1}{(\pi\sqrt{2}\sigma_0\sigma)^3} \exp\left[-\frac{(\mathbf{p} - \langle \mathbf{p} \rangle)^2}{\sigma^2} - \frac{(\mathbf{x} - \langle \mathbf{x} \rangle - (\mathbf{p} - \langle \mathbf{p} \rangle)t/m)^2}{2\sigma_{0x}^2} \right],$$

which reproduces the classical result in the limit $\sigma \to 0$, $\sigma_{0x} \to 0$. Since Eq. (35) is simultaneously a classical and a quantum-mechanical law, we must take care to guarantee that our particular solution is indeed consistent with quantum mechanics. This is achieved by selecting σ and σ_{0x} such that Heisenberg's inequality $\sigma_{(W)x}^2 \sigma_{(W)p}^2 \geqslant \hbar^2/4$ holds for any time, or, as follows directly from P_W,

$$\tfrac{1}{2}\sigma^2(\sigma_{0x}^2 + \tfrac{1}{2}\sigma^2 t^2/m^2) \geqslant \hbar^2/4.$$

Hence, it suffices to take $\sigma^2\sigma_{0x}^2 \geqslant \hbar^2/2$ to guarantee consistency with quantum mechanics. For times longer than $\hbar m/\sigma$ (which for $\sigma \sim \sigma_p$ will be of the order of t_H), the time-dependent contribution to $\sigma_x(t)$ is dominant and, as time grows, σ_{0x} becomes negligible. This observation is important, since in the SED calculation we have arbitrarily restricted ourselves to the case in which the initial dispersions of $\mathbf{x}$ and $\mathbf{p}$ are both zero and have even made the approximation valid for $t \gg \beta\hbar$, which allows us to write P in the simplified form of Eq. (33). Hence, in order to be able to compare both theories without having to complicate the calculations, we restrict ourselves to times $t \gg t_H$ in the quantum-mechanical case, which allows us to neglect σ_{0x}. In this approximation, the Wigner distribution becomes

$$P_W(\mathbf{x}, \mathbf{p}, t) = (\pi\sigma^2)^{-3/2} e^{-(\mathbf{p} - \langle \mathbf{p} \rangle)^2/\sigma^2} \delta(\mathbf{x} - \langle \mathbf{x} \rangle - (\mathbf{p} - \langle \mathbf{p} \rangle)t/m). \tag{36}$$

This restriction, while very convenient, is not crucial to the development of the theory which follows. We must still select a value for σ^2, which we do by demanding that P and P_W have as many desired properties in common as possible. In particular, since the averages of x_i and p_i are the same in both cases, we pay attention to their correlations. Identifying the averages calculated with the aid of P_W with an index W, we easily get from Eq. (36)

$$\langle p_i p_j \rangle_W = \langle p_i \rangle \langle p_j \rangle + \tfrac{1}{2}\lambda\delta_{ij}\sigma_p^2, \tag{37a}$$

$$\langle x_i p_j \rangle_W = \langle x_i \rangle \langle p_j \rangle + \lambda\delta_{ij}(\sigma_p^2/2m)t, \tag{37b}$$

$$\langle x_i x_j \rangle_W = \langle x_i \rangle \langle x_j \rangle + \lambda\delta_{ij}(\sigma_p^2/2m^2)t^2, \tag{37c}$$

where we have written for convenience

$$\sigma^2 = \lambda\sigma_p^2, \tag{38}$$

with λ to be fixed. From these results it immediately follows that

$$\sigma_{(W)x}^2(t) = (\sigma_{(W)p}^2/m^2)t^2, \tag{39}$$

which is just Eq. (34) and justifies our previous discussion. Comparing Eqs. (37) and Eqs. (30), we see that no value of λ exists that may give the same second-order moments in both theories; the best we can do is to take $\lambda = 1$, i.e., $\sigma^2 = \sigma_p^2$, a choice that we accept in what follows. With this selection, P_W predicts an asymptotic dispersion of the momentum equal to one-half of the one given by SED [see Eq. (45b) below], both theories predicting, however, the same value for σ_x^2 and Γ_{xp}. Despite the fact that Γ_{xp} is the same in both theories, the correlation factor, defined as

$$\rho(\mathbf{x}, \mathbf{p}, t) = \Gamma_{xp}(t)/\sigma_x(t)\sigma_p(t) \tag{40}$$

is not the same, due to the different value of $\sigma_p(t)$. In particular, for SED one gets from Eqs. (30) $\rho = 1/\sqrt{2}$, whereas for quantum mechanics one gets from Eqs. (39) $\rho_{(W)} = 1$. The relationship between ρ and $\rho_{(W)}$ is actually independent of the value adopted for λ, since quite generally Eqs. (37) give

$$\rho_{(W)} = \frac{\Gamma_W}{(\sigma_{Wx}^2\sigma_{Wp}^2)^{1/2}} = \frac{\lambda\Gamma}{(\lambda\sigma_x^2\lambda\sigma_p^2/2)^{1/2}} = \sqrt{2}\rho. \tag{41}$$

The surprising result $\rho_{(W)} = 1$ means that in the Wigner distribution the variables x_i and p_i are totally correlated, i.e., that one determines the other, a result obviously incompatible with a stochastic theory as SED, thus showing through a particular instance that QM cannot admit in general a strictly statistical interpretation, or, what is equivalent, that any stochastic theory will necessarily give some predictions at variance with QM—with only "accidental" exceptions, as is the case of the harmonic oscillator.[16] In some sense, this assertion may be considered the stochastic version of von Neumann's theorem.

The origin of the result $\rho_{(W)} = 1$ lies in the δ factor of the Wigner distribution. To see this, consider the auxiliary quantity $\langle [\alpha(p_i - \langle p_i \rangle) + (x_i - \langle x_i \rangle)]^2 \rangle_W$, which is obviously nonnegative for real α; developing this, we get

$$\alpha^2\sigma_{(W)p}^2 + 2\alpha\Gamma_{(W)xp} + \sigma_{(W)x}^2 \geqslant 0.$$

For $\Gamma_{(W)} = \rho_{(W)}\sigma_{(W)x}\sigma_{(W)p} = \sigma_{(W)x}\sigma_{(W)p}$, the minimum of this inequality occurs for $\alpha = -\sigma_{(W)x}/\sigma_{(W)p}$. With this value we get

$$\int [(x_i - \langle x_i \rangle) - (\sigma_{(W)x}/\sigma_{(W)p})(p_i - \langle p_i \rangle)]^2 \times P_W(\mathbf{x}, \mathbf{p})\, d^3x\, d^3p = 0,$$

which demands that

$$P_W(\mathbf{x}, \mathbf{p}, t) \sim \delta\left[\mathbf{x} - \langle \mathbf{x} \rangle - (\sigma_{(W)x}/\sigma_{(W)p})(\mathbf{p} - \langle \mathbf{p} \rangle)\right],$$

as asserted above.

C. Relation between the Wigner and the SED distribution

This relationship is easily established by calculating the following integral, with a a real parameter to be determined:

$$\int P_{\rm W}(\mathbf{x},\mathbf{p}',t)e^{-(\mathbf{p}'-\mathbf{p})^2/a^2}\,d^3p'$$
$$=(\pi a^2)^{3/2}\left(\frac{m}{\pi\sigma_p at}\right)^3\exp\left[-\frac{(\mathbf{p}-\langle\mathbf{p}\rangle)^2}{a^2}\right.$$
$$+\frac{2m}{a^2t}(\mathbf{x}-\langle\mathbf{x}\rangle)\cdot(\mathbf{p}-\langle\mathbf{p}\rangle)$$
$$\left.-\frac{m^2}{t^2}\left(\frac{1}{\sigma_p^2}+\frac{1}{a^2}\right)(\mathbf{x}-\langle\mathbf{x}\rangle)^2\right].$$

Comparison with Eq. (33) shows that by selecting $a=\sigma_p$ we get the desired result, namely,

$$P(\mathbf{x},\mathbf{p},t)=(\pi\sigma_p^2)^{-3/2}\int P_{\rm W}(\mathbf{x},\mathbf{p}',t)e^{-(\mathbf{p}'-\mathbf{p})^2/\sigma_p^2}\,d^3p'. \tag{42}$$

Thus the SED phase-space density is given by a convolution of the Wigner distribution with a Gaussian in velocity space. As stated previously, this kind of relationship has been suggested repeatedly[28] as a means of avoiding some of the problems that appear when $P_{\rm W}(\mathbf{x},\mathbf{p})$ is directly interpreted as a phase-space density.

As a first application of this result, we establish the general relationship between the average of a dynamical variable $A(\mathbf{x},\mathbf{p})$ as calculated in quantum mechanics, $\langle A\rangle_{\rm W}$, or, according to SED, $\langle A\rangle$, of which Eqs. (37) are particular examples. We get

$$\langle A(\mathbf{x},\mathbf{p})\rangle=(\pi\sigma_p^2)^{-3/2}\left\langle\int A(\mathbf{x},\mathbf{p}')e^{-(\mathbf{p}'-\mathbf{p})^2/\sigma_p^2}\,d^3p'\right\rangle_{\rm W}. \tag{43}$$

Developing $A(\mathbf{x},\mathbf{p}')$ in a Taylor series around $\mathbf{p}$ and performing the integrations, we get

$$\langle A(\mathbf{x},\mathbf{p})\rangle=\langle A(\mathbf{x},\mathbf{p})\rangle_{\rm W}+\tfrac{1}{4}\sigma_p^2\langle\nabla_p^2A(\mathbf{x},\mathbf{p})\rangle_{\rm W}+\cdots. \tag{44}$$

Incidentally, we notice that Eq. (44) applies also for the $\mathbf{x}$-conditioned averages, i.e., when the integral with respect to $\mathbf{x}$ is not performed. From this result it follows that both theories predict the same value for the average of any dynamical variable that either does not contain $\mathbf{p}$ (i.e., any function of $\mathbf{x}$), or depends linearly on $\mathbf{p}$. Hence, in particular,

$$\langle\mathbf{x}\rangle_{\rm W}=\langle\mathbf{x}\rangle,\quad\langle\mathbf{p}\rangle_{\rm W}=\langle\mathbf{p}\rangle,$$
$$\sigma^2_{({\rm W})x}=\sigma_x^2,\quad\Gamma_{({\rm W})xp}=\Gamma_{xp}, \tag{45}$$

in accordance with previous results. However, if $A(\mathbf{x},\mathbf{p})$ contains terms that depend quadratically or on higher powers of $\mathbf{p}$, the averages will differ from one another. For example, for $A=p_ip_j$ one gets

$$\langle p_ip_j\rangle=\langle p_ip_j\rangle_{\rm W}+\tfrac{1}{2}\sigma_p^2\delta_{ij}, \tag{46a}$$

and, using Eq. (31a), we get the already discussed relation

$$\sigma^2_{({\rm W})p}=\tfrac{1}{2}\sigma_p^2. \tag{46b}$$

These differences are more clearly reflected in the fact that whereas both phase-space distributions P and $P_{\rm W}$ lead to the same marginal distribution in configuration space, i.e.,

$$\int P(\mathbf{x},\mathbf{p},t)\,d^3p=\int P_{\rm W}(\mathbf{x},\mathbf{p},t)\,d^3p, \tag{47}$$

the marginal distributions in momentum space are different, as follows by integrating Eq. (42) with respect to $\mathbf{x}$. The explicit form of the marginal distribution in configuration space is

$$\rho_x(\mathbf{x},t)=\rho_{({\rm W})x}(\mathbf{x},t)$$
$$=\left[\frac{m}{\sqrt{\pi}\sigma_p t}\right]^3\exp\left[-\frac{m^2}{\sigma_p^2t^2}(\mathbf{x}-\langle\mathbf{x}\rangle)^2\right], \tag{48}$$

whereas for the momentum space one gets

$$\rho_{({\rm W})p}(\mathbf{p},t)=(\pi\sigma_p^2)^{-3/2}e^{-(\mathbf{p}-\langle\mathbf{p}\rangle)^2/\sigma_p^2} \tag{49}$$

and

$$\rho_p(\mathbf{p},t)=(2\pi\sigma_p^2)^{-3/2}e^{-(\mathbf{p}-\langle\mathbf{p}\rangle)^2/2\sigma_p^2}. \tag{50}$$

The relationship between $\rho_{({\rm W})p}$ and ρ_p follows immediately from Eq. (42) and is

$$\rho_p(\mathbf{p},t)=(\pi\sigma_p^2)^{-3/2}\int\rho_{({\rm W})p}(\mathbf{p}',t)e^{-(\mathbf{p}-\mathbf{p}')^2/\sigma_p^2}\,d^3p'. \tag{51}$$

Hence the two theories predict the same behavior of the system in configuration space, a point that we will confirm below more explicitly. Now, in quantum mechanics the configuration space distribution is most easily determined by solving the Schrödinger equation. This means that the Schrödinger wavefunction will also furnish an exact—though incomplete—description of the SED system. To explore more precisely the role of these functions in SED is the purpose of the next section.

V. THE PHASE-SPACE DISTRIBUTION IN TERMS OF SCHRÖDINGER AMPLITUDES

A. The quantum-mechanical case

We want to express explicitly the phase-space density presented by SED in terms of solutions to the Schrödinger equation. Since this is quite easily achieved for the Wigner distribution by following the rules of quantum mechanics and using afterwards Eq. (42), we first concentrate on the quantum-mechanical problem. We shall first construct the density matrix in the $\mathbf{x}$ representation associated with $P_{\rm W}$ and then express the result in terms of Schrödinger amplitudes. As is well known, the density matrix $\rho_{\rm W}(\mathbf{r},\mathbf{r}',t)$ is related to $P_{\rm W}(\mathbf{x},\mathbf{p},t)$ by the Weyl transform,[34,35] which consists of the following two successive steps: (i) First, we take the Fourier transform of $P_{\rm W}(\mathbf{x},\mathbf{p},t)$ with respect to the variable $\mathbf{p}$, to construct the function $\tilde{P}_{\rm W}(\mathbf{x},\mathbf{z},t)$:

$$\tilde{P}_{\rm W}(\mathbf{x},\mathbf{z},t)=\int P_{\rm W}(\mathbf{x},\mathbf{p},t)e^{(2i/\hbar)\mathbf{p}\cdot\mathbf{z}}\,d^3p. \tag{52}$$

(ii) Secondly, in the new function we make the change of variables

$$\mathbf{r}=\mathbf{x}+\mathbf{z},\quad\mathbf{r}'=\mathbf{x}-\mathbf{z} \tag{53}$$

and define

$$\rho_{\rm W}(\mathbf{r},\mathbf{r}',t)=\tilde{P}_{\rm W}((\mathbf{r}+\mathbf{r}')/2,(\mathbf{r}-\mathbf{r}')/2t). \tag{54}$$

The resulting expresion for the density matrix is

$$\rho_{\rm W}(\mathbf{r},\mathbf{r}',t)=\left(\frac{m}{\sqrt{\pi}+\sigma_p}\right)^3\exp\left[-\frac{m^2}{4\sigma_p^2t^2}(\mathbf{r}+\mathbf{r}'-2\langle\mathbf{r}\rangle)^2\right.$$
$$\left.+\frac{im}{2\hbar t}(\mathbf{r}-\mathbf{r}')\cdot\left(\mathbf{r}+\mathbf{r}'-2\langle\mathbf{r}\rangle+\frac{2t}{m}\langle\mathbf{p}\rangle\right)\right]. \tag{55}$$

To express this density matrix in terms of Schrödinger amplitudes, we consider the orthogonal and complete set of functions

$$u_k(r) = (2\pi)^{-3/2} e^{i\mathbf{k}\cdot\mathbf{r}} \tag{56}$$

and develop the density matrix in terms of it

$$\rho_W(\mathbf{r}, \mathbf{r}', t) = \int C_{k,k'}(t, T) u_k(\mathbf{r}) u_{k'}^*(\mathbf{r}')\, d^3k\, d^3k', \tag{57}$$

where we have explicitly indicated the temperature dependence coming from σ_p^2. The functions $C_{k,k'}(t, T)$ are the matrix elements of the density matrix in the momentum representation and are given by

$$\begin{aligned} C_{k,k'}(t, T) &= \int \rho_W(\mathbf{r}, \mathbf{r}', t) u_k^*(\mathbf{r}) u_{k'}(\mathbf{r}')\, d^3r d^3r' \\ &= \left(\frac{\hbar}{\sqrt{\pi}\sigma_p}\right)^3 \\ &\quad \times \exp\left\{ -\frac{1}{\sigma_p^2}\left[\mathbf{p}_0 + \mathbf{F}t - \frac{\hbar}{2}(\mathbf{k}+\mathbf{k}')\right]^2 \right. \\ &\quad \left. + i(\mathbf{k}'-\mathbf{k})\cdot\left[\mathbf{r}_0 - \frac{\mathbf{F}}{2m}t^2 + \frac{\hbar}{2m}(\mathbf{k}'+\mathbf{k})t\right]\right\}, \end{aligned} \tag{58}$$

where we used Eqs. (10) for $\langle \mathbf{r} \rangle$ and $\langle \mathbf{p} \rangle$. We want to separate explicitly the temperature from the time dependence in this expression; for this purpose, it is convenient to analyze separately two cases.

1. The free particle

If in Eq. (58) we put $F_i = 0$, the coefficients $C_{k,k'}$ separate quite naturally into a factor $\Delta_{k+k'}(\mathbf{p}_0, T)$ that depends on the temperature through σ_T but not on time, and factors that depend on time but not on temperature. Calling the latter $a_k(\mathbf{x}_0, t)$, we may write

$$C_{k,k'}(t, T) = \Delta_{k+k'}(\mathbf{p}_0, T) a_k(\mathbf{x}_0, t) a_{k'}^*(\mathbf{x}_0, t) \tag{59}$$

with

$$\Delta_k(\mathbf{p}_0, T) = (\hbar/\sqrt{\pi}\sigma_p)^3 \exp\left[-(1/\sigma_p^2)(\mathbf{p}_0 - \tfrac{1}{2}\hbar\mathbf{k})^2 \right] \tag{60}$$

and

$$a_k(\mathbf{x}_0, t) = \exp\left[-(i\hbar\mathbf{k}^2/2m)t - i\mathbf{k}\cdot\mathbf{x}_0 \right]. \tag{61}$$

Thus Eq. (57) takes a form in which the temperature dependence of each term has been separated out:

$$\begin{aligned} &\rho_W(\mathbf{r}, \mathbf{r}', t) \\ &\quad = \int \Delta_{k+k'}(\mathbf{p}_0, T) \psi_k(\mathbf{r} - \mathbf{x}_0, t) \psi_{k'}^*(\mathbf{r}' - \mathbf{x}_0, t)\, d^3k d^3k', \end{aligned} \tag{62}$$

where the functions

$$\begin{aligned} \psi_k(\mathbf{r} - \mathbf{r}_0, t) &\equiv a_k(\mathbf{r}_0, t) u_k(\mathbf{r}) \\ &= (2\pi)^{-3/2} e^{-(i\hbar\mathbf{k}^2/2m)t + i\mathbf{k}\cdot(\mathbf{r} - \mathbf{r}_0)} \end{aligned} \tag{63}$$

are solutions to the Schrödinger equation for a free particle

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \psi. \tag{64}$$

2. The particle in a homogeneous field of force

When $F_i \neq 0$ the procedure of separation of the variables T and t becomes somewhat more involved. To simplify matters, it is convenient to first introduce the variable

$$\mathbf{q} = \mathbf{k} - (\mathbf{F}/\hbar)t \tag{65}$$

into Eq. (58), to get

$$\begin{aligned} C_{k,k'}(t, T) &= \left(\frac{\hbar}{\sqrt{\pi}\sigma_p}\right)^3 \int d^3q\, d^3q'\, e^{-(1/\sigma_p^2)[\mathbf{p}_0 - (\hbar/2)(\mathbf{q}+\mathbf{q}')]^2} \\ &\quad \times \exp\left[-i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}_0 + \frac{i\hbar^2}{6m} \sum_{j=1}^{3} \frac{1}{F_j} \right. \\ &\quad \left. \times (q_j^3 - q_j'^3 + k_j'^3 - k_j^3) \right] \\ &\quad \times \delta\left[\mathbf{q} - \left(\mathbf{k} - \frac{\mathbf{F}}{\hbar}t\right)\right] \delta\left[\mathbf{q}' - \left(\mathbf{k} - \frac{\mathbf{F}}{\hbar}t\right)\right]. \end{aligned} \tag{66}$$

Using the Fourier development of the δ function,

$$\delta(\mathbf{q} - \mathbf{k}) = (1/8\pi^3) \int e^{i(\mathbf{q}-\mathbf{k})\cdot\mathbf{y}}\, d^3y,$$

and interchanging the order of the integrations, we get

$$\begin{aligned} C_{k,k'}(t, T) &= \int w_{y,y'}(\mathbf{p}_0, T) \phi_y(\mathbf{k}, \mathbf{x}_0, t) \\ &\quad \times \phi_{y'}^*(\mathbf{k}', \mathbf{x}_0, t)\, d^3y\, d^3y', \end{aligned} \tag{67}$$

where

$$\begin{aligned} w_{y,y'}(\mathbf{p}_0, T) &= \int \Delta_{q+q'}(\mathbf{p}_0, t) \phi_y^*(\mathbf{q}, 0, 0) \\ &\quad \times \phi_{y'}(\mathbf{q}', 0, 0)\, d^3q\, d^3q' \end{aligned} \tag{68}$$

and

$$\begin{aligned} \phi_y(\mathbf{k}, \mathbf{x}_0, t) &= \frac{1}{(2\pi)^{3/2}} e^{-(i/\hbar)\mathbf{F}\cdot\mathbf{y}t} \\ &\quad \times \exp\left[i(\mathbf{y} - \mathbf{x}_0)\cdot\mathbf{k} - \frac{i\hbar^2}{6m} \sum_{j=1}^{3} \frac{k_j^3}{F_j} \right]. \end{aligned} \tag{69}$$

In Eq. (7) the dependence on temperature comes only through the function $w_{y,y'}$ that does not depend on time, while the time enters only in the functions $\phi_y(\mathbf{k}, \mathbf{x}_0, t)$, which do not depend on temperature. The latter functions are solutions to the Schrödinger equation for a constant force in the momentum representation:

$$i\hbar \frac{\partial \phi}{\partial t} = \frac{\hbar^2\mathbf{k}^2}{2m} \phi - \mathbf{F}\cdot(i\nabla_k - \mathbf{x}_0)\phi. \tag{70}$$

Substituting Eq. (67) into Eq. (57), we get the desired development of the density matrix in the $\mathbf{x}$ representation:

$$\begin{aligned} \rho_W(\mathbf{r}, \mathbf{r}', t) &= \int W_{y,y'}(\mathbf{p}_0, T) \psi_y(\mathbf{r} - \mathbf{x}_0, t) \\ &\quad \times \psi_{y'}^*(\mathbf{r}' - \mathbf{x}_0, t)\, d^3y\, d^3y', \end{aligned} \tag{71}$$

where the Schrödinger amplitudes are given by the Fourier transform of the ϕ_y,

$$\psi_y(\mathbf{x} - \mathbf{x}_0, t) = (2\pi)^{-3/2} \int \phi_y(\mathbf{k}, \mathbf{x}_0, t) e^{i\mathbf{k}\cdot\mathbf{x}}\, d^3k \tag{72}$$

and are solutions to the Schrödinger equation

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\psi - \mathbf{F}\cdot(\mathbf{x}-\mathbf{x}_0)\psi, \tag{73}$$

which is the Fourier transform of Eq. (70).

B. The SED case

With the previous results at hand one can almost immediately express $P(\mathbf{x}, \mathbf{p}, t)$ in terms of Schrödinger amplitudes. We start from the Fourier transform of Eq. (42), which reads

$$\tilde{P}(\mathbf{x}, \mathbf{z}, t) = \tilde{P}_W(\mathbf{x}, \mathbf{z}, t)e^{-(\sigma_p^2/\hbar^2)\mathbf{z}^2}, \tag{74}$$

rewrite $\tilde{P}_W$ in terms of ρ_W with the help of Eq. (54), and express the density matrix in terms of Schrödinger amplitudes; we will work out the case $F_j \neq 0$ only for the sake of brevity, but the procedure is equally well applicable to the free-particle case. The result is obviously

$$\tilde{P}(\mathbf{x}, \mathbf{z}, t) = e^{-(\sigma_p^2/\hbar^2)\mathbf{z}^2}\int W_{y,y'}(\mathbf{p}_0, T) \times\psi_y(\mathbf{x}-\mathbf{x}_0+\mathbf{z}, t)\psi_{y'}^*(\mathbf{x}-\mathbf{x}_0-\mathbf{z}, t)\,d^3y\,d^3y'. \tag{75}$$

Performing an inverse Fourier transform, we may recast these results into the form

$$P(\mathbf{x}, \mathbf{p}, t) = (2\pi^2\hbar)^{-3}\int d^3y\,d^3y'\,W_{y,y'}(\mathbf{p}_0, T) \times\int d^3k\int d^3k'\,\phi_{y'}^*(\mathbf{k}', \mathbf{x}_0, t)\phi_y(\mathbf{k}, \mathbf{x}_0, t)e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}} \times\int e^{-(\sigma_p^2/\hbar^2)\mathbf{z}^2 + i[\mathbf{k}+\mathbf{k}'-(2/\hbar)\mathbf{p}]\cdot\mathbf{z}}\,d^3z,$$

where we have expressed the amplitudes $\psi_y(\mathbf{x})$ in terms of their Fourier transform $\phi_y(\mathbf{k})$, as given by Eq. (72). The integral over $\mathbf{z}$ is

$$\int e^{-(\sigma_p^2/\hbar^2)\mathbf{z}^2 + i[\mathbf{k}+\mathbf{k}'-(2/\hbar)\mathbf{p}]\cdot\mathbf{z}}\,d^3z = \pi^3\Delta_{k+k'}(\mathbf{p}, T),$$

so our final result is

$$P(\mathbf{x}, \mathbf{p}, t) = (2\pi\hbar)^{-3}\int d^3y\,d^3y'\,d^3k\,d^3k'\,W_{y,y'}(\mathbf{p}_0, T) \times\Delta_{k+k'}(\mathbf{p}, T)e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}}\phi_y(\mathbf{k}, \mathbf{x}_0, t)\phi_{y'}^*(\mathbf{k}', \mathbf{x}_0, t). \tag{76}$$

The density matrix of SED is simpler to write, since this is just Eq. (75) with the change of variables given by Eq. (53); therefore,

$$\rho(\mathbf{r}, \mathbf{r}', t) = e^{-(\sigma_p^2/4\hbar^2)(\mathbf{r}-\mathbf{r}')^2}\rho_W(\mathbf{r}, \mathbf{r}'), \tag{77}$$

where the quantum-mechanical density matrix is given by Eq. (71). The complicated structure of Eq. (76) tends to hide more than it reveals; however, a simple but very important result may be obtained either by integrating with respect to $\mathbf{p}$ or else by taking $\mathbf{r} = \mathbf{r}'$ in Eq. (77). In fact, since the relationship between $P(\mathbf{x}, \mathbf{p})$ and $\rho(\mathbf{r}, \mathbf{r}')$ is just the same as between P_W and ρ_W [Eq. (54)], in SED just as in quantum mechanics the marginal distribution in configuration space is obtained by taking $\mathbf{z} = 0$ in P_W [see Eq. (52)], which by Eqs. (53) amounts to taking $\mathbf{r} = \mathbf{r}' = \mathbf{x}$. Hence, one gets for the marginal distribution in configuration space from Eq. (77)

$$\rho(\mathbf{x}, t) = \rho_W(\mathbf{x}, \mathbf{x}) = \int W_{y,y'}(\mathbf{p}_0, T)\psi_y(\mathbf{x}-\mathbf{x}_0, t)\psi_{y'}^*(\mathbf{x}-\mathbf{x}_0, t)\,d^3y\,d^3y'. \tag{78}$$

This result is easily derived from Eq. (76) by noticing that

$$\int\Delta_{k+k'}(\mathbf{p}, T)\,d^3p = \hbar^3$$

and using once more Eqs. (72). Equation (78) is just the same as Eq. (71) with $\mathbf{x} = \mathbf{r} = \mathbf{r}'$, i.e., the quantum-mechanical density in x-space. We are recovering Eq. (48), but this time it is obvious that all the information about the distribution in *configuration* space as predicted by SED is contained in the quantum-mechanical wavefunctions $\psi(\mathbf{x}, t)$.

These results show also that the transition from the marginal distributions $\rho(\mathbf{x})$ and $\rho(\mathbf{p})$ to the phase-space density $P(\mathbf{x}, \mathbf{p})$ is a highly nontrivial problem. Hence the almost purely mathematical attempts to go from usual quantum mechanics over to a truly statistical description in phase space[33,36] are almost certainly condemned to failure as long as they are not furthered by physical argument, supported in its turn by a deeper theory, as might be, e.g., stochastic electrodynamics, as in the present case.

VI. ENERGY BALANCE

We recall that $\sigma_p^2(t)$ evolves from the initial value $\sigma_p^2(0) = 0$ to a stationary value which is positive definite. At first sight it would thus seem that the free particle attains a stationary situation, in which it does not interchange any energy with the background field. We will show in this section that actually a more complicated situation holds, in which a net interchange of energy does exist between field and particle at each separate frequency.

Multiplying Eq. (1) by $\mathbf{v} = \dot{\mathbf{x}}$ and rearranging, we get an equation for the evolution of the energy:

$$\frac{d}{dt}\left[\frac{\mathbf{p}^2}{2m} - \mathbf{F}\cdot\mathbf{x} - m\tau\mathbf{v}\cdot\dot{\mathbf{v}}\right] = e\mathbf{E}\cdot\mathbf{v} - m\tau\dot{\mathbf{v}}^2. \tag{79}$$

Within brackets we have the kinetic, the potential, and the Schott terms, whereas on the rhs we have the sum of the rate of energy absorbed by the particle from the field and the Larmor term for the power radiated by the accelerated particle. We will interpret the Schott term as the part of the electromagnetic energy bounded to the charged particle and, hence, as part of the energy content of the particle.[37] Hence the expression in brackets gives the total energy $\mathscr{E}(t)$ of the particle, and Eq. (79) expresses the rate of change of this energy as the difference between the absorbed and the radiated power. With the help of Eqs. (6), (10), (30), and (79) we get for the average instantaneous energy of the particle

$$\langle\mathscr{E}(t)\rangle = \frac{\mathbf{p}_0^2}{2m} - \mathbf{F}\cdot\mathbf{x}_0 + \frac{3\sigma_p^2}{2m} - \frac{\tau}{m}\mathbf{F}\cdot\langle\mathbf{p}\rangle + \frac{3\tau}{2m}\frac{d\sigma_p^2}{dt}, \tag{80}$$

where the last term has been calculated with the help of Eqs. (3) and (8) by noticing that

$$e^2\int_0^t \langle \mathbf{E}_m(t)\cdot\mathbf{E}_m(t')\rangle\, dt'$$

$$= 3e^2\int_0^\infty \frac{S_m(\omega)}{\omega}\sin\omega t\, d\omega = \frac{3}{2}\frac{d\sigma_p^2}{dt}.$$

Alternatively, Eq. (80) can be written as

$$\langle \mathscr{E}(t)\rangle = \langle \mathscr{E}(0)\rangle + \frac{3\sigma_p^2}{2m} - \frac{3\tau}{m}\frac{d\sigma_p^2}{dt} - \frac{\tau}{m}\mathbf{F}^2 t. \quad (81)$$

Thus for $t \gg \beta\hbar$ a free particle gains a net average energy equal to $3\sigma_p^2/2m$, as was already discussed; this is just the term related to the Boyer effect.[5,7] This clearly means that on the average the free particle absorbs more energy from the field than it radiates from $t = 0$ to t. This may be easily demonstrated by the following calculation.

The average radiated energy from time $t = 0$ to t is (we take $F_j = 0$)

$$W_R(t) = \frac{\tau}{m}\int_0^t \langle \dot{\mathbf{p}}^2(T')\rangle\, dt' = \frac{\tau e^2}{m}\int_0^t \langle \mathbf{E}_m^2(t')\rangle\, dt' \quad (82)$$

and is due entirely to the stochastic acceleration to which the particle is subject. On the other side, the average energy absorbed from time $t = 0$ to t is

$$W_A(t) = \frac{e}{m}\int_0^t \langle \mathbf{E}(t')\cdot\mathbf{p}(t')\rangle\, dt'$$

$$= \frac{e^2}{m}\int_0^t dt_1\int_0^{t_1} \langle \mathbf{E}(t_1)\cdot\mathbf{E}_m(t_2)\rangle\, dt_2.$$

Since from Eq. (7) it follows that

$$\mathbf{E}(t) = \mathbf{E}_m(t) - \tau\dot{\mathbf{E}}_m(t),$$

by integrating the term that contains $\mathbf{E}_m$ we may reexpress W_A as

$$W_A(t) = \frac{e^2}{m}\int_0^t dt'\int_0^{t'} dt''\, \langle \mathbf{E}_m(t')\cdot\mathbf{E}_m(t'')\rangle$$

$$- \frac{\tau e^2}{m}\int_0^t \langle (\mathbf{E}_m(t) - \mathbf{E}_m(t'))\cdot\mathbf{E}_m(t')\rangle\, dt'.$$

By changing variables and taking the stationary character of the field into account, we can express the double integral as follows:

$$e^2\int_0^t dt'\int_0^{t'} dt''\, \langle \mathbf{E}_m(t')\cdot\mathbf{E}_m(t'')\rangle$$

$$= e^2\int_0^t dt_1\int_{t_1}^t dt_2\, \langle \mathbf{E}_m(t_1)\cdot\mathbf{E}_m(t_2)\rangle = \tfrac{1}{2}\sigma_p^2(t).$$

Substituting this expression and its derivative in the last result, we get

$$W_A(t) = \frac{3}{2m}\sigma_p^2(t) - \frac{3\tau}{2m}\frac{d\sigma_p^2}{dt} + W_R(t), \quad (83)$$

which is just Eq. (81) (for $F_j = 0$), but seen from a different viewpoint. The fact that for sufficiently large times $\langle \mathscr{E}(t)\rangle$ attains a constant value and thus the particle radiates as much energy as it absorbs in the average does not imply, however, that the system reaches an equilibrium state. In fact, the net average power transferred from the stochastic field to the particle, $\dot{W}_A - \dot{W}_R = d\langle \mathscr{E}\rangle/dt$, becomes zero; but a frequency analysis reveals that at all times, however large, the free particle continues to pump energy from some field modes to others, interchanging them periodically, in such a form that the net power transferred from or to any mode averages out, in time, to zero. Obviously, this net average power transferred to or from any mode integrates to zero over the whole spectrum for $t\to\infty$. We will see below that this exchange of energy between field and particle in any frequency interval $d\omega$ is quite intense, even in the limit $t\to\infty$.

To see this, we follow a procedure similar to the one used above to calculate W_R and W_A; omitting irrelevant details, we get

$$\dot{W}_R(t) = (3\tau e^2/m)\int_0^\infty S_m(\omega)\, d\omega \quad (84)$$

and

$$\dot{W}_A(t) = \frac{e^2}{m}\int_0^t \langle \mathbf{E}_m(t)\cdot\mathbf{E}_m(t')\rangle\, dt'$$

$$- \frac{\tau e^2}{m}\int_0^t \frac{d}{dt}\langle \mathbf{E}_m(t)\cdot\mathbf{E}_m(t')\rangle\, dt'$$

or

$$\dot{W}_A(t) = \frac{3e^2}{m}\int_0^\infty S_m(\omega)\left[\frac{1}{\omega}\sin\omega t + \tau(1-\cos\omega t)\right] d\omega, \quad (85)$$

where we have made use of Eq. (3). Defining the power spectrum $w(\omega)$ so that

$$\dot{W}_N(t) = \int_0^\infty w_N(\omega, t)\, d\omega, \quad (86)$$

where N stands for A or R, we get from Eqs. (84) and (85)

$$w_R(\omega) = (3\tau e^2/m)S_m(\omega), \quad (87a)$$

$$w_A(\omega, t) = w_R(\omega) + (3e^2/m)S_m(\omega)[(\sin\omega t)/\omega - \tau\cos\omega t]. \quad (87b)$$

We see that in fact $w_A(\omega, t)$ does not reach any definite limit as $t\to\infty$, but oscillates around the time average value $w_R(\omega)$; also, for $t\to\infty$ the oscillating terms integrate to zero over the entire spectrum, thus showing that the particle reradiates all the absorbed energy, as corresponds to the stationary average energy. However, the average powers radiated and absorbed in a frequency interval $d\omega$ differ from each other by

$$(w_A - w_R)\, d\omega = (3e^2/m)[S_m(\omega)/\omega]\sin\omega t\, d\omega \quad (88)$$

(we neglect the small corrections proportional to τ since for all $\omega < \omega_C$, $\omega\tau \leqslant 4\alpha/3 \ll 1$), a quantity that oscillates both in time for fixed frequency and with frequency for fixed time. This is the ceaseless pumping effect referred to above. The energies involved in Eq. (88) are quite high; during a time interval of order τ we have $(e^2/m)\tau\int_0^{\omega_c}[S_m(\omega)/\omega]\, d\omega \sim \alpha^2 mc^2$ and for atomic times $t \sim \hbar/\alpha^2 mc^2 \sim \tau/\alpha^3$ the exchanged energy may highly exceed the rest energy of the electron.

Since the free particle is an obviously nonergodic system, no equilibrium state exists; hence it is not surprising that $w_A(\omega)$ depends on time for all times and that $w_A \neq w_R$ even for $t\to\infty$, so that no balance at every frequency ever exists. More interesting seems to be the fact that a similar (but time-independent) phenomenon of energy pumping has been discovered in SED by Boyer,[29] even for systems that classically do reach an equilibrium situation. Boyer sees in this another sign of failure if SED is to be a successful theory

for atomic systems.[29] Here it seems to be acceptable in principle, yet somewhat worrying by its magnitude.

It is clear that energy balance in SED is a delicate matter and that the understanding of this as well as of other problems as those mentioned in the Introduction demands further development and deepening of the theory.

[1]A recent review may be found in T. H. Boyer, in *Foundations of Radiation Theory and Quantum Electrodynamics*, edited by A. O. Barut (Plenum, New York, 1980).

[2]M. Planck, Verh. Dtsch. Phys. Ges. **13**, 138 (1911); Ann. Physik **37**, 642 (1912).

[3]W. Nernst, Verh. Dtsch. Phys. Ges. **18**, 83 (1916).

[4]P. Claverie and S. Diner, Int. J. Quantum Chem. **12**, Suppl. 1, 44 (1977).

[5]T. H. Boyer, Phys. Rev. **182**, 1374 (1969); **186**, 1304 (1969).

[6]J. L. Jiménez, L. de la Peña, and T. A. Brody, Am. J. Phys. **48**, 840 (1980).

[7]A. Rueda, "Behaviour of classical particles immersed in the classical electromagnetic zero-point field," Preprint AS1349, Universidad de los Andes, Bogotá, Colombia.

[8]T. W. Marshall, Nuovo Cimento **38**, 206 (1965); L. L. Henry and T. W. Marshall, Nuovo Cimento **41**, 188 (1966).

[9]T. H. Boyer, Phys. Rev. **174**, 1764 (1968); **180**, 19 (1968); A **11**, 1650 (1975) and references therein.

[10]P. Braffort and C. Tzara, C. R. Acad. Sci. (Paris) **239**, 1779 (1954).

[11]T. W. Marshall, Proc. R. Soc. London A **276**, 475 (1963); Proc. Camb. Philos. Soc. **61**, 537 (1965).

[12]E. Santos, Nuovo Cimento **19** B, 57 (1974).

[13]L. de la Peña and A. M. Cetto, Phys. Lett. A **47**, 183 (1974); Rev. Mex. Fís. **25**, 1 (1976).

[14]T. H. Boyer, Phys. Rev. D **11**, 809 (1975).

[15]L. Pesquera and E. Santos, Lett. Nuovo Cimento **20**, 308 (1977); **25**, 287 (1979).

[16]L. de la Peña and A. M. Cetto, J. Math. Phys. **20**, 469 (1979).

[17]M. Surdin, Ann. Inst. H. Poincaré **15**, 203 (1971).

[18]L. de la Peña and A. M. Cetto, J. Math. Phys. **18**, 1612 (1977); Found. Phys. **8**, 191 (1978).

[19]P. Claverie and S. Diner, Ann Fond. L. de Broglie **1**, 73 (1976).

[20]T. W. Marshall and P. Claverie, J. Math. Phys. **21**, 1819 (1980); P. Claverie, L. de la Peña, and S. Diner, "Stochastic electrodynamics of nonlinear systems," (to be published); P. Claverie, L. Pesquera, and F. Soto, Phys. Lett. A **80**, 113 (1980).

[21]P. Julg, Folia Chim. Theor. Latina **6**, 99 (1978).

[22]L. Pesquera, in *Dynamical Systems and Microphysics*, edited by A. Blaquière, F. Fer and A. Marzollo, CISM, Courses and Lectures No. 261 (Springer-Verlag, Vienna, New York, 1980).

[23]R. Blanco and E. Santos, Lett. Nuovo Cimento **25**, 360 (1979).

[24]D. J. Kaup, Phys. Rev. **152**, 1130 (1966); H. M. França, G. L. Marques, and A. J. da Silva, Nuovo Cimento A **48**, 65 (1978).

[25]L. de la Peña, J. L. Jiménez, and R. Montemayor, "The classical motion of an extended particle revisted," Preprint IFUNAM 80-20, UNAM, Mexico, 1980, Nuovo Cimento (to be published).

[26]L. de la Peña, "Stochastic Electrodynamics for particles with structure," Preprint IFUNAM 81-18, UNAM, Mexico, (to be published).

[27]J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1962).

[28]F. Bopp, Ann. Inst. H. Poincaré **15**, 81 (1956); N. D. Cartwright, Physica A **83**, 210 (1976).

[29]T. H. Boyer, Phys. Rev. D **13**, 2832 (1976); A **18**, 1228 (1978).

[30]O. Ivanenko and A. A. Sokolov, *Klassische Feldtheorie*, (Akademie-Verlag, Berlin, 1953) (translated from the Russian edition, 1949).

[31]I. S. Gradshteyn and J. N. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1980).

[32]L. Landau and E. Lipshitz, *The Classical Theory of Fields* (Addison-Wesley, Cambridge, MA, 1951).

[33]See, e.g., J. E. Moyal, Proc. Camb. Philos. Soc. **45**, 99 (1949).

[34]S. R. de Groot and L. G. Suttorp, *Foundations of Electrodynamics* (North-Holland, Amsterdam, 1972).

[35]B. Leaf, J. Math. Phys. **9**, 65 (1968).

[36]See, e.g., H. Margenau and L. Cohen, in *Quantum Theory and Reality*, edited by M. Bunge (Springer, New York, 1967); L. Cohen and Zaparovany, J. Math. Phys. **21**, 784 (1980).

[37]See, e.g., C. Teitelboim, D. Villarroel, and Ch. van Weert, Nuovo Cimento **3**(9) (1980).

Some properties of the eigenfunctions of the dilated model Hamiltonians with complex potentials

Piotr Froelich
Quantum Chemistry Institute, Uppsala University, Box 518, Uppsala, Sweden

(Received 20 July 1982; accepted for publication 18 November 1982)

A pair of operators $H(\theta)$ and $H(\theta^*)$ obtained by dilation into opposite directions of a model Hamiltonian with nonreal potential is considered. Relations between the resonant eigenfunctions of $H(\theta)$ and $H(\theta^*)$ are studied.

PACS numbers: 03.65.Ge

INTRODUCTION

The complex-coordinate method continues its inmarch into different branches of atomic and molecular physics,[1] and the need for an efficient treatment of the truly many-body system with this technique is becoming increasingly apparent. One of the difficulties is that while complex coordinates are used, the effects of poor convergence strike very quickly as the number of particles is increased. To overcome this obstacle, the recent effort has gone into various directions; different types of basis sets are tested and compared in the configuration interaction (CI)-type approaches,[2] stabilization-type implementations of the variational method appeared,[3] and approximate treatments based on the effective Hamiltonians[4,5] and propagators[6–8] are being attempted.

Although the exact physical Hamiltonians are based on real potentials, the model and/or effective operators do not have to share this property. The origin of complex potentials in model and effective operators might be quite different. We may mention the (complex) potential energy in the nuclear motion B-O Hamiltonian studied by Moiseyev,[9] the self-energy term in the Layzer operator,[10,6] or the generalized optical potential in Feshbach's formalism[11] as quite different cases. To avoid a misunderstanding, we emphasize that it is the complexity of the effective potential without any rotation present that we address here.

The purpose of this paper is to investigate certain features of the resonant solutions of the dilated Schrödinger-type operators with complex effective/optical potentials. In particular, we will be interested in the question under what conditions operators $H(\theta)$ and $H(\theta^*)$ can have conjugated resonant eigensolutions. [$H(\theta)$ denotes, as usual, the dilated operator $H(\theta) = U(\theta)HU^{-1}(\theta)$, where $U(\theta)$ is the dilation operator: $\theta = \alpha + i\beta$.] The acquired information might be of importance for the implementation of the approximate treatments.

$H(\theta)$, $H(\theta^*)$, AND THEIR RESONANT SOLUTIONS

In what follows we will demonstrate that for the Hermitian *and* real operator, say, the exact Hamiltoinan $H(0)$, the resonant eigenfunctions $\{\psi_i\}$ and $\{\varphi_i\}$ associated with $H(\theta)$ and $H(\theta^*)$, respectively, satisfy the condition $\psi_i = \varphi_i^*$, but the latter condition does not, in general, hold for nonreal operators (even if they are Hermitian).

Let's consider the operator $H = H_0 + V_{\text{eff}}$, with H being some "unperturbed" Hamiltonian containing kinetic energy T and one-particle potential energy V, and V_{eff} being the two-particle potential energy, or some approximation to it. We will also assume that H_0 itself does not show spectral concentrations, i.e., that the nonreal discrete spectrum of $H_0(\theta)$ is empty; so V_{eff} is the resonance-producing part.

If V_{eff} is real, then for $\theta \in \mathbb{R}$, the entire operator $H(\theta)$ is also real:

$$H(\theta)f = (H(\theta)f^*)^*, \quad \forall f \in D(H(\theta)), \quad \theta \in \mathbb{R}. \tag{1}$$

If V_{eff} is also dilation analytic in the strip ϑ_β around the real axis, then for all $f \in D(H(\theta)) = D(H(0))$, $H(\theta)f$ and $(H(\theta^*)f^*)^*$ are analytic in ϑ_β. Using the latter fact and the coincidence property (1), one has

$$H(\theta)f = (H(\theta^*)f^*)^*, \quad \forall f \in D(H(\theta)), \quad \theta \in \vartheta_\beta. \tag{2}$$

In particular, for any function from the null space of $(H(\theta) - \lambda)$, e.g., for the resonance eigenfunction associated with a complex eigenvalue ϵ, the property (2) gives

$$H(\theta^*)\psi^* = (H(\theta)\psi)^* = (\epsilon\psi)^* = \epsilon^*\psi^*, \quad \psi \in N(H(\theta) - \epsilon). \tag{3}$$

Thus for a real operator, in addition to

$$H(\theta)\psi = \epsilon\psi, \tag{4}$$

one also has

$$H(\theta^*)\psi^* = \epsilon^*\psi^*. \tag{5}$$

It is thus clear that, in this case, if $\{\psi_i\}$ is a basis of $N(H(\theta) - \epsilon)$, then $\{\psi_i^*\}$ is a basis of $N(H(\theta^*) - \epsilon^*)$.

However, if V_{eff} is nonreal, relation (2) cannot be obtained since (1) does not, in general, hold;

$$H(\theta)f \neq (H(\theta)f^*)^*, \quad \exists f \in D(H(\theta)), \quad \theta \in \mathbb{R}. \tag{6}$$

Hence Eqs. (4) and (5) cannot, in general, be satisfied simultaneously. Still, they might be satisfied for *some* f's from $D(H(\theta))$. We thus have to check whether the functions of interest, i.e., the resonant states associated with the complex point spectra of $H(\theta)$, are excluded from this category.

Let's assume the contrary, i.e., that there are f's such that

$$H_0(\theta)f + V_{\text{eff}}(\theta)f = \epsilon f, \quad f \in D(H(\theta)), \quad \epsilon \neq \epsilon^*, \tag{7}$$

$$H_0(\theta^*)f^* + V_{\text{eff}}(\theta^*)f^* = \epsilon^* f^*. \tag{8}$$

This is only possible when

 0022-2488/83/122762-02$02.50

$$(V_{\text{eff}}(\theta^*)f^*)^* = V_{\text{eff}}(\theta)f = 0, \tag{9}$$

and, consequently,

$$H_0(\theta)f = \epsilon f, \quad f \in D(H(\theta)), \quad \epsilon \neq \epsilon^*, \tag{10}$$

which is impossible for $f \in D(H(\theta)) = D(H_0(\theta)) \subset L^2$, since it contradicts the assumption that the nonreal discrete spectrum of $H_0(\theta)$ is empty.

Thus we have proven that the operators $H(\theta)$ and $H(\theta^*)$ can have conjugated resonant eigenfunctions associated with λ and λ^*, respectively, if and ony if the potential V_{eff} is real valued.

EXAMPLE

An illustrative case is provided by considering V_{eff} of type $V_{\text{eff}} \sim |f\rangle\langle f|$, f being dilation analytic but not real valued: $f(\theta) \neq f^*(\theta^*)$ or $f(0) \neq f^*(0)$. The dilated operator of this type becomes

$$V_{\text{eff}}(\theta) = U(\theta)|f\rangle\langle f|U^{-1}(\theta) = |f(\theta)\rangle\langle f(\theta^*)|, \tag{11}$$

which is easily seen by noting that $U^{-1}(\theta) = U(-\theta)$ and $U^\dagger(\theta) = U(-\theta^*)$. The latter relation holds since

$$U^\dagger(\theta)f = (e^{iA\theta})^\dagger f = e^{-iA\theta^*}f = U(-\theta^*)f, \tag{12}$$

where A is the infinitesimal generator of the dilation group.[12]

We see that $V_{\text{eff}}(\theta)$ of type (11) is an oblique projector,[13]

$$V_{\text{eff}}^\dagger(\theta) \neq V_{\text{eff}}(\theta), \quad \text{Im}(\theta) \neq 0, \tag{13}$$

but that it becomes symmetric for $\text{Im}(\theta) = 0$; so the basic premises of the dilation analytic theory are intact.[14] V_{eff} also satisfies the property

$$V_{\text{eff}}^\dagger(\theta) = V_{\text{eff}}(\theta^*), \quad \theta \in \vartheta_\beta, \tag{14}$$

but this is not enough for the two equations for $H(\theta)$ and $H^\dagger(\theta) = H(\theta^*)$ to be satisfied with the conjugated eigenfunctions. In fact, if

$$H(\theta)\varphi = H_0(\theta)\varphi + |f(\theta)\rangle\langle f(\theta^*)|\varphi = \epsilon\varphi \tag{15}$$

and

$$H^\dagger(\theta)\varphi^* = H_0(\theta^*)\varphi^* + |f(\theta^*)\rangle\langle f(\theta)|\varphi^* = \epsilon^*\varphi^*, \tag{16}$$

then, since $|f^*(\theta^*)\rangle \neq |f(\theta)\rangle$, one must have

$$\langle f^*(\theta)|\varphi = \langle f(\theta^*)|\varphi = 0, \tag{17}$$

which leads to the contradictory result (10).

In this simple case it is also easily seen that, in agreement with the general result of the previous section, Eqs. (15) and (16) can be satisifed if and only if $f^*(\theta^*) = f(\theta)$, i.e., if and only if the function entering V_{eff} is real valued: $f^*(0) = f(0)$.

SUMMARY

The results of the previous sections indicate that in situations where the potential V_{eff} is required (or appears) to be complex, the bivariational[15] treatments based on the stationarity of the "intermediate" expectation value

$$\epsilon = \frac{\langle\varphi|H(\theta)|\psi\rangle}{\langle\varphi|\psi\rangle} \tag{18}$$

will split into two sets of equations

$$H(\theta)\psi = \epsilon\psi, \quad H(\theta^*)\varphi = \epsilon^*\varphi \tag{19}$$

whose solutions cannot be obtained from each other by means of the conjugation operation.

ACKNOWLEDGMENTS

I am indebted to Professor E. Balslev for valuable discussions on the subject.

[1]See, e.g., the review article by W. P. Reinhardt in Ann. Rev. Phys. Chem. **33** (1982); or Int. J. Quantum Chem. Symp. **14** (1978) and the references therein.
[2]N. Moiseyev, P. R. Certain, and F. Weinhold, Phys. Rev. **24**, 1254 (1981).
[3]B. R. Junker, Phys. Rev. Lett. **44**, 1487 (1980).
[4]P. Forelich and E. Brändas, "Some Analytical Properties of the Dilated SCF Equations," to appear in *Proceedings of the Fourth International Congress in Quantum Chemistry*, Int. J. Quantum Chem. **23**, 91 (1983); M. Mishra, Y. Öhrn, and P. Froelich, Phys. Lett. **84**, 4 (1981).
[5]C. W. McCurdy, J. Lauderdale, and R. Mowrey, J. Chem. Phys. **75**, 1835 (1981).
[6]M. Mishra, P. Froelich, and Y. Öhrn, Chem. Phys. Lett. **81**, 339 (1981).
[7]P. Winkler, R. Yaris, and R. Lovett, Phys. Rev. A **23**, 1787 (1981).
[8]R. Donelly and J. Simons, J. Phys. Chem. **73**, 2858 (1980).
[9]N. Moiseyev, Mol. Phys. **42**, 129 (1981).
[10]A. J. Layzer, Phys. Rev. **129**, 897 (1963).
[11]H. Feshbach, Ann. Phys. **19**, 287 (1962).
[12]M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1972).
[13]T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1966).
[14]E. Balslev and J. M. Combes, Commun. Math. Phys. **22**, 280 (1971).
[15]P. -O. Lowdin, "Set Theory and Linear Algebra," Technical Note of the Uppsala Quantum Chemistry Group.

Evolution theorem for a class of perturbed envelope soliton solutions

E. W. Laedke and K. H. Spatschek
Fachbereich Physik, Universität Essen, D-4300 Essen, Federal Republic of Germany

L. Stenflo
Department of Plasma Physics, Umeå University, S-90187 Umeå, Sweden

(Received 8 February 1983; accepted for publication 25 May 1983)

Envelope soliton solutions of a class of generalized nonlinear Schrödinger equations are investigated. If the quasiparticle number N is conserved, the evolution of solitons in the presence of perturbations can be discussed in terms of the functional behavior of $N(\eta^2)$, where η^2 is the nonlinear frequency shift. For $\partial_{\eta^2} N > 0$, the system is stable in the sense of Liapunov, whereas, in the opposite region, instability occurs. The theorem is applied to various types of envelope solitons such as spikons, relatons, and others.

PACS numbers: 03.65.Ge, 52.35. − g

I. INTRODUCTION

In the past, many systems have been shown to possess solitary wave solutions even if an inverse scattering solution is unknown. Solitary waves are defined as quasistationary solutions of nonlinear wave equations under the restriction that the physically relevant quantities are localized. In many physical situations, the envelopes of the waves are localized, and the corresponding solitary waves are usually called envelope solitons. A typical example is the Langmuir soliton,[1] which, in its simplest form, is governed by a cubic nonlinear Schrödinger equation.[2]

A central question in soliton theory is the stability of a localized wave (packet) during collisions. If this problem cannot be solved by constructing an inverse scattering solution, the simpler question of stability against small perturbations becomes important. This paper deals with latter situation for envelope solitons. A general necessary and sufficient stability criterion is presented.

The question of stability for nonenvelope solitons governed by relativistically invariant field equations was discussed in the literature by many authors,[3–6] and a so-called Q-theorem[5] was obtained. Here we derive a corresponding N-theorem[7] for Schrödinger-type envelope solitons and prove that the theorem is necessary and sufficient for stability: If a (conserved) number N exists, then the soliton (index s) is stable if and only if the value of N_s increases with the nonlinear frequency shift η_s^2, i.e.,

$$\frac{dN_s}{d\eta_s^2} > 0. \tag{1}$$

The results can be applied to a large class of envelope solitons. In the last section we shall discuss the three examples of spikons, relatons, and envelope solitons of the derivative nonlinear Schrödinger equation.

II. THE MODEL

To be as general as possible, including all the applications[8–19] we have in mind, let us start with the Lagrangian density

$$\mathcal{L} = \tfrac{1}{2}i(\psi\,\partial_t\psi^* - \psi^*\,\partial_t\psi) - U(|\psi|^2) - \tfrac{1}{2}\beta\,[\partial_x|\psi|^2]^2 + \partial_x\psi\,\partial_x\psi^*, \tag{2}$$

where ψ is a complex envelope, the potential U is real, and β is a positive constant. The equation of motion follows by the principle of least action,

$$i\,\partial_t\psi + \partial_x^2\psi + U'\psi - \beta(\partial_x^2|\psi|^2)\psi = 0. \tag{3}$$

Here the prime denotes a derivative with respect to the argument. Note that Eq. (3) is a generalized nonlinear Schrödinger equation; special forms of (3) have been derived for various physical situations.[8–19]

In the following, we assume that localized and analytic quasistationary solutions exist. The latter we write in the form

$$\psi_s = G(x, \eta_s^2)\exp(i\eta_s^2 t), \tag{4}$$

where G follows from

$$-\eta_s^2 G + \partial_x^2 G + U'(G^2)G - \beta(\partial_x^2 G^2)G = 0. \tag{5}$$

The form of the solutions G will be discussed later [when evaluating the criterion (1) for various model equations]. Here we only note that, after multiplying by $\partial_x G$, Eq. (5) can be integrated once, i.e.,

$$(\partial_x G)^2(1 - 2\beta G^2) - \eta_s^2 G^2 + U(G^2) = 0. \tag{6}$$

Equation (6) shows that the soliton solutions follow by simple integration.

Next, we study the constants of motion for Eq. (3). Multiplying with ψ^*, subtracting the complex conjugate, and integrating over space, we find

$$\partial_t N = 0, \tag{7}$$

where the quasiparticle number

$$N = \int_{-\infty}^{+\infty} dx\,|\psi|^2 \tag{8}$$

has been introduced.

Multiplying both sides of Eq. (3) with $\partial_t\psi^*$, adding the complex conjugate (c.c), and integrating over space, we find the energy conservation

$$\partial_t E = 0, \tag{9}$$

where

$$E = \int dx\left[|\partial_x\psi|^2 - U(|\psi|^2) - \frac{\beta}{2}(\partial_x|\psi|^2)^2\right]. \tag{10}$$

 0022-2488/83/122764-06$02.50

In the following we shall only need these two conserved quantities; because of translational invariance, the momentum conservation law is not important.

III. INSTABILITY CRITERION AND GROWTH RATE

Now we perturb the soliton in the form

$$\psi = (G + a + ib)\exp(i\eta_s^2 t) \tag{11}$$

and study the evolution of the perturbations a and b. In general, the analysis should be nonlinear. Here we show only the linear part since the basic conclusions are not changed when a nonlinear calculation is performed. [A general nonlinear instability calculation will be presented elsewhere.[20]]

After some algebra, we obtain

$$\partial_t a = H_+ b \tag{12}$$

and

$$\partial_t b = -H_- a, \tag{13}$$

where the Schrödinger operators H_+ and H_- are defined as

$$H_+ = -\partial_x^2 + \eta_s^2 - U' + \beta(\partial_x^2 G^2), \tag{14}$$

$$H_- = H_+ - 2G^2 U'' + 2\beta G\,\partial_x^2 G. \tag{15}$$

It is not necessary to know the explicit form of G in order to discuss some spectral properties of H_+ and H_-. All we need are the relations

$$H_+ G = 0, \tag{16}$$

$$H_- \partial_x G = 0, \tag{17}$$

and the assumption that G is a bell-shaped soliton (without any nodes). Then H_+ is positive semidefinite, and H_- has one negative eigenvalue, since $\partial_x G$ has one node. Obviously, H_+ and H_- are symmetric with respect to the scalar product $\langle u|v\rangle \equiv \int_{-\infty}^{+\infty} dx\, uv$ [note also the definition $\langle u|H|v\rangle = \int_{-\infty}^{+\infty} dx\, uHv$]. Since $\partial_t \langle a|G\rangle = 0$ due to (12) and (16), it is sufficient to treat perturbations with $\langle a|G\rangle = 0$.

In the case of instability, a variational principle can be derived for the maximum growth rate γ. Using the results of Refs. 21 and 22, we consider functions ξ being perpendicular to the kernel of H_+ and set $D \equiv 0$, $N \equiv H_+^{-1}$, and $F \equiv -H_-$. Here and in the following the components of ξ being parallel to G (note $H_+ G = 0$) vanish, i.e., $\langle \xi|G\rangle = 0$.

Then, we have

$$\gamma^2 = \sup_{\substack{\xi \\ \langle \xi|G\rangle = 0}} \frac{-\langle \xi|H_-|\xi\rangle}{\langle \xi|H_+^{-1}|\xi\rangle}. \tag{18}$$

To derive an instability criterion, we construct a ξ, with $\langle \xi|G\rangle = 0$, such that $\langle \xi|H_-|\xi\rangle < 0$.

Let us consider

$$\xi = \langle \xi_-|G\rangle H^{-1}G - \langle G|H^{-1}|G\rangle \xi_-, \tag{19}$$

where $H^{-1}G$ is defined by

$$H^{-1}G = -\partial_{\eta_s^2} G, \tag{20}$$

and ξ_- is an arbitrary function with $\langle \xi_-|H_-|\xi_-\rangle < 0$.

The latter always exists since H_- has a negative eigenvalue.

Using Eq. (19) we find

$$\langle \xi|H_-|\xi\rangle = -\langle G|H^{-1}|G\rangle(\langle \xi_-|G\rangle^2 - \langle G|H^{-1}|G\rangle\langle \xi_-|H_-|\xi_-\rangle|). \tag{21}$$

Thus instability can occur provided

$$\langle G|H^{-1}|G\rangle > 0, \tag{22}$$

and from Eq. (20) it is obvious that this means

$$\partial_{\eta_s^2} N_s < 0, \tag{23}$$

where N_s is the soliton quasiparticle number

$$N_s = \int_{-\infty}^{+\infty} dx\, G^2. \tag{24}$$

So far, condition (23) is only sufficient for instability. In the next section we shall derive a sufficient criterion for stability which covers just the opposite region.

IV. STABILITY REGION

To discuss the stability properties, we construct a Liapunov functional out of the constants of motion. We choose

$$L = \int_{-\infty}^{+\infty} dx \left[|\partial_x \psi|^2 - U(|\psi|^2) - \frac{\beta}{2}(\partial_x |\psi|^2)^2 - (\partial_x G)^2 + U(G^2) + \frac{\beta}{2}(\partial_x G^2)^2 \right] + \eta_s^2 \left(\frac{N}{N_s}\right)^{\theta} \int_{-\infty}^{+\infty} dx\, [|\psi|^2 - G^2], \tag{25}$$

where N and N_s are defined by (8) and (24). The parameter θ is any large positive number; its lower limit will be determined later.

Since L is built out of constants of motion,

$$\partial_t L = 0 \tag{26}$$

follows. Furthermore,

$$L = 0 \quad \text{for } \psi = G\exp(i\eta_s^2 t) \tag{27}$$

is trivially satisfied. All we have to show is $L > 0$ in a small neighborhood of the stationary solution; in other words, L should have a minimum at the stationary point.

Calculating the first variation, we find

$$\delta L = \int_{-\infty}^{+\infty} dx\, [-\partial_x^2 G - U'G + \beta(\partial_x^2 G)G + \eta_s^2 G]\delta\psi^* + \text{c.c.} = 0, \tag{28}$$

because of Eq. (5). Using again the notation (11), (14), and (15), the second variation reads

$$\delta^2 L = \int_{-\infty}^{+\infty} dx\, (aH_- a + bH_+ b) + 4\theta \frac{\eta_s^2}{N_s}\left[\int_{-\infty}^{+\infty} dx\, aG\right]^2. \tag{29}$$

The stability theorem of Liapunov[23] requires that L can be estimated in terms of the norm. Here, we only treat the second-order term $\delta^2 L$ explicitly; the higher-order contributions can be estimated easily by means of Sobolev inequalities.[24]

A. Definiteness of $\delta^2 L$

Since H_+ is nonnegative, we investigate $\langle G|H_-|a\rangle$. In the following we shall always assume

$$\langle G|H_-^{-1}|G\rangle < 0 \tag{30}$$

since otherwise instability was shown already [see criterion (21)].

Let us first consider perturbations with $\langle a|G\rangle = 0$. We expand all functions in terms of eigenfunctions of H_-, remembering that H_- has only one negative eigenvalue λ_- corresponding to the eigenfunction e_-. A function a has a component a_- parallel to that eigenfunction (subscript $-$) and a component $a_\perp$ perpendicular to the eigenfunction e_- (subscript $\perp$). We then have

$$\langle a|H_-|a\rangle = -|\lambda_-|\langle a_-|a_-\rangle + \langle a_\perp|H_-|a_\perp\rangle. \tag{31}$$

Furthermore, abbreviating

$$F = H_-^{-1}G, \tag{32}$$

we get

$$0 = -|\lambda_-|\langle a_-|F_-\rangle + \langle a_\perp|H_-|F_\perp\rangle. \tag{33}$$

Using Eq. (33) and Schwarz inequality, one finds

$$\langle a_\perp|H_-|a_\perp\rangle \geqslant |\lambda_-|^2\langle a_-|F_-\rangle^2/\langle F_\perp|H_-|F_\perp\rangle. \tag{34}$$

For

$$\langle G|H^{-1}|G\rangle = \langle F|H_-|F\rangle < 0 \tag{35}$$

we obtain

$$\langle |F_\perp|H_-|F_\perp|\rangle < |\lambda_-|\langle F_-|F_-\rangle. \tag{36}$$

Combining (31), (34), and (36), we get the desired result

$$\langle a|H_-|a\rangle \geqslant 0. \tag{37}$$

Next, we have to allow for perturbations with $\langle a|G\rangle \neq 0$. Then[21]

$$\inf_{\substack{a \\ \langle a|G\rangle \neq 0}} \frac{\langle a|H_-|a\rangle}{\langle a|G\rangle^2} = \langle G|H_-^{-1}|G\rangle^{-1}, \tag{38}$$

implying

$$\langle a|H_-|a\rangle + 4\theta(\eta_s^2/N_s)\langle a|G\rangle^2 \geqslant [\langle G|H_-^{-1}|G\rangle^{-1} + 4\theta\eta_s^2/N_s]\langle a|G\rangle^2. \tag{39}$$

Thus choosing

$$\theta > -\langle G|H_-^{-1}|G\rangle N_s/4\eta_s^2 \tag{40}$$

again $\delta^2 L \geqslant 0$ follows. Inequality (40) determines a proper lower limit for θ which we should choose in the definition of L, Eq. (25).

From this rough discussion we conclude that L is nonnegative in the neighborhood of the stationary soliton state. Region (30) is indeed the stability region as will be proved next.

B. Stability with respect to form

From (29), we can immediately recognize that two modes can make $\delta^2 L = 0$: (i) $b = G$ and $a = 0$ or (ii) $a = \partial_x G$ and $b = 0$. We call these the rotation and translation modes, respectively. The motivation for this nomenclature is the following. If we translate the original soliton and rotate the phase,

$$\psi_s = G(x)e^{i\eta_s^2 t} \to \psi_s(\alpha_1, \alpha_2) = G(x-\alpha_1)e^{i\eta_0^2 t + i\alpha_2}, \tag{41}$$

we get a soliton of the same form. Stability with respect to form[3] only requires that the perturbed soliton does not deviate significantly in time from a soliton of the same form compared with the original stationary state. Physically, it would be also unjustified to call a solitary wave unstable when the distance (metric) between the perturbed and unperturbed states grows in time. For example, an amplitude perturbation can lead to a different translation velocity whereas the form of the soliton is only slightly changed. An invariant set consisting of all functions which differ from the original unperturbed state by arbitrary translations and rotations in phase should be introduced. The distance between a perturbed state and the invariant set will be found by minimizing with respect to the shift parameters. This defines the so-called closest solitary wave (reference state) being identical in form to the original unperturbed state.

Since the perturbed state may show in its time evolution a translation and rotation, we write it in the form

$$\psi = \varphi e^{i\eta_s^2 t} \equiv [G(x-x_0) + a(x-x_0, t) + ib(x-x_0, t)]e^{i\gamma_0}e^{i\eta_s^2 t}. \tag{42}$$

The reference state is then determined by

$$\frac{\partial}{\partial \alpha_i}\left|\left|\varphi - G(x-\alpha_1)e^{i\alpha_2}\right|\right|^2\Big|_{\substack{\alpha_1 = x_0 \\ \alpha_2 = \gamma_0}} = 0, \quad i = 1, 2, \tag{43}$$

where we use the usual Sobolev norm[24]

$$\|A\|^2 = \int_{-\infty}^{+\infty} dx\, [|\partial_x A|^2 + \eta_s^2|A|^2]. \tag{44}$$

Evaluating (43), we find for $i = 1$,

$$\int_{-\infty}^{+\infty} dx\, a\, \partial_x \left(\frac{\partial U}{\partial G^2}G - \beta\frac{\partial^2 G}{\partial x^2}G\right) = 0, \tag{45}$$

and for $i = 2$,

$$\int_{-\infty}^{+\infty} dx\, b\left(\frac{\partial U}{\partial G^2}G - \beta\frac{\partial^2 G}{\partial x^2}G\right) = 0. \tag{46}$$

It is now obvious that the consistency relations (45) and (46) forbid the (i) rotation and (ii) translation modes (and their linear combinations).

Accepting the constraints (45) and (46), which follow from the concept of form stability, it is now possible to estimate $\delta^2 L$ in terms of the norm.

C. Estimate in terms of the norm

It is needed[23] to construct upper and lower bounds for $\delta^2 L$ in terms of the norm. Upper bounds are trivially found; we concentrate on the lower bounds.

Let us first consider $\langle b|H_+|b\rangle$. The consistency relation (46) can be written in the form

$$\eta_s^2\langle b|G\rangle = \langle b|\partial_x^2 G\rangle. \tag{47}$$

Splitting b into two parts,

$$b = b_s G + b_\perp \equiv b_\parallel + b_\perp, \tag{48}$$

and remembering that the continuum of H_+ starts at a value

larger than zero, we get

$$\langle b\,|H_+|\,b\rangle \geqslant \epsilon_1 \langle b_\perp\,|b_\perp\rangle, \tag{49}$$

where ϵ_1 is a small positive number. (Here, and in the following, ϵ_ν are positive quantities which can be chosen appropriately.) Using (47) and Schwarz inequality, one gets

$$b_s^2(\eta_s^2\langle G\,|G\rangle + \langle \partial_x G\,|\partial_x G\rangle)^2 \leqslant \langle b_\perp\,|b_\perp\rangle\langle \partial_x^2 G\,|\partial_x^2 G\rangle, \tag{50}$$

so that the parallel component $\langle b_\parallel\,|b_\parallel\rangle$ can be estimated in terms of $\langle b_\perp\,|b_\perp\rangle$. Combining (49) and (50), we find

$$\langle b\,|H_+|\,b\rangle \geqslant \epsilon_2 \langle b\,|\,b\rangle. \tag{51}$$

To estimate $\langle b\,|H_+|\,b\rangle$ in terms of the norm (44), we need another inequality where $\langle \partial_x b\,|\partial_x b\rangle$ appears. The latter follows for large numbers n [so that $(n-1)\epsilon_2$ is larger than the absolute value of the minimum of $-U' + \beta\partial_x^2 G$] when the first term of the right-hand side of

$$n\langle b\,|H_+|\,b\rangle \geqslant \langle b\,|H_+ b\rangle + (n-1)\epsilon_2\langle b\,|\,b\rangle \tag{52}$$

is written out explicitly. Therefore, the estimate

$$\langle b\,|H_+|\,b\rangle \geqslant \epsilon_3 \|b\|^2 \tag{53}$$

is sufficiently established.

Similar estimates have to be performed for $\langle a|H_-|a\rangle + 4\theta\eta_s^2\langle a|G\rangle^2 N_s^{-1}$. Treating first the case $\langle a|G\rangle = 0$, we can recalculate the steps (31)–(37) for $\langle G\,|H_-^{-1}|G\rangle \leqslant -\epsilon_4$. Here, we split a into the parts

$$a = \tilde{a}_- e_- + \tilde{a}_0\,\partial_x G + a_\perp \equiv a_- + a_0 + a_\perp. \tag{54}$$

Then we obtain, instead of (35),

$$\langle a|H_-|a\rangle \geqslant \epsilon_5 \langle a_-|a_-\rangle. \tag{55}$$

Writing

$$\langle a|H_-|a\rangle = (1-\epsilon_6)\langle a|H_-|a\rangle + \epsilon_6\langle a|H_-|a\rangle \tag{56}$$

and inserting (55) for the first term on the right-hand side, we obtain after some straightforward algebra

$$\langle a|H_-|a\rangle \geqslant \epsilon_7[\langle a_-|a_-\rangle + \langle a_\perp|a_\perp\rangle]. \tag{57}$$

The consistency relation (45) is now needed to include the part $\langle a_\parallel\,|a_\parallel\rangle$. We have

$$\eta_s^2\langle a|\partial_x G\rangle = \langle a|\partial_x^3 G\rangle. \tag{58}$$

Decomposing $\partial_x^3 G$ into a form similar to (54), i.e.,

$$\partial_x^3 G = \gamma_- + \gamma_0 + \gamma_\perp, \tag{59}$$

we find, after some algebra (when the Schwarz inequality has also been applied),

$$\langle a_0|a_0\rangle\langle \eta_s^2\,\partial_x G - \gamma_0|\eta_s^2\,\partial_x G - \gamma_0\rangle \leqslant \langle \gamma_- + \gamma_\perp\,|\gamma_- + \gamma_\perp\rangle \times [\langle a_-|a_-\rangle + \langle a_\perp|a_\perp\rangle]. \tag{60}$$

Rearranging the constants, we can estimate $\langle a_0|a_0\rangle$ in terms of $\langle a_-|a_-\rangle + \langle a_\perp|a_\perp\rangle$. Combining with (57), we get

$$\langle a|H_-|a\rangle \geqslant \epsilon_8\langle a|a\rangle. \tag{61}$$

The arguments by which the term $\langle \partial_x a|\partial_x a\rangle$ [which supplements the right-hand side of (61) to $\|a\|^2$] can be estimated parallel those of the steps (52) and (53).

On the other hand, for $\langle a|G\rangle \neq 0$, and

$$a = \tilde{a}_\parallel G + a_p \equiv a_\parallel + a_p \tag{62}$$

we immediately find

$$\langle a|H_-|a\rangle + 4\theta\eta_s^2\langle a|G\rangle^2/N_s \geqslant \delta^2\tilde{a}_\parallel^2 - 2|\tilde{a}_\parallel|\,|\langle H_-G\,|H_-G\rangle^{1/2}\langle a_p\,|a_p\rangle^{1/2} + \epsilon_8\|a_p\|^2. \tag{63}$$

For large θ, the constant δ^2 becomes very large. Keeping part of the first term on the right-hand side of (63), combining the rest with the second term to a perfect square, and writing

$$\tilde{a}_\parallel^2 = \epsilon_9\langle a_\parallel\,|a_\parallel\rangle + (1-\epsilon_9\langle G\,|G\rangle) \times \langle \partial_x G\,|\partial_x G\rangle^{-1}\langle \partial_x a_\parallel\,|\partial_x a_\parallel\rangle, \tag{64}$$

we finally arrive at

$$\langle a|H_-|a\rangle + 4\theta\eta_s^2\langle a|G\rangle^2/N_s \geqslant \epsilon_{10}\|a\|^2. \tag{65}$$

The relations (53), (61), and (65) show that a lower bound of L in terms of the norm exists. Together with the (trivially to construct) upper bound and the fact $\partial_t L = 0$, stability in the sense of Liapunov follows for (30), i.e.,

$$\partial_{\eta_s^2} N_s > 0. \tag{66}$$

V. THE CASE $\partial_{\eta_s^2} N_s = 0$

The case $\langle G\,|H_-^{-1}|G\rangle = 0$ is not covered by the previous sections. We shall prove now that in this case unstable modes exist. The modes will grow quadratically in time during their initial evolution.

When stability with respect to form is considered, the linearized equations for the perturbations are

$$\partial_t a = H_+ b + \dot{x}_0\,\partial_x G, \tag{67}$$

$$\partial_t b = -H_- a - \dot{y}_0 G. \tag{68}$$

Note that Eqs. (67) and (68) differ from Eqs. (12) and (13) since the ansatz (42) is used.

Investigating even perturbations a and b, the consistency relation (45) demands

$$\dot{x}_0 = 0 \tag{69}$$

in Eq. (67), whereas Eqs. (46) and (68) lead to

$$\dot{y}_0 = -\left\langle H_-\left(\frac{\partial U}{\partial G^2}\,G - \beta\frac{\partial^2 G}{\partial x^2}\,G\right)\middle|\,a\right\rangle \times \left\langle G^2\middle|\frac{\partial U}{\partial G^2} - \beta\frac{\partial^2 G}{\partial x^2}\right\rangle. \tag{70}$$

Let us investigate solutions of the form

$$a = a_0 + a_1 t + a_2 t^2, \tag{71}$$

$$b = b_0 + b_1 t + b_2 t^2. \tag{72}$$

Collecting equal powers of t, we successively find

$$a_2 = H_-^{-1} G, \tag{73}$$

$$b_2 = 0, \tag{74}$$

$$a_1 = 0, \tag{75}$$

$$b_1 = 2H_+^{-1}H_-^{-1}G - 2\left\langle H_+^{-1}H_-^{-1}G\middle|\frac{\partial U}{\partial G^2}\,G - \beta\frac{\partial^2 G}{\partial x^2}\,G\right\rangle\left\langle\frac{\partial U}{\partial G^2} - \beta\frac{\partial^2 G}{\partial x^2}\middle|\,G^2\right\rangle^{-1} G, \tag{76}$$

and

$$a_0 = -2H_-^{-1}H_+^{-1}H_-^{-1}G, \tag{77}$$

$$b_0 = 0. \tag{78}$$

Note that the solvability condition requires

$$\langle G | H_-^{-1} | G \rangle = -\tfrac{1}{2}\partial_{\eta_s^2} N_s = 0. \tag{79}$$

Furthermore, it should be mentioned that other modes, being constant in time or varying linearly with time, exist so that the solutions (71)–(78) are not unique. Nevertheless, we have explicitly constructed a growing mode which proves instability in the case $\partial_{\eta_s^2} N_s = 0$. Although, within a linear instability calculation the modes seem to be only weakly growing, nonlinearly the instability can be quite fast, e.g., leading to a collapse in a finite time.

VI. SUMMARY AND APPLICATIONS

In this paper, we have derived a necessary and sufficient stability criterion for envelope soliton solutions of generalized Schrödinger equations. All previous investigations were aiming for a sufficient stability criterion. Comparing with the often called Q-theorem for relativistically invariant solitons, there is no need to restrict the perturbations in any manner. For example, it is not necessary to allow only for perturbed states which have the same quasiparticle number N as the soliton, i.e., $N = N_s$ (as it is often unnecessarily done for Q-stability).

For applications it is useful to note that usually the stability criterion can be evaluated without calculating the explicit form of $G(x, \eta_s^2)$. Let us assume that G is an even function of x, and that G is monotonically decreasing for $0 \leqslant x < \infty$. Making use of Eq. (6) and defining

$$I = -\int_{G_{\max}}^{0} \frac{G^2[1 - 2\beta G^2]^{1/2}}{[\eta^2 G^2 - U(G^2)]^{1/2}}\, dG, \tag{80}$$

the system is stable if

$$\frac{dI}{d\eta} > 0. \tag{81}$$

We shall discuss now three interesting applications.

First, for spikons[13–15] determined by

$$i\,\partial_t \psi + \partial_x^2 \psi + |\psi|^m \psi = 0, \quad m > 0, \tag{82}$$

analytic solutions exist. We have $\beta = 0$ and

$$U = (\tfrac{1}{2}m + 1)^{-1} |\psi|^{m+2}. \tag{83}$$

The value $G_{\max}$ follows from Eq. (6) for $\partial_x G|_{\max} = 0$, i.e.,

$$G_{\max} = [\eta^2(\tfrac{1}{2}m + 1)]^{1/m}. \tag{84}$$

Then the integral I (we shall use the index 1 for this case) follows from Eq. (80),

$$I_1 = -\eta^{4/m - 1} \int_{[m/2+1]^{1/m}}^{0} \frac{z\, dz}{[1 - (\tfrac{1}{2}m + 1)^{-1} z^m]^{1/2}}. \tag{85}$$

The criterion (81) can be immediately evaluated to yield stability for $0 < m < 4$ and instability for $m \geqslant 4$.

Next, we investigate a well-known model[25] for solitons formed by relativistic mass variation. The generalized Schrödinger equation is

$$i\,\partial_t \psi + \partial_x^2 \psi - [(1 + |\psi|^2)^{-1/2} - 1]\psi = 0. \tag{86}$$

Again $\beta = 0$, but

$$U = -2(1 + |\psi|^2)^{1/2} + 2 + |\psi|^2 \tag{87}$$

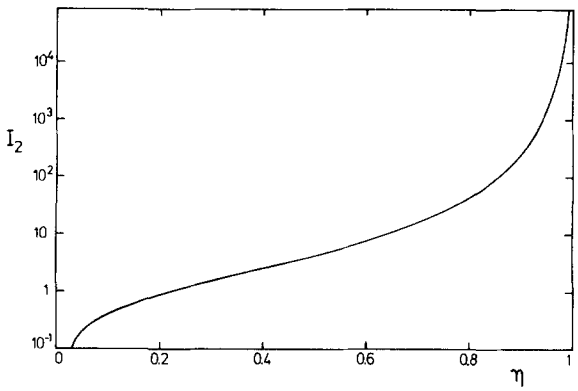


FIG. 1. Plot of I_2 [see Eq. (89)] vs η. According to criterion (81), solitons (relatons) are stable.

and

$$G_{\max} = 2\eta/(\eta^2 - 1) \tag{88}$$

with $\eta^2 < 1$. The integral (80) can be evaluated analytically,

$$I_2 = [\tfrac{1}{2}\pi + \pi\eta^2 + 3\eta\sqrt{1 - \eta^2} - (1 + 2\eta^2)\arcsin\sqrt{1 - \eta^2}\,](1 - \eta^2)^{-5/2}. \tag{89}$$

In Fig. 1, I_2 is plotted as a function of η. One clearly sees that I_2 is monotonically increasing with η leading to (longitudinal) stability for all η.

In the last application we investigate a model equation which has been derived for various modes in plasmas,[12,25]

$$i\,\partial_t \psi + \partial_x^2 \psi + |\psi|^2 \psi - \beta(\partial_x^2 |\psi|^2)\psi = 0, \tag{90}$$

for $\beta > 1$. We have

$$U = \tfrac{1}{2}|\psi|^4 \tag{91}$$

and

$$G_{\max} = \sqrt{2}\eta. \tag{92}$$

Note that analytic solutions exist for

$$4\beta\eta^2 < 1. \tag{93}$$

In that regime, the integral (80) yields

$$I_3 = \frac{1}{2\beta^{1/2}}\left[2\beta^{1/2}\eta + \frac{1}{2}(1 - 4\beta\eta^2)\ln\frac{1 + 2\beta^{1/2}\eta}{1 - 2\beta^{1/2}\eta}\right]. \tag{94}$$

The functional dependence $I_3 = I_3(\eta)$ is depicted in Fig. 2 for various parameter values β. A critical η_c exists such

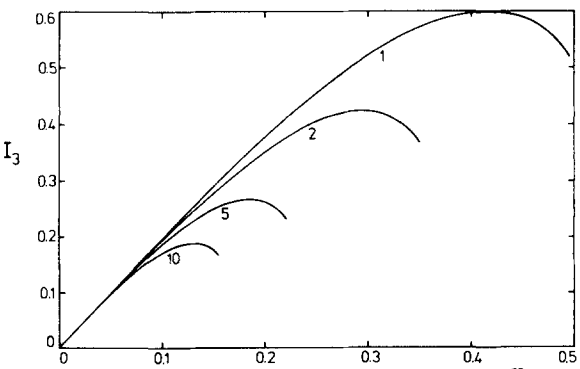


FIG. 2. Plot of I_3 [see Eq. (94)] vs η for various parameters β (shown as labels on the curves). According to criterion (81), a transition from stable to unstable behavior occurs when the sign of the first derivative changes.

that for $\eta > \eta_c$ the system is unstable. Approximately, we have

$$\sqrt{\beta}\ \eta_c \approx 0.42. \tag{95}$$

This stability region was also calculated by Litvak and Sergeev[11] using results of Kolokolov[7] for $\beta = 0$. The case $\beta \neq 0$ and, more important, the proof of instability for $\eta \geqslant \eta_c$ is treated here first.

Other examples of nonlinear envelope wave equations can be investigated in a similar way. Generalizations to three-dimensional situations are in progress.

ACKNOWLEDGMENT

Part of this work is supported by the Deutsche Forschungsgemeinschaft through SFB 162.

[1]V. N. Tsytovich, Physica C **82**, 141 (1976).
[2]V. E. Zakharov, Zh. Eksp. Teor. Fiz. **62**, 1745 (1972) [Sov. Phys. JETP **35**, 908 (1972)].
[3]T. B. Benjamin, Proc. Roy. Soc. London Ser. A **328**, 153 (1972).
[4]R. Friedberg, T. D. Lee, and A. Sirlin, Phys. Rev. D **13**, 2739 (1976).
[5]V. G. Makhankov, Phys. Rep. **35**, 1 (1978).
[6]E. W. Laedke and K. H. Spatschek, Physica D **5**, 227 (1982).
[7]A. A. Kolokolov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **17**, 1332 (1974).
[8]M. Porkolab and M. V. Goldman, Phys. Fluids **19**, 872 (1976).
[9]M. Y. Yu and P. K. Shukla, Plasma Phys. **19**, 889 (1977).
[10]P. K. Shukla, M. Y. Yu, and K. H. Spatschek, Phys. Lett. A **62**, 332 (1977).
[11]A. G. Litvak and A. M. Sergeev, Pis'ma Zh. Eksp. Teor. Fiz. **27**, 549 (1978) [JETP Lett. **27**, 517 (1978)].
[12]M. Y. Yu, P. K. Shukla, and K. H. Spatschek, Phys. Rev. A **18**, 1591 (1978).
[13]Yu. V. Katyshev, N. V. Makhalkdiani, and V. G. Makhankov, Phys. Lett. A **66**, 456 (1978).
[14]L. Stenflo and N. L. Tsintsadze, Astrophys. Space Sci. **64**, 513 (1979).
[15]E. W. Laedke and K. H. Spatschek, Phys. Lett. A **74**, 205 (1979).
[16]D. J. Kaup and A. C. Newell, J. Math. Phys. **19**, 798 (1978).
[17]V. I. Berezhiani, Sov. J. Plasma Phys. **7**, 365 (1981).
[18]M. Y. Yu, P. K. Shukla, and N. L. Tsintsadze, Phys. Fluids **25**, 1049 (1982).
[19]K. V. Kotetishvili, P. K. Kaw, and N. L. Tsintsadze, Sov. J. Plasma Phys. (1982) (to be published).
[20]E. W. Laedke and K. H. Spatschek, in *Topology—Calculus of Variations and Their Applications* (Marcel Dekker, New York, 1983).
[21]E. W. Laedke and K. H. Spatschek, J. Math. Phys. **23**, 460 (1982).
[22]E. W. Laedke and K. H. Spatschek, J. Plasma Phys. **28**, 469 (1982).
[23]V. I. Zubov, *Methods of A. M. Liapunov and Their Application* (Noordhoff, Groningen, 1964).
[24]R. A. Adams, *Sobolev Spaces* (Academic, New York, 1975).
[25]N. L. Tsintsadze, (1982) (to be published).

Extension of Fuda's off-shell analysis to screened Coulomb potentials for arbitrary l and limiting relations

Ranabir Dutt
Department of Physics, Visva-Bharati University, Santiniketan-731235, West Bengal, India
Y. P. Varshni
Department of Physics, University of Ottawa, Ottawa, Canada K1N 6N5

(Received 9 March 1983; accepted for publication 29 June 1983)

The Ecker–Weizel approximation technique is applied to the Schrödinger equation for a class of screened Coulomb potentials (Yukawa, Exponential cosine screened Coulomb and Hulthén) for any arbitrary angular momentum l. We find that the centrifugal term can be combined with the central screening potential to generate an effective Eckart potential with energy dependent strength parameters for which the s-wave Schrödinger equation is exactly solvable. Using this effective s-wave potential in the formalism of Fuda and Whiting for off-shell analysis, we obtain a closed expression for the off-shell Jost solution $f_{S,l}(k,q,r)$ in which k is the on-shell momentum, q is the off-shell momentum and the subscript S means screening. It turns out that for nonzero angular momentum, usual Jost function $f_{S,l}(k,q)$ can not be defined for finite screening parameter λ. However, we find that the Jost solution, as well as the Jost function defined in the limit $\lambda \to 0$, show discontinuities at the on-shell point $q = k$, similar to the observation made by van Haeringen [Phys. Rev. A **18**, 56 (1978)] for the s-wave Hulthén potential. For the $l = 0$ case, we obtain explicit expressions for the off-shell and on-shell Jost solutions and Jost functions which possess the limiting behaviors discussed by van Haeringen for the Hulthén potential only. Our results not only extend previous works to higher partial waves, but at the same time indicate that certain limiting properties of the Jost solutions and the Jost functions are generally true for a class of screened Coulomb potentials.

PACS numbers: 03.65.Nk

I. INTRODUCTION

During the last two decades, extensive investigations[1–11] have been carried out on the analytic properties of the Jost solution and the Jost function which are important ingredients in the theory of nonrelativistic two-body scattering by spherically symmetric potentials. In a many-particle system, the various pairs of particles do not scatter elastically from each other, and therefore one needs off-shell quantities, in particular, the off-shell Jost solution and the Jost function from which one constructs T- and K-matrices. Following the approach of van Leeuwen–Reiner to the T-matrix,[12] Fuda and Whiting[6] have proposed a method in which closed analytic expressions for both off-shell Jost solution and Jost function are obtainable. In fact, several authors[6–11] have obtained analytic formulas for these quantities for a number of potentials, mostly for the s-wave case. Working in this line, van Haeringen[9] has shown that for the s-wave Hulthén potential, the off-shell and the on-shell Jost solutions and the Jost functions possess certain interesting limiting behaviors. One of the interesting aspects of his work is that although the off-shell Hulthén–Jost solution and the Jost function smoothly go over to the corresponding Coulomb quantities in the limit of a vanishing screening parameter, such limits do not exist at the on-shell point $q = k$ due to the discontinuity arising out of the long-range nature of the Coulomb potential.

We conjecture that the limiting properties of the Jost solution and the Jost function shown by van Haeringen may not be a special feature of the Hulthén potential only, but these may be inherent properties of a class of screened Coulomb potentials. Furthermore, it was worthwhile to study whether such analytic properties of the Jost solution and the Jost function are retained by higher angular momentum states. With this motivation, we have reexamined the work of van Haeringen for arbitrary l for three potentials: static screened Coulomb or Yukawa (SSCP),[13] exponential cosine screened Coulomb (ECSC),[14] and the Hulthén potential.[15] These potentials have wide applications in atomic scattering and nuclear and solid state physics. In Sec. II, we demonstrate that using the Ecker–Weizel approximation (EWA) procedure,[16] the centrifugal term of the radial Schrödinger equation can be combined with the central screening potential to generate an effective Eckart potential[17] with energy dependent strength parameters. Since it is well known that the Schrödinger equation for the Eckart potential is solvable for the s-wave, we simulate the l-dependence of the original problem through the s-wave solution of the reduced Eckart potential. This procedure is found to work well at least for the bound states.[18]

In Sec. III, we use our reduced Eckart potential in the framework of Fuda and Whiting[6] for obtaining the off-shell Jost solution which satisfies an inhomogeneous differential equation. We find that although a compact expression for the off-shell Jost solution $f_{S,l}(k,q,r)$ is obtainable, it is difficult to define the off-shell Jost function $f_{S,l}(k,q)$ in the conventional way[1,9] for the finite screening coefficient λ. However, certain limiting conditions can be achieved in the limit

of a vanishing screening parameter. It is found that the Jost solution and the Jost function in the limit $\lambda \to 0$ exhibit discontinuity at the point where the off-shell and the on-shell momentums are equal. Similar discontinuity was noted by van Haeringen[9] for the s-wave Hulthén potential problem.

In Sec. IV, we study the limiting properties of both off-shell and on-shell Jost solutions and Jost functions for the s-wave. Since our work includes the Hulthén problem which has been thoroughly investigated by van Haeringen,[9] our observations are relevant for the other two potentials, i.e., SSCP and ECSC. It is very interesting to notice that the limiting behaviors of the Jost solutions and the Jost functions are independent of the form of the screening potential or of the approximation scheme. This clearly indicates a generality in the analytic properties of the scattering quantities for a class of potentials having the same singularity structure at the origin. In Sec. V, we make a few concluding remarks.

II. REDUCED ECKART POTENTIAL FROM EWA

For our discussion, we consider the following screened Coulomb potentials:

$$V_S(r) = \begin{cases} V_0 e^{-\lambda r}/r & \text{SSCP}, \\ V_0 e^{-\lambda r}\cos(\lambda r)/r & \text{ECSC}, \\ V_0 \lambda e^{-\lambda r}/(1-e^{-\lambda r}) & \text{Hulthén}, \end{cases} \tag{2.1}$$

where λ is the screening parameter.[19] In the limit $\lambda \to 0$, these potentials smoothly go over to the Coulomb form

$$V_C(r) = V_0/r = 2\gamma k/r, \tag{2.2}$$

in which γ is the Sommerfeld parameter.[20] In order to demonstrate the applicability of the EWA procedure, we consider a specific case, say, the SSCP. The radial Schrödinger equation for this potential is

$$\frac{d^2\chi_l(r)}{dr^2} + \left[E - V_0 e^{-\lambda r}/r - \frac{l(l+1)}{r^2}\right]\chi_l(r) = 0, \tag{2.3}$$

where we have used the units such that $\hbar = 2m = 1$.

Following the standard substitution

$$\chi_l(r) = \exp(-\alpha r) v_l(r) \tag{2.4}$$

with

$$\alpha = \sqrt{-E},$$

and using the transformation of the variable

$$x = e^{-\lambda r}, \tag{2.5}$$

Eq. (2.3) becomes

$$x(1-x)v_l'' + (1 + 2\alpha/\lambda)(1-x)v_l' + [(V_0/\lambda)f(x) - (l(l+1)/(1-x))g(x)]v_l(x) = 0, \tag{2.6}$$

in which

$$f(x) = (1-x)/\log x, \tag{2.7}$$

and

$$g(x) = (1-x)^2/x \log^2 x. \tag{2.8}$$

For screened potentials with a definite value of the screening parameter λ, it is reasonable to assume that the effective maximum range of the radial coordinate $r_{\max}$ is of the order $\sim 1/\lambda$. Consequently, from (2.5), we get $x_{\min} \approx e^{-1}$. According to the prescription of EWA,[16] the function $f(x)$ which varies slowly within the range $e^{-1} \leqslant x \leqslant 1$ (corresponding to $0 \leqslant r \leqslant \lambda^{-1}$) may be assumed to be a constant $\gamma = -(1 - e^{-\lambda \bar{r}})/(\lambda \bar{r})$, where $\bar{r}$ is some mean radial distance in the appropriate quantum state. The quantity $g(x)$ is equal to $((1 - e^{-\lambda \bar{r}})/\lambda \bar{r})^2 e^{\lambda \bar{r}}$ and this too is a slowly varying function, and thus $g(x)$ may also be considered to be a constant to the first approximation.

The advantage of considering both $f(x)$ and $g(x)$ to be approximately constants is that Eq. (2.6) can be recast into the standard hypergeometric equation. This approximation amounts to the fact that one essentially works with an effective potential which is in the Eckart form.[17] For the potentials in (2.1), we thus obtain the effective potential

$$V_E(r) = V_1 \frac{e^{-\lambda r}}{1 - e^{-\lambda r}} + V_2 \frac{e^{-\lambda r}}{(1 - e^{-\lambda r})^2}, \tag{2.9}$$

with

$$V_1 = \begin{cases} \lambda V_0 (1 - e^{-\lambda \bar{r}})/(\lambda \bar{r}) & \text{SSCP}, \quad (2.10a) \\ \lambda V_0 [(1 - e^{-\lambda \bar{r}})/\lambda \bar{r}] \cos(\lambda \bar{r}) & \text{ECSC}, \quad (2.10b) \\ \lambda V_0 & \text{Hulthén}, \quad (2.10c) \end{cases}$$

and

$$V_2 = \lambda^2 l(l+1)((1 - e^{-\lambda \bar{r}})/\lambda \bar{r})^2 e^{\lambda \bar{r}}. \tag{2.10d}$$

For the bound state problem (with $V_0 = -1$), the Schrödinger equation for the potential (2.9) has been solved to obtain the eigenenergies (in atomic units)

$$E_{nl} = \begin{cases} -\frac{1}{2}\left[\frac{1}{(n-l+\beta)}\left(\frac{1-e^{-\lambda \bar{r}}}{\lambda \bar{r}}\right) - \frac{(n-l+\beta)\lambda}{2}\right]^2 & \text{SSCP}, \quad (2.11a) \\ -\frac{1}{2}\left[\frac{1}{(n-l+\beta)}\left(\frac{1-e^{-\lambda \bar{r}}}{\lambda \bar{r}}\right)\cos(\lambda \bar{r}) - \frac{(n-l+\beta)\lambda}{2}\right]^2 & \text{ECSC}, \quad (2.11b) \\ -\frac{1}{2}\left[\frac{1}{(n-l+\beta)} - \frac{(n-l+\beta)\lambda}{2}\right]^2 & \text{Hulthén}, \quad (2.11c) \end{cases}$$

with

$$\beta = -\tfrac{1}{2} + [l(l+1)((1 - e^{-\lambda \bar{r}})/\lambda \bar{r})^2 e^{\lambda \bar{r}} + \tfrac{1}{4}]^{1/2} \tag{2.12}$$

and the unnormalized eigenfunctions

$$\chi_{nl}(r) = e^{-\alpha r}(1-e^{-\lambda r})^{\beta+1} \times {}_2F_1(-n+l+1,\ n-l+1+2\alpha/\lambda+2\beta;\ 1+2\alpha/\lambda,\ e^{-\lambda r}). \tag{2.13}$$

It turns out from Eq. (2.12) that $\beta > l$ for any finite value of λ, and β equals l only when λ vanishes. In this limiting case, Eqs. (2.11) give the Coulomb result $E_n = -1/2n^2$.

From Eq. (2.13), one restores the boundary condition

$$\chi_{nl}(r) \underset{r\to 0}{\sim} r^{(l+1)+O(\lambda)}, \quad \chi_{nl}(\infty) = 0. \tag{2.14}$$

It is important to point out that $\chi_{nl}(r)$ provides for a slightly greater repulsion near the origin than is allowed in an exact treatment of the problem. This is the price we pay for obtaining a simple analytic solution. One of the weaknesses of the EWA procedure is that $\bar{r}$ cannot be determined from first principles. However, it has been observed[18,21] that the choice, $\bar{r} = 1/\langle 1/r\rangle = n^2 + O(\lambda)$ gives very good numerical agreement between the predicted energy values and the exact ones obtained from the numerical integration of the Schrödinger equation for various values of n, l, and λ. This gives us confidence that the replacement of the $[V_S(r) + l(l+1)/r^2]$ term in the Schrödinger equation by an effective Eckart potential (2.9) with coefficients depending on the energy-dependent parameter $\bar{r}$, can approximately simulate the effect of nonzero l.

III. OFF-SHELL JOST SOLUTION: LIMITING BEHAVIOR

For obtaining the off-shell Jost solutions for the screened Coulomb potentials in (2.1) for nonzero l, we find it enough to consider the radial s-wave van Leeuwen–Reiner equation[6,12] with our effective Eckart potential (2.9):

$$\left[\frac{d^2}{dr^2} + k^2 - V_E(r)\right] f_{S,l}(k,q,r) = (k^2 - q^2)e^{iqr}. \tag{3.1}$$

Here, $\mathbf{k}$ denotes the on-shell momentum related to the energy $E = k^2 + i\epsilon$, $\epsilon \ll 1$, and $\mathbf{q}$, an off-shell momentum. The off-shell Jost solution $f_{S,l}(k,q,r)$ is a solution of Eq. (3.1) with the prescribed asymptotic behavior:

$$\lim_{r\to\infty} e^{-iqr} f_{S,l}(k,q,r) = 1. \tag{3.2}$$

The on-shell Jost solution is related to the off-shell one by

$$f_{S,l}(\pm k,r) = \lim_{q\to k} f_{S,l}(k, \pm q, r). \tag{3.3}$$

Making the following substitutions stepwise,

$$f_{S,l}(k,q,r) = e^{ikr} v_l(r), \tag{3.4a}$$

$$e^{-\lambda r} = x, \tag{3.4b}$$

and

$$v_l(x) = x^{\mu}(1-x)^{\nu} U_l(x). \tag{3.4c}$$

Equation (3.1) becomes

$$\begin{aligned} x(1-x)U_l'' + U_l'\left[\left(2\mu+1-\frac{2ik}{\lambda}\right) - x\left(2\mu+2\nu+1-\frac{2ik}{\lambda}\right)\right] \\ + U_l\left[\mu(\mu-1)\frac{1}{x} - \mu(\mu-1) - 2\mu\nu - \nu(\nu-1) + \frac{\nu(\nu-1)}{1-x}\right. \\ \left. + \left(1-\frac{2ik}{\lambda}\right)\frac{\mu}{x} - \mu\left(1-\frac{2ik}{\lambda}\right) - \nu\left(1-\frac{2ik}{\lambda}\right) - \frac{V_1}{\lambda^2} - \frac{V_2}{\lambda^2}\frac{1}{1-x}\right] \\ = \frac{(k^2-q^2)}{\lambda^2}\left[x^{i(k-q)/\lambda - 1} - x^{i(k-q)/\lambda}\right]. \end{aligned} \tag{3.5}$$

To identify Eq. (3.5) with the inhomogeneous hypergeometric equation,[22] we require that the last bracket on the left-hand side be independent of x. We thus get

$$\mu = 0 \tag{3.6a}$$

and

$$\mu = \tfrac{1}{2}[1 - (1 + 4V_2/\lambda^2)^{1/2}]. \tag{3.6b}$$

Using relations (3.6) in Eq. (3.5), one obtains

$$x(1-x)U_l'' + U_l'[C - (A+B+1)x] - ABU_l(x) = ((k^2-q^2)/\lambda^2)[x^{\sigma-1} - x^{\sigma}], \tag{3.7}$$

where

$$A = \nu - ik/\lambda + i(k^2 + V_1)^{1/2}/\lambda,$$
$$B = \nu - ik/\lambda - i(k^2 + V_1)^{1/2}/\lambda, \tag{3.8}$$
$$C = 1 - 2ik/\lambda,$$

and

$$\sigma = i(k - q)/\lambda.$$

A particular solution of Eq. (3.7) is given by[22]

$$U_l(x) = ((k^2 - q^2)/\lambda^2)[f_\sigma(A,B;C;x) - f_{\sigma+1}(A,B;C;x)], \tag{3.9}$$

in which the function $f_\sigma(A, B; C; x)$ is related to a generalized hypergeometric function by

$$f_\sigma(A,B;C;x) = \frac{x^\sigma}{\sigma(\sigma + C - 1)}\,{}_3F_2(1,A + \sigma,B + \sigma;1 + \sigma,C + \sigma;x). \tag{3.10}$$

This series converges when $|x| < 1$; it converges when $|x| = 1$ provided that $\mathrm{Re}(C - A - B) > 0$. In our case, $|x| = |e^{-\lambda r}| < 1$ for any finite λ. For a vanishing screening parameter, $|x| = 1$, and in that case,

$$\mathrm{Re}(C - A - B) = \mathrm{Re}(1 - 2\nu) = (1 + 4V_2/\lambda^2)^{1/2} > 0.$$

Using the recurrence relation[22]

$$(A + \sigma)(B + \sigma)f_{\sigma+1}(A,B;C;z) = \sigma(C + \sigma - 1)f_\sigma(A,B;C,z) - z^\sigma, \tag{3.11}$$

we finally obtain, from Eqs. (3.4a), (3.4c), (3.6), and (3.9)–(3.11), a closed expression for the off-shell Jost solution for arbitrary l:

$$f_{S,l}(k,q,r) = (1 - e^{-\lambda r})^\nu e^{iqr}\left[1 + e^{-\lambda r}\left\{(A + \sigma)(B + \sigma) - \left(\frac{k^2 - q^2}{\lambda^2}\right)\right\}\right.$$
$$\left.\times\frac{1}{(1 + \sigma)(C + \sigma)}\,{}_3F_2(1,1 + A + \sigma,1 + B + \sigma;2 + \sigma,1 + C + \sigma;e^{-\lambda r})\right]. \tag{3.12}$$

It is easy to see that our solution (3.12) gives the correct asymptotic behavior in Eq. (3.2). For the on-shell case, $k = q$, i.e., $\sigma = 0$. Thus the on-shell Jost solution for the screened Coulomb potentials for any l is given by

$$f_{S,l}(k,r) = (1 - e^{-\lambda r})^\nu e^{ikr}\left[1 + e^{-\lambda r}\frac{AB}{(1 + \sigma)(C + \sigma)}\,{}_3F_2(1,1 + A,1 + B;2,1 + C;e^{-\lambda r})\right]. \tag{3.13}$$

Although we have been able to obtain analytic expressions for the off-shell and on-shell Jost solutions for all l, it is difficult to define the corresponding Jost functions for the $l > 0$ situation because of the factor $(1 - e^{-\lambda r})^\nu$.

We shall now obtain the limiting relation of $f_{S,l}(k,r)$ for a vanishing screening coefficient, analogous to Eq. (34) of Ref. 9. As $\lambda \to 0$, one finds, from Eqs. (3.6b) and (3.8),

$$\nu \sim -l,$$
$$A \sim -l + iV_0/2k = -l + i\gamma,$$
$$B \sim -l - 2ik/\lambda - i\gamma, \tag{3.14}$$
$$C \simeq 1 - 2ik/\lambda.$$

Applying the connecting formula[5]

$${}_3F_2(1,a,b;2,c;z) = \frac{(c - 1)}{z(a - 1)(b - 1)}$$
$$\times[{}_2F_1(a - 1,b - 1;c - 1;z) - 1], \tag{3.15}$$

and using (3.14), we obtain, from Eq. (3.13) after some simplifications in the limit of the vanishingly small screening parameter λ,

$$\lim_{\lambda\to 0} f_{S,l}(k,r)$$
$$\sim e^{ikr}(\lambda r)^{-l}\,{}_2F_1(-l + i\gamma, -l - 2ik/\lambda - i\gamma;$$
$$1 - 2ik/\lambda;e^{-\lambda r}). \tag{3.16}$$

Further, using the relation[9]

$$\lim_{c\to\infty} c^{-\delta}\,{}_2F_1(\delta,b + c;c;1 - z/c) = U(\delta,\delta + b + 1,z), \tag{3.17}$$

where U is an irregular solution of the confluent hypergeometric equation, we derive, from small λ,

$$(-2ik/\lambda)^{l - i\gamma}\,{}_2F_1(-l + i\gamma, -l - 2ik/\lambda - i\gamma;$$
$$1 - 2ik/\lambda;e^{-\lambda r})$$
$$\simeq U(-l + i\gamma, -2l,(-2ikr)). \tag{3.18}$$

Thus from Eqs. (3.16) and (3.18), we get

$$(-2ik/\lambda)^{l - i\gamma} f_{S,l}(k,r)$$
$$\simeq (\lambda r)^{-l} e^{ikr} U(-l + i\gamma, -2l, -2ikr). \tag{3.19}$$

It is known that[23]

$$U(\beta,c,z) = z^{1-c}U(\beta + 1 - c,2 - c,z) \tag{3.20}$$

and so we obtain, from Eqs. (3.19) and (3.20) for small λ,

$$\begin{aligned}(-2ik/\lambda)^{l-i\gamma}f_{S,l}(k,r) &\sim (\lambda r)^{-l}e^{ikr}(-2ikr)^{2l+1}\\ &\times U(l+1+i\gamma,2l+2,-2ikr)\\ &= (-2ik/\lambda)^{l}e^{-\pi\gamma/2}[e^{ikr+\pi\gamma/2}(-2ikr)^{l+1}\\ &\times U(l+1+i\gamma,2l+2,-2ikr)].\end{aligned} \tag{3.21}$$

The bracketed term on the right-hand side of Eq. (3.21) is the Coulomb on-shell Jost solution[9]

$$\begin{aligned}f_{C,l}(k,r) &= e^{ikr+\pi\gamma/2}\\ &\times(-2ikr)^{l+1}U(l+1+i\gamma,2l+2,-2ikr).\end{aligned} \tag{3.22}$$

It is now easy to check that Eqs. (3.21) and (3.22) give

$$\lim_{\lambda\to 0}(2k/\lambda)^{-i\gamma}f_{S,l}(k,r) = f_{C,l}(k,r), \tag{3.23}$$

which indicates that the on-shell Jost solution for the screened Coulomb potentials has no limit for $\lambda \to 0$ even when $l \neq 0$. Our Eq. (3.23) is the generalized version of the van Haeringen result [Eq. (34) of Ref. 9] for any screened Coulomb potential for arbitrary l.

From Eq. (3.23), it is also clear that we can define the l nonzero Jost function only under the limiting situation $\lambda \to 0$. We find

$$\begin{aligned}&\lim_{\lambda\to 0}\lim_{r\to 0}(2k/\lambda)^{-i\gamma}f_{S,l}(k,r)\\ &= \lim_{\lambda\to 0}(2k/\lambda)^{-i\gamma}f_{S,l}(k) = f_{C,l}(k),\end{aligned} \tag{3.24}$$

where $f_{C,l}(k)$ is the on-shell Coulomb Jost function. A similar relation was obtained by van Haeringen [Eq. (40h) of Ref. 9] for the s-wave case, and that is only for the Hulthén potential.

IV. s-WAVE JOST SOLUTION, JOST FUNCTION AND LIMITING RELATIONS

Here, we derive explicit expressions for the s-wave off-shell and on-shell Jost solutions and Jost functions for three different screened Coulomb potentials given in (2.1). Since van Haeringen discussed the Hulthén problem elaborately, our observations are mainly concerned with SSCP and ECSC potentials. For the s-wave, setting $l = 0$ we have

$$\begin{aligned}&\nu = 0,\\ &A = -ik/\lambda + i(k^2+V_1)^{1/2}/\lambda,\\ &B = -ik/\lambda - i(k^2+V_1)^{1/2}/\lambda,\\ &C = 1 - 2ik/\lambda.\end{aligned} \tag{4.1}$$

From Eq. (3.12), we obtain the off-shell Jost solution (suppressing the suffix for l)

$$\begin{aligned}&f_S(k,q,r)\\ &= e^{iqr}\left[1 + e^{-\lambda r}\frac{AB}{(1+\sigma)(C+\sigma)}\right.\\ &\left.\times {}_3F_2(1,1+A+\sigma,1+B+\sigma;2+\sigma,1+C+\sigma,e^{-\lambda r})\right]\end{aligned} \tag{4.2}$$

and the off-shell Jost function

$$f_S(k,q) = \lim_{r\to 0} f_S(k,q,r) = \frac{\Gamma(1+\sigma)\Gamma(C+\sigma)}{\Gamma(1+A+\sigma)\Gamma(1+B+\sigma)}. \tag{4.3}$$

In arriving at Eq. (4.2) from Eq. (4.1), we have applied the technique of Baheti and Fuda.[5] For the on-shell case, one has $\sigma = 0$. Thus from Eqs. (4.2) and (4.3), we obtain the on-shell Jost solution and the Jost function

$$\begin{aligned}f_S(k,r) &= e^{ikr}[1 + e^{-\lambda r}(AB/C)\\ &\times {}_3F_2(1,1+A,1+B;2,1+C;e^{-\lambda r})],\end{aligned} \tag{4.4}$$

$$f_S(k) = \frac{\Gamma(C)}{\Gamma(1+A)\Gamma(1+B)}. \tag{4.5}$$

Our expressions (4.2)–(4.5) are identical to the corresponding analytic formulas derived by van Haeringen[9] for the Hulthén potential, except the fact that our A and B depend on V_1 which is different for different screened Coulomb potentials. So for the finite screening parameter λ, one essentially gets different results for different potentials.

We have already shown in Eqs. (3.23) and (3.24) that the on-shell Jost solution and the Jost function show discontinuities in the limit $\lambda \to 0$. The same results are therefore true for Eqs. (4.2) and (4.5). Since there is no well-defined off-shell Jost function for $l > 0$, we could not study its limiting behavior. However for the s-wave, we have obtained the off-shell Jost function for finite λ for various screened Coulomb potentials and it is worthwhile to study its limiting property. From Eq. (4.3), we find

$$\begin{aligned}&\lim_{\lambda\to 0} f_S(k,q)\\ &= \frac{\Gamma[1+i(k-q)/\lambda]\Gamma[1-i(k+q)/\lambda]}{\Gamma(1+iV_0/2k+i(k-q)/\lambda)\Gamma(1-iV_0/2k-i(k+q)/\lambda)}.\end{aligned} \tag{4.6}$$

Using the property

$$\begin{aligned}&\frac{\Gamma(z+\alpha)}{\Gamma(z+\beta)} = z^{\alpha-\beta}[1+O(z^{-1})] \quad \text{as} \quad z\to\infty,\\ &0 < \arg z < \pi,\end{aligned} \tag{4.7}$$

we get

$$\lim_{\lambda\to 0} f_S(k,q) = \left(\frac{q+k}{q-k}\right)^{iV_1/2k}. \tag{4.8}$$

Since, from Eq. (2,2), $V_0 = 2\gamma k$, we finally obtain

$$\lim_{\lambda\to 0} f_S(k,q) = \left(\frac{q+k}{q-k}\right)^{i\gamma} = f_C(k,q), \tag{4.9}$$

where $f_C(k,q)$ is the s-wave Coulomb off-shell Jost function.[9] An identical expression was obtained by van Haeringen for the Hulthén potential.

For the Hulthén off-shell Jost solution, van Haeringen conjectured that it should smoothly go to the Coulomb off-shell Jost solution [see Eq. (40k) of Ref. 9] when the screening is switched off. We believe that this happens to other screened Coulomb potentials also, and hence we expect

$$\lim_{\lambda\to 0} f_S(k,q,r) = f_C(k,q,r), \quad q \neq k. \tag{4.10}$$

Finally, we find that Eq. (4.2) admits the following interchangeability of the limits:

$$\lim_{r\to 0}\lim_{\lambda\to 0} f_S(k,q,r) = \lim_{\lambda\to 0}\lim_{r\to 0} f_S(k,q,r) = f_C(k,q)$$

and

$$\lim_{r\to 0}\lim_{q\to k} f_S(k,q,r) = \lim_{q\to k}\lim_{r\to 0} f_S(k,q,r) = f_S(k),$$

which were shown to be true only in the case of the Hulthén potential.

V. CONCLUSIONS

We have made an attempt to generalize the work of van Haeringen[9] on the limiting behaviors of the s-wave, off-shell, and on-shell Jost solutions and Jost functions for a class of screened Coulomb potentials. Extension to higher partial waves has been achieved through the use of Ecker–Weizel approximation procedure by means of which we construct an effective s-wave Eckart potential for which the problem is analytically solvable. Our analytic formula for the off-shell Jost solution for any l has the same limiting behavior shown earlier by van Haeringen for the s-wave Hulthén problem. For nonzero l, the second term of the Eckart potential in Eq. (2.9) remains, which means that there is a singularity $\sim 1/r^2$ at the origin, and hence the usual definition of the Jost function is not applicable. However, for the s-wave, we obtain closed analytic expressions for the off-shell and on-shell Jost solutions as well as Jost functions for all the screened potentials considered in this paper. We have been able to demonstrate that certain limiting behaviors of these scattering quantities are inherent properties of a variety of screened Coulomb potentials. Finally we would like to mention that there is no physical ambiguity in pursuing Ecker–Weizel approximation to the off-energy-shell region.

ACKNOWLEDGMENT

This work was supported in part by a research grant from the Natural Sciences and Engineering Research Council of Canada to one of the authors (Y.P.V.).

[1]R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).
[2]M. L. Goldberger and K. M. Watson, *Collision Theory* (Wiley, New York, 1964); V. De Alfaro and T. Regge, *Potential Scattering* (North-Holland, Amsterdam, 1965).
[3]M. G. Fuda, Phys. Rev. C **1**, 1910 (1970).
[4]M. G. Fuda, J. Math. Phys. **12**, 1163 (1971).
[5]O. P. Baheti and M. G. Fuda, J. Math. Phys. **12**, 2076 (1971).
[6]M. G. Fuda and J. S. Whiting, Phys. Rev. C **8**, 1255 (1973).
[7]M. G. Fuda and B. A. Girard, Phys. Rev. C **16**, 2445 (1977).
[8]H. van Haeringen and R. van Wageningen, J. Math. Phys. **16**, 1441 (1975); H. van Haeringen, Nuovo Cimento B **34**, 53 (1976); H. van Haeringen, J. Math. Phys. **17**, 995 (1976).
[9]H. van Haeringen, Phys. Rev. A **18**, 56 (1978).
[10]B. Talukdar, M. N. Sinha Roy, N. Mallick, and D. K. Nayek, Phys. Rev. C **12**, 370 (1975); M. N. Sinha Roy, B. Talukdar, and D. Chattarji, J. Math. Phys. **17**, 1763 (1976); B. Talukdar, R. N. Chaudhuri, U. Das, and P. Banerjee, J. Math. Phys. **19**, 1654 (1978).
[11]U. Das, S. Chakravarty, and B. Talukdar, J. Math. Phys. **20**, 887 (1979).
[12]J. M. J. van Leeuwen and A. S. Reiner, Physica (Utrecht) **27**, 99 (1961).
[13]H. Yukawa, Proc. Phys. Math. Soc. Jpn. **17**, 48 (1935); P. Debye and E. Hückel, Phys. Z. **24**, 185 (1923); F. J. Rogers, H. C. Graboshe, Jr., and D. J. Harwood, Phys. Rev. A **1**, 1577 (1970); C. S. Lam and Y. P. Varshni, Phys. Rev. A **4**, 1875 (1971); **19**, 413 (1979).
[14]V. L. Bonch-Bruevich and V. B. Glasko, Sov. Phys. Dokl. **4**, 147 (1959); V. L. Bonch-Bruevich and Sh. M. Kogan, Sov. Phys. Solid State **9**, 993 (1967).
[15]L. Hulthén, Ark. Math. Astron. Fys. **28A**, No. 5, 1 (1942).
[16]G. Ecker and W. Weizel, Ann. Phys. (Leipzig) **17**, 126 (1956).
[17]C. Eckart, Phys. Rev. **35**, 1303 (1930).
[18]R. Dutt, A. Ray, and P. P. Ray, Phys. Lett. A **83**, 65 (1981).
[19]The screening parameter 'a' in Ref. 9 is the reciprocal of our λ.
[20]A. Messiah, *Quantum Mechanics* (North-Holland, Amsterdam, 1961), Vol. I.
[21]C. S. Lam and Y. P. Varshni, Phys. Lett. A **59**, 363 (1976); R. Dutt, Phys. Lett. A **73**, 310 (1979); P. P. Ray and A. Ray, Phys. Lett. A **78**, 443 (1980).
[22]A. W. Babister, *Transcendental Functions Satisfying Non-homogeneous Linear Differential Equations* (Macmillan, New York, 1967), Chap. 6.
[23]W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer, New York, 1966).

A new semiclassical interpretation of the Lamb shift

John T. F. Barwick[a)]
Westinghouse Defense and Electronics Center, P. O. Box 746 M. S. 434, Baltimore, Maryland 21203

(Received 29 January 1983; accepted for publication 29 June 1983)

A modification of a previous semiclassical explanation of the Lamb shift is shown to be applicable to all levels in hydrogenic ions. The phenomenon responsible for the level shifts has not been considered explicitly in other quantum electrodynamic or semiclassical theories, but it is shown that it should be a source of observable energy changes. An approximate calculation for hydrogen s states gives $\Delta E_{1s} = 0.25748\ \text{cm}^{-1}$ (experimental $\Delta E_{1s} = 0.2722\ \text{cm}^{-1}$), $\Delta E_{2s} = 0.03755\ \text{cm}^{-1} = 1125.7$ MHz (experimental $\Delta E_{2s} = 0.03528\ \text{cm}^{-1}$), and $\Delta E_{3s} = 0.01166\ \text{cm}^{-1}$ (experimental $\Delta E_{3s} = 0.0088\ \text{cm}^{-1}$), but more important is the demonstration of an effect which should apparently be involved in any theory of the Lamb shift.

PACS numbers: 03.65.Sq

There has been great interest in the Lamb shift ever since the original confirmation of the small energy difference between the $2S_{1/2}$ and $2P_{1/2}$ states in hydrogen by Lamb and Retherford in 1947, and its almost simultaneous theoretical calculation by Bethe by the methods of quantum electrodynamics. It long remained a phenomenon which was considered purely as a manifestation of quantum electrodynamic effects. The quantitative success of the QED calculations, however, did not eliminate the desire for a simple intuitive picture of the phenomenon. Along with the more recent resurgence of interest in semiclassical ideas, there have appeared a number of discussions concerning the physical origin of the Lamb shift.[1] These have included both new physical interpretations of the QED derivation and new semiclassical theories based on several alternative concepts of its source.[2] The purpose of the following discussion is to suggest another, different semiclassical interpretation of the Lamb shift, which, although similar in spirit to recent neoclassical ideas, is based on electromagnetic effects which have been previously excluded from consideration. It is suggested that the true source of the Lamb shift may be found in the interactions involving the longitudinal electromagnetic fields of the atom, rather than in the transverse field effects almost universally assumed in both QED and semiclassical theories. This is not to say, however, that the shifts predicted by quantum electrodynamics, neoclassical theory, or random electrodynamics are incorrect, or that additional effects must be included in them, but possibly that none of these theories has really correctly identified the actual source of the shifts that they calculate. As indicated below, it is necessary to mesh these hypothesized longitudinal field effects with any effects of semiclassical radiation fields, which will of course still be present during atomic transitions.

Modern QED interpretations of the Lamb shift can be related to different ways of ordering atomic and field operators in the Heisenberg equations of motion for the operators. It is thus possible to interpret the shift as a radiation reaction effect, a vacuum field fluctuation effect, or a combination of both.[1] The semiclassical analog of this QED calculation is very similar and leads to an identical level shift which in this case can be ascribed only to radiation reaction. The semiclassical calculation, however, predicts some higher order differences from QED, such as frequency "chirps" in the emitted radiation and different emission line shapes. In fact, the frequency variation obtained for a single two-level transition is such that at its midpoint, the point of maximum emission, there is no frequency shift at all.[3] Thus *complete* two-level transitions show no net shift due to this semiclassical radiation effect. The fact that there may be no *overall* observable shift of energy levels as a result of atomic transition radiation is important in the present context, however, since it allows for the joint existence of both the transitory radiation reaction effect and the semiclassical effect described below, which is a permanent shift of energy levels unrelated to radiation reaction and which will be observable, even when the radiation field effects, although definitely present, will not. Just as the neoclassical radiation reaction shift is the semiclassical analog of the modern QED interpretation, the present theory can be considered in some ways to be the semiclassical analog of Bethe's original QED interpretation of the Lamb shift as an effect of shrouding an electron in a particular atomic state with its electromagnetic field.

In a previous paper,[4] it was found that the calculation of atomic state energies in terms of the fields set up by the charge and current densities[5]

$$\rho = e\psi\psi^*,$$
$$\mathbf{j} = (e\hbar/2im)(\psi^*\nabla\psi - \psi\nabla\psi^*) - (e^2/mc)(k_e/k_m)^{1/2}\psi\psi^*\mathbf{A}, \tag{1}$$

where ψ was, in general, the transition wave function

$$\psi = \sum_i c_i\,\psi_i, \tag{2}$$

provided a good way to examine certain energy adjustments such as the Lamb shift. The energies were obtained from the fields produced by the sources in Eq. (1) by means of

$$U_E = \frac{1}{8\pi k_e}\int \mathbf{E}^2\,d^3r,$$
$$U_B = \frac{1}{8\pi k_m}\int \mathbf{B}^2\,d^3r. \tag{3}$$

All electromagnetic fields of a single electron charge density were included in Eqs. (3); however, radiation fields, pro-

[a)] Previous address: Roane State Community College, Harriman, TN 37748.

 0022-2488/83/122776-04$02.50

duced while a transition was actually occurring,[3] made no contribution to these integrals. The fields involved here were thus solely stationary state fields.

For the particular case of the hydrogen $2S_{1/2} - 2P_{1/2}$ Lamb shift, the orbiting electron in the $2p$ state was assumed to create its magnetic field $\mathbf{B}$ by carrying its electric field $\mathbf{E}$ at some velocity $\mathbf{v}$. The fields created by this motion were thus

$$\mathbf{E}_1' = \gamma[\mathbf{E}_1 + (1/c)(k_e/k_m)^{1/2}\mathbf{v}\times\mathbf{B}] \cong \mathbf{E}_1(1 + \tfrac{1}{2}\beta^2),$$

and

$$\mathbf{B}_1' = \gamma[\mathbf{B}_1 - (1/c)(k_m/k_e)^{1/2}\mathbf{v}\times\mathbf{E}]$$

$$B_1' = -\gamma(1/c)(k_m/k_e)^{1/2}vE = -\gamma(k_m/k_e)^{1/2}\beta E, \quad (4)$$

where $\gamma = (1-\beta^2)^{-1/2}$. It was assumed that the electric field of the electron itself was that of the $u_{21\pm1}$ state and that it had the same form in the electron and atomic systems (aside from the γ factor). The magnetic field was calculated from $\mathbf{j}_{21\pm1}$. Both fields were then perpendicular to $\mathbf{v}_{21\pm1}$ and therefore fitted Eqs. (4). The motion of the electron thus created $\mathbf{B}_{211}$ and increased $\mathbf{E}_{211}$ by $\frac{1}{2}\beta^2\mathbf{E}_{211}$. The energy shift represented by these field changes was suggested as the source of the Lamb shift, lowering the $2P_{1/2}$ state in energy by an amount equivalent to $\Delta\nu = 1064$ MHz $= 0.0355$ cm^{-1}.

This simple line of argument is now revised to account for s-level shifts, where the orbital current densities in Eq. (1) are zero,[3] and to allow for inclusion of the nuclear field. The need for such an extension is illustrated by the recent measurements of the $1S_{1/2}$–$2S_{1/2}$ transition in hydrogen by two photon spectroscopy.[6] Although Jaynes "neoclassical" theory, as mentioned previously, could possibly have been envisioned as a "dynamic" radiation reaction complement to the previous stationary state field theory for the nS shift in some nS–$n'P$ transitions,[7,8] it could not be included as easily in this $1S$–$2S$ case. In addition to this S state problem, there exists a disagreement of the field velocity calculation as outlined above with the results of recent $2S_{1/2} - 2P_{1/2}$ Lamb shift measurements on various hydrogenic ions,[9] which show roughly a Z^4 dependence on atomic number Z, whereas Eqs. (1)–(3) would imply

$$U_E \propto Z, \quad U_B \propto Z^3. \quad (5)$$

The Z dependence of the Lamb shift, however, suggests the proper generalization of the classical calculation. The nuclear Coulomb field, $\mathbf{E}_+ = k_e\, Ze\mathbf{n}/r^2$, contributes a factor Z in its interactions with the electron fields, and its effects are certainly substantial since, for example, $\mathbf{E}_+$ essentially cancels the electron $\mathbf{E}_-$ over a large region of space around the atom, which classically represents a large energy change. This energy is, of course, easily shown to be equivalent to the ordinary potential energy of the positive-nucleus negative-electron combination.[10] That is,

$$W = \int \rho_- V_+ \, d^3r = \frac{1}{8\pi k_e}\int 2\mathbf{E}_+\cdot\mathbf{E}_- \, d^3r. \quad (6)$$

It is therefore necessary for a classical calculation to begin with a complete expression for the fields in the vicinity of the ion. If it is again assumed that a nonzero expectation value of $\mathbf{v}$ or $\mathbf{L}$ for a hydrogenic state actually implies some sort of movement of the charge density $\rho = e\psi\psi^*$ and that classical electrodynamics applies to the resultant fields, then a magnetic field $\mathbf{B}'$ will be created and the electric field $\mathbf{E}_1'$ will be increased by a small amount as discussed previously [see Eq. (4)]. These changes will affect the energies contained in the total fields, energies which should reflect various potential and interaction energies of the system. The total electric field energy density is, with the integration being over increments of electron charge $\rho_-(\mathbf{r})dV'$,

$$u_E(\mathbf{r}) = \frac{1}{8\pi k_e}\left[\mathbf{E}_+ + \int\left(\mathbf{E}_\parallel' + \mathbf{E}_1' + \frac{1}{2}\beta'^2\mathbf{E}_1'\right)dV'\right]^2$$

$$= \frac{1}{8\pi k_e}\left[\mathbf{E}_+^2 + \left(\int \mathbf{E}'\,dV'\right)^2 + \left(\int \frac{1}{2}\beta'^2\mathbf{E}'\,dV'\right)^2 + 2\mathbf{E}_+\cdot\int \mathbf{E}'\,dV' + \mathbf{E}_+\cdot\int \beta'^2\mathbf{E}_1'\,dV' + \int \mathbf{E}'dV'\cdot\int \beta'^2\mathbf{E}_1'\,dV'\right]. \quad (7)$$

Also,

$$u_B = (1/8\pi k_m)\,\mathbf{B}_-^2 = (1/8\pi k_e)\,\beta^2\mathbf{E}_-^2. \quad (8)$$

The various terms in Eqs. (7) and (8) can now be related to mechanical energies of the system as follows:

$\mathbf{E}_+^2$	= self potential energy of nuclear ρ due to its own field.
$(\int \mathbf{E}'dV')^2$	= self potential energy of electron ρ.
$(\int \beta'^2\mathbf{E}'dV')^2$	= adjustment to above.
$\int \mathbf{E}'dV'\cdot\int \beta'\mathbf{E}'dV'$	= electric field correction to e^- kinetic energy.
$(k_m/k_e)\mathbf{B}_-^2$	= magnetic field correction to e^- kinetic energy.
$2\mathbf{E}_+\cdot\int \mathbf{E}'\,dV'$	= usual interaction potential energy between nucleus and electron.
$\mathbf{E}_+\cdot\int \beta'\mathbf{E}_1'\,dV'$	= correction to the interaction energy of nucleus and electron.

The fourth and fifth terms in this list yield the hydrogen Lamb shift value previously suggested. It is now apparent that these terms should be grouped with the first three as energies which do not affect the solution of the Schrödinger equation for the relative motion and which are unobservable. The sixth and seventh terms are related to the potential energy term which appears in the Schrödinger equation, and the sixth at least must affect the wave function frequencies, since it is the usual potential energy of an eigenstate. This identification can easily be checked by a direct calculation. For $\mathbf{E}_- = \mathbf{E}_{211}$, for example,[3]

$$\frac{1}{8\pi k_e}\int\int\int 2\mathbf{E}_+\cdot\mathbf{E}_{211}\, r^2\sin\theta\, dr\, d\theta\, d\phi = (-k_e\, Z^2e^2)/(4a_0) = -Z^2(6.8)\text{ eV} = \langle V\rangle_{211}. \quad (9)$$

The last energy density listed above should then also affect the wave function frequencies since it is evidently a part of this same potential energy V. But calculations such as that of Eq. (9) show that it has *not* been included. The energy value calculated from this term,

$$\Delta U_{\pm} = \frac{1}{8\pi k_e} \int \mathbf{E}_{+} \cdot \int \beta'^2 \mathbf{E}'_{1\beta'}\, d^3r'\, d^3r, \tag{10}$$

will depend on Z^4 and, because of $\beta^2 = (v/c)^2$, will be of the same order of magnitude as the velocity dependent energies previously calculated. Equation (10) is now hypothesized to be the principal source of the Lamb shift.

Now the discussion so far leaves unresolved the problem of a velocity which would, through Eq. (10), provide an explanation of an S state Lamb shift, a shift which must be recognized to exist and to be in good agreement with QED calculations. Further consideration of the classical fields involved suggests that there does appear to be one important source of electromagnetic fields which has been omitted in the previous discussion, but which might produce significant effects. This is, of course, the *spin* of the electron. The spin magnetic field is as large as the orbital magnetic field but, although some of its effects are accounted for in the fine and hyperfine structure of the energy levels, the charge density velocities and field velocities that it *must* represent have not been examined. It is shown below that such velocities will produce energy level shifts which will account for the S state Lamb shift in hydrogenic ions.

Does the electron spin magnetic moment represent some current density which can be evaluated simply? There is such a current density which can be derived from the Dirac equation for the electron. The Dirac equation, written in two-component form, with

$$\psi = \begin{pmatrix} \phi \\ \chi \end{pmatrix},$$

is

$$c[\,\mathbf{p} - (k_e/k_m)^{1/2}(e/c)\mathbf{A}]\cdot\hat{\sigma}\chi + (eV + mc^2)\,\phi = E\phi,$$
$$c[\,\mathbf{p} - (k_e/k_m)^{1/2}(e/c)\mathbf{A}]\cdot\hat{\sigma}\phi + (eV + mc^2)\chi = E\,\chi, \tag{11}$$

with a current density given by

$$\mathbf{j} = ce\psi\dagger\hat{\alpha}\psi$$
$$= ce(\,\phi^*\chi^*)\begin{pmatrix} 0 & \hat{\sigma} \\ \hat{\sigma} & 0 \end{pmatrix}\begin{pmatrix} \phi \\ \chi \end{pmatrix}. \tag{12}$$

For a positive energy electron, the χ spinor is small and can be expressed, from Eq. (11), in terms of the large component ϕ as

$$\chi = \frac{c[\,\hat{\mathbf{p}} - (k_e/k_m)^{1/2}(e/c)\mathbf{A}]\cdot\hat{\sigma}\phi}{(E + mc^2 - eV)}. \tag{13}$$

Substitution of Eq. (13) in Eq. (12), simplification, and rearrangement yield

$$\mathbf{j} = (ie\hbar/2m)[(\nabla\phi)\dagger\phi - \phi\dagger\nabla\phi]$$
$$- (e^2/mc^2)(k_e/k_m)^{1/2}\phi\dagger\mathbf{A}\phi + (e\hbar/2m)\nabla\times(\,\phi\dagger\hat{\sigma}\phi). \tag{14}$$

The second term here is the vector potential term which provides for a semiclassical account of the Compton effect[4] and spontaneous emission,[2] and the last is an expression for the spin current density. It can be obtained nonrelativistically as the curl of the average magnetic moment density $\mathbf{m} = (e\hbar/2mc)\,(k_e/k_m)^{1/2}\psi\hat{\sigma}\psi$.[11]

There will thus be electromagnetic field motion and consequent energy level shifts for the $u_{n1m} = u_{n00}$ s states in hydrogen, even though the orbital current densities of all these states are zero. The current densities for the 1s, 2s, and 3s states, from Eqs. (1) and (14), are

$$\mathbf{j}_{100} = (e\hbar Z/2ma_0)K_1^2 2e^{-2x}\sin\theta\hat{e}_\phi$$
$$\mathbf{j}_{200} = (e\hbar Z/2ma_0)K_2^2(x-2)(x-4)e^{-x}\sin\theta\hat{e}_\phi,$$
$$\mathbf{j}_{300} = (e\hbar Z/2ma_0)K_3^3(2x^2 - 18x + 27)$$
$$\times(\tfrac{4}{3}x^2 - 20x + 54)e^{-2/3x}\sin\theta\hat{e}_\phi, \tag{15}$$

where $K_1^2 = \pi^{-1}(Z/a_0)^3$, $K_2^2 = (32\pi)^{-1}(Z/a_0)^3$, $K_3^2 = (81)^{-2}(3\pi)^{-1}(Z/a_0)^3$, $x = Zr/a_0$, and $\hat{e}_\phi$ is a unit vector in the ϕ direction. It is now possible to obtain the electric and magnetic fields produced by these states and use them directly in Eq. (10) to determine the resulting changes in energy, but it is mathematically simpler to transform Eq. (10) to a form involving potentials rather than fields. Thus by the use of the relations $\beta'^2\mathbf{E}'_1 = -\beta'\times(\,\beta'\times\mathbf{E}')$, $\mathbf{E}' = -\nabla V'$, and $\nabla^2 V'(\mathbf{r}) = -4\pi k_e\,\rho'\,d^3r'\delta^3(\mathbf{r}-\mathbf{r}')$, Eq. (10) can be converted to

$$\Delta U_{\pm} = \frac{1}{2}\int \beta'^2 V'_{+}\,\rho' d^3r'. \tag{16}$$

This is the same expression as that obtained by using the Lorentz transformed $\rho_- = \gamma(\rho_0 - c^{-1}\beta j_0) \cong \rho_0(1 + \frac{1}{2}\beta^2)$ in Eq. (6).

Equation (16) can now be used to find the energy level shifts if a reasonable form for β_{ns} can be found. Classically one would expect that $\beta = \mathbf{j}/\rho c$, which, for the 1$s$ state, for example, would give

$$\beta_{1s} = \mathbf{j}_{1s}/\rho_{1s}c = (Z\hbar/mca_0)\sin\theta\hat{e}_\phi. \tag{17}$$

The expressions obtained in this way, however, do not agree with a Z component of angular momentum of $\frac{1}{2}\hbar$ and in most cases they become infinite at various points. So if these charge and current densities are to be thought of classically, there must be something more complicated involved than just a single motion of an *entire* charge density, and exact expressions must therefore await the correct description of the internal spin motions of the electron.[12] A reasonable approximate expression for β_{ns} may however be obtained from the expectation value of the spin velocity operator $\hat{\mathbf{v}}_{ns}$ and the form of Eq. (17). For the 2s state as a reference point, then

$$\langle v_{\text{spin}}\rangle_{2s} = \int (\hbar/2m)|\nabla\times\psi^*_{2s}\,\hat{\sigma}\,\psi_{2s}|\,d^3r$$
$$= \int j_\phi/ed^3r = \pi Z\hbar/16ma_0. \tag{18}$$

Thus if it is assumed that $\beta_{2s} = (\pi Z\hbar/16mca_0)\sin\theta\hat{e}_\phi$, Eq. (16) gives

$$\Delta U_{2s} = \frac{1}{2}\int \beta_{2s}^2(k_e\,Ze/r)\,\rho_{2s}\,d^3r$$
$$= 0.037\,55\ \text{cm}^{-1} = 1125.7\ \text{MHz}, \tag{19}$$

which is very close to the QED and experimental values for the 2s Lamb shift.

Now for the whole sequence of states 1s, 2s, 3s, ns, it is necessary to examine how β^2 and $\int \rho_- V_+ d^3r$ must vary in Eq. (16). $\langle\hat{v}^2\rangle$ and $\int \rho_- V_+ d^3r$ both vary as n^{-2}, however the expectation value of v^2 should actually be weighted by $V_+ \propto r^{-1}$. The average value of v^2 will then vary according to

TABLE I. Comparison of S level Lamb shifts.[a]

ns	ΔU	exp.	QED
$1s$	0.2575 cm^{-1}	0.2722 cm^{-1}	0.2718 cm^{-1}
$2s$	0.03755 cm^{-1}	0.0373 cm^{-1}	0.0353 cm^{-1}
$3s$	0.01166 cm^{-1}	0.0083 cm^{-1}	0.0105 cm^{-1}

[a] See Refs. 2 and 6.

$$(v^2)_{ave} = \langle \hat{v}^2/x\rangle/\langle 1/x\rangle = v_0^2(2n^2\langle x^{-2}\rangle - 1/n^2), \tag{20}$$

where $v_0 = Z\hbar/ma_0$. This gives $(v^2)_{1s} = 3v_0^2$, $(v^2)_{2s} = 1.75v_0^2$, and $(v^2)_{3s} = 1.2222v_0^2$. If the β_{ns}^2 values for $n = 1$ and 3 are adjusted according to the ratios of these numbers, Eq. (16) give the Lamb shifts ΔU_{ns} shown in Table I. Experimental and QED values are also shown.

In order to provide an idea of the magnitudes of the $2S_{1/2} - 2P_{1/2}$ Lamb shift for various hydrogenic ions and the corresponding predictions of the present Z^4 relationship, the results are compared in Table II. The numbers for this theory show simply the Z^4 variation and are unadjusted for higher order effects which, as in the QED calculations, could make important corrections.[13] It should be emphasized that $\Delta U_{\pm}$ depends sensitively on β and the values obtained here are meant mainly to show that this simple semiclassical method does give reasonable answers.

Finally, how does one reconcile the fact that both S and P states now seem to exhibit a shift? The answer comes from the fact that the spin effects must also be included in the $nP_{1/2}$ states. This can easily be done as follows. The total magnetic moment of the electron in any state, including both orbital and spin motion, can be written as[14]

$$\boldsymbol{\mu} = (e/2mc)(k_e/k_m)^{1/2}(\mathbf{j} + \mathbf{s}). \tag{21}$$

Now for the $nP_{1/2}$ state $l = 1, s = 1/2, j = 1/2$, and therefore $\mathbf{s}$ and $\mathbf{j}$ are in opposite directions ($\mathbf{j} = \mathbf{l} + \mathbf{s}$). Thus $\mu_{np} = 0$. The orbital effects are cancelled by the spin effects, implying that they are equal in magnitude. But then the net result is that the $nP_{1/2}$ state is actually unshifted. For the $nS_{1/2}$ state $l = 0, s = \frac{1}{2}, j = \frac{1}{2}$, and thus $\mathbf{s}$ and $\mathbf{j}$ are in the same direction, implying that $\mu_{ns} = (k_e/k_m)^{1/2}(e\hbar/2mc)$. Therefore only the $nS_{1/2}$ states actually have any net field motion and accompanying energy shifts.

TABLE II. Lamb shift frequencies for the $2^2P_{1/2} - 2^2S_{1/2}$ transition in hydrogenic atoms and ions.

Ion	Experiment[13]	This theory[a]
hydrogen (H)	1 057.8 MHz	1 125.7 MHz
deuterium (D)	1 059 MHz	1 125.7 MHz
helium	14 040 MHz	18 011 MHz
lithium ($^{6}_{3}Li^{+2}$)	62 800 MHz	91 182 MHz
carbon ($^{12}_{6}C^{+5}$)	780 GHz	1 459 GHz
oxygen ($^{16}_{8}O^{+7}$)	2 210 GHz	4 611 GHz
fluorine ($^{20}_{10}F^{+9}$)	3 339 GHz	11 257 GHz

[a] Values are uncorrected for any higher order terms, vacuum polarization, or nuclear size effect.

[1] P. W. Milonni, Phys. Rep. **25**, 1–81 (1976).
[2] M. D. Crisp and E. T. Jaynes, Phys. Rev. **179**, 1253 (1969); J. H. Eberly, "Unified View of Spontaneous Emission," in *Foundations of Radiation Theory and Quantum Electrodynamics*, edited by A. O. Barut (Plenum, New York, 1980), p. 31.
[3] J. T. F. Barwick, Ph. D. dissertation (Univ. of Tennessee, 1979) (unpublished).
[4] J. Barwick, Phys. Rev. A **17**, 1912 (1978).
[5] All equations are written with whatever combination of the constants k_e and k_m is necessary to make them compatible with all systems of electromagnetic units. For SI, $k_e = 1/4\pi\epsilon_0$, $k_m = \mu_0/4\pi$; Gaussian, $k_e = 1$, $k_m = 1$; Heavyside–Lorentz, $k_e = 1/4\pi$, $k_m = 1/4\pi$; esu, $k_e = 1$, $k_m = 1/c^2$; emu, $k_e = c^2$, $k_m = 1$.
[6] T. W. Hansch, S. A. Lee, and C. Wieman, Phys. Rev. Lett. **35**, 1262 (1975); T. W. Hansch, A. L. Schawlow, and G. W. Series, Sci. Am. **240**(3), 94 (1979).
[7] G. Herzberg, Proc. R. Soc. London, Ser A **234**, 516 (1956).
[8] S. Triebwasser, E. S. Dayhoff, and W. E. Lamb, Phys. Rev. **89**, 98 (1953).
[9] P. J. Mohr, "Lamb Shift in Hydrogenlike Ions," in *Beam Foil Spectroscopy*, edited by I. Sellin and D. Pegg (Plenum, New York, 1976), Vol. 1, p. 89.
[10] J. D. Jackson, *Classical Electrodynamics*, 1st ed. (Wiley, New York, 1962), Chap. 1.
[11] D. I. Blokhintsev, *Principles of Quantum Mechanics* (Allyn and Bacon, Boston, 1964), p. 246.
[12] Note the necessity of using all components (χ) of the electron spinor to get any j_{spin} at all.
[13] Mohr, Ref. 9, p. 90.
[14] R. B. Leighton, *Principles of Modern Physics* (McGraw-Hill, New York, 1959), p. 277.

B^*-algebra representations in a quaternionic Hilbert module[a)]

A. Soffer and L. P. Horwitz
Tel Aviv University, Ramat Aviv, 69978 Tel Aviv, Israel

(Received 24 June 1982; accepted for publication 26 November 1982)

It is shown that the Gel'fand–Naimark–Segal (GNS) construction can be generalized to real B^*-algebras containing an algebra *-isomorphic to the quaternion algebra by the use of quaternion linear functionals and Hilbert Q-modules. An extension of the Hahn–Banach theorem to such functionals is proved.

PACS numbers: 03.70. + k, 11.15. − q

1. INTRODUCTION

In recent years, considerable effort has been made in the development of quantum field theories with nonabelian gauge fields.[1] Since the fundamental objects described by theories of this type, the quanta of the gauge fields and matter fields, are not observed in experiment directly, it is a logical requirement on these theories that they do not admit direct observation of these objects. The lack of decisive results on this problem of confinement has led to the suggestion that achieving a semiclassical understanding of the dynamics of such systems would be a useful first step.[2]

It is difficult to define and study a semiclassical limit for a field theory of this type in three-space and one-time dimensions, since there is no natural scale for achieving such a limit. Khriplovich,[2] Giles and McLerran,[3] and Adler[4] have suggested a direction for the development of a semiclassical understanding of theories of this type, and in a series of papers Adler[5] has worked out a systematic procedure for obtaining dynamical equations describing field configurations, and the static potentials, in a semiclassical framework.

The correspondence between Adler's construction and the usual approach taken in quantum field theory has not yet been clarified. On the other hand, Biedenharn, Sepunaru, and Horwitz[6] have shown that the algebraic structures associated by Adler's construction with the special case of an underlying U(2) gauge group can be obtained from quaternionic quantum theory.

Originally proposed by Birkhoff and von Neumann,[7] some quantum mechanical aspects of vector spaces over quaternion multipliers (that is, vector spaces which are also modules) were worked out by Finkelstein, Jauch, Schiminovich, and Speiser[8]; a more complete study describing a hierarchy of scalar products linear over real and complex subalgebras, as well as quaternion linear, and the projection operators into corresponding linear manifolds, was carried out by Horwitz and Biedenharn.[9] The decomposition of the space into complex linear subspaces was utilized[9] to construct a tensor product and a procedure for second quantization.

A direct construction of a Hilbert Q-module[10] of the type carried out by Horwitz and Biedenharn[9] involves a somewhat *ad hoc* definition of the quantum state, and is justified formally through the Gleason theorem.[11] The approach taken by Cassinelli, Truini, and Biedenharn[12] utilizes Mackey's theory of induced representations of imprimitivity systems on a Hilbert Q-module, but the same assumptions are made on the structure of the states. In this paper, we shall start with a B^* algebra over the reals, which contains, as additional structure, a subalgebra *-isomorphic (under the same * operation) to the quaternions. With the help of the positive linear functionals which map the B^*-algebra into the quaternions, and are two-sided linear (through the *-isomorphism), we show that Gel'fand–Naimark–Segal (GNS) type construction can be used to represent the B^*-algebra in a Hilbert Q-module. This procedure provides a deeper insight into the meaning of physical states in a Hilbert Q-module. We prove the Hahn–Banach theorem, which shows that these states provide a Hausdorf topology on the B^*-algebra.

2. QUATERNION VALUED STATES

Consider a B^*-algebra A over the reals, which contains a subalgebra A_Q *-isomorphic to the real quaternions; i.e.,

$$i: A_Q \to Q \tag{2.1}$$

is a *-isomorphism. There may be many such subalgebras; in the following, we pick one of them.

Lemma 2.1: i is an isometry:

$$\|A_q\| = |q|, \tag{2.2}$$

where $A_q \in A_Q$, $i(A_q) = q \in Q$, $\|a\|$ is the B^*-norm of a in A, and $|q|$ is the usual quaternion norm, the real positive number $\sqrt{q^*q}$. The proof follows from the observation that $i(A_q^* A_q) = |q|^2$ implies $A_q^* A_q = |q|^2 \cdot 1$ and the fact that A is a B^*-algebra (in a B^*-algebra,[12] $\|x^*x\| = \|x\|^2$).

Let ρ be a linear mapping $\rho: A \to Q$ (we shall also use the term functional to describe a mapping into Q).

Definition 2.1: We shall say that ρ is two-sided quaternion linear (relative to A_Q) if

$$\rho(A_q a A_{q'}) = q\rho(a)q' \quad \forall A_q, A_{q'} \in A_Q.$$

It is positive if $\rho(a^*a) \geqslant 0$ for all $a \in A$. For positive ρ, if $\rho(I) = 1$, we shall call it a state.

It follows that, for quaternion linear ρ,

$$\rho(A_q) = q. \tag{2.3}$$

Lemma 2.2: If ρ is positive and two-sided quaternion linear, then

$$\rho(a^*) = \rho(a)^* \quad \forall a \in A.$$

Proof: $\rho[(a^* + b^*)(a + b)] = \rho(a^*a) + \rho(b^*b) + \rho(a^*b) + \rho(b^*a) \geqslant 0$.

It follows that

$$\operatorname{Im}\rho(a^*b) = -\operatorname{Im}\rho(b^*a), \tag{2.4}$$

[a)] Research supported in part by the Binational Science Foundation (BSF), Jerusalem.

 0022-2488/83/122780-03$02.50

where we define

$$\operatorname{Im} q = \tfrac{1}{2}(q - q^*),$$
$$\operatorname{Re} q = \tfrac{1}{2}(q + q^*). \tag{2.5}$$

Now, replace b in Eq. (2.4) by bA_q. Equation (2.4) then becomes

$$\operatorname{Im}[\rho(a^*b)q + q^*\rho(b^*a)] = 0. \tag{2.6}$$

Let us take q to be a pure imaginary quaternion (the real component does not provide new information). Then, using Eq. (2.4), we see that

$$\operatorname{Im}\{[\operatorname{Im}\rho(a^*b)]q - q[\operatorname{Im}\rho(b^*a)]\}$$
$$= \operatorname{Im}\{[\operatorname{Im}\rho(a^*b)]q + q[\operatorname{Im}\rho(a^*b)]\} = 0,$$

since the symmetric product of two imaginary quaternions is real. We are thus left, in Eq. (2.6), with

$$\operatorname{Im}\{[\operatorname{Re}\rho(a^*b)]q - q\operatorname{Re}[\rho(b^*a)]\}$$
$$= q\{\operatorname{Re}\rho(a^*b) - \operatorname{Re}\rho(b^*a)\} = 0.$$

Since Q is a division ring, it follows that

$$\operatorname{Re}\rho(a^*b) - \operatorname{Re}\rho(b^*a) = 0.$$

Together with Eq. (2.4), this completes the proof of the lemma.

Lemma 2.3: Let ρ be a positive Q-linear mapping. Then, the Schwarz inequality

$$|\rho(a^*b)|^2 \leqslant \rho(a^*a)\rho(b^*b) \quad \forall a,b \in A \tag{2.7}$$

is valid.

Proof: $\rho(a^*a) \geqslant 0$ implies $\rho(a^*a) + \lambda^2\rho(b^*b) + 2\lambda \operatorname{Re}\rho(a^*b) \geqslant 0$, for λ real, and hence

$$[\operatorname{Re}\rho(a^*b)]^2 \leqslant \rho(a^*a)\rho(b^*b).$$

From quaternion linearity, we may replace b by bA_q to obtain

$$(\operatorname{Re}(\rho(a^*b)q))^2 \leqslant \rho(a^*a)\rho(b^*b)$$

for $|q|^2 = 1$. Let us write $\rho(a^*b) = |\rho(a^*b)|u$, where $|u|^2 = 1$. The maximum of $\operatorname{Re}(uq)$ is unity, since $|uq|^2 = 1$, and hence Eq. (2.7) follows.

3. GNS CONSTRUCTION

Let V be a vector space over the reals which is also a right Q-module, i.e., a vector space over the quaternions Q so that if $x, y \in V$, $q_1, q_2 \in Q$,

$$xq_1 + yq_2 \in V,$$
$$(x + y)q = xq + yq,$$
$$x(q_1q_2) = (xq_1)q_2,$$

and

$$x(q_1 + q_2) = xq_1 + xq_2.$$

Suppose, moreover, that there exists a binary mapping (x, y) of $V \times V$ into Q with the following properties:

(i) $(x,y)^* = (y,x)$,

where $*$ is the involution in Q,

(ii) $(x + y,z) = (x,z) + (y,z)$,

(iii) $(x, yq) = (x, y)q$,

(iv) $(x,x) = \|x\|^2 \geqslant 0$,

and is zero if and only if $x = 0$.

Definition 3.1: A right Q-module V with properties (i)–(iv), which is closed under the topology defined by the norm $\|x\|$, will be called a Hilbert Q-module.

Definition 3.2: The mapping $\hat{A}: \mathscr{H} \to \mathscr{H}$ of a Hilbert Q-module into itself will be called a quaternion linear (Q-linear) operator if $\hat{A}(xq) = (\hat{A}x)q$ and $\hat{A}(x + y) = \hat{A}x + \hat{A}y$ for all $q \in Q$ and all x, y in the domain of $\hat{A}$.

We are now in a position to state Theorem 3.1.

Theorem 3.1: Let A be a B^*-algebra over the reals which contains a subalgebra A_Q *-isomorphic to the real quaternions Q, and ρ a two-sided quaternion linear state on A. Then there exists a representation $\pi_\rho: A \to \mathscr{B}(\mathscr{H}_\rho)$, where $\mathscr{H}_\rho$ is a Hilbert Q-module and $\mathscr{B}(\mathscr{H}_\rho)$ is the set of bounded Q-linear operators on $\mathscr{H}_\rho$.

Proof: Let $\mathscr{I}$ be the left ideal consisting of all elements $a \in A$ for which $\rho(a^*a) = 0$. Then $A \xrightarrow{h} A/\mathscr{I}$ is a homomorphism for which $h(A)$ is a Q-module over $h(A_Q)$. We define a scalar product in $A/\mathscr{I}$ by

$$(\eta,\xi) = \rho(a_\eta^* a_\xi), \tag{3.1}$$

where a_η, a_ξ are elements of the equivalent classes η, ξ in $A/\mathscr{I}$.

We define the multiplication ηq as the equivalence class $\{aA_q\}$. The scalar product (3.1) then has the properties

$$\begin{aligned} &(\eta,\xi + \chi) = (\eta,\xi) + (\eta,\chi),\\ &(\eta,\xi q) = (\eta,\xi)q,\\ &(\eta,\xi)^* = (\xi,\eta),\\ &(\eta,\eta) = \|\eta\|^2 \geqslant 0, \end{aligned} \tag{3.2}$$

and is zero if and only if $\eta = 0$. The first and second properties of Eq. (3.2) follow from the quaternion linearity of ρ, and the third property from Lemma 2.2. The last property follows from the definition of $A/\mathscr{I}$. We therefore complete $A/\mathscr{I}$ in the topology provided by the norm $\|\eta\|$ to obtain a Hilbert Q-module which we denote by $\mathscr{H}_\rho$. Elements a of A are mapped onto operators $\hat{A}$ on $\mathscr{H}_\rho$ by

$$\{aa_\eta\} = \hat{A}\eta \tag{3.3}$$

for $\eta \in \mathscr{H}_\rho$ and a_η any element of the equivalence class which defines η. The operator $\hat{A}$ is bounded by the norm of a in the B^*-algebra (the functional ρ is continuous; the proof is as in the complex case[13]). $\hat{A}$ can therefore be extended to a bounded (linear) mapping of $\mathscr{H}_\rho$ into itself.

4. HAHN–BANACH THEOREM

The validity of the Hahn–Banach theorem implies that the set of linear functionals ρ that can be constructed on the B^*-algebra is separating.

Theorem 4.1: Let A be a B^*-algebra over the reals which contains a subalgebra A_Q, *-isomorphic to the real quaternions. Let λ be a two-sided quaternion linear functional defined on a subspace $Y \subset A$, which is an A_Q-module and is bounded by the real function $p(x) = \|x\|$. Then λ has a two-sided quaternion linear extension Λ, which is also bounded by $p(x)$ and coincides with λ on Y.

Proof: We first remark that $p(x)$ is quaternion convex,

since

$$p(A_\alpha x + A_\beta y) = \|A_\alpha x + A_\beta y\| \leqslant p(A_\alpha x) + p(A_\beta y)$$
$$\leqslant p(A_\alpha)p(x) + p(A_\beta)p(y)$$
$$= |\alpha| p(x) + |\beta| p(y) \quad \forall A_\alpha, A_\beta \in A_Q. \tag{4.1}$$

The last equality follows from Lemma 2.1. Let

$$l(x) = \mathrm{Re}\,\lambda(x). \tag{4.2}$$

Then $l(x)$ is real linear and

$$\lambda(x) = \sum_{i=0}^{3} e_i l(A^*_{e_i} x). \tag{4.3}$$

Equation (4.3) follows from the fact that

$$l(A^*_{e_i} x) = \mathrm{Re}\,\lambda(A^*_{e_i} x) = \mathrm{Re}(e_i^* \lambda(x)),$$

which projects the four real-valued components of $\lambda(x)$.

By the real Hahn–Banach theorem, $l(x)$ has an extension $L(x)$ to all of A, which is real linear and satisfies

$$|L(x)| \leqslant p(x) \quad \forall x \in A. \tag{4.4}$$

We define the two-sided quaternion linear extension of $\lambda(x)$ by

$$\Lambda(x) = \sum_{i=0}^{3} e_i L\left[P_0(A^*_{e_i} x)\right], \tag{4.5}$$

where

$$P_0 y = \tfrac{1}{4} \sum_{i=0}^{3} A^*_{e_i} y A_{e_i} \tag{4.6}$$

commutes with all $A_q \in A_Q$. The left linearity of $\Lambda(x)$ follows from the fact that the replacement $x \to A_{e_j} x$ induces a permutation of the functionals in the sum (4.5) that is compensated by the extraction of a factor e_j from each of the e_i. For example,

$$\begin{aligned}\Lambda(A_{e_1} x) = &- L\left[P_0(A^*_{e_1} x)\right] + e_1 L\left[P_0(x)\right] \\ &- e_2 L\left[P_0(A^*_{e_3} x)\right] + e_3 L\left[P_0(A^*_{e_2} x)\right] \\ = e_1 \{ &L\left[P_0(x)\right] + e_1 L\left[P_0(A^*_{e_1} x)\right] \\ &+ e_2 L\left[P_0(A^*_{e_2} x)\right] + e_3 L\left[P_0(A^*_{e_3} x)\right] \}.\end{aligned} \tag{4.7}$$

Since

$$P_0(A^*_{e_i} x) = P_0(x A^*_{e_i}), \tag{4.8}$$

the same argument illustrated in Eq. (4.7) shows that $\Lambda(x)$ is right linear also. Since $\Lambda(x)$ is linear, it follows that for some $q(|q| = 1)$,

$$|\Lambda(x)| = q\Lambda(x) = \Lambda(A_q x) = \mathrm{Re}\,\Lambda(A_q x)$$
$$= L(P_0(A_q x)) \leqslant p(A_q x).$$

Furthermore, since $p(A_q x) \leqslant |q| p(x) = p(x)$ and

$$p(x) = p(A^*_q A_q x) \leqslant |q^*| p(A_q x) = p(A_q x),$$
$$p(A_q x) = p(x), \tag{4.9}$$

i.e., $\Lambda(x)$ is bounded by $p(x)$.

We now show that $\Lambda(x)$ coincides with $\lambda(x)$ for $x \in Y$. Since Y is an A_Q-module, $A^*_{e_i} x A_{e_i} \in Y$ if $x \in Y$. Hence

$$L\left[P_0(A^*_{e_i} x)\right] = l\left[P_0(A^*_{e_i} x)\right] \tag{4.10}$$

for $x \in Y$. However,

$$l[P_0(x)] = \mathrm{Re}\,\lambda[P_0(x)]$$
$$= \tfrac{1}{4}\mathrm{Re} \sum_{i=0}^{3} \lambda(A^*_{e_i} x A_{e_i})$$
$$= \tfrac{1}{4}\mathrm{Re} \sum_{i=0}^{3} e_i^* \lambda(x) e_i.$$

From the quaternion multiplication table, it follows that

$$\tfrac{1}{4} \sum_{i=0}^{3} e_i^* \lambda(x) e_i = \mathrm{Re}\,\lambda(x). \tag{4.11}$$

Hence

$$l[P_0(x)] = l(x), \tag{4.12}$$

which completes the proof of our theorem. We remark that one could alternatively decompose every x in A uniquely into a combination of four terms of type (4.6) with coefficients A_{e_i}, and use the real Hahn–Banach theorem for the extension of real linear functionals on elements of type Eq. (4.6) in Y to the subspace of elements of this type on the full algebra A. One can then construct a two-sided quaternion linear extension of $\lambda(x)$ of the form Eq. (4.5).

Corollary 4.1: The weakest topology for which all of the two-sided quaternion linear functionals are continuous is Hausdorf.

ACKNOWLEDGMENTS

We wish to thank Professor M. Flato for suggesting the problem, and for his hospitality at the University of Dijon where this work was initiated. We also thank him and Professor D. Sternheimer for discussions on this and related problems.

[1]For a recent review, see, for example, C. Itzykson and J. B. Zuber, *Quantum Field Theory* (McGraw-Hill, New York, 1980).
[2]I. B. Khriplovich, Sov. Phys. JETP **47**, 1 (1978) [Zh. Eksp. Teor. Fiz. **74**, 37 (1978)].
[3]R. Giles and L. McLerran, Phys. Lett. B **79**, 447 (1978); Phys. Rev. D **19**, 3732 (1979); Phys. Rev. D **21**, 1672 (1980).
[4]S. L. Adler, Phys. Rev. D **17**, 3212 (1978).
[5]S. L. Adler, Phys. Rev. D **18**, 411 (1978); Phys. Lett. B **86**, 203 (1979); Phys. Rev. D **19**, 1168 (1979); Phys. Rev. D **20**, 1386 (1979).
[6]L. C. Biedenharn, D. Sepunaru, and L. P. Horwitz, *IX International Colloquium on Group Theoretical Methods*, Cocoyoc, Jan. 23–27, 1980, edited by K. B. Wolf, *Lecture Notes in Physics*, Vol. 135 (Springer-Verlag, Berlin, 1981), p. 51; L. P. Horwitz and L. C. Biedenharn, *Weak Interaction as Probes of Unification*, edited by G. B. Collins, L. N. Chang, and J. R. Ficenec, AIP Conf. Proc. No. 72, Particles and Fields Subseries 23 (AIP, New York, 1981), p. 553.
[7]G. Birkhoff and J. von Neumann, Ann. Math. **37**, 823 (1936).
[8]D. Finkelstein, J. M. Jauch, S. Schiminovich, and D. Speiser, J. Math. Phys. **3**, 207 (1962); **4**, 788 (1963).
[9]L. P. Horwitz and L. C. Biedenharn, "Quaternion Quantum Mechanics and Second Quantization," Duke University preprint 1981. See also, J. Rembieliński, J. Phys. A: Math. Gen. **13**, 15 (1980); **13**, 23 (1980).
[10]See, for example, M. J. Dupré and P. A. Fillmore, in *Topics in Modern Operator Theory*, edited by Apostol *et al.*, *Operator Theory: Advances and Applications*, Vol. 2 (Birkhäuser Verlag, Basel, Boston, Stuttgart, 1981), and references contained therein.
[11]A. M. Gleason, J. Math. Mech. **6**, 885 (1957). See also, C. Piron, *Foundations of Quantum Physics* (Benjamin, New York, 1976).
[12]G. Cassinelli and P. Truini, "Quantum Mechanics of the Quaternionic Hilbert Space based upon the Imprimitivity Theorem," Duke University preprint 1981; P. Truini, L. C. Biedenharn, and G. Cassinelli," Imprimitivity Theorem and Quaternionic Quantum Mechanics," Duke University preprint 1981.
[13]M. A. Naimark, *Normed Rings* (Noordhoff, Groningen, 1964), p. 189.

A one-fixed-point Killing parameter transform

J. P. Krisch
Department of Physics, University of Michigan, Ann Arbor, Michigan 48109

(Received 14 September 1982; accepted for publication 5 August 1983)

A single fixed-point transformation which generates solutions to the field equations is discussed. The method is applied to several examples.

PACS numbers: 04.20.Jb

I. INTRODUCTION

There has been much recent interest in generating new solutions to the vacuum field equations by transforming known solutions.[1–5] One very useful transformation method was developed by Geroch,[4] who generalized the work of Ehlers[2] and Harrison.[3] The original method is applicable to spaces which have one timelike Killing vector. Given a metric g_{ab} with timelike Killing vector ξ^a, this transformation technique will produce a new metric g'_{ab} with the same Killing vector. As described by Geroch,[4] the new metric is generated from the base metric by projective transformations on the scalar norm λ and scalar twist ω of the Killing vector, where

$$\lambda = \xi^a \xi_a, \qquad \omega_a = \epsilon_{abcd} \xi^b \nabla^c \xi^d = D_a(\omega). \tag{1}$$

The transformations are performed in the three-dimensional manifold defined by the Killing trajectories. The covariant derivative in this space is D_a.

The transformation is expressed in terms of a complex Ernst potential $\tau = \omega + i\lambda$. The transformed potential is given by

$$\tau' = (a\tau + b)/(c\tau + d). \tag{2}$$

In this original formulation, a particular parametrization was chosen, $a = d = \cos\gamma$ and $b = -c = \sin\gamma$. This choice is one of the simplest to make. It also has the nice physical consequence of making the transform a rotation of potential functions in the orbit space.[6]

This choice of parametrization has some other consequences. Any bilinear transform leaves up to three points fixed. The single parameter form has two fixed points corresponding to $\tau = (\omega, \lambda) = (0, \pm 1)$. One of the fixed points can be identified as infinity, where λ takes on its asymptotic Minkowski value. The second fixed point is difficult to interpret. Using Schwarzschild parameters, for example, the second point occurs at $r = M$, a point inside the event horizon. Because of the ambiguity in the second fixed point, it is of interest to examine the one-fixed-point form. The purpose of this paper is to discuss the one-fixed-point transformation.

The next section contains a brief review of the formalism and the one-fixed-point transform is written down. The parameters of the transform are discussed. In this section we derive the differential equation obeyed by the parameters. In the last part of the paper, the transformation is applied to some specific examples.

II. THE TRANSFORMATION

A. The formalism

Start with a vacuum solution g_{ab} possessing a single timelike Killing vector ξ^a. The norm λ and twist ω_a of the Killing vector are given by Eq. (1). The solution g_{ab} is described by a set of equations on a four-dimensional space M: g_{ab}. Geroch[4] has shown that g_{ab} is also described by a set of equations written on the three-dimensional manifold H: h_{ab} of Killing trajectories $h_{ab} = \lambda(g_{ab} - \xi_a \xi_b/\lambda)$:

$$R_{ab} = -2(\tau - \bar{\tau})^{-2}(D_{(a}\bar{\tau} D_{b)}\tau), \tag{3a}$$

$$D^2\tau = 2(\tau - \bar{\tau})^{-1}(D\tau)\cdot(D\tau), \tag{3b}$$

where $\tau = \omega + i\lambda$, and D is the covariant derivative in H. Ernst[7] has demonstrated that Eq. (3b) is derivable from an action and is equivalent to the field equations in the axially symmetric case. He gives a prescription for generating metric components from potentials satisfying this Ernst equation.

To generate a new metric g'_{ab} from g_{ab}, one may go to H and look for a new solution τ' of Eq. (3b). Geroch's τ' is of the form (2) which he writes as

$$\tau' = (\cos(\gamma)\tau + \sin(\gamma))/(-\sin(\gamma)\tau + \cos(\gamma)). \tag{4}$$

It is easily verified that τ' will satisfy the Ernst equation for constant γ. Using τ', new metric components can be constructed.[4]

The fixed points corresponding to Eq. (4) are found by setting $\tau' = \tau$. One obtains $\tau_0 = \tau'_0 = \pm i$. The positive fixed point corresponds to infinity, $\lambda = 1$. The negative one is difficult to interpret. The choice of a fixed point at infinity is a good one since it ensures the asymptotic behavior of the Killing norm and twist. Instead of Eq. (4), write down Eq. (2) with the single fixed point $\tau'_0 = \tau_0 = (0,1)$. One obtains[8]

$$1/(\tau' - \tau_0) = 1/(\tau - \tau_0) + \beta', \tag{5}$$

where β' is possibly complex. This equation can be put into a linear form by defining the Ernst function

$$\xi = (i - \bar{\tau})/(i + \bar{\tau}). \tag{6}$$

With this substitution, Eq. (5) becomes

$$\xi' = \xi + i\beta, \tag{7}$$

$\beta = 2\beta'$. The usual projective transform has β a constant. In the next section we will discuss the conditions that β must meet in order that ξ' represent a solution to the field equations. We will find that allowing β to be coordinate dependent leads to interesting solutions.

 0022-2488/83/122783-03$02.50

B. The parameter

1. Differential equation for the parameter

Equation (7) is to generate a new solution ξ' to the equivalent field equations. It is necessary that ξ' satisfy an Ernst equation equivalent to Eq. (3b),[7]

$$(\xi'^*\xi' - 1)D'^2\xi' = 2\xi^*'D'\xi'\cdot D'\xi'. \tag{8}$$

This requirement can be used to determine β. Substituting Eq. (7) one obtains

$$(\psi^2 + \beta_R^2 - 1)D'^2\psi = 2\psi(D'\psi\cdot D'\psi - D'\beta_R\cdot D'\beta_R) + 4\beta_R D'\psi\cdot D'\beta_R, \tag{9a}$$

$$(\psi^2 + \beta_R^2 - 1)D'^2\beta_R = -2\beta_R(D'\psi\cdot D'\psi - D'\beta_R + D'\beta_R) + 4\psi D'\psi\cdot D'\beta_R, \tag{9b}$$

where $\psi = \xi - \mathrm{Im}(\beta), \beta_R = \mathrm{Re}(\beta)$. We have assumed ξ real for simplicity. The covariant derivative in the transformed space H' with $h_{ab} = h'_{ab}$ is D'.

In order that ξ' be a solution to the field equations, it is necessary that β satisfy Eq. (9). One immediately notices the only constant β solution is the trivial transformation $\beta_I = \text{const}, \beta_R = 0$. Physically significant solutions will have β coordinate dependent. This is a broad generalization of the usual constant parameter projective transform. In order that Eq. (5) still represents a fixed point at ∞ we require $\lim_{\to\infty}(\beta/r) = 0$. The fixed-point condition is satisfied in this limit.

2. Interpretation of β

In order to understand the physical significance of the real part of β, examine the asymptotic form of Eq. (5). Assuming β real, the imaginary part of Eq. (5) is

$$\omega' = \frac{\omega(1 + \beta'\omega) + \beta'(\lambda - 1)^2}{(1 + \beta'\omega)^2 + (\beta')^2(\lambda - 1)^2}. \tag{10}$$

Consider H: h_{ab} to be asymptotically flat in the sense of Geroch[9] and Ashtekar and Ashtekar.[10,11] This means there exists a conformally related manifold H_0: $\Omega^2 h_{ab}$, which at Λ, the point at infinity, is smooth on the completed manifold. Choose the conformal factor to be $\Omega = (\lambda - 1)^2$, $\Omega' = (\lambda' - 1)^2$,[12] with $\lim_{\to\Lambda}\Omega = \lim_{\to\Lambda}\Omega' \sim 1/r^2$. Defining asymptotic twists $\omega_0 = \lim_{\to\Lambda}\omega/\Omega$, and $\omega_0' = \lim_{\to\Lambda}\omega'/\Omega'$, and noting the imposed convergence of β', $\lim_{\to\Lambda}\beta'/r = 0$ implies $\lim \beta'\omega = 0$, we have

$$\omega_0' = \omega_0 + \lim_{\to\Lambda}\beta'. \tag{11}$$

The one-fixed-point transform is a simple translation of a scalar twist defined at infinity. In the case where the base space is static, $\lim_{\to\Lambda}\beta'$ can be identified as a projected scalar twist at infinity. This identification helps in understanding the coordinate dependence of β'. Adding rotation to a static space could, for example, reduce the symmetry from spherical to axial. A coordinate dependent β' accomplishes this.

III. APPLICATIONS

A. Schwarzschild metric

We will take the base space to be the Schwarzschild metric in prolate spheroidal coordinates. The convenience of this choice has been stressed by Vorhees.[12] In these coordinates $\xi = x$, with $x = r/m - 1$, r is the usual polar radius and $y = \cos\theta$. In their usual form, the coordinates are normalized to unit distance between foci. This is acceptable for one space, say the base space, but is an overly restrictive assumption to impose throughout the transformation. In general we have

$$D'^2A = \frac{1}{(x^2 - d^2y^2)}\left(\frac{\partial}{\partial x}(x^2 - d^2)\frac{\partial A}{\partial x} + \frac{\partial}{\partial y}(1 - y^2)\frac{\partial A}{\partial y}\right),$$
$$D'A\cdot D'B = \frac{1}{(x^2 - d^2y^2)}\left((x^2 - d^2)\frac{\partial A}{\partial x}\frac{\partial B}{\partial x} + (1 - y^2)\frac{\partial A}{\partial y}\frac{\partial B}{\partial y}\right). \tag{12}$$

Substituting into Eq. (9), one finds a solution for β, $\mathrm{Im}\,\beta = 0$, $\mathrm{Re}\,\beta = cy = c\cos\theta$, with $c^2 + d^2 = 1$.

This solution generates the Kerr metric with $c = -a/2r_0$.

B. $\xi = \xi(x), \beta = \beta(y), \beta_{\text{real}}$

Again using prolate spheroidal coordinates we assume ξ is a general function of x, and β a real general function of y. We will investigate what kinds of base spaces satisfying this will generate new solutions $\xi'(x, y)$.

Using Eq. (12) we see the last terms of Eq. (9) vanishes. Equation (9) becomes

$$D'^2\xi(x)/\xi(x) = -D'^2\beta(y)/\beta(y) = \text{const} = c_1, \tag{13}$$

which is Legendre's equation. $\xi(x)$ and $\beta(y)$ will both then satisfy a Legendre's equation in their own coordinate with $c_1 = L(L + 1)$. We have

$$\beta_L(y) = a_L P_L(y) + b_L Q_L(y),$$
$$\xi_L(x) = f_L P_L(x) + g_L Q_L(x). \tag{14}$$

We can then say that for any space whose ξ are either polynomial or logarithmic in x, we can generate new solutions. The Schwarzschild solution of part A is obviously a special case of this with $L = 1$ and imposed asymptotic flatness and regularity.[13]

C. Slow rotation

The identification of real β with an angular speed allows Eq. (9) to be written in a slow rotation approximation, to first order in β, assuming β real. We obtain

$$(\xi^2 - 1)D'^2\xi = 2\xi D'\xi\cdot D'\xi,$$
$$(\xi^2 - 1)D'^2\beta = 2\beta D'\xi\cdot D'\xi + 4\xi D'\xi\cdot D'\beta. \tag{15}$$

The first equation merely says that in the slow rotation limit, ξ will continue to satisfy an Ernst equation in the new metric. The second equation determines β.

For example, using the solutions formed from $\xi = \xi(x)$,

and general $\beta(x, y) = \beta_x(x)\beta_y(y)$, we have two separated equations from Eq. (15). One gives

$$\beta_y(y) = \sum_L a_L P_L(y) + b_L Q_L(y), \tag{16}$$

and the other is

$$\frac{d^2\beta_x}{dx^2} - 4\xi \frac{d\xi}{dx}\frac{(x^2-1)}{\xi^2-1}\,d\beta_x + \left(\frac{2(x^2-1)}{\xi^2-1}\frac{d\xi}{dx}\xi - L(L+1)\right)\beta_x = 0. \tag{17}$$

When these equations are applied to the general Weyl solutions, $\xi = ((x+1)^\delta + (x-1)^\delta)/((x+1)^\delta - (x-1)^\delta)$, the slow rotation solution of Tomimatsu–Sato[14] is reproduced.

In conclusion we have presented a one-fixed-point method of generating solutions to the field equations. We have shown that the method is especially adapted to base spaces where $\xi(x)$ is polynomial or logarithmic in the distance coordinate. An equation determining new solutions in the slow rotation limit for general $\xi = \xi(x)$ is derived.

The one-fixed-point method is significant not only because it generates new solutions but also because of the insights it provides about the importance of asymptotic behavior. The matching point of the base and new space-times is conformal infinity. At conformal infinity the transformation is a simple translation of the angular speed.

[1]H. Buchdahl, Quart. J. Math. **5**, 116 (1954).
[2]J. Ehlers, in *Les theories relatavistics de la gravitation* (CNRS, Paris, 1959).
[3]B. K. Harrison, J. Math. Phys. **9**, 1744 (1968).
[4]R. Geroch, J. Math. Phys. **12**, 918 (1971); **13**, 394 (1972).
[5]W. Kinnersley, J. Math. Phys. **18**, 1529 (1977); W. Kinnersley and D. M. Chitre, J. Math. Phys. **18**, 1538 (1977); **19**, 1926, 2037 (1978); Phys. Rev. Lett. **40**, 1608 (1978).
[6]R. O. Hansen, J. Math. Phys. **15**, 46 (1974).
[7]Frederick J. Ernst, Phys. Rev. **167**, 1175 (1968); **168**, 1415 (1968).
[8]J. Irving and N. Mullineux, *Mathematics in Physics and Engineering* (Academic, New York, 1959), p. 461.
[9]R. Geroch, J. Math. Phys. **11**, 2580 (1970).
[10]Abhay Ashtekar and Anne Magnon-Ashtekar, J. Math. Phys. **20**, 793 (1979).
[11]Since $n_a = \nabla_a\lambda = \nabla_a\sqrt{\Omega} = \nabla_a(\lambda - 1)$, this is a reasonable choice close to Λ_0.
[12]B. H. Vorhees, Phys. Rev. D **2**, 2119 (1970).
[13]C. Reina and A. Treves, Gen. Relativ. Gravit. **1**, 817 (1976).
[14]Akira Tomimatsu and Humitaka Sato, Progr. Theoret. Phys. **50**, 95 (1973).

Solutions of Einstein's equations involving arbitrary functions

Peter Szekeres
Department of Mathematical Physics, University of Adelaide, G.P.O. Box 498, Adelaide, South Australia 5001

(Received 8 April 1983; accepted for publication 29 July 1983)

Necessary and sufficient conditions are derived for a solution of Einstein's vacuum equations to depend on an arbitrary function of some scalar function ϕ. Unlike the case of the scalar wave equation the constant surfaces of the function ϕ need not be null. This apparent anomaly is discussed.

PACS numbers: 04.20.Jb, 02.30. + g

1. INTRODUCTION

While many exact solutions of Einstein's equations have been discovered by now, these generally depend on a finite, or at best countably infinite, number of parameters. The only exception known to the author are the plane-fronted waves,[1] which can depend on an arbitrary function of a null coordinate u. Yet it is a feature of hyperbolic differential equations that they often have classes of solutions with arbitrary functions in them. It is the purpose of this paper to investigate this situation in general relativity. We shall find necessary and sufficient conditions for a one-parameter family of metrics to be *generalizable*, by which we mean that the metrics are still solutions of Einstein's equations after the parameter is replaced by an arbitrary function of some given function ϕ.

Section 2 sets the scene. The scalar wave equation in Minkowski space is discussed in detail, conditions being derived for a solution of the form $F(x,f(\phi))$, where f is an arbitrary function of a given function ϕ: a similar treatment of Maxwell's equations is outlined. Two main features are worth noticing in these cases. Firstly ϕ must satisfy the null condition

$$\phi_{,\mu}\phi_{,\nu}g^{\mu\nu}=0,$$

i.e., the surfaces $\phi=$ const are null hypersurfaces. Secondly, no solutions of the form

$$F=F(x,f(\phi,\psi))$$

depending in an arbitrary way on two functions ϕ, ψ can exist.

In Sec. 3 we consider a one-parameter family of metrics $g_{\mu\nu}(x,\lambda)$. The Riemann tensor is computed for its *generalization*, which consists of replacing λ by an arbitrary function $f(\phi)$. Conditions are then derived for the one-parameter family to be *Riemann generalizable*, i.e., in order that the resulting Riemann tensor should be obtained from the original one-parameter family of Riemann tensors by simply the same replacement of λ by $f(\phi)$.

Section 4 is the key section. Here the same procedure is applied to the Ricci tensor, and somewhat surprisingly it is found that the conditions for Ricci generalizability are identical with those for Riemann generalizability, at least in the case of Riemannian or Lorentzian metrics. We now have all the conditions necessary for a one-parameter family of vacuum metrics to be generalizable.

In Sec. 5 we discuss one-parameter families of vacuum metrics which arise by simply applying a one-parameter family of coordinate transformations to a given metric. The example of plane waves shows that this seemingly trivial procedure may lead to nonequivalent generalizations.

Section 6 discusses a particular specialization of the conditions derived in Sec. 4, in complete detail. All exact solutions are found which fall into this subcase, but unfortunately all turn out to depend on the arbitrary function in a trivial way whenever $\phi_{,\mu}$ is non-null.

Finally in Sec. 7 we discuss in greater detail the unexpected conclusion that the generalizing function ϕ is *not* necessarily a null coordinate in the case of general relativity. At first sight this seems to be at variance with what is known of the characteristic surfaces of Einstein's equations. It is shown that while in the case of the scalar wave equation or Maxwell's equations the constant surfaces of the generalizing function ϕ must be characteristic surfaces, this conclusion does not follow in the case of Einstein's equations. The way is therefore open for ϕ to be spacelike or timelike. However nontrivial examples will have to be more complicated than any analyzed in this paper.

2. GENERALIZABLE SOLUTIONS OF THE WAVE EQUATION AND MAXWELL'S EQUATIONS

Suppose we search for solutions of the wave equation

$$\Box\psi\equiv\psi^{,\mu}_{,\mu}=0 \tag{1}$$

of the form

$$\psi=F(x,f(\phi)), \tag{2}$$

where $f(\phi)$ is an arbitrary function (subject to suitable differentiability conditions) of some scalar function ϕ on Minkowski space (the argument x is shorthand for the four arguments x^0, x^1, x^2, x^3).

On setting $f=\lambda=$ const we see that

$$\psi=F(x,\lambda) \tag{3}$$

is a one-parameter family of solutions of the wave equation

$$\Box F=0.$$

We shall say that such a one-parameter family of solutions (3) is *generalizable by a* (nonconstant) *function* ϕ if $\bar{\psi}=F(x,f(\phi))$ is also a solution of the wave equation for arbitrary functions f.

What conditions must $F(x,\lambda)$ satisfy in order for it to be generalizable by ϕ? From (2) we clearly have

$$\psi_{,\mu}=F_{,\mu}(x,f(\phi))+\frac{\partial F}{\partial\lambda}(x,f(\phi))f'(\phi)\phi_{,\mu},$$

where

$$f'(\phi) = \frac{df}{d\phi}, \quad \phi_\mu = \phi_{,\mu}.$$

A more compact way of writing this equation is

$$\bar{\psi}_\mu = \bar{F}_\mu + \bar{K} f'(\phi)\phi_\mu, \tag{4}$$

where

$$F_\mu = \frac{\partial F(x,\lambda)}{\partial x^\mu},$$

$$K(x,\lambda) = \frac{\partial}{\partial \lambda} F(x,\lambda), \tag{5}$$

and placing a bar over a function involving the parameter λ means that all occurrences of λ are replaced by an arbitrary function $f(\phi)$.

Differentiating again and substituting in (1) gives

$$0 = \overline{\Box F} + f'(\phi)(2\bar{K}_\mu \phi^\mu + \bar{K}\Box\phi) + f'^2(\phi)\frac{\overline{\partial K}}{\partial \lambda}\phi_\mu \phi^\mu + f''(\phi)\bar{K}\phi_\mu \phi^\mu. \tag{6}$$

Since $f(\phi)$ is to be arbitrary it is clear that in order for $F(x,\lambda)$ to be generalizable by ϕ the coefficients of $f'(\phi), f'^2(\phi), f''(\phi)$ must vanish separately. This results in the following conditions:

$$F^\mu_\mu = 0, \tag{7}$$

$$\phi_\mu \phi^\mu = 0, \tag{8}$$

$$2K_\mu \phi^\mu + K\phi^\mu_\mu = 0. \tag{9}$$

Equation (7) is of course merely a restatement of the fact that F is a one-parameter family of solutions of the wave equation. Equation (8) says that the ϕ = const surfaces are null hypersurfaces, i.e., characteristic surfaces. This feature is not surprising in view of the fact that discontinuities of solutions of the wave equation (e.g., choosing f'' discontinuous) can only occur across such characteristic surfaces. Equation (9) is a first-order linear partial differential equation for K, which can be solved for any given ϕ. Note that it has been possible to remove bars from the equations arising out of Eq. (6), since for any one-parameter set of equations it is evidently true that

$$H(x,\lambda) = 0 \Leftrightarrow \bar{H} = H(x, f(\phi)) = 0.$$

Some typical examples of generalizable solutions are

(i) $F = F(y,z,\lambda)$ any one-parameter solution of the two-dimensional Laplace equation

$$F_{yy} + F_{zz} = 0$$

is generalizable by the functions $\phi = t \pm x$.

(ii) $F = \lambda r^{-1}$, where $r = (x^2 + y^2 + z^2)^{1/2}$ is generalizable by functions

$$\phi = t \pm r.$$

(iii) $F = u(x,y,z)\lambda$, $\phi = t - v(x,y,z)$. This general case has been solved by Friedlander.[2]

If we had posed the above problem with f an arbitrary function of two independent variables ϕ,ψ we would have obtained from the coefficients of $f_{\phi\phi}$, $f_{\psi\psi}$, and $f_{\phi\psi}$, respectively

$$\phi_\mu \phi^\mu = \psi_\mu \psi^\mu = \phi_\mu \psi^\mu = 0.$$

Since no two linearly independent null vectors may be orthogonal to each other in a Lorentzian metric, this condition is impossible to fulfill. Thus no solution of the wave equation may contain a general function of two variables. In non-Lorentzian metrics this is not true, since

$$\psi_{tt} + \psi_{ww} - \psi_{xx} - \psi_{yy} = 0$$

has solutions

$$\psi = f(t \pm x, w \pm y), \quad f \text{ arbitrary.}$$

A similar treatment for Maxwell's equations

$$F_{\mu\nu,\rho} + F_{\nu\rho,\mu} + F_{\rho\mu,\nu} = 0,$$

$$F_{\mu\nu}{}^{,\nu} = 0,$$

yields that a one-parameter family of solutions $F_{\mu\nu}(x,\lambda)$ is generalizable by a function ϕ if

$$K_{\mu\nu} \equiv \frac{\partial F_{\mu\nu}}{\partial \lambda} = -K_{\nu\mu}$$

satisfies

$$K_{\mu\nu}\phi_\rho + K_{\nu\rho}\phi_\mu + K_{\rho\mu}\phi_\nu = 0$$

and

$$K_{\mu\nu}\phi^\nu = 0.$$

It follows from these equations again that

$$\phi_\rho \phi^\rho = 0$$

and also that

$$K_{\mu\nu} = \phi_{[\mu} q_{\nu]} \equiv \tfrac{1}{2}(\phi_\mu q_\nu - \phi_\nu q_\mu),$$

where q_ν is a spacelike vector satisfying

$$\phi_\mu q^\mu = 0.$$

3. RIEMANN-GENERALIZABLE METRICS

Let $g_{\mu\nu} = g_{\mu\nu}(x,\lambda)$ be a one-parameter family of general-relativistic metrics (i.e., four-dimensional with signature $- + + +$). Again we adopt the convention that if ϕ is a given function on the space time we set

$$\bar{g}_{\mu\nu}(x) = g_{\mu\nu}(x, f(\phi)),$$

where f is an arbitrary function [i.e., $\bar{g}_{\mu\nu}$ actually represents a class of metrics arising from $g_{\mu\nu}(x,\lambda)$ and ϕ].

We set

$$K_{\mu\nu}(x,\lambda) = \frac{1}{2}\frac{\partial}{\partial \lambda} g_{\mu\nu}(x,\lambda) \tag{10}$$

and raise and lower indices by $g_{\mu\alpha}$, $g^{\mu\alpha}$:

$$K^\mu{}_\nu = g^{\mu\alpha}K_{\alpha\nu},$$

$$K^{\mu\nu} = g^{\mu\alpha}g^{\nu\beta}K_{\alpha\beta} = -\frac{1}{2}\frac{\partial}{\partial \lambda} g^{\mu\nu}.$$

Then

$$\frac{\partial}{\partial \lambda}\Gamma^\rho_{\mu\nu} = P^{\rho\alpha}_{\mu\nu;\alpha}, \tag{11}$$

where

$$P^{\rho\alpha}_{\mu\nu} = K^\rho{}_\mu \delta^\alpha_\nu + K^\rho{}_\nu \delta^\alpha_\mu - K_{\mu\nu}g^{\rho\alpha}. \tag{12}$$

Now if $F = F(x,\lambda)$ is any function, then clearly

$$\bar{F}_{,\rho} = \overline{F_{,\rho}} + \frac{\overline{\partial F}}{\partial \lambda} f'(\phi)\phi_\rho, \tag{13}$$

whence

$$\Gamma^{\rho}_{\mu\nu}(\bar{g}) \equiv \tfrac{1}{2}\bar{g}^{\rho\sigma}(\bar{g}_{\sigma\mu,\nu} + \bar{g}_{\sigma\nu,\mu} - \bar{g}_{\mu\nu,\rho}) = \bar{\Gamma}^{\rho}_{\mu\nu} + \bar{P}^{\rho\alpha}_{\mu\nu}\phi_\alpha f'(\phi).$$

Applying (13) again to $\Gamma^{\rho}_{\mu\nu,\rho}(\bar{g})$ and using (11) we obtain the following expression for the generalized Riemann tensor:

$$\begin{aligned} R^{\rho}_{\mu\sigma\nu}(\bar{g}) = \bar{R}^{\rho}_{\mu\sigma\nu}(g) &+ f'(\phi)\left[P^{\rho\alpha}_{\mu\nu;\alpha}\phi_\sigma - P^{\rho\alpha}_{\mu\sigma;\alpha}\phi_\nu + (P^{\rho\alpha}_{\mu\nu}\phi_\alpha)_{;\sigma} - (P^{\rho\alpha}_{\mu\sigma}\phi_\alpha)_{;\nu}\right] \\ &+ f'^2(\phi)\phi_\alpha\phi_\beta\left[P^{\tau\alpha}_{\mu\nu}P^{\rho\beta}_{\tau\sigma} - P^{\tau\alpha}_{\mu\sigma}P^{\rho\beta}_{\tau\nu} + \frac{\partial P^{\rho\alpha}_{\mu\nu}}{\partial\lambda}\delta^{\beta}_{\sigma} - \frac{\partial P^{\rho\alpha}_{\mu\sigma}}{\partial\lambda}\delta^{\beta}_{\nu}\right] \\ &+ f''(\phi)\phi_\alpha\phi_\beta\left[P^{\rho\alpha}_{\mu\nu}\delta^{\beta}_{\sigma} - P^{\rho\alpha}_{\mu\sigma}\delta^{\beta}_{\nu}\right]. \end{aligned} \tag{14}$$

We call the family of metrics *Riemann generalizable* by the function ϕ if

$$R^{\rho}_{\mu\sigma\nu}(\bar{g}) = \bar{R}^{\rho}_{\mu\sigma\nu}(g), \tag{15}$$

i.e., if replacing λ by an arbitrary function $f(\phi)$ implies the Riemann tensor components are obtained simply by replacing all occurrences of λ by $f(\phi)$.

Since f is arbitrary we may set the coefficients of f'', f'^2, and f' separately to zero, and as in Sec. 2, we may remove all bars in these equations. Using Eq. (12) the coefficient of $f''(\phi)$ results in

$$\phi_{[\mu}K_{\rho][\nu}\phi_{\sigma]} = 0, \tag{16}$$

which holds if and only if there exists a vector field $\psi_\nu(x,\lambda)$ such that

$$K_{\mu\nu} = \phi_\mu\psi_\nu + \psi_\mu\phi_\nu. \tag{17}$$

The last two terms in the coefficient of f'^2 are just $\partial/\partial\lambda$ applied to the coefficient of f'', hence they vanish as a consequence of Eq. (16). On the other hand, from Eq. (17) we obtain

$$\phi_\alpha P^{\tau\alpha}_{\mu\nu} = 2\psi^\tau\phi_\mu\phi_\nu, \tag{18}$$

whence the first two terms of the coefficient of f'^2 also vanish. Hence no further information arises by setting the coefficient of f'^2 to zero.

Turning now to the coefficient of $f'(\phi)$, when Eqs. (17) and (18) are substituted in this equation we obtain

$$\phi_\mu L^{\rho}_{[\nu}\phi_{\sigma]} = L_{\mu[\nu}\phi_{\sigma]}\phi^{\rho}, \tag{19}$$

where

$$L_{\mu\nu} = \psi_{\mu;\nu} + \psi_{\nu;\mu} \equiv \pounds_\psi g_{\mu\nu}. \tag{20}$$

That is, $L_{\mu\nu}$ satisfied the same equation as $K_{\mu\nu}$, whence

$$L_{\mu\nu} = \phi_\mu\alpha_\nu + \alpha_\mu\phi_\nu \tag{21}$$

for some vector field

$$\alpha_\mu = \alpha_\mu(x,\lambda).$$

4. RICCI-GENERALIZABLE METRICS

Contracting Eq. (14) over ρ and σ we obtain conditions for a one-parameter family of metrics $g_{\mu\nu}(x,\lambda)$ to be *Ricci generalizable* by a function ϕ, i.e., for

$$R_{\mu\nu}(\bar{g}) = \bar{R}_{\mu\nu}(g). \tag{22}$$

In this case the coefficient of $f''(\phi)$ gives rise to the equation

$$\phi_\beta K^{\beta}_{\ \mu}\phi_\nu + \phi_\beta K^{\beta}_{\ \nu}\phi_\mu - K_{\mu\nu}\phi_\alpha\phi^{\alpha} - K\phi_\mu\phi_\nu = 0, \tag{23}$$

where

$$K = K^{\alpha}_{\ \alpha}.$$

If $\phi_\alpha\phi^{\alpha} \neq 0$, then it follows at once that

$$K_{\mu\nu} = \phi_\mu\psi_\nu + \psi_\mu\phi_\nu \tag{24}$$

for some $\psi_\mu = \psi_\mu(x,\lambda)$.

If $\phi_\alpha\phi^{\alpha} = 0$ however, we can only conclude from Eq. (23) that ϕ_μ is a (null) eigenvector of $K_{\mu\nu}$,

$$\phi_\rho K^{\rho}_{\ \mu} = \tfrac{1}{2}K\phi_\mu. \tag{25}$$

Again the last two terms of the coefficient of f'^2 vanish on taking $\partial/\partial\lambda$ of the f'' equation, while the first two terms give rise to the equation

$$\phi_\alpha\phi_\beta(P^{\tau\alpha}_{\mu\nu}P^{\rho\beta}_{\tau\rho} - P^{\tau\alpha}_{\mu\rho}P^{\rho\beta}_{\tau\nu}) = 0.$$

In the case $\phi_\alpha\phi^{\alpha} \neq 0$, substitution of (24) immediately guarantees that this equation is fulfilled, but in the case $\phi_\alpha\phi^{\alpha} = 0$ one obtains from (25) the condition

$$(\tfrac{1}{2}K^2 - K_{\alpha\beta}K^{\alpha\beta})\phi_\mu\phi_\nu = 0. \tag{26}$$

If we assume the metric is Lorentzian (signature $n-2$, where n = dimension of space) then (24) again follows from (25) and (26). To see this, let σ_ν be a second null vector normalized such that

$$\sigma_\nu\sigma^\nu = \phi_\nu\phi^{\nu} = 0, \quad \sigma_\nu\phi^{\nu} = 1,$$

and let $e^{\mu}_i (i = 2, \ldots, n-1)$ be an orthonormal basis of the tangent subspace orthogonal to ϕ_ν and σ_ν, i.e.,

$$\sigma_\mu e^{\mu}_i = \phi_\mu e^{\mu}_i = 0, \quad e_{i\mu}e^{\mu}_j = \delta_{ij}.$$

It is clear that $\phi_\mu, \rho_\mu, e^{\mu}_i$ form a basis of the tangent space and $K_{\mu\nu}$ may be expanded in this basis

$$\begin{aligned} K_{\mu\nu} = K_{00}\phi_\mu\phi_\nu &+ K_{01}(\phi_\mu\sigma_\nu + \sigma_\mu\phi_\nu) + K_{11}\sigma_\mu\sigma_\nu \\ &+ \sum_i K_{0i}(\phi_\mu e_{i\nu} + e_{i\mu}\phi_\nu) + K_{1i}(\sigma_\mu e_{i\nu} + e_{i\mu}\sigma_\nu) \\ &+ \sum_{i,j} K_{ij}e_{i\mu}e_{j\nu}. \end{aligned}$$

Equation (25) then gives

$$K_{11} = K_{1i} = 0, \quad \sum_i K_{ii} = 0,$$

while (26) implies

$$\sum_{i,j} K^2_{ij} = 0.$$

Hence $K_{ij} = 0$ for all i,j and Eq. (24) holds with

$$\psi_\mu = \frac{1}{2}K_{00}\phi_\mu + K_{01}\sigma_\mu + \sum_i K_{0i}e_{i\mu}.$$

It is easy to convince oneself that if the signature were other than Lorentzian (or Riemannian) the conclusion (24) would not in general be justified. In general relativity ($n = 4$, Lorentzian) we may however adopt Eq. (24) as being equivalent to the f'' and f'^2 equations.

There remains the equation arising from setting the coefficient of $f'(\phi)$ equal to zero,

$$P^{\rho\alpha}_{\mu\nu;\alpha}\phi_\rho - P^{\rho\alpha}_{\mu\rho;\alpha}\phi_\nu + (P^{\rho\alpha}_{\mu\nu}\phi_\alpha)_{;\rho} - (P^{\rho\alpha}_{\mu\rho}\phi_\alpha)_{;\nu} = 0.$$

On substituting Eqs. (12) and (24) we obtain

$$\phi_\rho\,\phi^{\,\rho}L_{\mu\nu}-\phi_\nu L_{\mu\rho}\phi^{\,\rho}-\phi_\mu L_{\nu\rho}\phi^{\,\rho}+L_\rho^{\,\rho}\phi_\mu\phi_\nu=0\,,$$

where $L_{\mu\nu}$ is given by Eq. (20). That is, $L_{\mu\nu}$ satisfies the same equation as $K_{\mu\nu}$ [Eq. (23)] and again we may conclude that $L_{\mu\nu}$ has the form given in (21). In summary then

Theorem 1: A one-parameter family of Riemannian or Lorentzian metrics $g_{\mu\nu}(x,\lambda)$ is Ricci generalizable by a function ϕ if and only if there exist vector fields $\psi_\mu(x,\lambda)$ and $\rho_\mu(x,\lambda)$ such that

$$K_{\mu\nu}\equiv\frac{1}{2}\frac{\partial}{\partial\lambda}g_{\mu\nu}=\phi_\mu\psi_\nu+\psi_\mu\phi_\nu \tag{27}$$

and

$$L_{\mu\nu}\equiv\psi_{\mu;\nu}+\psi_{\nu;\mu}=\phi_\mu\rho_\nu+\rho_\mu\phi_\nu. \tag{28}$$

Since these conditions are exactly the same as those obtained in Sec. 3 we also have the following result:

Theorem 2: A one-parameter family of Riemannian or Lorentzian metrics $g_{\mu\nu}(x,\lambda)$ is Ricci generalizable by a function ϕ if and only if it is Riemann generalizable.

Perhaps the most surprising aspect of these theorems is that ϕ_μ is not necessarily a null vector. We shall return to this point later. To conclude this section we just wish to remark that as in the case of the scalar wave equation, no metrics are Ricci generalizable by arbitrary functions of two independent variables $f(\phi,\psi)$ (with ϕ_μ, ψ_μ linearly independent vector fields).

This follows by setting up the equations (22) and setting to zero all coefficients of f_ϕ, f_ψ, etc. The coefficients of $f_{\phi\phi}, f_{\psi\psi}$, $f_{\phi\psi}$ yield, respectively

$$\phi_\mu\alpha_\nu+\alpha_\mu\phi_\nu-K\phi_\mu\phi_\nu=K_{\mu\nu}\phi_\alpha\phi^{\,\alpha}, \tag{29}$$

$$\psi_\mu\beta_\nu+\beta_\mu\psi_\nu-K\psi_\mu\psi_\nu=K_{\mu\nu}\psi_\alpha\psi^\alpha, \tag{30}$$

$$\psi_\mu\alpha_\nu+\alpha_\mu\psi_\nu+\phi_\mu\beta_\nu+\beta_\mu\phi_\nu-K(\phi_\mu\psi_\nu+\psi_\mu\phi_\nu)=2K_{\mu\nu}\phi_\alpha\psi^\alpha, \tag{31}$$

where

$$\alpha_\mu=K_{\mu\rho}\,\phi^{\,\rho},\quad \beta_\mu=K_{\mu\rho}\psi^\rho.$$

By taking suitable linear combinations of ϕ and ψ it is easy to see that for Riemannian or Lorentzian metrics there is no loss of generality in assuming that

$$\phi_\alpha\phi^{\,\alpha}\neq0,\quad \psi_\alpha\psi^\alpha\neq0.$$

Equations (29) and (30) give at once that α_μ and β_μ must be linear combinations of ϕ_μ and ψ_μ

$$\alpha_\mu=a\phi_\mu+b\psi_\mu,\quad \beta_\mu=c\phi_\mu+d\psi_\mu$$

and substituting back in (29) and (30) results in

$$K_{\mu\nu}=B(\phi_\mu\psi_\nu+\psi_\mu\phi_\nu),$$

where

$$b=B\phi_\alpha\phi^{\,\alpha},\ a=d=B\phi_\alpha\psi^\alpha,\ c=B\psi_\alpha\psi^\alpha.$$

Finally substituting into (31) results in $B=0$, i.e.,

$$K_{\mu\nu}=0,$$

which proves the desired result.

The most interesting case arises in general relativity when $R_{\mu\nu}(g)=0$ leads to $\bar{R}_{\mu\nu}(\bar{g})=0$; we call this situation *vacuum generalizable*. Theorems 1 and 2 clearly apply to this case.

5. COORDINATE TRANSFORMATIONS

A particularly simple way of generating a one-parameter family of vacuum metrics is to apply a one-parameter family of coordinate transformations

$$y^\mu=y^\mu(x,\lambda),$$

with inverse

$$x^\mu=x^\mu(y,\lambda)$$

to a given parametrized vacuum metric $\mathring{g}_{\mu\nu}(x,\lambda,\mu,\ldots)$($\mathring{g}_{\mu\nu}$ may of course depend on no parameters at all). This new family

$$g_{\mu\nu}(y,\lambda,\mu,\ldots)=\mathring{g}_{\alpha\beta}(x(y,\lambda),\lambda,\mu,\ldots)\frac{\partial x^\alpha}{\partial y^\mu}(y,\lambda)\frac{\partial x^\beta}{\partial y^\nu}(y,\lambda) \tag{32}$$

is of course geometrically equivalent to the original family $\mathring{g}_{\mu\nu}$, but if it is vacuum generalizable there is no guarantee that its generalizations by a function ϕ are so equivalent.

Unfortunately Theorem 2 implies that the simplest possible procedure, namely, to apply a one-parameter coordinate transformation to flat (Minkowski) space does not lead to anything new. For in this case the Riemann tensor vanishes for the entire family, and hence all its generalizations also have vanishing Riemann tensor and are therefore flat. However, for curved vacuum metrics the procedure may result in new metrics.

Differentiating (32) with respect to λ gives the transformation of $K_{\mu\nu}$

$$K_{\mu\nu}(y,\lambda)=\mathring{K}_{\alpha\beta}(x(y,\lambda),\lambda)\frac{\partial x^\alpha}{\partial y^\mu}\frac{\partial x^\beta}{\partial y^\nu}+\xi_{(\mu;\nu)}\,, \tag{33}$$

where

$$\xi^{\,\beta}(y,\lambda)=-\frac{\partial y^\beta}{\partial\lambda}=\frac{\partial x^\alpha}{\partial\lambda}\frac{\partial y^\beta}{\partial x^\alpha},$$

and ";" refers to covariant derivative with respect to $g_{\mu\nu}$ (all other parameters μ, ... have been suppressed). A similar analysis for any tensor $\mathring{T}^{\mu\nu\ldots}_{\rho\sigma\ldots}(x,\lambda)$ yields the general result

$$\frac{\partial}{\partial\lambda}T^{\mu\nu\ldots}_{\rho\sigma\ldots}(y,\lambda)=\left(\frac{\partial}{\partial\lambda}\mathring{T}^{\alpha\beta\ldots}_{\gamma\delta\ldots}\right)\frac{\partial y^\mu}{\partial x^\alpha}\cdots\frac{\partial x^\gamma}{\partial y^\rho}\cdots+\pounds_\xi T^{\alpha\beta\ldots}_{\gamma\delta\ldots},$$

where

$$T^{\mu\nu\ldots}_{\rho\sigma\ldots}(y,\lambda)=\mathring{T}^{\alpha\beta\ldots}_{\gamma\delta\ldots}(x(y,\lambda),\lambda)\frac{\partial y^\mu}{\partial x^\alpha}\cdots\frac{\partial x^\gamma}{\partial y^\rho}\cdots$$

Examples

(1) Suppose $g_{\mu\nu}(y,\lambda)$ is a one-parameter family of vacuum metrics, generalizable by a function ϕ, i.e., suppose Eqs.(27) and (28) hold. If $\rho_\mu=0$, then

$$\psi_{\mu;\nu}+\psi_{\nu;\mu}=0,$$

i.e., ψ_μ is a Killing vector and from (27) we have

$$K_{\mu\nu}=\xi_{(\mu;\nu)},$$

where

$$\xi_\mu=2\phi\psi_\mu.$$

Hence $g_{\mu\nu}(y,\lambda)$ is obtained by applying a coordinate transformation to $\mathring{g}_{\mu\nu}(x)=g_{\mu\nu}(x,0)$, given by

$$\frac{\partial y^\mu}{\partial\lambda}=-\xi^{\,\mu}=-2\phi(y)\psi^\mu(y,\lambda)$$

subject to

$$y^\mu(x,0) = x^\mu.$$

(2) Plane-fronted waves[1]:

$$ds^2 = -2Hdu^2 + 2dudv + dx^2 + dy^2,$$

where

$$H = H(u,x,y)$$

is subject to

$$H_{,22} + H_{,33} = 0 \tag{34}$$

($x^0 = u$, $x^1 = v$, $x^2 = x$, $x^3 = y$). All such metrics satisfy Einstein's vacuum equations. Thus, it is clear that $H = H(\lambda,x,y)$ gives rise to a one-parameter family of metrics which are vacuum generalizable by $\phi = u$. All conditions of Theorem 1 are satisfied since

$$K_{\mu\nu} = -\frac{\partial H}{\partial \lambda} u_{,\mu} u_{,\nu},$$

whence Eq. (27) holds with

$$\psi_\mu = -\frac{1}{2}\frac{\partial H}{\partial \lambda} u_{,\mu}$$

and (28) follows with

$$\rho_\mu = -\frac{\partial^2 H}{\partial x^\mu \partial \lambda}.$$

Theorem 2 is easily verified by computing the Riemann tensor, whose only surviving components are

$$R_{OAOB} = H_{,AB} (A,B = 2,3).$$

Clearly λ is replaced by $f(u)$ in these components, hence these solutions are all Riemann generalizable. Flat space occurs if H is linear in x and y. Plane waves arise if H is quadratic in x and y, and may be regarded as generalizations by $\phi = u$ of the two-parameter metric having

$$H = \lambda_1(x^2 - y^2) - 2\lambda_2 xy.$$

If we set $\lambda_1 = \cos 2\lambda$, $\lambda_2 = \sin 2\lambda$, then

$$K_{\mu\nu} = \xi_{(\mu,\nu)},$$

where $\xi_\mu = (0,0,-y,x)$. In this case the one-parameter family is obtained from

$$dS^2 = -2(X^2 - Y^2)du^2 + 2\,du\,dv + dX^2 + dY^2$$

by performing the one-parameter family of coordinate transformations

$$x = X\cos\lambda + Y\sin\lambda$$

$$y = -X\sin\lambda + Y\cos\lambda.$$

However the generalizations obtained by setting $\lambda = f(u)$ are not in general equivalent to $\mathring{g}_{\mu\nu}$, since the generalizations are plane waves with variable phase and amplitude while $\mathring{g}_{\mu\nu}$ has constant phase and amplitude. So here is an example where the original one-parameter family of metrics are all equivalent but their generalizations result in genuinely inequivalent metrics.

6. A SPECIAL CASE

It turns out that if we set ψ_μ proportional to ϕ_μ, the generalizable metrics can all be written down explicitly. The analysis is rather long and we will only outline the main steps here.

If ϕ_μ is a null vector field, then Eq. (28) immediately shows that it is shear-free and twist-free. All such metrics have been discovered by Kundt[3] and need not be discussed further here.

If ϕ_μ is non-null, let us postulate it to be timelike (the spacelike case is similarly analyzed) and set $x^0 = \phi$. From Eq. (27) it is easily shown that coordinates $x^i (i = 1,2,3)$ may be found such that

$$ds^2 = -N^2(x,\lambda)(dx^0)^2 + g_{ij}(x)dx^i\,dx_j.$$

Since $\psi_\mu = f(x,\lambda)\phi_\mu$, Eq. (28) shows that

$$0 = \phi_{(i;j)} = -\Gamma^0_{ij} = -\tfrac{1}{2}N^{-2}g_{ij,0},$$

whence $g_{ij} = g_{ij}(x^1, x^2, x^3)$. Now the spatial part of the Einstein field equations gives

$$R^{(3)}_{ij} = N^{-1}N_{|ij},$$

where $|$ refers to covariant derivative with respect to the three-metric g_{ij}.

Since $R^{(3)}_{ij}$, being constructed from the three-metric g_{ij}, has no λ dependence it can be verified that

$$\xi_i = N^2\left(N^{-1}\frac{\partial N}{\partial \lambda}\right)_{,i}$$

satisfies

$$\xi_{(i,j)} = 0$$

and is therefore a hypersurface-orthogonal Killing three-vector. We can therefore cast the metric in a Weyl form

$$ds^2 = -e^{2\alpha + 2\mu}(dx^0)^2 + e^{2\mu}(dx^1)^2 + e^{2\nu}\,d\xi\,d\bar{\xi}, \tag{35}$$

where

$$\xi = x^2 + ix^3, \quad \bar{\xi} = x^2 - ix^3,$$

$$\mu = \mu(\xi,\bar{\xi}), \quad \gamma = \gamma(\xi,\bar{\xi}), \quad \alpha = \alpha(x^0, x^1, \xi, \bar{\xi}).$$

The Einstein equations may now be written out in full:

$$\alpha_{1\xi} + \alpha_1\alpha_\xi = 0, \tag{36}$$

$$\alpha_{11} + \alpha_1^2 + e^{2\mu - 2\nu}(4\mu_{\xi\bar{\xi}} + 8\mu_\xi\mu_{\bar{\xi}} + 2\alpha_\xi\mu_{\bar{\xi}} + 2\alpha_{\bar{\xi}}\mu_\xi) = 0, \tag{37}$$

$$\alpha_{\xi\bar{\xi}} + \alpha_\xi\alpha_{\bar{\xi}} + \alpha_\xi\mu_{\bar{\xi}} + \alpha_{\bar{\xi}}\mu_\xi = 0, \tag{38}$$

$$\alpha_{\xi\xi} + \alpha_\xi^2 + 2\alpha_\xi(\mu_\xi - \nu_\xi) + 2\mu_{\xi\xi} + 2\mu_\xi^2 - 4\nu_\xi\mu_\xi = 0, \tag{39}$$

and

$$\nu_{\xi\bar{\xi}} + \mu_{\xi\bar{\xi}} + \mu_\xi\mu_{\bar{\xi}} = 0. \tag{40}$$

Now Eq. (36) implies that

$$e^\alpha = G(x^0, x^1) + F(x^0,\xi,\bar{\xi}). \tag{41}$$

Case (i): $\alpha_1 = G_1 \equiv \dfrac{\partial G}{\partial x^1} \neq 0.$

Substitution of (41) into (39) gives

$$\mu_{\xi\xi} + \mu_\xi^2 - 2\nu_\xi\mu_\xi = 0.$$

Together with the results of differentiating Eqs. (38) and (39) with respect to ξ and $\bar{\xi}$, respectively, this results in

$$2\nu_{\xi\bar{\xi}} - \mu_{\xi\bar{\xi}} - \mu_\xi\mu_{\bar{\xi}} = 0,$$

which is clearly only consistent with Eq. (40) if $\nu_{\xi\bar{\xi}} = 0$. Now

at this point a computation of the Riemann tensor components would show them all to vanish, so there is no need to proceed further with this case as it can only result in flat space.

Case (ii): $\alpha_1 = 0$.

In this case, Eqs. (37) and (38) result in

$$(e^{\alpha+2\mu})_{\zeta\bar{\zeta}} = 0,$$

whence

$$e^{\alpha} = e^{-2\mu}[f(x^0,\zeta) + \bar{f}(x^0,\bar{\zeta})]$$

and

$$2\mu_{\zeta\bar{\zeta}}(f+\bar{f}) + \mu_{\bar{\zeta}} f_{\zeta} + \mu_{\zeta}\bar{f}_{\bar{\zeta}} = 0. \tag{42}$$

Let $\phi = f(0,\zeta)$ and set

$$K(x^0,\zeta,\bar{\zeta}) = (f+\bar{f})/(\phi+\bar{\phi}). \tag{43}$$

We are clearly only interested in the case $K \neq$ const, else there is no x^0 dependence in the metric at all. Equation (42) is clearly equivalent to the pair of equations

$$2\mu_{\zeta\bar{\zeta}}(\phi+\bar{\phi}) + \mu_{\bar{\zeta}}\phi_{\zeta} + \mu_{\zeta}\bar{\phi}_{\bar{\zeta}} = 0 \tag{44}$$

and

$$K_{\zeta}\mu_{\bar{\zeta}} + K_{\bar{\zeta}}\mu_{\zeta} = 0, \tag{45}$$

while Eq. (39) implies

$$2(\phi+\bar{\phi})\mu_{\zeta} - 2(\nu_{\zeta}+\mu_{\zeta})\phi_{\zeta} + \phi_{\zeta\zeta} = 0 \tag{46}$$

and

$$(e^{-2\tau}K_{\zeta})_{\zeta} = 0, \tag{47}$$

where

$$\tau = \nu + \mu - \ln(\phi+\bar{\phi}).$$

Hence

$$K_{\zeta} = \bar{p}(x^0,\bar{\zeta})e^{2\tau}, \quad K_{\bar{\zeta}} = p(x^0,\zeta)e^{2\tau}$$

for some function $p(x^0,\zeta)$. From Eq. (45) one sees that $p = P(x^0)b(\zeta)$ for some real function P and complex function $b(\zeta)$. Defining a complex function $Z(\zeta)$ by

$$\frac{dZ}{d\zeta} = -\frac{i}{b(\zeta)}$$

we see that $\mu = \mu(x)$ where we have set $Z = x + iy$. By changing the complex coordinates $\zeta,\bar{\zeta}$ to $Z,\bar{Z}$ it is now a relatively straightforward matter to integrate Eqs. (43)–(47). Apart from some removable arbitrary constants, there are three distinct cases, arising from

$$(\mu')^{-1} = x, \ \sin x, \ \text{or} \ \sinh x.$$

The first case leads only to flat space whilst the second results in the metric

$$ds^2 = (1+\cos x)^2\{-[A(x^0)e^{y} + B(x^0)e^{-y}]^2(dx^0)^2 + dx^2 + dy^2\} + \frac{(1-\cos x)}{(1+\cos x)}(dx')^2.$$

At first sight it appears that one has here a vacuum metric of the desired kind, exhibiting arbitrary functions A, B of a timelike coordinate x^0. However, further inspection reveals that the two-metric

$$-[A(x^0)e^{y} + B(x^0)e^{-y}]^2(dx^0)^2 + dy^2$$

has constant curvature (curvature scalar = 1). As any such two-metric can be brought to a canonical form, it is clear that coordinate transformations exist which eliminate the arbitrary functions. The resulting space time is then a very special Weyl static axisymmetric solution. The case $\mu' = (\sinh x)^{-1}$ is similar.

7. RELATION TO CHARACTERISTICS AND CONCLUSIONS

The structure of solutions to Einstein's equations involving arbitrary functions has been shown to possess remarkably simple properties. Perhaps the most surprising aspect of the results in Theorems 1 and 2 is that the gradient of the function ϕ need not be a null vector, as is the case for the scalar wave equation or Maxwell's equations. This is especially surprising in view of the fact that characteristic surfaces of Einstein's equations (i.e., surface ϕ = const across which the curvature tensor has a discontinuity) are known to be null surfaces.[4,5] Since such discontinuities can apparently be generated by setting the arbitrary function f to be discontinuous, it seems at first sight paradoxical that this conclusion does not follow. A detailed look at Pirani's treatment of characteristics reveals the reason for this apparent discrepancy.

In the case of the scalar wave equation, a discontinuity in $f''(\phi)$ gives rise, via Eq. (6), to the usual characteristic condition (8). Similarly in Maxwell's equations a discontinuity in $f'(\phi)$ gives rise to the characteristic conditions

$$\Delta F_{\mu\nu} = \phi_{[\mu}q_{\nu]}, \ \phi_{\mu}\phi^{\mu} = 0, \ \phi_{\mu}q^{\mu} = 0$$

obtained on replacing $K_{\mu\nu}$ by $\Delta F_{\mu\nu}$.

For Einstein's equations it is not strictly allowable to set $f(\phi)$ discontinuous since this violates the usual C^2-differentiability conditions. If one sets $f''(\phi)$ discontinuous, this does indeed result in the discontinuity of the Riemann tensor

$$\Delta R_{\mu\nu\rho\sigma} = 2\phi_{[\mu}K_{\rho][\nu}\phi_{\sigma]}\Delta f''(\phi),$$

while the condition $\Delta R_{\mu\nu} = 0$ results in

$$0 = \Delta R_{\mu\rho} = (K_{\mu\rho}\phi_{\nu}\phi^{\nu} - \phi_{\mu}K_{\rho\nu}\phi^{\nu} - \phi_{\rho}K_{\mu\nu}\phi^{\nu} + K\phi_{\mu}\phi_{\rho})\Delta f''(\phi),$$

which implies in turn

$$\Delta R_{\rho\mu\sigma\nu}\phi_{\alpha}\phi^{\alpha} = 0.$$

At this point it is argued that if $\Delta R_{\rho\mu\sigma\nu} \neq 0$ (i.e., the surface ϕ = const is a characteristic), then $\phi_{\alpha}\phi^{\alpha} = 0$. However, our discussion has shown that it is precisely the case $\Delta R_{\rho\mu\sigma\nu} = 0$ which occurs here since $R_{\mu\nu\rho\sigma}$ depends only on $f(\phi)$ and not its derivatives. While it is tempting on this account to permit $f(\phi)$ to be discontinuous there appears to be no guarantee that this violation of Lichnerowicz conditions can be undone by a (discontinuous) coordinate transformation. It should furthermore be recognized that these continuity conditions are an integral part of the proof of the nullness of characteristic surfaces.

Nevertheless one's feeling that this problem is intimately connected with characteristics goes deep, and this paper only goes a little way to disturbing it. The detailed example given in Sec. 6, at first looks most promising in its goal of

obtaining a family of solutions depending in an arbitrary way on a non-null function, only to dissolve in the final analysis through a series of coordinate transformations. Whether the same phenomenon would occur in general, without the restrictive ansatz $\psi_\mu \propto \phi_\mu$, is impossible to say at this point. In conclusion, it is perhaps worth pointing out the existence[6] of a family of solutions of the Einstein dust equations, $G_{\mu\nu} = \rho u_\mu u_\nu$, which have in them arbitrary functions of a spacelike coordinate. It is not inconceivable that interior solutions like this, which permit a certain amount of arbitrary variation of the matter distribution, could not lead to exterior solutions similarly dependent on an arbitrary function.

ACKNOWLEDGMENTS

The author would like to express his appreciation to Dr. J. Ehlers and Dr. R. Penrose for some useful discussions.

[1] J. Ehlers and W. Kundt, in *Gravitation, an Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962).
[2] F. G. Friedlander, Proc. Cambridge Philos. Soc. **43**, 360 (1947).
[3] W. Kundt, Z. Phys. **163**, 77 (1961).
[4] J. L. Synge, *Relativity: the General Theory* (North-Holland, Amsterdam, 1960).
[5] F. A. E. Pirani, in *Brandeis Lectures on General Relativity*, edited by S. Deser and K. W. Ford (Prentice–Hall, Englewood Cliffs, NJ, 1964).
[6] P. Szekeres, Comm. Math. Phys. **41**, 55 (1975).

The Cauchy problem for the $R + R^2$ theories of gravity without torsion

P. Teyssandier and Ph. Tourrenc
Laboratoire de Physique Théorique, E.R.A. 533 associée au CNRS Institut Henri Poincaré, 75005 Paris, France

(Received 20 July 1982; accepted for publication 18 November 1982)

The exterior Cauchy problem is discussed for the fourth-order theories of gravity derived from the Lagrangian densities $\mathscr{L} = \sqrt{-g}\,(R + \frac{1}{2}aR^2 + bR_{\mu\nu}R^{\mu\nu}) - \kappa\mathscr{L}_m$. When $b \neq 0$, the Cauchy problem can be solved by the standard method already used in general relativity. When $b = 0$, the problem cannot be formulated as in the case where $b \neq 0$, since the corresponding fourth-order theory is shown to be equivalent to a second-order scalar–tensor theory. This scalar–tensor theory is proved to coincide with one of the models of gravity proposed by O'Hanlon in order to present a covariant version of the massive dilaton theory suggested by Fujii. This result is generalized: The models of O'Hanlon are shown to be indistinguishable from the fourth-order theories derived from the Lagrangian densities $\mathscr{L} = \sqrt{-g}\,F(R) - \kappa\mathscr{L}_m$, where F is any real function such that $F''(R)$ does not identically vanish.

PACS numbers: 04.50. + h

I. INTRODUCTION

In recent years, the quantization of the gravitational field has given rise to much interest in the fourth-order theories of gravity derived from the Lagrangian densities[1,2]

$$\mathscr{L} = \sqrt{-g}(R + \tfrac{1}{2}aR^2 + bR_{\mu\nu}R^{\mu\nu}) - \kappa\mathscr{L}_m, \tag{1}$$

where R denotes the scalar curvature, $R_{\mu\nu}$ the Ricci tensor, $\mathscr{L}_m$ the matter Lagrangian, κ a coupling constant similar to the Einstein constant, and a, b two parameters with dimension of a squared length. Moreover, the effects on cosmological solutions of adding quadratic terms in the curvature tensor to the usual Einstein–Hilbert action have been considered by a number of authors.[3]

The purpose of the present work is to study the exterior Cauchy problem for the two-parameter family of field equations derived from (1). Stelle[4] has already touched on this problem for the linearized equations by using the de Rham transverse-traceless decomposition of the metric. Although this procedure presents the advantage of exhibiting the various helicity components of the gravitational potentials, we employ in our investigations the classical method developed by Lichnerowicz[5] and others for general relativity. Indeed this method brings very fruitful information about the intrinsic structure of the theories examined here, particularly in the case where $b = 0$.[6]

Section II is concerned with the exterior Cauchy problem in the case where $b \neq 0$.

The other sections are devoted to the field equations obtained when $b = 0$. For these equations, the Cauchy problem cannot be solved if one formulates it as in Sec. II. To remove this difficulty, we are led to construct a second-order scalar–tensor model equivalent to the fourth-order equations (see Sec. III). Using the new system of equations, we reformulate and solve the Cauchy problem in Sec. IV. In Sec. V the scalar–tensor model constructed in Sec. III is compared with one of the models of gravity proposed by O'Hanlon[7] in order to get a covariant formulation of the massive dilaton theory suggested by Fujii.[8] It is then proved that any model of O'Hanlon is equivalent to a theory of gravity involving fourth-order derivative terms. Section VI contains our conclusions.

II. THE EXTERIOR CAUCHY PROBLEM FOR $b \neq 0$

In this work, we consider a Lorentzian space–time V_4 in which the connection is the Christoffel connection formed from the metric tensor $g_{\mu\nu}$. The variation of the Lagrangian density (1) with respect to the $g^{\mu\nu}$ yields the fourth-order field equations[1,3,4]

$$G_{\mu\nu} = \kappa T_{\mu\nu} \quad (\mu,\nu = 0,1,2,3), \tag{2}$$

where

$$\begin{aligned} G_{\mu\nu} = {} & bR_{\mu\nu;\lambda}{}^{;\lambda} + (a + \tfrac{1}{2}b)g_{\mu\nu}\Box R \\ & - (a + b)R_{;\mu;\nu} + aR(R_{\mu\nu} - \tfrac{1}{4}Rg_{\mu\nu}) \\ & + 2b(R_{\mu\rho\nu\sigma} - \tfrac{1}{4}g_{\mu\nu}R_{\rho\sigma})R^{\rho\sigma} + R_{\mu\nu} - \tfrac{1}{2}Rg_{\mu\nu}, \end{aligned} \tag{3}$$

$$T_{\mu\nu} = \frac{1}{\sqrt{-g}}\frac{\delta\mathscr{L}_m}{\delta g^{\mu\nu}}.$$

The Ricci tensor $R_{\mu\nu}$ and the scalar curvature R are defined by

$$R_{\mu\nu} = R^{\lambda}{}_{\mu\lambda\nu}, \quad R = g^{\mu\nu}R_{\mu\nu},$$

$$R^{\alpha}{}_{\mu\beta\nu} = \Gamma^{\alpha}_{\mu\nu,\beta} - \Gamma^{\alpha}_{\mu\beta,\nu} + \Gamma^{\lambda}_{\mu\nu}\Gamma^{\alpha}_{\lambda\beta} - \Gamma^{\lambda}_{\mu\beta}\Gamma^{\alpha}_{\lambda\nu}.$$

We use the signature $(+,-,-,-)$ for the metric tensor. The $\Gamma^{\lambda}_{\mu\nu}$ are the Christoffel symbols of the second kind formed from the metric $g_{\mu\nu}$. The symbol $(\ )_{,\alpha}$ denotes the ordinary derivative and $(\ )_{;\alpha}$ the covariant derivative: $\Box$ is the d'Alembertian operator acting on a scalar function.

Contracting the field equations (2), one finds

$$(3a + 2b)\Box R - R = \kappa T, \tag{4}$$

where T is the trace of the energy–momentum tensor $T_{\mu\nu}$.

The geometrical tensor $G_{\mu\nu}$ satisfies the conservation identities

$$G^{\mu}{}_{\nu;\mu} = 0 \tag{5}$$

so that $T_{\mu\nu}$ satisfies the conservation law $T^{\mu}{}_{\nu;\mu} = 0$ when the field equations are verified.

In our approach, we suppose that the space–time is a

 0022-2488/83/122793-07$02.50

differentiable manifold V_4 of class C^4 and piecewise of class C^6. Moreover, the metric components $g_{\mu\nu}$ are assumed to be of class C^3 and piecewise continuous of class C^5. Our aim is to study the Cauchy initial value problem in an empty space–time. Then, the field equations are

$$G_{\mu\nu} = 0. \tag{6}$$

Let Σ be a hypersurface oriented everywhere in space and described, at least locally, by $x^0 = 0$. The Cauchy data consist of the values on Σ of the metric components $g_{\mu\nu}$ and of their partial derivatives $g_{\mu\nu,0}$, $g_{\mu\nu,00}$, and $g_{\mu\nu,000}$. The spacelike orientation of Σ is equivalent to the condition $g^{00} > 0$ on Σ. The index of any partial derivative of a potential $g_{\mu\nu}$ is defined as the number of times this potential is differentiated with respect to x^0 [5]. The Cauchy data of index $n \leqslant 3$ are assumed to be of class C^{5-n}.

The fourth-order derivatives of the potentials of index $\leqslant 3$ can be directly calculated on Σ from the Cauchy data by differentiation. They are continuous. In order to examine the behavior of the derivatives of index 4, let us write the field equations (6) so that these derivatives appear explicitly. If we define $\overset{*}{g}{}^{ij}$ and A by

$$\overset{*}{g}{}^{ij} = g^{ij} - g^{0i}g^{0j}/g^{00} \quad (i,j,=1,2,3), \tag{7}$$

$$A = \overset{*}{g}{}^{ij} g_{ij,0000}, \tag{8}$$

we get

$$G_{ij} \equiv -\tfrac{1}{2}(g^{00})^2 [\, b g_{ij,0000} + (2a+b) A g_{ij}\,] + F_{ij}(\text{d.C.}) = 0, \tag{9}$$

$$G_{i0} \equiv \tfrac{1}{2} g^{00} [\, b g^{oj} g_{ij,0000} - (2a+b) A g^{00} g_{i0}\,] + F_{i0}(\text{d.C.}) = 0, \tag{10}$$

$$G_{00} \equiv -\tfrac{1}{2} g^{00} \{ b g^{ij} g_{ij,0000} + [(2a+b) g_{00} g^{00} - 2(a+b)] A \} + F_{00}(\text{d.C.}) = 0, \tag{11}$$

where $F_{\mu\nu}(\text{d.C.})$ denote quantities which can be determined on Σ from the Cauchy data by algebraic calculations and by differentiations.

There is no derivative $g_{\mu 0,0000}$ in the above equations. This fact is not surprising, since there exist coordinate transformations of class C^4 and piecewise continuous of class C^6 which alter neither the Cauchy data nor the derivatives $g_{ij,0000}$, but reduce the $g_{\mu 0,0000}$ to any specified values (see Appendix A). This implies in particular that the discontinuities in the derivatives $g_{\mu 0,0000}$ have no physical meaning and do not play an essential role in the problem.

Given the quantities G_{ij} and $G_{0\alpha}$, it is possible to determine the G^0_α. Conversely, if $g^{00} \neq 0$, one can derive the quantities $G_{0\alpha}$ from G_{ij} and G^0_α by using the relations

$$g^{00} G_{i0} = G^0_i - g^{0j} G_{ij},$$
$$(g^{00})^2 G_{00} = g^{00} G^0_0 - g^{0i} G^0_i + g^{0i} g^{0j} G_{ij}.$$

It follows from these relations that if $g^{00} \neq 0$, the system of Eqs. (9) and (10)–(11) is equivalent to the system formed by Eqs. (9) and by the four equations

$$G^0_\alpha = 0. \tag{12}$$

Let us suppose Eqs. (9) to be satisfied. We can prove as follows that Eqs. (12) remain valid in a neighborhood of Σ once they are satisfied on Σ.

Identities (5) may be written as

$$G^0_{\alpha,0} + \Gamma^\lambda_{0\lambda} G^0_\alpha - \Gamma^\lambda_{0\alpha} G^0_\lambda + G^i_{\alpha,i} + \Gamma^\lambda_{i\lambda} G^i_\alpha - \Gamma^\lambda_{i\alpha} G^i_\lambda = 0. \tag{13}$$

When Eqs. (9) are satisfied, the quantities G^i_λ are connected with the G^0_α by the relations

$$g^{00} G^i_j = g^{i0} G^0_j,$$

$$g^{00} G^i_0 = g^{i0} G^0_0 + \overset{*}{g}{}^{ik} G^0_k. \tag{14}$$

Since we have assumed g^{00} continuous in a neighborhood of Σ and > 0 on Σ, we have $g^{00} > 0$ in a neighborhood of Σ. Hence relations (14) determine G^i_j and G^i_0 from G^0_α. Then identities (13) become four linear and homogeneous partial differential equations of first order in G^0_α:

$$G^0_{\alpha,0} = A^{i\lambda}{}_\alpha G^0_{\lambda,i} + B^\lambda_\alpha G^0_\lambda, \tag{15}$$

where the coefficients $A^{i\lambda}{}_\alpha$ and B^λ_α are continuous. For the initial conditions $G^0_{\alpha|\Sigma} = 0$, the only solution of Eqs. (15) is $G^0_\alpha = 0$. Thus it is proved that the field equations (6) are in involution in the sense of Cartan.

A straightforward calculation shows that the quantities G^0_α contain no derivative of index 4. The conditions

$$G^0_{\alpha|\Sigma} = 0 \tag{16}$$

must be therefore imposed to the Cauchy data: They constitute a set of four initial constraint equations.

Hence the initial exterior Cauchy problem is split into two distinct parts:

(1) the problem of finding Cauchy data which satisfy the constraint equations (16) on Σ;

(2) the problem of integrating the dynamical equations (9) for these Cauchy data.

Let us now assume the constraint equations (16) to be fulfilled by the Cauchy data (problem 1) and let us try to solve problem 2. Using the relations

$$\overset{*}{g}{}^{kl} g_{km} = \delta^l_m \tag{17}$$

and contracting Eqs. (9) by $\overset{*}{g}{}^{ij}$, we get

$$\overset{*}{g}{}^{ij} G_{ij} \equiv -(3a+2b)(g^{00})^2 A + \overset{*}{g}{}^{ij} F_{ij}(\text{d.C.}) = 0. \tag{18}$$

A. Case $3a + 2b \neq 0$

If $3a + 2b \neq 0$, Eq. (18) determines A on Σ from the Cauchy data. Inserting the expression obtained for A in Eqs. (9), we can determine the fourth-order derivatives $g_{ij,0000}$ on Σ from the Cauchy data, since we assume $b \neq 0$ and $g^{00}{}_{|\Sigma} > 0$. Thus it is proved that the $g_{ij,0000}$ are continuous across Σ.

If we restrict the solution to the class of analytic functions, the Cauchy–Kowaleska theorem shows that the second problem admits a real analytic solution which is unique to within a transformation of coordinates leaving unaltered the coordinates of any point on Σ and the Cauchy data on Σ. It should be noted that the $g_{\mu\nu}$ can be analytic in empty

regions of the space–time free from gravitational shock waves, but that discontinuities would appear in the fourth-order derivatives of the $g_{\mu\nu}$ across hypersurfaces separating the matter from the vacuum. However, more work is needed on this question, and we shall not go far away in this direction.

B. Case $3a + 2b = 0$

If $3a + 2b = 0$, A is not determined by (18) since this equation is reduced to

$$\overset{*}{g}^{ij}G_{ij} \equiv \overset{*}{g}^{ij}F_{ij}(\text{d.C.}) = 0. \tag{19}$$

This relation constitutes a new constraint on the Cauchy data, which is different from the four constraints (16). This fact is easy to understand. If $3a + 2b = 0$, the trace equation (4) implies in the vacuum

$$R = 0. \tag{20}$$

Hence, the values of R on Σ being completely determined by the Cauchy data, we must impose the additional constraint

$$R_{|\Sigma} = 0. \tag{21}$$

But in the case of $3a + 2b = 0$, (21) is equivalent to (19) provided that Eqs. (16) are satisfied. Indeed, the following identity is valid for any symmetric tensor $G_{\mu\nu}$,

$$\overset{*}{g}^{ij}G_{ij} = G - g^{0\alpha}G^{0}_{\alpha}/g^{00}, \quad G = g^{\alpha\beta}G_{\alpha\beta}, \tag{22}$$

and $G = -R$ for the tensor (3) when $3a + 2b = 0$.

Moreover, (20) implies $R_{,0} = 0$ and $R_{,0}$ is determined by the Cauchy data on Σ. Hence the Cauchy data must also satisfy the supplementary constraint

$$R_{,0|\Sigma} = 0. \tag{23}$$

We demonstrate in Appendix B that the exterior Cauchy problem can be split into the following parts:

1. The problem of finding Cauchy data which satisfy the constraint equations (16), (21), and (23) on Σ.

2. The problem of integrating the dynamical equations

$$\tilde{G}_{ij} = 0, \tag{24}$$

where

$$\begin{aligned}\tilde{G}_{ij} &\equiv G_{ij} + \tfrac{1}{3}(g^{0\alpha}G^{0}_{\alpha}/g^{00})g_{ij} + \tfrac{1}{6}b\,\Box R g_{ij} \\ &\equiv -\tfrac{1}{2}b\,(g^{00})^2 g_{ij,0000} + \tilde{F}_{ij}(\text{d.C.}),\end{aligned} \tag{25}$$

the $\tilde{F}_{ij}$ (d.C.) being quantities which can be calculated on Σ from the Cauchy data.

Since we assume $b\,(g^{00})^2 \neq 0$ on Σ, Eqs. (24) yield the $g_{ij,0000}$ on Σ from the Cauchy data. As in the case of $3a + 2b \neq 0$, these derivatives are continuous across Σ.

This study shows that there is no possibility of determining the $g_{ij,0000}$ if $g^{00}{}_{|\Sigma} = 0$; Σ is then a characteristic hypersurface. Thus we see that any characteristic hypersurface in the vacuum is described by a local equation of the form

$$f(x^{\alpha}) = 0,$$

where f is a solution of the equation

$$g^{\alpha\beta}f_{,\alpha}f_{,\beta} = 0.$$

Therefore, the characteristic hypersurfaces are null hypersurfaces, exactly as in general relativity.

III. FIELD EQUATIONS WHEN $b = 0$

When $b = 0$, the Cauchy problem cannot be solved if one formulates it as in Sec. II since the $g_{ij,0000}$ are not present in the field equations (9). However, we shall demonstrate in this section that the fourth-order theory may be replaced by an equivalent second-order scalar–tensor theory. In this new version, the Cauchy problem is formulated differently and can be solved properly.

In the case where $b = 0$, the field equations (2) are

$$G_{\mu\nu} \equiv (1 + aR)S_{\mu\nu} + \tfrac{1}{4}aR^{2}g_{\mu\nu} + ag_{\mu\nu}\Box R - aR_{;\mu;\nu} = \kappa T_{\mu\nu}, \tag{26}$$

where $S_{\mu\nu} = R_{\mu\nu} - \tfrac{1}{2}Rg_{\mu\nu}$.

Now let us consider the set (g,Q) constituted by a metric tensor g and a scalar field Q. We associate to (g,Q) the scalar field K and the symmetric tensor $L_{\mu\nu}$ defined by

$$K = a\Box Q - \tfrac{1}{3}Q, \tag{27}$$

$$L_{\mu\nu} = (1 + aQ)S_{\mu\nu} + (\tfrac{1}{3}Q + \tfrac{1}{4}aQ^{2})g_{\mu\nu} - aQ_{;\mu;\nu} + Kg_{\mu\nu}. \tag{28}$$

The tensor $L_{\mu\nu}$ and the scalar K verify the identities

$$L - 3K = (1 + aQ)(Q - R), \quad L = g^{\alpha\beta}L_{\alpha\beta}, \tag{29}$$

$$L^{\mu}{}_{\nu;\mu} = \tfrac{1}{2}a(Q - R)Q_{;\nu}. \tag{30}$$

It is easy to show that if we except a very particular case, the fourth-order equations (26) can be replaced by the following set of second-order equations relating g and Q to the source of the gravitational field,

$$L_{\mu\nu} = \kappa T_{\mu\nu}, \tag{31}$$

$$K = \tfrac{1}{3}\kappa T. \tag{32}$$

Let us suppose that a metric g satisfies the field equations (26). Then, we can associate to it a set (g,Q) solution of Eqs. (31)–(32), where Q is given by

$$Q = R. \tag{33}$$

If (33) holds, we have indeed

$$L_{\mu\nu} = G_{\mu\nu},$$

and Eq. (32) becomes the equation obtained by contracting (26).

Conversely, let (g,Q) be a solution of Eqs. (31)–(32). In order to demonstrate that g is a solution of (26), it is sufficient to show that (33) is valid.

Taking into account identity (29), Eqs. (31)–(32) imply that at any point of V_4

$$(1 + aQ)(Q - R) = 0. \tag{34}$$

Differentiating this equation we get

$$(1 + aQ)(Q - R)_{;\mu} + a(Q - R)Q_{;\mu} = 0. \tag{35}$$

Let us suppose now that there exists a point $p \in V_4$ such that $Q \neq R$. Then it follows from (34) and (35) that at this point

$$1 + aQ = 0, \quad Q_{;\mu} = 0. \tag{36}$$

Differentiating (35) and taking (36) into account, we get at p

$$Q_{;\mu;\nu} = 0.$$

Therefore, at the point p, $\Box Q = 0$, and Eqs. (31)–(32) are reduced to

$$(1/4a)g_{\mu\nu} = \kappa T_{\mu\nu}, \quad 1/a = \kappa T. \tag{37}$$

Since the cosmological constant is taken here to be zero, such an expression of $T_{\mu\nu}$ is excluded in a vacuum ($T_{\mu\nu} = 0$) or in a pure electromagnetic field ($T = 0$). The same conclusion is true for a matter fluid described by

$$T_{\mu\nu} = (\rho + p)\mu_\mu\mu_\nu - pg_{\mu\nu} + \tau_{\mu\nu}$$

with the usual condition $\tau_{\mu\nu}\mu^\mu = 0$, μ^μ being the unit 4-velocity of the fluid and ρ the matter–energy rest density. Multiplying $T_{\mu\nu}$ by $\mu^\mu\mu^\nu$ and comparing with (37), we find

$$\kappa\rho = 1/4a. \tag{38}$$

But it results from (27) and (32) that a must be assumed < 0 in order to avoid a tachyonic propagation for the scalar field Q. On the other hand, the coupling constant κ must be > 0 in order to ensure an attractive predominant force of gravity. Hence (38) contradicts the natural condition $\rho > 0$, which proves that (37) is not possible for a realistic fluid when there is no tachyonic field.

Finally, we can conclude that if the no-tachyon condition $a < 0$ is realized, then Eq. (33) is valid everywhere for any usual energy–momentum tensor, and particularly for $T_{\mu\nu} = 0$. The same conclusion could be drawn with a cosmological constant $\Lambda \neq (4a)^{-1}$.

It should be noted that if Eqs. (31)–(32) are satisfied, the energy–momentum tensor $T_{\mu\nu}$ is divergence-free. Taking into account identity (30) and Eq. (35), the field equations (31) imply

$$\begin{aligned}\kappa T^\mu{}_{\nu;\mu} &= \tfrac{1}{2}a(Q - R)Q_{;\nu} \\ &= -\tfrac{1}{2}(1 + aQ)(Q - R)_{;\nu}.\end{aligned}$$

Now applying (34) we get immediately $T^\mu{}_{\nu;\mu} = 0$.

IV. THE EXTERIOR CAUCHY PROBLEM FOR $b = 0$

We turn now to study the exterior Cauchy problem for the system of second-order equations (31)–(32). Given a spacelike hypersurface Σ ($x^0 = 0$), the Cauchy data consist of the values on Σ of $g_{\mu\nu}$, $g_{\mu\nu,0}$, Q, and $Q_{,0}$. The differential properties of the space–time and of the metric g are supposed to be the same as in Sec. II.

We assume g^{00} continuous and $g^{00}{}_{|\Sigma} > 0$. Hence g^{00} is > 0 in a neighborhood of Σ. By a method similar to that which has been already employed in Sec. II, one can easily see that in this neighborhood of Σ, Eqs. (31)–(32) are equivalent to the set of equations

$$L_{ij} \equiv -\tfrac{1}{2}g^{00}(1 + aQ)(g_{ij,00} - Bg_{ij}) + Kg_{ij} + H_{ij}(\text{d.C.}) = 0, \tag{39}$$

$$L^0_\alpha \equiv H^0_\alpha(\text{d.C.}) = 0, \tag{40}$$

$$K \equiv ag^{00}Q_{,00} + E(\text{d.C.}) = 0, \tag{41}$$

where

$$B = g^{kl}g_{kl,00} \tag{42}$$

and $H_{ij}(\text{d.C.})$, $H^0_\alpha(\text{d.C.})$, and $E(\text{d.C.})$ denote quantities which can be computed on Σ from the Cauchy data.

There are no derivatives $g_{\mu 0,00}$ in the above equations. As in general relativity, these derivatives do not play an essential role in the problem since they can be reduced to any specified values on Σ (provided that they are continuous across Σ) by an admissible local coordinate transformation of the form

$$x^{\lambda'} = x^\lambda + \tfrac{1}{6}(x^0)^3\varphi^{(\lambda)}(x^i),$$

where $\varphi^{(\lambda)}(x^i)$ are functions of class C^4 and piecewise continuous of class C^6. (It is easy to verify that neither the Cauchy data nor the derivatives $g_{ij,00}$ are modified by such a coordinate transformation.) The indetermination of the $g_{\mu 0,00}$ can be supplied by imposing four coordinate conditions consistent with the Cauchy data.

The quantities L^0_α contain no derivatives of index 2. Therefore, the initial conditions

$$L^0_{\alpha|\Sigma} = 0 \tag{43}$$

constitute a set of four constraint equations which the Cauchy data have to verify.

Let us prove now that Eqs. (39)–(41) are sufficient to maintain the validity of Eqs. (40) in the neighborhood of Σ, once the constraint equations (43) are satisfied. First, the quantities L^i_λ are connected with the L^0_α by relations similar to (14), where the G^α_β have to be replaced by L^α_β. As a consequence, identities (30) yield the equations

$$L^0_{\alpha,0} = C^{i\lambda}{}_\alpha L^0_{\alpha,i} + D^\lambda_\alpha L^0_\lambda + \tfrac{1}{2}aUQ_{,\alpha}, \tag{44}$$

where U is defined by

$$U = Q - R$$

and $C^{i\lambda}{}_\alpha$ and D^λ_α denote continuous quantities built from $g_{\alpha\beta}$ and $g_{\alpha\beta,\gamma}$. Secondly, differentiating identity (29) with respect to x^0 we get

$$L_{,0} = (1 + aQ)U_{,0} + aUQ_{,0}. \tag{45}$$

But Eqs. (39) imply

$$L = g^{0\alpha}L^0_\alpha/g^{00}.$$

Thus, Eq. (45) may be written as

$$(1 + aQ)U_{,0} + aUQ_{,0} = (g^{0\alpha}/g^{00})L^0_{\alpha,0} + (g^{0\alpha}/g^{00})_{,0}L^0_\alpha. \tag{46}$$

At this stage, let us assume that $1 + aQ$ is continuous and

$$1 + aQ_{|\Sigma} \neq 0. \tag{47}$$

Consequently, $1 + aQ \neq 0$ in a neighborhood of Σ. Hence L^0_α and U satisfy five linear and homogeneous partial differential equations of first order, namely Eqs. (44)–(46), which are resoluble in $L^0_{\alpha,0}$ and $U_{,0}$. For the initial conditions $L^0_{\alpha|\Sigma} = 0$ and $U_{|\Sigma} = 0$ [this last condition being a direct consequence of identity (29) and of the field equations on Σ], the only solution of Eqs. (44)–(46) is

$$L^0_\alpha = 0, \quad U = 0.$$

Therefore, Eqs. (40) are maintained once they are satisfied on Σ, provided that (47) is realized. Moreover, we recover $Q = R$ in the particular case of a vacuum.

The assumption (47) is essential. Indeed, if we suppose $1 + aQ = 0$ at a point of Σ, the coefficient of the $g_{ij,00}$ vanishes in Eqs. (39): In this case, the evolution of the metric is not determined by the initial conditions.

We conclude from these results that the initial exterior Cauchy problem must be split into the following parts:

1. The problem of finding Cauchy data on Σ which satisfy the constraint equations (43) and the condition (47).

2. The problem of integrating the dynamical equations (39)–(41).

If the problem 1 is solved, Eq. (41) enables us to calculate $Q_{,00}$ on Σ from the Cauchy data, since $g^{00}{}_{|\Sigma}$ is assumed $\neq 0$. Then inserting the value of $Q_{,00}$ in Eqs. (39) and contracting by $\overset{*}{g}{}^{ij}$, we get

$$\overset{*}{g}{}^{ij}L_{ij} \equiv g^{00}(1+aQ)B + \overset{*}{g}{}^{ij}H_{ij}(\text{d.C.}) = 0.$$

This equation determines B on Σ from the Cauchy data. Hence the derivatives $g_{ij,00}$ can be determined on Σ.

When $g^{00}{}_{|\Sigma} = 0$, it is impossible to determine the $g_{ij,00}$. Consequently, the characteristic hypersurfaces coincide with the null hypersurfaces, as in general relativity.

V. THE SCALAR–TENSOR VERSION OF THE THEORY WHEN $b = 0$

In order to solve the Cauchy problem when $b = 0$, we have been led to replace the primitive fourth-order theory by a second-order scalar–tensor theory. This equivalent version can be deduced from a variational principle. In fact, the Lagrangian density

$$\mathscr{L} = \sqrt{-g}\,[(1+aQ)R - \tfrac{1}{2}aQ^2] - \kappa\mathscr{L}_m \tag{48}$$

yields the field equations

$$(1+aQ)S_{\mu\nu} + \tfrac{1}{4}aQ^2 g_{\mu\nu} + a g_{\mu\nu}\Box Q - aQ_{;\mu;\nu} = \kappa T_{\mu\nu}, \tag{49}$$

$$Q = R, \tag{50}$$

which are strictly equivalent to the starting equations (26).

The Lagrangian density (48) belongs to the class of Lagrangian densities proposed by O'Hanlon[7] in order to present a covariant model of the massive dilaton theory suggested by Fujii.[8] Starting from densities of the kind[9]

$$\mathscr{L} = \sqrt{-g}\,[\varphi R + m^2 f(\varphi)] - 8\pi\mathscr{L}_m, \tag{51}$$

where φ is a scalar field, O'Hanlon takes for the potential $f(\varphi)$:

$$f(\varphi) = 3(2\varphi_0)^{-1}(\varphi - \varphi_0)^2. \tag{52}$$

With this potential, the Lagrangian density (51) is equivalent to (48). It is easily verified by defining

$$\varphi = 8\pi\kappa^{-1}(1+aQ),$$

$$\varphi_0 = 8\pi\kappa^{-1}, \quad m^2 = -1/3a.$$

Such an equivalence shows the purely geometrical nature of the theory corresponding to (52). This result is the more interesting as it is not isolated. Indeed, we will now demonstrate that the class of the scalar–tensor theories derived from (51) coincide with the class of the fourth-order theories yielded by the Lagrangian densities

$$\mathscr{L} = \sqrt{-g}F(R) - \kappa\mathscr{L}_m, \tag{53}$$

where F is any real function of class C^3 different from the affine function.

Varying (53) with respect to $g^{\mu\nu}$, we get the field equations[10]

$$F'R_{\mu\nu} - \tfrac{1}{2}Fg_{\mu\nu} + F''(g_{\mu\nu}\Box R - R_{;\mu;\nu}) + F'''(g_{\mu\nu}R_{;\lambda}R^{;\lambda} - R_{;\mu}R_{;\nu}) = \kappa T_{\mu\nu}, \tag{54}$$

where a prime denotes a derivation with respect to R. Using the identities

$$[F'(R)]_{;\mu;\nu} = F''(R)R_{;\mu;\nu} + F'''(R)R_{;\mu}R_{;\nu},$$

Eqs. (54) become[11]

$$F'R_{\mu\nu} - \tfrac{1}{2}Fg_{\mu\nu} + g_{\mu\nu}\Box F' - F'_{;\mu;\nu} = \kappa T_{\mu\nu}. \tag{55}$$

Now the variational principle applied to (51) leads to the equations

$$\varphi R_{\mu\nu} - \tfrac{1}{2}[R\varphi + m^2 f(\varphi)]g_{\mu\nu} + g_{\mu\nu}\Box\varphi - \varphi_{;\mu;\nu} = 8\pi T_{\mu\nu}, \tag{56}$$

$$R = -m^2 f'(\varphi), \tag{57}$$

where the prime denotes here a derivation with respect to φ. We may note that the propagation equation for the scalar φ given by O'Hanlon is obtained by contracting Eqs. (56) and inserting the expression (57) of R:

$$3\Box\varphi + m^2(\varphi f' - 2f) = 8\pi T. \tag{58}$$

Equations (55) are changed into Eqs. (56) under the Legendre transformation $(R,F)\rightarrow(\varphi,f)$ defined by

$$\kappa\varphi = 8\pi F'(R), \tag{59a}$$

$$-\kappa m^2 f(\varphi) = 8\pi(RF' - F). \tag{59b}$$

This transformation is formally regular if $F''(R)$ does not identically vanish. [Note that $F''(R)$ identically null corresponds to the Lagrangian function of general relativity.] Differentiating (59b) and taking into account (59a), we get

$$-m^2\,df = R\,d\varphi,$$

relation equivalent to Eq. (57).

Conversely, Eqs. (56) are changed into Eqs. (55) under the dual transformation of (59),

$$R = -m^2 f'(\varphi),$$

$$8\pi F(R) = -\kappa m^2(\varphi f' - f).$$

VI. CONCLUSIONS

The solution of the Cauchy problem for the fourth-order theories derived from the family of Lagrangian densities (1) depends upon the values of the parameter b.

A. When $b \neq 0$, the Cauchy problem can be solved for Cauchy data consisting of the potentials $g_{\mu\nu}$ and of their derivatives $g_{\mu\nu,0}, g_{\mu\nu,00}, g_{\mu\nu,000}$ on a spacelike hypersurface Σ locally described by $x^0 = 0$. Two cases have to be carefully distinguished:

1. If $3a + 2b \neq 0$, the Cauchy data must satisfy four initial constraint equations, exactly as in general relativity.

2. If $3a + 2b = 0$, two supplementary constraint equations, $R_{|\Sigma} = 0$ and $R_{,0|\Sigma} = 0$, are implied by the field equations. Moreover the six primitive dynamical equations must be written in a modified form consistent with these supplementary constraints.

In each case, the constraint equations are propagated by the dynamical equations once they are fulfilled on Σ.

B. When $b = 0$, the Cauchy problem cannot be solved for the Cauchy data specified in the case where $b \neq 0$ since it

is impossible to determine the $g_{ij,0000}$ from the dynamical equations. As a matter of fact, the fourth-order theory obtained when $b = 0$ is equivalent to a second-order scalar–tensor model of gravity. The unknown fields of this model are the ten potentials $g_{\mu\nu}$ and a scalar field Q which happens to coincide with the scalar curvature R as a consequence of the field equations. Hence the Cauchy data consist of the values of $g_{\mu\nu}$, $g_{\mu\nu,0}$, Q, and $Q_{,0}$ on the initial hypersurface Σ. These data must satisfy the condition $1 + aQ_{|\Sigma} \neq 0$ and four initial constraint equations. Once they are fulfilled on Σ, the constraint equations are propagated by the dynamical equations of the model, and consequently the equality $Q = R$ holds in a neighborhood of Σ. In that theory, the scalar curvature R plays the role of a massive scalar field, and its values must be given on the initial hypersurface Σ together with its normal derivative in order to ensure the determinism.

Moreover, we have shown that the scalar–tensor model which corresponds to $b = 0$ can be identified with one of the theories of O'Hanlon. We have generalized this conclusion: The class of the scalar–tensor models of gravity proposed by O'Hanlon has been proved to be indistinguishable from the class of the fourth-order theories derived from the Lagrangian densities $\mathscr{L} = \sqrt{-g}\,F(R) - \kappa\mathscr{L}_m$, where $F(R)$ is any real function of R such that $F''(R)$ does not identically vanish. This result should be very useful to treat the Cauchy problem related to this class of theories.

APPENDIX A

We have to show that there exist some coordinate transformations of class C^4 and piecewise continuous of class C^6 which reduce the derivatives $g_{\mu 0,0000}$ to any specified values on Σ but affect neither the Cauchy data nor the $g_{ij,0000}$ on Σ.

Let Ω be an open subset of V_4 such that $\Omega \cap \Sigma \neq \emptyset$. We call Ω^+ (resp. Ω^-) the part of the open Ω corresponding to $x^0 > 0$ (resp. $x^0 < 0$). For a function $f: \Omega \to \mathbb{R}$, which may be discontinuous across Σ, we use the classical notation

$$f_\Sigma^+(x^i) = \lim_{x^0 \to +0} f(x^0,x^i), \quad f_\Sigma^-(x^i) = \lim_{x^0 \to -0} f(x^0,x^i).$$

Let us now consider the coordinate transformation defined on Ω:

$$\begin{aligned} x^{\lambda'} &= x^\lambda + [(x^0)^5/120]\varphi^{+(\lambda)}(x^i) && \text{on } \Omega^+, \\ x^{\lambda'} &= x^\lambda && \text{on } \Omega \cap \Sigma, \\ x^{\lambda'} &= x^\lambda + [(x^0)^5/120]\varphi^{-(\lambda)}(x^i) && \text{on } \Omega^-, \end{aligned} \tag{A1}$$

where the $\varphi^{+(\lambda)}(x^i)$ [resp. $\varphi^{-(\lambda)}(x^i)$] are functions of the x^i of class C^4 and piecewise continuous of class C^6 on Ω^+ (resp. Ω^-). This transformation does not change the coordinates of the points on Σ, since Σ is described by $x^0 = 0$.

Clearly the coefficients of transformation associated with (A1),

$$A^{\alpha'}_{\lambda} = \frac{\partial x^{\alpha'}}{\partial x^\lambda},$$

are of class C^3 and piecewise continuous of class C^5 on Ω. On the hypersurface Σ, we have indeed

$$A^{\alpha'}_{\lambda|\Sigma} = (A^{\alpha'}_{\lambda})^+_\Sigma = (A^{\alpha'}_{\lambda})^-_\Sigma = \delta^\alpha_\lambda, \tag{A2}$$

$$A^{\alpha'}_{\lambda,\mu_1\cdots\mu_n|\Sigma} = (A^{\alpha'}_{\lambda,\mu_1\cdots\mu_n})^+_\Sigma = (A^{\alpha'}_{\lambda,\mu_1\cdots\mu_n})^-_\Sigma = 0 \quad \text{for } n \leqslant 3.$$

Consequently the coordinate transformation (A1) is of class C^4 and piecewise continuous of class C^6.

The metric components are transformed as

$$g_{\mu\nu} = A^{\alpha'}_{\mu} A^{\beta'}_{\nu} g_{\alpha'\beta'}. \tag{A3}$$

Differentiating (A3) with respect to x^0 an appropriate number of times, then substituting ∂_0 with $A^{\alpha'}_0 \partial_{\alpha'}$, and using (A2), we find that the Cauchy data are not modified by the transformation (A1):

$$g_{\mu\nu|\Sigma} = g_{\mu'\nu'|\Sigma}, \ldots, g_{\mu\nu,000|\Sigma} = g_{\mu'\nu',0'0'0'|\Sigma}.$$

Now, the derivatives of index 4 are transformed as

$$\begin{aligned} g_{\mu\nu,0000} = {} & A^{\alpha'}_{\mu} A^{\beta'}_{\nu} A^{\gamma'}_0 A^{\epsilon'}_0 A^{\eta'}_0 A^{\lambda'}_0 g_{\alpha'\beta',\gamma'\epsilon'\eta'\lambda'} \\ & + (A^{\alpha'}_{\mu,0000} A^{\beta'}_{\nu} + A^{\alpha'}_{\mu} A^{\beta'}_{\nu,0000}) g_{\alpha'\beta'} + [\cdots], \end{aligned}$$

where the symbol $[\cdots]$ denotes quantities which vanish on Σ according to (A2). Taking into account the following relations,

$$\begin{aligned} & A^{\alpha'}_{\lambda,\mu_1\mu_2\mu_3 i|\Sigma} = (A^{\alpha'}_{\lambda,\mu_1\mu_2\mu_3 i})^+_\Sigma = (A^{\alpha'}_{\lambda,\mu_1\mu_2\mu_3 i})^-_\Sigma = 0, \\ & A^{\alpha'}_{i,\mu_1\mu_2\mu_3\lambda|\Sigma} = (A^{\alpha'}_{i,\mu_1\mu_2\mu_3\lambda})^+_\Sigma = (A^{\alpha'}_{i,\mu_1\mu_2\mu_3\lambda})^-_\Sigma = 0, \\ & (A^{\alpha'}_{0,\mu_1\mu_2\mu_3 0})^+_\Sigma = \delta_{\mu_1 0}\delta_{\mu_2 0}\delta_{\mu_3 0}\varphi^{+(\alpha)}, \\ & (A^{\alpha'}_{0,\mu_1\mu_2\mu_3 0})^-_\Sigma = \delta_{\mu_1 0}\delta_{\mu_2 0}\delta_{\mu_3 0}\varphi^{-(\alpha)}, \end{aligned}$$

we get

$$\begin{aligned} & g_{ij,0000|\Sigma} = g_{i'j',0'0'0'0'|\Sigma}, \\ & (g_{\mu 0,0000})^+_\Sigma = (g_{\mu'0',0'0'0'0'})^+_\Sigma + \delta_{\mu 0}\varphi^{+(\alpha)} g_{\alpha 0} + \varphi^{+(\beta)} g_{\mu\beta}, \\ & (g_{\mu 0,0000})^-_\Sigma = (g_{\mu'0',0'0'0'0'})^-_\Sigma + \delta_{\mu 0}\varphi^{-(\alpha)} g_{\alpha 0} + \varphi^{-(\beta)} g_{\mu\beta}. \end{aligned}$$

Therefore, the derivatives $g_{\mu 0,0000}$ can be reduced to specified values on each side of the hypersurface Σ by an appropriate transformation (A1), while the derivatives $g_{ij,0000}$ remain unchanged. In particular, the functions $\varphi^{+(\alpha)}$ and $\varphi^{-(\alpha)}$ can always be chosen so that the $g_{\mu 0,0000}$ are continuous across Σ.

APPENDIX B

In the case of $3a + 2b = 0$, let us suppose solved the problem of finding Cauchy data which satisfy the constraint equations (16), (21), and (23) on Σ. Our purpose is to prove that the field equations (24) are equivalent to Eqs. (9) when $a = -\frac{2}{3}b$.

1. Let us assume that Eqs. (9) are satisfied by the metric. Since the constraint equations (16) are supposed to be satisfied on Σ, Eqs. (12) hold in the neighborhood of Σ. (The analysis of Sec. II is valid even in the case where $3a + 2b = 0$.) Hence $R = 0$ in the neighborhood of Σ. Consequently, it follows from the definition (25) that

$$\tilde{G}_{ij} = G_{ij} = 0.$$

Therefore, Eqs. (24) are satisfied by the metric.

2. Conversely, let us admit that Eqs. (24) are verified by the metric. Then the identity

$$\overset{*}{g}{}^{ij}\tilde{G}_{ij} \equiv -R + \tfrac{1}{2}b\,\Box R$$

yields the equation

$$-R + \tfrac{1}{2}b\,\Box R = 0.$$

The unique solution of this equation satisfying the initial conditions (21) and (23) is

$$R = 0,$$

which implies that $\Box R = 0$. So Eqs. (24) are reduced to

$$G_{ij} = -\tfrac{1}{3}(g^{0\alpha}G^0_\alpha/g^{00})g_{ij}. \tag{B1}$$

It results from (B1) that the G^i_λ are related to the G^0_α by the relations

$$g^{00}G^i_j = g^{0i}G^0_j - \tfrac{1}{3}g^{0\alpha}G^0_\alpha\delta^i_j,$$

$$g^{00}G^i_0 = \tfrac{4}{3}g^{0i}G^0_0 + (g^{ij} - \tfrac{2}{3}g^{0i}g^{0j}/g^{00})G^0_j. \tag{B2}$$

Because of these relations, the conservation identities (13) give four equations similar to Eqs. (15):

$$G^0_{\alpha,0} = A'^{i\lambda}{}_\alpha G^0_{\lambda,i} + B'^\lambda_\alpha G^0_\lambda, \tag{B3}$$

where the coefficients $A'^{i\lambda}{}_\alpha$ and B'^λ_α are continuous quantities built from the $g_{\alpha\beta}$ and their first derivatives. [As in the discussion of Eqs. (15), we suppose $g^{00} > 0$.] For the initial conditions (16), the only solution of (B3) is

$$G^0_\alpha = 0.$$

Hence Eqs. (B1) are reduced to Eqs. (9) in the case where $3a + 2b = 0$, and Eqs. (12) are satisfied. Q.E.D.

[1] R. Utiyama and B. S. De Witt, J. Math. Phys. **3**, 608 (1962); A. D. Sakharov, Dokl. Akad. Nauk SSSR **177**, 70 (1968) [Sov. Phys. Dokl. **12**, 1040 (1968)]; G. t'Hooft and M. Veltman, Ann. Inst. H. Poincaré **20**, 69 (1974); S. Deser and P. van Nieuwenhuizen, Phys. Rev. D **10**, 401, 411 (1974); K. S. Stelle, Phys. Rev. D **16**, 953 (1977).

[2] Because of the Gauss–Bonnet theorem in four dimensions, (1) is the most general algebraic Lagrangian which is at most quadratic in the curvature tensor.

[3] T. V. Ruzmaikina and A. A. Ruzmaikin, Zh. Eksp. Teor. Fiz. **57**, 680 (1969) [Sov. Phys. JETP **30**, 372 (1970)]; H. Nariai, Progr. Theor. Phys. **46**, 433 (1971); H. Nariai and K. Tomita, Progr. Theor. Phys. **46**, 776 (1971); A. A. Ruzmaikin, Astrophysics (translation of Astrofizica, U.R.S.S.) **13**, 186 (1977); K. Tomita, T. Azuma, and H. Nariai, Progr. Theor. Phys. **60**, 403 (1978); K. I. Macrae and R. J. Riegert, Phys. Rev. D **24**, 2555 (1981).

[4] K. S. Stelle, Gen. Rel. Grav. **9**, 353 (1978).

[5] A. Lichnerowicz, *Théories relativistes de la gravitation et de l'électromagnétisme* (Masson, Paris, 1955). See also A. Z. Petrov, *Einstein Spaces* (Pergamon, Oxford, 1969) (translated from the Russian).

[6] The case where $b = 0$ is very significant since the no-tachyon constraints ($3a + 2b \leqslant 0$ $b \geqslant 0$) and the no-ghost constraint $b \leqslant 0$ imply that b must be zero. See D. E. Neville, Phys. Rev. D **21**, 867 (1980). The condition $b = 0$ is also required to have the positivity of the energy. See Ref. 4.

[7] J. O'Hanlon, Phys. Rev. Lett. **29**, 137 (1972).

[8] Y. Fujii, Nature Phys. Sci. (London) **234**, 5 (1971); Ann. Phys. (N.Y.) **69**, 494 (1972).

[9] We replace the term $-m^2 f(\varphi)$ in the original Lagrangian of O'Hanlon by $m^2 f(\varphi)$ because we use a metric of signature -2.

[10] H. A. Buchdahl, Mon. Not. Roy. Astron. Soc. **150**, 1 (1970).

[11] N. M. Polievktov-Nikoladze, Zh. Eksp. Teor. Fiz. **52**, 1360 (1967) [Sov. Phys. JETP **25**, 904 (1967)]; B. N. Breizman, V. Ts. Gurovich, and V. P. Sokolov, Zh. Eksp. Teor. Fiz. (1971) [Sov. Phys. JETP **32**, 155 (1971)]; H. Nariai, Progr. Theor. Phys. **51**, 613 (1974).

A probabilistic rejection test for multivariable sensitivity analysis

Y. Ronen and A. Dubi
Department of Nuclear Engineering, Ben-Gurion University of the Negev, Beer Sheva, Israel

(Received 18 May 1981; accepted for publication 28 January 1983)

A probabilistic rejection test for multivariable sensitivity analysis is presented. The test is applied by randomly changing all the assumed unimportant (those having low sensitivity values) input parameters simultaneously and calculating the appropriate response. It is shown that by repeating this procedure N times, where N is much smaller than the number of input parameters, it is possible to assign a probability limit to the assumption that a high sensitivity parameter exists. The application of the test is demonstrated in a nuclear waste disposal problem.

PACS numbers: 06.20.Dk, 02.50.Cw

1. INTRODUCTION

The development of large computer codes is enabling us to model in great detail many physical systems. Such codes provide solutions while taking into account large numbers of physical effects. However, due to the fact that there are many input parameters involved and the cost of running the computer is high, there are difficulties in performing survey studies. These difficulties are reflected in sensitivity questions which arise in all design and safety analyses. In such analyses, the effect of changes in the input parameters on the calculated results is important, namely, sensitivity analysis.

There are three approaches to sensitivity analysis. The first one is based on the use of adjoint functions.[1] Although this approach was successfully applied to many engineering and physics problems, it has the drawbacks that it is necessary to solve an additional set of equations for each response studied. For large and complicated codes which consist of many equations, an additional set of equations requires a considerable amount of effort. The second approach is one of the forms of the response surface method.[2] A variety of experimental design theories[3–5] are used in order to change the input parameters. These altered data sets are used for calculating the changes in the response of interest. These calculated results are then used to construct the response surface, which is fitted with a simple functional from the sensitivities' derivations. The advantage of this approach is that the original computer code is used for calculating the sensitivities. The main limitation is the number of computer runs needed, which in practice limits the number of input parameters that can be considered. The third approach is the "brute force" one. By this approach, each of the input parameters is changed slightly and the change in the response is calculated and so is the sensitivity component. The advantage of this approach is the straightforward usage of the code. The limitations are the number of runs which is dependent on the number of input parameters, and it is limited to small variations of the input parameters.

In many practical situations, we have systems with hundreds or thousands of input parameters. Treating such cases with the current response surface or "brute force" method is impractical, without screening of the important input parameters.

In most of the physical systems, the number of important parameters is small. By important parameters we mean those parameters with high sensitivity values. Physical systems with many important input parameters tend to be unstable.

In many cases, intuition or previous knowledge enable us to know in advance which of the input parameters is important. So instead of dealing with many input parameters, we can consider only the few important ones, by using the response surface or the "brute force" methods. However, intuition might fail and previous knowledge might be incomplete. *However, we are assuming that the probability that there are no important parameters after screening is not vanishingly small.* As a result, one can rely on such screening only if we have a test to validate our choice of important input parameters. Such a test was suggested,[6] however, without mathematical rigorousness.

This test was inspired by the "statistical screening" procedure suggested by Durston and Krieger[7] and later by Krieger.[8,9] Although the procedures of the "statistical screening" procedure and the suggested rejection test are the same, the fundamental bases of the two procedures are different.

The purpose of this paper is to present such a rejection test with all the mathematical rigorousness. The application of the test is demonstrated by an example from nuclear waste disposal analysis.

2. THE REJECTION TEST

Given the problem of *estimating* the relative sensitivity of a model response R to many modeling parameters α_i, the following methodology can be developed. Assume first

$$\delta R \simeq \sum_{i=1}^{M} \frac{\partial R}{\partial \alpha_i} \delta \alpha_i, \tag{1}$$

where $\delta R = R - R_0$, R is the perturbed response value, and R_0 is the base case response value; $\delta \alpha_i = \alpha_i - \alpha_{0i}$, and α_i is the perturbed parameter value and α_{0i} is its base case value; and the number of the input parameters is M.

Equation (1) can be rewritten in relative terms as

$$\frac{\delta R}{R_0} \simeq \sum_{i=1}^{M} \frac{\partial R / R_0}{\partial \alpha_i / \alpha_{0i}} \frac{\delta \alpha_i}{\alpha_{0i}} = \sum_{i=1}^{M} S_i a_i, \tag{2}$$

where $S_i = (\partial R / R_0)/(\partial \alpha_i / \alpha_{0i})$ is called the sensitivity of R to α and $a_i = \delta \alpha_i / \alpha_i$.

 0022-2488/83/122800-09$02.50

Consider that by previous knowledge we have screened P_1 important parameters. The number of remaining parameters is $K = M - P_1$. These are the parameters to which we assign low sensitivity, namely, S_i are small. Due to the fact that our previous knowledge is not complete, there is a nonzero probability that among the K parameters there are some parameters which have high sensitivity. The aim of the rejection test is to find the probability that there are high sensitivity coefficients among the K parameters. Changing the input parameters by $+1\%$ or -1% randomly, Eq. (2) will have the form

$$\frac{\delta R}{R_0} \times 100 = \sum_{i=1}^{K} S_i r_i, \tag{3}$$

where $r_i = \pm 1$. Making N different runs so that in each run the input parameters are changed by $\pm 1\%$, we will have

$$\frac{\delta R_j}{R_0} \times 100 = \sum_{i=1}^{K} S_i r_{ji} = \mathbf{S}\mathbf{r}_j, \quad j = 1, ..., N \quad N < K. \tag{4}$$

If for N runs $\delta R_j / R_0 \simeq 0$, what is the probability that there are S_i components different from zero?

Assume that there are I components of the vector $\mathbf{S}$ which are different from zero. The value $I \leqslant K$ is unknown.

The nonzero components of the vector $\mathbf{S}$ can be considered as a new vector $\mathbf{S}^1$ with a dimension I satisfying the equation

$$\mathbf{S}^{(1)} \cdot \mathbf{a}_j = \sum_{n=1}^{I} S_n a_{jn} = 0, \tag{5}$$

where $j = 1,...,N$ and $a_{jn} = +1$ or -1. The vectors $\mathbf{a}_j$ are subvectors of the vectors $\mathbf{r}_j$. For $j = 1$, we have

$$\mathbf{S}^{(1)} \cdot \mathbf{a}_1 = \sum_{n=1}^{I} S_n a_{1n} = 0. \tag{6}$$

For a given vector $\mathbf{S}^{(1)}$ there are other vectors with components of ± 1 beside $\mathbf{a}_1$ which are orthogonal to $\mathbf{S}^{(1)}$. For example, $\mathbf{a}_2 = -\mathbf{a}_1$ will also be orthogonal to $\mathbf{S}^{(1)}$. The number of vectors of the type $\mathbf{a}$ which are orthogonal to $\mathbf{S}^{(1)}$ depends on the particular nature of $\mathbf{S}^{(1)}$. Define a value Q_I as the number of vectors of the type $\mathbf{a}$ which are orthogonal to $\mathbf{S}^{(1)}$. Since the nature of the vector $\mathbf{S}^{(1)}$ is not known, the value Q_I is unknown. The number of all possible vectors of type $\mathbf{a}$ is 2^I. Thus the probability of obtaining Eq. (6) when the vector $\mathbf{a}_1$ is chosen randomly is

$$p(I, 1) = Q_I / 2^I. \tag{7}$$

The probability of obtaining Eq. (5) for N random runs will be

$$p(I, N) = (Q_I / 2^I)^N, \tag{8}$$

namely, the probability that a certain vector $\mathbf{S}^{(1)}$ with I nonzero components will satisfy Eq. (5).

The value Q_I will be the largest when the vector $\mathbf{S}^{(1)}$ has the property of $S_1 = S_2 = S_3 = \cdots = S_I$ for I even, and $S_2 = S_3 = \cdots = S_I$, $S_1 = 2S_2$ for I odd. The proofs of these statements are given in the next chapter. Thus,

$$Q_I \leqslant 2\binom{I-1}{(I+1)/2} \quad \text{for } I \text{ odd}, \tag{9}$$

$$Q_I \leqslant \binom{I}{I/2} \quad \text{for } I \text{ even}. \tag{10}$$

Thus the probability that any vector $\mathbf{S}^{(1)}$ with I nonzero components which satisfies Eq. (5) will be bounded by

$$p(I, N) \leqslant [Q(I)/2^I]^N. \tag{11}$$

Thus the probability that any vector with nonzero components will satisfy Eq. (5) is

$$P(K, N) = \sum_{I=2}^{K} p(I, N) \leqslant \sum_{l=1}^{K/2} \left\{ \left[\binom{2l}{l} \Big/ 2^{2l} \right]^N + \left[\binom{2l}{l+1} \Big/ 2^{2l} \right]^N \right\} - \left[\binom{K}{(K/2)+1} \Big/ 2^K \right]^N = R(K, N) \tag{12}$$

for K even, and

$$P(K, N) = \sum_{I=3}^{K} p(I, N) \leqslant \sum_{l=1}^{(K-1)/2} \left[\binom{2l}{l} \Big/ 2^{2l} \right]^N + \left[\binom{2l}{l+1} \Big/ 2^{2l} \right]^N = T(K, N) \tag{13}$$

for K odd.

The values of $R(K, N)$ and $T(K, N)$ are given in Table I. From these values, it is obtained that for $N \geqslant 10$ the upper bound on the probabilities practically does not depend on the number of input parameters. Furthermore, for $N \geqslant 10$, the probability of obtaining nonzero coefficients with the rejection test is very small.

Assume that a given system can have two possible states: A_0, A_1, $(A_0 \cap A_1 = \varnothing)$ such that the probability ratio $P(A_1)/P(A_0)$ is known to have an upper bound $P(A_1)/P(A_0) \leqslant \gamma$. Let C_N denote an "event," i.e., a possible outcome of an experiment E done on/with the system such that the conditional probabilities $P(C_N|A_0) = 1$ and $P(C_N|A_1) \leqslant \alpha_N$ and C_N can be chosen such that $\alpha_N \gamma \ll 1$. One now wants to reject the hypothesis that the system is in state A_1 by stating that if $P(A_1|C_N) \leqslant \beta$ (β being a predetermined limit, say, 10^{-3}), then A_1 is rejected. Now

TABLE I. Rejection criterion.

N Number of trials	K Number of input parameters	The rejection criteria R or T
6	15	0.021 121 8
8	15	0.004 494 6
6	20	0.020 251 0
8	20	0.004 445 6
10	20	0.001 044 2
12	20	0.000 253 0
14	20	0.000 062 2
6	25	0.021 207 4
8	25	0.004 497 7
10	25	0.001 047 1
12	25	0.000 253 2
14	25	0.000 062 2
6	30	0.020 251 0
8	30	0.004 445 6
10	30	0.001 044 2
12	30	0.000 253 0
14	30	0.000 062 2

$$P(A_1|C_N)P(C_N) = P(C_N|A_1)P(A_1), \tag{14}$$

$$P(A_0|C_N)P(C_N) = P(C_N|A_0)P(A_0), \tag{15}$$

and, upon division,

$$\frac{P(A_1|C_N)}{1-P(A_1|C_N)} = P(C_N|A_1)\frac{P(A_1)}{P(A_0)} \tag{16}$$

yields

$$P(A_1|C_N) = \frac{P(C_N|A_1)P(A_1)/P(A_0)}{1+P(C_N|A_1)P(A_1)/P(A_0)} \leqslant \alpha_N \gamma. \tag{17}$$

One then conducts the experiment choosing C_N to be such that $\alpha_N < \beta/\gamma$. If C_N comes out in the experiment, A_1 is rejected, and the system is assumed to be in state A_0.

Let us now proceed one step further: Let

$$\{C_N\}_{N=1}^{\infty}$$

be a set of possible outcomes of the experiment E with the conditional probabilities $P(C_N|A_0) = 1\ \forall\ N$ and $P(C_N|A_1) = \alpha_N, N = 1,...,\infty$ known. One may then set a limit β (a rejection limit) and conduct the experiment E. If the outcome C_N is such that $\alpha_N \leqslant \beta/\gamma$, the possibility that the system is in state A_1 is rejected. Now suppose that A_1 is a state composed of the union of K mutually exclusive substates

$$A_1 = \bigcup_{i=1}^{K} B_I, \quad B_I \cap B_J = \varnothing, \quad \forall\ I \neq J$$

and suppose that $P(C_N|B_I)$ is known. Then

$$P(C_N|A_1) = \sum_{I=1}^{K} P(C_N|B_I). \tag{18}$$

This is the general structure of the suggested test, with the following identifications:

A_0-(the vector $\mathbf{S}^{(1)} = 0$, all its components are zero), A_1-(at least one component of $\mathbf{S}^{(1)}$ differs from zero), B_I-(exactly I components of $\mathbf{S}^{(1)}$ differ from zero), E-the experiment (pick up N vectors $\mathbf{a}_j = (a_1^j,...,a_I^j)$ such tha $a_n^j = \pm 1$ with probability $\frac{1}{2}$ and calculate the N sums $D_j = \mathbf{S}^{(1)}\mathbf{a}_j$), C_N-the event (All sums are zeros, i.e., $D_j = 0\ \forall\ j$.

To identify the above parameters with the previously mentioned quantities, it is easy to see from the definitions that $P(C_N|B_I)$ is identical to $p(I,N)$ of Eq. (7) and, consequently, $p(K,N)$ is identical to $P(C_N|A_1)$.

Thus we have two events: A_1-there are nonzero elements in the vector S. A_0-there are only zero elements in the vector S. *A priori* we have $P(A_1)$ and $P(A_0)$. It might be that $P(A_1)$ is much larger than $P(A_0)$. However, on the basis of the test, we have the assumption that $P(A_1)/P(A_0)$ is finite. Also, although γ may be as large as we choose, it nevertheless has to be finite. The estimation of the value of γ $[P(A_1)/P(A_0) \leqslant \gamma]$ is determined by our confidence in the screening of the important parameters. The more confidence we have in our screening procedure, the lower is the value of γ. The value α_N in Eq. (17) is bounded by

$$\alpha_N \leqslant R(K,N), \quad K \text{ even}, \tag{18a}$$

$$\alpha_N \leqslant T(K,N), \quad K \text{ odd}. \tag{18b}$$

It can now be seen from Eq. (17) and the following discussion that

$$P(A_1|C_N) \leqslant P(C_N|A_1)[P(A_1)/P(A_0)] \leqslant P(C_N|A_1)\gamma$$
$$= \sum_{I=1}^{K} P(C_N|B_I)\gamma = \sum_{I=2}^{K} p(I,N)\gamma$$
$$= P(K,N)\gamma \leqslant \alpha_N \gamma. \tag{19}$$

Thus the rejection test can be stated as follows: the assumption is that the vector S fulfills Eq. (5) and has $0 < I \leqslant K$ nonzero components. $P(K,N)\gamma$ is an upper bound for the probability of the assumption being true. If $P(K,N)\gamma$ is smaller than a predetermined value β, we reject that assumption. What is actually checked against β is $R(K,N)\gamma$ or $T(K,N)\gamma$, yet if they are smaller than β, so is $P(K,N)\gamma$.

A simple example of the test is given in Appendix A.

3. COMBINATION OF VECTORS

Let $\mathbf{S}_n = (s_1,s_2,...,s_n)$ be an n-dimensional vector. We define a combination of a vector as a number obtained from $\mathbf{S}_n$ by adding and/or subtracting all its terms, i.e., let $\mathbf{a}_n = (a_1,...,a_n)$ be a vector such that $|a_i| = 1, i = 1,...,n$, then a combination of $\mathbf{S}_n$ is the number $x = \mathbf{S}_n \cdot \mathbf{a}$. In general, there are 2^n different possible $\mathbf{a}$'s and thus 2^n possible combinations. Given a number x one may ask how many combinations of a vector U yield x. This number is denoted by $N(\mathbf{U}_n, x)$ and we can then ask: For a given x, which is the vector $\mathbf{U}_n^0$ for which $N(\mathbf{U}_n^0, x)$ is maximum. We thus denote

$$K_n(x) = \operatorname*{Max}_{\mathbf{U}_n \in E_n} \{N(\mathbf{U}_n, x)\}, \tag{20}$$

where E_n is the n-dimensional Euclidean space. If $N(\mathbf{S}_n, x) = K_n(x)$, we say that $\mathbf{S}_n$ is maximum for x.

We first show that in order to discuss $K_n(x)$, it is sufficient to consider only positive component vectors.

Lemma 1: Given a vector $\mathbf{S}_n = (s_1,...,s_n)$ and the number of combinations $\mathbf{S}_n$ that yield x is $N(\mathbf{S}_n, x)$, then for the vector $\mathbf{S}_n^a = (|s_1|,|s_2|,...,|s_n|)$, we get $N(\mathbf{S}_n^a, x) = N(S_n, x)$.

Proof: Let $\mathbf{a}$ be such that $\mathbf{S}_n \cdot \mathbf{a} = x$. Then $(|a_i| = +1)$ $\mathbf{S}_n \cdot \mathbf{a} = s_1a_1 + a_2s_2 + \cdots + a_ns_n = a_1\delta_1 s_1 + a_2\delta_2 s_2 + \cdots + a_n\delta_n|s_n| = x$, where $\delta_i = s_i/|s_i| = \pm 1$. It is clear that the vector $\mathbf{a}^a = (a_1\delta_1,...,a_n\delta_n)$ fulfills $\mathbf{S}_n^a a^a = x$. Thus to every combination of $\mathbf{s}_n$ that yields x, there exists a unique combination of $\mathbf{S}_n^a$ that yields x. The opposite is also trivial to show.

From now on, we will consider only vectors with positive components.

Lemma 2: $K_n(g)$ is independent of $g \neq 0$.

Proof: Let $\mathbf{S}_n$ be the maximum for g_1 and let E_a be the set of all vectors $\mathbf{a}$ such that $\mathbf{S}_n \cdot \mathbf{a} = g_1$. Let $g_2 \neq g_1$ and $\alpha = g_2/g_1$. Consider the vector $\mathbf{S}_n^1 = \alpha \cdot \mathbf{S}_n$. Then if $\mathbf{a} \in E_a$, we get $\mathbf{S}_n^1 \cdot \mathbf{a} = \alpha \mathbf{S}_n \cdot \mathbf{a} = \alpha g_1 = g_2$. This means that $N(\mathbf{S}_n^1, g_2) = K_n(g_1)$. If we assume that $K_n(g_2) > N(\mathbf{S}_n^1, g_2)$, then there exists a vector $\mathbf{W}_n$ and a set of $K_n(g_2)$ different vectors $\mathbf{a}^2$ such that for each $\mathbf{a}^2$, $\mathbf{W}_n\mathbf{a}^2 = g_2$. In that case, let us consider the vector $\mathbf{W}_n^2 = W_n/\alpha$; clearly for each $\mathbf{a}^2$, $\mathbf{W}_n^2 \cdot \mathbf{a}^2 = g_1$. Thus $N(W_n^2, g_1) = K_n(g_2)$ and this leads to $N(W_n^2, g_1) > N(S_n, g_2) = K_n(g_1)$, which is an obvious contradiction. Thus $K_n(g)$ is independent of g.

The conclusion of the above lemma is that the maximum number of combinations to yield a nonzero number is

the same for any such number (though a different vector may be maximum for different numbers). We will thus denote that maximum by $K_n(1)$ and, for the case that $g=0$ (which is not included in the above lemma), we denote by $K_n(0)$.

The objective of the following discussion is to find $K_n(1)$ and $K_n(0)$. Noting now that if $\mathbf{S}_n \cdot a = g$, then $\mathbf{S}_n(-a) = -g$, we also conclude that $N(\mathbf{U},g) = N(\mathbf{U},-g)$.

Lemma 3: If $\mathbf{S}_n = (s_1 \cdots s_n)$ is maximum for g, then the vector $\mathbf{S}_{n+1} = (s_1 \cdots s_n, g)$ is maximum for zero.

Proof: Let us denote any combination of $\mathbf{S}_n$ by C_n and a combination of $\mathbf{S}_{n+1}$ by C_{n+1}. Let us consider the combinations of $\mathbf{S}_{n+1}$ that equal zero.

$$C_{n+1} = C_n + g = 0 \rightarrow C_n = -g$$

or

$$C_{n+1} = C_n - g = 0 \rightarrow C_n = +g.$$

Thus we may write

$$N(\mathbf{s}_{n+1}, 0) = N(\mathbf{S}_n, g) + N(S_n, -g) = K_n(g) + K_n(g) = 2K_n(g).$$

Let us next consider any general $(n+1)$-dimensional vector $U_{n+1}(u_1 \cdots u_{n+1})$ and $\mathbf{U}_n = (u_1 \cdots u_n)$. Let us denote a combination of $\mathbf{U}_{n+1}$ by C_{n+1} and a combination of $\mathbf{U}_n$ by C_n. Then the combinations of $\mathbf{U}_{n+1}$ that equal zero may be obtained by

$$C_{n+1} = C_n + u_{n+1} = 0, \quad C_n = -u_{n+1},$$

or

$$C_{n+1} = C_n - u_{n+1} = 0, \quad C_n = u_{n+1}.$$

Thus

$$N(\mathbf{U}_{n+1}, 0) = N(\mathbf{U}_n, u_{n+1}) + N(\mathbf{U}_n, -u_{n+1}) = 2N(\mathbf{U}_n, u_{n+1}) \leqslant 2K_n(1).$$

Corollary: $K_{n+1}(0) = 2K_n(1)$.

A. Combinations that yield more than one number

In order to proceed, it is now necessary to consider sums of combinations that yield more than one number. Let $\mathbf{U}_n$ be any n-dimensional vector and let $x_p = (x_1, x_2, \ldots, x_p)$ be a set of p positive different numbers. Let, as before, $N(\mathbf{U}, x_i)$ be the number of combinations of $\mathbf{U}$ that yield x_i. We now define the quantity

$$B(\mathbf{U}, x_1, x_2, \ldots, x_p) = B(\mathbf{U}_n, \mathbf{x}_p) = \sum_{i=1}^{p} N(\mathbf{U}_n, x_i), \quad (21)$$

which is simply the total number of combinations of $\mathbf{U}$ that yield any component of $\mathbf{x}_p$. For a given vector $\mathbf{x}_p$, we may inquire for which vector $\mathbf{U}_n$, $B(\mathbf{U}_n, \mathbf{x}_p)$ is maximum and denote

$$B_m(\mathbf{x}_p) = \max_{\mathbf{U}_n}\{B(\mathbf{U}_n, \mathbf{x}_p\}.$$

Then we may ask for which $\mathbf{x}_p$ $B_m(\mathbf{x}_p)$ is maximum and denote

$$D_n(p) = \max_{\mathbf{x}_p}\{\max_{\mathbf{U}_n}(B(\mathbf{U}_n, \mathbf{x}_p))\}. \quad (22)$$

$D_n(p)$ is the maximum possible number of combinations of an n-dimensional vector that yields p positive values. Clearly $D_n(p) \leqslant p \cdot K_n(1)$.

Equality will imply that a certain vector can simultaneously be a maximum for p different numbers. At this point, an important observation should be made. Suppose we divide $\mathbf{x}_p$ into two separate vectors $\mathbf{x}_{p_1}$, $\mathbf{x}_{p_2}$ such that $p_1 + p_2 = p$. Then by virtue of the fact that

$$\max\{B(\mathbf{U}, \mathbf{x}_p)\} = \max\{B(\mathbf{U}, \mathbf{x}_{p_1}) + B(\mathbf{U}, \mathbf{x}_{p_2})\} \leqslant \max\{B(\mathbf{U}, \mathbf{x}_{p_1})\} + \max\{B(\mathbf{U}, \mathbf{x}_{p_2})\},$$

we obtain the relation

$$D_n(p) \leqslant D_n(p_1) + D_n(p_2), \quad p_1 + p_2 = p. \quad (23)$$

With the aid of the above concepts, we now prove the following lemma:

Lemma 4: Let $\mathbf{S}_{2n-1}$ be a $(2n-1)$-dimensional vector with the property

$$D_{2n-1}(p) = B(\mathbf{S}_{2n-1}, g, 3g, 5g, \ldots, (2p-1)\cdot g).$$

That is, the maximum possible combinations to get p positive different numbers for any $(2n-1)$-dimensional vector is equal to the number of combinations of the vector $\mathbf{S}_{n-1}$ that yields any one of the p positive different numbers g, $3g, \ldots, (2p-1)\cdot g$. Note that putting $p=1$ implies that $\mathbf{S}_{2n-1}$ is also maximum for g. Let $\mathbf{S}_{2n+1}$ be the vector obtained from $\mathbf{S}_{2n-1}$ by adding two components each of which equals g, i.e., $\mathbf{S}_{2n+1} = (\mathbf{S}_{2n-1}, g, g)$, then $\mathbf{S}_{2n+1}$ also has the property

$$D_{2n+1}(p) = B(\mathbf{S}_{2n+1}, g, 3g, \ldots, (2p-1)\cdot g). \quad (24)$$

Proof: Let us first consider the combinations of $\mathbf{S}_{2n+1}$ that yield any component of the set $(g, 3g, \ldots, (2p-1)\cdot g)$. Let us denote by C_{2n+1} a combination of $\mathbf{S}_{2n+1}$. Then for each combination of $\mathbf{S}_{2n-1}(C_{2n-1})$, there are four combinations of $\mathbf{S}_{2n+1}$, namely, $C_{2n+1} = C_{2n-1} \pm (g \pm g)$. Thus g is obtained when

$$C_{2n+1} = C_{2n-1} + (g+g) = g \rightarrow C_{2n-1} = -g$$

or

$$C_{2n+1} = C_{2n-1} + (g-g) = g \rightarrow C_{2n-1} = g$$

or

$$C_{2n+1} = C_{2n-1} - (g-g) = g \rightarrow C_{2n-1} = g$$

or

$$C_{2n+1} = C_{2n-1} - (g+g) = g \rightarrow C_{2n-1} = 3g$$

Following the same procedure for $3g, 5g, \ldots, (2p-1)\cdot g$, one verifies that any of these values is obtained by a different combination of $\mathbf{S}_{2n+1}$, when C_{2n-1} takes any one of the following values:

$-g$	g	g	$3g$	to get	g
g	$3g$	$3g$	$5g$	to get	$3g$
$3g$	$5g$	$5g$	$7g$	to get	$5g$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
$(2p-3)g$,	$(2p-1)g$,	$(2p-1)g$,	$(2p+1)g$	to get	$(2p-1)g$.

With the above considerations and using the maximum property of $\mathbf{S}_{2n-1}$, we may now write

$$B(\mathbf{S}_{2n+1}, g, 3g, ..., (2p-1)g) = B(\mathbf{S}_{2n-1}, g, 3g, ..., (2p-3)g) + B(\mathbf{S}_{2n-1}, g, 3g, ..., (2p-1)g) + B(\mathbf{S}_{2n-1}, -g, 3g, ..., (2p+1)g) = D_{2n-1}(p-1) + 2D_{2n-1}(p) + D_{2n-1}(p+1). \tag{25}$$

Equation (24) states the fact that the p different values of C_{2n+1} are obtained from $4p$ values of C_{2n-1}, which are not all different. $4p$ is divided into four sets of $p-1, p, p, p+1$, values such that in each set, all values are different. Now let $\mathbf{U}_{2n+1}$ be any $(2n+1)$-dimensional vector and $(x_1,...,x_p)$ any set of p *positive different* numbers. We then wish to show that

$$B(\mathbf{U}_{2n+1}, x_1, ..., x_p) \leqslant B(\mathbf{S}_{2n+1}, g, 3g, ..., (2p-1)g).$$

Let us denote the last two components of $\mathbf{U}_{2n+1}$ by a and b and assume with no loss of generality $a \geqslant b$. And let us denote a combination of the first $(2n-1)$ terms of $\mathbf{U}_{2n+1}$ by C_{2n-1}. Then a number x_1 may be obtained as a combination of $\mathbf{U}_{2n+1}$ by the following four possibilities:

$$\begin{aligned} &C_{2n-1} + (a+b) = x_1, \\ &C_{2n-1} + (a-b) = x_1, \\ &C_{2n-1} - (a-b) = x_1, \\ &C_{2n-1} - (a+b) = x_1. \end{aligned}$$

Denoting $a+b=\beta>0$ and $a-b=\alpha\geqslant 0$ $(\alpha<\beta)$, we get for $x_1,...,x_p$

$$\begin{aligned} &C_{2n-1} + \beta = x_1 \to C_{2n-1} = x_1 - \beta = b_1^1, \\ &C_{2n-1} + \alpha = x_1 \to C_{2n-1} = x_1 - \alpha = b_2^1, \\ &C_{2n-1} - \alpha = x_1 \to C_{2n-1} = x_1 + \alpha = b_3^1, \\ &C_{2n-1} - \beta = x_1 \to C_{2n-1} = x_1 + \beta = b_4^1, \\ &C_{2n-1} + \beta = x_2 \to C_{2n-1} = x_2 - \beta = b_1^2, \\ &C_{2n-1} + \alpha = x_2 \to C_{2n-1} = x_2 - \alpha = b_2^2, \\ &\quad\vdots \qquad\qquad\qquad\quad \vdots \\ &C_{2n-1} + \beta = x_p \to C_{2n-1} = x_p - \beta = b_1^p, \\ &C_{2n-1} + \alpha = x_p \to C_{2n-1} = x_p - \alpha = b_2^p, \\ &C_{2n-1} - \alpha = x_p \to C_{2n-1} = x_p + \alpha = b_3^p, \\ &C_{2n-1} - \beta = x_p \to C_{2n-1} = x_p + \beta = b_4^p. \end{aligned}$$

We see that the p values of $x_1,...,x_p$ are obtained in different combinations when C_{2n-1} takes any one of the $4p$ values of $b_1^1 b_2^1,...,b_4^p$. If all 4 values are different, then we immediately obtain

$$\begin{aligned} B(\mathbf{U}_{2n+1}, x_1, ..., x_p) &= B(\mathbf{U}_{2n-1}, b_1^1, ..., b_4^p) \leqslant D_{2n-1}(4p) \leqslant D_{2n-1}(p-1) + 2D_{2n-1}(p) + D_{2n-1}(p+1) \\ &= B(\mathbf{S}_{2n+1}, g, ..., (2p-1)g). \end{aligned}$$

Thus we have to find what are the maximum possible identical (in absolute value) numbers in the set $b_1^1 \cdots b_4^p$. This under the constraints that $x_1 \cdots x_p$ are all positive and different and $\beta > \alpha \geqslant 0$. A close examination of all the possibilities reveals that the maximal number of identical (in absolute value) terms is obtained if the following choice is made:

$$\begin{array}{lllll} x_1 = a = b = s > 0 \ \beta = 2s \ \alpha = 0 & \to b_1^1 = -s & b_2^1 = s & b_3^1 = s & b_4^1 = 3s \text{ then} \\ x_2 = x_1 + 2s = 3s & \to b_1^2 = s & b_2^2 = 3s & b_3^2 = 3s & b_4^2 = 5s \\ x_3 = x_2 + 2s = 5s & \to b_1^3 = 3s & b_2^3 = 5s & b_3^3 = 5s & b_4^3 = 7s \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_p = x_{p-1} + 2s = (2p-1)s & \to b_1^p = (2p-3)s, & b_2^p = (2p-1)s, & b_3^p = (2p-1)s, & b_4^p = (2p+1)s. \end{array}$$

(The choice $x_1 = a = b$ is made with no loss of generality and any other x_i could be taken instead. The same holds for every step.) Thus we may write, arranging $b_1^1 \cdots b_4^p$ in four groups,

$$\begin{aligned} &B(\mathbf{U}_{2n+1}, x_1, ..., x_p) \\ &\quad = B(\mathbf{U}_{2n-1}, b_1^1, b_1^3, b_1^4, ..., b_1^p) \\ &\qquad + B(\mathbf{U}_{2n-1}, b_2^1, b_2^2, ..., b_2^p) \\ &\qquad + B(\mathbf{U}_{2n-1}, b_3^1, b_3^2, ..., b_3^p) \\ &\qquad + B(\mathbf{U}_{2n-1}, b_1^2, b_4^1, b_4^1, b_4^2, ..., b_4^p) \\ &\quad \leqslant D_{2n-1}(p-1) + 2D_{2n-1}(p) + D_{2n-1}(p+1) \\ &\quad = B(\mathbf{S}_{2n+1}, g, 3g, ..., (2p-1)g). \end{aligned} \tag{26}$$

Lemma 5: If the vector $\mathbf{S}_n = (s_1,...,s_n)$ has the property $B(\mathbf{S}_n, 0, 2g) = D_n(2)$, then the vector $\mathbf{S}_{n+1} = (s_1 \cdots s_n, g)$ is maximum for g.

Proof: Again with the notation of C_{n+1}, C_n being combinations of $\mathbf{S}_{n+1}$ and $\mathbf{S}_n$ correspondingly, we obtain g from $\mathbf{S}_{n+1}$ by

$$C_{n+1} = C_n + g = g \to C_n = 0$$

or

$$C_{n+1} = C_n - g = g \to C_n = 2g.$$

Thus

$$N(\mathbf{S}_{n+1}, g) = N(\mathbf{S}_n, 0) + N(\mathbf{S}_n, 2g)$$
$$= B(\mathbf{S}_n, 0, 2g) = D_n(2),$$

and it is easy to verify (as in Lemma 3) that the number of combinations of any $(n+1)$-dimensional vector that yields a positive given number cannot exceed $D_n(2)$.

Theorem: (a) For any odd integer $(2n-1)$, the identity vector

$$\mathbf{S}_{2n-1}(g) = (g, g, g, ..., g, g)$$

(all components equal to g) is maximum for g and

$$B(\mathbf{S}_{2n-1}(g), g, 3g, ..., (2p-1)g) = D_{2n-1}(p), \tag{27}$$

and the vector

$$\mathbf{S}_{2n-1}(0) = (2g, g, g, ..., g)$$

(first component equals $2g$, all others $= g$) is maximum for zero.

(b) For any even integer $2n$, the identity vector

$$\mathbf{S}_{2n}(0) = (g, g, ..., g) \quad s_i \equiv g \; i = 1,...,2n$$

is maximum for zero and

$$\mathbf{S}_{2n}(g) = (2g, g, ..., g), \quad s_1 = 2g, \; s_i \equiv g, \; i \neq 1$$

is maximum for g.

Proof: The above theorem is proved using the following induction.

Let us assume that the vectors $\mathbf{S}_{2n-1}(g)$ and $\mathbf{S}_{2n-1}(0)$ are given as in the theorem with the properties stated in the theorem. Given this, we will prove that $\mathbf{S}_{2n}(0)$ and $\mathbf{S}_{2n}(g)$ have the properties stated in the theorem and the identity vectors $S_{2n+1}(g) = (g,g,...,g)$, $s_i = g$ and $S_{2n+1}(0)$ $= (2g,g,...,g)$, $s_1 = 2g$, $s_i = g$, $i = 2,...,2n+1$ have the same properties as $S_{2n-1}(g)$ and $S_{2n-1}(0)$.

From Eq. (21), it is clear that

$$B(\mathbf{S}_{2n-1}(g), g, 3g) = D_{2n-1}(2);$$

also

$$B(\mathbf{S}_{2n-1}(g), g, 3g) = N(\mathbf{S}_{2n-1}(g), g) + N(\mathbf{S}_{2n-1}(g), 3g)$$
$$= \binom{2n-1}{n} + \binom{2n-1}{n+1} = \frac{2n}{n+1}\binom{2n-1}{n}.$$

Let us now consider

$$B(\mathbf{S}_{2n-1}(0),0,2g) = N(\mathbf{S}_{2n-1}(0), 0) + N(\mathbf{S}_{2n-1}(0), 2g)$$
$$= 2\binom{2n-2}{n} + \left[\binom{2n-2}{n-1} + \binom{2n-2}{n+1}\right]$$
$$= \frac{2n}{n+1}\binom{2n-1}{n}.$$

Thus

$$B(\mathbf{S}_{2n-1}(0), 0, 2g) = D_{2n-1}(2). \tag{28}$$

From Eq. (28) and Lemma 5, we immediately conclude that the vector

$$\mathbf{S}_{2n}(g) = \overbrace{(2g,g,g,g,...,g)}^{2n} = (\mathbf{S}_{2n-1}(0), g)$$

is maximum for g. And since $\mathbf{S}_{2n-1}(g)$ is maximum for g, we conclude, using Lemma 3, that the vector $\mathbf{S}_{2n}(0) = \overbrace{(g,...,g)}^{2n}$ is maximum for zero.

We now note that

$$K_{2n}(0) = \binom{2n}{n}$$

and

$$K_{2n}(1) = \binom{2n}{n+1}; \text{ thus } K_{2n}(1) < K_{2n}(0).$$

Also we get

$$N(\mathbf{S}_{2n}(0), 2g) = \binom{2n}{n+1} = K_{2n}(1).$$

Thus $\mathbf{S}_{2n}(0)$ is both maximum for zero and $2g$. Thus $B(\mathbf{S}_{2n}(0))(\mathbf{S}_{2n}(0), 0, 2g) = D_{2n}(2)$. And again using Lemma 5, we get the result that

$$\mathbf{S}_{2n+1}(g) = \overbrace{(g,g,...,g)}^{2n+1}, \quad s_i = g, \; i = 1,...,2n+1$$

is maximum for g. And using Lemma 3 on $\mathbf{S}_{2n}(g)$, we get that

$$S_{2n+1}(0) = (2g, g, ...g), \quad s_1 = 2g, \; s_i = g, \; i = 2,...,2n+1,$$

is maximum for zero. Furthermore, using Lemma 4 we see that $\mathbf{S}_{2n+1}(0)$ has the property

$$B(\mathbf{S}_{2n+1}(g), g, 3g, ..., (2p-1)g) = D_{2n+1}(p).$$

This concludes the transfer from $(2n-1)$ to $(2n+1)$ through $2n$. Let us now consider the case $n = 2$, $2n - 1 = 3$. Clearly, $\mathbf{S}_3(g) = (g,g,g)$. It is maximum for g. And $B(\mathbf{S}_3(g), g, 3g) = 4$, which is the largest possible number. Thus, $B(\mathbf{S}_3(g),g,3g,5g,...,(2p-1)g) = D_3(p)$. Also, $\mathbf{S}_3(0) = (2g,g,g)$ is maximum for zero!

Starting from $(2n-1)$, we may now proceed by the proven induction to see that the theorem holds for every n.

Conclusion: From the above results, we have

$$K_{2n}(0) = \binom{2n}{n}, \tag{29a}$$

$$K_{2n}(1) = \binom{2n}{n+1}, \quad K_{2n}(0) > K_{2n}(1), \tag{29b}$$

$$K_{2n+1}(0) = 2\binom{2n}{n+1}, \tag{29c}$$

$$K_{2n+1}(1) = \binom{2n+1}{n+1}, \quad K_{2n+1}(1) > K_{2n+1}(0), \tag{29d}$$

$$D_{2n-1}(p) = \sum_{k=n+1}^{2n+1}\binom{2n+1}{k}. \tag{29e}$$

4. APPLICATION OF THE TEST TO A WASTE DISPOSAL PROBLEM

The test described above has been applied to a heat transfer problem in a hypothetical waste repository.[6] The temperature profile within salt in the repository was determined to be the quantity of most interest from a design point of view for this test problem. This response was selected to provide the focal point of the sensitivity demonstration problem. A near-field single-level storage repository of spent fuel in salt[6] was chosen to model the thermal behavior of a generic design. The disposal horizon of the repository was assumed to have a thermal loading density of 60 kW/acre (15 W/m^2) and its time behavior, as well as all the input parameters, are the same as those used in ORNL/ENG/TM-14,[10] and are also given in Appendix B. The computer code used

to calculate the thermal conductivity for this analysis was HEATING 5.[11] This code uses finite-difference methods to solve heat transfer equations and solves both steady-state and transient heat conduction problems with spatially- and temperature-dependent thermal properties and time-dependent heat generation rates.

For this problem, 49 input parameters were considered. These parameters included all the material properties, boundary conditions, and modeling parameters in the repository (see Appendix B). The sensitivity coefficients S_i, $i = 1,...,49$, were estimated for $R_0 = T_s$, where T_s is the midplane temperature at the interface between the spent fuel assembly and the salt five years after burial.

The numerical accuracy of the temperature profile calculation by HEATING 5 with convergence criterion of 10^{-5} was about 0.01 °F. Since the value of R_0 is 155.69 °F (68.72 °C), the error in any response change is no better than $\delta R/R_0 = +6.42\times10^{-5}$. Therefore, the minimum value of S_i that was termed to be useful was $+6.42\times10^{-3}$ and importance in terms of sensitivity coefficients was defined to be $|S_i| \geqslant 0.01$. Eleven sensitivity coefficients were assumed to be important.

The exact sensitivity coefficients for these parameters were then obtained exactly by rerunning the code for each individual parameter change. These were the important sensitivities:

(1) sensitivity to the salt density ρ,
(2) sensitivity to the heat capacity C_p,
(3) sensitivity to the salt conductivity at 32 °F (0 °C) [k(32°)],
(4) sensitivity to the salt conductivity at 122 °F (50 °C) [k(122°)],
(5) sensitivity to the salt conductivity at 212 °F (100 °C) [k(212°)],
(6) sensitivity to the initial temperature at 2000 ft (610 m) [T_0 (2000 ft)],
(7) sensitivity to the distance between canisters D_s,
(8) sensitivity to the canister radius R_s,
(9) sensitivity to the thermal loading P,
(10) sensitivity to the initial temperature at 2250 ft (686 m) [T_0 (2250 ft)],
(11) sensitivity to the salt thermal emissivity h_r.

[Note: For (3), (4), and (5), the conductivity is given at discrete temperature points.]

In the second stage of analysis, these eleven parameters were held constant in all cases, and the rest of the parameters $(M - P_1) = 38$ were then randomly changed. Eight such runs were made to analyze the response temperatures for these runs.

Thus values obtained for $R' = \delta R/R_0 \times 100$ for this set of runs were $R'^{(1)} \simeq 0$, $R'^{(2)} \simeq 0$, $R'^{(3)} \simeq 6\times10^{-5}$, $R'^{(4)} \simeq -6\times10^{-5}$, $R'^{(5)} \simeq 0$, $R'^{(6)} \simeq 0$, $R'^{(7)} \simeq 0$, and $R'^{(8)} \simeq 0$.

The practically R' can be considered as zero. The rejection test was applied to this case. By this test, we have found from the upper bound to the probability $-\alpha_N$ that there are sensitivity components different from zero which will yield eight runs with $R' \simeq 0$ less than 0.5% (see Table I).

Due to our experience with this type of a problem, we were quite confident that we had screened out the important sensitivities. Thus we estimated γ to be less than 2. Thus by the rejection test we found that the probability that there are nonzero sensitivity elements is less than 0.01. We rejected this possibility. We have randomly chosen some of the input parameters and calculated their sensitivities. In all our checks, we indeed found negligible sensitivity values.

5. SUMMARY

We have found expressions for the probabilities that a certain unknown vector has nonzero components. It was found that by applying the rejecting test, it is enough to obtain a very low probability for the existence of a vector with nonzero components. It was found that for 10 trials, the probability of obtaining nonzero component vectors is less than 1.1×10^{-3} and for 16 trials the probability is less than 1.6×10^{-5} with a rapid reduction in the probability with the increased number of trials. It should be noted that for a vector with a number of components greater than the number of trials, the probability is almost *independent* of the number of the vector's components. Namely, the advantage of the rejection test lies in the fact that by making very few trials, we gain information on the input parameters involved regardless of how many we have.

The input for the N runs in the rejection test are almost identical to those used in the "statistical screening" procedure.[7–9] A discussion on the "statistical screening" procedure was published by Perey.[12] The difference between the runs in the two procedures pertains only to the variables that are being subjected to the random increments. In the rejection test they are the variables being checked for having zero sensitivity coefficients, while in the "statistical screening" procedure they are all the variables that one believes can be important, namely, those having large sensitivity coefficients. Thus, if $\mathbf{S}\cdot\mathbf{a} \neq 0$, the rejection test runs are not completely wasted. These runs, in some cases, can be used for the screening procedure on the variable tested.

In the presented test, we required orthogonality between the vectors $\mathbf{S}$ and $\mathbf{a}_j$. In a practical application, one will always find $\mathbf{S}\cdot\mathbf{a}_j = \epsilon_j$, where $|\epsilon_j|$ is a small but finite quantity. We believe that setting $\epsilon_j = 0$ when $|\epsilon_j|$ is below some small threshold value will not invalidate the test. So far we have not succeeded in supporting this assertion by a rigorous mathematical proof. This question is a subject for further investigations.

It may be interesting to note that the problem of the maximum combinations of a vector underlying the rejection test has interpretations in geometry and the theory of games. In the geometrical aspect, the problem can be interpreted as finding a vector (with nonzero elements) in n-dimensional Euclidean space; that is, perpendicular to the maximum number of edges of an n-dimensional cube centered at the origin. The edges of an n-dimensional cube centered at the origin are of the form $\mathbf{a} = (a_1,a_2,...,a_n)$, $a_i = \pm 1$. We want to find a vector $\mathbf{S} = (s_1,s_2,...,s_n)$, $s_i \neq 0$, such that $\mathbf{S}\cdot\mathbf{a} = 0$ for the maximum number of edges. The theorems discussed in this paper solve this problem. Thus, the maximum number of cube edges that will be orthogonal to the vector is given by Eqs. (29a) and (29c). The form of such a vector will be

$\mathbf{S} = (g, g, ..., g)$ for n even,

$\mathbf{S} = (2g, g, ..., g)$ for n odd

(this is not necessarily the only possibility). For example, in a three-dimensional space, the vector $\mathbf{S} = (2,1,1)$ is perpendicular to $(1, -1, -1) - (1,1,1)$. This is the maximum number of cube edges that can be perpendicular to any vector with nonzero components.

In the theory of strategic games, we may consider the following game. A set of n-steps $(s_1,...,s_n)$ is chosen. N "scouts" start from the origin and follow the n steps in two directions; (b) steps must be followed in the order that they appear in the sequence; (c) no two "scouts" are allowd to follow exactly the same route; (d) all scouts must finish their route at a predetermined point of distance g from the origin. The problem is now how to choose the set of steps such that the number of "scouts," N, is maximum. The solution of this problem is given in Eqs. (29b) and (29d).

ACKNOWLEDGMENTS

Thanks for stimulating discussions are due to E. M. Oblow and F. G. Perey of Oak Ridge National Laboratory and M. Kalai of the Hebrew University, Jerusalem. Thanks are also due to the referee for his useful comments.

APPENDIX A: A SIMPLE EXAMPLE FOR THE REJECTION TEST

Consider a die that can be either normal die or a loaded die. We thus have

$$A_1 = \{\text{normal die}\} \tag{A1}$$

or

$$A_0 = \{\text{loaded die}\}. \tag{A2}$$

Let us assume that

$$P(A_1)/P(A_0) \leqslant \gamma = 10^3, \tag{A3}$$

thus assuming that the chances of a fair play are large.

Let us choose a rejection criterion $\beta = 10^{-5}$, and E = experiment: N throws of the die. C_N-in all throws, face up shows six dots. (A4)

Thus we have

$$P(C_N|A_0) = 1,$$
$$P(C_N|A_1) = (\tfrac{1}{6})^N. \tag{A5}$$

Demand

$$P(A_1|C_N) \leqslant 10^{-5}. \tag{A6}$$

Thus

$$(\tfrac{1}{6})^N \times 10^3 \leqslant 10^{-5}, \quad N > 10.$$

Thus, throw the die 11 times. If C_{11} occurs, reject A_1. Hence decide the die is loaded.

APPENDIX B: THE INPUT PARAMETER DATA FOR THE PROBLEM

This appendix includes all input parameters used in the problem design. The parameters are related to material properties, initial values, and modeling parameters.

All temperatures are given in °F, distances in ft, conductivities in Btu/yr ft °F, densities in lb/ft^3, heat capacities in Btu/lb °F, and geothermal flux in Btu/yr ft^2:

(1) concrete conductivity	4380,
(2) concrete density	125,
(3) concrete C_p	0.21,
(4) soil conductivity	8760,
(5) soil density	120,
(6) soil C_p	0.2,
(7) salt density	165,
(8) salt C_p	0.2,
(9) earth surface temperature	60,
(10) geothermal flux	140,
(11) radiative coefficient	4.886×10^{-6}

(Note: Stefan–Boltzman constant is multiplied by the Radiation shape factor.)

Salt conductivity	Temperature
(12) 30 900	32
(13) 25 400	122
(14) 21 300	212
(15) 18 200	302
(16) 15 800	392
(17) 14 000	482
(18) 12 600	572
(19) 11 700	662
(20) 10 500	752

Initial temperature	Depth (ft)
(21) 60	0
(22) 63.20	200
(23) 66.39	400
(24) 69.59	600
(25) 72.78	800
(26) 73.77	1000
(27) 75.01	1250
(28) 76.25	1500
(29) 77.49	1750
(30) 78.74	2000
(31) 79.99	2250
(32) 81.24	2500
(33) 82.50	2750
(34) 83.76	3000
(35) 85.02	3250
(36) 86.29	3500
(37) 87.56	3750
(38) 88.83	4000
(39) 90.11	4250
(40) 91.39	4500
(41) 92.68	4750
(42) 93.97	5000

(43) distance between spent fuel assemblies	40,
(44) canister radius	0.6667,
(45) repository depth	5000,
(46) overburden thickness	800,
(47) room width	10,
(48) concrete thickness	8,
(49) room thickness	25.

[1]C. R. Weisbin, R. W. Peelle, J. H. Marable, P. Collins, E. Kujauski, E. Greenspan, and G. de Saussure, "Sensitivity and Uncertainty Analysis of Reactor Performance Parameters," in *Advances in Nuclear Science and Technology, Vol. 14* (Plenum, New York, 1982).
[2]R. H. Myers, *Response Surface Methodology* (Allyn and Bacon, Boston, 1971).
[3]N. D. Cox, Nucl. Sci. Eng. **64**, 258 (1977).
[4]J. K. Vaurio and C. Mueller, "A Probabilistic/Deterministic Procedure for Analyzing LMFBR Core Disruptive Accidents," *Proceedings of the International Meeting on Fast Reactor Safety and Related Physics* (Chicago, 1976).
[5]M. D. McKay, W. J. Conover, and D. E. Whiteman, "Report on the Application of Statistical Techniques to the Analysis of Computer Codes," LA-NUREG-6526-MS, Los Alamos Scientific Laboratory, 1976.
[6]Y. Ronen, J. L. Lucius, and E. M. Oblow, "A Statistical Sensitivity Analysis of a Simple Nuclear Waste Repository Model," ORNL/TM-7310, Oak Ridge National Laboratory, 1980.
[7]C. Durston, T. Krieger, and D. Albright, "Sensitivity Analysis Method Application to Code Validation," Brookhaven National Laboratory Report BNL-NUREG-23930, 1977.
[8]T. J. Krieger *et al.*, Trans. Amer. Nucl. Soc. **23**, 515 (1978).
[9]T. J. Krieger, Nucl. Sci. Eng. **81**, 281 (1982).
[10]F. G. Perey, "Contributions to Screening Methodology," Oak Ridge National Laboratory Report ORNL-5744, 1981.
[11]R. A. Just, "Heat Transfer Studies in Salt and Granite," ORNL/ENG/TM-14, Oak Ridge National Laboratory, 1978.
[12]W. D. Turner, D. C. Elrod, and I. Simon-Tov, "HEATING 5—An IBM 360 Heat Conduction Program," ORNL/CSD/TM-15, Oak Ridge National Laboratory, 1977.

A dense set of cyclic vectors for quantum field polynomial algebras

Wulf Driessler and Stephen J. Summers
Fachbereich Physik, Universität Osnabrück, D-4500 Osnabrück, Federal Republic of Germany

(Received 13 September 1982; accepted for publication 10 December 1982)

It is shown that in the Hilbert space of a quantum field theory with a nonzero mass gap there exists a dense set of vectors, each entire analytic for the energy–momentum operators, that are cyclic for the polynomial algebra $\mathscr{P}(\mathbb{R}^d)$ [and for the local polynomial algebras $\mathscr{P}(\mathscr{O})$, for any nonempty $\mathscr{O} \subseteq \mathbb{R}^d$]. It is proven that for every vector Φ from this dense set there exists an element $Q \in \mathscr{P}(\mathbb{R}^d)$ such that $Q\Phi = \Omega$, where Ω is the vacuum, and $Q\Omega = 0$. A similar, stronger result is proven for free field theories (including mass zero).

PACS numbers: 11.10.Cd

I. INTRODUCTION

As part of a program[1] intending to prove, for a large class of quantum field theories, that symmetric field operators are essentially self-adjoint on their usual domain of definition and that their unique self-adjoint extensions are local, this paper proves a technical result, which we believe to be of independent interest in its own right, which is employed crucially in that endeavor. We first indicate what this technical result is and point out some of its immediate consequences. There shall surely be other applications for this result besides the one being developed in Ref. 1.

If $\mathscr{P}(\mathbb{R}^d)$ denotes the polynomial algebra of smeared field operators, we shall call $P \in \mathscr{P}(\mathbb{R}^d)$ vacuum reducible if there exists a $Q \in \mathscr{P}(\mathbb{R}^d)$ such that $Q\Omega = 0$ and $QP\Omega = \Omega$, where Ω signifies the vacuum vector (assumed unique). Q will be called an annihilator of P. We shall show that in a quantum field theory with a nonzero mass gap there are enough vacuum reducible $P \in \mathscr{P}(\mathbb{R}^d)$ so that the set

$$\hat{\mathscr{E}} = \{P\Omega \,|\, P \in \mathscr{P}(\mathbb{R}^d) \text{ is vacuum reducible}\}$$

is dense in the Hilbert space $\mathscr{H}_W$ associated with the quantum field. It then follows easily that every vector in $\hat{\mathscr{E}}$ is cyclic for $\mathscr{P}(\mathbb{R}^d)$ (since Ω is). Furthermore, every vector in $\hat{\mathscr{E}}$ that is an analytic vector for the energy (and we show that this subset is dense) is also cyclic for the local polynomial algebra $\mathscr{P}(\mathscr{O})$, for any nonempty, open $\mathscr{O} \subseteq \mathbb{R}^d$.

We recall that it is well known that every vector in a Hilbert space $\mathscr{H}$ carrying an irreducible *-algebra $\mathscr{A}$ of bounded operators is cyclic for $\mathscr{A}$. Indeed, the cyclicity of every vector for $\mathscr{A}$ is necessary and sufficient for the irreducibility of $\mathscr{A}$. The same, however, cannot be said for a *-algebra $\mathscr{P}$ of unbounded operators, viewed as being defined on a common, dense, invariant, linear domain D. To make this clear, let us define the commutant $\mathscr{P}'$ of the algebra $\mathscr{P}$ as all those bounded operators B for which

$$\langle \Phi, BP\Psi \rangle = \langle P^*\Phi, B\Psi \rangle, \quad \forall\ \Phi, \Psi \in D, \ \forall\ P \in \mathscr{P}\ .$$

This definition provides the largest commutant $\mathscr{P}'$ (and $\mathscr{P}'$ is not necessarily an algebra). One has taken to calling such algebras $\mathscr{P}$ irreducible[2–5] if $\mathscr{P}'$ is composed solely of multiples of the identity operator I on $\mathscr{H}$, in analogy to one of the many equivalent formulations of irreducibility for algebras of bounded operators. But, if $\mathscr{M} \subset D$ is a linear manifold that reduces $\mathscr{P}$, i.e., $\mathscr{P}\mathscr{M} \subset \mathscr{M}$, it is not true in general that the self-adjoint projection $E_{\mathscr{M}}$ onto the closure of $\mathscr{M}$ in $\mathscr{H}$ is in $\mathscr{P}'$. Thus, even if $\mathscr{P}$ is "irreducible," there could be a proper reducing subspace for $\mathscr{P}$ in $\mathscr{H}$. Concretely, there are examples of irreducible $\mathscr{P}$ for which "many" vectors $\Phi \in D$ are not cyclic. (For further details on these matters, see Refs. 5 and 4.)

Therefore, although we are concerned here only with irreducible algebras $\mathscr{P}$, it is not a triviality in general to produce a dense set of cyclic vectors for $\mathscr{P}$, much less show they are generated by vacuum reducible operators, which is what we shall actually need in Ref. 1. This will be possible for the particular algebras handled here either because of the straightforward structure of the algebra of the canonical commutation relations in Fock space or because of the special information provided by the mathematical structure of the general theory of relativistic quantum fields. In fact, as we show in Sec. II, it is easy to demonstrate in the former case, for the examples of quantum mechanics with arbitrarily many degrees of freedom and of free (boson) quantum field theories with mass $m \geqslant 0$, that every operator in $\mathscr{P}$ is vacuum reducible. But the proof for interacting fields is not quite so simple. In Sec. III we introduce the definitions, notation, and some previously known results that we need to prove, for a real, scalar quantum field theory satisfying Wightman's axioms generalized to localizable ultradistributions and having a nonzero mass gap, that the set $\hat{\mathscr{E}}$ defined above is dense. An exact statement of the main result is given in Theorem 4.1. The vectors of $\hat{\mathscr{E}}$ that we explicitly construct in Sec. V are all entire analytic for the energy-momentum operator, enabling us to prove (in Sec. IV) the consequences mentioned above. The proof of Theorem 4.1 is given in Sec. V.

We comment that the requirement of a nonzero mass gap is an artifice of the proof in Sec. V (it is not necessary for free fields in the proof in Sec. II) and that, although we restrict outselves to giving details only for the real scalar field, a similar argument can be applied to arbitrary spin fields.

II. QUANTUM MECHANICS AND FREE FIELDS

The aim of this chapter is to prove the existence of a dense set $\hat{\mathscr{E}}$ of cyclic vectors for *-algebras of unbounded operators with particularly simple algebraic relations—the polynomial algebra of position and momentum operators in

 0022-2488/83/122809-11$02.50

quantum mechanics with arbitrarily many degrees of freedom and the polynomial algebra of (generalized) free scalar quantum fields—by showing that each operator P in these algebras is vacuum reducible, in the sense defined in the Introduction. We shall also show that Wick monomials of free scalar fields are vacuum reducible. The generalization to free quantum fields with other spins will be clear after one has seen the proof in the scalar case.

We begin with the simplest algebra—that of the polynomial algebra of the position and momentum operators for one degree of freedom. Let A and $A^\dagger$ be the maps of $\mathscr{S}(\mathbb{R})$, the Schwartz space of tempered test functions on $\mathbb{R}$, into itself given by

$$A = 2^{-1/2}\left(x + \frac{d}{dx}\right), \quad A^\dagger = 2^{-1/2}\left(x - \frac{d}{dx}\right),$$

so that

$$x = 2^{-1/2}(A + A^\dagger) \quad \text{and} \quad i\frac{d}{dx} = i2^{-1/2}(A - A^\dagger). \tag{2.1}$$

They are to be understood as linear operators on the dense domain $\mathscr{S}(\mathbb{R})$ in the Hilbert space $L_2(\mathbb{R})$. Let

$$\phi_0 = \pi^{-1/4}e^{-x^2/2} \quad \text{and} \quad \phi_n = (n!)^{-1/2}(A^\dagger)^n\phi_0, \quad n \in \mathbb{N}, \tag{2.2}$$

be the Hermite functions, which form an orthonormal basis for $L_2(\mathbb{R})$. One has

$$AA^\dagger - A^\dagger A \equiv [A, A^\dagger] = I \tag{2.3}$$

on $\mathscr{S}(\mathbb{R})$, where I is the identity operator on $L_2(\mathbb{R})$, and

$$A^\dagger\phi_n = \sqrt{n+1}\,\phi_{n+1}, \quad A\phi_n = \sqrt{n}\,\phi_{n-1}, \quad n \geqslant 1, \tag{2.4}$$

with $A\phi_0 = 0$. For obvious reasons A is called the annihilation operator and $A^\dagger$ the creation operator. The symmetric operator $N = A^\dagger A$ satisfies $N\phi_n = n\phi_n$ and is essentially self-adjoint on

$$D = \operatorname{span}\{(A^\dagger)^n\phi_0\}_{n\in\mathbb{N}},$$

which is, by (2.2), dense in $L_2(\mathbb{R})$. By (2.1), ϕ_0 is a cyclic vector for $\mathscr{P}$, the algebra of all polynomials with complex coefficients of the operators x and id/dx. It is well known that $\mathscr{P}$ is irreducible in $L_2(\mathbb{R})$ in the sense defined in the Introduction (see, e.g., Ref. 5). (2.1) also implies that any $P(x, id/dx) \in \mathscr{P}$ can be written as a polynomial $Q(A, A^\dagger)$ in A and $A^\dagger$, which in turn, using (2.3), can be rewritten as a sum of normal-ordered monomials in A and $A^\dagger$, i.e., all creation operators in a summand stand to the left of all annihilation operators. Since $A\phi_0 = 0$, we have

$$P\left(x, i\frac{d}{dx}\right)\phi_0 = Q(A, A^\dagger)\phi_0 = \sum_{n=0}^{d(P)} c_n(A^\dagger)^n\phi_0,$$

for some $\{c_n\}_{n=0}^{d(P)} \subset \mathbb{C}$, where $d(P)$ denotes the degree of P. Therefore, using (2.3) to commute the annihilation operators through to the vacuum ϕ_0, if m is the largest integer for which $c_m \neq 0$,

$$A^m P\left(x, i\frac{d}{dx}\right)\phi_0 = m!c_m\phi_0,$$

since

$$A^n(A^\dagger)^n\phi_0 = n!\phi_0,$$

and

$$A^n(A^\dagger)^m\phi_0 = 0,$$

if $n > m$. Thus, we have

Proposition 2.1: Given any nonzero vector Φ in the dense set $\mathscr{P}\phi_0 = D$, there exists a $c \in \mathbb{C}$ and an $n \in \mathbb{N}$ so that

$$cA^n\Phi = c2^{-n/2}\left(x + \frac{d}{dx}\right)^n\Phi = \phi_0.$$

Therefore, any such Φ is cyclic for $\mathscr{P}$ in $L_2(\mathbb{R})$.

It is clear how one may generalize this simple argument to any number of degrees of freedom.

We next show that such straightforward considerations yield similar results for free field theories and their Wick-ordered monomials. For the reader's benefit we summarize those properties of the free scalar field of mass $m \geqslant 0$ in d space–time dimensions that will be of relevance. For further details see, e.g., Ref. 6. [We remark that for the special case $d = 2$, $m = 0$, one must use a smaller test function space than $\mathscr{S}(\mathbb{R}^2)$—see Ref. 7—but that otherwise the arguments are the same.]

If $\mathscr{F}_n$ denotes the space of symmetric L_2 functions on $\mathbb{R}^{n(d-1)}$ and $\mathscr{F}_0 = \mathbb{C}$, let

$$\mathscr{F} = \bigoplus_{n=0}^{\infty} \mathscr{F}_n \quad \text{and} \quad \Omega = 1 \in \mathscr{F}_0.$$

Let S_n be the symmetrizing projection of $L_2(\mathbb{R}^{n(d-1)})$ onto $\mathscr{F}_n$ and D be the algebraic span in $\mathscr{F}$ of Ω and vectors of the form $f(x_1,\ldots,x_n) = S_n(f(x_1)\cdots f(x_n))$. If $a^*(\mathbf{p})$ and $a(\mathbf{p})$ $(\mathbf{p} \in \mathbb{R}^{d-1})$ are the momentum space creation and annihilation "operators," and defining for any $f \in \mathscr{S}_R(\mathbb{R}^{d-1})$ [real-valued functions $\mathscr{S}(\mathbb{R}^{d-1})$]

$$b^*(f) = \int a^*(-\mathbf{p})\tilde{f}(\mathbf{p})\,d^{d-1}p$$

and

$$b(f) = \int a(\mathbf{p})\tilde{f}(\mathbf{p})\,d^{d-1}p,$$

where

$$\tilde{f}(\mathbf{p}) = (2\pi)^{-(d-1)/2}\int e^{-i\mathbf{p}\cdot\mathbf{x}}f(\mathbf{x})\,d^{d-1}x,$$

we have, for any $f, g \in \mathscr{S}_R(\mathbb{R}^{d-1})$, $b^*(f)D \subset D$, $b(f)D \subset D$, and

$$[b(f), b^*(g)] = \langle f, g\rangle_{L_2(\mathbb{R}^{d-1})}I, \quad [b(f), b(g)] = 0, \tag{2.5}$$

on D (I is the identity operator on $\mathscr{F}$). Furthermore, for any $f \in D \cap \mathscr{F}_n$,

$$(b^*(g)f)(\mathbf{x}_1,\ldots,\mathbf{x}_{n+1}) = (n+1)^{1/2}S_{n+1}[g(\mathbf{x}_{n+1})f(\mathbf{x}_1,\ldots,\mathbf{x}_n)], \tag{2.6a}$$

$$(b(g)f)(\mathbf{x}_1,\ldots,\mathbf{x}_{n-1}) = n^{1/2}\int g(\mathbf{x}_n)f(\mathbf{x}_1,\ldots,\mathbf{x}_n)\,d^{d-1}x_n, \tag{2.6b}$$

and

$$b(g)\Omega = 0, \quad \forall\, g \in \mathscr{S}_R(\mathbb{R}^{d-1}).$$

D is thus the algebraic span of vectors $b^*(f_1)\cdots b^*(f_n)\Omega$,

$n \in \mathbb{N}$, $\{f_i\}_{i=1}^n \subset \mathscr{S}_R(\mathbb{R}^d)$, and

$$b^*(g)\colon \mathscr{F}_n \cap D \to \mathscr{F}_{n+1} \cap D,$$
$$b(g)\colon \mathscr{F}_n \cap D \to \mathscr{F}_{n-1} \cap D.$$

If $\omega = (-\Delta + m^2)^{1/2}$, where Δ is the Laplacian in $d-1$ dimensions, then the time-zero field $\varphi_0(x)$ and its canonical conjugate $\pi_0(x)$ are given by

$$\varphi_0(f) = 2^{-1/2}(b^*(\omega^{-1/2}f) + b(\omega^{-1/2}f)) \tag{2.7}$$

and

$$\pi_0(f) = i[H_0, \varphi_0(f)] = i2^{-1/2}(b^*(\omega^{1/2}f) - b(\omega^{1/2}f)), \tag{2.8}$$

for any $f \in \mathscr{S}_R(\mathbb{R}^{d-1})$, where H_0 is the free Hamiltonian. The field at other times is given by

$$\varphi(\mathbf{x}, t) = e^{iH_0 t}\varphi_0(\mathbf{x})e^{-iH_0 t},$$

so that

$$\varphi(\mathbf{x}, t) = 2^{-1/2}(2\pi)^{-d/2}\int e^{-i\mathbf{p}\cdot\mathbf{x}} \times (e^{i\omega(\mathbf{p})t}a^*(\mathbf{p}) + e^{-i\omega(\mathbf{p})t}a(-\mathbf{p}))\omega(\mathbf{p})^{-1/2}\, d^{d-1}p, \tag{2.9}$$

where $\omega(\mathbf{p}) = (\mathbf{p}\cdot\mathbf{p} + m^2)^{1/2}$, and

$$\pi(\mathbf{x}, t) = i2^{-1/2}(2\pi)^{-d/2}\int e^{-i\mathbf{p}\cdot\mathbf{x}} \times (e^{i\omega(\mathbf{p})t}a^*(\mathbf{p}) - e^{-i\omega(\mathbf{p})t}a(-\mathbf{p}))\omega(\mathbf{p})^{1/2}\, d^{d-1}p.$$

We note that (2.5) and (2.6) imply

$$b(f) = 2^{-1/2}(\varphi_0(\omega^{1/2}f) + i\pi_0(\omega^{-1/2}f)), \tag{2.10}$$

so that $b(f) \in \mathscr{P}_0(\mathbb{R}^d)$, the polynomial algebra generated by $\{\varphi_0(g), \pi_0(g) | g \in \mathscr{S}_R(\mathbb{R}^d)\}$, i.e., the time-zero polynomial algebra of the d space–time dimensional free field, which is irreducible in the sense defined in the introduction.

Let P_n be the projection in $\mathscr{F}$ onto $\mathscr{F}_n$. Given any $n \in \mathbb{N}$, $\{f_i\}_{i=1}^n \subset \mathscr{S}(\mathbb{R}^d)$ such that

$$\prod_{i=1}^n \varphi(f_i)\Omega = \prod_{i=1}^n \left\{\int \varphi(\mathbf{x}, t) f_i(\mathbf{x}, t)\, d^{d-1}x dt\right\}\Omega \neq 0;$$

let $n \geqslant m \geqslant 0$ be the largest integer for which

$$P_m \prod_{i=1}^n \varphi(f_i)\Omega \neq 0.$$

Since D is dense in $\mathscr{F}$, we can find a collection $\{h_i\}_{i=1}^m \subset \mathscr{S}_R(\mathbb{R}^{d-1})$ so that

$$\left\langle \prod_{i=1}^m b^*(h_i)\Omega, \prod_{i=1}^n \varphi(f_i)\Omega \right\rangle \neq 0 \tag{2.11}$$

[recall $\mathscr{F}_n \perp \mathscr{F}_m$ for $n \neq m$, and $\Pi_{i=1}^m b^*(h_i)\Omega \in \mathscr{F}_m$]. We assert

$$P_0 \prod_{i=m}^1 b(h_i) \prod_{i=1}^n \varphi(f_i)\Omega = \prod_{i=m}^1 b(h_i) \prod_{i=1}^n \varphi(f_i)\Omega \neq 0. \tag{2.12}$$

This follows since

$$\begin{aligned} P_0 \prod_{i=m}^1 b(h_i) \prod_{i=1}^n \varphi(f_i)\Omega &= P_0\left[\prod_{i=m}^1 b(h_i)\right]\left[\sum_{i=0}^\infty P_i\right]\left[\prod_{i=1}^n \varphi(f_i)\right]\Omega \\ &= P_0\left[\prod_{i=m}^1 b(h_i)\right]P_m\left[\prod_{i=1}^n \varphi(f_i)\right]\Omega \\ &= \left[\prod_{i=m}^1 b(h_i)\right]P_m\left[\prod_{i=1}^n \varphi(f_i)\right]\Omega \\ &= \left[\prod_{i=m}^1 b(h_i)\right]\sum_{i=0}^m P_i\left[\prod_{i=1}^n \varphi(f_i)\right]\Omega \\ &= \prod_{i=1}^m b(h_i) \prod_{i=1}^n \varphi(f_i)\Omega, \end{aligned} \tag{2.13}$$

where we have used (2.5), (2.6), and (2.9), along with the fact that

$$\sum_{i=0}^m P_i\left[\prod_{i=1}^n \varphi(f_i)\right]\Omega = \prod_{i=1}^n \varphi(f_i)\Omega.$$

(2.12) then follows from (2.13) and (2.11). Therefore, the proof of the following proposition is clear.

Proposition 2.2: Given any nonzero vector Φ of the form

$$\sum_{j=1}^l \left(\prod_{i=1}^{n_j} \varphi(f_{ij})\right)\Omega, \quad l \in \mathbb{N}, \ \{n_j\}_{j=1}^l \subset \mathbb{N},$$
$$\{f_{ij}\}_{i=1,j=1}^{n_j, l} \subset \mathscr{S}(\mathbb{R}^d),$$

or of the form

$$\sum_{j=1}^l \left(\prod_{i=1}^{n_j} \varphi_0^\#(f_{ij})\right)\Omega, \quad l \in \mathbb{N}, \ \{n_j\}_{j=1}^l \subset \mathbb{N},$$
$$\{f_{ij}\}_{i=1,j=1}^{n_j, l} \subset \mathscr{S}(\mathbb{R}^{d-1}),$$

where $\varphi_0^\#(f)$ can stand for either $\varphi_0(f)$ or $\pi_0(f)$, there exists an operator $Q \in \mathscr{P}_0(\mathbb{R}^d)$ such that $Q\Phi = \Omega$ and $Q\Omega = 0$. Therefore, every such vector is cyclic for $\mathscr{P}_0(\mathbb{R}^d)$ in $\mathscr{F}$.

Proof: One chooses m to be the largest integer such that $P_m\Phi \neq 0$, and argues as before.

Remark: The set of such vectors is dense in $\mathscr{F}$.

To close this section, we wish to extend this result to vectors obtained from the vacuum by products of Wick-ordered free scalar fields. Given any $n \in \mathbb{N}$, $\{f_i\}_{i=1}^n \subset \mathscr{S}(\mathbb{R}^d)$, we define

$$:\varphi(f_1)\cdots\varphi(f_n):\Omega = P_n\varphi(f_1)\cdots\varphi(f_n)\Omega,$$

so that

$$:\varphi(f_1)\cdots\varphi(f_n):\Omega = \prod_{j=1}^n \left\{2^{-1/2}(2\pi)^{-d/2}\int e^{-i\mathbf{p}\cdot\mathbf{x}}e^{i\omega(\mathbf{p})t} \times a^*(\mathbf{p})\omega(\mathbf{p})^{-1/2}\, d^{d-1}p f_j(\mathbf{x}, t)\, d^{d-1}x dt\right\}\Omega.$$

As we have just seen, for any set $\{h_i\}_{i=1}^n \subset \mathscr{S}_R(\mathbb{R}^{d-1})$ such that

$$\left\langle \prod_{i=1}^n b^*(h_i)\Omega, :\prod_{i=1}^n \varphi(f_i):\Omega \right\rangle \neq 0,$$

one has

$$P_0 \prod_{i=1}^n b(h_i) :\prod_{i=1}^n \varphi(f_i):\Omega = \prod_{i=n}^1 b(h_i) :\prod_{i=1}^n \varphi(f_i):\Omega \neq 0. \tag{2.14}$$

The nth Wick power of $\varphi(x)$ can be defined formally by

$$:\varphi^n:(x) = \lim_{x_1,\ldots,x_n \to x} :\varphi(x_1)\cdots\varphi(x_n):$$

(see Ref. 6 for a rigorous definition) and gives a field satisfying the Wightman axioms.[6] Moreover, for any $n \in \mathbb{N}$, $g \in \mathscr{S}(\mathbb{R}^d)$, there exists a sequence of vectors

$$\{:\varphi(f_1^N)\cdots\varphi(f_n^N):\Omega\}_{N\in\mathbb{N}}$$

converging strongly in $\mathscr{F}$ to $:\varphi^n:(g)\Omega$ and such that $\{Q:\Pi_{i=1}^n\varphi(f_i^N):\Omega\}$ is strongly Cauchy for any polynomial Q in the free field (see Ref. 8). Therefore, $:\varphi^n:(g)\Omega$ is in the domain of the closure of $Q \restriction D$, and, in particular, in the domain of the closure of $[\Pi_{i=n}^1 b(h_i)] \restriction D$, for any $n \in \mathbb{N}$, $\{h_i\}_{i=1}^n \subset \mathscr{S}_R(\mathbb{R}^{d-1})$.

There exists a set $\{h_i\}_{i=1}^n \subset \mathscr{S}_R(\mathbb{R}^{d-1})$ such that

$$\left\langle \prod_{i=1}^n b^*(h_i)\Omega,\ :\varphi^n:(g)\Omega \right\rangle \neq 0, \tag{2.15}$$

so that there exists an $N_0 \in \mathbb{N}$ for which

$$\left\langle \prod_{i=1}^n b^*(h_i)\Omega,\ :\varphi(f_1^N)\cdots\varphi(f_n^N):\Omega \right\rangle \neq 0, \quad \forall\ N \geqslant N_0 . \tag{2.16}$$

(2.14) then implies

$$\prod_{i=n}^1 b(h_i):\prod_{i=1}^n \varphi(f_i^N):\Omega = c_N\Omega ,$$

where c_N is equal to the scalar product in (2.16). Using the already mentioned strong convergence, it follows that

$$\overline{\left[\prod_{i=n}^1 b(h_i)\right]}\ :\varphi^n:(g)\Omega = c_\infty\Omega , \tag{2.17}$$

where c_∞ is equal to the scalar product in (2.15) and the bar above the operator in (2.17) designates the closure on the domain D. We have therefore proven the following result.

Proposition 2.3: Given any nonzero vector Φ of the form $:\varphi^n:(g)\Omega$, $n \in \mathbb{N}$, $g \in \mathscr{S}(\mathbb{R}^d)$, there exists a closable operator $Q \in \mathscr{P}_0(\mathbb{R}^d)$ such that $\overline{Q}\Phi = \Omega$ and $Q\Omega = 0$, where $\overline{Q} = \overline{[Q \restriction D]}$. Thus, every such vector is cyclic for $\overline{\mathscr{P}}_0(\mathbb{R}^d)$, in $\mathscr{F}$, where $\overline{\mathscr{P}}_0(\mathbb{R}^d)$ contains the appropriate closed operator.

Remarks: (1) By using expansion formulas in Appendix A of Ref. 8 and the arguments above, it is easy to extend Proposition 2.3 sums of products of Wick powers of free fields applied to the vacuum.

(2) All Arguments given above can be applied immediately to generalized free fields to obtain for them analogous results.

III. BASIC ASSUMPTIONS AND NOTATION: GENERAL THEORY OF QUANTIZED FIELDS

In the preceding chapter we have shown that a few special, irreducible *-algebras of unbounded operators have a dense set of cyclic vectors and that the operators in these algebras are vacuum reducible. The proofs we have given depend on the especially straightforward algebra of the canonical commutation relations in Fock space. In the remainder of this paper we wish to prove similar results for a large class of irreducible *-algebras of unbounded operators, arising out of relativistic quantum field theory, whose algebraic relations are far from being simple, but which have an involved and beautiful structure that we can employ to obtain the desired results. The purpose of this chapter is to establish notation and to recall those results of axiomatic quantum field theory that we shall have particular use for. We shall admit quantum fields that are more singular than the tempered distributions of the standard Wightman theory.

We recall here the axioms of a strictly localizable quantum field theory, although the arguments in Sec. V, with one exception, would apply to any scalar field theory that can be defined on a nuclear test function space containing test functions with compact support in momentum space. In that case the locality condition given below is then replaced by an appropriate generalization (see Refs. 9 and 10) and the crucial asymptotic decay property of the truncated vacuum expectation values (Theorem 3.6) is known to hold again.[9,10] However, the proof of the density of the resulting vacuum reducible vectors that we give does use strict localizability; so aside from these remarks, we restrict our attention to test function spaces with a dense set of elements with compact support in configuration space, i.e., strictly localizable fields. We take the trouble to handle this problem in this generality because there are concrete field models with fields that are not tempered distributions (see Refs. 11–13) and there is reason to believe that other physically interesting models will be able to be constructed, if at all, only in this framework (see references in Refs. 12 and 10).

The most general setting for strictly localizable fields would seem to be the theory of ultradistributions (see Refs. 14 and 15), which subsumes as special cases both Jaffe spaces[11] and the S-type spaces of Gel'fand and Shilov.[16] We recall the necessary definitions and briefly indicate the salient properties. Let $\omega:[0,\infty) \to \mathbb{R}$ be a function satisfying the following conditions:

(a) $\exp[\omega(x^2)]$ is a real entire function on $\mathbb{R}^d$ ($x^2 = \Sigma_{i=0}^{d-1} x_i^2$);
(b) $\omega(x+y) \leqslant \omega(x) + \omega(y)$, $\forall\ x, y \in [0,\infty)$,
(c) $\int_0^\infty [\omega(t^2)/(1+t^2)]\,dt < \infty$;
(d) $\omega(x^2) \geqslant \ln(1+|x|^2)$, $\forall\ x \in \mathbb{R}^d$,
(e) $2\omega(x^2) \leqslant \omega(Ax^2) + C$, $\forall\ x \in \mathbb{R}^d$,

for some constants A and C. Condition (c), known as Carleman's criterion, assures that there will be sufficiently many local test functions in the space; condition (d) entails that the test functions are in $\mathscr{S}(\mathbb{R}^d)$, and condition (e) will give the nuclearity of the space. (This condition can be somewhat weakened—see Ref. 17). In the general theory of ultradistributions, condition (e) is omitted; we, however, shall need the nuclearity. Following Ref. 17, we call any such function ω a Jaffe indicatrix.

Given a Jaffe indicatrix ω, we define $\mathscr{M}_\omega(\mathbb{R}^d)$ to be the space of all infinitely differentiable functions f on $\mathbb{R}^d$ such that

$$P_{\alpha,\lambda}(f) \equiv \|f\|_{\alpha,\lambda}^{(\mathscr{M}_\omega)} \equiv \sup_{p\in\mathbb{R}^d}\{e^{\lambda\omega(|p^2|)}|D^\alpha f(p)|\} < \infty ,$$

for each multiindex α and each constant $\lambda > 0$. The topology in $\mathscr{M}_\omega$ is then defined canonically in terms of these semin-

orms. The space $\mathscr{C}_\omega(\mathbb{R}^d)$ of inverse Fourier transforms of $\mathscr{M}_\omega$, i.e., $\mathscr{C}_\omega = \mathscr{F}^{-1}(\mathscr{M}_\omega)$ has a canonical topology determined by the seminorms

$$\pi_{\alpha,\lambda}(f) \equiv \|f\|^{(\mathscr{C})}_{\alpha,\lambda} \equiv \sup_{p\in\mathbb{R}^d}\{e^{\lambda\omega(p^2)}|D^\alpha \tilde{f}(p)|\}\,,$$

where

$$\tilde{f}(p) = (\mathscr{F}f)(p) = (2\pi)^{-d/2}\int e^{-ix\cdot p}f(x)\,d^dx\,. \tag{3.1}$$

$\mathscr{F}:\mathscr{C}_\omega(\mathbb{R}^d)\to\mathscr{M}_\omega(\mathbb{R}^d)$ is a continuous isomorphism in these topologies. If $\mathscr{O}\subset\mathbb{R}^d$ is compact and $\mathscr{D}_\omega(\mathscr{O})$ is the set of all L_1-functions with support in $\mathscr{O}$ such that the seminorms

$$\|f\|_\lambda = \int |\tilde{f}(p)|e^{\lambda\omega(p^2)}\,d^dp$$

are finite, for every $\lambda>0$, then $\mathscr{D}_\omega(\mathscr{O})$ is dense in $\mathscr{D}(\mathscr{O})$ (Schwartz's test function space of C^∞-functions with support in $\mathscr{O}$). And defining $\mathscr{D}_\omega(\mathbb{R}^d)$ as the inductive limit of $\mathscr{D}_\omega(\mathscr{O}_n)$, $\mathscr{O}_n\uparrow\mathbb{R}^d$, $\mathscr{O}_n\subset\mathscr{O}_{n+1}$, $\forall\ n$, the same is true for $\mathscr{D}_\omega(\mathbb{R}^d)\subset\mathscr{D}(\mathbb{R}^d)$. These assertions follow from condition (c). The elements of the dual space $\mathscr{D}'_\omega$ of $\mathscr{D}_\omega$ are called ultradistributions. $\mathscr{C}'_\omega$ is a space of ultradistributions and $\mathscr{M}'_\omega$ is a distribution space, i.e., can be viewed as a subspace of $\mathscr{D}'$.[18] We define $\mathscr{K}_\omega(\mathbb{R}^d)=\mathscr{F}^{-1}(\mathscr{D}_\omega(\mathbb{R}^d))$.

For easy reference in this paper and in Ref. 1, we collect here some known results in a series of lemmas.

Lemma 3.1[14,18]: The spaces $\mathscr{M}_\omega$ and $\mathscr{C}_\omega$ are topological algebras under pointwise multiplication and convolution.

Lemma 3.2[14,18]: $\mathscr{D}_\omega$ is dense in $\mathscr{M}_\omega$ and $\mathscr{C}_\omega$.

Lemma 3.3[18]: The translation operator τ_a, $a\in\mathbb{R}^d$:

$$(\tau_a f)(x) = f(x-a)$$

and the multiplication by $\exp(ix\cdot a)$ are continuous operators on $\mathscr{M}_\omega$ and $\mathscr{C}_\omega$.

Lemma 3.4[15,17]: The spaces $\mathscr{M}_\omega$, $\mathscr{C}_\omega$, and $\mathscr{D}_\omega$ are nuclear.

Lemma 3.5[15]: Corresponding to every open covering in $\mathbb{R}^d$, there exists a partition of unity in $\mathscr{D}_\omega(\mathbb{R}^d)$ [or $\mathscr{M}_\omega(\mathbb{R}^d)$ or $\mathscr{C}_\omega(\mathbb{R}^d)$, as the case may be].

The Wightman axioms (Hilbert space structure, covariance of the fields under the Poincaré group, the spectrum condition, and locality) can then be formulated for a general spin quantum field theory as usual (see Refs. 3 and 19), with the configuration space test functions $\mathscr{S}(\mathbb{R}^d)$ replaced by $\mathscr{C}_\omega(\mathbb{R}^d)$.[11] Thus, for any vectors Φ, $\Psi\in D_{\mathrm{W}}$, the invariant common domain of the fields $\varphi(f)$, $f\in\mathscr{C}_\omega(\mathbb{R}^d)$, in the Hilbert space $\mathscr{H}_{\mathrm{W}}$ associated with the fields, the form $\langle\Phi,\varphi(f)\Psi\rangle$ is continuous in f in the topology of $\mathscr{C}_\omega$, i.e., $\langle\Phi,\varphi(\cdot)\Psi\rangle\in\mathscr{C}_\omega(\mathbb{R}^d)'$. We shall refer to this axiom as the continuity axiom or Wightman continuity. It is due to Lemma 3.2, i.e., due to the fact that in configuration space there are enough test functions of compact support, that the locality property does not need to be generalized. Jaffe (unpublished) and Constantinescu and Thalheimer[17,18] (see also Refs. 9 and 10 for even smaller test function spaces) have shown that all the standard tools and results of axiomatic quantum field theory (including analytic continuation of Wightman functions and edge-of-the-wedge theorems that we shall use many times in the following) are valid for such strictly localizable fields.

We shall denote with $\mathscr{P}(\mathbb{R}^d)$ the $*$-algebra (defined on D_{W}) of polynomials generated by $\{\varphi(f)|\,f\in\mathscr{C}(\mathbb{R}^d)\}$ (we shall henceforth suppress the index ω, since all formulations and results in the sequel will be valid for any Jaffe indicatrix), with $\mathscr{P}_{\mathrm{c}}(\mathbb{R}^d)$ that generated by the operators $\{\varphi(f)|\,f\in\mathscr{K}(\mathbb{R}^d)\}$, and with $\mathscr{P}_{\mathrm{loc}}(\mathbb{R}^d)$ that generated by $\{\varphi(f)|\,f\in\mathscr{D}(\mathbb{R}^d)\}$. If $\mathscr{O}\subseteq\mathbb{R}^d$, $\mathscr{P}(\mathscr{O})$ will designate the $*$-algebra of polynomials generated by $\{\varphi(f)|\,f\in\mathscr{D}(\mathscr{O})\}$. Included in the Wightman axioms is the assumption that the vacuum Ω is cyclic for $\mathscr{P}(\mathbb{R}^d)$ in $\mathscr{H}_{\mathrm{W}}$, and because $\mathscr{K}(\mathbb{R}^d)$ and $\mathscr{D}(\mathbb{R}^d)$ are dense in $\mathscr{C}(\mathbb{R}^d)$ [using Lemma 3.2 and $\mathscr{F}:\mathscr{C}(\mathbb{R}^d)\to\mathscr{M}(\mathbb{R}^d)$ a continuous isomorphism], the continuity axiom implies that $\mathscr{P}_{\mathrm{c}}(\mathbb{R}^d)\Omega$ and $\mathscr{P}_{\mathrm{loc}}(\mathbb{R}^d)\Omega$ are also dense in $\mathscr{H}_{\mathrm{W}}$. It is well known that the uniqueness of the vacuum, which also is assumed in the Wightman axioms, entails the irreducibility of $\mathscr{P}(\mathbb{R}^d)$ [thus of $\mathscr{P}_{\mathrm{c}}(\mathbb{R}^d)$ and $\mathscr{P}_{\mathrm{loc}}(\mathbb{R}^d)$] in the sense given in the Introduction.[3] We shall also be assuming below the existence of a positive mass gap, i.e., the existence of a lowest nonzero eigenvalue for the mass operator $P\cdot P=(P^0)^2-\sum_{k=1}^{d-1}(P^k)^2$ (where $P=(P^0,P^1,\ldots,P^{d-1})$ is the generator of the space–time translations on $\mathscr{H}_{\mathrm{W}}$).

Let us recall the connection between the Wightman functions

$$W_n(x_1,\ldots,x_n) = \langle\Omega,\varphi(x_1)\cdots\varphi(x_n)\Omega\rangle$$

and the truncated vacuum expectation values (TVEV's) $W^{\mathrm{T}}_n(x_1,\ldots,x_n)$. The TVEV's are defined recursively by

$$W_n(x_1,\ldots,x_n) = \sum_{\text{part.}} W^{\mathrm{T}}_{r_1}(x_{l_{1,1}},\ldots,x_{l_{1,r_1}})W^{\mathrm{T}}_{r_2}(x_{l_{2,1}},\ldots,x_{l_{2,r_2}})\cdots W^{\mathrm{T}}_{r_s}(x_{l_{s,1}},\ldots,x_{l_{s,r_s}})\,, \tag{3.2}$$

where the sum on the right-hand side runs over all partitions of the indices $1,\ldots,n$, and in each subset $1_{k,1},\ldots,1_{k,r_k}$, the indices are taken in natural order. Since the Wightman functions and TVEV's are translation-invariant, one may define the difference value Wightman functions and TVEV's:

$$\mathscr{W}^{(\mathrm{T})}_n(\xi_1,\ldots,\xi_n)\equiv W^{(\mathrm{T})}_{n+1}(x_0,\ldots,x_n),\quad \xi_k\equiv x_{k-1}-x_k\,. \tag{3.3}$$

The Fourier transforms of the difference variable Wightman functions satisfy

$$\operatorname{supp}\tilde{\mathscr{W}}_n(q_1,\ldots,q_n)\subset\{q_k\in\overline{V}_+,\ k=1,\ldots,n\}\,,$$

where supp denotes support and $\overline{V}_+$ the closure in $\mathbb{R}^d$ (with Minkowski metric) of $\{p\in\mathbb{R}^d\,|\,p\cdot p\geqslant 0\}$, and, since in the TVEV the vacuum contribution is symmetrically subtracted out, if there is a mass gap $m>0$,

$$\operatorname{supp}\tilde{\mathscr{W}}^{\mathrm{T}}_n(q_1,\ldots,q_n)\subset\{q_k\in\overline{V}^m_+,\ k=1,\ldots,n\},$$[19]

where $\overline{V}^m_+$ is the closure of $\{p\in\mathbb{R}^d\,|\,p\cdot p\geqslant m^2\}$ and

$$\tilde{\mathscr{W}}_n^{(T)}(q_1,\ldots,q_n) = (2\pi)^{-nd/2}\int \exp\Big(i\sum_{k=1}^{n} p_k\cdot\xi_k\Big) \times \mathscr{W}_n^{(T)}(\xi_1,\ldots,\xi_n)\prod_{i=1}^{n} d^d\xi_i\,. \tag{3.4}$$

It might be worthwhile pointing out for later use that with our conventions (3.1), (3.3), and (3.4),

$$W_n(f_1,\ldots,f_n) = \int \tilde{\mathscr{W}}_{n-1}(p_1,\ldots,p_{n-1}) \times \tilde{f}_1(p_1)\tilde{f}_2(p_2-p_1)\cdots\tilde{f}_n(-p_{n-1})\prod_{i=1}^{n-1} d^d p_i\,. \tag{3.5}$$

We shall have need below of the following theorem about the asymptotic behavior of (generalized) TVEV's on spacelike hyperplanes. Let

$$Q_k(0) = \int f_k(x_1,\ldots,x_{n_k})\varphi(x_1)\cdots\varphi(x_{n_k})\prod_{j=1}^{n_k} d^d x_j\,,$$

where $f_k \in \mathscr{C}(\mathbb{R}^{dn_k})$, $k=1,\ldots,N$. Let $Q_k(x_k)$ be the translate of $Q_k(0)$, i.e.,

$$Q_k(x_k) = e^{iP\cdot x_k}Q_k(0)e^{-iP\cdot x_k},\quad x_k=(0,\mathbf{x}_k)\in\mathbb{R}^d\,.$$

We define the generalized TVEV

$$F(\bar{x}) = \langle Q_1(x_1)\cdots Q_N(x_N)\rangle^T$$

in analogy to (3.2) and let

$$\rho = \max_{k,k'=1,\ldots,N}|\mathbf{x}_k-\mathbf{x}_{k'}|\,.$$

Then using ideas dating back to Ruelle,[2] the following theorem can be proven.

Theorem 3.6[9,10]**:** Let $\varphi(x)$ be a real scalar quantum field theory satisfying the axiom structure discussed above and having a nonzero mass gap. Then for any set $\{Q_k(0)\}_{k=1}^N$ defined as above, $|F(\bar{x})|$ converges to zero faster than any power of ρ^{-1} as $\rho\to\infty$.

Remark: Although the proofs in Refs. 9 and 10 are carried out in detail only for the TVEV's defined in (3.2), it is easy to apply the argument to the generalized TVEV's we have defined.

IV. THE PRIMARY RESULT AND ITS IMMEDIATE CONSEQUENCES

In this section we state the central technical result of this paper, the proof of which is presented in the following section, and demonstrate its immediate consequences. As previously mentioned, the main theorem, Theorem 4.1, will have other applications in our study of the essential self-adjointness of quantum field operators. And there will surely be other uses for the results of this section.

In the following, a "quantum field" shall be understood to mean one satisfying the axioms discussed in the previous section.

Theorem 4.1: Let $\varphi(x)$ be a real scalar quantum field with mass gap $m>0$. Then there exists a set $\mathscr{E}$ of vectors, dense in the Hilbert space $\mathscr{H}_W$ associated with the field, of the form

$$\sum_{j=1}^{N}\Big(\prod_{i=1}^{n_j}\varphi(f_{ij})\Big)\Omega,$$

$$N\in\mathbb{N},\ \{n_j\}_{j=1}^N\subset\mathbb{N},\ \{f_{ij}\}\subset\mathscr{K}(\mathbb{R}^d),$$

such that there exists, for any such nonzero vector of this form, an $n_{j_0}\in\mathbb{N}$ and a set $\{g_i\}_{i=1}^{n}\subset\mathscr{K}(\mathbb{R}^d)$ so that the following equations are satisfied:

(i) $\prod_{i=1}^{n_{j_0}}\varphi(g_i)\prod_{k=1}^{n_{j_0}}\varphi(f_{kj_0})\Omega=\Omega$,

(ii) $\varphi(g_i)\Omega=0,\quad i=1,\ldots,n_{j_0}$,

and

(iii) $\prod_{i=1}^{n_{j_0}}\varphi(g_i)\prod_{k=1}^{n_j}\varphi(f_{kj})\Omega=0,\quad \forall\, j\neq j_0$.

Remarks: (1) The set $\mathscr{E}$ is defined in the next section [see, in particular, (5.15)].

(2) The $\varphi(g_i)$'s clearly act in an analogous manner to the annihilation operators in Sec. II, and for this reason we have taken to calling them annihilators.

(3) Since the test functions in the definition of the vectors in $\mathscr{E}$ have Fourier transforms with compact support, $\mathscr{E}$ is a subset of all finite energy–momentum vectors in $\mathscr{H}_W$. Therefore, vectors in $\mathscr{E}$ are entire analytic vectors for the energy–momentum operators—that is to say, given any $\Phi\in\mathscr{E}$,

$$\sum_{n=1}^{\infty}\frac{\|P^n\Phi\|}{n!}t^n<\infty\,, \tag{4.1}$$

for all $t\in\mathbb{C}$, where P represents any of the energy–momentum operators P^μ.

An immediate consequence of Theorem 4.1 is made clear by the following remark. Given any vector $\Phi=\Sigma_{j=1}^N\big[\Pi_{i=1}^{n_j}\varphi(f_{ij})\big]\Omega\in\mathscr{E}$, one has

$$\mathscr{P}_c\Phi\supset\mathscr{P}_c\prod_{i=1}^{n_{j_0}}\varphi(g_i)\sum_{j=1}^{N}\Big(\prod_{i=1}^{n_j}\varphi(f_{ij})\Big)\Omega=\mathscr{P}_c\Omega \tag{4.2}$$

$[\mathscr{P}_c=\mathscr{P}_c(\mathbb{R}^d)]$, which is dense in $\mathscr{H}_W$. Therefore, we have:

Corollary 4.2: Let $\varphi(x)$ be a real scalar quantum field with mass gap $m>0$. There exists a dense set $\mathscr{E}$ of vectors, each of whose elements is cyclic for $\mathscr{P}_c(\mathbb{R}^d)$, $\mathscr{P}(\mathbb{R}^d)$, and $\mathscr{P}_{loc}(\mathbb{R}^d)$.

Proof: (4.2) yields the claim for $\mathscr{P}_c$, and since, by Wightman continuity,

$$\overline{\mathscr{P}_c(\mathbb{R}^d)\Phi}=\overline{\mathscr{P}(\mathbb{R}^d)\Phi}=\overline{\mathscr{P}_{loc}(\mathbb{R}^d)\Phi},\quad\forall\ \Phi\in D_W$$

(see Sec. III), the other cases follow at once.

Since, as already noted, the vectors in $\mathscr{E}$ are entire analytic for the energy, one may adapt a standard argument to obtain:

Corollary 4.3: Let $\varphi(x)$ be a real scalar quantum field with mass gap $m>0$. There exists a dense set $\mathscr{E}$ of vectors, each of whose elements is cyclic and separating for $\mathscr{P}(\mathscr{O})$, given any open $\mathscr{O}\subset\mathbb{R}^d$ such that $\mathscr{O}'$, the causal complement of $\mathscr{O}$, is not empty.

Proof: As the argument for the case $\Phi=\Omega$ is well known (see Ref. 3), we shall only sketch the proof (see also Ref. 20 for the case of bounded algebras). Given such an $\mathscr{O}$ and $\Phi\in\mathscr{E}$, assume there exists a $\Psi\in\mathscr{H}_W$ such that

$$\langle \Psi, P\Phi \rangle = 0, \quad \forall\ P \in \mathscr{P}(\mathscr{O}).$$

Let $\mathscr{O}_1$ be a nonempty open subset of $\mathscr{O}$ such that $\mathscr{O}_1 + \mathscr{N} \subset \mathscr{O}$, where $\mathscr{N}$ is a neighborhood of the origin in $\mathbb{R}^d$. Then we have (we continue to suppress ω)

$$\left\langle \Psi, \prod_{j=1}^{n} e^{ia_j P} \varphi(h_j) e^{-ia_j P} \Phi \right\rangle = 0,$$

$$\forall\ \{h_j\}_{j=1}^{n} \subset \mathscr{D}(\mathscr{O}_1), \quad \forall\ \{a_j\}_{j=1}^{n} \subset \mathscr{N}. \tag{4.3}$$

Since Φ is entire analytic for P^{μ}, $\mu = 0,..., d-1$, we may employ the spectrum condition to analytically continue (4.3) to

$$\left\langle \Psi, \prod_{j=1}^{n} e^{iz_j P} \varphi(h_j) e^{-iz_j P} \Phi \right\rangle, \tag{4.4}$$

where $z_j = a_j + ib_j$, $\{a_j\}_{j=1}^{n} \subset \mathscr{N}$, $b_1 \in \overline{V}_+$, $b_{j+1} - b_j \in \overline{V}_+$, $j = 1,..., n$. There is no further restriction on b_n since Φ is an entire vector. But the analytic function (4.4) coincides in the specified domain of analyticity with the analytic function

$$\left\langle \Psi, \prod_{j=1}^{n} e^{iz_j P} \varphi(h_j) e^{-iz_j P} \Phi \right\rangle$$

with domain $z_j = c_j + ib_j$, $\{c_j\}_{j=1}^{n} \subset \mathbb{R}$, $b_1 \in \overline{V}_+$, $b_{j+1} - b_j \in \overline{V}_+$, $j = 1,...,n$. Since (4.4) vanishes in an open real neighborhood [by (4.3)], one may then conclude that

$$\left\langle \Psi, \prod_{j=1}^{n} e^{ib_j P} \varphi(h_j) e^{-ib_j P} \Phi \right\rangle = 0,$$

$$\forall\ \{h_j\}_{j=1}^{n} \subset \mathscr{D}(\mathscr{O}_1), \quad \forall\ \{b_j\}_{j=1}^{n} \subset \mathbb{R},$$

using the edge-of-the-wedge theorem in the form of Theorem 2.17 in Ref. 3. (For confirmation that the edge-of-the-wedge theorem holds also for ultradistributions, see Ref. 17, which uses Lemmas 3.2 and 3.4). Thus, if $\{\mathscr{O}_n\}$ is a countable covering of $\mathbb{R}^d$ composed of translates of $\mathscr{O}_1$,

$$\left\langle \Psi, \sum_{i=1}^{k} P_{n_i} \Phi \right\rangle = 0, \quad \forall\ \{n_i\}_{i=1}^{n} \subset \mathbb{N}, \quad \forall\ P_{n_i} \in \mathscr{P}(\mathscr{O}_{n_i}). \tag{4.5}$$

Since there is a partition of unity in $\mathscr{C}(\mathbb{R}^d)$ (Lemma 3.5) corresponding to each open covering $\mathbb{R}^d$, there exists a sequence $\{\chi_n\}_{n=1}^{\infty} \subset \mathscr{C}(\mathbb{R}^d)$ such that $\chi_n \in \mathscr{D}(\mathscr{O}_n)$, for each n, and $\Sigma_{n=1}^{\infty} \chi_n f = f$, given any $f \in \mathscr{C}(\mathbb{R}^d)$, where the convergence is in the topology of $\mathscr{C}(\mathbb{R}^d)$ (by Lemma 3.1, $\chi_n f \in \mathscr{C}(\mathbb{R}^d)$, $\forall\ n$). Thus, by the linearity of $\varphi(f)$ on $\mathscr{C}(\mathbb{R}^d)$ and the strong continuity of $\Pi_{i=1}^{k} \varphi(f_i)\Phi$ in the topology of $\mathscr{C}(\mathbb{R}^d)$ (follows from Wightman continuity), (4.5) entails that

$$\langle \Psi, P\Phi \rangle = 0, \quad \forall\ P \in \mathscr{P}(\mathbb{R}^d),$$

which, with Corollary 4.2, completes the sketch of the proof.

We note here that the same arguments may be applied to those cyclic vectors constructed in Sec. II for the free field that are entire analytic for the energy–momentum operators, to conclude that they are cyclic for the local polynomial algebras, as well.

V. VACUUM REDUCIBLE OPERATORS IN QUANTUM FIELD THEORIES WITH MASS GAP

The object of this chapter is to prove Theorem 4.1. It will be clear from the proof that we do not demonstrate the strongest possible result, even using only the relatively crude methods we have employed here. But it is not obvious that one would not already have passed the point of diminishing returns by squeezing out a few more operators with annihilators. Nevertheless, it would be an interesting extension to show Theorem 4.1 to be true for all $P \in \mathscr{P}_c(\mathbb{R}^d)$ (we have results to this effect in the weak-coupling limit, and, as already seen, it is true for free field theories)—but see Ref. 21.

We must begin by defining the special classes of test functions with which we shall be working. Let

$$\mathscr{A}_1(\mathbb{R}^d) = \{ f \in \mathscr{K}(\mathbb{R}^d) | \tilde{f}(p) = 0 \Leftrightarrow \tilde{f}(-p) = 0 \}$$

(we continue to suppress the Jaffe indicatrix ω), and

$$\mathscr{A}_2(\mathbb{R}^d) = \{ f \in \mathscr{A}_1(\mathbb{R}^d) | (I - P_\Omega)\varphi(f)\Omega \neq 0 \},$$

where P_Ω is the projection in $\mathscr{H}_W$ onto the one-dimensional subspace spanned by Ω, and $\varphi(x)$ is a real scalar quantum field. Let $T_r = \{ p \in \mathbb{R}^d | 0 \leqslant p^0 \leqslant r \}$ [$p = (p^0, p^1,..., p^{d-1})$] and, for any $f \in \mathscr{K}(\mathbb{R}^d)$,

$$r_f = \inf\{ r | \operatorname{supp} \tilde{f} \cap \overline{V}_+ \subset T_r \}.$$

Given any $f \in \mathscr{A}_2(\mathbb{R}^d)$, let $\theta_f(x)$ be any element of $\mathscr{A}_1(\mathbb{R}^d)$ whose Fourier transform $\tilde{\theta}_f(p)$ satisfies

$$\operatorname{supp} \tilde{\theta}_f(p) = \{ p \in \mathbb{R}^d | -r_f \leqslant p^\mu \leqslant r_f,\ \mu = 0,..., d-1 \} \setminus (\operatorname{supp} \tilde{f})^0,$$

where a set with a superscript 0 signifies its interior. Then for any $f \in \mathscr{A}_2(\mathbb{R}^d)$ and for any $\epsilon \geqslant 0$,

$$\|\varphi(f + \epsilon\theta_f)\Omega\|^2 = \|\varphi(f)\Omega\|^2 + \epsilon^2 \|\varphi(\theta_f)\Omega\|^2 > 0$$

and

$$\|(I - P_\Omega)\varphi(f + \epsilon\theta_f)\Omega\|^2 = \|(I - P_\Omega)\varphi(f)\Omega\|^2 + \epsilon^2 \|(I - P_\Omega)\varphi(\theta_f)\Omega\|^2 > 0,$$

as is easily seen, since $(\operatorname{supp} \tilde{\theta}_f)^0 \cap (\operatorname{supp} \tilde{f})^0 = \varnothing$. Furthermore,

$$\tilde{f}(p) + \epsilon\tilde{\theta}_f(p) = 0 \quad \Leftrightarrow \quad \tilde{f}(-p) + \epsilon\tilde{\theta}_f(-p) = 0.$$

We define now the set

$$\mathscr{B}(\mathbb{R}^d) = \{ h(x) \in \mathscr{K}(\mathbb{R}^d) | h(x) = f(x) + \epsilon\theta_f(x), \ f \in \mathscr{A}_2(\mathbb{R}^d),\ \epsilon > 0,\ \theta_f \text{ defined as above} \}.$$

One notes at once that every element of $\mathscr{A}_2(\mathbb{R}^d)$ can be arbitrarily well approximated [in the topology of $\mathscr{C}(\mathbb{R}^d)$] by functions in $\mathscr{B}(\mathbb{R}^d)$.

The first step in the proof of Theorem 4.1 is to verify that it holds for vectors of the form $\varphi(f)\Omega$, with $f \in \mathscr{B}(\mathbb{R}^d)$.

Lemma 5.1: Let $\varphi(x)$ be a real scalar quantum field with mass gap $m \geqslant 0$. Given any $f \in \mathscr{B}(\mathbb{R}^d)$, there exists a family $\{g_\delta\}_{\delta \in (0,m)} \subset \mathscr{K}(\mathbb{R}^d)$ such that

(i) $\varphi(g_\delta)\varphi(f)\Omega = \Omega$

and

(ii) $\varphi(g_\delta)\Omega = 0$,

for each $\delta \in (0, m)$.

Proof: The hypothesis entails that $(\operatorname{supp}\tilde{f})^0 \cap \operatorname{supp}\tilde{\mathscr{W}}_1 \cap \overline{V_+^m} \neq \emptyset$, since $\operatorname{supp}\tilde{\mathscr{W}}_1 \subseteq \{0\} \cup \overline{V_+^m}$ and

$$\begin{aligned} 0 &< \|(I - P_\Omega)\varphi(f)\Omega\|^2 \\ &= \langle \Omega, \varphi(f^*)\varphi(f)\Omega\rangle - |\langle \Omega, \varphi(f)\Omega\rangle|^2 \\ &= \langle \Omega, \varphi(f^*)\varphi(f)\Omega\rangle^T \\ &= \int_{p^2 \geq m^2} \tilde{\mathscr{W}}_1^{\mathrm{T}}(p)|\tilde{f}(-p)|^2\, dp\,, \end{aligned}$$

using (3.2) and the support properties of $\tilde{\mathscr{W}}_1^{\mathrm{T}}(p)$ mentioned in Sec. III. The fact that $\tilde{f}(p) = 0$ if and only if $\tilde{f}(-p) = 0$, $\forall\, f \in \mathscr{B}(\mathbb{R}^d)$, completes the proof of the assertion.

Let

$$\mu \in \partial(\operatorname{supp}\tilde{f}) \cap \partial T_{r_f} \cap \operatorname{supp}\tilde{\mathscr{W}}_1\,, \tag{5.1}$$

where ∂ applied to a set denotes its boundary. This intersection is nonempty, since $\tilde{\mathscr{W}}_1(p)$ is a Lorentz-invariant distribution and because $(\operatorname{supp}\tilde{f})^0 \cap \operatorname{supp}\tilde{\mathscr{W}}_1 \cap \overline{V_+^m} \neq \emptyset$, $\forall\, f \in \mathscr{B}(\mathbb{R}^d)$. The rather curious set $\mathscr{B}(\mathbb{R}^d)$ was chosen expressly so that one could be assured of finding such a μ for each $f \in \mathscr{B}(\mathbb{R}^d)$ (but see the remark following this proof). Of course, for such functions one can find in general many such μ, but for our purposes it does not matter which one one chooses. We henceforth presume that such a μ has been chosen for each $f \in \mathscr{B}(\mathbb{R}^d)$.

Choose a $\delta \in (0, m)$ and define $g_{\delta,\mu} \in \mathscr{K}(\mathbb{R}^d)$ by its Fourier transform:

$$\tilde{g}_{\delta,\mu}(p) = \chi_{\delta,\mu}(p)\tilde{f}(-p)^*\,,$$

where $1 \geqslant \chi_{\delta,\mu}(p) \geqslant 0$ is a C^∞-function that vanishes outside of $B_\delta(\mu) = \{\, p \in \mathbb{R}^d \mid\ |p - \mu| \leqslant \delta\,\}$. We have then (suppressing the subscript μ),

$$\|\varphi(g_\delta)\Omega\|^2 = \int \tilde{\mathscr{W}}_1(p)\widetilde{(g_\delta^*)}(p)\tilde{g}_\delta(-p)\, dp = 0\,,$$

since $\operatorname{supp}\tilde{\mathscr{W}}_1 \subset \overline{V}_+$. Thus,

$$\varphi(g_\delta)\Omega = 0\,, \tag{5.2}$$

for any $\delta \in (0, m)$. In addition,

$$\begin{aligned} W_2(g_\delta, f) &= \int \tilde{\mathscr{W}}_1(p)\tilde{g}_\delta(p)\tilde{f}(-p)\, dp \\ &= \int \tilde{\mathscr{W}}_1(p)\chi_{\delta,\mu}(p)|\tilde{f}(-p)|^2\, dp > 0\,, \end{aligned}$$

since $\tilde{\mathscr{W}}_1(p) \geqslant 0$ and μ satisfies (5.1). We may then normalize g_δ by multiplying by a positive constant, which, of course, depends on δ, so that, retaining the symbol g_δ for the normalized function,

$$W_2(g_\delta, f) = \langle \Omega, \varphi(g_\delta)\varphi(f)\Omega\rangle = 1\,. \tag{5.3}$$

Defining

$$F(p,q) = \tilde{g}(q-p)\tilde{f}(-q) = \chi_{\delta,\mu}(q-p)\tilde{f}(p-q)^*\tilde{f}(-q)\,,$$

we note that for any $q \in \overline{V}_+$,

$$(\operatorname*{supp}_{\mathrm{P}} \mathrm{F}(\mathrm{p}, \mathrm{q})) \cap \overline{V_+^m} = \phi\,, \tag{5.4}$$

by construction. Let $\{h_i\}_{i=1}^n \subset \mathscr{C}(\mathbb{R}^d)$. Equation (3.2) tells us that

$$W_{n+2}(h_1, \ldots, h_n, g, f) = \sum_{\text{part.}} W_{r_1}^{\mathrm{T}}(\cdots)W_{r_2}^{\mathrm{T}}(\cdots)\cdots W_{r_s}^{\mathrm{T}}(\cdots)\,,$$

where the sum is over partitions of $\{h_1, \ldots, h_n, g, f\}$, and in each term of the sum, the arguments of the TVEV's are the functions in the suitable elements of the given partition, ordered naturally. But (5.2) and (5.4) entail that the only nonzero terms in this sum are subpartitions of the partition $\{h_1, \ldots, h_n\}\{g, f\}$, and by (3.2) this resulting sum is equal to

$$W_n(h_1, \ldots h_n)W_2(g, f) = W_n(h_1, \ldots, h_n)\,, \tag{5.5}$$

using (5.3). Therefore, given any $\Phi = \Pi_{i=1}^n \varphi(h_i)\Omega$, $\{h_i\}_{i=1}^n \subset \mathscr{C}(\mathbb{R}^d)$, we have

$$\begin{aligned} &\langle \Phi, [\varphi(g)\varphi(f) - I]\Omega\rangle \\ &\quad = \langle \Phi, \varphi(g)\varphi(f)\Omega\rangle - \langle \Phi, \Omega\rangle \\ &\quad = W_{n+2}(h_n^*, \ldots, h_1^*, g, f) - W_n(h_n^*, \ldots, h_1^*) = 0\,, \end{aligned}$$

by (5.5). Since such vectors are total in $\mathscr{H}_{\mathrm{W}}$,

$$\varphi(g_{\delta,\mu})\varphi(f)\Omega = \Omega\,, \tag{5.6}$$

for all $g_{\delta,\mu}$ as constructed.

Remark: The freedom of choosing any $\delta \in (0, m)$ will be important below.

Remark 5.2: In Lemma 5.1 we have restricted our attention to test functions $f \in \mathscr{B}(\mathbb{R}^d)$ for pedagogical reasons. We wish to point out that, in this special case of one field operator on the vacuum, more general test functions may be admitted. If for example, $f \in \mathscr{K}_R(\mathbb{R}^d)$ such that $(I - P_\Omega)\phi(f)\Omega \neq 0$, then $\tilde{f}(p) = 0$ if and only if $\tilde{f}(-p) = 0$ and, as before, $(\operatorname{supp}\tilde{f})^0 \cap \operatorname{supp}\tilde{\mathscr{W}}_1 \cap \overline{V_+^m} \neq \emptyset$. However, the supports of $\tilde{f}$ and $\tilde{\mathscr{W}}_1$ can be such that the intersection in (5.1) is empty. In that case, since $\varphi(h)\Omega = 0$ for any test function h such that $\operatorname{supp}\tilde{h}(-p) \cap \overline{V_+^m} = \emptyset$, one can find a test function $h \in \mathscr{K}(\mathbb{R}^d)$ such that $\tilde{h}(p)$ vanishes for all $p \in \mathbb{R}^d$ for which $p^0 \leqslant 0$ [so that $\varphi(h)\Omega = 0$] and r_{f+h} is such that

$$\partial(\operatorname{supp}(\tilde{f} + \tilde{h})) \cap \partial T_{r_{f+h}} \cap \operatorname{supp}\tilde{\mathscr{W}}_1 \neq \emptyset\,.$$

Then one may carry through the construction for the test function $f + h$ to find a $g \in \mathscr{K}(\mathbb{R}^d)$ such that $\varphi(g)\Omega = 0$ and

$$\Omega = \varphi(g)\varphi(f + h)\Omega = \varphi(g)\varphi(f)\Omega\,.$$

This argument is limited to only one field operator on the vacuum, and we are interested in finding a dense subset of $\mathscr{C}(\mathbb{R}^d)$ for which the construction may be pushed through for arbitrarily many field operators on the vacuum. Hence the choice of $\mathscr{B}(\mathbb{R}^d)$.

Let now $\{f_i\}_{i=1}^n \subset \mathscr{B}(\mathbb{R}^d)$ and define the corresponding g_i's as in the proof above. We shall need to show that, for suitably many $f_i \in \mathscr{B}(\mathbb{R}^d)$, the vector $\Pi_{i=1}^n \varphi(g_i)\Pi_{i=1}^n \varphi(f_i)\Omega$ is nonzero. Given any $f \in \mathscr{K}(\mathbb{R}^d)$, we define, for any $t \in \mathbb{C}$, $f_t(x) \in \mathscr{K}(\mathbb{R}^d)$ as the function with Fourier transform

$$\tilde{f}_t(p) = e^{itp^1}\tilde{f}(p)\,,$$

where $p = (p^0, p^1, \ldots, p^{d-1})$. We first note the following obvious fact.

Lemma 5.3: Let $n \in \mathbb{N}$, $\{f_i\}_{i=1}^n \subset \mathscr{K}(\mathbb{R}^d)$, and $\{t_i\}_{i=1}^n \subset \mathscr{C}$ be arbitrary. Then

$$\left\langle \Omega, \prod_{i=1}^{n} \varphi(f_{i,t_i})\Omega \right\rangle = W_n(f_{1,t_1},\ldots,f_{n,t_n})$$

is entire analytic in $(t_1,\ldots,t_n)\in\mathbb{C}^n$.

Proof: Obvious, due to the compact support of $\tilde{f}_i$, $i=1,\ldots,n$.

Corollary 5.4: For $n\in\mathbb{N}$, $\{f_i\}_{i=1}^n \subset \mathscr{B}(\mathbb{R}^d)$ arbitrary and $\{g_i\}_{i=1}^n \subset \mathscr{K}(\mathbb{R}^d)$ as specified above, the set of all $(t_1,\ldots,t_{n-1})\in\mathbb{R}^{n-1}$ such that

$$\prod_{i=1}^{n}\varphi(g_{i,t_i})\prod_{i=1}^{n}\varphi(f_{i,t_i})\Omega = 0 \quad (t_n=0)$$

is nowhere dense in $\mathbb{R}^{n-1}$.

Remarks: (1) If $f\in\mathscr{B}(\mathbb{R}^d)$, then so is f_t, $\forall\ t\in\mathbb{C}$. (2) A choice of $\delta_i\in(0,m)$, $i=1,\ldots,n$, is tacitly assumed. The result holds for each such choice.

Proof: By Lemma 5.3,

$$\left|\left|\prod_{i=1}^{n}\varphi(q_{i,t_i})\prod_{i=1}^{n}\varphi(f_{i,t_i})\Omega\right|\right|^2 = W_{4n}((f_n^*)_{t_n},(f_{n-1}^*)_{t_{n-1}},\ldots,(g_1^*)_{t_1},g_{1,t_1},\ldots,f_{n,t_n}) \quad (5.7)$$

is entire analytic in $(t_1,\ldots,t_n)\in\mathbb{C}^n$. By Theorem 3.6 and (3.2), after taking the limits $t_i\to\infty$, $i=1,\ldots,n-1$, one by one, (5.7) becomes

$$\prod_{i=1}^{n} W_4(f_i^*,g_i^*,g_i,f_i) = \prod_{i=1}^{n}\|\varphi(g_i)\varphi(f_i)\Omega\|^2 = 1\,,$$

using Lemma 5.1 in the second equality. Therefore, given any open set $\mathscr{U}\subset\mathbb{R}^{n-1}$, (5.7) cannot vanish for all $(t_1,\ldots,t_{n-1})\in\mathscr{U}$. Indeed, since (5.7) is continuous in t_i, $\forall$ i, it cannot vanish for all $(t_1,\ldots,t_{n-1})$ in a set dense in $\mathscr{U}$. Thus, the conclusion of the corollary follows. In fact, using Theorem 24 in Ref. 22, i.e., since locally the zeros of an analytic function of n variables form finitely many $n-1$ (complex)-dimensional manifolds, one can show that the mentioned set has Lebesgue measure zero in $\mathbb{R}^{n-1}$.

To proceed further, we prove a straightforward result relying on the support properties of the functions $\{f_i\}_{i=1}^n \subset \mathscr{B}(\mathbb{R}^d)$ and of the corresponding functions $\{g_i\}_{i=1}^n$ constructed according to the proof of Lemma 5.1. For such functions, let $e_i=g_i$, $i=1,\ldots,n$, and $e_i=f_{i-n}$, $i=n+1,\ldots,2n$, and denote by $\Pi(2n)$ the set of all partitions of the set $\{1,\ldots,2n\}$ (which we identify with any ordered set of $2n$ elements).

Lemma 5.5: Let $\pi\in\Pi(2n)$ and $\sigma=\{i_1,\ldots,i_l\}\in\pi$. If

$$\sum_{i\in\sigma\cap\{1,\ldots,n\}}(r_i-\delta_i) > \sum_{i+n\in\sigma\cap\{n+1,\ldots,2n\}} r_i\,,$$

where $r_i=r_{f_i}$, $i=1,\ldots,n$, then $\Phi=\Pi_{j=1}^{l}\varphi(e_{i_j})\Omega=0$.

Proof: For all $p_i\in\overline{V}_+$, $i=0,\ldots,s-1$, the product

$$\tilde{e}_{j_1}(p_1-p_0)\tilde{e}_{j_2}(p_2-p_1)\cdots\tilde{e}_{j_s}(-p_{s-1})\,,$$

$$j_1,\ldots,j_s\in\{n+1,\ldots,2n\}\,,$$

vanishes if

$$p_0^0 > \sum_{n=1}^{s} r_{j_n}\,, \quad (5.8)$$

as may easily be seen. And for $p_0,\ldots,p_k\in\overline{V}_+$, the product

$$\tilde{e}_{j_1}(p_1-p_0)\tilde{e}_{j_2}(p_2-p_1)\cdots\tilde{e}_{j_k}(p_k-p_{k-1})\,,$$

$$j_1,\ldots,j_k\in\{1,\ldots,n\}\,,$$

vanishes if

$$p_k^0 < \sum_{n=1}^{k}(r_{j_n}-\delta_{j_n})\,, \quad (5.9)$$

by definition of the functions g_i. Thus, for any $p_1,\ldots,p_{l-1}\in\overline{V}_+$,

$$(\operatorname{supp}_{p_0}\tilde{e}_{i_1}(p_1-p_0)\tilde{e}_{i_2}(p_2-p_1)\cdots\tilde{e}_{i_l}(-p_{l-1}))\cap\overline{V}_+ = \varnothing\,,$$

given the hypothesis of the lemma. This implies, however [see (3.5)], by the support properties of the Fourier transforms of the difference-variable Wightman functions, that the inner product of Φ with any vector in $\mathscr{P}(\mathbb{R}^d)\Omega$ vanishes. In particular, $\|\Phi\|^2=0$.

Corollary 5.6: Let $\varphi(x)$ be a real scalar quantum field with mass gap $m>0$. If $n\in\mathbb{N}$ and $\{f_i\}_{i=1}^n \subset \mathscr{A}_1(\mathbb{R}^d)$ are arbitrary, let $\rho=\Sigma_{i=1}^n r_{f_i}$. Let further $\{h_i\}_{i=1}^k \subset \mathscr{A}_2(\mathbb{R}^d)$ by any set satisfying $\Sigma_{i=1}^k r_{h_i} > \rho+\Delta$, $\Delta>0$. Then if $\{g_i\}_{i=1}^k$ is any set of annihilators for $\{h_i\}_{i=1}^k$, as constructed in Lemma 5.1, satisfying the additional constraint $\Sigma_{i=1}^k\delta_i<\Delta$, we have

$$\prod_{i=1}^{k}\varphi(g_i)\prod_{j=1}^{n}\varphi(f_j)\Omega = 0\,.$$

Proof: The argument proceeds as in the proof of Lemma 5.5.

Remark: We have employed our freedom to choose $\delta_i\in(0,m)$, $i=1,\ldots,k$, such that the additional constraint $\Sigma_{i=1}^k\delta_i<\Delta$, for any given $\Delta>0$, may be satisfied.

Proposition 5.7: Let $\varphi(x)$ be a real scalar quantum field with mass gap $m>0$. Let $0<\Delta<m$ and $n\in\mathbb{N}$ be arbitrary and $\{f_i\}_{i=1}^n \subset \mathscr{B}(\mathbb{R}^d)$ such that, with g_i, $i=1,\ldots,n$, as constructed in Lemma 5.1 under the additional constraint that $\Sigma_{i=1}^n\delta_i<\Delta$,

(i) $$\prod_{i=1}^{n}\varphi(g_i)\prod_{i=1}^{n}\varphi(f_i)\Omega \neq 0\,;$$

then there exists a set $\{\hat{g}_i\}_{i=1}^n \subset \mathscr{K}(\mathbb{R}^d)$ such that

(ii) $$\prod_{i=1}^{n}\varphi(\hat{g}_i)\prod_{i=1}^{n}\varphi(f_i)\Omega = \Omega\,,$$

(iii) $$\varphi(\hat{g}_i)\Omega = 0, \quad i=1,\ldots,n\,,$$

and

(iv) $$\prod_{i=1}^{n}\varphi(\hat{g}_i)\prod_{i\in\mathscr{N}}\varphi(f_i)\Omega = 0, \quad \forall\ \mathscr{N}\subsetneq\{1,\ldots,n\}\,.$$

Remark: Here we have again used our freedom of choice of $\delta_i\in(0,m)$ that the construction in Lemma 5.1 allows us, in order to satisfy the additional constraint, for fixed $n\in\mathbb{N}$. Thus, the annihilator g_j for the element f_j in the set $\{f_i\}_{i=1}^n \subset \mathscr{B}(\mathbb{R}^d)$ depends on n, as well as on f_j.

Proof: $\{\hat{g}_i\}_{i=1}^n$ is simply chosen to be $\{c_i g_i\}_{i=1}^n$, with $\{c_i\}_{i=1}^n \subset \mathbb{C}$ selected so that

$$\left|\left|\prod_{i=1}^{n}\varphi(\hat{g}_i)\prod_{i=1}^{n}\varphi(f_i)\Omega\right|\right| = 1\,, \quad (5.10)$$

which, by (i), is possible. Then (iii) follows from Eq. (5.2). We next verify (ii).

If there exists a $\sigma_{k_0}\in\pi\in\Pi(2n)$ such that

$$\sum_{i\in\sigma_{k_0}\cap\{1,\ldots,n\}}(r_i-\delta_i)+m\leqslant\sum_{i+n\in\sigma_{k_0}\cap\{n+1,\ldots,2n\}}r_i\,, \tag{5.11}$$

then, clearly, there must exist a $\sigma_{k_1}\in\pi$ such that

$$\sum_{i\in\sigma_{k_1}\cap\{1,\ldots,n\}}(r_i-\delta_i)>\sum_{i+n\in\sigma_{k_1}\cap\{n+1,\ldots,2n\}}r_i \tag{5.12}$$

[note that (5.11) cannot be satisfied by $\sigma_{k_0}=\{1,\ldots,2n\}$, so there must always exist another $\sigma_{k_1}\in\pi$, if (5.11) obtains], since otherwise

$$\sum_{i\in\sigma_k\cap\{1,\ldots,n\}}(r_i-\delta_i)\leqslant\sum_{i+n\in\sigma_k\cap\{n+1,\ldots,2n\}}r_i,$$
$$\forall\ \sigma_k\in\pi\quad\text{with}\quad\sigma_k\neq\sigma_{k_0}\,,$$

implies with (5.11)

$$\sum_{i=1}^{n}\delta_i\geqslant m\,,$$

which contradicts the hypothesis. As in (5.8) and (5.9) in Lemma 5.5, one sees that for any $\delta_k\in\pi$ with $\delta_k\cap\{1,\ldots,n\}=\{i_1,\ldots,i_l\}$ and $\sigma_k\cap\{n+1,\ldots,2n\}=\{i_{l+1},\ldots,i_{l+s}\}$, if $p_k\in\overline{V}_+$, $k=1,\ldots,l+s-1$,

$$\left(\operatorname*{supp}_{p_0}\tilde{e}_{i_1}(p_1-p_0)\tilde{e}_{i_2}(p_2-p_1)\cdots\tilde{e}_{i_{l+s}}(-p_{l+s-1})\right)\cap\overline{V}_+^{\,m}=\varnothing\,, \tag{5.13}$$

unless

$$\sum_{j=1}^{l}(r_j-\delta_j)+m\leqslant\sum_{j=l+1}^{l+s}r_j\,.$$

But in this latter case there exists a $\sigma_{k_1}\in\pi$ such that (5.12) holds, which implies, by Lemma 5.5, that the associated product on the vacuum vanishes.

We are thus in the position to prove that for any $\{h_i\}_{i=1}^{l}\subset\mathscr{C}(\mathbb{R}^d)$,

$$\left\langle\Omega,\prod_{i=1}^{l}\varphi(h_i)\prod_{i=1}^{n}\varphi(\hat{g}_i)\prod_{i=1}^{n}\varphi(f_i)\Omega\right\rangle=\left\langle\Omega,\prod_{i=1}^{l}\varphi(h_i)\Omega\right\rangle\left\langle\Omega,\prod_{i=1}^{n}\varphi(\hat{g}_i)\prod_{i=1}^{n}\varphi(f_i)\Omega\right\rangle. \tag{5.14}$$

To see this, we use (3.2) to decompose

$$W_{2n+l}(h_1,\ldots,h_l,\hat{g}_1,\ldots,\hat{g}_n,f_1,\ldots,f_n)=\sum_{\text{part.}}W_{r_1}^{\mathrm{T}}(\cdots)W_{r_2}^{\mathrm{T}}(\cdots)\cdots W_{r_s}^{\mathrm{T}}(\cdots)$$

into a sum, over all partitions of $\{h_1,\ldots,h_l,\hat{g}_1,\ldots,\hat{g}_n,f_1,\ldots,f_n\}$, of products of the appropriate TVEV's. Since the Fourier transforms $\tilde{\mathscr{W}}_n^{\mathrm{T}}(p_1,\ldots,p_n)$ of the difference-variable TVEV's $\mathscr{W}_n^{\mathrm{T}}(\xi_1,\ldots,\xi_n)$ have support only in $(\overline{V_+^m})^n$, the only nonzero terms in the sums on the right-hand side of the equation above are terms arising from subpartitions of the partition

$$\{h_1,\ldots h_l\}\{\hat{g}_1,\ldots,\hat{g}_n,f_1,\ldots,f_n\}\,.$$

But the sum over such subpartitions, again by (3.2), is equal to

$$W_l(h_1,\ldots,h_l)W_{2n}(\hat{g}_1,\ldots,\hat{g}_n,f_1,\ldots,f_n)\,,$$

which yields (5.14). Then, as in the proof of Lemma 5.1, one uses (5.14) and (5.10) to show that (ii) is indeed correct.

(iv) follows from Lemma 5.5, once one notices that

$$\sum_{i=1}^{n}(r_i-\delta_i)>\sum_{i\in\mathscr{N}}r_i\,,$$

for any $\mathscr{N}\overset{\subset}{\neq}\{1,\ldots,n\}$, since $r_i>m$, $\forall\ i$, and $\Sigma_{i=1}^{n}\delta_i<m$. This completes the proof of the proposition.

We can now define the set of vectors $\mathscr{E}$ mentioned in Theorem 4.1 and prove the assertions made about it. Let $\mathscr{E}_1$ be the set of all vectors

$$\prod_{i=1}^{n}\varphi(f_i)\Omega,\quad n\in\mathbb{N},\quad\{f_i\}_{i=1}^{n}\subset\mathscr{B}(\mathbb{R}^d)\,,$$

satisfying condition (i) in Proposition 5.7—that is to say, for fixed $n\in\mathbb{N}$, $\{f_i\}_{i=1}^{n}\subset\mathscr{B}(\mathbb{R}^d)$, the vector $\Pi_{i=1}^{n}\varphi(f_i)\Omega$ is in $\mathscr{E}_1$ if for some choice of $0<\Delta<m$ and $\{\delta_i\}_{i=1}^{n}\subset(0,m)$ satisfying $\Sigma_{i=1}^{n}\delta_i<\Delta$, and with g_{i,δ_i} an annihilator, corresponding to f_{i,δ_i}, constructed in Lemma 5.1, the vector

$$\prod_{i=1}^{n}\varphi(g_{i,\delta_i})\prod_{j=1}^{n}\varphi(f_j)\Omega$$

does not vanish. With each set $\{f_i\}_{i=1}^{n}\subset\mathscr{B}(\mathbb{R}^d)$ we associate the number $\rho=\Sigma_{i=1}^{n}r_{f_i}$. We define the set $\mathscr{E}$ as the set of all vectors

$$\sum_{j=1}^{N}\left(\prod_{j=1}^{n_j}\varphi(f_{i_j})\right)\Omega,\quad N\in\mathbb{N},$$
$$\{n_j\}_{j=1}^{N}\subset\mathbb{N},\quad\{f_{i_j}\}\subset\mathscr{B}(\mathbb{R}^d)\,, \tag{5.15}$$

such that $\Pi_{i=1}^{n_j}\varphi(f_{i_j})\Omega\in\mathscr{E}_1$, $\forall\ j=1,\ldots,N$, and such that there exists a $j_0\in\{1,\ldots,N\}$ so that $\rho_{j_0}>\rho_j$, $\forall\ j\neq j_0$, $j=1,\ldots,N$, and so that the Δ $(0<\Delta<m)$ for which condition (i) in Proposition 5.7 is satisfied by the vector $\Pi_{i=1}^{n_{j_0}}\varphi(g_{i_{j_0}})\Pi_{i=1}^{n_{j_0}}\varphi(f_{i_{j_0}})\Omega$ satisfies $\Delta<\rho_{j_0}-\rho_j$, $\forall\ j\neq j_0$. One sees that Proposition 5.7 and Corollary 5.6 then yield assertions (i), (ii), and (iii) of Theorem 4.1 for each such vector. It remains, then, only to show that $\mathscr{E}$ is dense in $\mathscr{H}_{\mathrm{w}}$.

Lemma 5.8: Let $\varphi(x)$ be a real scalar quantum field with mass gap $m>0$. Then the set $\mathscr{E}$ defined above is dense in $\mathscr{H}_{\mathrm{w}}$.

Proof: By Corollary 5.4 and the strong continuity in t_i of vectors of the form $\Pi\varphi(f_{i,t_i})\Omega$, and because, as already mentioned, Corollary 5.4 holds for each choice of $0<\Delta<m$, all the vectors in the set

$$\left[\sum_{j=1}^{N}\left(\prod_{i=1}^{n_j}\varphi(f_{i_j})\right)\Omega\,|\,N\in\mathbb{N},\ \{n_j\}_{j=1}^{N}\subset\mathbb{N},\ \{f_{i_j}\}\subset\mathscr{B}(\mathbb{R}^d)\right]$$

are limits of strongly convergent sequences from $\mathscr{E}$. Since, as previously remarked, every element of $\mathscr{A}_2(\mathbb{R}^d)$ can be arbitrarily well approximated in the topology of $\mathscr{C}(\mathbb{R}^d)$ by elements of $\mathscr{B}(\mathbb{R}^d)$, the vectors in the set

$$\mathscr{G}=\left[\sum_{j=1}^{N}\left(\prod_{i=1}^{n_j}\varphi(f_{i_j})\right)\Omega\,|\,N\in\mathbb{N},\ \{n_j\}_{j=1}^{N}\subset\mathbb{N},\ \{f_{i_j}\}\subset\mathscr{A}_2(\mathbb{R}^d\right]$$

are, as well, limits of strongly convergent sequences from $\mathscr{E}$.

We recall next that, for any compact $\mathcal{O} \subset \mathbb{R}^d$ and any $h \in \mathcal{D}(\mathcal{O})$, the vacuum Ω is separating for $\varphi(h)$. Thus, if the vector

$$\prod_{i=1}^{n} \varphi(h_i)\Omega, \quad \{h_i\}_{i=1}^{n} \subset \bigcup_{\mathcal{O} \subset \mathbb{R}^d} \mathcal{D}(\mathcal{O}),$$

is nonzero, one must have $\varphi(h_i)\Omega \neq 0, \ \forall \ i = 1,..., n$. Indeed, unless

$$\varphi(h_i) = \langle \Omega, \varphi(h_i)\Omega \rangle I,$$

one must have $(I - P_\Omega)\varphi(h_i)\Omega \neq 0, \ \forall \ i = 1,..., n$. Since $\mathcal{A}_1(\mathbb{R}^d)$ is dense in $\mathcal{K}(\mathbb{R}^d)$ and every function $h \in \mathcal{D}(\mathbb{R}^d)$ is a limit of some sequence $\{f_i\} \subset \mathcal{K}(\mathbb{R}^d)$ in the topology of $\mathcal{E}(\mathbb{R}^d)$ [since $\mathcal{K}(\mathbb{R}^d)$ is dense in $\mathcal{E}(\mathbb{R}^d)$], it follows at once that every vector of the form

$$\sum_{j=1}^{N} \left(\prod_{i=1}^{n_j} \varphi(f_{i_j}) \right) \Omega, \quad N \in \mathbb{N},$$
$$\{n_j\}_{j=1}^{N} \subset \mathbb{N}, \quad \{f_{i_j}\} \subset \mathcal{D}(\mathbb{R}^d), \tag{5.16}$$

is a limit of a strongly convergent sequence from $\mathcal{G}$. But vectors of the form (5.16) are dense in $\mathcal{H}_W$, since $\mathcal{D}(\mathbb{R}^d)$ is dense in $\mathcal{E}(\mathbb{R}^d)$.■

Remark: This is the sole result whose proof depends on the existence of (sufficiently many) strictly localizable test functions. If there were another class $\mathcal{J}$ of test functions dense in $\mathcal{E}(\mathbb{R}^d)$ such that Ω is separating for $\{\varphi(g) | \ g \in \mathcal{J}\}$, then one could do without strict localizability.

[1] W. Driessler and S. J. Summers, "On the Self-adjointness of Quantum Field Operators," work in progress.
[2] D. Ruelle, "On the Asymptotic Condition in Quantum Field Theory," Helv. Phys. Acta **35**,147 (1962).
[3] R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics, and All That* (Benjamin, New York, 1964).
[4] A. N. Vasil'ev, "Theory of Representations of a Topological (Non-Banach) Involutory Algebra," Theor. Math. Phys. **2**, 113 (1970).
[5] R. T. Powers, "Self-Adjoint Algebras of Unbounded Operators," Commun. Math. Phys. **21**, 85 (1971).
[6] A. S. Wightman and L. Garding, "Fields as Operator-Valued Distributions in Relativistic Quantum Theory," Ark. Fysik. **28**, 129 (1965).
[7] A. S. Wightman, "Introduction to Some Aspects of the Relativistic Dynamics of Quantized Fields," in *Cargèse Lectures in Theoretical Physics, 1964*, edited by M. Lévy (Gordon and Breach, New York, 1967).
[8] A. Jaffe, "Entire Functions of the Free Field," Ann. Phys. (N. Y.) **32**, 127 (1965).
[9] J. Bümmerstede and W. Lücke, "Haag-Ruelle-Hepp Scattering Formalism for Essentially Local Non-Localizable Fields," Commun. Math. Phys. **37**, 121 (1974).
[10] V. Ya. Fainberg and M. A. Soloviev, "How Can Local Properties Be Described in Field Theories Without Strict Locality?," Ann. Phys. (N. Y.) **113**, 421 (1978).
[11] A. Jaffe, "High-Energy Behavior in Quantum Field Theory," Phys. Rev. **158**, 1454 (1967).
[12] A. Rieckers, "Power Series of the Free Field as Operator-Valued Functionals on Spaces of Type S," Int. J. Theor. Phys. **4**, 55 (1971).
[13] W. Lücke, "Functions of the Free Field: Examples for Essentially Local Non-Localizable Fields," J. Phys. A **7**, 2258 (1974).
[14] A. Björck, "Linear Partial Differential Opertors and Generalized Distributions" Ark. Mat. **6**, 351 (1966).
[15] H. Komatsu, "Ultradistributions, I," J. Fac. Sci. Univ. Tokyo, Sect. 1A **20**, 25 (1973).
[16] I. M. Gel'fand and G. E. Shilov, *Generalized Functions* (Academic, New York, 1968), Vol. 2.
[17] F. Constantinescu and W. Thalheimer, "Euclidean Green's Functions for Jaffe Fields," Commun. Math. Phys. **38**, 299 (1974).
[18] F. Constantinescu and W. Thalheimer, "Ultradistributions and Quantum Fields: Fourier-Laplace Transforms and Boundary Values of Analytic Functions," Rep. Math. Phys. **16**, 167 (1979).
[19] R. Jost, *The General Theory of Quantized Fields* (American Mathematical Society, Providence, R. I., 1965).
[20] H. J. Borchers, "On the Vacuum State in Quantum Field Theory, II," Commun. Math. Phys. **1**, 57 (1965).
[21] After this paper was submitted for publication and distributed in preprint form, we received a letter from Professor D. Buchholz (Hamburg) in which our argument was improved at a crucial point—where the product of the annihilators applied to $\Pi\varphi(f_i)\Omega$ is shown to be nonzero—yielding a proof that every element of $\mathcal{P}_c(\mathbb{R}^d)$ is vacuum reducible.
[22] H. Behnke and P. Thullen, *Theorie der Funktionen mehrerer komplexer Veränderlichen* (Springer, Berlin, 1970), 2nd ed.

Galilean field theories and nonlocal S-matrix symmetries [a)]

Wolf-Dieter Garber
Institut für Theoretische Physik, Bunsenstr. 9, D-3400 Göttingen, West Germany

(Received 27 July 1982; accepted for publication 17 December 1982)

Any operator that commutes with the S matrix and is additive, i.e., transforms an asymptotic incoming n-particle state as a sum of its constituent one-particle states, is called a symmetry of the S matrix. The structure of local S-matrix symmetries in Galilean field theories is known. In this paper, S-matrix symmetries that are *nonlocal*, i.e., may transform asymptotic fields into nonlocal operators, are investigated under the assumption that there is a finite number of such symmetries in the theory.

PACS numbers: 11.10.Lm, 11.10.Jj, 11.20.Fm

1. INTRODUCTION

In Ref. 1, symmetries of the S matrix have been considered in the context of Galilean invariant field theories. A symmetry of the S matrix Q in a theory of asymptotic fields $\phi_\alpha^{\rm ex}(x)$ with corresponding annihilation and creation operators $a^{\rm ex}(\mathbf{p},\alpha)$, $(a^{\rm ex}(\mathbf{p},\alpha))^*$is defined to be an additive operator, i.e., one that is bilinear in incoming fields,

$$Q=\int d^3p\, d^3q\, Q^{\alpha\beta}(\mathbf{p},\mathbf{q})(a^{\rm in}(\mathbf{p},\alpha))^*\, a^{\rm in}(\mathbf{q},\beta) \tag{1.1}$$

(summation convention) and which, in addition, commutes with the S matrix

$$(S^*\Psi^{\rm in}/Q\Phi^{\rm in})=(Q^*\Psi^{\rm in}/S\Phi^{\rm in}) \tag{1.2}$$

as a form on a suitable domain. Those S-matrix symmetries that acted locally, i.e., for which $Q,\phi_\alpha^{\rm ex}(x)$ is again a local Galilei field, had special properties:

1. (i) $Q^{\alpha\beta}=0$ for $m_\alpha\neq m_\beta$, where m_α is the mass of the state $(a^{\rm ex}(\mathbf{p},\alpha))^*\Omega$.

(ii) $Q^{\alpha\beta}=q^{\alpha\beta}(\mathbf{p},\partial)\ \delta(\mathbf{p}-\mathbf{q})$, where $q^{\alpha\beta}$ is a polynomial in $\mathbf{p}$ and derivatives in p_k, $k=1,2,3$.

Furthermore, in a theory with scalar particles and with nontrivial scattering in the sense that the particles can be so ordered that the elastic two-particle scattering amplitude, for any two consecutive particles, does not vanish in some open subset of the sets of momenta allowed by energy and momentum conservation (some particles may be counted more than once!),

$$\begin{aligned}2.\ q^{\alpha\beta}=\delta^{\alpha\beta}\{&a(2m)^{-1}\mathbf{p}^2+\mathbf{bp}+\mathbf{c}\partial\\&+d^{ik}(p_i\,\partial_k-p_k\,\partial_i)\\&+e[\,p_i,\partial_i\,]_+ +f\partial^2\}+g^{\alpha\beta}\end{aligned} \tag{1.3}$$

with constants a, b_k, c_k, d^{ik}, e, f, and $g^{\alpha\beta}$. Here $m=m_\alpha=m_\beta$ since, by (1.1), it suffices to consider a single mass multiplet. Thus, Q is a linear combination of generators of the 12 parameter Schrödinger group S together with translation invariant generators (given by $g^{\alpha\beta}$) which, however, need not commute with the Hamiltonian. The situation is thus fairly analogous to the relativistic case where S is replaced by the Poincaré group P in statement 2.

[a)] This work contains parts of the author's "Habilitatiensschrift", accepted by the Physics Department, University of Göttingen.

Now, it is easy to construct *nonlocal* S-matrix symmetries, at least in a free field theory. Consider a single scalar free Schrödinger field $\phi(x)$ with corresponding annihilation operator $a(\mathbf{p})$ and define

$$Q=\int d^3p\, d^3q a^*(\mathbf{p})\, a(\mathbf{q})\exp(\mathbf{p}-\mathbf{q})^2, \tag{1.4}$$

which is additive but for which $[Q,\phi(x)]$ is not a field obeying Galilean locality. Statement 1 is thus manifestly false.

Even for nonlocal S-matrix symmetries, one could try to retain the physical idea that a finite set of quantum numbers should suffice to label free particles. This would correspond to the assumption that only a finite number of S-matrix symmetries should be present in the theory. It will be shown in the next sections that this assumption is quite restrictive. With it, statement 1 can again be proven so that, in fact, Q turns out to be local and statement 2 is valid, too. The finiteness assumption can even be relaxed somewhat. Statement 1 is true under the weaker hypothesis that infinitely many S-matrix symmetries are allowed which must, however, be grouped into finite-dimensional spaces invariant under the Galilei group. This will now be shown.

2. POLYNOMIAL DEPENDENCE IN DERIVATIVES

Consider a theory of Galilei covariant asymptotic incoming fields $\phi_\alpha^{\rm in}(x)$ with creation and annihilation operators $(a^{\rm in}(\mathbf{p},\alpha))^*$, $a^{\rm in}(\mathbf{p},\alpha)$ obeying Galilei covariant commutation relations

$$[a^{\rm in}(\mathbf{p},\alpha),(a^{\rm in}(\mathbf{q},\beta))^*]_{\pm}=\delta_{\alpha\beta}\,\delta(\mathbf{p}-\mathbf{q}), \tag{2.1}$$

which transform as follows:

$$U^{-1}(\mathbf{a})\,a^{\rm in}(\mathbf{p},\alpha)\,U(\mathbf{a})=\exp\{-i\mathbf{pa}\}\,a^{\rm in}(\mathbf{p},\alpha), \tag{2.2}$$

$$\begin{aligned}&U^{-1}(b)\,a^{\rm in}(\mathbf{p},\alpha)\,U(b)\\&\quad=\exp\{ib((2m_\alpha)^{-1}\mathbf{p}^2+W_\alpha)\}\,a^{\rm in}(\mathbf{p},\alpha),\end{aligned} \tag{2.3}$$

$$U^{-1}(\mathbf{v})\,a^{\rm in}(\mathbf{p},\alpha)\,U(\mathbf{v})=a^{\rm in}(\mathbf{p}-m_\alpha\mathbf{v},\alpha), \tag{2.4}$$

$$U^{-1}(R)\,a^{\rm in}(\mathbf{p},\alpha)\,U(R)=(D^s(R))_{\alpha\beta}\,a^{\rm in}(R^{-1}\mathbf{p},\beta) \tag{2.5}$$

under space translations, time translations, boosts, and rotations, respectively. Here, m_α is the mass of $\phi_\alpha^{\rm in}(x)$ and the constant W_α is the inner energy; D^s denotes the irreducible $(2s+1)$-dimensional representation of the rotation group.

Any S-matrix symmetry of the form (1.1) has, by Bargmann's mass superselection rule,[2] vanishing matrix elements

 0022-2488/83/122820-03$02.50

between states of different mass multiplets. This fact, proved again as Lemma 2.1 of Ref. 1, has nothing to do with locality as was emphasized there; part (i) of statement 1 holds in general. It is therefore sufficient to consider, from now on, only fields of the same mass multiplet $m_\alpha = m$.

To prove part (ii) of statement 1, it will be assumed in the following that the space of all S-matrix symmetries decomposes into subspaces that are invariant under the Galilei group and are finite dimensional.

Let E be such a finite-dimensional space of S-matrix symmetries, and consider an S-matrix symmetry $Q \in E$. Since any Galilean transform of Q is again of the form (1.1) by (2.2)–(2.5), and since E is invariant under the Galilei group G, $U^{-1}(g)\,QU(g)$ is again in E for any $g \in G$. Hence, there exists an operator $D(g)$: $E \to E$ with

$$U^{-1}(g)\,QU(g) = D(g)\,Q, \quad g \in G.$$

In any basis $Q_1,\ldots, Q_n \in E$, $D(g)$ is simply a finite-dimensional matrix of complex numbers depending on g. It is even a representation of G since $U(g)$ is. By (2.1), the kernel $Q^{\alpha\beta}$ of Q is, as a distribution, given by

$$Q^{\alpha\beta}(\mathbf{p},\mathbf{q}) = ((a^{in}(\mathbf{p},\alpha))^{*}\Omega\,/\,Q\,(a^{in}(\mathbf{q},\beta))^{*}\Omega)$$

and $D(g)$ can be evaluated on $Q^{\alpha\beta}$ for different elements of $g \in G$ if (2.2)–(2.5) are used:

$$D(\mathbf{a})\,Q^{\alpha\beta}(\mathbf{p},\mathbf{q}) = \exp\{i(\mathbf{p}-\mathbf{q})\,\mathbf{a}\}\;Q^{\alpha\beta}(\mathbf{p},\mathbf{q}), \tag{2.2a}$$

$$D(b)\,Q^{\alpha\beta}(\mathbf{p},\mathbf{q}) = \exp\{-ib\,((2m)^{-1}(\mathbf{p}^2-\mathbf{q}^2) + W_\alpha - W_\beta)\}\;Q^{\alpha\beta}(\mathbf{p},\mathbf{q}), \tag{2.3a}$$

$$D(\mathbf{v})\,Q^{\alpha\beta}(\mathbf{p},\mathbf{q}) = Q^{\alpha\beta}(\mathbf{p}-m\mathbf{v},\mathbf{q}-m\mathbf{v}), \tag{2.4a}$$

$$D(R)\,Q^{\alpha\beta}(\mathbf{p},\mathbf{q}) = D^{s(\alpha)}(R)_{\alpha\gamma}\;\overline{D^{s(\beta)}(R)}_{\beta\delta}\;Q^{\gamma\delta}(R^{-1}\mathbf{p},R^{-1}\mathbf{q}). \tag{2.5a}$$

Since rotations and translations are continuous automorphisms of the test function space $\mathcal{S}$,[3] the representation D is infinitely often differentiable in the group parameters.

First, the translation subgroups of G will be considered which have the following structure:

Lemma 2.1: Let D be an arbitrarily often differentiable representation of the time translation subgroup of G in E, i.e., D fulfills

$$D(c)\,D(d) = D(c+d), \quad c,d \in \mathbb{R}^1. \tag{2.6}$$

Then there is an operator A:$E \to E$ with

$$D(c) = \exp(icA). \tag{2.7}$$

Proof: (2.6) implies $D(0) = 1$. Now differentiate (2.6) with respect to c for $c = 0$,

$$D'(d) = iAD(d), \tag{2.8}$$

where $A := i^{-1}D'(0)$. But the linear differential equation with the constant coefficient matrix A has a unique solution given by (2.7). ∎

The three-dimensional analog is valid, too, and gives the structure for the translation subgroup:

Lemma 2.2: Let D be an infinitely often differentiable representation of the space translation subgroup of G in E, i.e.,

$$D(\mathbf{a})\,D(\mathbf{b}) = D(\mathbf{a}+\mathbf{b}), \quad \mathbf{a},\mathbf{b} \in \mathbb{R}^3. \tag{2.9}$$

Then there are three commuting operators A_k: $E \to E$ with

$$D(\mathbf{a}) = \exp(ia^kA_k), \quad \mathbf{a} = (a^1, a^2, a^3). \tag{2.10}$$

Proof: Choose a basis $\mathbf{e}_k$, $k = 1,2,3$, of $\mathbb{R}^3$. By Lemma 2.1, for each k there are operators A_k such that

$$D(a\mathbf{e}_k) = \exp(iaA_k).$$

(2.10) implies

$$D(a\mathbf{e}_k)\,D(b\mathbf{e}_l) = D(a\mathbf{e}_k + b\mathbf{e}_l) = D(b\mathbf{e}_l)\,D(a\mathbf{e}_k)$$

so that the operators A_k, A_e will commute, too. Hence

$$D(\mathbf{a}) = \exp(ia^k A_k). \quad ∎$$

Of course, in general, the matrices A_k will not be Hermitian.

Next, the specific action of $D(\mathbf{a})$ will be considered. If in (2.2a) new variables $\mathbf{p}$ and $\mathbf{r}$: $= \mathbf{p} - \mathbf{q}$ are introduced and a distribution $\hat{Q}^{\alpha\beta}$ is defined so that

$$\hat{Q}^{\alpha\beta}(\mathbf{r},\mathbf{p}) := Q^{\alpha\beta}(\mathbf{p},\mathbf{q}), \tag{2.11}$$

then all distributions have to be characterized fulfilling

$$D(\mathbf{a})\,\hat{Q}^{\alpha\beta}(\mathbf{r},\mathbf{p}) = \exp(i\mathbf{ra})\,\hat{Q}^{\alpha\beta}(\mathbf{r},\mathbf{p}). \tag{2.12}$$

This is done in

Lemma 2.3: Let D be an infinitely often differentiable representation of the translation subgroup of G in E. For the kernel $\hat{Q}^{\alpha\beta}(\mathbf{r},\mathbf{p})$ of a symmetry $Q \in E$, assume that

$$D(\mathbf{a})\,\hat{Q}^{\alpha\beta}(\mathbf{r},\mathbf{p}) = \exp(i\mathbf{ra})\,\hat{Q}^{\alpha\beta}(\mathbf{r},\mathbf{p}). \tag{2.13}$$

Then all distributions $\hat{Q}^{\alpha\beta}(\mathbf{r},\mathbf{p})$ have support in $\mathbf{r}$ at most in a set B that consists of finitely many vectors in $\mathbb{R}^3$ and is independent of $\mathbf{p},\alpha,\beta$.

Proof: By combining (2.13) with Lemma 2.2,

$$\exp(i\mathbf{aA})\,\hat{Q}^{\alpha\beta}(\mathbf{r},\mathbf{p}) = \exp(i\mathbf{ra})\,\hat{Q}^{\alpha\beta}(\mathbf{r},\mathbf{p})$$

and differentiation with respect to a_k gives

$$A_k\,\hat{Q}^{\alpha\beta}(\mathbf{r},\mathbf{p}) = r_k\,\hat{Q}^{\alpha\beta}(\mathbf{r},\mathbf{p}), \quad k = 1,2,3.$$

Hold k fixed. The matrix A_k can be brought into the Jordan normal form: The space E splits into a direct sum of subspaces E_i ($i = 1,\ldots,m$, say) and within each subspace, there is a basis $Q^i_1,\ldots,Q^i_{n(i)}$ of symmetries such that

$$\sum_m (A_k)^m_l\,Q^i_m = a^i_k\,Q^i_l + Q^i_{l+1}, \quad l = 1,\ldots,n(i)-1,$$

$$\sum_m (A_k)^m_{n(i)}\,Q^i_m = a^i_k\,Q^i_{n(i)}.$$

By taking scalar products with one-particle states, these equations remain valid if Q^i_m is replaced everywhere by the corresponding kernel $(\hat{Q}^i_m)^{\alpha\beta}(\mathbf{r},\mathbf{p})$. For $l = n(i)$,

$$(a^i_k - r_k)(\hat{Q}^i_{n(i)})^{\alpha\beta}(\mathbf{r},\mathbf{p}) = 0$$

so that $(\hat{Q}^i_{n(i)})^{\alpha\beta}(\mathbf{r},\mathbf{p})$ has only support for $r_k = a^i_k$. More generally,

$$(a^i_k - r_k)^{n(i)-l+1}\,(\hat{Q}^i_l)^{\alpha\beta}(\mathbf{r},\mathbf{p}) = 0$$

as one can prove by deduction from $l+1$ to l: Since for $l \neq n(i)$

$$(a^i_k - r_k)\,(\hat{Q}^i_l)^{\alpha\beta}(r,p) + (\hat{Q}^i_{l+1})^{\alpha\beta}(\mathbf{r},\mathbf{p}) = 0,$$

by multiplication with $(a^i_k - r_k)^{n(i)-l}$ the assertion follows, as the second term is zero by the deduction assumption.

Hence, all distributions $(\hat{Q}^{\,i}_{l})^{\alpha\beta}(\mathbf{r},\mathbf{p})$ have their support at most for $r_k = a^i_k$. The $\hat{Q}^{\,\alpha\beta}$ are linear combinations of the $(\hat{Q}^{\,i}_{l})^{\alpha\beta}$ and have their support, in r_k, only in the points $a^1_k,\ldots,a^m_k$. For another index l, choose another basis in E such that A_l is in the Jordan normal form, and proceed as above. ■

One can even show that B consists only of the zero vector:

Let $B_0 := \{\mathbf{r}_i \in B / \mathbf{r}_i \neq 0\}$. Under rotations, by (2.5a),

$$\hat{Q}^{\,\alpha\beta}(\mathbf{r},\mathbf{p}) = D^{s(\alpha)}(R)_{\alpha\gamma}\, D^{s(\beta)}(R)_{\beta\delta}\, D(R^{-1})\, \hat{Q}^{\,\gamma\delta}(R^{-1}\mathbf{r},R^{-1}\mathbf{p}). \tag{2.14}$$

Proceed indirectly and assume that for some $\mathbf{p}_0$ there is a distribution $\hat{Q}^{\,\alpha\beta}(\mathbf{r},\mathbf{p}_0)$ which has, in $\mathbf{r}$, support in B_0. Choose $\mathbf{r}_0 \in B_0$ so that the left-hand side does not vanish. Then the right-hand side of (2.14) is nonzero for $\mathbf{p} = \mathbf{p}_0$, $\mathbf{r} = \mathbf{r}_0$, and all rotations R. On the other hand, if R varies, $R^{-1}\mathbf{r}_0$ varies over the whole surface of a ball in $\mathbb{R}^3$ with radius $|\mathbf{r}_0| \neq 0$. Therefore, there exists a rotation R_0 which rotates the finite number of points in B_0 in such a way that no rotated point lies again in B_0: $R_0^{-1}B_0 \cap B_0 = \emptyset$. For this rotation R_0 however, the right-hand side is zero. This contradiction proves that any distribution $\hat{Q}^{\,\alpha\beta}(\mathbf{r},\mathbf{p})$ has support at most for $\mathbf{r} = 0$ which results in

Theorem 2.4: Let E be a finite-dimensional, G-invariant space of S-matrix symmetries of the form (1.1). Then, the kernel $Q^{\,\alpha\beta}$ of any $Q \in E$ fulfills (i) $Q^{\,\alpha\beta} = 0$ for $m_\alpha \neq m_\beta$, (ii) $Q^{\,\alpha\beta}(\mathbf{p},\mathbf{q}) = 0$ for $\mathbf{p} \neq \mathbf{q}$.

3. POLYNOMIAL DEPENDENCE IN MOMENTA

By Theorem 2.4, $Q^{\,\alpha\beta}$ has the representation

$$Q^{\,\alpha\beta}(\mathbf{p},\mathbf{q}) = q^{\alpha\beta}_{\nu}(\mathbf{p})\, \delta^{(\nu)}(\mathbf{p} - \mathbf{q}), \tag{3.1}$$

where the multi-index $\nu = (\nu_1,\nu_2,\nu_3)$ in the δ function denotes ν_k derivatives with respect to $p_k - q_k$. The coefficient $q^{\alpha\beta}_{\nu}(\mathbf{p})$ can still be an arbitrary distribution in $\mathbf{p}$. It will be shown in this section that $q^{\alpha\beta}_{\nu}(\mathbf{p})$ is, in fact, a polynomial in $\mathbf{p}$.

Consider the boost subgroup of G. Combining (3.1) and (2.4a) gives

$$D(\mathbf{v})\, q^{\alpha\beta}_{\nu}(\mathbf{p}) = q^{\alpha\beta}_{\nu}(\mathbf{p} - m\mathbf{v}), \tag{3.2}$$

and Fourier transformation in the variable $\mathbf{p}$ results in

$$D(\mathbf{v})\, \tilde{q}^{\,\alpha\beta}_{\nu}(\mathbf{x}) = e^{im\mathbf{v}\mathbf{x}}\, \tilde{q}^{\,\alpha\beta}_{\nu}(\mathbf{x}). \tag{3.3}$$

On the other hand, the boost subgroup is isomorphic to the translation subgroup. Hence, Lemmas 2.2 and 2.3 are applicable to (3.3), which shows that the distributions $\tilde{q}^{\,\alpha\beta}_{\nu}(\mathbf{x})$ have support, in $\mathbf{x}$, in a set C that is independent of ν,α,β and consists of a finite number of vectors only.

As in Sec. 2, it can be shown that C consists of the zero vector only: By (3.1) and (2.14),

$$D(R)\, q^{\alpha\beta}_{\nu}(\mathbf{p})\, \delta^{(\nu)}(\mathbf{p} - \mathbf{q}) = (D^{s(\alpha)}(R))_{\alpha\gamma}(D^{s(\beta)}(R))_{\beta\delta}\, q^{\gamma\delta}_{\nu}(R^{-1}\mathbf{p})\, \delta^{(\nu)}(R^{-1}(\mathbf{p} - \mathbf{q})). \tag{3.4}$$

Operators of differentiation transform under rotations as vectors so that

$$\delta^{(\nu)}(R^{-1}(\mathbf{p} - \mathbf{q})) = (D(R))^{\nu}_{\mu}\, \delta^{(\mu)}(\mathbf{p} - \mathbf{q}), \tag{3.5}$$

where D is some tensor product of $D^{\,1}$-representations of the rotation group. Insert (3.5) into (3.4), compare terms with the same number of derivatives, and perform a Fourier transformation in $\mathbf{p}$,

$$\tilde{q}^{\,\alpha\beta}_{\mu}(\mathbf{x}) = D(R^{-1})(D^{s(\alpha)}(R))_{\alpha\gamma} \times (D^{s(\beta)}(R))_{\beta\delta}(D(R))^{\nu}_{\mu}\, \tilde{q}^{\,\gamma\delta}_{\nu}(R^{-1}\mathbf{x}), \tag{3.6}$$

so that the summation in ν is now over terms with $\Sigma\, \nu_k = \Sigma\, \mu_k$ only. From now on, one can proceed as in (2.14). Let C_0 consists of the nonzero vectors of C, and assume indirectly that there is a distribution $\tilde{q}^{\,\alpha\beta}_{\mu}(\mathbf{x})$ with support in some vector $\mathbf{x}_0 \in C_0$. Again, one can find a rotation R_0 such that $R_0^{-1}C_0$ and C_0 have an empty intersection so that, for $\mathbf{x} = \mathbf{x}_0$ and $R = R_0$, (3.6) gives a contradiction. Hence, $\tilde{q}^{\,\alpha\beta}_{\nu}$ has support only at zero, and thus $q^{\alpha\beta}_{\nu}(\mathbf{p})$ is a polynomial in $\mathbf{p}$. Together with Theorem 2.4, this proves

Theorem 3.1: Assume that the space of S-matrix symmetries E decomposes into finite-dimensional subspaces E_i invariant under the Galilei group G. Then, the kernel $Q^{\,\alpha\beta} = 0$ of any $Q \in E$ fulfills (i) $Q^{\,\alpha\beta} = 0$ for $m_\alpha \neq m_\beta$, (ii) $Q^{\,\alpha\beta}(\mathbf{p},\mathbf{q}) = q^{\alpha\beta}\delta(\mathbf{p} - \mathbf{q})$, where $q^{\alpha\beta}$ is a polynomial in $\mathbf{p}$ and p_k-derivatives $k,1,2,3$.

Thus, any $Q \in E$ is, in fact, local so that the results of Ref. 1 for local S-matrix symmetries in theories with nontrivial scattering (statement 2 of the Introduction) can be used:

Theorem 3.2: Assume a theory with scalar asymptotic particles which can be so ordered that the elastic two-particle scattering amplitude, for any two consecutive particles, does not vanish in some open neighborhood of the sets of momenta allowed by energy and momentum conservation. Then, under the assumptions of Theorem 3.1, $Q \in E$ has the form (1.3).

ACKNOWLEDGMENTS

The author wishes to thank Klaus Baumann and Helmut Reeh for their constant encouragement.

[1] W. D. Garber, "Local S-matrix Symmetries in Galilean Field Theories," J. Math. Phys. (to be published).
[2] V. Bargmann, Ann. Math. **59**, 1–46 (1954).
[3] R. Jost, *The general theory of quantized fields* (Amer. Math. Soc., Providence, RI, 1965).

A Lagrangian of Bargmann–Wigner equations for massive particles of spin 2

M. Lorente and M. A. Rodriguez
Departamento de Métodos Matemáticos de la Física, Universidad Complutense, Madrid-3, Spain

(Received 1 March 1983; accepted for publication 24 June 1983)

Following the ideas of Guralnik and Kibble and those of Larsen and Repko, we introduce a general method to calculate the first-order Lagrangian of Bargmann–Wigner equations (BWE) of arbitrary spin, and make explicit calculations in case of spin 2. Finally, some considerations on the motivation of this method and on the invariance of the Lagrangian under the symmetric group and the general Lorentz group are discussed.

PACS numbers: 11.10.Qr

1. INTRODUCTION

It is well known that the solutions of the BWE are fields that transform under the symmetric group as totally symmetric multispinors of definite mass and spin.[1]

In order to construct a first-order Lagrangian of the BWE for spin greater than one it is necessary to use auxiliary fields of mixed symmetry that transform as irreducible representations of the symmetric group. These multispinor fields of mixed symmetry also have the properties of definite mass and spin.[2]

Guralnik and Kibble[3] were able to write down a Lagrangian for massive particles of spin $\frac{3}{2}$ with totally symmetric spinor of rank 3 and auxiliary fields, without any extra conditions. This Lagrangian leads to the BWE for the totally symmetric spinors and to the vanishing of the auxiliary fields of mixed symmetry.

Larsen and Repko[4] have generalized the ideas of Guralnik and Kibble. They found the expression for the operator connecting two arbitrary fields which transform as two irreducible representations of the symmetric group S_n. But they did not give a general method to obtain the coefficients in the Lagrangian leading to the correct Euler–Lagrange equations, the solution of which are the BWE of definite mass and spin. They only succeeded in the trivial case of a Lagrangian for particles of spin 0 using multispinors of rank 4.

We present now a general method for finding the Lagrangian of BWE which can be applied to any spin and we have calculated it for particles of spin 2. Our ideas are based in the work of Larsen and Repko, combined with that of Guralnik and Kibble. We refer those papers to the reader, where both methods are given in detail.

2. COUPLING BETWEEN MULTISPINORS OF DEFINITE SYMMETRY

Let F and G be two multispinors of rank n, that transform as two irreducible representations of the symmetric group S_n. The kinetic terms in the Lagrangian are of the form $\overline{F}\hat{O}G$, where $\hat{O}$ is an operator matrix of the required dimension to connect the representations F and G, and its elements are linear combinations of Dirac operators,

$$(\gamma p)_k = 1\times 1\times\cdots\times\gamma p)\times\cdots\times 1$$

with (γp) in the k-position.

If F belongs to the (λ) representation and G to the (μ) representation of S_n, the js-components of $\hat{O}$ are given by

$$[\hat{O}(\lambda,\mu)]_{js} = \sum_{k=1}^{n-1} \beta_k \,\langle \lambda j,\ (n-1,1)k \mid \mu s\rangle,$$

where β_k are the operators

$$\beta_k = \frac{1}{\sqrt{n(n-1)}\sqrt{k(k+1)}}\left\{\sum_{i=1}^{k}(\gamma p)_i - k(\gamma p)_{k+1}\right\},$$

$(n-1,1)$ is the representation of S_n corresponding to the same partition and $\langle \lambda j(n-1,1)k \mid \mu s\rangle$ is the Clebsh–Gordan coefficient connecting the (λ) and the $(n-1,1)$ representation to the (μ) representation.

If $(\lambda)=(\mu)$ then there is also another operator α connecting the F-multispinor with itself, namely

$$\alpha = \frac{1}{n}\sum_{k=1}^{n}(\gamma p)_k.$$

As Larsen and Repko pointed out, the symmetry of $[\hat{O}G]$ under a permutation of indices is the same as that of $\overline{F}$ under the same permutation; therefore the Lagrangian is invariant under the symmetric group S_n, and $\overline{F}$ may be varied as if all of its components were independent.

In the case of S_4 we write down the components of the fields and the coupling among them.

The partition of 4 are

$$(4),\quad (3,1),\quad (2^2),\quad (2,1^2),\quad (1^4),$$

or in short form,

$$(4),\quad (3),\quad (2),\quad (\bar{3}),\quad (\bar{4}).$$

The fields, using the notation of Larsen and Repko, are

(4): $\psi_{(\alpha\beta\gamma\delta)}$,

$$(3)\colon \chi_{\alpha\beta\gamma\delta} = \begin{pmatrix} \chi'_{(\alpha\beta\gamma)\delta} \\ \chi''_{(\alpha\beta)[\gamma\delta]} \\ \chi'''_{[\alpha\beta](\gamma\delta)} \end{pmatrix},$$

$$(2)\colon \phi_{\alpha\beta\gamma\delta} = \begin{pmatrix} \phi'_{(\alpha\beta)(\gamma\delta)} \\ \phi''_{[\alpha\beta][\gamma\delta]} \end{pmatrix},$$

$$(\bar{3})\colon \zeta_{\alpha\beta\gamma\delta} = \begin{pmatrix} \zeta'_{(\alpha\beta)[\gamma\delta]} \\ \zeta''_{[\alpha\beta](\gamma\delta)} \\ \zeta'''_{[\alpha\beta\gamma]\delta} \end{pmatrix},$$

$(\bar{4})$: $\Omega_{[\alpha\beta\gamma\delta]}$.

 0022-2488/83/122823-05$02.50

For the couplings we have

$$\alpha = \tfrac{1}{4}\{(\gamma p)_1 + (\gamma p)_2 + (\gamma p)_3 + (\gamma p)_4\},$$
$$\beta_1 = \tfrac{1}{12}\{(\gamma p)_1 + (\gamma p)_2 + (\gamma p)_3 - 3(\gamma p)_4\},$$
$$\beta_2 = (1/6\sqrt{2})\{(\gamma p)_1 + (\gamma p)_2 - 2(\gamma p)_3\},$$
$$\beta_3 = (1/2\sqrt{6})\{(\gamma p)_1 - (\gamma p)_2\}.$$

The coupling among the fields are given (up to an arbitrary factor) by

$$\hat{O}(4,3) = (\beta_1 \beta_2 \beta_3),$$

$$\hat{O}(3,3) = \begin{pmatrix} -2\beta_1 & \beta_2 & \beta_3 \\ \beta_2 & \beta_1 - \sqrt{2}\beta_2 & \sqrt{2}\beta_3 \\ \beta_3 & \sqrt{2}\beta_1 & \beta_1 + \sqrt{2}\beta_2 \end{pmatrix},$$

$$\hat{O}(3,2) = \begin{pmatrix} -\beta_2 & -\beta_3 \\ -\beta_1 - (1/\sqrt{2})\beta_2 & (1/\sqrt{2})\beta_3 \\ (1/\sqrt{2})\beta_3 & -\beta_1 + (1/\sqrt{2})\beta_2 \end{pmatrix},$$

$$\hat{O}(3,\bar{3}) = \begin{pmatrix} \beta_2 & \beta_3 & 0 \\ -\beta_1 & 0 & \beta_3 \\ 0 & -\beta_1 & -\beta_2 \end{pmatrix},$$

$$\hat{O}(2,\bar{3}) = \begin{pmatrix} \beta_1 - (1/\sqrt{2})\beta_2 & (1/\sqrt{2})\beta_3 & \beta_3 \\ (1/\sqrt{2})\beta_3 & \beta_1 + (1/\sqrt{2})\beta_2 & -\beta_2 \end{pmatrix},$$

$$\hat{O}(\bar{3},\bar{3}) = \begin{pmatrix} \beta_1 + \sqrt{2}\beta_2 & -\sqrt{2}\beta_3 & \beta_3 \\ -\sqrt{2}\beta_3 & \beta_1 - \sqrt{2}\beta_2 & -\beta_2 \\ \beta_3 & -\beta_2 & -2\beta_1 \end{pmatrix},$$

$$\hat{O}(\bar{3},\bar{4}) = \begin{pmatrix} \beta_3 \\ -\beta_2 \\ \beta_1 \end{pmatrix}.$$

The rest of the operators are obtained by interchanging rows by columns.

The Lagrangian for massive particles of spin 2 will be given by the kinetic terms and the mass terms with arbitrary constants, namely,

$$\begin{aligned}\mathscr{L} = {} & \bar{\psi} a \psi + a_1 \bar{\psi} \hat{O}(4,3)\chi + a_2 \bar{\chi} \hat{O}(3,4)\psi + a_3 \bar{\chi} \alpha \chi + a_3' \bar{\chi} \hat{O}(3,3)\chi \\ & + a_4 \bar{\chi} \hat{O}(3,2)\phi + a_5 \bar{\chi} \hat{O}(3,\bar{3})\zeta + a_6 \bar{\phi} \hat{O}(2,3)\chi + a_7 \bar{\phi} \alpha \phi + a_8 \bar{\phi} \hat{O}(2,\bar{3})\zeta \\ & + a_9 \bar{\zeta} \hat{O}(\bar{3},3)\chi + a_{10} \bar{\zeta} \hat{O}(\bar{3},2)\phi + a_{11} \bar{\zeta} \alpha \zeta + a_{11}' \bar{\zeta} \hat{O}(\bar{3},\bar{3})\zeta \\ & + a_{12} \bar{\zeta} \hat{O}(\bar{3},\bar{4})\Omega + a_{13} \bar{\Omega} \hat{O}(\bar{4},\bar{3})\zeta - b_1 m \bar{\psi}\psi - b_2 m \bar{\chi}\chi - b_3 m \bar{\phi}\phi \\ & - b_4 m \bar{\zeta}\zeta - b_5 m \bar{\Omega}\Omega. \end{aligned} \tag{1}$$

After variation with respect to the independent components of the multispinors we obtain the Euler–Lagrange equations, which can be written in the matrix form

$$\begin{pmatrix} \alpha & c_1\hat{O}(4,3) & 0 & 0 & 0 \\ c_2\hat{O}(3,4) & c_3\alpha + c_3'\hat{O}(3,3) & c_4\hat{O}(3,2) & c_5\hat{O}(3,\bar{3}) & 0 \\ 0 & c_6\hat{O}(2,3) & c_7\alpha & c_8\hat{O}(2,\bar{3}) & 0 \\ 0 & c_9\hat{O}(\bar{3},3) & c_{10}\hat{O}(\bar{3},2) & c_{11}\alpha + c_{11}'\hat{O}(\bar{3},\bar{3}) & c_{12}\hat{O}(\bar{3},\bar{4}) \\ 0 & 0 & 0 & c_{13}\hat{O}(\bar{4},\bar{3}) & 0 \end{pmatrix} \begin{pmatrix} \psi \\ \chi \\ \phi \\ \zeta \\ \Omega \end{pmatrix} = m \begin{pmatrix} \psi \\ \chi \\ \phi \\ \zeta \\ \Omega \end{pmatrix}, \tag{2}$$

where the coefficients b_i have been absorbed in the c_i, for instance,

$$c_1 = a_1/b_1, \quad c_2 = a_2/b_2, \quad c_3 = a_3/b_2, \quad c_3' = a_3'/b_2, \text{ etc.}$$

We can express the matrix equation (2) in the symbolic form

$$AX = mX.$$

3. THE PROJECTION OPERATORS

The method we introduce now is to find out the arbitrary constants a_i, b_i, c_i necessary to deduce the BWE for the totally symmetric field ψ and to annihilate the auxiliary fields χ,ϕ,ζ,Ω. This method is based in the alternative expression of the BWE written by Guralnik and Kibble.

With the help of the projection operators

$$\Lambda_{\pm}(p) = \tfrac{1}{2}\left(1 \pm \frac{\gamma p}{p}\right),$$

where $p = (p^2)^{1/2}$ $(p^2 \neq 0)$, they introduce the mutually orthogonal projection operators

$$P_{\pm}(p) = \Lambda_{\pm} \times \Lambda_{\pm} \times \cdots \times \Lambda_{\pm},$$
$$P_0(p) = 1 - P_+(p) - P_-(p),$$

where each operator $\Lambda_{\pm}(p)$ acts on different tensor indices.

The BWE are equivalent to

$$(p-m)P_+(p)\psi=0, \tag{3a}$$

$$-m\,P_0(p)\psi=0, \tag{3b}$$

$$(-p-m)P_-(p)\psi=0. \tag{3c}$$

In fact, since $(\gamma p)_k\Lambda_{\pm}=\pm p\Lambda_{\pm}$, we have

$$(\gamma p)_k\,P_{\pm}(p)=\pm p\,P_{\pm}(p),$$

and from (3b) $(\gamma p)_k\psi=p(P_+(p)-P_-(p))\psi$.

Therefore $(\gamma p)_1\psi=(\gamma p)_2\psi=\cdots=(\gamma p)_n\psi$. From (3a) and (3b) we get

$$(p-m)[\Lambda_+(p)]_k\psi=0,$$

$$(-p-m)[\Lambda_-(p)]_k\psi=0.$$

Adding both equations, finally we obtain

$$p[\Lambda_+(p)-\Lambda_-(p)]_k=m[\Lambda_+(p)+\Lambda_-(p)]_k$$

or

$$(\gamma p)_k\psi=m\psi.$$

Equations (3a) and (3c) for the totally symmetric tensor fields and the auxiliary fields are easily obtained applying the projector $P_+(p)$ and $P_-(p)$ to the five equations (2):

$$\left.\begin{aligned}(p-m)\,P_+(p)\psi=0\\(-p-m)\,P_-(p)\psi=0\end{aligned}\right\}, \tag{4}$$

$$\left.\begin{aligned}(c_3p-m)\,P_+(p)\chi=0\\(-c_3p-m)\,P_+(p)\chi=0\end{aligned}\right\}, \tag{5}$$

$$\left.\begin{aligned}(c_7p-m)\,P_+(p)\phi=0\\(-c_7p-m)\,P_-(p)\phi=0\end{aligned}\right\}, \tag{6}$$

$$\left.\begin{aligned}(c_{11}p-m)\,P_+(p)\zeta=0\\(-c_{11}p-m)\,P_-(p)\zeta=0\end{aligned}\right\}, \tag{7}$$

$$\left.\begin{aligned}-m\,P_+(p)\Omega=0\\-m\,P_-(p)\Omega=0\end{aligned}\right\}. \tag{8}$$

The operator $P_0(p)$ can be decomposed in the following way:

$$\begin{aligned}P_0(p)=&\Lambda_+\Lambda_+\Lambda_+\Lambda_-+\Lambda_+\Lambda_+\Lambda_-\Lambda_+\\&+\Lambda_+\Lambda_-\Lambda_+\Lambda_++\Lambda_-\Lambda_+\Lambda_+\Lambda_+\\&+\Lambda_+\Lambda_+\Lambda_-\Lambda_-+\Lambda_+\Lambda_-\Lambda_+\Lambda_-\\&+\Lambda_+\Lambda_-\Lambda_-\Lambda_++\Lambda_-\Lambda_-\Lambda_+\Lambda_+\\&+\Lambda_-\Lambda_+\Lambda_-\Lambda_++\Lambda_-\Lambda_+\Lambda_+\Lambda_-\\&+\Lambda_+\Lambda_-\Lambda_-\Lambda_-+\Lambda_-\Lambda_+\Lambda_-\Lambda_-\\&+\Lambda_-\Lambda_-\Lambda_+\Lambda_-+\Lambda_-\Lambda_-\Lambda_-\Lambda_+,\end{aligned}$$

where the direct product among the Λ's is understood. (This expression comes from the decomposition of the identity operator in terms of the mutually orthogonal projection operators.)

In order to have $P_0(p)\psi=0$, we apply each of the terms in the last decomposition to the first equation (2) and require that the left side of the resulting expression should vanish. But this is impossible, since, for instance,

$$\Lambda_+\Lambda_+\Lambda_+\Lambda_-\alpha\psi=\tfrac{1}{2}p\Lambda_+\Lambda_+\Lambda_+\Lambda_-\psi\neq0.$$

We can avoid this difficulty if we iterate the matrix equation (2), namely

$$A^2X=m^2X, \tag{9}$$

and then apply each of the projectors

$$\Lambda_+\Lambda_+\Lambda_+\Lambda_-,\ \Lambda_+\Lambda_+\Lambda_-\Lambda_+,$$

$$\Lambda_+\Lambda_-\Lambda_+\Lambda_+,\ \Lambda-\Lambda_+\Lambda_+\Lambda_+$$

and

$$\Lambda_+\Lambda_-\Lambda_-\Lambda_-,\ \Lambda_-\Lambda_+\Lambda_-\Lambda_-,$$

$$\Lambda_-\Lambda_-\Lambda_+\Lambda_-,\ \Lambda_-\Lambda_-\Lambda_-\Lambda_+$$

to the five equations (9). In order that the right side should vanish, we require that the coefficients of the projection operators applied to the tensor fields on the left should vanish, giving some conditions on the coefficients c's.

Next, we iterate again the matrix equation, i.e.,

$$A^3X=m^3X, \tag{10}$$

apply to each one of the five equations the projectors

$$\Lambda_+\Lambda_+\Lambda_-\Lambda_-,\quad \Lambda_+\Lambda_-\Lambda_+\Lambda_-,\quad \Lambda_+\Lambda_-\Lambda_-\Lambda_+,$$

$$\Lambda_-\Lambda_+\Lambda_+\Lambda_-,\quad \Lambda_-\Lambda_+\Lambda_-\Lambda_+,\quad \Lambda_-\Lambda_-\Lambda_+\Lambda_+,$$

and require that all the coefficients of the projectors applied to the tensor fields on the left vanish, giving new relations among the c's.

The conditions that the c's must satisfy in order that

$$P_0(p)\psi=P_0(p)\chi=P_0(p)\phi=P_0(p)\zeta=P_0(p)\Omega=0$$

are written down below. Before each equation, the symbol, $(A^2)_{ij}$ or $(A^3)_{ij}$ is expressed denoting which matrix element has been used after applying each of the projector operators independently to Eqs. (9) and (10).

$$(A^2)_{11}{:}c_1c_2+1=0, \tag{11}$$

$$(A^2)_{12}{:}c_1(1-2c_3'+c_3)=0, \tag{12}$$

$$(A^2)_{21}{:}c_2(1-2c_3'+c_3)=0, \tag{13}$$

$$(A^2)_{22}{:}(c_3+c_3')^2+c_4c_6+c_5c_9=0, \tag{14a}$$

$$c_1c_2+(2c_3'-c_3)^2=0, \tag{14b}$$

$$(A^2)_{23}{:}c_4(c_3+c_3'+c_7)+c_5c_{10}=0, \tag{15}$$

$$(A^2)_{24}{:}c_4c_8+c_5(c_3+c_3'+c_{11}+c_{11}')=0, \tag{16}$$

$$(A^2)_{32}{:}c_6(c_3+c_3'+c_7)+c_8c_9=0, \tag{17}$$

$$(A^2)_{33}{:}c_4c_6+c_8c_{10}+c_7^2=0, \tag{18}$$

$$(A^2)_{34}{:}c_8(c_{11}+c_{11}'+c_7)+c_5c_6=0, \tag{19}$$

$$(A^2)_{42}{:}c_6c_{10}+c_9(c_3+c_3'+c_{11}+c_{11}')=0, \tag{20}$$

$$(A^2)_{43}{:}c_{10}(c_{11}+c_{11}'+c_7)+c_4c_9=0, \tag{21}$$

$$(A^2)_{44}{:}(c_{11}+c_{11}')^2+c_8c_{10}+c_5c_9=0, \tag{22a}$$

$$c_{12}c_{13}+(2c_{11}'-c_{11})^2=0, \tag{22b}$$

$$(A^2)_{54}{:}c_{13}(1-2c_{11}'+c_{11})=0, \tag{23a}$$

$$(A^2)_{55}{:}c_{12}c_{13}=0. \tag{23b}$$

The last two equations are not necessary since

$$\Lambda_+\Lambda_+\Lambda_+\Lambda_-\Omega=\Lambda_+\Lambda_+\Lambda_-\Lambda_+\Omega=\Lambda_+\Lambda_-\Lambda_+\Lambda_+\Omega$$
$$=\Lambda_-\Lambda_+\Lambda_+\Lambda_+\Omega=0$$

because of the antisymmetry of the indices of Ω.

The same is true for (22b) since this constraint comes from the coefficient of $\Lambda_{+++-}\zeta'''_{[\alpha\beta\gamma]\delta}$; but this is zero because of the antisymmetry of the indices $[\alpha\,\beta\gamma]$.

$$(A^{3})_{12}: c_1(c_1c_2 + 2c_4c_6) = 0, \tag{24}$$

$$(A^{3})_{21}: c_2(c_1c_2 + 2c_4c_6) = 0, \tag{25}$$

$$(A^{3})_{22}: c_3'^3 + \tfrac{2}{3}c_3'c_5c_9 - \tfrac{1}{3}c_5c_9c_{11}' = 0, \tag{26}$$

$$(A^{3})_{23}: c_4(c_1c_2 + 2c_4c_6) = 0, \tag{27}$$

$$(A^{3})_{24}: c_5(c_3'^2 + c_{11}'^2 + \tfrac{1}{3}c_5c_9 - c_3'c_{11}') = 0, \tag{28}$$

$$(A^{3})_{32}: c_6(c_1c_2 + 2c_4c_6) = 0, \tag{29}$$

$$(A^{3})_{34}: c_8(c_{12}c_{13} + 2c_8c_{10}) = 0, \tag{30}$$

$$(A^{3})_{42}: c_9(c_3'^2 + c_{11}'^2 - c_3'c_{11}' + \tfrac{1}{3}c_5c_9) = 0, \tag{31}$$

$$(A^{3})_{43}: c_{10}(c_{12}c_{13} + 2c_8c_{10}) = 0, \tag{32}$$

$$(A^{3})_{44}: c_{11}'^3 + \tfrac{2}{3}c_5c_9c_{11}' - \tfrac{1}{3}c_3'c_5c_9 = 0, \tag{33}$$

$$(A^{3})_{45}: c_{12}(c_{12}c_{13} + 2c_8c_{10}) = 0, \tag{34}$$

$$(A^{3})_{54}: c_{13}(c_{12}c_{13} + 2a_8c_{10}) = 0. \tag{35}$$

From (11) we obtain $c_1c_2 = -1$.

From (26) and (31) we obtain $c_3' = c_{11}'$ and $c_5c_9 = -3c_3'^2$.

From (24) we get $c_4c_6 = -\frac{1}{2}c_1c_2 = \frac{1}{2}$.

From (14) we get $c_3 = 0$, $c_3' = \frac{1}{2}$.

From (16), (20), and (22a) we get $c_8c_{10} = -\frac{3}{2}$ and $c_{11} = -2$.

From (15), (17), (19), and (21) we get $c_7 = 1$.

From (35) we have $c_{12}c_{13} = -2c_8c_{10} = 3$.

The rest of the conditions are consistent with these solutions, except (22b), (23a), and (23b) which are not necessarily zero because the projectors applied to the corresponding multispinors are zero.

Equations (14)–(35) imply that each of the tensor fields satisfies the following equations:

$$\left.\begin{aligned}(+p-m)P_+(p)\psi &= 0\\ (-p-m)P_+(p)\psi &= 0\\ P_0(p)\psi &= 0\end{aligned}\right\},$$

equivalent to $(\gamma p)_k\psi = m\psi$;

$$P_+(p)\chi = P_-(p)\chi = P_0(p)\chi = 0,$$

equivalent to $\chi = 0$;

$$\left.\begin{aligned}(p-m)P_+(p)\phi &= 0\\ P_0(p)\phi &= 0\\ (-p-m)P_-(p)\phi &= 0\end{aligned}\right\},$$

equivalent to $(\gamma p)_k\phi = m\phi$;

$$\left.\begin{aligned}(-2p-m)P_+(p)\xi &= 0\\ P_0(p)\xi &= 0\\ (2p-m)P_-(p)\xi &= 0\end{aligned}\right\},$$

equivalent to $(\gamma p)_k\xi = (m/2)\xi$,

$$P_+(p)\Omega = P_-(p)\Omega = P_0(p)\Omega = 0,$$

equivalent to $\Omega = 0$.

In order to have $\phi = 0$ we exclude in the Lagrangian those components of the tensor ϕ which have only the indices 1 or 2, since tensor components with indices 3 or 4 are annihilated by the BWE. (It means that we exclude only the tensor component $\phi^{[12][12]}$ from the 15 independent components of the tensor $\phi^{[\alpha\beta][\gamma\delta]}$ and similarly in the tensor $\phi^{(\alpha\beta)(\gamma\delta)}$.) The tensor fields ξ are also annihilated by the BWE because of the antisymmetry[5] of the three indices of $\xi^{[\alpha\beta\gamma]\dot{\delta}}$.

Finally we want the Lagrangian to be Hermitian; therefore the matrix A in (2) should be Hermitian and hence the coefficient c's are real and those coefficients symmetric with respect to the principal diagonal should be equal. This is impossible with the c's but we still have at our disposal the arbitrary coefficient b's. If we take

$$b_1 = 1, \quad b_2 = -1, \quad b_3 = -1, \quad b_4 = 1, \quad b_5 = 1$$

we have the desired result, namely,

$$a_1^2 = a_2^2 = 1, \quad a_3 = 0, \quad a_3' = -\tfrac{1}{2},$$
$$a_4^2 = \tfrac{1}{2}, \quad a_5^2 = \tfrac{3}{4}, \quad a_6^2 = \tfrac{1}{2}, \quad a_7 = -1,$$
$$a_8^2 = \tfrac{3}{2}, \quad a_9^2 = \tfrac{3}{4}, \quad a_{10}^2 = \tfrac{3}{2}, \quad a_{11} = -2,$$
$$a_{11}' = \tfrac{1}{2}, \quad a_{12}^2 = a_{13}^2 = 3.$$

Substituting these values, we get for the Lagrangian[6] of massive particle of spin 2,

$$\begin{aligned}\mathscr{L} = {} & \psi\alpha\psi + [\bar{\psi}\hat{O}(4,3)\chi + \bar{\chi}\hat{O}(3,4)\psi] - \tfrac{1}{2}\bar{\chi}\hat{O}(3,3)\chi\\ & + (1/\sqrt{2})[\bar{\chi}\hat{O}(3,2)\phi + \bar{\phi}\hat{O}(2,3)\chi] + (\sqrt{3}/2)[\bar{\chi}\hat{O}(3,\bar{3})\zeta\\ & + \bar{\zeta}\hat{O}(\bar{3},3)\chi] - \bar{\phi}\alpha\phi + (\sqrt{3}/2)\\ & \times[\bar{\phi}\hat{O}(2,\bar{3})\zeta + \bar{\zeta}\hat{O}(\bar{3},2)\phi] - 2\bar{\zeta}\alpha\zeta + \tfrac{1}{2}\bar{\zeta}\hat{O}(\bar{3},\bar{3})\zeta\\ & + \sqrt{3}[\bar{\zeta}\hat{O}(\bar{3},\bar{4})\Omega + \bar{\Omega}\hat{O}(\bar{4},\bar{3})\zeta]\\ & - m\bar{\psi}\psi + m\bar{\chi}\chi + m\bar{\phi}\phi - m\bar{\zeta}\zeta - m\bar{\Omega}\Omega.\end{aligned}$$

4. DISCUSSION

The use of the iteration of the matrix equations (9) and (10) is required in order to have necessary and sufficient conditions to obtain the BWE for a totally symmetric multispinor of rank 4. [In the case of rank 3, it is easily proved that one needs only to iterate twice the matrix equation.[7]] This suggests that we have to iterate $(n-1)$ times in order to obtain the BWE for multispinors of rank n.

The solution for the coefficients is unique, but this implies in the case of spin 2, that the fields ψ and ϕ satisfy the BWE, and the rest are zero. But if we want to eliminate also the auxiliary field ϕ, as desired, we can do it by suppressing one component field in the Lagrangian from the beginning, and then the rest of the components of ϕ will be zero at the end.

In the last case, the Lagrangian will be invariant under the symmetric group, namely, under any permutation of the spinorial indices, but it will not be invariant under the complete Lorentz group, since to each spinorial index we have to apply the generators of the Dirac representation.

[1]V. Bargmann and E. Wigner, Proc. Natl. Acad. Sci. USA **34**, 211 (1948).
[2]M. A. Rodriguez and M. Lorente, J. Math. Phys. **22**, 1283 (1981).
[3]G. S. Guralnik and T. W. B. Kibble, Phys. Rev. **139**, B712 (1965).
[4]M. L. Larsen and W. W. Repko, J. Math. Phys. **19**, 930 (1978).
[5]See Ref. 2, p. 1286.

[6]If we take $\psi = \chi = 0$ and then $a_1 = a_2 = a_3 = a_3' = a_4 = a_5 = a_6 = a_9 = b_1 = b_2 = 0$ in the Lagrangian (1), we obtain, from (11)–(35), $a_7 = a_8 = a_{10} = a_{11} = 1$, $a_{11}' = 0$, $a_{12} = a_{13} = \sqrt{2}$, $b_3 = 1$, $b_4 = b_5 = -1$ which corresponds to the Lagrangian (after redefinition of the coupling operators $\hat{O}$) found by Larsen and Repko for BWE of rank-4 multispinors representing massive particles of spin 0.

[7]This iteration leads to the same operator equations for the totally symmetric multispinors as those found by Larsen and Repko (see Ref. 4). In case of rank-3 spinors the β_k annihilate the operator $\alpha^2 - \frac{1}{4}\beta^2$, which appears in the component $(A^2)_{11}$ of the matrix equation $A^2X = m^2X$. In the case of rank-4 spinors the β_k annihilate the operator $\alpha(\alpha^2 - \beta_1^2 - \beta_2^2 - \beta_3^2)$ which appears in the component $(A^3)_{11}$ of the matrix equation $A^3X = m^3X$.

Conformal QED

B. Binegar, C. Fronsdal, and W. Heidenreich
Department of Physics, University of California, Los Angeles, California 90024

(Received 11 March 1983; accepted for publication 29 June 1983)

A conformally invariant quantum electrodynamics is constructed. The setting is realistic space-time (rather than Euclidean), and a complete Gupta–Bleuler quantization scheme is carried out. Conformal invariance of the quantum field theory (as opposed to either classical field theory or to a theory defined by its Feynman rules) requires a richer Gupta–Bleuler structure than has been considered previously. Yet the essential features of this structure are preserved. The requirement that the wave equation be of second order fixes a unique action that already contains the gauge-fixing terms that are required in any complete quantum field theory. The "Lorentz condition" turns out to be the transversality condition $y_\alpha a_\alpha(y) = 0$ (in the manifestly covariant six-dimensional notation); this condition has to be treated in the same way as the Lorentz condition $\partial_\mu A_\mu(x) = 0$ (four-dimensional notation), as a boundary condition on the physical states.

PACS numbers: 12.20.Ds

I. INTRODUCTION

This paper deals with the problem of setting up a formulation of quantum electrodynamics that is manifestly invariant under the transformations of the full conformal group. To put our work in perspective, a very short review of the history of the subject may be useful.

The invariance of Maxwell's equations has been known since the early papers of Cunningham and Bateman.[1] The six-dimensional formalism that makes the invariance manifest was invented by Dirac[2] and developed by Mack and Salam[3] and many others.[4] As far as classical electrodynamics is concerned, we have little to add. Conformal invariance in quantum field theory was exploited by Baker and Johnson,[5] by Mack and Symanzik,[6] by Adler,[7] and by many others,[8] but all these authors were primarily concerned with the Green's functions of the theory, or with the Feynman rules. Here we do have something new, namely, a complete Gupta–Bleuler quantization procedure that is manifestly conformally invariant. Much of the work cited was carried out on the Euclidean version of electrodynamics; this may be convenient for the study of Green's functions, but it would completely distort the salient features of the Gupta–Bleuler quantization scheme. This paper is concerned exclusively with real, Minkowski space-time electrodynamics.

Motivation for the present investigation comes from suggestions that have repeatedly been put forward, concerning conformally covariant theories of gravity and supergravity. While Einstein's action may be dominant at low energies, other terms in the action may be responsible for high energy behavior. The conformally invariant action of Weyl[9] is particularly tantalizing. Indeed, theories that contain both the Weyl action and the Einstein–Hilbert action have sometimes been called "manifestly renormalizable." The idea is that the conformally invariant Weyl action dominates at high energies, providing an ultraviolet cutoff that makes the theory renormalizable, while the conformal symmetry-breaking Einstein action dominates at low energies and accounts for observed phenomena.[10] It is believed[11] that such theories contain ghosts, because of the appearance of higher order derivatives. So it may be, but it seems that the question has never been fully investigated. The work presented here lays the groundwork for a thorough investigation of conformally invariant quantum gravity.

The spectacular development of modern gauge theories is another source of inspiration. Some of these theories are conformally invariant in the classical formulation.[12] The Gupta–Bleuler structure has been extended by the introduction of Faddeev–Popov ghosts[13] and play a major role in the BRS quantization scheme[14] for field theories of the Yang–Mills type, including gravity and supergravity.[15] The Gupta–Bleuler structure appears to be very fundamental and further studies seem to be called for, particularly in the context of nonconventional gauge theories[16] and in conformally invariant field theories.

The connection between renormalizability (both ultraviolet and infrared) and dilatation invariance is easily understood in terms of power counting. There is a strong suspicion, however, that full conformal invariance may play a deeper role in the renormalization program. This is especially evident in the case of the problems of the infrared and the phenomenon of anomalous dimensions (see Todorov *et al.*, Ref. 8). But most previous work on conformal invariance has been concerned with the properties of Green's functions, and not with quantum fields. As we shall show here, the full incorporation of conformal invariance in a canonical theory of quantum fields requires the introduction of additional gauge fields. The effect that these new fields may have on the renormalization program has not yet been examined. We suggest that such an investigation may be fruitful, not only in QED, but perhaps especially in nonabelian gauge theories.

The idea that massless particles may be composites[17] is gaining ground. In one of these modern versions of the "neutrino theory of light," the constituents are described by a new type of field theory.[18] The physical state space of these fields is dramatically reduced in comparison with conventional fields, and for this reason an equivalent three-dimensional formulation may be preferable. This leads to three-dimensional, conformally invariant field theories,[19] which provides another area for application of the ideas of this paper.

 0022-2488/83/122828-19$02.50

The main results of this paper are included in the following summary. Section II is a brief outline of the familiar Gupta–Bleuler quantization scheme for quantum electrodynamics. The purpose of including this well-known material is to establish a terminology that will be applied later in another context; in particular, the notion of Gupta–Bleuler triplets is presented in such a manner as will facilitate generalizations.

Section III initiates the study of conformal QED with an account of the relevant representations of SO(4,2). These are representations with positive energy, and they are therefore characterized (not always uniquely) by a minimal weight. The phenomenon that characterizes gauge theories, and leads to indefinite-metric quantization, is the appearance in a field theory of representations that "leak," that is, indecomposable representations. Among representations with minimal weight, opportunities for leakage occur only in very exceptional cases, and a fairly complete classification of the possibilities is carried out. A very simple derivation of sufficient (and, as it turns out, necessary) conditions for the unitarity of minimal weight representations is also given.

Section IV introduces Dirac's manifestly covariant formalism of symmetric tensor fields on the 6-cone. Conditions for the existence of conformally invariant wave equations (of any order) are given. Invariant subspaces of fields are defined by means of subsidiary conditions: transversality, divergencelessness, tracelessness. These subspaces usually carry irreducible (projective) representations of the conformal group; however, in exceptional cases they contain invariant subspaces of "gauge fields." There are two main types, "current type gauge fields" and "gradient type gauge fields." These gauge subspaces are not invariantly complemented and signal the presence of a gauge structure. An intimate correspondence is established between these exceptional gauge structures and the nondecomposable representations conjectured in Sec. III, and the existence of some of the latter is demonstrated. In physical terms, what becomes known at this stage are the "physical photons" and the "gauge photons." The "scalar photons" (the third component of the Gupta–Bleuler triplet, needed for quantization) have not yet appeared.

The "scalar photons" may be found by analyzing the propagators, and this is done in Secs. V and VI. The analysis is carried to completion in the cases that are of interest for electrodynamics. The scalar photons are described. The propagator has precisely the structure that guarantees that only physical photons propagate between conserved currents. The "Lorentz condition," that is, the condition that removes the scalar photons, turns out, surprisingly, to be the condition of transversality, $y_\alpha a_\alpha(y) = 0$, and not, as expected, the condition of divergencelessness, $\partial_\alpha a_\alpha = 0$. Consequently, $y_\alpha a_\alpha = 0$ is a condition that cannot be imposed on the quantized field. Instead, it must be enforced as a boundary condition on the physical states. The suitability of this boundary condition is confirmed by the fact that $y_\alpha a_\alpha$ is a free field.

An action principle is formulated in Sec. VII. Under the requirement of strict conformal invariance there is no second-order, gauge invariant wave equation. For the formulation of a classical field theory, where full gauge invariance of the field equations may be desirable, there is a gauge and conformal invariant wave equation of the third order. To formulate a quantum field theory, one needs to introduce gauge fixing terms in the action. The simplest choice is to drop the third-order terms; one is thus led to a unique second-order, conformally invariant wave equation. In this gauge the theory is easily quantized in the Gupta–Bleuler manner. Previous treatments of the Feynman rules and the Green's functions of conformal electrodynamics have restricted the vector potential by imposing transversality, $y\cdot a = 0$, from the outset. This is why Adler,[7] for example, must use a propagator that is not covariant. It is thus crucial to recognize that $y\cdot a = 0$ plays the role of Lorentz condition for conformal QED. When interactions are included, then one finds, in addition to the gradient-type gauge phenomena associated with the potential, that current-type gauge transformations also appear.

Section VIII contains all our main conclusions about conformal QED in the ordinary, four-dimensional notation. It has been written so as to be as self-contained as possible, and it will therefore not be summarized in this place. The main feature is the appearance of two spinless fields, and an extension of the indefinite metric space of one-particle states. Section IX discusses the breakdown of conformal symmetry by the introduction of a causal structure, and suggests future developments.

II. GUPTA–BLEULER TRIPLET IN QED

Classical electrodynamics employs a fully gauge invariant action. In QED, however, it is necessary to choose a propagator, and this implies a choice of gauge. According to the modern view, the propagator is fixed by the choice of an action that includes gauge fixing terms, as originally suggested by Fermi.[20] The simplest choice is

$$L = \int dx\, (\tfrac{1}{2}A_\mu \Box A_\mu - A\cdot J)\,. \tag{2.1}$$

The current will be treated, for the present, as fixed and external. It must be conserved, $\partial\cdot J = 0$. The field equation is $\Box A = J$.

To prepare for quantization one begins with free fields with positive energy satisfying

$$\Box A_\mu = 0\,, \tag{2.2}$$

$$\|A\|^2 \equiv -\int d^3x\, A_\mu \overleftrightarrow{i\partial_0} A_\mu < \infty\,. \tag{2.3}$$

The norm is indefinite, and this is the crux of the problem. The Poincaré group acts on the vector fields in the usual way, preserving both (2.2) and (2.3). [The restriction to positive energies will be maintained, although this will not always be emphasized, throughout this paper.]

The Gupta–Bleuler triplet $\mathscr{V}' \supset \mathscr{V} \supset \mathscr{V}_g$ will now be defined. The space $\mathscr{V}_g$ of "gauge states" or "longitudinal photons" consists of all (positive energy) solutions of (2.2), (2.3) of the form $A_\mu = \partial_\mu \Lambda$. This space is invariant for the action of the Poincaré group. The space $\mathscr{V}$ consists of all solutions of (2.2), (2.3) that satisfy the Lorentz condition

$$\partial\cdot A = 0 \text{ (Lorentz condition).} \tag{2.4}$$

Although $\mathscr{V}_g$ is an invariant subspace of $\mathscr{V}$, it is not invar-

iantly complemented. The quotient space $\mathscr{V}/\mathscr{V}_g$ of equivalence classes is the space of "physical states," or "transverse photons." Finally, $\mathscr{V}'$ is the space of all solutions of (2.2), (2.3). Although $\mathscr{V}$ is an invariant subspace of $\mathscr{V}'$, it has no invariant complement in $\mathscr{V}'$. The quotient space $\mathscr{V}'/\mathscr{V}$ is the space of "scalar states" or "scalar photons." The appearance of this triplet seems to be universal in gauge theories, and crucial for quantization. Most of the following also appears to be very general, and the terminology (gauge states, physical states, scalar states, Lorentz condition) will be applied in a wider context.

The norm (2.3) is indefinite. When it is restricted to $\mathscr{V}$ it becomes semidefinite. The radical of (2.3) in $\mathscr{V}$ (the set of all fields in $\mathscr{V}$ orthogonal to $\mathscr{V}$) is $\mathscr{V}_g$; a positive-definite norm is thus induced on $\mathscr{V}/\mathscr{V}_g$, and completion turns this space of physical states into a Hilbert space $\mathscr{H}$. The action of the Poincaré group in $\mathscr{H}$ is the unitary representation $D(0,1)\oplus D(0,-1), D(0,\lambda)$ being the UIR with zero mass and discrete helicity λ. The subspace $\mathscr{V}_g$ may also be turned into a Hilbert space, in which the action of the Poincaré group is $D(0,0)$. Finally, a representation equivalent to $D(0,0)$ is induced in $\mathscr{V}'/\mathscr{V}$. Altogether, the space of (positive energy) solutions of (2.2), (2.3) carries a representation that is indecomposable and equivalent to the triplet

$$D(0,0)\rightarrow[D(0,1)\oplus D(0,-1)]\rightarrow D(0,0)\,. \tag{2.5}$$

The arrow denotes semidirect sum. It points towards the invariant subrepresentation; that is, in the direction of the leak.

In Gupta–Bleuler quantization[21] all three kinds of states are associated with quantum excitations, and all three contribute to the quantum field operator $\hat{A}_\mu(x)$. This is defined in terms of creation and destruction operators acting in an indefinite metric Fock space:

$$\hat{A}_\mu(x)=\sum_i A_\mu^{(i)}(x)b^{(i)}+\text{h.c.} \tag{2.6}$$

The "homogeneous propagator," or "reproducing kernel," is the vacuum expectation value

$$K_{\mu\nu}(x,x')=\langle 0|\hat{A}_\mu(x)\hat{A}_\nu(x')|0\rangle\,. \tag{2.7}$$

This is regarded as the distribution kernel of an operator K in (2.3). It has a modal decomposition

$$K_{\mu\nu}(x,x')=\sum_i A_\mu^{(i)}(x)\,\overline{A_\nu^{(i)}(x')}\,. \tag{2.8}$$

The Lorentzian metric makes this operator nonpositive. Broken up into photon types, the decomposition (2.8) may be rendered symbolically by

$$K=\sum P\overline{P}+\sum S\overline{G}+\sum G\overline{S}\,. \tag{2.9}$$

The first term represents the contribution of the physical states. It is positive, but noninvariant because of the leak into gauge states. In the other terms, G stands for gauge fields and S for scalar-photon fields. These terms make K invariant but destroy positivity. It is therefore important, in order that the theory be unitary, that the S, G terms do not contribute propagating waves. This comes about because current conservation, $\partial\cdot J=0$ is equivalent to

$$(J,G)\equiv\int d^4x\,J_\mu(x)\,\partial_\mu\Lambda(x)=-\int d^4x\,\Lambda(x)\,\partial\cdot J(x)=0\,. \tag{2.10}$$

Wave propagation between a source J and a detector J' is expressed by (J,KJ'), and if both J and J' are conserved, then only the $P\overline{P}$ terms contribute. The crucial property of K is thus the absence of $S\overline{S}$ terms in (2.9).

To sum up: In Gupta–Bleuler quantization one obtains a covariant propagator by quantizing all modes, including the scalar modes. The Lorentz condition (2.4) is thus not satisfied by the quantum field operator: $\partial\cdot\hat{A}\neq 0$. Nevertheless, the scalar and gauge photons do not interact. This may be seen in another way, by noting that the field equation $\Box\hat{A}=J$ and current conservation $\partial\cdot J=0$ give

$$\Box\partial\cdot\hat{A}=O\,, \tag{2.11}$$

so that the scalar field $\partial\cdot\hat{A}$ is a free field.

The essential aspect of the Gupta–Bleuler triplet is displayed by (2.5). Physical state space occupies a central position between two similar unphysical objects. Under Poincaré transformations, states with helicity ± 1 leak to gauge states. To obtain a covariant propagator, one needs a dual set of states, the scalar photons, that leak to the physical states. The expansion (2.8) shows that the representations associated with gauge and scalar states must be equivalent to each other. This is why the two outer elements of (2.5) are the same.

III. MINIMAL WEIGHT REPRESENTATIONS OF SO(4,2)

The UIR's $D(0,1)$ and $D(0,-1)$ of the Poincaré group that are associated with the physical photons of QED have unique extensions[22] to unitary irreducible representations of SO(4,2), the double covering of the conformal group. In the notation to be introduced below, these two UIR's of SO(4,2) will appear as $D(2,1,0)$ and $D(2,0,1)$. They have the remarkable property of remaining irreducible when restricted to the Poincaré subgroup,[23] and it is evident that they must play an important role in the description of the physical states of conformal QED. The representation $D(0,0)$ of the Poincaré group also has a unique extension to SO(4,2), but the triplet (2.5) does not. Therefore, one does not yet know what representations of SO(4,2) are associated with the gauge and scalar photons in conformal QED.

The representations of SO(4,2) will not be investigated directly, but indirectly by means of the representations of the Lie algebra so(4,2). A representation of so(4,2) will be said to be unitary if it can be integrated to a unitary representation of a covering of SO(4,2). For calculations a basis for so(4,2) will be used, denoted $\{L_{\alpha\beta}\}$ with $\alpha,\beta=0,1,\ldots,5$ and $L_{\alpha\beta}=-L_{\beta\alpha}$. The commutation relations are

$$[L_{\alpha\beta},L_{\gamma\delta}]=i(\delta_{\beta\gamma}L_{\alpha\delta}+\delta_{\alpha\delta}L_{\beta\gamma}-\delta_{\beta\delta}L_{\alpha\gamma}-\delta_{\alpha\gamma}L_{\beta\delta}), \tag{3.1}$$

$$(\delta_{\alpha\beta})=\text{diag}(+1,-1,-1,-1,-1,+1)\,. \tag{3.2}$$

A maximum compact subalgebra is generated by L_{05}, that spans so(2), and by $\{L_{ij}\}$, $i,j=1,2,3,4$, that span so(4). The generator L_{05} is interpreted physically as an energy, and so its spectrum must be positive in all representations associat-

ed with physical states. Interest consequently focuses on representations with minimal weights.

Roots and weights will be referred to the ordered basis

$$H_1 = L_{05},\ H_2 = \tfrac{1}{2}(L_{12} + L_{34}),\ H_3 = \tfrac{1}{2}(L_{12} - L_{34}) \quad (3.3)$$

of a Cartan subalgebra. A weight or a root is thus labeled by three real numbers and written (E, j_1, j_2); it is positive if $E > 0$, or if $E = 0$ and $j_1 < 0$, or if $E = j_1 = 0$ and $j_2 < 0$. With this convention, a "minimal weight" will have the smallest value of E and among the weights with the same value of E it will have the *largest* value of j_1 and j_2. Any irreducible representation of so(4,2) that contains a minimal weight (E_0, j_1, j_2) is determined up to equivalence by that weight and will be denoted $D(E_0, j_1, j_2)$. All representations considered here, whether irreducible or not, are K-finite, which implies that each subspace of fixed E (every eigenspace of L_{05}) contains only finite-dimensional representations of so(4) (see Appendix A). The weight (E_0, j_1, j_2) will be said to be K-finite if $D(E_0, j_1, j_2)$ is K-finite.

All representations directly associated with physical states must be unitary, but gauge theories are characterized by nondecomposable representations in which the physical representations appear as (unitary) subquotients. It is important, therefore, to know what irreducible representations can be combined to form nondecomposable representations. If A and B are representations and if we denote by

$$A \rightarrow B$$

a semidirect sum of representations, acting in a space $V = V_A \oplus V_B$ with B acting in the invariant subspace V_B, then V_A will be said to leak into V_B and A will be said to leak into B. (This is because a vector initially lying in V_A will acquire a component lying in V_B after some group action.) The notation $A \rightarrow B$ will be used only for representations that are not equivalent to $A \oplus B$. It is an elementary fact that if A and B are irreducible, then a necessary condition for the existence of a nondecomposable representation $A \rightarrow B$ is that A and B have the same values for all the Casimir operators.[24] If A and B are irreducible representations with minimal weights w_A and w_B, then this condition is the same as the requirement that w_A is related to w_B by a Weyl reflection (see Appendix A). In this case we say that the two minimal weights are Weyl-equivalent. The requirement that w_A be K-finite brings a further restriction, so that not all weights that are Weyl-equivalent are relevant. [In our context, at least, leakage into a representation that is not K-finite cannot occur.]

Let $\{R(E_0, j_1, j_2)\}$, $E_0 \in R$, denote a family of indecomposable or irreducible K-finite representations of so(4,2) with multiplicity-free minimal weight (E_0, j_1, j_2). The K-finiteness of $R(E_0 j_0, j_2)$ implies that $2j_1$ and $2j_2$ are nonnegative integers. It is well known that if $(E_0, j_1. j_2)$ is dominant, then there is no Weyl-equivalent weight lying within the weight diagram of $R(E_0, j_1, j_2)$ and so $R(E_0, j_1, j_2)$ cannot contain an invariant subrepresentation. Precisely, there is no K-finite weight equivalent to and higher than (E_0, j_1, j_2) if

$$E_0 > \begin{cases} j_1 + j_2 + 2, & j_1 > 0 \text{ and } j_2 > 0, \\ j_1 + j_2 + 1, & j_1 = 0 \text{ or } j_2 = 0, \end{cases} \quad (3.4)$$

and $R(E_0, j_1, j_2)$ is irreducible whenever the inequality holds. It turns out that this condition is also sufficient for unitarity. The derivation of (3.4) as well as a demonstration of unitarity is relegated to Appendix A.

So $R(E_0, j_1, j_2)$ is unitary and irreducible as long as the inequality (3.4) holds. Suppose we gradually decrease E_0 until the limit is reached. There then appears within the weight diagram of $R(E_0, j_1, j_2)$ a pair of weights Weyl equivalent to $w_A = (E_0, j_1, j_2)$, namely,

$$w_{B,1} = (j_1 + j_2 + 3, j_1 - \tfrac{1}{2}, j_2 - \tfrac{1}{2}), \quad j_1, j_2 > 0, \quad (3.5)$$

or, if either j_1 or j_2 equals zero,

$$w_{B,2} = (j_1 + j_2 + 3, j - 1, 0), \quad j = j_1 + j_2. \quad (3.6)$$

There is thus an opportunity, but no assurance, for $R(E_0, j_1; j_2)$ to become indecomposable.

It will be shown in Sec. IV that when $j_1 = j_2 = s/2 > 0$ and the limit

$$E = j_1 + j_2 + 2 \quad (3.7)$$

is reached, then there does indeed exist an indecomposable representation $A \rightarrow B$ where A and B have minimal weights

$$w_A = (s + 2, s/2, s/2), \quad w_B = (s + 3, s/2 - \tfrac{1}{2}, s/2 - \tfrac{1}{2}). \quad (3.8)$$

Representations of this type can be associated with "current type gauge fields."

When E_0 is decreased further, below (3.4), then $R(E_0, j_1, j_2)$ again becomes irreducible (although nonunitary) unless E_0 is integral. But when E_0 reaches the values

$$E_0 = j_1 + j_2 + 3 - n, \quad n \leqslant \min\{2j_1, 2j_2\} \quad (3.9)$$

then a Weyl-equivalent weight

$$w_{B,n}(j_1 + j_2 + 3, j_1 - n/2, j_2 - n/2) \quad (3.10)$$

appears within the weight lattice of $R(E_0, j_1, j_2)$ and so there is again an opportunity for an indecomposable structure within $R(E_0, j_1, j_2)$. Representations of this type may be associated with "current type gauge fields of order n."

A qualitatively different phenomenon occurs when E_0 reaches the value

$$E_0 = |j_1 - j_2| + 1. \quad (3.11)$$

At this point another Weyl-equivalent weight appears above $(E_0 j_1 j_2)$:

$$w_C = (|j_1 - j_2| + 2, j_> - \tfrac{1}{2}, j_< + \tfrac{1}{2}), \quad (3.12)$$

where $j_>$ ($j_<$) is the greater (lesser) of the pair $\{j_1, j_2\}$. In the special case when $j_1 = j_2 = s/2 > 0$, we obtain the following set of K-finite Weyl equivalent weights when E_0 reaches unity:

$$w_A = (1, s/2, s/2),$$
$$w_{B,s} = (s + 3, 0, 0),$$
$$w_C = (3, s/2, s/2),$$
$$w_D = (2, s/2 + \tfrac{1}{2}, s/2 - \tfrac{1}{2}),$$
$$w_E = (2, s/2 - \tfrac{1}{2}, s/2 + \tfrac{1}{2}).$$

Thus, we may have, for example,

$$[D(2, s/2 + \tfrac{1}{2}, s/2 - \tfrac{1}{2}) \oplus D(2, s/2 \rightarrow D(1, s/2, s/2) - \tfrac{1}{2}, s/2 + \tfrac{1}{2})].$$

It will be shown in the next section that this structure does, in fact, occur and describes "gradient type gauge fields."

IV. CONFORMAL FIELDS

The manifestly covariant 6-cone formalism was invented by Dirac[2] and developed by Mack and Salam[3] and many others. It is sufficiently well known that detailed explanations can be omitted. The main point is that the action of SO(4,2) in Minkowski space is linearized by introducing two extra coordinates that are subsequently made redundant by a constraint and a projection.

In R^6 introduce coordinates (y_α), $\alpha = 0,1,2,3,4,5$, and the pseudo-Euclidean metric δ defined by $\delta(y) \equiv y^2 \equiv y_\alpha y_\alpha \equiv y_0^2 - \mathbf{y}^2 + y_5^2$. Note that $\mathbf{y}$ stands for (y_1, y_2, y_3, y_4) in this and subsequent sections, and $\mathbf{y}^2 = y_1^2 + y_2^2 + y_3^2 + y_4^2$. The 6-cone is the subset defined by $y^2 = 0$; it is, of course, five-dimensional. Reduction to four dimensions is achieved by projection; that is, by fixing the degrees of homogeneity of all the fields. The resulting projective cone can be identified with compactified Minkowski space. [The projective cone is the homogeneous space $SO(4,2)/H \otimes Z_2$, where H is the 11-parameter Weyl group.]

The field that will be studied in greatest detail in this paper is the electromagnetic potential. The traditional approach associates this with (the restriction to $y^2 = 0$ of) a vector field on R^6. Except for equivalent spinorial formulations, this seems to be the only one that has been investigated. In this section it will be useful to widen the scope somewhat and deal with symmetrical tensor fields of arbitrary degree. In order to avoid cumbersome indices, a set (z_α), $\alpha = 0,1,2,3,4,5$, of auxiliary variables will be introduced, and the symmetric tensor field h with components $h_{\alpha\cdots}$ will be replaced by the polynomial

$$\psi(y,z) = h_{\alpha_1\cdots\alpha_s}(y) z_{\alpha_1}\cdots z_{\alpha_s} . \tag{4.1}$$

The degree of homogeneity in y will be denoted N; thus

$$\hat{N}\psi = N\psi, \quad \hat{n}\psi = s\psi , \tag{4.2}$$

$$\hat{N} \equiv y\cdot\partial_y, \quad \hat{n} \equiv z\cdot\partial_z . \tag{4.3}$$

The number s, sometimes identified with spin, is a nonnegative integer. The number N is real and, in the most important cases, a negative integer. The fields may be multivalued on the cone, and are thus to be understood as sections of a line bundle over the cone, the bundle being determined by N.

Wave equations, subsidiary conditions, etc., must be expressed in terms of operators that are defined intrinsically on the cone. An intrinsically defined differential operator acts on functions over R^6 in such a way as to leave invariant the subspace of functions that vanish on the cone. Thus, if D is an intrinsic differential operator, then there is another operator D' such that $Dy^2 = y^2 D'$. It will be convenient to set $z^2 = 0$ also, and consider operators that are intrinsic to $y^2 = z^2 = 0$. Invariant, intrinsic operators are sums of products of the following:

$$\mathrm{Tr} = y\cdot z\ \partial_z^2 - (2\hat{n} + d + 1)\, y\cdot\partial_z , \tag{4.4}$$

$$\mathrm{Div} = y\cdot z\ \partial_y\cdot\partial_z - z\cdot\partial_y\ y\cdot\partial_z , \tag{4.5}$$

$$\mathrm{Grad} = y\cdot z\ \partial_y^2 - (2\hat{N} + d + 1)\, z\cdot\partial_y . \tag{4.6}$$

The parameter d is equal to 3 in the case of present interest. More generally, d is the number of space dimensions, $y^2 = 0$ is a cone in R^{d+3}, and the conformal group is SO(d + 1,2). This simple generalization will save the labor of repeating the whole discussion when the case $d = 2$ comes up later.

The Casimir operators of so(d + 1,2) can be expressed in terms of Tr, Div and Grad. In particular,

$$Q \equiv \tfrac{1}{2}\sum L_{\alpha\beta}L_{\alpha\beta} = \hat{N}(\hat{N} + d + 1) + \hat{n}(\hat{n} + d - 1) - 2\,\mathrm{Div} . \tag{4.7}$$

Thus, for fields that satisfy the subsidiary condition Div $\psi = 0$ (divergenceless), Q is fixed by the degrees N and s. If Tr ψ = Div $\psi = 0$ and N,s are fixed, then all the Casimir operators reduce to multiplication by constants. It sometimes happens, for the most degenerate representations, that other elements of the enveloping algebra are fixed. Examples are

$$C^{\mu\nu\cdots} \equiv \epsilon^{\alpha\beta\gamma\delta\mu\nu\cdots} L_{\alpha\beta}L_{\gamma\delta} , \tag{4.8}$$

that may vanish if $d = 2$ and may reduce to a multiple of $L^{\mu\nu}$ if $d = 3$, and

$$C'_{\alpha\beta} \equiv L_{\alpha\gamma}L_{\gamma\beta} + L_{\beta\gamma}L_{\gamma\alpha} \quad \{ = [-4/(d+3)]Q\delta_{\alpha\beta} \} . \tag{4.9}$$

The values given in parenthesis are the only eigenvalues possible. One has, up to a numerical factor,

$$\sum C^{\mu\nu\cdots} C^{\mu\nu\cdots}$$
$$\propto [2\,\mathrm{Div} + (d-1)^2 + (2\hat{n} + d - 1)(2\hat{N} + d + 1)]\,\mathrm{Div}$$
$$- 2[z\cdot y\ \partial_y^2 - (2\hat{N} + d + 1)\, z\cdot\partial_y]\,\mathrm{Tr}$$
$$+ 2\hat{n}(\hat{n} + d - 1)(\hat{N} + 1)(\hat{N} + d) . \tag{4.10}$$

Therefore, if Tr and Div vanish, then (4.8) cannot vanish unless either $s = 0$ (trivial case) or $N = -1$ or $N = -d$. The possibility that C or C' may be fixed will not be explored systematically, but it will be discussed later on a case by case basis (see the end of Sec. V).

Wave equations

The only candidate for a second order wave operator on the cone is ∂_y^2, and this is intrinsic only if $2N = 1 - d$, as may be seen from Eq. (4.6). More generally, it is easy to check that $(\partial_y^2)^k$ is intrinsic only if it acts on fields that are homogeneous of degree $N = k - \frac{1}{2}(d + 1)$ in (y_α). In particular, for $d = 3$, ∂_y^2 is intrinsic if $N = -1$ and $(\partial_y^2)^2$ if $N = 0$. For $d = 2$, ∂_y^2 is intrinsic if $N = -\frac{1}{2}$ and $(\partial_y^2)^2$ if $N = \frac{1}{2}$.

Subsidiary conditions

The action of so(d + 1,2) on h is highly reducible even if N is fixed and the wave equation (when it is applicable) is imposed. To improve on this situation, one imposes subsidiary conditions on h: transversality, $y_\alpha h_{\alpha\cdots} = 0$; divergencelessness, $\partial_\alpha h_{\alpha\cdots} = 0$; tracelessness, $h_{\alpha\alpha\cdots} = 0$. In terms of $\psi(y,z)$ these conditions read $y\cdot\partial_z\psi = 0$, $\partial_z\cdot\partial_z\psi = 0$, and $\partial_z^2\psi = 0$, respectively. If $y^2 = z^2 = 0$, then the trace condition becomes redundant, and the other two take the form

$$\mathrm{Tr}\ \psi = 0 \quad (y_\alpha h_{\alpha\cdots} = 0) , \tag{4.11}$$

$$\mathrm{Div}\ \psi = 0 \quad (y_\alpha \partial_\beta h_{\beta\cdots} - \partial_\alpha y_\beta h_{\beta\cdots} = 0) . \tag{4.12}$$

The space of solutions of (4.11), (4.12) and the wave equation (when applicable) can sometimes be partially reduced by recognizing the existence of certain invariant subspaces. These

invariant subspaces are not invariantly complemented; they are therefore associated with indecomposable representations and with gauge theories.

Gauge fields

There are two main types. "Current type" gauge fields are of the form $\psi = y{\cdot}z\Lambda$, where Λ is a polynomial in (z_α). Equations (4.11), (4.12) give these restrictions on Λ:

$$\mathrm{Tr}\,\Lambda = 0, \quad [\mathrm{Div} + \hat{N} + \hat{n} + d + 1]\Lambda = 0\,. \tag{4.13}$$

Now there are several possibilities. If $\mathrm{Div}\,\Lambda = 0$, then $(\hat{N} + \hat{n} + d - 1)\psi = 0$. Such fields will be called first-order (current type) gauge fields. They appear as gauge photons in de Sitter electrodynamics, and in connection with the conformal electromagnetic current. Alternatively, Λ may also be a current type gauge field, $\Lambda = (y{\cdot}z)\Sigma$ or $\psi = (y{\cdot}z)^2\Sigma$. In general, a (current type) gauge field of order k has the form $\psi = (y{\cdot}z)^k\Sigma$, with $\mathrm{Tr}\,\Sigma = \mathrm{Div}\,\Sigma = 0$; it satisfies the subsidiary conditions (4.11), (4.12) only if $N + s + d = k$. Evidently, $k \leqslant s$ so that $N + d \leqslant 0$, and the wave equation is not applicable to current type gauge fields.

If one should want additional justification for calling these fields gauge fields, then it may be pointed out that a de Rham type cohomology can be built up, starting with the notion that an "exact" field is the form $(y{\cdot}z)\Lambda$ and a "closed" field is one that satisfies $\mathrm{Tr}\,\psi = 0$. This becomes particularly evident after performing a Fourier transformation. There is an application of this notation in connection with the conformal currents.

"Gradient type" gauge fields are of the form $\psi = \mathrm{Grad}\,\Lambda$. Equations (4.11), (4.12) impose constraints on Λ:

$$\mathrm{Tr}\,\psi = \mathrm{Grad}\,\mathrm{Tr}\,\Lambda + (\hat{n} - \hat{N})[4\,\mathrm{Div} - (2\hat{n} + d + 1) \times (2\hat{N} + d - 1)]\Lambda = 0\,, \tag{4.14}$$

$$\mathrm{Div}\,\psi = \mathrm{Grad}(\mathrm{Div} + \hat{n} - \hat{N})\Lambda = 0\,. \tag{4.15}$$

Again, there are several possibilities. If $\mathrm{Tr}\,\Lambda = \mathrm{Div}\,\Lambda = 0$, then $(\hat{N} - \hat{n} + 2)\psi$ must vanish. In this case $\psi = \mathrm{Grad}\,\Lambda$ will be called a first-order (gradient type) gauge field. Such fields appear in conformal electrodynamics and also in de Sitter electrodynamics. Alternatively, Λ itself may be a gauge field. It would be straightforward to classify all possibilities.

If $\psi = \mathrm{Grad}\,\Lambda$ is a vector field ($s = 1$), then $N = -1$ and the wave equation $\partial_y^2\psi = 0$ makes sense if $d = 3$. In this case Λ is of degree zero in (y_α) as well as in (z_α), and must satisfy the equation

$$(\partial_y^2)^2\Lambda = 0\,. \tag{4.16}$$

This is the gauge field that appears in conformal QED.

Gauge fields and representations of so(4,2)

The appearance of gauge structures for fields with certain specific degrees of homogeneity is related to the indecomposable representations conjectured in Sec. III. The existence of some of those representations will now be demonstrated by actual construction.

Consider the special field[25]

$$\psi_0 = y_+^{N-s}M(\boldsymbol{\eta}), \quad \boldsymbol{\eta} = y_+\mathbf{z} - z_+\mathbf{y}\,, \tag{4.17}$$

$$y_+ = y_5 + iy_0, \quad z_+ = z_5 + iz_0\,, \tag{4.18}$$

where M is an s-linear, symmetric, traceless form. This field satisfies all the subsidiary conditions; in fact $y{\cdot}\partial_z\psi_0 = \partial_y{\cdot}\partial_z\psi_0 = \partial_z^2\psi_0 = 0$. In addition, ψ_0 is annihilated by the lowering operators $L_{i5} + iL_{i0}$, $i = 1,\ldots,d+1$; see Appendix A. The action of the Lie algebra on ψ_0 therefore generates a module on which ψ_0 is cyclic and of minimal weight. The minimal weight is $(E_0,s/2,s/2)$ for $d = 3$ and (E_0,s) for $d = 2$, with $E_0 = -N$. It was found in Sec. III that the highest value of E_0 for which this module can become indecomposable is $E_0 = s + d - 1$ ($s \geqslant 1$), or $N + s + d = 1$. This is precisely the value of N for which first-order current type gauge fields appear. A straightforward calculation confirms that the representation generated from ψ_0 is

$$D(s+2,s/2,s/2) \to D(s+3,(s-1)/2,(s-1)/2) \tag{4.19}$$

if $d = 3$, and[25]

$$D(s+1,s) \to D(s+2,s-1), \quad s \geqslant 1\,, \tag{4.20}$$

if $d = 2$. The invariant subspace consists of fields of the form $\psi = y{\cdot}z\Lambda$. (Of course, the ground state ψ_0 is not of this form!)

The next lower value of E_0 at which an indecomposable representation can exist was found (Sec. III) to be at $E_0 = s + d - 2$, $s \geqslant 2$, or $N + s + d = 2$. This is the value of N for which second-order current type gauge fields appear. The representation would be

$$D(s+1,s/2,s/2) \to D(s+3,s/2-1,s/2-1) \tag{4.21}$$

for $d = 3$ and $D(s+1,s) \to D(s+3,s-2)$ for $d = 2$; however, whether this is actually realized in the space generated from ψ_0 has not been checked. Similar representations are expected to be associated with current type gauge fields of order $k \leqslant s$. No physical applications are known for $k > 1$. Figure 1 illustrates the general case.

Something qualitatively different happens at $E_0 = 1$, for at this point other elements of the Weyl group, and other equivalent weights, begin to play a role. For $d = 3$ and $s = 1$ the weights equivalent to $(1,\frac{1}{2},\frac{1}{2})$ and inside the weight lattice are $(2,1,0)$, $(2,0,1)$, $(3,\frac{1}{2},\frac{1}{2})$, and $(4,0,0)$. For $d = 2$ and $s = 1$ the weights equivalent to $(1,1)$ and inside the weight lattice are $(2,1)$ and $(3,0)$. The field ψ_0 has the weight $(1,\frac{1}{2},\frac{1}{2})$ if $d = 3$ and $(1,1)$ if $d = 2$, provided M is linear and $N = -1$. This is just the value of N for which first-order gradient type gauge fields should occur. A straightforward calculation shows, however, that the space generated from ψ_0 is irreducible! What happens in this case is that there are indeed nondecomposable representations involving the several equivalent minimal weights, but this time ψ_0 is not cyclic. Only the invariant subspace is generated from ψ_0. Indeed, ψ_0 is a gradient type gauge field in this case:

$$\psi_0 = y_+^{-2}\mathbf{M}{\cdot}\boldsymbol{\eta} = z{\cdot}\partial_y y_+^{-1}\mathbf{M}{\cdot}\mathbf{y} \propto \mathrm{Grad}(y_+^{-1}\mathbf{M}{\cdot}\mathbf{y})\,. \tag{4.22}$$

A nondecomposable representation is generated from

$$(\Psi_{0\rho})_{ij} = y_+^{-2}(y_iz_j - y_jz_i), \quad i,j = 1,\ldots,d+1\,. \tag{4.23}$$

For $d = 2$, these fields carry an irreducible representation of $\mathrm{so}(2) \oplus \mathrm{so}(3)$ including the weight $(2,1)$. For $d = 3$ the self-dual and anti-self-dual parts are associated with $(2,1,0)$ and

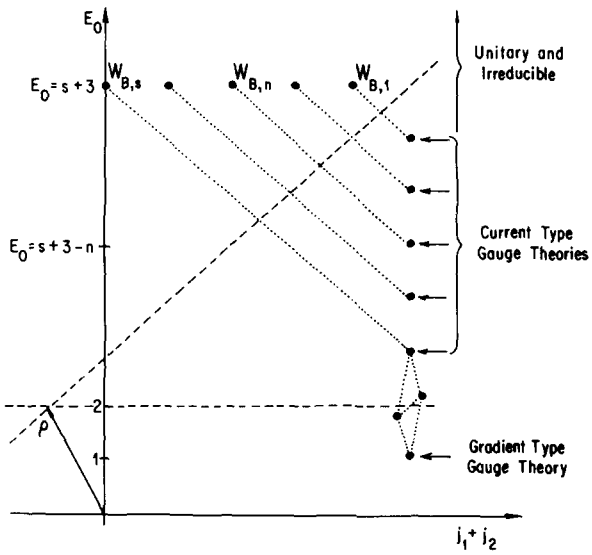


FIG. 1. The principal reduction points for $j_1 = j_2 = s/2 > 0$. The planes of reflection are indicated by dashed lines and equivalent weights are connected by dotted lines. If the lowest weight (on the line $j_1 + j_2 = s$) is $w_A = (E_0,s/2,s/2)$, then the first reduction point occurs at $E_0 = s + 2$ with equivalent weight $w_{B,1} = (s + 3,(s/2) - (1/2),(s/2) - (1/2))$. The nth ($n \leqslant s$) occcurs at $E_0 = s + 3 - n$ with equivalent weight $w_{B,n} = (s + 3,\frac{1}{2}(s - n),\frac{1}{2}(s - n))$. The weight labeled $w_{B,s+1}$ is Weyl-equivalent to $(2,s/2,s/2)$ but it is not K-finite. The weight $w_A = (1,s/2,s/2)$ has four K-finite equivalent weights $w_{B,s} = (s + 3,0,0)$, $w_C = (3,s/2,s/2)$, $w_D = (2,(s/2) + (1/2),(s/2) - (1/2))$, and $w_E = (2,(s/2) - (1/2),(s/2) + (1/2))$.

with (2,0,1). The space generated from (4.23) includes the space of gauge fields generated from (4.17) as an invariant subspace, and the complete representation is

$$[D(2,1,0) \oplus D(2,0,1)] \to D(1,\tfrac{1}{2},\tfrac{1}{2}) \tag{4.24}$$

when $d = 3$, and

$$D(2,1) \to D(1,1) \tag{4.25}$$

when $d = 2$.

Our investigation of alternate gauge structures ends here, since those that are of direct interest to electrodynamics have already been described. To sum up, two possibilities are offered for conformal electrodynamics, the representation (4.21) with $s = 1$, and (4.24). The first will turn out to be relevant in connection with the currents. The other is a very strong candidate for describing the electromagnetic field, for two reasons. Since $N = -1$, the wave operator ∂_y^2 is available to form an invariant action and covariant wave equation. The representations $D(2,1,0)$ and $D(2,0,1)$ that appear in (4.24) are precisely those representations of SO(4,2) that remain irreducible when restricted to the Poincaré subgroup. [Two possibilities are available for de Sitter space electrodynamics also. In another paper it will be shown that both of these appear when (4.24) is restricted to SO(3,2).] There is also a third possibility, pointed out by Dirac.[2] Conformal photons may be described by a field strength; then one is dealing with $[D(2,1,0) \oplus D(2,0,1) \to D(3,\frac{1}{2},\frac{1}{2})$.]

The representation (4.24) contains two parts of the Gupta–Bleuler triplet of conformal electrodynamics. To discover the extension to the full triplet, including the representation associated with the scalar photons, that is needed to carry out conformally invariant quantization, we shall calculate the (homogeneous) propagator.

V. HOMOGENEOUS PROPAGATORS

In Sec. II the original Gupta–Bleuler quantization paradigm was described. All components of $A_\mu(x)$ were quantized; that means that the set of one-particle states (including all the ghosts) transforms as the direct product $D_4 \otimes D(0,0)$, where D_4 is the finite-dimensional vector representation that is trivial on the translations. In fact, this direct product is equivalent to (2.5). The homogeneous propagator was then defined by (2.7) and can be evaluated subsequently. The result is that $K_{\mu\nu}(x,x') = \delta_{\mu\nu} D(x,x')$, where $D(x,x')$ is the propagator for the scalar field and thus for the representation $D(0,0)$ of the Poincaré group, while $(\delta_{\mu\nu})$ can be regarded as the propagator for D_4. A guess based on analogy would lead directly to the propagator (5.14) for conformal QED. However, a more systematic approach is worthwhile.

Return to the context and notations of Sec. IV. Let the degrees N,s of the field, in (y_α), resp. (z_α), be fixed. Set $z^2 = y^2 = 0$. The homogeneous propagator K is a distribution in y,z and y',z' that is homogeneous of degree N in (y_α) and in (y'_α), a polynomial of degree s in (z_α) and in (z'_α), and invariant under the action of so($d + 1,2$). The most general ansatz is

$$K = \sum_{a,b} C_{ab} (y\cdot y')^{N-a-b} (z\cdot z')^{s-a-b} [y\cdot z'\, y'\cdot z]^a [y\cdot z\, y'\cdot z']^b . \tag{5.1}$$

The sum is over $a,b = 0,1,\cdots$ and the coefficients C_{ab} are real numbers to be determined. The factor $(y\cdot y')^{N-a-b}$ is a distribution and needs to be defined. When $s = 0$, then the propagator is

$$(2y\cdot y')^N = (YY')^N \sum_L e^{-i\tau(L-N)} C_L^{-N}(\hat{y}\cdot\hat{y}') , \tag{5.2}$$

where C_L^{-N} is a Gegenbauer polynomial and

$$y_+ = y_5 + iy_0 = Y e^{it}, \quad \mathbf{y} = Y\hat{y}, \quad \tau = t - t' . \tag{5.3}$$

This is a Fourier expansion in which the L th term corresponds to the eigenvalue $L - N$ of L_{05}, and the lowest value of the energy ($= L_{05}$) is $-N$. The distributions $(y\cdot y')^{N-a-b}$ must be understood in this sense, so that K will be the propagator for a representation in which the energy has a lower bound. That lower bound is $-N - s$ in general (Appendix C), but a special choice of the coefficients in (5.1) may make it higher.

"Massive" representations of so(4,2) are those minimal weight representations for which the inequality (3.4) holds, so that no invariant subspaces can occur. [Massive representations of so(3,2) are those $D(E_0,s)$ for which $E_0 > s + 1$.] Hence $D(E_0,j_1,j_2)$ is massive if the weight (E_0,j_1,j_2) lies above the first reduction point. Recall that, in this case, $D(E_0,j_1,j_2)$ is irreducible. For such representations the free fields, *and therefore also the homogeneous propagator*, satisfy the subsidiary conditions. Ordinary massive vector fields, transforming irreducibly under the Poincaré group, furnish the standard analogy. The fields satisfy the divergence condition $\partial\cdot A = 0$. After four-dimensional Fourier transforma-

tion, $\tilde{A}_\mu(p) = \mathrm{FT}\, A_\mu(x)$, this reads $p\cdot\tilde{A} = 0$. The propagator becomes

$$K_{\mu\nu} = (\delta_{\mu\nu} - p_\mu p_\nu/m^2)\delta(p^2 - m^2) \tag{5.4}$$

and $p_\mu K_{\mu\nu} = 0$.

The first problem is to determine K for the "massive" case. The subsidiary conditions are $\mathrm{Tr}\, K = 0$, $\mathrm{Div}\, K = 0$, with Tr and Div defined by (4.4) and (4.5). These are solved in Appendix B. The result, for $s = 1$, is

$$K = (y\cdot y')^{N-1}\{\, y\cdot y'\, z\cdot z' - y\cdot z'\, y'\cdot z + [(1-N)/(N+d)]\, y\cdot z\, y'\cdot z'\,\}\,. \tag{5.5}$$

A close look at this function gives yet another method for discovering nondecomposable representations and gauge structures.

Current-type gauge theory

The minimal energy in the Fourier expansion of (5.5) is $E_0 = -N$. The minimal weight, in the case $d = 3$, is $(E_0, \frac{1}{2}, \frac{1}{2})$. Both of these statements are proved in Appendix C. As E_0 is decreased, gauge phenomena are expected as E_0 reaches the limit set by (3.4); that is, as E_0 reaches the value 3, or $N = -3 = -d$. This is revealed by (5.5) since the coefficient of the last term blows up at $N = -d$. The analogy with (5.4) as m^2 is decreased to zero is striking and illuminating.

When $E_0 = -N = d$ there is only one invariant propagator that satisfies the subsidiary conditions. It is obtained by taking the limit of $(N+d)K$ and is proportional to

$$K_{g+} = (y\cdot z)(y\cdot y')^{-d-1}(y'\cdot z')\,. \tag{5.6}$$

But the fields that contribute to the factorization

$$K_{g+} = \sum_\psi \psi(y,z)\psi^*(y',z') \tag{5.7}$$

are all gauge fields of the form $\psi = y\cdot z\Lambda$. To retain the full nondecomposable representation one takes the limit (q fixed, real) of

$$K - [(1-N)/(N+d) + q]K^N_{g+}\,,$$
$$K^N_{g+} = y\cdot z(y\cdot y')^{N-1}y'\cdot z'\,. \tag{5.8}$$

This yields the following propagator for $N = -d$:

$$K^+_q = (y\cdot y')^{-d-1}[\, y\cdot y'\, z\cdot z' - y\cdot z'\, y'\cdot z] - qK_{g+}\,. \tag{5.9}$$

The constant q is arbitrary and is ultimately fixed by convenience. Analogous treatment of (5.4) gives the familiar result; thus, in the Feynman gauge ($q = 0$), $k_{\mu\nu} = \delta_{\mu\nu}\delta(p^2)$. Spectral decomposition of (5.9) into products of fields, which will reveal the Gupta–Bleuler triplet, will be carried out in Sec. VI.

Gradient-type gauge theory

Consider the propagator $(y\cdot y')^{N+1}$ for a field Λ with $s = 0$. The propagator for the first-order gradient type gauge field $\psi = \mathrm{Grad}\,\Lambda$ is

$$\mathrm{Grad\,Grad}'(y\cdot y')^{N+1}\{\, y\cdot y'\, z\cdot z' + N\, z\cdot y'\, z'\cdot y - [2N/(2N+d+1)]\, y\cdot z\, y'\cdot z'\,\}\,. \tag{5.10}$$

The operator Grad was defined by (4.6). According to a result of Sec. IV, this propagator satisfies the subsidiary conditions only if $N = -1$, in which case it is

$$K_{g-} = (y\cdot y')^{-2}\{\, y\cdot y'\, z\cdot z' - z\cdot y'\, z'\cdot y + [2/(d-1)]\, y\cdot z\, y'\cdot z'\,\}\,. \tag{5.11}$$

Now this is exactly the same as (5.5) when $N = -1$. Thus, in the limit $N \to -1$, the propagator K given by (5.5) propagates only gauge fields. This time the gauge fields are of gradient type. To retain the full, nondecomposable representation one must take the limit of

$$(N+1)^{-1}(K^N_{g-} - K), \quad K^N_{g-} = \mathrm{Grad\,Grad}'(y\cdot y')^{N+1}\,, \tag{5.12}$$

as N tends to -1. With the addition of an arbitrary multiple of K_{g-} the result is

$$K^-_q = (y\cdot y')^{-2}\{z\cdot y'\, z'\cdot y - [(d+1)/(d-1)^2]\, y\cdot z\, y'\cdot z'\} + qK_{g-}\,. \tag{5.13}$$

With one notable exception, $K^\pm_q$ does not factorize as in $(y\cdot y')^{-\nu}z\cdot z'$, so the direct product structure that was discussed in the beginning of this section does not appear. The one exception is that, when $d = 3$ only, the choice $q = 1$ leads to $K^-_1 = D$ with

$$D = (y\cdot y')^{-1}z\cdot z' \quad (d = 3, \text{ Feynman gauge}). \tag{5.14}$$

This is the preferred propagator for conformal electrodynamics. The Gupta–Bleuler triplet will be found by inspection of its modal decomposition, in Sec. VI.

The subsidiary conditions

The massive propagator (5.5), and the gauge field propagators (5.6) and (5.11), all satisfy the subsidiary conditions, $\mathrm{Tr}\, K = \mathrm{Div}\, K = 0$. However, the gauge field propagator K^N_{g+} that was needed in (5.8) to obtain the physical propagator K^+_q satisfies only $\mathrm{Tr}\, K^N_{g+} = 0$. Consequently,

$$\mathrm{Tr}\, K^+_q = 0, \quad \text{but} \quad \mathrm{Div}\, K^+_q \neq 0\,. \tag{5.15}$$

For very similar reasons,

$$\mathrm{Tr}\, K^-_q \neq 0, \quad \mathrm{Div}\, K^-_q \neq 0\,. \tag{5.16}$$

This has implications for the scalar states as will be seen below.

Let us now study the operator (4.8), in the case of gradient type gauge fields, $N = -1$ and $s = 1$. The ground state ψ_0 of (4.17) is annihilated by $C^{\alpha\beta}$, and so is the gauge field propagator (5.11). As for the physical ground state (4.23), one can verify that (when $d = 3$)

$$(C^{\alpha\beta} + 8L^{\alpha\beta})\,\psi^+_{op} \simeq 0, \quad (C^{\alpha\beta} - 8L^{\alpha\beta})\,\psi^-_{op} \simeq 0\,. \tag{5.17}$$

Here ψ^+_{op} (ψ^-_{op}) is the self-dual (anti-self-dual) part of ψ_{op}, and $\psi \simeq 0$ means that ψ is a gauge field. The states ψ^+_{op} and ψ^-_{op} are the ground states of $D(2,1,0)$ and $D(2,0,1)$, respectively, and (5.17) shows that neither of these representations can be extended to the other; that is, $D(2,0,1) \to D(2,1,0)$ does not exist. It is very interesting that, in spite of this, both leak to the same representation $D(1, \frac{1}{2}, \frac{1}{2})$, and both must be included in the quantization. There is no invariant propagator that propagates one without the other.

As for (4.9), one may check that $C'_{\alpha\beta}$, applied to the physical ground state (4.23) is a gauge field,

$$C'_{\alpha\beta}\,\psi_{op} \simeq 0\,. \tag{5.18}$$

Hence $C'_{\alpha\beta}$ vanishes on the physical (quotient) representa-

tions. It is well known that $C'_{\alpha\beta} - 2\delta_{\alpha\beta}$ vanishes in $D(1,0,0)$ and that $C'_{\alpha\beta} - \delta_{\alpha\beta}$ vanishes in $D(\frac{1}{2},0)$, representations of SO(4,2) and SO(3,2), respectively.

VI. CONFORMAL GUPTA–BLEULER TRIPLETS

It is instructive to examine, in somewhat greater detail, the cancellations between terms of order $(N+d)^{-1}$ in (5.8). The q term is irrelevant, so let $q=0$. The second term has a decomposition in terms of gauge fields, Eq. (5.7), that may be abbreviated as

$$\frac{1-N}{N+d} K^{N}_{g+} = \epsilon^{-1} \sum_{g} gg^*, \quad \epsilon = N+d. \tag{6.1}$$

Similarly, the massive propagator (5.5) has a decomposition

$$K = \sum_{p} pp^* + \epsilon^{-1} \sum_{\phi} \phi\phi^*. \tag{6.2}$$

Since the ϵ^{-1} terms cancel, $\phi - g$ must be of order ϵ, and

$$K - \frac{1-N}{N+d} K^{N}_{g+} = \sum pp^* + \sum (gs^* + sg^*) + \epsilon \sum ss^*, \tag{6.3}$$

where $s = (\phi - g)/\epsilon$ remains finite as $\epsilon \to 0$. This field that has been denoted s, is the third member of the Gupta–Bleuler triplet. It is easy to see, in fact, that, under the action of so(4.2), s leaks to p and p leaks to g. *The important point is that* ss^* terms are of order ϵ and disappear in the limit. Compare Eq. (2.8) and the discussion that follows it. Naturally, this rough sketch is not intended to prove anything.

Recall that (6.2) satisfies the subsidiary conditions; therefore so do the fields ϕ. The gauge fields g satisfy the subsidiary conditions at $\epsilon = 0$ only. Since $s = -\partial g/\partial\epsilon$ at $\epsilon = 0$, it follows that the scalar fields may not satisfy the subsidiary conditions: Tr $s \neq 0$ and Div $s \neq 0$. Now the absence of ss^* terms from (6.3) can be tested simply by verifying that Tr K and Div K are gauge fields in the second variable. Thus, in the case of the propagator $D = (y\cdot y')^{-1} z\cdot z'$, one has Tr $D = (y\cdot y')^{-1} y\cdot z' = z'\cdot\partial_{y'} \ln y\cdot y'$. However, we prefer a more detailed proof that gives specific information about the nature of the scalar fields.

Current-type gauge theory

The propagator (5.9), with $q=0$, is dissected in Appendix C, and only the results will be quoted here. The propagator is considered as the integration kernel of an operator, and spectral resolution of this operator by projection to the eigenstates of L_{05} yields the following.

The lowest energy is $E_0 + d$, and the projection of K_0^+ on this subspace is

$$(K_0^+)_0(y,z;y',z') = \mathbf{p}_0(y,z)\cdot \mathbf{p}_0^*(y',z'), \tag{6.4}$$

$$\mathbf{p}_0(y,z) = y_+^{-d-1}\boldsymbol{\eta}, \quad \boldsymbol{\eta} = y_+ \mathbf{z} - z_+ \mathbf{y}. \tag{6.5}$$

The field $\mathbf{p}_0$ is annihilated by the lowering operators and is an eigenstate for a true minimal weight. This vector is *not* cyclic for the entire space of fields that occurs in the spectral decomposition of K_0^+.

The projection of K_0^+ on the next higher eigenspace is found to contain a contribution of the form $s_0(y,z)\, g_0^*(y',z') + g_0(y,z) s_0^*(y',z')$, with

$$s_0(y,z) = y_+^{-d-2}\, \mathbf{y}\cdot\boldsymbol{\eta}, \quad g_0(y,z) = y_+^{-d-1}\, y\cdot z. \tag{6.6}$$

The field s_0 is cyclic for the whole space $\mathscr{Y}''$ of fields that occur in the spectral decomposition of K_0^+. The field p_0 generates an invariant subspace $\mathscr{Y}$ of $\mathscr{Y}''$ and g_0 generates an invariant subspace $\mathscr{Y}_g$ of $\mathscr{Y}$. The weights of s_0, p_0, and g_0 are the minimal weights of $\mathscr{Y}''/\mathscr{Y}$, $\mathscr{Y}/\mathscr{Y}_g$, and $\mathscr{Y}_g$, respectively; when $d=3$ they are (4,0,0), $(3,\frac{1}{2},\frac{1}{2})$ and (4,0,0). The Gupta–Bleuler triplet is therefore

$$D(4,0,0) \to D(3,\tfrac{1}{2},\tfrac{1}{2}) \to D(4,0,0). \tag{6.7}$$

[When $d=2$, the triplet is $D(3,0) \to D(2,1) \to D(3,0)$.]

As expected, there are no ss^* terms in K_0^+; therefore, to assure that only $D(3,\frac{1}{2},\frac{1}{2})$ propagate between curents, it is sufficient that

$$(g,j) = \int dy\, (y_\alpha \Lambda)^* j_\alpha = \int dy\, \Lambda^* y\cdot j = 0. \tag{6.8}$$

The current must be transverse, $y\cdot j = 0$, but it need not be divergenceless. [Actually, the terminology is misleading, for in the physical applications the roles of fields and currents will be reversed; see Sec. VII.]

As was pointed out in Sec. V (last section), Tr $K_0^+ = 0$, so this subsidiary condition holds in $\mathscr{Y}''$. If $\psi(y,z) = z_\alpha h_\alpha(y)$, then

$$y_\alpha h_\alpha = 0, \quad \text{all fields } h \text{ in } \mathscr{Y}''. \tag{6.9}$$

On the other hand, Div $K_0^+ \neq 0$, so the subsidiary condition Div $\psi = 0$ holds only in $\mathscr{Y}$. This is therefore the Lorentz condition. In terms of h,

$$y_\alpha\, \partial\cdot h - \partial_\alpha\, y\cdot h = 0 \quad \text{(Lorentz condition, current type)}. \tag{6.10}$$

The field $y_\alpha\, \partial\cdot h - \partial_\alpha\, y\cdot h$ is a gauge field, for (6.9) implies that $\partial_\alpha\, y\cdot h = y_\alpha\, \Lambda$. The map $\psi \to \text{Div}\,\psi$ is therefore a map of $\mathscr{Y}''$ onto $\mathscr{Y}'_g$ with kernel $\mathscr{Y}$. The second order Casimir operator Q is the same as -2 Div in $\mathscr{Y}''$ [Eq. (4.7)]; therefore,

$$Q\mathscr{Y}'' = \mathscr{Y}'_g, \quad Q\mathscr{Y} = 0, \tag{6.11}$$

and the Casimir operator is nondiagonalizable on $\mathscr{Y}''$.

Gradient-type gauge theory

The analysis of the propagator K_q^-, Eq. (5.13), is a good deal more complicated. The details are in Appendix C. The results are as follows.

The lowest energy is 0. The projection of K_0^- on this eigenspace is

$$(K_0^-)_0 = -\psi_0\psi_0^*, \quad \psi_0(y,z) = y_+^{-1} z_+. \tag{6.12}$$

This ψ_0 is interesting. Locally, it is a gauge field, $\psi_0 = \text{Grad}\,\Lambda$, but with $\Lambda = \ln y_+$. The term (6.12) is therefore not totally negligible; more about this later. Anyway, the space generated from ψ_0 consists entirely, with the sole exception of ψ_0 itself, of true gauge fields. This space may be denoted $\mathscr{Y}_1$.

The next energy is 1, and the projection of K_0^- on it is

$$(K_0^{-1})_1 = 2(\psi_g\cdot\psi_s^* + \psi_s\cdot\psi_g^*), \tag{6.13}$$

$$\psi_s = y_+^{-2}\, \mathbf{y} z_+, \quad \psi_g = y_+^{-2}\boldsymbol{\eta} = \text{Grad}\, y_+^{-1}\mathbf{y}. \tag{6.14}$$

The field ψ_s is cyclic for the whole space $\mathscr{V}'$ of fields that contribute to K_0^-. The field ψ_g is a gauge field and generates an invariant subspace $\mathscr{V}_g \subset \mathscr{V}'$.

The next energy is 2 and the projection of K_0^- on it has the form

$$(K_0^-)_2 = \psi_p \cdot \psi_p^* - \sum(\psi_s \psi_g^* + \psi_g \psi_s^*), \quad \psi_p = y_+^{-2}\, \mathbf{y} \wedge \mathbf{z}\,. \tag{6.15}$$

The ψ_g are gauge fields and belong to $\mathscr{V}_g$. The field ψ_p generates an invariant subspace $\mathscr{V}_2$ of $\mathscr{V}'$ that includes $\mathscr{V}_g$ (but not $\mathscr{V}_1$). Let $\mathscr{V} = \mathscr{V}_1 \cup \mathscr{V}_2$. The decomposition of the total space $\mathscr{V}'$ is then summed up by the inclusions $\mathscr{V}' \supset \mathscr{V} \supset \mathscr{V}_g$. Neither subspace is invariantly complemented. The weights of ψ_s and ψ_g are the (equal) minimal weights of $\mathscr{V}'/\mathscr{V}$ and $\mathscr{V}_g$. The quotient $\mathscr{V}/\mathscr{V}_g$ has two minimal weights, the weights of ψ_0 and ψ_p. When $d = 3$ the total representation is the following "augmented triplet:"

$$D(1,\tfrac{1}{2},\tfrac{1}{2}) \begin{cases} \nearrow [D(2,1,0) \oplus D(2,0,1)] \\ \searrow D(0,0,0) \end{cases} \begin{cases} \searrow \\ \nearrow \end{cases} D(1,\tfrac{1}{2},\tfrac{1}{2})\,. \tag{6.16}$$

For $d = 2$ one has instead

$$D(1,1) \begin{cases} \nearrow D(2,1) \\ \searrow D(0,0) \end{cases} \begin{cases} \searrow \\ \nearrow \end{cases} D(1,1)\,. \tag{6.17}$$

The representations in the middle are found on $\mathscr{V}/\mathscr{V}_g$.

There are several reasons for setting the trivial representation apart from the physical representation. The propagator K_0^- is a positive-definite operator on $\mathscr{V}_2/\mathscr{V}_g$, on which the group action is the physical representation $D(2,1,0) \oplus D(2,0,1)$ or $D(2,1)$. It is negative on $\mathscr{V}_1/\mathscr{V}_g$, on which the group action is trivial. This may be seen by inspecting (6.12) and (6.15). Therefore, the space of physical states is $\mathscr{V}_2/\mathscr{V}_g$ only. In addition, the zero energy state ψ_0 is, strictly speaking, not a gauge field, as was pointed out below (6.12).

The representation (6.16) is (algebraically) equivalent to the direct product representation $D_6 \otimes D(1,0,0)$ of SO(4,2), as was mentioned in connection with (5.14). The representation (6.17) of SO(3,2) is *not* equivalent to $D_5 \otimes D(1,0)$, since this direct product contains (6.17) as well as $D(2,0)$. [D_6 and D_5 are the six- and five-dimensional natural representations of SO(4,2) and SO(3,2).] [If $d = 3$, then Q^2 vanishes in the direct product $D_6 \otimes D(1,0,0)$, since in that case Div Div vanishes. If $d = 2$, then the minimal polynomial for Q in $D_5 \otimes D(1,0)$ is $Q^2(Q+2)$. The fact that $Q^2 = 0$ on $\mathscr{V}'$ is revealed by $(\partial_z \cdot \partial_y - \partial_y^2\, y \cdot \partial_z)\, K_q^- = 0$ (for $d = 2$).]

There are no ss^* terms in K_q^-, and so, to ensure that only the physical states propagate between currents, it is *almost* sufficient that

$$(g,j) = \int dy\, (\mathrm{Grad}\, \Lambda)_\alpha^*\, j_\alpha = 0\,. \tag{6.18}$$

This will be shown (in Appendix D) to be equivalent to a conservation law for j. The reason for the qualification "almost" is that (6.18) does not eliminate the contribution of the zero-energy state. The current will be discussed in the next section.

As was pointed out in Section V, Eq. (5.16), Tr $K_q^- \neq 0$ and Div $K_q^- \neq 0$. Therefore, Tr $\psi =$ Div $\psi = 0$ holds only for physical and gauge photons. More precisely, Tr $\psi = 0$ in $\mathscr{V}_2$ (but Tr $\psi_0 =$ const), while Div $\psi_0 = 0$ in $\mathscr{V} = \mathscr{V}_1 \cup \mathscr{V}_2$ (thus Div $\psi_0 = 0$). Since the zero energy state is unphysical (having negative norm), the Lorentz condition must be Tr $\psi = 0$ rather than Div $\psi = 0$. If $\psi(y,z) = z_\alpha a_\alpha(y)$, then it is

$$y \cdot a = 0 \quad \text{(Lorentz condition, gradient type).} \tag{6.19}$$

The scalar (i.e., spinless) field $y \cdot a$ represents the scalar photons *and* the zero energy state.

It should be noted that the wave equation

$$\partial^2 a = 0 \tag{6.20}$$

holds in $\mathscr{V}'$. Applying Grad to $y \cdot a$, one confirms, using (6.20), that Div $\psi = 0$ in $\mathscr{V}_2$. The map $\psi \to$ Div ψ maps $\mathscr{V}'$ onto $\mathscr{V}_g$, and (6.11) holds in this case also.

VII. CONFORMAL QED

The preparations have now been completed. It is clear that the gradient type gauge theory is the most reasonable candidate for application to electrodynamics, because (a) there is a second-order invariant wave operator in this case only and (b) the center of the triplet (6.16) is precisely (except for the additional trivial representation) the representation that, when restricted to the Poincaré subgroup, describes massless particles with helicities + 1 and − 1. In this section $d = 3$, of course.

Free wave equation

There is no conformally invariant wave operator of second order that is also gauge-invariant. Recall that a gradient type vector gauge field is of the form Grad Λ, where Λ is a scalar field. A gauge transformation is thus a substitution $\psi \to \psi +$ Grad Λ. [In this context, "gauge invariance" means invariance under general gauge transformations, with no restrictions on the gauge parameter Λ.] If $\psi(y,z) = z \cdot a(y)$, then this is the same as

$$a_\alpha \to a_\alpha + \mathrm{Grad}_\alpha\, \Lambda\,. \tag{7.1}$$

$$\mathrm{Grad}_\alpha = y_\alpha\, \partial^2 - (2\hat{N} + 4)\, \partial_\alpha\,. \tag{7.2}$$

Since Grad_α commutes with Grad_β, one can build a gauge-invariant "field strength"; since $\hat{N}\partial_\alpha a_\beta = -2\partial_\alpha a_\beta$, one has

$$f_{\alpha\beta} = \mathrm{Grad}_\alpha\, a_\beta - \mathrm{Grad}_\beta\, a_\alpha = y_\alpha\, \partial^2 a_\beta - y_\beta \partial^2 a_\alpha\,. \tag{7.3}$$

[The quantity $\partial_\alpha a_\beta - \partial_\beta a_\alpha$ is not intrinsically defined, since $\partial_\mu(y^2 \Lambda_\beta) - \partial_\beta(y^2 \Lambda_\alpha)$ does not vanish at $y^2 = 0$; see Dirac.[1]] However, $f_{\alpha\beta} f_{\alpha\beta}$ is not a suitable action density. Another gauge invariant quantity is the completely antisymmetric tensor

$$f_{\alpha\beta\gamma} = \sum(\pm)\, y_\alpha\, \partial_\beta a_\gamma\,; \tag{7.4}$$

but $f_{\alpha\beta\gamma} f_{\alpha\beta\gamma}$ is not even of the correct degree to serve as an action density.

The simplest gauge invariant action for the free field is the following:

$$\mathscr{L}_1 = \frac{1}{2}\int dy\, a\cdot\partial^2\,\mathrm{Div}\, a\,. \tag{7.5}$$

This is gauge-invariant because Div Grad $\Lambda = 0$. It is a special case ($c = 1$) of the one-parameter family (c real)

$$\mathscr{L}_c = \frac{1}{2}\int dy\, a_\alpha \partial^2(\delta_{\alpha\beta} + icM_{\alpha\beta})\, a_\beta,$$

$$M_{\alpha\beta} = i(y_\alpha\partial_\beta - y_\beta\partial_\alpha)\,. \tag{7.6}$$

Note that the field equation derived from (7.5) contains derivatives up to the third order.

A gauge invariant action may be convenient in classical electrodynamics, although gauge invariance is usually sacrificed by imposition of the Lorentz condition. In quantum field theory it is necessary to choose a propagator, and this amounts to giving up gauge invariance in the action. The obvious choice of action, with gauge fixing already built in, is the unique conformally invariant action that leads to a second-order wave equation, namely, (7.6) with $c = 0$:

$$\mathscr{L} = \int dy\,(\tfrac{1}{2}a_\alpha\partial^2 a_\alpha - j\cdot a)\,. \tag{7.7}$$

An interaction with an external source has been included.

The current

Though the kinetic term need not, and indeed cannot, be gauge-invariant in quantum field theory, the interaction functional must be gauge-invariant to ensure physical unitarity. As was pointed out in the preceding section, this implies that the integral (6.18) must vanish. It is shown in Appendix D that the necessary and sufficient conditions on j are

$$(\hat{N}+3)j_\alpha = 0, \quad \partial^2 y\cdot j + 2\partial\cdot j = 0\,. \tag{7.8}$$

It is not difficult to construct field theoretical models that furnish examples of such currents. Thus, consider a spinor field χ of degree -2, with the free action proposed by Dirac

$$\mathscr{L}_\chi = \frac{1}{2}\int dy\,\bar{\chi}\gamma_\alpha\gamma_\beta M_{\alpha\beta}\chi\,. \tag{7.9}$$

It is not obvious that the usual minimal substitution $\partial_\alpha \to \partial_\alpha - iea_\alpha$ is consistent with invariance under the gauge transformations (7.1), but here it amounts to

$$M_{\alpha\beta}\to M_{\alpha\beta} + e\mathscr{A}_{\alpha\beta}, \quad \mathscr{A}_{\alpha\beta}\equiv y_\alpha a_\beta - y_\beta a_\alpha\,. \tag{7.10}$$

The effect of the gauge transformation (7.1) is

$$\mathscr{A}_{\alpha\beta}\to\mathscr{A}_{\alpha\beta} + (y_\alpha\partial_\beta - y_\beta\partial_\alpha)\Lambda\,. \tag{7.11}$$

The unpleasant first order term in (7.2) cancels out and the substitution (7.10) gives the same result as $\partial_\alpha\to\partial_\alpha - iea_\alpha$, that is, precisely the change that can be cancelled by changing the phase of the field χ.

The modified action

$$\mathscr{L}_{\chi,a} = \frac{1}{2}\int dy\,\bar{\chi}\gamma_\alpha\gamma_\beta(M_{\alpha\beta} + e\mathscr{A}_{\alpha\beta})\chi \tag{7.12}$$

yields the current

$$j_a = \mathscr{J}_{\alpha\beta}\, y_\beta, \quad \mathscr{J}_{\alpha\beta} = \tfrac{1}{2}e\bar{\chi}[\gamma_\alpha,\gamma_\beta]\chi\,. \tag{7.13}$$

The conservation law (7.8) may be written

$$M_{\alpha\beta}\mathscr{J}_{\alpha\beta} = 0\,. \tag{7.14}$$

In fact, it should be written this way, so as to make explicit its intrinsic nature. When j has the form (7.13), then $y\cdot j$ vanishes everywhere, and (7.8) reduces to $\partial\cdot j = 0$. But the intrinsic nature of this equation is not explicit, since ∂_α is not defined intrinsically on the cone.

It is possible, and indeed reasonable, to develop the entire canonical formalism in terms of the intrinsic vector fields $M_{\alpha\beta}$, rather than the ∂_α. The skew tensor $\mathscr{J}$ then appears as the natural object, rather than the vector j. In particular, the conservation law (7.14) emerges directly. In other words, it seems quite natural to expect that the current j can always be expressed as in (7.13), in terms of an antisymmetric tensor field $\mathscr{J}$ that is conserved in the sense of (7.14). In that case one has

$$y\cdot j = 0, \quad \partial\cdot j = 0\,. \tag{7.15}$$

As was first shown by Dirac, the spinor field χ is not directly relevant. The fields χ and $\chi + \gamma\cdot y\lambda$ represent the same physical state. This is the origin of "current gauge transformations." The effect of the substitution $\chi\to\chi + \gamma\cdot y\lambda$ on the current (7.13) is a current type gauge transformation

$$j_\alpha \to j_\alpha + y_\alpha\Lambda\,,$$

with $\Lambda\propto\bar{\psi}\lambda$. This transformation must be physically unobservable. The effect on the interaction density $a\cdot j$ is to add a term $\Lambda\, y\cdot a$. This vanishes if the Lorentz condition holds, that is, when a_α is an external field. It will be seen below that, even though the field operator $y\cdot a$ does not vanish, it is in any case a free field. The density Λ couples to a free field only and has no observable effects. More will be said about this later.

Classical field equations

For classical field theory it is possible to maintain complete gauge invariance as well as manifest conformal invariance. The total action for photons and massless fermions is the sum of (7.6), with $c = 1$, and (7.12):

$$\int dy\,\{\tfrac{1}{2}a_\alpha\partial^2 a_\alpha + \tfrac{1}{2}ia_\alpha\partial^2 M_{\alpha\beta}a_\beta - \tfrac{1}{2}\bar{\chi}\gamma_\alpha\gamma_\beta(M_{\alpha\beta} + e\mathscr{A}_{\alpha\beta})\chi\}\,. \tag{7.16}$$

The tensor $\mathscr{A}$ was defined by (7.10). The coupled field equations are

$$\partial^2 a_\alpha + i\partial^2 M_{\alpha\beta}a_\beta = j_\alpha \equiv \tfrac{1}{2}e\bar{\chi}[\gamma_\alpha,\gamma\cdot y]\chi\,, \tag{7.17}$$

$$\gamma_\alpha\gamma_\beta(M_{\alpha\beta} + e\mathscr{A}_{\alpha\beta})\chi = 0\,. \tag{7.18}$$

The current satisfies (7.15). The gauge may be "fixed" by imposing the Lorentz condition $y\cdot a = 0$, without compromising manifest conformal invariance.

Quantization

Let $[a_\alpha^{(i)}]$, $i = 1,2,\cdots$, be a set of solutions of the free wave equation $\partial^2 a_\alpha = 0$, with positive energy, such that—compare (5.14)—

$$D_{\alpha\beta}(y,y') = \delta_{\alpha\beta}(y\cdot y')^{-1} = \sum_i a_\alpha^{(i)}(y)\,\overline{a_\beta^{(i)}(y)}\,. \tag{7.19}$$

With the help of these solutions, define the quantum field operator

$$\hat{a}_\alpha(y) = \sum_i \left(a_\alpha^{(i)}(y)\hat{b}_i + \overline{a_\alpha^{(i)}(y)}\hat{b}_i^*\right), \tag{7.20}$$

where $\hat{b}_i$ and $\hat{b}_i^*$ are destruction and creation operators acting on a Fock space. There is a unique vacuum state $|0\rangle$, annihilated by $\hat{b}_i$, and the repeated action of the $\hat{b}_i^*$ builds up the whole space. The following commutation relations are postulated, $[\hat{b}_i,\hat{b}_j] = 0 = [\hat{b}_i^*,\hat{b}_j^*]$ and

$$[\hat{b}_i,\hat{b}_j^*] = \delta_{ij}. \tag{7.21}$$

With these definitions one has

$$D_{\alpha\beta}(y,y') = \langle 0|\hat{a}_\alpha(y)\hat{a}_\beta(y')|0\rangle. \tag{7.22}$$

This is the Gupta–Bleuler quantization paradigm.

Quantum field theory

From now on, the field under discussion is the quantum field operator (7.20), but for convenience this operator will be denoted $a_\alpha(x)$ without the caret. The action (7.7) leads to

$$\partial^2 a_\alpha = j_\alpha, \tag{7.23}$$

where $j_\alpha(x)$ also stands for a quantum field operator from now on. The conservation law for j, in the weak form (7.8), yields

$$\partial^2\partial^2\, y\cdot a = 0. \tag{7.24}$$

The field $y\cdot a$ that, according to the discussion in Sec. VI, describes the scalar photons is thus a free field. (As will be seen in Sec. VIII, $y\cdot a$ is a dipole ghost. What is important is the fact that the wave equation for $y\cdot a$ is unaffected by the interaction.) This information expresses the decoupling of the scalar modes and is equivalent to what was learned about the propagator K_q^- in the same section [absence of $\psi_s\psi_s^*$ terms in (6.15)]. Recall that $y\cdot a = 0$ is the "Lorentz condition"; the quantum field operator does not satisfy this condition, but it does satisfy (7.24), in strict analogy with the equation $\Box\partial\cdot A = 0$ in ordinary QED. Equation (7.24) shows that the scalar photons have no interactions.

No attempt will be made at this point to define an S matrix or any other observables; this will be done after the theory has been translated into Minkowski space notation.

VIII. CONFORMAL QED IN MINKOWSKI SPACE-TIME

Resumé

Our construction of conformal QED in the manifestly covariant formalism may be briefly summarized as follows. We restrict ourselves to the simplest (Feynman) gauge. The Gupta–Bleuler quantization procedure (as generalized) led to the wave equation

$$\partial^2 a_\alpha = j_\alpha, \tag{8.1}$$

where a_α has degree of homogeneity -1. Gauge invariance leads to the conservation law

$$\partial_\alpha j_\alpha + \tfrac{1}{2}\partial^2 y_\alpha j_\alpha = 0. \tag{8.2}$$

The Lorentz condition that eliminates the scalar photons and thus projects onto the subspace of transverse and longitudinal photons is

$$y_\alpha a_\alpha = 0. \tag{8.3}$$

The longitudinal photon fields (gauge fields) are of the form

$$a_\alpha = \mathrm{Grad}_\alpha\, \Lambda = y_\alpha\partial^2\Lambda - 2\partial_\alpha\Lambda, \tag{8.4}$$

with $\partial^2\partial^2\Lambda = 0$. Finally, one must require invariance under the current gauge transformation

$$j_\alpha \to j_\alpha + y_\alpha\Lambda. \tag{8.5}$$

We must now rewrite all this in the conventional, Minkowski notation.

Minkowski coordinates

Let $x_+ = y_4 + y_5$, and introduce the inhomogeneous coordinates

$$x_\mu = y_\mu/x_+, \quad B = y^2/x_+^2, \quad \mu = 0,1,2,3. \tag{8.6}$$

On the projective cone $B = 0$ and $(x_\mu,x_+) \simeq (x_\mu,\lambda x_+)$ for $\lambda \neq 0$; this is a compactification of Minkowski space, coordinated by (x_μ). The electromagnetic potential is the 1-form

$$a_\alpha\, dy_\alpha = A_\mu\, dx_\mu + (A_+/x_+)\, dx_+ + A_B\, dB. \tag{8.7}$$

(One replaces z_α of Sec. IV by dy_α.) The fields A_μ, A_+, A_B are independent of x_+. Explicitly,

$$a_\mu = x_+^{-1}(A_\mu + 2x_\mu A_B),$$
$$A_\mu = x_+[a_\mu - x_\mu(a_4 + a_5)], \tag{8.8}$$
$$a_4 = -x_+^{-1}[-x\cdot A + A_+ - (1+x^2)A_B],$$
$$A_B = \tfrac{1}{2}x_+(a_4 + a_5), \tag{8.9}$$
$$a_5 = x_+^{-1}[-x\cdot A + A_+ + (1-x^2)A_B],$$
$$A_+ = y\cdot a. \tag{8.10}$$

For the currents one has similar relations, except that the factor x_+^{-1} is replaced by x_+^{-3}.

Current conservation

The conservation law (8.2) takes the form

$$\Box J_+ + 2\partial\cdot J = 0. \tag{8.11}$$

This is an ordinary conservation law:

$$\partial\cdot J' = 0, \quad J'_\mu \equiv J_\mu + \tfrac{1}{2}\partial_\mu J_+. \tag{8.12}$$

The current gauge transformation (8.5) affects only J_B, so this component of the current must effectively decouple.

Lorentz condition

The Lorentz condition (8.3) becomes

$$A_+ = 0 \quad \text{(conformal Lorentz condition)}, \tag{8.13}$$

and (A_μ,A_+,A_B) is a gauge field if, in addition,

$$A_\mu = \partial_\mu\Lambda, \quad A_B = -\tfrac{1}{4}\Box\Lambda, \tag{8.14}$$

with $\Box^2\Lambda = 0$. Let us now study the action and the field equations.

Field equations

The action (7.7) takes the form

$$\int d^4x \left(\tfrac{1}{2}A_\mu \Box A_\mu - 4A_B \partial\cdot A + 2A_+ \Box A_B - 8A_B A_B - A\cdot J - 2A_+ J_B - 2A_B J_+\right). \quad (8.15)$$

The field equations are thus

$$\Box A_\mu + 4\partial_\mu A_B = J_\mu\,, \quad (8.16)$$

$$\Box A_+ - 2\partial\cdot A - 8A_B = J_+\,, \quad (8.17)$$

$$\Box A_B = J_B\,. \quad (8.18)$$

By setting $J_\mu = J_B = J_+ = A_+ = 0$, one obtains the conformally invariant spin one equations of Mayer and Bayen and Flato.[26] The necessity of the current J_B becomes apparent as soon as one tries to reintroduce a source J_μ to (8.16); one finds that the resulting system of equations will not be conformally invariant unless one also returns J_B to (8.18). To find the consequences of current conservation, Eq. (8.11) or (8.12), take the divergence of (8.16) and combine with (8.17) to find that

$$\Box^2 A_+ = 0\,. \quad (8.19)$$

The scalar photons, characterized by $A_+ \neq 0$, are thus dipole ghosts, as are the gauge photons of Eq. (8.14).

The Lorentz condition (8.13) is not satisfied by the quantum field operator, but one can choose initial conditions such that (see *Notes added in proof*)

$$A_+ |\Psi\rangle = 0\,, \quad (8.20)$$

which implies that

$$\Phi |\Psi\rangle = 0, \quad \Phi \equiv \Box A_+ = 2\partial\cdot A + 8A_B + J_+\,. \quad (8.21)$$

Since Φ is a free field, $\Phi|\psi\rangle$ remains zero at all times; hence $\Box A_+ |\Psi\rangle$ vanishes, and the condition $A_+|\Psi\rangle = 0$ is also preserved at all times. These conditions are conformally invariant, since (8.20) is the same as $y_\alpha a_\alpha |\Psi\rangle = 0$; and Eq. (8.21) is the same as $y_\alpha(\partial^2 a_\alpha - j_\alpha)|\Psi\rangle = 0$.

Eliminating A_B from (8.16)–(8.18), one gets

$$\Box A_\mu - \partial_\mu \partial\cdot A = J'_\mu \equiv J_\mu - \tfrac{1}{2}\partial_\mu \Phi\,, \quad (8.22)$$

$$\Box \partial\cdot A = \tfrac{1}{2}\Box J_+ - 4J_B, \quad \Box\Phi = 0\,. \quad (8.23)$$

Recall that J_B is the pure current type gauge field. Setting $J_B = 0$ violates conformal invariance unless J_+ and J_μ vanish as well. To show that J_B is, nevertheless, irrelevant, it is enough to define a new vector potential $A'_\mu \equiv A_\mu + 4\Box^{-1}\partial_\mu A_B$; then J_B disappears from (8.23), and the action takes the form

$$\int d^4x \left(-\tfrac{1}{2}A'_\mu \Box A'_\mu + A'\cdot J' + 4A_B \Box^{-1}\partial\cdot J' + 2A_+ J_B - 2A_+ \Box A_B\right). \quad (8.24)$$

Variation of A_B gives $\Box A_+ = 2\Box^{-1}\partial\cdot J'$. The ambiguity inherent in the choice of inverse of $\Box$ is resolved by the boundary condition $A_+|\Psi\rangle = 0$. One sees that the unfamiliar fields decouple completely, leaving ordinary QED.

Conformal transformations

Equations (8.1)–(8.3) are invariant under the SO(4,2) transformations

$$T_\Lambda : a \to a', \quad a'_\alpha(y) = \Lambda_{\alpha\beta} a_\beta(y\Lambda), \quad \Lambda \in \mathrm{SO}(4,2)\,. \quad (8.25)$$

This transformation law can be expressed in terms of the Minkowski fields A_μ, A_B, and A_+, with the help of (8.6)–(8.10) and the usual rule

$$A'_\alpha(x') = \frac{\partial x_\beta}{\partial x'_\alpha} A_\beta(x)\,. \quad (8.26)$$

One then finds that under an infinitesimal dilatation, $1 + i\epsilon L_{56}$

$$\delta A_\mu(x) = \epsilon \hat{D} A_\mu(x) = -\epsilon(x\cdot\partial + 1) A_\mu(x)\,, \quad (8.27)$$

$$\delta A_B(x) = \epsilon \hat{D} A_B(x) = -\epsilon(x\cdot\partial + 2) A_B(x)\,, \quad (8.28)$$

$$\delta A_+(x) = \epsilon \hat{D} A_+(x) = -\epsilon(x\cdot\partial) A_+(x)\,, \quad (8.29)$$

which indicates that the conformal degrees of A_μ, A_B, and A_+ are, respectively, -1, -2, and 0. Under an infinitesimal special conformal tranformation $1 + i\epsilon^\mu(L_{5\mu} - L_{6\mu})$

$$\delta A_\mu(x) = \{[x^2(\epsilon\cdot\partial) + 2(\epsilon\cdot x)\hat{D}]\delta_\mu{}^\nu + 2(x_\mu \epsilon^\nu - \epsilon_\mu x^\nu)\} A_\nu(x) + 2\epsilon_\mu A_+(x)\,, \quad (8.30)$$

$$\delta A_B(x) = [x^2(\epsilon\cdot\partial) + 2(\epsilon\cdot x)\hat{D}] A_B(x) - \epsilon^\mu A_\mu(x)\,, \quad (8.31)$$

$$\delta A_+(x) = [x^2(\epsilon\cdot\partial) + 2(\epsilon\cdot x)\hat{D}] A_+(x)\,. \quad (8.32)$$

It is apparent from (8.27)–(8.32) that, while A_+ itself transforms homogeneously under the action of the conformal group, it also leaks into A_μ and that A_μ, in turn, leaks into A_B. The indecomposable nature of the conformal vector field is thus manifest in this coordinate system. The invariance of the "Lorentz condition" $A_+ = 0$ is also evident. The currents J_μ, J_B, and J_+ transform similarly except that their conformal degrees are, respectively, -3, -4, and -2. One can easily check the invariance of (8.15)–(8.18)

There is one point that needs elaboration. Previous formulations of conformal electrodynamics have imposed the conditions $y\cdot a$ on the field and $y\cdot j$ on the current. With these restrictions the field equations and the transformations simplify, since A_+ and J_+ vanish. It is perhaps worthwhile to repeat at this point the reason why A_+ cannot vanish in a conformal quantum field theory. The quantum field operator was defined by (7.20). [The caret on $\hat{a}_\alpha(y)$ was subsequently dropped.] In the sum (7.20) there is a contribution from the scalar photons; without such field modes it is impossible to construct a covariant propagator. What does not emerge clearly in (7.21), but what is evident in (6.3), for example, is that the scalar photons are canonically conjugate to the longitudinal photons. Conformal invariance requires that the "longitudinal modes" be quantized along with the transverse modes, but this is impossible without the help of the scalar modes.

Propagator

The definition (7.20) of the quantum field operator a_α, in terms of wave functions and a set of creation and destruction operators, can, of course, be rewritten in terms of the new notation introduced by (8.8)–(8.10). Equations (7.19) and (7.22) leads to the following homogeneous propagators for (A_μ, A_+, A_B):

$$D_{\mu\nu}(x,x') = \langle 0|\, A_\mu(x) A_\nu(x')|0\rangle = \delta_{\mu\nu} D(x,x')$$
$$\equiv -2\delta_{\mu\nu}/(x-x')^2\,, \tag{8.33}$$

and similarly

$$D_{\mu +}(x,x') = (x'_\mu - x_\mu) D(x,x')\,, \tag{8.34}$$
$$D_{+B}(x,x') = \tfrac{1}{2} D(x,x')\,, \tag{8.35}$$
$$D_{\mu B} = D_{BB} = 0, \quad D_{++} = 1\,. \tag{8.36}$$

The fact that $D_{++} = 1$ reveals the presence of the zero energy mode; that is, the trivial representation $D(0,0,0)$ in (6.16).

Representations

It remains only to clarify the group representations. The restrictions to the Poincaré group are conjectured to be as follows (see *Notes added in proof*):

$$D_6|\mathscr{P} = D_1 \to D_4 \to D_1\,, \tag{8.37}$$
$$D(1,0,0)|\mathscr{P} = D(0,0)\,, \tag{8.38}$$
$$D(1,1/2,1/2)|\mathscr{P} = D(0,0) \to D(0,0)\,, \tag{8.39}$$
$$D(2,1,0)|\mathscr{P} = D(0,1)\,, \tag{8.40}$$
$$D(2,0,1)|\mathscr{P} = D(0,-1)\,. \tag{8.41}$$

The augmented triplet (6.16) becomes

$$[D(0,0) \to D(0,0)] \xrightarrow{[D(0,1)\oplus D(1,0)]} [D(0,0) \to D(0,0)]$$
$$\xrightarrow{D_1}$$

$$\simeq D(0,0) \otimes [D_1 \to D_4 \to D_1]$$
$$\simeq D(0,0) \to [D(0,0) \otimes D_4] \to D(0,0)\,. \tag{8.42}$$

The equivalences expressed here are algebraic equivalences. The last expression is especially illuminating. The center bracket is the usual triplet (2.5). The extra scalar field on the left is Φ, and the extra gauge field on the right is $\Box A_B$. These extra fields distinguish conformal QED. They play no actual role in the evaluation of the S matrix (except perhaps in the renormalization program), but they participate in the construction of the indefinite metric Fock space. The larger Fock space is required in order that the conformal group act on it.

IX. FINAL REMARKS

1. We have formulated an action principle and quantization rules for a conformally invariant field theory, but no reference was made to the possibility of a physical interpretation. No conformal field theory is known that has a direct physical application. Masses can and must be introduced by hand, but this is not the only source of breakdown of conformal invariance.

To have a physical interpretation, one must decide what are the observables. This is very difficult to do within a context in which all the particles are massless, but one may attempt to begin by discussing scattering experiments. This is beset with well-known difficulties, but it may be possible in perturbation theory at least. The boundary conditions appropriate to scattering are incompatible with the structure of the projective cone, for this space has no global causal structure[27] and no "infinity." Infinity in the sense of Minkowski space is introduced artificially through the choice of coordinates. Massless particles interacting with other massless particles in a conformally invariant field theory do not know where infinity is and therefore they do not stop interacting with each other as they become infinitely separated (in Minkowski coordinate sense). The introduction of a mass term $m\bar{\psi}\psi$ in the Minkowski space action is equivalent to adding a term of the form $(y_4 + y_5)^{-1} m\bar{\psi}\psi$ to the conformal action. A massive particle therefore knows where infinity is—at $y_4 + y_5 = 0$—and the stage is set for describing scattering. Whether masses are introduced or not, it is necessary to choose the location of Minkowski infinity on the projective cone, before the conventional postulates of scattering theory make any sense.[28]

Conventionally, the way this is done, is to define a perturbative S matrix through the imposition of boundary conditions of the inhomogeneous propagator. The choice

$$D^F(x,x') = -2i[(x-x'^2) + i\epsilon]^{-1} \tag{9.1}$$

exploits the global, causal (but not conformally invariant) structure of Minkowski space. At first sight, this expression may seem to be conformally invariant. Indeed, if (9.1) is introduced into (8.26)–(8.29), in place of the homogeneous propagator $D(x,x')$, and the result rewritten in terms of the six-dimensional coordinates, then one obtains the tensor operator

$$D^F{}_{\alpha\beta}(y,y') = i\delta_{\alpha\beta}(y\cdot y' + i\epsilon)^{-1}\,.$$

The problem is that this is not a distribution over the projective cone. (Recall that the projective cone is defined by the projection $y_\alpha \simeq \lambda y_\alpha, \lambda \neq 0$. Distributions on R^6 are interpretable as distributions on the projective cone only if they are homogeneous. This problem could be solved by replacing the projective cone by its double covering defined by $y_\alpha \simeq \lambda y_\alpha, \lambda > 0$.[29])

2. The fact that conformal invariance is broken, by the causal structure and by the introduction of masses, does not mean that potential benefits of the conformal structure is lost. On the contrary, the lesson learned by the application of soft symmetry breaking to nonabelian gauge theories is that the partial preservation of symmetry by Ward–Takahashi identities is of great utility in renormalization. In order to determine whether conformal invariance (though broken) is especially beneficial, it was necessary first to formulate a conformally invariant field theory. We have seen that this leads to a richer ghost structure than that of conventional QED, but we have not yet examined the effect that this may have on the removal of infrared and ultraviolet divergences. We suggest that it may be useful to do so, and that conformally invariant nonabelian gauge theories should also be studied in this manner. Note in this connection the appearance of the zero-energy ghost, and the related result $\langle 0|\, A_+(x) A_+(x')|0\rangle = 1$; does this have anything to do with the Higgs–Kibble field and its nonvanishing vacuum expectation value? The reduction to Minkowski notation was done in the simplest gauge only: $c = 0$ in Eq. (7.6). In any other gauge the field equations contain third order derivatives. It seems unlikely that this improves the ultraviolet behavior, but it is worth looking into. See in this connection Ref. 30.

3. The techniques developed in this paper will be applied to a study of conformal gravity (see *Notes added in proof*).

4. The construction of unitary representations of semisimple groups on quotient spaces is being rapidly developed by mathematicians; see especially papers by Blattner and Rawnsley,[31] Schmid and Wolf,[32] and Rawnsley, Schmid, and Wolf,[33] and Schmid's Berlin address.[34]

Notes added in proof: 1. Equation (8.20) and similar subsequent equations are stated incorrectly. It is the annihilating part of A_+ that kills the physical state $|\Psi\rangle$.

2. The ideas of this paper have been applied to de Sitter QED (to appear in Annals of Physics) and to linear conformal quantum gravity [Phys. Rev. D **27**, 2249 (1983)].

ACKNOWLEDGMENTS

We thank Moshe Flato for many stimulating discussions and for collaboration in developing the main ideas of this paper. We are indebted to Jean-Pierre Labesse for an inspiring lecture on Gupta–Bleuler quantization. We thank Joseph Wolf for interesting discussions and for information about the work of Enright, Howe, and Wallach.[35] C. F. and B. B. thank the Research Committee of the Academic Senate of UCLA, and the National Science Foundation, for support, W. H. thanks Lutz Castell for discussions and DAAD for support.

APPENDIX A: INDECOMPOSABLE MODULES

Let us first recall some basic facts concerning minimal weight representations of semisimple groups. Our interest in minimal weight representations stems from the physical requirement of a lower bound on the energy spectrum (without which a vacuum state cannot be defined). We shall consider here only minimal weight representations which are K-finite, that is:

Definition: A representation with minimal weight λ is said to be *K-finite* if in its associated weight diagram the multiplicity of each weight is finite and if only finite-dimensional irreducible representations of the maximal compact subgroup occur. We shall also say that a weight λ is K-finite if it is the minimal weight of a K-finite representation.

Now if g is a semisimple Lie algebra of rank l, then there exist l independent Casimir operators in the enveloping algebra of g. If D is a finite-dimensional representation of g, then the values of these Casimirs on D can be used to uniquely characterize the representation. However, if D is a minimal weight representation (i.e., bounded only from below), then it is possible for other (inequivalent) minimal weight representations of g to have the same values for the Casimir operators. If this is the case, then the minimal weights of these representations must be related by Weyl reflection. More precisely, two inequivalent irreducible representations D_λ and D_μ with minimal weights λ and μ, respectively, possess the same values for the Casimir operators if and only if

$$\lambda - \rho = w(\mu - \rho), \tag{A1}$$

where w is an element of the Weyl group of g and ρ is one-half the sum of the negative roots of g. In such a case, both the minimal weights and the irreducible representations themselves are said to be *Weyl-equivalent*.

Finally, we point out that a precondition for the occurrence of a nondecomposable representation

$$D_\lambda \to D_\mu$$

is the equality of the values of the Casimir operators on D_λ and D_μ. For if a Casimir operator C had different values in D_λ and in D_μ, then the operator C in $D_\lambda \to D_\mu$ could be diagonalized and the representation then decomposed via Schur's lemma.[24]

The Cartan–Weyl decomposition of so(4,2) by the Cartan subalgebra spanned by (3.3) yields 12 nonzero roots

$$(\pm 1, \pm \tfrac{1}{2}, \pm \tfrac{1}{2}), \quad (0, \pm 1, 0), \quad (0, 0, \pm 1). \tag{A2}$$

Using the ordering described in Sec. III, we compute ρ to be

$$\rho = (2, -\tfrac{1}{2}, -\tfrac{1}{2}). \tag{A3}$$

The 12 roots generate six independent Weyl reflections through planes through the point ρ in weight space

$$\begin{aligned}
&w(1, -\tfrac{1}{2}, -\tfrac{1}{2})(E j_1 j_2) = (j_1 + j_2 + 3, \tfrac{1}{2}(E + j_1 - j_2 - 2) - \tfrac{1}{2}, \tfrac{1}{2}(E - j_1 + j_2 - 2) - \tfrac{1}{2}),\\
&w(1, -\tfrac{1}{2}, \tfrac{1}{2})(E, j_1, j_2) = (j_1 - j_2 + 2, \tfrac{1}{2}(E + j_1 + j_2 - 1) - \tfrac{1}{2}, \tfrac{1}{2}(E - j_1 - j_2 - 3) - \tfrac{1}{2}),\\
&w(1, \tfrac{1}{2}, -\tfrac{1}{2})(E, j_1, j_2) = (j_2 - j_1 + 2, \tfrac{1}{2}(E - j_1 - j_2 - 3) - \tfrac{1}{2}, \tfrac{1}{2}(E + j_1 + j_2 - 1) - \tfrac{1}{2}),\\
&w(1, \tfrac{1}{2}, \tfrac{1}{2})(E, j_1, j_2) = (1 - j_1 - j_2, \tfrac{1}{2}(E - j_1 + j_2 - 2) - \tfrac{1}{2}, \tfrac{1}{2}(E + j_1 - j_2 - 2) - \tfrac{1}{2}),\\
&w(0,1,0)(E, j_1 j_2) = (E, -1 - j_1, j_2),\\
&w(0,0,1)(E, j_1, j_2) = (E, j_1, -1 - j_2).
\end{aligned} \tag{A4}$$

When compounded, these six reflections produce a lattice which may contain as many as 24 different weights.

Yet only a few of these Weyl equivalent weights are K-finite. For a minimal weight representation $D(E_0, j_1 j_2)$ is not K-fintte unless both $2j_1$ and $2j_2$ are nonnegative integers.

The symmetry of the lattice generated by the six reflections allows one to infer that if a weight $(a + 2, b - \tfrac{1}{2}, c - \tfrac{1}{2})$ occurs, then so do all those weights obtained from $(a + 2, b - \tfrac{1}{2}, c - \tfrac{1}{2})$ when one replaces a by $-a$, b by $-b$, and/or c by $-c$. We can thus draw out of the six simple

reflections the following transformations:

$$w_0(E,j_1,j_2) = (E,j_1,j_2)\,, \tag{A5}$$

$$w_1(E,j_1,j_2) = (j_1+j_2+3,\tfrac{1}{2}|E+j_1-j_2-2|-\tfrac{1}{2},\ \tfrac{1}{2}|E-j_1+j_2-2|-\tfrac{1}{2})\,, \tag{A6}$$

$$w_2(E,j_1,j_2) = (j_1-j_2+2,\tfrac{1}{2}|E+j_1+j_2-1|+\tfrac{1}{2},\ \tfrac{1}{2}|E-j_1-j_2-3|+\tfrac{1}{2})\,, \tag{A7}$$

$$w_3(E,j_1,j_2) = (j_2-j_1+2,\tfrac{1}{2}|E-j_1-j_2-3|-\tfrac{1}{2},\ \tfrac{1}{2}|E+j_1+j_2-1|-\tfrac{1}{2})\,, \tag{A8}$$

$$w_4(E,j_1,j_2) = (1-j_1-j_2,\tfrac{1}{2}|E-j_1+j_2-2|-\tfrac{1}{2},\ \tfrac{1}{2}|E+j_1-j_2-2|-\tfrac{1}{2})\,, \tag{A9}$$

$$w_5(E,j_1,j_2) = (4-E,j_1j_2)\,. \tag{A10}$$

These new weights will be K-finite so long as the expressions within the absolute value signs are integral and non-vanishing. In fact, these weights account for *all* the Weyl equivalents of (E,j_1,j_2) that are K-finite.

Let $R\,(E_0,j_1,j_2)$ be any indecomposable or irreducible K-finite representation of so(4,2) with minimal multiplicity-free weight (E_0,j_1,j_2) ($E_0\in R$; $2j_1$ and $2j_2$ nonnegative integers). From Eqs. (A5)–(A10) we see that if

$$E_0 > j_1+j_2+2\,, \tag{A11}$$

then there is no K-finite Weyl equivalent of (E,j_1,j_2) that lies within the weight diagram of $R\,(E_0,j_1,j_2)$. But this implies that there can be no indecomposable structure within $R\,(E_0,j_1,j_2)$ and so it must be irreducible. This condition is also sufficient for unitarity, as will now be demonstrated.

Consider a continuous family $\mathscr{J}(j_1,j_2)$ of indecomposable or irreducible representations

$$\mathscr{J}(j_1,j_2) = \{R\,(E_0,j_1,j_2);E_0\in R\}\,.$$

Let L be any element of the complex extension of so(4,2) with the property that

$$[L_{05},L] = L\,.$$

Then there are complex numbers a_k, $k=1,2,3,4$, such that $L = \sum a_k(L_{k5}-iL_{k0})$. If the a_k are real, then in a unitary representation $L^+ = \sum a_k(L_{k5}+iL_{k0})$. The operator L, acting on an eigenvector of L_{05}, increases the eigenvalue by one and L^+ decreases it by 1. If the operator L^+L is positive definite for one set of (real) a_k, then it is positive definite for all such sets [because the co-adjoint action of so(4) on $a_1,\ldots,a_4$ is irreducible]. Now $R\,(E_0,j_1,j_2)$ is unitary if E_0 is large enough, and L^+L is positive definite. As the parameter E_0 is decreased, there comes a critical point at which L^+L becomes singular. For that value of E_0, $R\,(E_0,j_1,j_2)$ must be indecomposable; as the singularity of L^+L implies that there exists a state above (E_0,j_1,j_2) which cannot be returned to (E_0,j_1,j_2). Thus two subrepresentations result, characterized by unequal but Weyl equivalent minimal weights $\alpha=(E_0,j_1,j_2)$ and α' with $\alpha'>\alpha$. But this cannot happen if (A11) holds; therefore, (A11) is a sufficient condition for unitary. In fact, it is also necessary if $j_1j_2\neq 0$.[32,21] In the special case where either j_1 or j_2 is zero, then the condition (A11) may be extended to

$$E_0 > j_1+j_2+1 \tag{A12}$$

since the relevant transformation (A16) will not produce a K-finite weight above $\alpha=(E_0,j_1,j_2)$ until $E_0=j_1+j_2+1$ (see Fig. 1).

Now what happens to $R\,(E_0,j',j_2)$ if we allow E_0 to fall below the limit of (A11) [or (A12)]? From the above argument we infer that L^+L will become indefinite and thus the subrepresentation $D\,(E_0,j_1j_2)$ will be nonunitary. It is apparent from (A6) that a Weyl equivalent weight will continue to reside above (E_0,j_1,j_2); however, this weight will not be K-finite unless E_0 is integral. But when

$$(E_0,j_1,j_2) = (j_1+j_2+3-n,j_1j_2),\quad n=1,2,\ldots,$$

the transformation (A6) yields

$$(j_1+j_2+3,j_1-n/2;j_2-n/2)\,,$$

which *is* K-finite so long as $2j_1$ and $2j_2$ are greater than or equal to n. Thus indecomposable representations of the form

$$D\,(j_1+j_2+3-n,j_1,j_2)\rightleftarrows D\,(j_1+j_2+3,j_1-n/2,j_2-n/2)$$

are possible. Indeed, when $j_1=j_2=s/2$, such repesentations may be connected with "current type gauge fields of order n" (see Fig. 1).

From (A7) and (A8) we see that another Weyl-equivalent weight will appear above (E_0,j_1,j_2) when E_0 reaches

$$E_0 = |j_1-j_2|+1\,, \tag{A13}$$

namely,

$$(j_1-j_2+2,j_1-\tfrac{1}{2},j_2-\tfrac{1}{2})\,,$$

where we have assumed without loss of generality that $j_1>j_2$. If $j_1=j_2=s/2>0$, then the three transformations (A7), (A8), and (A10) all come into play when E_0 reaches the value (A13). These transformations together with (A6) put a total of four K-finite Weyl equivalent weights

$$(s+3,0,0)\,,$$

$$(3,s/2,s/2)\,,$$

$$(2,(s/2)+(1/2),(s/2)-(1/2))\,,$$

$$(2,(s/2)-(1/2),(s/2)+(1/2))$$

above (1,s/2,s/2). Among the many possibilities for indecomposable representations with these minimal weights is

$$[D\,(2,(s/2)+\tfrac{1}{2},(s/2)-\tfrac{1}{2})\oplus D\,(2,(s/2)-\tfrac{1}{2},(s/2+\tfrac{1}{2})]$$
$$\rightarrow D\,(1,s/2,s/2)\,,$$

which describes "first-order gradient type gauge fields."

If $j_1=j_2=0$, then the transformations (A7) and (A8) do not produce K-finite equivalent weights above (E_0,j_1,j_2) until E_0 reaches the value 0. But here (A9) also comes into play and so one obtains the following complete set of K-finite Weyl equivalent representations:

$$D\,(4,0,0)\,,$$

$$D\,(3,\tfrac{1}{2},\tfrac{1}{2})\,,$$

$$D\,(2,1,0),\quad D\,(2,0,1)$$

$$D\,(1,\tfrac{1}{2},\tfrac{1}{2})\,,$$

$$D\,(0,0,0)\,.$$

This particular case is important since it exposes the maximal structure of an indecomposable representation involving the physical photon $D\,(2,1,0)\oplus D\,(2,0,1)$ (see Fig. 2).

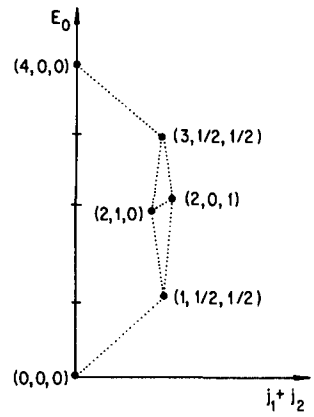


FIG. 2. K-finite weights Weyl equivalent to (2,1,0). In conformal QED, (0,0,0) is the zero energy ghost, (1,1/2,1/2) is the ground state of the longitudinal and scalar photons, (2,1,0) and (2,0,1) are the ground states of the transverse photons, and (3,1/2,1/2) and (4,0,0) do not appear.

APPENDIX B: PROPAGATORS

The constraint Tr $K = 0$, when applied to (5.1), reads

$$[y\cdot\partial_z + (1-d-2s)\,y\cdot\partial_z]\,K = 0\,, \tag{B1}$$

and gives the constraint

$$[(3-d)/2+b-s]^{-1}C_{ab} + (s-a-b)^{-1}C_{a,b+1} + a^{-1}C_{a-1,b+1} = 0\,. \tag{B2}$$

This fixed C_{ab} in terms of C_{ob}, the latter remaining arbitrary. The constraint Div $K = 0$ gives the recursion relation

$$A_{ab}C_{ab} + (N+1-a-b)(s+1-a-b)\times[C_{a,b-1} - C_{a-1,b}] + (a+1)^2C_{a+1,b-1} - (a+1)^2C_{a+1,b} = 0\,, \tag{B3}$$

$$A_{ab} \equiv (b-a)(N+s-2a-2b) + b(b+2a+d+1) - s\,. \tag{B4}$$

This is consistent with (B2) and determines C_{ab} in terms of $C_{00} = 1$. For $s = 1$ one obtains the coefficients shown in (5.5).

APPENDIX C: MODAL DECOMPOSITION

To find the decomposition of a propagator such as (5.5) one expands each inner product in a Fourier series, as in

$$2\,y\cdot z' = y_+z'_- - 2\,\mathbf{y}\cdot\mathbf{z}' + y_-z'_+\,.$$

The notation was explained in (4.18) and (5.3). The first term has the lowest energy: -1. In the case of nonintegral powers of $y\cdot y'$, the expansion (5.2) has to be used. In (5.5), the first factor has lowest energy $1-N$. The first two terms in the bracket have lowest energy -2, but there is a cancellation so that the combination has lowest energy -1. The third term has energy 0; hence K has a Fourier expansion in powers of $\exp(-it)$ with exponents $-N, -N+1, \cdots$. The first term is (up to a numerical factor)

$$(y_+y'_-)^{N-1}\boldsymbol{\eta}\cdot\boldsymbol{\eta}'\,. \tag{C1}$$

with $\eta = y_+\mathbf{z} - z_+\mathbf{y}$ and $\eta' = y'_-\mathbf{z}' - z'_-\mathbf{y}'$. Under the transformations of the compact subgroup SO($d+1$), the $d+1$ functions $y_+^{N+1}\boldsymbol{\eta}$ transform like the vector representation; $D(1)$ for $d=2$, $D(\frac{1}{2},\frac{1}{2})$ for $d=3$.

Consider (5.9) next. The interpretation of terms appearing in the expansion is simplified by choosing variables as follows. First express all inner products in terms of $y_+, y_-, \mathbf{y}$ and $z_+, z_-, \mathbf{z}$. Next eliminate z_- in favor of $y\cdot z$ and $\mathbf{z}$ in favor of $\boldsymbol{\eta} = y_+\mathbf{z} - z_+\mathbf{y}$. Then all dependence on z_+ cancels; this is because $y\cdot\partial_z\,K_q^+ = 0$. One finds in this way that

$$K_0^+ = 2^d\,(M_{pp} + M_{sg}\,y'\cdot z' + y\cdot z M_{gs} + y\cdot z\,M_{gg}\,y'\cdot z')\,, \tag{C2}$$

where M_{pp}, M_{sg}, M_{gs} are polynomials in $(y_+y'_-)^{-1}$ with coefficients that are polynomials in $y_-, y'_+, \mathbf{y}, \mathbf{y}', \boldsymbol{\eta}, \boldsymbol{\eta}'$. The terms of lowest energy are given by (if $q=0$)

$$M_{pp} = -(y_+^{-d-1}\boldsymbol{\eta})\cdot(y'^{-d-1}_+\boldsymbol{\eta}')^* + \cdots\,, \tag{C3}$$

$$M_{sg}\,y'\cdot z' = (y_+^{-d-2}\mathbf{y}\cdot\boldsymbol{\eta})(y'^{-d-1}_+\,y'\cdot z')^* + \cdots\,, \tag{C4}$$

where one recognizes the functions (6.5) and (6.6). The term M_{pp} contains the ground state, with energy $E_0 = d$. If $d=3$, then the minimal weight is $(3,\frac{1}{2},\frac{1}{2})$. The only weight that is equivalent to this one, and inside the weight lattice of K_0^+, is (4,0,0). The lower equivalent weights (2,1,0), (2,0,1), $(1,\frac{1}{2},\frac{1}{2})$, and (0,0,0) do not appear. This is related to the fact that K, and thus all the fields, satisfy one of the subsidiary conditions: Tr $\psi = y\cdot\partial_z\psi = 0$. When $d=2$, then the minimal weight is (2,1) and the only other relevant weight is (3,0). *Therefore, in both cases, one needs to investigate the levels $E_0 = d$ and $E_0 = d+1$ only.* The raising operators (for energy $= L_{05}$) are

$$L_i^+ = (y_-\partial_i + 2\,y_i\partial_+) + (z_-\partial_i + 2\,z_i\partial_+)\,. \tag{C5}$$

Application of L_i^+ to the physical ground states $y_+^{-d-1}\eta_i = p_{0i}$ yield $(d+1)^2$ states ψ_{ij}. Under the action of the compact subalgebra so($d+1$), they break up into three irreducible representations spanned by the traceless part of $\psi_{ij} + \psi_{ji}$, $\psi_{ij} - \psi_{ji}$, and $\Sigma\,\psi_{ii}$. The first two leak back to the ground state under the action of the lowering operators

$$L_i^- = (y_+\partial_i + 2\,y_i\partial_-) + (z_+\partial_i + 2\,z_i\partial_-)\,, \tag{C6}$$

but the third is a minimal weight vector: $L_i^-\,\Sigma\,\psi_{jj} = 0$. In fact, this function is the gauge field $y_+^{-d-1}y\cdot z = g_0$. The ground state $s_0 = y_+^{-d-2}\mathbf{y}\cdot\boldsymbol{\eta}$ of the scalar field appears as the coefficients of g_0^* in $M_{sg}\,y'\cdot z'$, Eq. (C4). It leaks to the absolute ground state and is a minimal weight vector in the quotient.

The propagator (5.13) may be analyzed in the same way. When $d=3$, the simplest choice of q is $q=1$, which leads to (5.14). Taking $q=0$, one finds an expansion like (C2),

$$K_0^- = N_{pp} + N_{sg} + N_{gs} + N_{gg}\,, \tag{C7}$$

where N_{sg} is a sum of terms of the form $\psi(\mathrm{y},\mathrm{z})\,z'\cdot\partial_{y'}\psi'(\mathrm{y}',\mathrm{z}')$, N_{gs} is a sum of terms like $\psi'(y',z')\,z\cdot\partial_y\psi(y,z)$, and N_{gg} can be expressed as $z\cdot\partial_y\,z'\cdot\partial_{y'}\Lambda(y,z,y',z')$. The lowest energies are given by

$$N_{pp} = -[y_+^{-2}(z_iy_j - z_iy_i)]\times[y'^{-2}_+(z'_jy'_j - z'_jy'_i)]^* + \cdots\,, \tag{C8}$$

$$N_{sg} = -2[y_+^{-2}y_iz_+][y'^{-2}_+\eta'_i]^* + \cdots\,, \tag{C9}$$

$$N_{gg} = [y_+^{-1}z_+][y_+^{-1}z'_-]^* + \cdots\,. \tag{C10}$$

When $d=3$, the equivalent weights are (0,0,0), $(1,\frac{1}{2},\frac{1}{2})$, [2,1,0], (2,0,1), $(3,\frac{1}{2},\frac{1}{2})$, and (4,0,0). The last two do not generate invariant subspaces or subquotients, as is easily seen in the gauge (5.14). As was pointed out in Sec. VI, after (6.17), the total representation is equivalent to $D_6 \otimes D(1,0,0)$, and

this product does not contain $D(3,\frac{1}{2},\frac{1}{2})$ or $D(4,0,0)$. When $d=2$, the equivalent weights are (0,0), (1,1), (2,1), and (3,0). The direct product $D_6 \otimes D(1,0)$ does not contain $D(3,0)$, so that the weight space (3,0) need not be examined. Thus, in either case, one must investigate the energies 0, 1, and 2. Only N_{gg} has $E=0$ terms. In N_{sg} the lowest energy is 1. The $E=2$ space is very complicated, with contributions from N_{sg}, N_{gs}, and N_{gg}, and the remainder shown by (C8). The detailed calculation will be omitted.

APPENDIX D: INTEGRALS

In R^{d+3} introduce coordinates (y_α), $\alpha = 0,1,\ldots,d+1,5$, and the pseudo-Euclidean metric δ defined by $\delta(y) = y^2 = y_0^2 + y_5^2 - y_1^2 - \cdots - y_{d+1}^2$, as in Sec. IV. On the cone $y^2=0$ identify any pair y,y' of points if there is a positive c such that $y_\alpha = cy'_\alpha$, $\alpha = 0,\ldots,5$. The resulting projective space is $S_1 \times S_d$. Coordinates t for S_1 and $\hat{y}$ for S_d were introduced in (5.3). By an integral over the (projective) cone one means an integral over $S_1 \times S_d$, with the standard volume element $dt\, d\hat{y}$ that is invariant under the induced action of the compact subgroup SO(2) ⊗ SO(4). Let L denote a scalar field on the cone, homogeneous of degree N, and let a (generally multivalued) function $\tilde{L}$ on $S_1 \otimes S_d$ be defined by $L(y) = Y^N L(t,\hat{y})$, in the notation of Eq. (5.3). Then the following is elementary and well known:

Proposition: The integral

$$\int dy\, L(y) \equiv \int_{S_1 \times S_d} dt\, d\hat{y}\, \tilde{L}(t,\hat{y}) \tag{D1}$$

is invariant under the transformations of $SO(d+1,2)$ if and only if $L = Y^N \tilde{L}$ is a scalar field and $N = -d-1$.

Proof: Let (L^σ), with σ real, be a family of homogeneous scalar fields on R^{d+3}, with

$$\hat{N} L^\sigma = (i\sigma - d - 1)L, \quad \hat{N} = y\cdot\partial_y \tag{D2}$$

and define

$$F(y) = \int_R d\sigma\, L^\sigma(y)\,. \tag{D3}$$

Now consider the invariant integral

$$\int_{R^{d+3}} [dy]\, \delta(y^2) F(y) \propto \int_{S_1 \times S_d} dt\, d\hat{y} \int_\infty^0 Y^d\, dY \int_R d\sigma\, L^\sigma(y)\,,$$

where $[dy]$ is the Lebesque measure. Set $L^\sigma(y) = Y^{i\sigma - d - 1} \tilde{L}^\sigma(t,\hat{y})$ and $Y = e^x$; then it becomes

$$\int_{S_1 \times S_2} dt\, d\hat{y} \int_{R_2} dx\, d\sigma\, e^{i\sigma x} \tilde{L}^\sigma(t,\hat{y})$$

$$\propto \int_{S_1 \times S_d} dt\, d\hat{y}\, \tilde{L}^{\,0}(t,\hat{y})\,.$$

Hence (D3) depends on $L^{\,0}$ only, and agrees (up to a numerical factor) with (D1), which confirms that the latter integral is invariant.

The integrand in (6.18) is

$$j_\alpha [\, y_\alpha \partial^2 - 2(y\cdot\partial + d + 1)\partial_\alpha\,]\Lambda$$
$$= \partial_\beta Q_\beta + \Lambda\, [\, \partial^2 y\cdot j - 2(\hat{N}+3)\partial\cdot j]$$

with

$$Q_\beta = y_\alpha (j_\alpha \partial_\beta - j_\beta \partial_\alpha)\Lambda$$
$$+ [2/(d+1)] y_\alpha (y_\beta \partial_\alpha - y_\alpha)\, j\cdot\partial\Lambda$$
$$- y_\alpha (\Lambda \partial_\beta j_\alpha - \Lambda \partial_\alpha j_\beta) + (\hat{N}+3)\Lambda j_\beta\,.$$

Since the last term is zero when $d=3$, one has in that case

$$\partial_\beta Q_\beta =$$
$$(\partial_\beta y_\alpha - \partial_\alpha y_\beta [\, j_\alpha \partial_\beta \Lambda - \tfrac{1}{2} y_\alpha \partial_\beta\, j\cdot\partial\Lambda + \Lambda \partial_\alpha j_\beta\,]$$
$$= -iM_{\alpha\beta} [\, j_\beta \leftrightarrow \partial_\beta \Lambda + \tfrac{1}{4} i M_{\alpha\beta}\, j\cdot\partial\Lambda\,]\,,$$

which reveals that $\int dy\, \partial\cdot Q = 0$, and (7.8) follows.

[1] E. Cunningham, Proc. Math. Soc. London **8**, 77 (1909); H. Bateman, *ibid.* **8**, 223 (1910)
[2] P. A. M. Dirac, Ann. Math. **37**, 429 (1936).
[3] G. Mack and A. Salam, Ann. Phys. **53**, 174 (1969).
[4] H. A. Kastrup, Phys. Rev. **150**, 1189 (1964); W. Heidenreich, Diplomarbeit, Munich, 1977 (unpublished).
[5] M. Baker and K. Johnson, Phys. Rev. D **3**, 2516, 2541 (1971).
[6] G. Mack and K. Symanzik, Commun. Math. Phys. **27**, 247 (1972).
[7] S. Adler, Phys. Rev. D **6**, 3445 (1972); **8**, 2400 (1973).
[8] M. Flato, J. Simon, and D. Sternheimer, Ann. Phys. **61**, 78 (1970); J. Niederle, *Proceedings of the VIIIth Winter School of Theoretical Physics in Karpacz* (Wroclaw, 1972), Vol. II, p. 44; A. O. Barut and W. E. Brittin (Eds.), *Lectures in Theoretical Physics* (Univ. of Colorado Press, Boulder, 1971), Vol. XIII. For additional references, see I. T. Todorov, M. C. Mintchev, and V. B. Petkova, "Conformal Invariance in Quantum Field Theory," Scuola Normale Superiore, Pisa, 1978.
[9] H. Weyl, *Space–Time–Matter* (Dover, New York, 1950).
[10] For a comprehensive review see S. L. Adler, Rev. Mod. Phys. (to be published).
[11] See, for example, S. Ferrara and B. Zumino, Nucl. Phys. B **134**, 301 (1978); K. S. Stelle, Phys. Rev. D **16**, 953 (1977).
[12] I. E. Segal, *Mathematical Cosmology and Extragalactic Astronomy* (Academic, New York, 1976); H. P. Jacobson *et al.*, Proc. Nat. Acad. Sci. USA **75**, 1609 (1978); J. Mickelsson and J. Niederle, Ann Inst. H. Poincaré **23**, 277 (1975).
[13] R. P. Feynman, Acta Phys. Polon. **24**, 697 (1963); B. S. DeWitt, *Relativity, Groups and Topology* (Blackie, London, 1964), pp. 587–820; L. D. Faddeev and V. N. Popov, Phys. Lett. B **25**, 29 (1967).
[14] C. Becchi, A. Rouet, and R. Stora, Ann. Phys. **98**, 287 (1976).
[15] T. Kugo and I. Ojima, Progr. Theor. Phys. **60**, 1869 (1978); **61**, 294 (1979); N. Nakanishi, Progr. Theor. Phys. **35**, 1111 (1966).
[16] C. Fronsdal, in *Lecture Notes in Physics*, Vol. 153 (Springer-Verlag, Berlin, 1982), p. 329–335.
[17] J. C. Pati and A. Salam, Phys. Rev. D **10**, 275 (1974); J. C. Pati, A. Salam, and J. Strathdee, Phys. Lett. B **59**, 265 (1975); K. Matsumoto, Prog. Theor. Phys. **52**, 1973 (1974).
[18] M. Flato and C. Fronsdal, Phys. Lett. B **97**, 236 (1980).
[19] M. Flato and C. Fronsdal, J. Math. Phys. **22**, 1100 (1981).
[20] E. Fermi, Rend. R. Accad. Lincei (6) **9**, 881 (1929).
[21] S. N. Gupta, Proc. Phys. Soc. A **63**, 681 (1950); K. Bleuler, Helv. Phys. Acta **23**, 567 (1950).
[22] E. Angelopoulos and M. Flato, Lett Math. Phys. **2**, 405 (1978).
[23] G. Mack and I. T. Todorov, J. Math. Phys. **10**, 2078 (1969).
[24] The general character of indecomposable representations has been discussed in I. N. Bernshtein, I. M. Gel'fand, and S. I. Gel'fand, Funct. Anal. Priozen **5**, 1 (1977) [Func. Anal. Appl. **5**, 1 (1971)]; G. Pinczon and J. Simon, Rep. Math. Phys. **16**, 49 (1979).
[25] C. Fronsdal, "Composite Massless Fields from Racs," UCLA preprint, 1981.
[26] D. H. Mayer, J. Math. Phys. **16**, 884 (1975); F. Bayen and M. Flato, J. Math. Phys. **17**, 1112 (1976).
[27] M. Lüscher and G. Mack, Commun. Math. Phys. **41**, 203 (1975).
[28] B. Binegar, M. Flato, C. Fronsdal, and S. Salamó, Czech. J. Phys. B **32**, 439 (1982).
[29] L. Castell, Nucl. Phys. B **13**, 231 (1969).

[30]S. Deser, R. Jackiw, and S. Templeton, MIT Preprint CTP #964, 1981 [submitted to Ann. Phys. (N.Y.)].
[31]R. J. Blattner and J. H. Rawnsley, preprint, University of Marwick, 1981.
[32]W. Schmid and J. A. Wolf, "Geometric Construction of Singular Unitary Representations of Reductive Lie Groups," preprint, 1982.
[33]J. Rawnsley, W. Schmid, and J. A. Wolf, "Singular Unitary Representations and Indefinite Harmonic Theory," J. Funct. Anal. **51**, 1–114 (1983).
[34]W. Schmid, in *Lecture Notes in Physics*, Vol. 153 (Springer-Verlag, Berlin, 1982), p. 348–355.
[35]T. J. Enright, R. Howe, and N. R. Wallach (to be published).
[36]G. Mack, Commun. Math. Phys. **55**, 1 (1977).

Cross sections with polarized spin-1/2 particles in terms of helicity amplitudes

Till B. Anders and Wolfgang Jachmann
Sektion Physik, Universität München, Munich, West Germany

(Received 5 April 1983; accepted for publication 18 July 1983)

The derivation of cross sections for collisions with polarized particles of spin $\frac{1}{2}$ can be simplified considerably if the scattering amplitude is calculated explicitly before the transition probability is obtained by a simple squaring. The states of the polarized particles are represented as superpositions of states with definite helicity. The coefficients of the superposition relate directly to the strength of the transversal and longitudinal polarization. The helicity amplitudes are products of helicity currents for which detailed formulas have been elaborated. Two-component spinors have been used. All the contractions of vector indices are done with a new set of general formulas. The resulting cross sections show terms with separate factors due to the polarization, the energy, and the directions of the particles. Therefore, the high energy approximation can be achieved very conveniently. Applications of the described method have been performed to the scattering of electrons by electrons or positrons including the exchange of Z_0 and Higgs particles.

PACS numbers: 13.85.Lg, 11.80. – m, 12.20. – m

1. THE USUAL METHOD OF CALCULATING CROSS SECTIONS

The usual method[1] of deriving cross sections of QED reactions avoids calculating explicitly the scattering amplitude

$$M_{fi} = \bar{u}^{Q'}(p', s')\Gamma u^{Q}(p, s) \tag{1.1}$$

between the states $u^{Q}(p, s)$ of polarized spin-$\frac{1}{2}$ particles. Instead of this the transition probability

$$|M_{fi}|^2 = [\bar{u}^{Q}(p,s)\gamma_0\Gamma^{\dagger}\gamma_0 u^{Q'}(p',s')][\bar{u}^{Q'}(p',s')\Gamma u^{Q}(p,s)] \tag{1.2}$$

is considered and expressed as a trace of Dirac matrices. Thereby the projection operators for normalized spin states[2] are used

$$\Lambda^{Q}(p,s) \equiv u^{Q}(p,s)\bar{u}^{Q}(p,s) = \frac{p + Qm}{2m} \cdot \frac{1}{2}(1 + \gamma_5 \not{\hat{s}}). \tag{1.3}$$

The quantity Q characterizes the fermions ($Q = +1$) and antifermions ($Q = -1$). The vector $\hat{s}^{\mu}$ represents the moving spin vector, which is gained from the spin vector s^{μ} in the rest frame of the particle

$$s^{\mu} = (0, \vec{s}), \quad |\vec{s}|^2 = 1 \tag{1.4}$$

through a Lorentz transformation. A separation of the rest spin vector in the components parallel and transversal to the direction of the momentum $\hat{p} = \vec{p}/|\vec{p}|$,

$$s = s_{\parallel}\,\hat{p} + \vec{s}_{\perp}, \quad s_{\parallel} = \vec{s}\cdot\hat{p}, \tag{1.5}$$

allows a simple presentation of the moving spin vector

$$\hat{s}^{\mu} = \hat{s}^{\mu}_{\parallel} + \hat{s}^{\mu}_{\perp}, \tag{1.6}$$

where

$$\hat{s}^{\mu}_{\parallel} = s_{\parallel}\left(\frac{p}{m}, \frac{E}{m}\hat{p}\right), \quad \hat{s}^{\mu}_{\perp} = (0, \vec{s}_{\perp}). \tag{1.7}$$

The transition probability now becomes a trace of Dirac matrices (more generally a product of such traces)

$$|M_{fi}|^2 = \mathrm{tr}[\gamma_0\Gamma^{+}\gamma_0\Lambda^{Q'}(p',s')\Gamma\Lambda^{Q}(p,s)]. \tag{1.8}$$

Such traces can be evaluated analytically by means of special computer systems. Let $\Gamma = \Gamma_n$ describe the kernel of a fermion current with n vertices. In this case at least $(4n + 1)!!$ terms arise by the evaluation of the trace, if the fermions are polarized, but $(4n - 1)!!$ terms with unpolarized particles.

Each polarization vector (1.7) enlarges the magnitude of the resulting terms by a factor of the order E/m. At high energy conditions, however, all these big magnitude terms cancel out. The cancelation can be avoided if the high energy approximation of the spin projection operator (1.3) is used[3]

$$\Lambda^{Q}(p,s) = (Q/4m)\not{p}(1 + Qs_{\parallel}\gamma_5 + \gamma_5 \not{\hat{s}}_{\perp}) + O(1). \tag{1.9}$$

Anyway, the formulas of the cross sections which are derived with the conventional method, presents a lot of terms in a completely unarticulated form, which allows no analytic insight into the actions of the various polarization states.

2. ALTERNATIVE WAYS TO EVALUATE SCATTERING AMPLITUDES WITH POLARIZED FERMIONS

For many cases it is very convenient first to evaluate a matrix element explicitly and thereafter to gain the transition probability by squaring it. This proceeding has been proposed by several authors.[4–10]

A further facilitation of the calculations is obtained if the amplitude is expanded with helicity states.[4–8] We will observe that the coefficients of this expansion are in a direct relation to the degrees of the longitudinal and transversal polarizations of the particles. Moreover, in this way the dependence on the spins and on the dynamical factors is separated, because the influence of the polarization is exclusively described by the expansion coefficients.

The explicit calculation of the amplitude can be done in different ways. Many authors use four-component spinors and the γ matrices in the original form of Dirac.[9,10] The transition to the two-component spinors of van der Waerden

 0022-2488/83/122847-08$02.50

and to Pauli matrices causes essential simplifications.[7] The present paper uses the two-component spinors which describe the spin directions in the rest frames of the particles. The restriction to a spin description in the rest frame allows to use only one type of van der Waerden spinors representing the rotation group. Spinors for moving particles are obtained by the application of boosts $\sqrt{p\sigma/m}$ and rotations U.[7,8] We show that it is convenient to use only boosts in a suitably defined z direction. With this restriction we can separate the dependence of the amplitude on the energies and on the directions of the moving particles because in the z direction the boosts are diagonal. With boosts in arbitrary directions no such separation can be obtained.

The separated energy factors allow a transparent application of high energy approximations just in this early state of the calculation. The high energy approximation causes a considerable simplification of the results because only few helicity amplitudes survive.

The scattering amplitudes consist of products of currents in such a way that all the vector indices are contracted. Several authors use an expansion of the currents into 16 basis currents according to the fact that the Dirac algebra shows 16 basis elements Γ_i

$$\Gamma = \sum_{i=1}^{16} A_i \Gamma_i, \quad A_i = \tfrac{1}{2}\mathrm{tr}[\Gamma\Gamma^i]. \tag{2.1}$$

The 16 basis currents are evaluated easily by different methods.[7–10] The coefficients of these expansions are traces containing vector indices. All these indices have to be contracted now according to the usual methods—Chisholm's formulas,[11] Kahane's algorithm,[12] etc. In this way the advantages as well as all disadvantages of the conventional trace methods return to the calculation. The expansion into the 16 basis currents can be avoided completely and replaced by a treatment solving the problem more directly.

The Feynman diagrams present the kernels of the currents as sums of products of γ matrices. The transition to two-component van der Waerden spinors yields terms containing a separate energy factor and a Pauli current including products of Pauli matrices with vector indices alternating their covariant or contravariant positions.

The vector indices of the Pauli currents now have to be contracted either with the indices of other Pauli currents of the reaction or with other external vectors, i.e., describing the polarization of the photons, etc. We give powerful formulas (3.8), (3.10), and (3.11) which make the task of contraction solvable in complete generality.

Finally the amplitudes are represented by a sum of terms, each of which shows an energy factor and several special spinor scalars formed by two van der Waerden spinors. The scalars describe the dependence on the directions of the momenta and spins.

We notice that quite general symmetries connect the helicity currents and amplitudes with the ones of reversed helicity indices (Sec. 4). These symmetry relations have been derived independently of the discrete Lorentz transformations. They reduce the number of the independent dynamical functions of the cross sections by a factor of 2.

3. THE DETAILS OF EVALUATING THE AMPLITUDES

Each scattering amplitude with spin-$\frac{1}{2}$ particles is composed by Dirac currents $\bar{u}^{Q'}(p',\hat{s}')\Gamma u^{Q}(p,\hat{s})$. Here the kernel Γ is a special 4×4 matrix representing the happenings of a particle passing the reaction. Generally it shows several vector indices which are contracted with the indices of momenta or polarization vectors of photons or other spin-1 particles, or with the vector indices of further Dirac currents of the reaction. The kernel can be involved in momentum loop integrals. It is given by a sum of products of γ matrixes.

The spinors $u^{Q}(p,\hat{s})$ are conveniently expanded by two helicity spinors (Appendix A)

$$u^{Q}(p,\hat{s}) = \alpha^{Q}_{+}(\hat{p},\hat{s})u^{Q}_{+}(p) + \alpha^{Q}_{-}(\hat{p},\hat{s})u^{Q}_{-}(p). \tag{3.1}$$

The coefficients do not depend on the energy of the particle. In this way each Dirac current is expanded by four helicity currents

$$\begin{aligned}\bar{u}^{Q'}(p',\hat{s}')\Gamma u^{Q}(p,\hat{s}) &= \alpha^{Q'}_{+}\alpha^{Q}_{+}(\bar{u}^{Q'}_{+}\Gamma u^{Q}_{+}) + \alpha^{Q'}_{+}\alpha^{Q}_{-}(\bar{u}^{Q'}_{+}\Gamma u^{Q}_{-}) \\ &\quad + \alpha^{Q'}_{-}\alpha^{Q}_{+}(\bar{u}^{Q'}_{-}\Gamma u^{Q}_{+}) + \alpha^{Q'}_{-}\alpha^{Q}_{-}(\bar{u}^{Q'}_{-}\Gamma u^{Q}_{-}).\end{aligned} \tag{3.2}$$

The four-component helicity spinors can be separated in a two-component spinor $|N,\hat{p}\rangle$ and energy factors $\eta_{\pm}$ (Appendix A)

$$u^{Q}_{N}(p) = \frac{1}{\sqrt{2}}\begin{pmatrix}\eta_{QN}\cdot|QN,\hat{p}\rangle \\ Q\eta_{-QN}\cdot|QN,\hat{p}\rangle\end{pmatrix}, \quad \eta_N = \left(\frac{E+Np}{m}\right)^{1/2}. \tag{3.3}$$

Accordingly, the γ matrices

$$\gamma^{\mu} = \begin{pmatrix}0 & \sigma_{\mu} \\ \sigma^{\mu} & 0\end{pmatrix}, \quad \gamma_5 = \begin{pmatrix}1 & 0 \\ 0 & -1\end{pmatrix} \tag{3.4}$$

and their products can be split into four Pauli matrices. We define the product

$$\sigma^{\nu_1}{}_{\nu_2}{}^{\nu_3\cdots}{}_{\cdots} = \sigma^{\nu_1}\sigma_{\nu_2}\sigma^{\nu_3}\cdots, \tag{3.5}$$

which results in a special Hermitian Pauli matrix. The use of this splitting brings the helicity currents of pure products of γ matrices into the following forms (here is $a = 1, 0$):

$$\begin{aligned}&\bar{u}^{Q'}_{Q'N'}(p')\prod_{i=1}^{2n}\gamma^{\nu_i}(\gamma_5)^a u^{Q}_{QN}(p) \\ &\quad = \frac{Q'}{2}\eta'_{-N'}\eta_N\langle N',\hat{p}'|\sigma_{\nu_1}{}^{\nu_2\cdots\nu_{2n}}{}_{\cdots}|N,\hat{p}\rangle \\ &\qquad + (-1)^a\frac{Q}{2}\eta'_{N'}\eta_{-N}\langle N',\hat{p}'|\sigma^{\nu_1\cdots}{}_{\nu_2\cdots\nu_{2n}}|N,\hat{p}\rangle,\end{aligned} \tag{3.6a}$$

$$\begin{aligned}&\bar{u}^{Q'}_{Q'N'}(p')\prod_{i=1}^{2n+1}\gamma^{\nu_i}(\gamma_5)^a u^{Q}_{QN}(p) \\ &\quad = \tfrac{1}{2}\eta'_{N'}\eta_N\langle N',\hat{p}'|\sigma^{\nu_1\cdots\nu_{2n+1}}{}_{\nu_2\cdots}|N,\hat{p}\rangle \\ &\qquad + (-1)^a\tfrac{1}{2}QQ'\eta'_{-N'}\eta_{-N}\langle N',\hat{p}'|\sigma_{\nu_1}{}^{\nu_2\cdots}{}_{\cdots\nu_{2n+1}}|N,\hat{p}\rangle.\end{aligned} \tag{3.6b}$$

The helicity currents of γ products are separated into energy factors and Pauli currents of the form

$$\langle N',\hat{p}'|\sigma^{\nu_1}{}_{\nu_2}{}^{\nu_3\cdots}{}_{\cdots}|N,\hat{p}\rangle. \tag{3.7}$$

The forms (3.6) help to derive general symmetry relations for the helicity currents and amplitudes (Sec. 4). They also give a basis for the applications of a high energy approximation which make only a few helicity amplitudes survive (Sec. 5).

The evaluation of the scattering amplitudes requires that all vector indices of the helicity currents have to be contracted. The contraction of the vector indices of the Pauli currents (3.7) with other external momenta can be treated with the formulas (Appendix A)

$$p_\mu \sigma^\mu = m \sum_N |N, \hat{p}\rangle \eta^2_{-N} \langle N, \hat{p}|,$$
$$p_\mu \sigma_\mu = m \sum_N |N, \hat{p}\rangle \eta^2_N \langle N, \hat{p}|. \tag{3.8}$$

Some cases require a contraction of Pauli currents with normalized polarization vectors of spin 1 particles $\epsilon^\mu_\pm(p)$ perpendicular to the momentum p^μ. We use the complex vector $\vec{g}$ given by (A27) in the helicity frame (Appendix A) and define in agreement with the literature[6,13,14]

$$\epsilon^\mu_+(p) = (1/\sqrt{2})(0, -\vec{g}), \quad \epsilon^\mu_-(p) = (1/\sqrt{2})(0, \vec{g}^*). \tag{3.9}$$

The contraction of Pauli currents with these polarization vectors is treated with the formula (A28) or

$$\epsilon^\mu_N(p)\sigma_\mu = -\epsilon^\mu_N(p)\sigma^\mu = N\sqrt{2}|N, \hat{p}\rangle\langle -N, \hat{p}|. \tag{3.10}$$

The vector contraction of two Pauli currents can be treated with the formulas (Appendix A)[15]

$$\langle N_1|a\sigma_\mu b\,|N_2\rangle\langle N_3|c\sigma_\mu d\,|N_4\rangle = 2\langle N_1|ad\,|N_4\rangle\langle N_3|cb\,|N_2\rangle,$$
$$\langle N_1|a\sigma_\mu b\,|N_2\rangle\langle N_3|c\sigma^\mu d\,|N_4\rangle = 2\langle N_1|ab\,|N_2\rangle\langle N_3|cd\,|N_4\rangle - 2\langle N_1|ad\,|N_4\rangle\langle N_3|cb\,|N_2\rangle. \tag{3.11}$$

Here a, b, c, and d mean any arbitrary 2×2 matrix. Some specializations of these formulas are found in Appendix A. In many cases we are interested in contracted products of two helicity currents. Appendix B shows explicitly how to treat this case.

The complete contraction of all vector indices makes all the Pauli currents (3.7) disappear. Instead of them we obtain products of scalars of two two-component spinors $\langle N_1, \hat{p}_1|N_2, \hat{p}_2\rangle$ which each can be expressed by the directions $\theta_i\phi_i$ of the vectors $\hat{p}_i$

$$\begin{aligned}
\langle +, \hat{p}_1| +, \hat{p}_2\rangle &= \cos\frac{\theta_1}{2}\cos\frac{\theta_2}{2} \\
&\quad + \sin\frac{\theta_1}{2} e^{-i\phi_1} \sin\frac{\theta_2}{2} e^{i\phi_2}, \\
\langle +, \hat{p}_1| -, \hat{p}_2\rangle &= -\cos\frac{\theta_1}{2}\sin\frac{\theta_2}{2} e^{-i\phi_2} \\
&\quad + \sin\frac{\theta_1}{2} e^{-i\phi_1}\cos\frac{\theta_2}{2}, \\
\langle -, \hat{p}_1| +, \hat{p}_2\rangle &= -\langle +, \hat{p}_1| -, \hat{p}_2\rangle^*, \\
\langle -, \hat{p}_1| -, \hat{p}_2\rangle &= \langle +, \hat{p}_1| +, \hat{p}_2\rangle^*.
\end{aligned} \tag{3.12}$$

We show in Appendix A that the coefficients used in the expansion (3.1) of general spinors by the helicity spinors can also be represented by these spinor scalars

$$\alpha^Q_N(\hat{p}, \hat{s}) = \langle QN, \hat{p}|Q, \hat{s}\rangle = N\,[\alpha^{-Q}_{-N}(\hat{p}, \hat{s})]^*. \tag{3.13}$$

The expansion coefficients have a direct relation to the polarization vectors of the incoming and outgoing spin-$\frac{1}{2}$ particles[5,16] (Appendix A)

$$\alpha^Q_N(\alpha^Q_N)^* = \tfrac{1}{2}(1 + Ns_\parallel), \quad \alpha^Q_N(\alpha^Q_{-N})^* = \tfrac{1}{2}Qs_\perp e^{-iQN\psi}. \tag{3.14}$$

These formulas (3.14) will directly apply when the cross section is obtained by the squaring of the scattering amplitude. The simple and transparent form of the cross section is shown in Sec. 6.

4. SYMMETRIES OF THE HELICITY AMPLITUDES

The Pauli currents obey simple symmetry relations which generalize the relations (A34) and (A35)

$$\langle N_1, p_1|\sigma^{\nu_1\cdots}_{\nu_2\cdots}|N_2, p_2\rangle^* = \langle N_2, p_2|\sigma^{\cdots\nu_1}_{\cdots\nu_2}|N_1, p_1\rangle, \tag{4.1}$$
$$\langle N_1, p_1|\sigma^{\nu_1\cdots}_{\nu_2\cdots}|N_2, p_2\rangle = N_1N_2\langle -N_2, p_2|\sigma^{\cdots\nu_1}_{\cdots\nu_2}|-N_1, p_2\rangle. \tag{4.2}$$

These relations and the forms (3.6) help to derive quite general symmetry relations for the helicity currents:

$$[\bar{u}^{Q'}_{N'}(p')\Gamma u^Q_N(p)]^* = NN'\bar{u}^{Q'}_{-N'}(p')\Gamma^{(*)}u^Q_{-N}(p). \tag{4.3}$$

Here we suppose that the kernel Γ is given as a sum of products of γ matrices with arbitrary coefficients. The kernel $\Gamma^{(*)}$ arises from the original expression by an operation $(*)$ which keeps all the factors γ^μ on its place and unchanged but takes the conjugate complex value of all the coefficients. Especially, the matrix $\gamma_5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ changes its sign.

The scattering amplitudes as described by a Feynman diagram is a product of several (say n) Dirac currents. Let F fermions be of equal type. In this case the general amplitude consists of a sum of 4^n different helicity amplitudes, each of which belongs to $F!$ different diagrams. The pairs of terms which show reversed helicity indices N_i are connected by a symmetry relation of the type (4.3), namely

$$\left\langle\left\langle \begin{matrix}\cdots Q_k\cdots \\ \cdots N_k\cdots\end{matrix}\right|\mathfrak{M}\left|\begin{matrix}\cdots Q_i\cdots \\ \cdots N_i\cdots\end{matrix}\right\rangle\right\rangle^* = \left\langle\left\langle \begin{matrix}\cdots Q_k\cdots \\ \cdots -N_k\cdots\end{matrix}\right|\mathfrak{M}^{(*)}\left|\begin{matrix}\cdots Q_i\cdots \\ \cdots -N_i\cdots\end{matrix}\right\rangle\right\rangle \cdot \prod_{i,k} N_iN_k. \tag{4.4}$$

As a consequence of the operation $(*)$ all the pseudovector or pseudoscalar couplings change their sign. Also the helicity indices of the emitted and absorbed photons are reversed, and a factor -1 is multiplied for each photon.

5. THE HIGH ENERGY APPROXIMATION

For high energies the ratio η_-/η_+ is small and approximately $m/2E$. Only two of the four helicity currents of the type (3.6a) or (3.6b) survive. For the currents including even products of γ^μ (type $\tau = 0$) the surviving currents are

$$\begin{aligned}
&\bar{u}^{Q_1}_{Q_1}(p_1)\prod_{i=1}^{2n}\gamma^{\nu_i}(\gamma_5)^a u^{Q_2}_{-Q_2}(p_2) \\
&\quad = (-1)^a\frac{Q_2}{2}\eta^{(1)}_+\eta^{(2)}_+\langle +, \hat{p}_1|\sigma^{\nu_1\cdots}_{\nu_2\cdots\nu_{2n}}|-, \hat{p}_2\rangle, \\
&\bar{u}^{Q_1}_{-Q_1}(p_1)\prod_{i=1}^{2n}\gamma^{\nu_i}(\gamma_5)^a u^{Q_2}_{Q_2}(p_2) \\
&\quad = \frac{Q_1}{2}\eta^{(1)}_+\eta^{(2)}_+\langle -, \hat{p}_1|\sigma_{\nu_1}{}^{\nu_2\cdots\nu_{2n}}_{\cdots}|+, \hat{p}_2\rangle.
\end{aligned} \tag{5.1}$$

For the currents including odd products of γ^μ (type $\tau = +1$) the surviving currents are

$$\bar{u}^{Q_1}_{Q_1}(p_1)\prod_{i=1}^{2n+1}\gamma^{\nu_i}(\gamma_5)^a u^{Q_2}_{Q_2}(p_2)$$
$$= \tfrac{1}{2}\eta^{(1)}_+\eta^{(2)}_+\langle +,\hat{p}_1|\sigma^{\nu_1\cdots\nu_{2n+1}}_{\nu_2\cdots}|+,\hat{p}_2\rangle, \tag{5.2}$$

$$\bar{u}^{Q_1}_{-Q_1}(p_1)\prod_{i=1}^{2n+1}\gamma^{\nu_i}(\gamma_5)^a u^{Q_2}_{-Q_2}(p_2)$$
$$= (-1)^a\frac{Q_1Q_2}{2}\eta^{(1)}_+\eta^{(2)}_+\langle -,\hat{p}_1|\sigma^{\nu_2\cdots}_{\nu_1\cdots\nu_{2n+1}}|-,\hat{p}_2\rangle.$$

In agreement with the symmetry (4.3) one of the surviving currents is the conjugate complex of the other—except eventually a sign.

A general Dirac kernel Γ of a current has either even ($\tau = 0$) or odd ($\tau = +1$) factors γ^μ in the high energy limit. We remember that the fermion propagators produce factors $p + m$, which reduce simply to p in the high energy limit. We conclude that only two of four helicity currents survive in the high energy limit.

We suppose that this statement is correct also when the currents are involved in any momentum loop integrals (Feynman integrals). For instance, the explicit QED formulas for the radiative corrections of the e–e scattering in the lowest order[17,18] show products of two helicity currents only of the type $\tau = +1$.

We have seen that the scattering amplitude is a special product of n Dirac currents, leading to 4^n different helicity amplitudes. The high energy approximation reduces this number drastically to 2^n many surviving helicity amplitudes.

The previous statements shall be exemplified. We consider the creation of a pair of muons (particle $3 = \mu^-$, particle $4 = \mu^+$) by the collision of an electron (particle $1 = e^-$) and a positron (particle $2 = e^+$). We suppose this reaction as running through the production of one or more photons and Z_0, i.e., intermediate spin-1 particles. In this case the amplitude has the following form

$$\mathfrak{M}_{e^-e^+\to(1)\to\mu^-\mu^+}$$
$$= -\alpha_1\beta_2\beta^*_3\alpha^*_4A^{\mathrm{I}} + \alpha_1\beta_2\alpha^*_3\beta^*_4A^{\mathrm{II}}$$
$$+\beta_1\alpha_2\beta^*_3\alpha^*_4A^{\mathrm{III}} - \beta_1\alpha_2\alpha^*_3\beta^*_4A^{\mathrm{IV}}. \tag{5.3}$$

Here we used a simplified notation for the superposition coefficients (3.1) or (3.13)

$$\alpha_i = \langle +,\hat{p}_i|+,\hat{s}_i\rangle, \qquad \beta_i = \langle -,\hat{p}_i|+,\hat{s}_i\rangle,$$
$$-\beta^*_i = \langle +,\hat{p}_i|-,\hat{s}_i\rangle, \quad \alpha^*_i = \langle -,\hat{p}_i|-,\hat{s}_i\rangle. \tag{5.4}$$

The dynamic functions A^i are the helicity amplitudes[19] surviving in the HEA

$$A^{\mathrm{I}} = \left\langle\left\langle \begin{matrix}- & + \\ 2 & 3\\ - & -\end{matrix}\middle|\mathfrak{M}\middle|\begin{matrix}+ & - \\ 1 & 4\\ + & +\end{matrix}\right\rangle\right\rangle, \qquad A^{\mathrm{II}} = \left\langle\left\langle \begin{matrix}- & + \\ 2 & 3\\ - & +\end{matrix}\middle|\mathfrak{M}\middle|\begin{matrix}+ & - \\ 1 & 4\\ + & -\end{matrix}\right\rangle\right\rangle,$$
$$A^{\mathrm{III}} = \left\langle\left\langle \begin{matrix}- & + \\ 2 & 3\\ + & -\end{matrix}\middle|\mathfrak{M}\middle|\begin{matrix}+ & - \\ 1 & 4\\ - & +\end{matrix}\right\rangle\right\rangle, \qquad A^{\mathrm{IV}} = \left\langle\left\langle \begin{matrix}- & + \\ 2 & 3\\ + & +\end{matrix}\middle|\mathfrak{M}\middle|\begin{matrix}+ & - \\ 1 & 4\\ - & -\end{matrix}\right\rangle\right\rangle. \tag{5.5}$$

The kernels of the constituting currents have odd factors γ^μ in the HEA. They are of the type (5.2) with $\tau = 1$.

If the creation of $\mu^-\mu^+$ by e^-e^+ runs through odd additional scalar Higgs particles,[20] a further amplitude has to be added, namely,

$$\mathfrak{N}_{e^+e^-\to(0)\to\mu^+\mu}$$
$$= \alpha_1\alpha_2\alpha^*_3\alpha^*_4B^{\mathrm{I}} - \alpha_1\alpha_2\beta^*_3\beta^*_4B^{\mathrm{II}}$$
$$-\beta_1\beta_2\alpha^*_3\alpha^*_4B^{\mathrm{III}} + \beta_1\beta_2\beta^*_3\beta^*_4B^{\mathrm{IV}}. \tag{5.6}$$

The dynamical functions B^i are the helicity amplitudes surviving in the HEA

$$B^{\mathrm{I}} = \left\langle\left\langle \begin{matrix}- & + \\ 2 & 3\\ + & +\end{matrix}\middle|\mathfrak{N}\middle|\begin{matrix}+ & - \\ 1 & 4\\ + & +\end{matrix}\right\rangle\right\rangle, \qquad B^{\mathrm{II}} = \left\langle\left\langle \begin{matrix}- & + \\ 2 & 3\\ + & -\end{matrix}\middle|\mathfrak{N}\middle|\begin{matrix}+ & - \\ 1 & 4\\ + & -\end{matrix}\right\rangle\right\rangle,$$
$$B^{\mathrm{III}} = \left\langle\left\langle \begin{matrix}- & + \\ 2 & 3\\ - & +\end{matrix}\middle|\mathfrak{N}\middle|\begin{matrix}+ & - \\ 1 & 4\\ - & +\end{matrix}\right\rangle\right\rangle, \qquad B^{\mathrm{IV}} = \left\langle\left\langle \begin{matrix}- & + \\ 2 & 3\\ - & -\end{matrix}\middle|\mathfrak{N}\middle|\begin{matrix}+ & - \\ 1 & 4\\ - & -\end{matrix}\right\rangle\right\rangle. \tag{5.7}$$

The kernels of the currents constituting this channel of the reaction have even factors γ^μ in the HEA. They are of the type (5.1) with $\tau = 0$.

The scattering reaction $e^-e^+\to e^-e^+$ contains similar amplitudes as (5.3) and (5.6), with particle $3 = e^-$ and particle $4 = e^+$. But we have to add (subtract) the amplitudes of the genuine scattering channel. The scattering channel of the type $\tau = 1$ (arbitrary many intermediate spin-1 particles and even spin-0 particles exchanged) yields the HEA

$$\mathfrak{M}'_{e^-e^+\to(1)\to e^-e^+}$$
$$= -\alpha_1\alpha_2\alpha^*_3\alpha^*_4C^{\mathrm{I}} - \alpha_1\beta_2\alpha^*_3\beta^*_4C^{\mathrm{II}}$$
$$-\beta_1\alpha_2\beta^*_3\alpha^*_4C^{\mathrm{III}} - \beta_1\beta_2\beta^*_3\beta^*_4C^{\mathrm{IV}}, \tag{5.8}$$

where the dynamical functions C^i are defined as

$$C^{\mathrm{I}} = \left\langle\left\langle \begin{matrix}+ & - \\ 3 & 2\\ + & +\end{matrix}\middle|\mathfrak{M}'\middle|\begin{matrix}+ & - \\ 1 & 4\\ + & +\end{matrix}\right\rangle\right\rangle, \qquad C^{\mathrm{II}} = \left\langle\left\langle \begin{matrix}+ & - \\ 3 & 2\\ + & -\end{matrix}\middle|\mathfrak{M}'\middle|\begin{matrix}+ & - \\ 1 & 4\\ + & -\end{matrix}\right\rangle\right\rangle,$$
$$C^{\mathrm{III}} = \left\langle\left\langle \begin{matrix}+ & - \\ 3 & 2\\ - & +\end{matrix}\middle|\mathfrak{M}'\middle|\begin{matrix}+ & - \\ 1 & 4\\ - & +\end{matrix}\right\rangle\right\rangle, \qquad C^{\mathrm{IV}} = \left\langle\left\langle \begin{matrix}+ & - \\ 3 & 2\\ - & -\end{matrix}\middle|\mathfrak{M}'\middle|\begin{matrix}+ & - \\ 1 & 4\\ - & -\end{matrix}\right\rangle\right\rangle. \tag{5.9}$$

The scattering channel of type $\tau = 0$ (odd spin-0 particles and arbitrary many intermediate spin-1 particles exchanged) yields the HEA

$$\mathfrak{N}'_{e^-e^+\to(0)\to e^-e^+}$$
$$= -\alpha_1\beta_2\beta^*_3\alpha^*_4D^{\mathrm{I}} - \alpha_1\alpha_2\beta^*_3\beta^*_4D^{\mathrm{II}}$$
$$-\beta_1\beta_2\alpha^*_3\alpha^*_4D^{\mathrm{III}} - \beta_1\alpha_2\alpha^*_3\beta^*_4D^{\mathrm{IV}}, \tag{5.10}$$

where the dynamical functions D^i are defined as

$$D^{\mathrm{I}} = \left\langle\left\langle \begin{matrix}+ & - \\ 3 & 2\\ - & -\end{matrix}\middle|\mathfrak{N}'\middle|\begin{matrix}+ & - \\ 1 & 4\\ + & +\end{matrix}\right\rangle\right\rangle, \qquad D^{\mathrm{II}} = \left\langle\left\langle \begin{matrix}+ & - \\ 3 & 2\\ - & +\end{matrix}\middle|\mathfrak{N}'\middle|\begin{matrix}+ & - \\ 1 & 4\\ + & -\end{matrix}\right\rangle\right\rangle,$$
$$D^{\mathrm{III}} = \left\langle\left\langle \begin{matrix}+ & - \\ 3 & 2\\ + & -\end{matrix}\middle|\mathfrak{N}'\middle|\begin{matrix}+ & - \\ 1 & 4\\ - & +\end{matrix}\right\rangle\right\rangle, \qquad D^{\mathrm{IV}} = \left\langle\left\langle \begin{matrix}+ & - \\ 3 & 2\\ + & +\end{matrix}\middle|\mathfrak{N}'\middle|\begin{matrix}+ & - \\ 1 & 4\\ - & -\end{matrix}\right\rangle\right\rangle. \tag{5.11}$$

We notice the symmetries as consequences of the relation (4.4)

$$A^{\mathrm{IV}(*)} = A^{\mathrm{I}*}, \quad A^{\mathrm{III}(*)} = A^{\mathrm{II}*} \tag{5.12}$$

and the similar relations for the other amplitudes B, C and D.

The helicity amplitudes for the Bhabba scattering and for the e^-e^- scattering with polarized fermions are easily calculated in the lowest order of the perturbation expansion, including Z_0 exchange. The results are found in a previous

paper.[21] The explicit calculation confirms the symmetries (5.12).

6. THE FORM OF THE CROSS SECTION

The absolute square of the scattering amplitude yields an expression proportional to the cross section. This expression shows products of the superposition coefficients of which two at a time belong to the same fermion. These products of two coefficients belonging to the same particle have a simple physical meaning. They relate directly to the strength of the transversal and longitudinal polarization according to the formulas (3.14). The cross section for the reaction $e^-e^+\to(1)\to\mu^-\mu^+$ with polarized fermions, for instance, has the following form in the HEA:

$$
\begin{aligned}
&d\sigma_{e^-e^+\to(1)\to\mu^-\mu^+}\\
&\quad\propto |A^{\mathrm{I}}|^2(1+s_{1\parallel})(1-s_{2\parallel})(1-s_{3\parallel})(1+s_{4\parallel})\\
&\qquad+|A^{\mathrm{IV}}|^2(1-s_{1\parallel})(1+s_{2\parallel})(1+s_{3\parallel})(1-s_{4\parallel})\\
&\qquad+|A^{\mathrm{II}}|^2(1+s_{1\parallel})(1-s_{2\parallel})(1+s_{3\parallel})(1-s_{4\parallel})\\
&\qquad+|A^{\mathrm{III}}|^2(1-s_{1\parallel})(1+s_{2\parallel})(1-s_{3\parallel})(1+s_{4\parallel})\\
&\qquad+2\mathrm{Re}\{-[A^{\mathrm{I}}A^{\mathrm{III}*}(1-s_{3\parallel})(1+s_{4\parallel})\\
&\qquad+A^{\mathrm{II}}A^{\mathrm{IV}*}(1+s_{3\parallel})(1-s_{4\parallel})]s_{1\perp}s_{2\perp}e^{i(\psi_2-\psi_1)}\\
&\qquad-[A^{\mathrm{I}}A^{\mathrm{II}*}(1+s_{1\parallel})(1-s_{2\parallel})\\
&\qquad+A^{\mathrm{III}}A^{\mathrm{IV}*}(1-s_{1\parallel})(1+s_{2\parallel})]s_{3\perp}s_{4\perp}e^{i(\psi_4-\psi_3)}\\
&\qquad+A^{\mathrm{I}}A^{\mathrm{IV}*}s_{1\perp}s_{2\perp}s_{3\perp}s_{4\perp}e^{i(\psi_2-\psi_1-\psi_3+\psi_4)}\\
&\qquad+A^{\mathrm{II}}A^{\mathrm{III}*}s_{1\perp}s_{2\perp}s_{3\perp}s_{4\perp}e^{i(\psi_2-\psi_1+\psi_3-\psi_4)}\}. \qquad (6.1)
\end{aligned}
$$

We notice that several pairs of the dynamical factors A^iA^{k*} in this formula are identical except the terms which violates the parity and which differ by a sign. This is a consequence of the symmetry (5.12).

The formula (6.1) is correct also for partly polarized particles. In this case the degree of polarization is described by the length of the spin vector, which is now smaller than 1. If the polarization of the final particles is not observed, we have to put the spin vector equal zero and to multiply the rest of the cross section by 2 for each final particle.

The cross section of the e^-e^- scattering and of the Bhabba scattering has been calculated with this method and given elsewhere.[21] The contributions to the scattering of polarized electrons and positrons due to the Higgs particles are calculated in Ref. 20.

ACKNOWLEDGMENTS

The authors thank Professor A. O. Barut and Professor H. Salecker for many constructive and critical remarks, which led to several improvements of this paper.

APPENDIX A: SPINORS

The four-component spinors $u^Q(p,\hat{s})$ in Weyl's form[22] are composed by two associated two-component spinors (or van der Waerden spinors[23]) u_k and v^k

$$
u^Q(p,\hat{s})=\frac{1}{\sqrt{2}}\begin{pmatrix}u_k(p,Q\hat{s})\\ Qv^k(p,Q\hat{s})\end{pmatrix},
\qquad (A1)
$$

$$
\bar{u}^Q(p,\hat{s})=\frac{1}{\sqrt{2}}(Q\bar{v}^k(p,Q\hat{s}),\bar{u}_k(p,Q\hat{s})).
$$

The two-component spinors are normalized by

$$
\bar{v}^ku_k=1,\quad \bar{u}_kv^k=1. \qquad (A2)
$$

These normalization relations reduce the eight real parameters of u_k and v^k to 6. Three of them are used to describe the momentum, two describe the spin direction, and one remains for an arbitrary phase.

The charge quantum number Q, the momentum p^μ and the moving spin vector $\hat{s}^\mu$ of a particle state are represented by

$$
Q=\bar{u}^Q(p,\hat{s})u^Q(p,\hat{s}), \qquad (A3)
$$

$$
p^\mu=m\bar{u}^Q(p,\hat{s})\gamma^\mu u^Q(p,\hat{s}), \qquad (A4)
$$

$$
\hat{s}^\mu=-Q\bar{u}^Q(p,\hat{s})\gamma_5\gamma^\mu u^Q(p,\hat{s}). \qquad (A5)
$$

The representation of the γ matrices

$$
\gamma^\mu=\begin{pmatrix}0&\sigma_\mu\\ \sigma^\mu&0\end{pmatrix}=\begin{pmatrix}0&\sigma^{\mu}{}_{k\dot{l}}\\ \sigma^{\mu\dot{k}l}&0\end{pmatrix},
\qquad (A6)
$$

$$
\gamma_5=\begin{pmatrix}1&0\\ 0&-1\end{pmatrix}
$$

allow us to write

$$
p^\mu=\tfrac{1}{2}m[\bar{u}_k\sigma^{\mu\dot{k}l}u_l+\bar{v}^k\sigma^{\mu}{}_{k\dot{l}}v^{\dot{l}}], \qquad (A7)
$$

$$
\hat{s}^\mu=\tfrac{1}{2}Q[\bar{u}_k\sigma^{\mu\dot{k}l}u_l-\bar{v}^k\sigma^{\mu}{}_{k\dot{l}}v^{\dot{l}}]. \qquad (A8)
$$

We choose the direction $(\theta\phi)$ for the momentum, and $(\vartheta\varphi)$ for the spin vector in the rest frame of a particle

$$
\hat{p}=(\sin\theta\cos\phi,\ \sin\theta\sin\phi,\ \cos\theta), \qquad (A9)
$$

$$
\hat{s}=(\sin\vartheta\cos\varphi,\ \sin\vartheta\sin\varphi,\ \cos\vartheta). \qquad (A10)
$$

In the rest frame the two associated two-component spinors u_k and v^k are identical

$$
u_k(0,N\hat{s})=v^k(0,N\hat{s})\equiv|N,\hat{s}\rangle. \qquad (A11)
$$

The rest frame spinor for electrons ($Q=+1$) with spin parallel ($N=+1$) or antiparallel ($N=-1$) to the unit vector $\hat{s}$ has the components

$$
|+,\hat{s}\rangle=\begin{pmatrix}c\\ s\end{pmatrix},\quad |-,\hat{s}\rangle=\begin{pmatrix}-s^*\\ c\end{pmatrix}, \qquad (A12)
$$

where

$$
c=\cos(\vartheta/2),\quad s=\sin(\vartheta/2)e^{i\varphi}. \qquad (A13)
$$

The state with spin antiparallel to $\hat{s}$ and the state with spin parallel to $-\hat{s}$ are equal up to the phase factor $-e^{i\varphi}$.

The spinors for moving states are obtained by the application of a boost $B(p)$:

$$
\begin{aligned}
&u_k(p,N\hat{s})=B(p)u_k(0,N\hat{s})=B(p)|N,\hat{s}\rangle,\\
&v^k(p,N\hat{s})=B^{-1,\dagger}(p)v^k(0,N\hat{s})=B^{-1,\dagger}(p)|N,\hat{s}\rangle.
\end{aligned}
\qquad (A14)
$$

The boost $B(p)$ is composed of a unitary matrix

$$U(\theta\phi) = e^{-i(\phi/2)\sigma_z} e^{-i(\theta/2)\sigma_y} e^{i(\phi/2)\sigma_z} = \begin{pmatrix} C & -S^* \\ S & C \end{pmatrix} \tag{A15}$$

with

$$C = \cos(\theta/2), \quad S = \sin(\theta/2)e^{i\phi} \tag{A16}$$

and a Hermitian (and diagonal) matrix

$$B(\bar{p}) = \begin{pmatrix} \eta_+ & 0 \\ 0 & \eta_- \end{pmatrix} \tag{A17}$$

with

$$\bar{p} = (E, 0, 0, p) \tag{A18}$$

and

$$\eta_\pm = [(E \pm p)/m]^{1/2}. \tag{A19}$$

The unitary matrix U tips the direction $(\theta\phi)$ of the momentum into the z direction.[7,13] The Hermitian matrix $B(\bar{p})$ moves the rest frame along the z axis. By evaluating the Hermitian matrix

$$B(p) = U(\theta\phi)B(\bar{p})U^{-1}(\theta\phi) \tag{A20}$$

we find

$$B(p) = \sum_N |N, \hat{p}\rangle \eta_N \langle N, \hat{p}|,$$
$$B^{-1,\dagger}(p) = \sum_N |N, \hat{p}\rangle \eta_{-N} \langle N, \hat{p}|, \tag{A21}$$

and

$$[B(p)]^2 = \frac{1}{m} p^\mu \sigma^\mu, \quad [B^{-1,\dagger}(p)]^2 = \frac{1}{m} p^\mu \sigma_\mu. \tag{A22}$$

The comparison of (A22) and (A21) yields the relation (3.8). The four-component spinors for helicity states with $\hat{s} = N\hat{p}$ are, according to (A1), (A14), and (A21),

$$u_N^Q(p) = \frac{1}{\sqrt{2}} \begin{pmatrix} \eta_{QN} |QN, \hat{p}\rangle \\ Q\eta_{-QN} |QN, \hat{p}\rangle \end{pmatrix},$$
$$\bar{u}_N^Q(p) = \frac{1}{\sqrt{2}} (Q\eta_{-QN} \langle QN, \hat{p}|, \eta_{QN} \langle QN, \hat{p}|). \tag{A23}$$

The general four-component spinors are superpositions of the two helicity spinors

$$u^Q(p, \hat{s}) = \sum_N \alpha_N^Q u_N^Q. \tag{A24}$$

The coefficients of this expansion

$$\alpha_N^Q(\hat{p}, \hat{s}) = \langle QN, \hat{p} | Q, \hat{s} \rangle \tag{A25}$$

are independent on the energy.

By tipping the primary coordinate system around the axis $\vec{e}_z \times \hat{p}$ according to (A15) and the literature[7,13] arises a helicity frame with the basical vectors $\vec{e}_{x'}$, $\vec{e}_{y'}$, and $\vec{e}_{z'} = \hat{p}$. The spin vector (A10) is given with regard to this helicity frame by

$$\hat{s} = s_\perp \cos\psi \, \vec{e}_{x'} + s_\perp \sin\psi \, \vec{e}_{y'} + s_\parallel \hat{p}. \tag{A26}$$

The unit vectors perpendicular to the momentum $\hat{p}$ can be gathered by a complex vector

$$\vec{g} = \vec{e}_{x'} + i\vec{e}_{y'} = (C^2 - S^2, i(C^2 + S^2), -2CS). \tag{A27}$$

The following relation can be verified with (A12) and (A27):

$$\vec{g}\vec{\sigma} = 2|+, \hat{p}\rangle\langle -, \hat{p}|. \tag{A28}$$

This relation and the high energy limit of (3.8)

$$1 + Q\hat{p}\vec{\sigma} = 2|Q, \hat{p}\rangle\langle Q, \hat{p}| \tag{A29}$$

allow to derive now Eqs. (3.14).

According to (A8) and (A11) the spin vector in the rest frame is given by

$$\hat{s} = N\langle N, \hat{s}|\vec{\sigma}|N, \hat{s}\rangle. \tag{A30}$$

We find with (A29) and (A28) now

$$1 + N\hat{p}\hat{s} = 2\langle QN, \hat{s}|Q, \hat{p}\rangle\langle Q, \hat{p}|QN, \hat{s}\rangle,$$
$$N\vec{g}\hat{s} = 2\langle N, \hat{s}|+, \hat{p}\rangle\langle -, \hat{p}|N, \hat{s}\rangle. \tag{A31}$$

These relations lead to (3.14) with the help of (A25). The two-component rest frame spinors $|N_i, \vec{e}_i\rangle \equiv |N_i\rangle$ have interesting features. We state the relation

$$1\langle N_i|N_k\rangle = |N_k\rangle\langle N_i| + N_iN_k|-N_i\rangle\langle -N_k| \tag{A32}$$

from which several equations can be derived, for instance, the following relation between the spinor scalars:

$$\langle N_i|N_k\rangle\langle N_l|N_m\rangle = \langle N_i|N_m\rangle\langle N_l|N_k\rangle + N_iN_k\langle N_l|-N_i\rangle\langle -N_k|N_m\rangle. \tag{A33}$$

We derive also the symmetry relations between Pauli currents

$$\langle N_i|\sigma^\mu|N_k\rangle^* = \langle N_k|\sigma^\mu|N_i\rangle, \tag{A34}$$
$$\langle N_i|\sigma^\mu|N_k\rangle = N_iN_k\langle -N_k|\sigma_\mu|-N_i\rangle. \tag{A35}$$

The last two equations are easily generalized to Eqs. (4.1) and (4.2).

The Pauli matrices σ^μ can be expressed with the help of the two-component rest frame spinors for spin parallel or antiparallel to the z axis:

$$|+_0\rangle \equiv |+, \vec{e}_z\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |-_0\rangle \equiv |-, \vec{e}_z\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{A36}$$

We find

$$\sum_\mu \sigma^\mu \odot \sigma_\mu = 2\sum_{M,N} MN\,|N_0\rangle\langle M_0|\odot|-N_0\rangle\langle -M_0|, \tag{A37}$$

$$\sum_\mu \sigma^\mu \odot \sigma^\mu = 2\sum_{M,N} |N_0\rangle\langle M_0|\odot|M_0\rangle\langle N_0|. \tag{A38}$$

With these formulas we now treat the contraction products of the two Pauli currents $\langle N_1|a\sigma_\mu b\,|N_2\rangle\langle N_3|c\sigma^\mu d\,|N_4\rangle$ and $\langle N_1|a\sigma_\mu b\,|N_2\rangle\langle N_3|c\sigma_\mu d\,|N_4\rangle$. Here a, b, c, and d are arbitrary 2×2 matrices. With the help of (A32) the spinors $|N_0\rangle$ (A36) can be completely eliminated. The very general formulas (3.11) for the contraction of a product of two Pauli currents are obtained. Here we notice two useful specializations which can be derived with (A32):

$$\langle N_1|a\sigma_\mu|N_2\rangle\langle N_3|\sigma^\mu d\,|N_4\rangle = -2N_2N_3\langle N_1|a|-N_3\rangle\langle -N_2|d\,|N_4\rangle, \tag{A39}$$

$$\langle N_1|\sigma_\mu|N_2\rangle\langle N_3|c\sigma^\mu d\,|N_4\rangle = 2N_1N_2\langle N_3|c|-N_1\rangle\langle -N_2|d\,|N_4\rangle. \tag{A40}$$

APPENDIX B: PRODUCTS OF TWO HELICITY CURRENTS

Let A and B be arbitrary Dirac matrices. Now we define the product of two helicity currents

$$A \times B \equiv \bar{u}^{Q_1}_{N_1}(p_1) A u^{Q_2}_{N_2}(p_2) \bar{u}^{Q_3}_{N_3}(p_3) B u^{Q_4}_{N_4}(p_4). \tag{B1}$$

We introduce some abbreviations

$$M_j = Q_j N_j, \tag{B2}$$

$$\eta^{(j)}_{N_j} = [(E_j + N_j p_j)/m_j]^{1/2}, \tag{B3}$$

$$\epsilon_{\alpha_1 \cdots \alpha_n \| \beta} = \eta^{(1)}_{\alpha_1} \cdots \eta^{(n)}_{\alpha_n} + \beta \eta^{(1)}_{-\alpha_1} \cdots \eta^{(n)}_{-\alpha_n}, \tag{B4}$$

$$Q = Q_1 Q_2 Q_3 Q_4. \tag{B5}$$

Now let A and B mean even or odd products of matrices. There are four different types of current products. We introduce an operator G which changes all upper vector indices into lower indices and reverse. The four types of current products are according to (3.6a) and (3.6b):

$$\prod_i^{2n} \gamma^{\mu_i}(\gamma_5)^a \times \prod_j^{2k} \gamma^{\nu_j}(\gamma_5)^b = \tfrac{1}{4} Q_1 Q_3 \epsilon_{-M_1 M_2 - M_3 M_4 \| (-1)^{a+b} QG} \langle M_1 | \sigma_{\mu_1} \cdots^{\mu_{2n}} | M_2 \rangle \langle M_3 | \sigma_{\nu_1} \cdots^{\nu_{2k}} | M_4 \rangle$$
$$+ (-1)^b \cdot \tfrac{1}{4} Q_1 Q_4 \epsilon_{-M_1 M_2 M_3 - M_4 \| (-1)^{a+b} QG} \langle M_1 | \sigma_{\mu_1} \cdots^{\mu_{2n}} | M_2 \rangle \langle M_3 | \sigma^{\nu_1} \cdots_{\nu_{2k}} | M_4 \rangle, \tag{B6}$$

$$\prod_i^{2n+1} \gamma^{\mu_i}(\gamma_5)^a \times \prod_j^{2k+1} \gamma^{\nu_j}(\gamma_5)^b = \tfrac{1}{4} \epsilon_{M_1 M_2 M_3 M_4 \| (-1)^{a+b} QG} \langle M_1 | \sigma^{\mu_1} \cdots^{\mu_{2n+1}} | M_2 \rangle \langle M_3 | \sigma^{\nu_1} \cdots^{\nu_{2k+1}} | M_4 \rangle$$
$$+ \tfrac{1}{4} (-1)^b Q_3 Q_4 \epsilon_{M_1 M_2 - M_3 - M_4 \| (-1)^{a+b} QG} \langle M_1 | \sigma^{\mu_1} \cdots^{\mu_{2n+1}} | M_2 \rangle \langle M_3 | \sigma_{\nu_1} \cdots_{\nu_{2k+1}} | M_4 \rangle, \tag{B7}$$

$$\prod_i^{2n+1} \gamma^{\mu_i}(\gamma_5)^a \times \prod_j^{2k} \gamma^{\nu_j}(\gamma_5)^b = \tfrac{1}{4} Q_3 \epsilon_{M_1 M_2 - M_3 M_4 \| (-1)^{a+b} QG} \langle M_1 | \sigma^{\mu_1} \cdots^{\mu_{2n+1}} | M_2 \rangle \langle M_3 | \sigma_{\nu_1} \cdots^{\nu_{2k}} | M_4 \rangle$$
$$+ \tfrac{1}{4} (-1)^b Q_4 \epsilon_{M_1 M_2 M_3 - M_4 \| (-1)^{a+b} QG} \langle M_1 | \sigma^{\mu_1} \cdots^{\mu_{2n+1}} | M_2 \rangle \langle M_3 | \sigma^{\nu_1} \cdots_{\nu_{2k}} | M_4 \rangle, \tag{B8}$$

$$\prod_i^{2n} \gamma^{\mu_i}(\gamma_5)^a \times \prod_j^{2k+1} \gamma^{\nu_j}(\gamma_5)^b = \tfrac{1}{4} Q_1 \epsilon_{-M_1 M_2 M_3 M_4 \| (-1)^{a+b} QG} \langle M_1 | \sigma_{\mu_1} \cdots^{\mu_{2n}} | M_2 \rangle \langle M_3 | \sigma^{\nu_1} \cdots^{\nu_{2k+1}} | M_4 \rangle$$
$$+ \tfrac{1}{4} (-1)^b Q_1 Q_3 Q_4 \epsilon_{-M_1 M_2 - M_3 - M_4 \| (-1)^{a+b} QG} \langle M_1 | \sigma_{\mu_1} \cdots^{\mu_{2n}} | M_2 \rangle \langle M_3 | \sigma_{\nu_1} \cdots_{\nu_{2k+1}} | M_4 \rangle. \tag{B9}$$

The specialization of these formulas and the application of (3.11) for the contraction of Pauli currents produce all current products required in the calculation of scattering amplitudes. We notice some of these as examples:

$$1 \times 1 = \tfrac{1}{4} Q_1 Q_2 \epsilon_{-M_1 M_2 - M_3 M_4 \| Q} \langle M_1 | M_2 \rangle \langle M_3 | M_4 \rangle$$
$$+ \tfrac{1}{4} Q_1 Q_4 \epsilon_{-M_1 M_2 M_3 - M_4 \| Q} \langle M_1 | M_2 \rangle \langle M_3 | M_4 \rangle, \tag{B10}$$

$$\gamma^\mu \times \gamma^\nu = \tfrac{1}{4} \epsilon_{M_1 M_2 M_3 M_4 \| QG} \langle M_1 | \sigma^\mu | M_2 \rangle \langle M_3 | \sigma^\nu | M_4 \rangle$$
$$+ \tfrac{1}{4} Q_3 Q_4 \epsilon_{M_1 M_2 - M_3 - M_4 \| QG} \langle M_1 | \sigma^\mu | M_2 \rangle \langle M_3 | \sigma_v | M_4 \rangle, \tag{B11}$$

$$\gamma_\mu \times \gamma^\mu = \tfrac{1}{4} \epsilon_{M_1 M_2 M_3 M_4 \| Q} \cdot (-2) M_2 M_3 \langle M_1 | -M_3 \rangle \langle -M_2 | M_4 \rangle$$
$$+ \tfrac{1}{4} Q_3 Q_4 \epsilon_{M_1 M_2 - M_3 - M_4 \| Q} \cdot 2 \langle M_1 | M_4 \rangle \langle M_3 | M_2 \rangle, \tag{B12}$$

$$\gamma^\mu \gamma^\nu \times \gamma^\rho \gamma^\tau = \tfrac{1}{4} Q_1 Q_3 \epsilon_{-M_1 M_2 - M_3 M_4 \| QG} \langle M_1 | \sigma_\mu \sigma^\nu | M_2 \rangle \langle M_3 | \sigma_\rho \sigma^\tau | M_4 \rangle$$
$$+ \tfrac{1}{4} Q_1 Q_4 \epsilon_{-M_1 M_2 M_3 - M_4 \| QG} \langle M_1 | \sigma_\mu \sigma^\nu | M_2 \rangle \langle M_3 | \sigma^\rho \sigma_\tau | M_4 \rangle, \tag{B13}$$

$$\gamma_\mu \gamma^\nu \times \gamma^\rho \gamma^\mu = \tfrac{1}{4} Q_1 Q_3 \epsilon_{-M_1 M_2 - M_3 M_4 \| QG} \cdot 2 \langle M_1 | M_4 \rangle \langle M_3 | \sigma_\rho \sigma^\nu | M_2 \rangle$$
$$+ \tfrac{1}{4} Q_1 Q_4 \epsilon_{-M_1 M_2 M_3 - M_4 \| QG} \cdot 2 M_3 M_4 \langle M_1 | \sigma_\rho | -M_3 \rangle \langle -M_4 | \sigma^\nu | M_2 \rangle, \tag{B14}$$

$$\gamma_\mu \gamma_\nu \times \gamma^\nu \gamma^\mu = \tfrac{1}{4} Q_1 Q_3 \epsilon_{-M_1 M_2 - M_3 M_4 \| Q} \cdot 8 \langle M_1 | M_4 \rangle \langle M_3 | M_2 \rangle$$
$$+ \tfrac{1}{4} Q_1 Q_4 \epsilon_{-M_1 M_2 M_3 - M_4 \| Q} \cdot 4 \langle M_1 | M_2 \rangle \langle M_3 | M_4 \rangle. \tag{B15}$$

Not all the vector indices of the currents are contracted internally. The remaining Pauli currents $\langle A | \sigma^\mu \sigma_\nu | B \rangle$, etc., of the amplitudes disappear by contractions with external momenta. In these cases the formulas (3.8) and (3.10) apply.

[1] J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics* (McGraw-Hill, New York, 1964), or any other textbook.
[2] L. Michel and A. S. Wightman, Phys. Rev. **98**, 1190 (1955).
[3] L. D. Landau and E. M. Lifschitz, *Lehrbuch der Theoretischen Physik Bd. IV* (Akademie-Verlag, Berlin, 1980).
[4] M. Carrassi and G. Passatore, Nuovo Cimento **13**, 944 (1959).
[5] W. Jachmann, "Explizite Berechnung von Streuamplituden und Wirkungsquerschnitten für polarisierte Dirac-Teilchen und Photonen," internal report, Sektion Physik, Universität München, 1980.
[6] J. G. Körner and D. H. Schiller, International Symposium on Lepton and Photon Interactions at High Energies, Bonn, 1981, p. 233.
[7] A. O. Barut, *The Theory of the Scattering Matrix* (MacMillan, New York, 1967).
[8] A. O. Barut, S. A. Baran, and B. C. Ünal, Fortschr. Phys. **21**, 609 (1973).

[9]M. Hofri and A. Peres, Nucl. Phys. **59**, 618 (1964).
[10]Yu. S. Perov, Sov. Phys. J. **16**, 1719 (1973).
[11]J. S. R. Chisholm, Proc. Cambridge Phil. Soc. **48**, 300 (1952); Nuovo Cimento **30**, 426 (1963).
[12]J. Kahane, J. Math. Phys. **9**, 1732 (1968).
[13]M. Jacob and G. C. Wick, Ann. Phys. **7**, 404 (1959).
[14]P. R. Auvil and J. J. Brehm, Phys. Rev. **145**, 1152 (1966).
[15]Special contraction formulas have been elaborated in the literature, e.g., Ref. 8, Eq. (3.5).
[16]H. A. Tolhoek, Rev. Mod. Phys. **28**, 277 (1956).
[17]Y. S. Tsai, Phys. Rev. **120**, 269 (1960).
[18]R. V. Mellenthin, "Radiative Corrections to Polarized e^+e^--Scattering," dissertation, Universität München, 1983, in progress.
[19]The notations of the terms of the helicity amplitudes are chosen so that the particles with indices at the nth position right form a current with the particle having its indices at the nth position left. The currents run from right to left. Upper signs mean the charges Q; lower signs, the helicities N.
[20]K. Wadan, "e^+e^--Scattering with Electro-Weak Interaction," Dissertation, Universität München, 1983, in progress.
[21]T. Anders, W. Jachmann, H. Salecker, and K. Wadan, in *Old and New Questions in Physics, Cosmology, Philosophy, and Theoretical Biology: Essays in Honor of Wolfgang Yourgrau*, edited by A. van der Merve (Plenum, New York, 1983), p. 411. (The paper was submitted to the International Symposium on Lepton and Photon Interactions at High Energies, Bonn, August 1981.)
[22]The dotted and undotted spinor indices have the same meaning as in Ref. 7. Few differences may be mentioned. We write simply $\sigma_\mu \equiv \sigma_\mu{}^{kl} = (1, -\vec{\sigma})$ instead of $\tilde{\sigma}_\mu \equiv \tilde{\sigma}_\mu{}^{kl} = (1, -\vec{\sigma})$. Similarly we write $\sigma^\mu \equiv \sigma_{\mu kl} = (1, \vec{\sigma})$ instead of $\sigma_\mu \equiv \sigma_{\mu kl} = (1, \vec{\sigma})$. Besides this, our positron states ($Q = -1$) for negative helicity have a reversed sign in comparison with what is usual.[1,7] The phase convention for the positron states in accordance with the usual charge conjugation $v = i\gamma^2\gamma^0(\bar{u})^T$ would cause complication of sign factors of the type $Q^{(1+QN)/2}$ in the relations (3.6) and (3.13) for the helicity currents and the expansion coefficients.
[23]B. L. van der Waerden, *Die Gruppentheoretische Methode in der Quantenmechanik* (Springer-Verlag, Berlin, 1932).

Quasiclassical trajectory-coherent states of a particle in an arbitrary electromagnetic field

V. G. Bagrov, V. V. Belov, and I. M. Ternov
Institute of High Current Electronics of the USSR Academy of Sciences Siberian Branch, pr. Akademichesky, 4, 634055, Tomsk, USSR

(Received 18 June 1981; accepted for publication 12 November 1982)

In this paper we show that for a nonrelativistic charged particle moving in an arbitrary external electromagnetic field there exist approximate solutions of the Schrödinger equation, such that the quantum-mechanical averages of the coordinates and the momenta with respect to these states are general exact solutions of the classical Hamiltonian equations. Such states are called trajectory-coherent states. The wave functions of the trajectory-coherent states are obtained by the complex germ method by V. P. Maslov. The simplest properties of these states are studied.

PACS numbers: 41.70. + t, 03.65.Ge

1. INTRODUCTION

The question of the correspondence between the results of classical and quantum mechanics is as old as classical mechanics itself. It is quite obvious that there is no universal method of obtaining arbitrary classical quantities from quantum-mechanical ones. Let us explain this in more detail. Assume that one wants, for the conservative system, to obtain a classical expression for energy (as a function of any parameters of the system), starting from a quantum-mechanical one. It is obvious that in this case one should determine the stationary states of the system and then, by the definite way, take the limit as $\hbar \to 0$ of the quantum-mechanical expression for the energy. On the other hand, if one is interested in obtaining the classical particle trajectory as a limit for $\hbar \to 0$ of some quantum-mechanical expression, the stationary states cannot lead to a reasonable classical result because for such states the coordinate and momentum averages do not depend on time at all. Hence, to derive the classical trajectory as a limit for $\hbar \to 0$ of a certain quantum-mechanical expression one should first of all take care of this problem, and determine an appropriate choice of the corresponding quantum-mechanical states. Thus, the mode of construction (and the existence of this mode itself) of quantum-mechanical states that are "close" to classical ones depends on the very classical and quantum-mechanical quantities we want proximity for (or coincidence, if possible).

Schrödinger[1] introduced an example of this kind for the harmonic oscillator: he constructed a set of states solving the Schrödinger equation for which the quantum-mechanical averages of the coordinates and the momentum are general solutions of the classical equations of motion. Glauber[2] discovered the same states for free electromagnetic field; in these states the quantum-mechanical average of the potential operator is a general solution of the Maxwell free equations. (The coherent states of electromagnetic field already appeared in Ref. 3; they were studied in Ref. 4.) Such states were called coherent states. These states are constructed in a quite simple way for quantum systems whose effective Hamiltonian is quadratic by the coordinate and momentum operators. The theory of these states has been developed in detail. (The most complete statement of this theory is presented in Ref. 5.)

At present the coherent state sequence is assumed to be determined as a complete system of quantum-mechanical states that are integrals of motion. (The discussion of this determination may be found in Refs. 6 and 7.) A quite reasonable conviction has developed that coherent states, in a certain sense, are the closest to classical. In particular, for the quadratic systems the coordinate and momentum quantum-mechanical averages are solutions of the corresponding classical Hamiltonian equations, and for the case of quadratic systems with constant coefficients the coherent states attain the minimum of Heisenberg uncertainty relations.

However, it follows from the Ehrenfest theorem that if the Hamiltonian operator of the system is not quadratic there are no states in which the mean quantum-mechanical trajectory would exactly coincide with a classical one. Thus, the question of constructing quantum-mechanical states giving mean trajectories that coincide with classical ones in the limit $\hbar \to 0$ is not a trivial one. Such states, if any, are called trajectory-coherent states (TCS).

The aim of this work is to prove the following statement: for a nonrelativistic charged particle moving in an arbitrary electromagnetic field there always exist approximate solutions of the Schrödinger equation, such that the coordinate and momentum quantum-mechanical averages are exact solutions of the corresponding classical nonrelativistic Hamiltonian equations. These solutions are constructed explicitly and an estimate of their accuracy is obtained. In other words, we construct approximate TCS which give an exact classical trajectory. In addition, we generalize the results obtained to the case of a nonrelativistic particle described by the Klein–Gordon equation.

The basis of our construction is the complex WKB method by V. P. Maslov that is the complex-germ theory.[8–10]

2. NOTATION

In order to avoid overcrowding further statements, we start by describing the basic notations that will be used below. We denote the coordinate and momentum vectors by $\mathbf{X}$ (X_1, X_2, X_3), $\mathbf{P}$ (P_1, P_2, P_3). In quantum theory the operator $\hat{\mathbf{p}}$ is given by $\hat{\mathbf{p}} = -ih\nabla$, $\nabla_i = \partial_{x_i}$. The partial derivatives are designated as $\partial_{x_i} = \partial/\partial x_i$, $\partial_{p_i} = \partial/\partial p_i$. The classical Hamiltonian of the system is

 0022-2488/83/122855-05$02.50

$$H(\mathbf{x}, \mathbf{p}) = (2m)^{-1}(\mathbf{p} - \mathbf{A})^2 + \phi. \tag{1}$$

The particle charge e and the velocity of light are included in the electromagnetic potentials $\mathbf{A} = \mathbf{A}(\mathbf{x}, t)$, $\phi = \phi(\mathbf{x}, t)$ which are taken in the Coulomb gauge div $A = 0$. The operator $\hat{H}(\mathbf{X}, \mathbf{P})$ is obtained from (1) by substitutions $\mathbf{P} \to \hat{\mathbf{P}}$. The time functions $X(t)$ and $P(t)$ are solutions of the classical equations

$$\dot{\mathbf{x}}(t) = \partial_{\mathbf{p}} H(\mathbf{x}, \mathbf{p}), \quad \dot{\mathbf{p}}(t) = -\partial_{\mathbf{x}} H(\mathbf{x}, p). \tag{2}$$

We suppose everywhere that the general solution, $\mathbf{X}(t)$ and $\mathbf{P}(t)$, of Eqs. (2) is known. The Schrödinger equation has the form

$$\hat{L}\Psi = 0, \quad \hat{L} = \hat{H}(\mathbf{x}, \mathbf{p}) - i\hbar\partial_t. \tag{3}$$

Furthermore we introduce the 6×6 matrix $\mathscr{H}(\mathbf{X}, \mathbf{P})$, written in the form of four 3×3 blocks

$$\mathscr{H}(\mathbf{x}, \mathbf{p}) = \begin{vmatrix} -H_{xp}(\mathbf{x}, \mathbf{p}) & -H_{xx}(\mathbf{x}, \mathbf{p}) \\ H_{pp}(\mathbf{x}, \mathbf{p}) & H_{px}(\mathbf{x}, \mathbf{p}) \end{vmatrix}, \tag{4}$$

where, for example, the blocks $H_{xx}(\mathbf{x}, \mathbf{p})$ and $H_{xp}(\mathbf{x}, \mathbf{p})$ are defined as

$$(H_{xx}(\mathbf{x}, \mathbf{p}))_{ij} = \partial_{x_i}\partial_{x_j} H(\mathbf{x}, \mathbf{p}), \quad (H_{xp}(\mathbf{x}, \mathbf{p}))_{ij} = \partial_{x_i}\partial_{p_j} H(\mathbf{x}, \mathbf{p}). \tag{5}$$

In the case of the Hamiltonian (1), it is easy to find

$$\begin{aligned} &(mH_{xp}(\mathbf{x}, \mathbf{p}))_{ij} = -\partial_{x_i} A_j, \quad (mH_{pp}(\mathbf{x}, \mathbf{p}))_{ij} = \delta_{ij}, \\ &(H_{xx}(\mathbf{x}, \mathbf{p}))_{ij} = \partial_{x_i}\partial_{x_j}\phi + m^{-1}\langle \partial_{x_i}\mathbf{A}, \partial_{x_j}\mathbf{A}\rangle \\ &\qquad - \langle \mathbf{x}(t), \partial_{x_i}\partial_{x_j}\mathbf{A}\rangle. \end{aligned} \tag{6}$$

The scalar product of three-dimensional vector is designated by French quotes $\langle\ \rangle$. If in some expression dependent on $\mathbf{X}$ and $\mathbf{P}$, after performing all the operations (e.g., after taking partial derivatives with respect to $\mathbf{x}$ and $\mathbf{p}$), the substitution of $\mathbf{X} = \mathbf{X}(t)$ and $\mathbf{P} = \mathbf{P}(t)$ is made, this expression will be designated by the same letter with the argument t [e.g., $H(t) = H(\mathbf{x}, \mathbf{p})$ with $\mathbf{X} = \mathbf{X}(t)$, $\mathbf{P} = \mathbf{P}(t)$].

The expression $R\mathbf{x}$, where R is a 3×3 matrix, $\mathbf{x}$ is a vector, designates the vector $\mathbf{y} = R\mathbf{x}$ with the components

$$y_i = \sum_{k=1}^{3} R_{ik} x_k. \tag{7}$$

3. THE VACUUM TCS

Introduce 3×3 matrices $B(t)$ and $C(t)$ that are the solutions of the system of differential equations

$$\begin{pmatrix} \dot{B} \\ \dot{C} \end{pmatrix} = \mathscr{H}(t) \begin{pmatrix} B \\ C \end{pmatrix}, \quad \begin{aligned} &B(0) = \operatorname{diag}(b_1, b_2, b_3), \\ &C(0) = E, \end{aligned} \tag{8}$$

where E is the unit 3×3 matrix, and b_i are complex numbers obeying the condition

$$\operatorname{Im} b_i > 0, \quad i = 1, 2, 3. \tag{9}$$

A simple argument borrowed from Ref. 9 (see also Ref. 11) shows that $C(t)$ in nonsingular: explicit calculation, using the special form (4) of $H(t)$, shows that $(C^+B - B^+C) = 0$, $\Rightarrow \forall x \in R^3$, $x \neq 0$:

$$(Cx, Bx) - (Bx, Cx) = (x, B(0)x) - (B(0)x, x) = 2i \sum_{k=1}^{3} \operatorname{Im} b_k |x_k|^2 \neq 0 \Rightarrow Cx \neq 0.$$

Analogous arguments show that $Q(t) = B(t)C^{-1}(t)$ is symmetric (i.e., equal to its transposed matrix and that

$$\operatorname{Im} Q(t) > 0. \tag{10}$$

For the matrix $Q(t)$ it is not difficult to obtain the equation

$$\dot{Q} + QH_{pp}(t)Q + QH_{px}(t) + H_{xp}(t)Q + H_{xx}(t) = 0 \tag{11}$$

and the equality

$$m\dot{I} = I \operatorname{tr} Q, \quad I = I(t) = \det C(t). \tag{12}$$

Equation (12) follows from the Lewill theorem[8] taking into account Eqs. (11) and (1).

We introduce also the function of WKB-solution type

$$\Psi_0(\mathbf{x}, t, \hbar) = N(\hbar)\Phi(t) \exp[i\hbar^{-1}S(\mathbf{x}, t)], \tag{13}$$

where the phase $S(\mathbf{x}, t)$ is the complex-valued function [in the standard WKB method (see e.g., Ref. 12) the phase is real valued].

$$\begin{aligned} S(\mathbf{x}, t) = \int_0^t &[\langle \mathbf{p}(t), \dot{\mathbf{x}}(t)\rangle - H(t)]dt \\ &+ \langle \mathbf{p}(t), \mathbf{x} - \mathbf{x}(t)\rangle \\ &+ \tfrac{1}{2}\langle \mathbf{x} - \mathbf{x}(t), Q(t)(\mathbf{x} - \mathbf{x}(t))\rangle \end{aligned} \tag{14}$$

while the amplitude and normalization are

$$\begin{aligned} &\Phi(t) = [I(t)]^{-1/2}, \\ &N(\hbar) = [(\pi\hbar)^{-3} \operatorname{Im} b_1 \operatorname{Im} b_2 \operatorname{Im} b_3]^{1/4}. \end{aligned} \tag{15}$$

On account of (10) we have $\operatorname{Im} S(\mathbf{x}, t) > 0$ for the phase $S(\mathbf{x}, t)$ and therefore the function (13) decreases exponentially in $|\mathbf{x}|^2$, at the fixed t, for $|\mathbf{x}| \to \infty$ ($|\mathbf{x}| = \sqrt{\langle \mathbf{x}, \mathbf{x}\rangle}$, $\mathbf{x} \in R^3$).

The function (13) is an approximate solution of the Schrödinger equation in the limit $\hbar \to 0$. The estimation of accuracy of this approximation is defined by the following theorem.

Theorem: Let the potentials $\mathbf{A}$, ϕ be smooth functions tending to zero as $|\mathbf{x}| \to \infty$, together with all their derivatives. Then

$$\hat{L}\Psi_0(\mathbf{x}, t, \hbar) = f_0(\mathbf{x}, t, \hbar) \tag{16}$$

with, in the limit $\hbar \to 0$, the following bound on the L^2-norm of $f_0(\mathbf{x}, t, h)$:

$$\|f_0(\mathbf{x}, t, \hbar)\|^2 \leqslant \hbar^3(C_1 + \hbar^2 C_2), \quad 0 \leqslant t \leqslant T,$$

$$\begin{aligned} C_1 = \max_t \max_{ijk} \{ &|\partial_{x_i}\partial_{x_j}\mathbf{A}(t)|, |\partial_{x_i}\partial_{x_j}\partial_{x_k}\mathbf{A}(t)|, \\ &|\partial_{x_i}\partial_{x_j}\partial_{x_k}\phi(t)|, |Q(t)|\}, \end{aligned} \tag{17}$$

$$C_2 = \max_t \max_{ijkl} \{ |\partial_{x_i}\partial_{x_j}\partial_{x_k}\partial_{p_l} H(t)|, |\partial_{x_i}\partial_{x_j}\partial_{x_k}\partial_{x_l} H(t)|\}.$$

There $|Q(t)|$ is a norm of the matrix $Q(t)$.

We now proceed to prove this theorem. We apply $\hat{L}$ [see Eq. (3)] to $\Psi_0(\mathbf{x}, t, h)$ and we obtain thus the equality

$$\begin{aligned} \hat{L}\Phi(t) \exp[i\hbar^{-1}S(\mathbf{x}, t)] = \exp[i\hbar^{-1}S(\mathbf{x}, t)]\{ &\Phi(t)[\partial_t S \\ &+ H(\mathbf{x}, \nabla S)] - i\hbar\hat{\tau}\Phi(t)\}, \end{aligned} \tag{18}$$

where $\hat{\pi}$ is the transfer operator

$$\hat{\pi} = \partial_t + m^{-1}(\langle \nabla S - \mathbf{A}, \nabla \rangle + \tfrac{1}{2}\Delta S - i\hbar\Delta). \tag{19}$$

We now expand the right side of the formula (18) into its Taylor series in the neighborhood of the point $\mathbf{X}(t)$ [i.e., in powers of $X_i - X_i(t)$] up to the 4th power inclusive, making use of the symmetry of the matrix $Q(t)$, of the Eqs. (11), (12), and of the equalities following from (6), (14)

$$\nabla S = \mathbf{p}(t) + Q(t)(\mathbf{x} - \mathbf{x}(t)),$$

$$\partial_t S = -H(t) + \langle \dot{\mathbf{p}}(t), \mathbf{x} - \mathbf{x}(t)\rangle - \langle \dot{\mathbf{x}}(t), Q(t)(\mathbf{x} - \mathbf{x}(t))\rangle + \tfrac{1}{2}\langle \mathbf{x} - \mathbf{x}(t), \dot{Q}(t)(\mathbf{x} - \mathbf{x}(t))\rangle. \tag{20}$$

As a result we have

$$f_0\langle \mathbf{x}, t, \hbar\rangle = (6m)^{-1}\Psi_0(\mathbf{x}, t, \hbar)\Big[\sum_{ijk} H_{ijk}(t)(x_i - x_i(t))(x_j - x_j(t))(x_k - x_k(t)) + \langle \mathbf{x} - \mathbf{x}(t), D(\mathbf{x} - \mathbf{x}(t))\rangle + R_4(\mathbf{x}, t)\Big],$$

$$H_{ijk}(t) = \langle \partial_{x_i}\mathbf{A}(t), \partial_{x_j}\partial_{x_k}\mathbf{A}(t)\rangle + \langle \partial_{x_j}\mathbf{A}(t), \partial_{x_i}\partial_{x_k}\mathbf{A}(t)\rangle + \langle \partial_{x_k}\mathbf{A}(t), \partial_{x_i}\partial_{x_j}\mathbf{A}(t)\rangle + \sum_{n_i} \partial_{x_i}^{n_1}\partial_{x_j}^{n_2}\partial_{x_k}^{n_3} \times [m\phi(t) - \langle \dot{\mathbf{x}}(t), \mathbf{A}(t)\rangle], \tag{21}$$

where the sum over the n_i is a sum over all sets of integer nonnegative numbers n_i $(i = 1, 2, 3)$ satisfying the condition $n_1 + n_2 + n_3 = 3$. The elements of the matrix D are determined by the relation

$$D_{ik} = \langle Q(t)(\mathbf{x} - \mathbf{x}(t)), \quad \partial_{x_i}\partial_{x_k}\mathbf{A}(t)\rangle. \tag{22}$$

The remainder term R_4 of the Taylor series is bounded by

$$|R_4(\mathbf{x}, t)| \leqslant C_3|\mathbf{x} - \mathbf{x}(t)|^4. \tag{23}$$

By integrating the square of the modulus of the expression (21) over the space R^3, taking into account the trivial integral

$$\int_{-\infty}^{\infty} x^{2n}\exp(-x^2/\hbar)\,dx = \hbar^{n+1/2}\Gamma(n + \tfrac{1}{2}) \tag{24}$$

and using smoothness of the potentials $\mathbf{A}$, ϕ and the Koshi–Buniakovsky inequality, the proof of the theorem is obtained.

In conclusion we note that analogous results were obtained in Ref. 14 for a particular case of $\mathbf{A} = 0$, $\phi = \phi(\mathbf{x})$.

4. EXCITED TCS

Let us denote the three-dimensional complex vectors that are the columns of the matrices $C(t)$ and $B(t)$ by $Z_i(t)$ and $W_i(t)$ $(i = 1, 2, 3)$, respectively. We introduce the "generalized" annihilation and creation operators

$$\hat{a}_i(t) = (2\hbar\,\mathrm{Im}\,b_i)^{-1/2}(\langle \mathbf{z}_i(t), \hat{\mathbf{p}}\rangle - \langle \mathbf{W}_i(t), \mathbf{x}\rangle),$$
$$\hat{a}_i^+(t) = (2\hbar\,\mathrm{Im}\,b_i)^{-1/2}(\langle \mathbf{z}_i^*(t), \hat{\mathbf{p}}\rangle - \langle \mathbf{W}_i^*(t), \mathbf{x}\rangle). \tag{25}$$

The basic properties of these operators

$$[\hat{a}_i, \hat{a}_j^+] = \delta_{ij}, \quad [\hat{a}_i, a_j] = [\hat{a}_i^+, \hat{a}_j^+] = 0 \tag{26}$$

can be checked by direct calculation. In the same simple way it is found that

$$\hat{a}_i(t)\Psi_0(\mathbf{x}, t, \hbar) = \alpha_i(t)\Psi_0(\mathbf{x}, t, \hbar), \tag{27}$$

where the functions $\alpha_i(t)$, which are eigenvalues of the operators $a_i(t)$, are the simplectical scalar products[13]

$$\alpha_i(t) = (2\hbar\,\mathrm{Im}\,b_i)^{-1/2}(\langle \mathbf{Z}_i(t), \mathbf{p}(t)\rangle - \langle \mathbf{W}_i(t), \mathbf{x}(t)\rangle). \tag{28}$$

The statement that the operators $\hat{a}_i(t)$ and $\hat{a}_i^+(t)$ are symmetries of the Schrödinger equation (3) in the limit of $\hbar \to 0$ is somewhat less obvious. We shall take this statement to be true if the commutational relationships

$$[\hat{a}_i(t), L]\Psi = O(\hbar^{3/2}), \quad [\hat{a}_i^+, L]\Psi = O(\hbar^{3/2}) \tag{29}$$

are satisfied over the set of approximate solutions of the Schrödinger equation

$$\Psi(\mathbf{x}, t, \hbar) = \phi(\mathbf{x}, t, \hbar)\exp[i\hbar^{-1}S(\mathbf{x}, t)], \tag{30}$$

where $\phi(\mathbf{x}, t, h) \in P$, P the set of polynominals in $\hbar^{-1/2}$ $[\mathbf{x} - \mathbf{x}(t)]$ with time-dependent coefficients.

From Eq. (18) it is easy to obtain

$$\exp[-i\hbar^{-1}S(\mathbf{x}, t)]\hat{L}\exp[i\hbar^{-1}S(\mathbf{x}, t)] = -i\hbar\hat{\pi} + \hat{O}(\hbar^{3/2}), \tag{31}$$

where $\hat{\pi}$ is defined by (19). Here and below we use the notation $\hat{O}(\hbar^{\alpha})$ for any operator which, when applied to $\Phi \in P$, yields a vector with L^2-norm $O(\hbar^{\alpha})$ when multiplied by $\exp[i\hbar^{-1}S(\mathbf{x}, t)]$:

$$\forall \Phi \in P: \|\exp[i\hbar^{-1}S(\mathbf{x}, t)\hat{O}(\hbar^{\alpha})\Phi\| = O(\hbar^{\alpha}).$$

Further we have

$$\exp[-i\hbar^{-1}S(\mathbf{x}, t)]\hat{a}_i(t)\exp[i\hbar^{-1}S(\mathbf{x}, t)] = \hat{A}_i(t) + \alpha_i(t),$$
$$\exp[-i\hbar^{-1}S(\mathbf{x}, t)]\hat{a}_i^+(t)\exp[i\hbar^{-1}S(\mathbf{x}, t)] = \hat{A}_i^+(t) + \alpha_i^*(t). \tag{32}$$

Here $\alpha_i(t)$ is defined by the formula (28), while

$$\hat{A}_i(t) = (2\hbar\,\mathrm{Im}\,b_i)^{-1/2}\langle \mathbf{Z}_i(t), \hat{\mathbf{p}}\rangle,$$
$$\hat{A}_i^+(t) = (2\hbar\,\mathrm{Im}\,b_i)^{-1/2}(\langle \mathbf{Z}_i^*(t), \hat{\mathbf{p}}\rangle - \langle \mathbf{W}_i^*(t) - Q(t)\mathbf{Z}_i^*(t), \mathbf{x} - \mathbf{x}(t)\rangle). \tag{33}$$

As an example, we calculate the commutator $[\hat{a}_i^+(t) - \alpha_i^*(t), \hat{L}]$, applied to the functions (30). Using (31)–(33) we have

$$[\hat{a}_i^+(t) - \alpha_i^*(t), \hat{L}]\Psi = \exp[i\hbar^{-1}S(\mathbf{x}, t)]\{-i\hbar[\hat{A}_i(t), \hat{\pi}] + [\hat{A}_i^+(t), O(\hbar^{3/2})]\}\phi(\mathbf{x}, t, \hbar)\}. \tag{34}$$

The class of the functions $\phi(\mathbf{x}, t, \hbar) \in P$ is invariant under the operators $\hat{A}(t)$, $\hat{A}^+(t)$. Hence, it follows

$$[\hat{A}_i^+(t), O(\hbar^{3/2})] = \hat{O}(\hbar^{3/2}).$$

Hence, it would be enough to show that the commutator $[\hat{A}_i(t), \hat{\pi}]$ is an operator of type $\hat{O}(\hbar^{1/2})$. By expanding this operator in a power series of $x - x(t)$ up to the 3rd term inclusive, using (11), (12), and (20) we have required statement.

From the property (29) and the invariance of the P-class relative to the operators $\hat{a}_i^+(t)$, $\hat{A}_i^+(t)$ it follows that the functions

$$\Psi_{n_i}(\mathbf{x}, t, \hbar) = \prod_{i=1}^{3}(n_i!)^{-1/2}[\hat{a}_i(t) - \alpha_i^*(t)]^{n_i}\Psi_0(\mathbf{x}, t, \hbar) \tag{35}$$

are approximate solutions of the Schrödinger equation with respect to mod $O(h^{3/2})$.

5. BASIC PROPERTIES OF TCS

1. The set of the functions $\Psi_{n_i}(\mathbf{x}, t, \hbar)$ is an orthonormal system. This statement is checked by the direct calculation.

2. The functions $\Psi_{n_i}(\mathbf{x}, t, \hbar)$ form a complete system. This analogy can be used to establish the completeness for the Ψ_{n_i}, which can now be proved by a word for word transcription of the corresponding standard proof for the completeness of the Hermitian functions $U_n(x)$,

$$U_n(x) = C_n \exp(-x^2/2) H_n(x), \quad H_n(x) = 2^n (b^+)^n \Phi_0,$$
$$2b = \frac{d}{dx} + 2x, \quad \Phi_0 = 1. \tag{36}$$

Using the relations (32) and (33) the functions $\Psi_{n_i}(\mathbf{x}, t, \hbar)$ can be represented in the form

$$\Psi_{n_i}(\mathbf{x}, t, \hbar) = C_n \Psi_0(\mathbf{x}, t, \hbar) \prod_i [\hat{\Lambda}_i^+(t)]^{n_i} \Phi_0, \quad \Phi_0 = 1. \tag{37}$$

Thus the $\Psi_{n_i}(\mathbf{x}, t, \hbar)$ constitute a three-dimensional generalization of the Hermitian functions U_n. The completeness relation may be established now by the method that is a word for word repetition of the corresponding calculations for oscillator functions.

3. The quantum-mechanical means $\bar{\mathbf{X}}$ and $\bar{\mathbf{P}}$ calculated by functions $\Psi_{n_i}(\mathbf{x}, t, \hbar)$ don't depend on n_i and coincide (exactly) with $\mathbf{x}(t)$ and $\mathbf{p}(t)$. One may easily check this property by direct calculation. Thus $\mathbf{x}(t)$ is the center of the wave packet $\Psi_{n_i}(\mathbf{x}, t, \hbar)$. By analogy, $\mathbf{p}(t)$ is a center of the wave packet $\Psi_{n_i}(\mathbf{p}, t, \hbar)$ of the TCS in p-representation. The form and properties of the functions $\Psi_{n_i}(\mathbf{p}, t, \hbar)$ are determined by the form and properties of the functions $\Psi_{n_i}(\mathbf{x}, t, \hbar)$ with the substitutions

$$\mathbf{x} \leftrightarrow \mathbf{p}, \quad \mathbf{x}(t) \leftrightarrow \mathbf{p}(t), \quad C(t) \to -B(t), \quad B(t) \to C(t). \tag{38}$$

4. In case the operator $H(x, p)$ is a quadratic function in the coordinates and momenta, with coefficients that are arbitrary functions of time, the TCS are exact solutions of the Schrödinger equations, and $\hat{a}_i(t)$ and $\hat{a}_i^+(t)$ are exact symmetries of this equation.

5. Quasiclassical trajectory-coherent states of a spinless relativistic particle in an arbitrary electromagnetic field.

Now we construct trajectory-coherent states for the Klein–Gordon equation describing the motion of a spinless relativistic charged particle in an arbitrary electromagnetic field. This equation, in view of the above notation, has the following form:

$$[(i\hbar\partial_t - e\phi)^2 - (ic\hbar\nabla - e\mathbf{A})^2 - m^2c^4]\Psi = 0. \tag{39}$$

Here, the electromagnetic field potentials $\mathbf{A}$ and ϕ are taken in the Lorentz gauge $\partial_t \phi + C \operatorname{div} \mathbf{A} = 0$.

Let us denote by $\Psi_+(\mathbf{x}, t)$ a positive-frequency normalized solution of the Schrödinger-type equation

$$\hat{L}_+ \Psi_+ = 0, \ \hat{L}_+ = i\hbar\partial_t - \hat{H}(\mathbf{x}, \mathbf{p}, t), \ \int \Psi_+^* \Psi_+ d^3x = 1, \tag{40}$$

where, in difference with Sec. 2, the pseudodifferential operator $\hat{H}(\mathbf{x}, \mathbf{p}, t)$, that is a nonlocal "$\hbar^{-1}$," is defined by the formula

$$\hat{H}(\mathbf{x}, \mathbf{p}, t)\Psi_+(\mathbf{x}, t) = (2\pi\hbar)^{-3/2} \int \exp[i\hbar^{-1}\langle \mathbf{x}, \mathbf{p}\rangle] H(\mathbf{x}, \mathbf{p}, t) \tilde{\Psi}_+(\mathbf{p}, \hbar, t) d^3p. \tag{41}$$

Here $\tilde{\Psi}_+(\mathbf{p}, \hbar, t)$ is a "$\hbar^{-1}$" Fourier transform for the function $\tilde{\Psi}_+(\mathbf{x}, t)$, that is

$$\tilde{\Psi}_+(\mathbf{p}, \hbar, t) = (2\pi\hbar)^{-3/2} \int \exp[-i\hbar^{-1}\langle \mathbf{x}, \mathbf{p}\rangle] \Psi_+(\mathbf{x}, t) d^3x. \tag{42}$$

The function $H(\mathbf{x}, \mathbf{p}, t)$ is a classical relativistic Hamiltonian function

$$H(\mathbf{x}, \mathbf{p}, t) = e\phi(\mathbf{x}, t) + \epsilon_+(\mathbf{x}, \mathbf{p}, t),$$
$$\epsilon_+(\mathbf{x}, \mathbf{p}, t) = ([c\mathbf{p} - e\mathbf{A}(\mathbf{x}, t)]^2 + m^2c^4)^{1/2}. \tag{43}$$

We suppose the general solution $\mathbf{x}(t)$ and $\mathbf{p}(t)$ of the classical Lorentz equations to be known. In the Hamiltonian form, these equations coincide with Eqs. (2) with Hamiltonian (43). We furthermore introduce the matrix $\mathscr{H}(\mathbf{x}, \mathbf{p})$ analogous to the matrix (4). The block elements of this matrix calculated in the points of the classical relativistic trajectory defined by $\mathbf{x}(t)$ and $\mathbf{p}(t)$ have now the form

$$(H_{px}(t))_{ij} = (H_{xp}(t))_{ji}^* = [e\dot{x}_i \langle \mathbf{x}(t), \nabla_x A_j(t)\rangle - ec^2 \partial_{x_j} A_i(t)] \times [c\epsilon_+(t)]^{-1},$$
$$(H_{xx}(t))_{ij} = e[c^2 \epsilon_+(t) \partial_{x_i}\partial_{x_j}\phi(t) + ec^2 \langle \partial_{x_i}\mathbf{A}(t), \partial_{x_j}\mathbf{A}(t)\rangle - c\epsilon_+(t)\langle \mathbf{x}(t), \partial_{x_i}\partial_{x_j}\mathbf{A}(t)\rangle - e\langle \dot{\mathbf{x}}(t), \partial_{x_i}\mathbf{A}(t)\rangle \times \langle \dot{\mathbf{x}}(t), \partial_{x_j}\mathbf{A}(t)\rangle] [c^2\epsilon_+(t)]^{-1},$$
$$(H_{pp}(t))_{ij} = [c^2\delta_{ij} - \dot{x}_i(t)\dot{x}_j(t)]\epsilon_+^{-1}(t). \tag{44}$$

Here $\epsilon_+(t) = \epsilon_+(\mathbf{x}(t), \mathbf{p}(t), t)$. Solving the set of equations (8) with the matrix (44) we construct the normalized vacuum relativistic TCS solution of WKB type $\Psi_{+0}(\mathbf{x}, t, \hbar)$ as follows:

$$\Psi_{+0}(\mathbf{x}, t, \hbar) = N(\hbar)\Phi(t) \exp[i\hbar^{-1} S_+(\mathbf{x}, t)],$$
$$N(\hbar) = [(\pi\hbar)^{-3} \operatorname{Im} b_1 \operatorname{Im} b_2 \operatorname{Im} b_3]^{1/4}. \tag{45}$$

Here the relativistic complex phase $S_+(\mathbf{x}, t)$ differs from (14) and is defined by the formula

$$S_+(\mathbf{x}, t) = \int_0^t \mathscr{L}(\mathbf{x}(t), \dot{\mathbf{x}}(t), t) dt + \langle \mathbf{p}(t), \mathbf{x} - \mathbf{x}(t)\rangle + \tfrac{1}{2}\langle \mathbf{x} - \mathbf{x}(t), Q(t)(\mathbf{x} - \mathbf{x}(t))\rangle, \tag{46}$$

where $\mathscr{L}$ is the relativistic Lagrangian on the trajectory

$$\mathscr{L}(\mathbf{x}, \dot{\mathbf{x}}, t) = -mc^2(1 - \beta^2)^{1/2} - e[\phi(t) - c^{-1}\langle \dot{\mathbf{x}}, \mathbf{A}(t)\rangle], \quad c\beta = \dot{\mathbf{x}}. \tag{47}$$

The amplitude $\Phi(t)$ and the matrix $O(t)$ are defined in Sec. 3.

Relations (25) introduce the annihilation and creation operators that are constructed by the formulas (35). $\Psi_{+n_k}(\mathbf{x}, t, \hbar)$ are excited relativistic TCS's. The precision with which the relativistic TCS constructed satisfy equation (40) is defined by the theorem from Sec. 3. The operator $\hat{L}_+$

on the solutions (45) can be estimated in the norm $L_2(R^3)$ as follows

$$\|\hat{L}_+ \Psi_{+n_k}\| \leqslant R\hbar^{3/2}, \tag{48}$$

where the constant R, analogous to C_1 from (17), depends linearly on the maximum, with respect to $t \in [0, T]$, of the second-order derivatives of electromagnetic field potentials derived for the points of the classical relativistic trajectory.

The relativistic TCS's constructed here satisfy the original Klein–Gordon equation (39) with the accuracy $h^{3/2}$ at $\hbar \to 0$, the probability density for Eq. (39) on the approximate solutions being, accurate up to the members $O(\sqrt{\hbar})$,

$$\rho = \gamma \epsilon_+(t) |\Psi_{+n_k}|^2, \quad \gamma = \text{const} > 0 \tag{49}$$

and from this it follows that $\rho > 0$.

It is interesting to note that, in difference with the nonrelativistic case, the TCS's constructed here are not exact solutions even in the case of stationary uniform fields or for a free particle. The well-known exact coherent states of a relativistic electron[7] are partially coherent and can be obtained by the complex germ method on the basis of the complex Hamiltonian formalism for narrow beams; however, this deviates from the framework of this paper.

[1]E. Schrödinger, Naturwissenschaften **14**, 664 (1926).
[2]R. J. Glauber, Phys. Rev. **130**, 2529 (1963); **131**, 2766 (1964).
[3]J. Schwinger, Phys. Rev. **91**, 728 (1953).
[4]P. K. Rashevskij, Uspekhi Mat. Nauk **13**, 3 (1958).
[5]I. A. Malkin and V. I. Man'ko, *Dinamicheskie simmetrii i kogerentnye sostoyaniya kvantovykh sistem* (Nauka, Moscow, 1979).
[6]*Kvantovaya electrodinamika s vneshnim polem. Sb. statey. Izd.* (Tomskogo Universiteta, Tomsk, 1977).
[7]V. G. Bagrov, D. M. Gitman, I. M. Ternov, V. R. Chalilov, and V. N. Schapovalov, *Exact solutions of the relativistic wave equations* (Nauka, Novosibirsk, 1982).
[8]V. P. Maslov and M. V. Fedorjuk, *Kvaziklassicheskoe priblizhenie dlya uravnenij kvantovoj mekhaniki* (Nauka, Moscow, 1976).
[9]V. P. Maslov, *Kompleksnyj metod WKB v nelinejnykh uravneniyakh* (Nauka, Moscow, 1977).
[10]V. P. Maslov, *Operational methods* (1976).
[11]V. M. Babich and V. S. Buldyrev, *Asimptoticheskie metody v zadachakh difraktsii korotkikh voln* (Nauka, Moscow, 1972).
[12]L. D. Landau and E. M. Lifshits, *Kvantovaya mekhanika* (Nauka, Moscow, 1974).
[13]V. A. Yakubovich and V. M. Starzhinskij, *Lineinye differentsial'nye uravneniya s pereodicheskimi koeffitsientami i ikh prilozheniya* (Nauka, Moscow, 1972).
[14]G. A. Hagedorn, Commun. Math. Phys. **71**, 77 (1980).

On the Ohm–Navier–Stokes system in magnetohydrodynamics

Zensho Yoshida[a] and Yoshikazu Giga[b]
Courant Institute of Mathematical Sciences, New York University, New York, New York 10012

(Received 19 May 1983; accepted for publication 24 June 1983)

We study a mathematical structure of the Ohm–Navier–Stokes system that describes the incompressible dissipative evolution of a plasma. We apply the nonlinear semigroup theory and construct a unique regular solution which satisfies the system at least locally-in-time. We show that, for small initial data, this solution solves the system globally-in-time. We also introduce another scheme to construct solutions for less regular initial data.

PACS numbers: 52.30. + r, 52.55.Ke, 02.30. − f

I. INTRODUCTION

The dynamics of magnetically confined plasmas are attracting strong interest, especially in their relation to nuclear fusion technology.[1] Here we discuss an analytical structure of the following Ohm–Navier–Stokes system that describes incompressible dissipative dynamics of a magnetofluid.[2,3] We consider the initial-boundary-value problem of the system (see Sec. II)

$$\partial_t \mathbf{B} = (\eta/\mu_0)\triangle \mathbf{B} + \nabla\times(\mathbf{v}\times\mathbf{B}), \quad \nabla\cdot\mathbf{B} = 0,$$

$$\partial_t \mathbf{v} = \nu\triangle \mathbf{v} - (\mathbf{v}\cdot\nabla)\mathbf{v} + [(\nabla\times\mathbf{B})\times\mathbf{B}/\mu_0 - \nabla p]/\rho, \quad \nabla\cdot\mathbf{v} = 0,$$

where $\mathbf{B}$ is the magnetic flux density, $\mathbf{v}$ is the macroscopic fluid velocity, p is the thermal pressure; η is the resistivity, ν is the kinematic viscosity, ρ is the mass density, and μ_0 is the vacuum permeability. We assume η, ν, and ρ are constants. This system is closed because it is decoupled with the mass and energy (entropy) equations, since we assume the incompressibility of the fluid. Since we consider temporally homogeneous boundary conditions for $\mathbf{B}$ and $\mathbf{v}$ [see (2.5) and (2.6)], this system is autonomous.

In this paper we apply the semigroup theory in functional analysis and construct regular solutions of the Ohm–Navier–Stokes system. For an autonomous system, the solution with initial value u_0, if it exists uniquely, can be written as $T(t)u_0$ and the solution operator $T(t)$ $(t\geqslant 0)$ has the semigroup property $T(t)\cdot T(s) = T(t+s)$ $(t,s\geqslant 0)$, $T(0) = I$ (identity operator). The semigroup theory studies the existence and analytical property of semigroups. The idea for this theory comes from the concept of Laplace transformation that is a tool used to solve autonomous ordinary differential equations. Therefore, to study autonomous systems it seems natural to apply the semigroup theory. A nonlinear theory for semigroups in Hilbert space has been developed by Kōmura[4,5]; see the Appendix. We apply this theory to construct solutions of the Ohm–Navier–Stokes system. This nonlinear semigroup theory has been applied to the usual Navier–Stokes system in Ref. 6. We will see, in Sec. IV, that the Ohm–Navier–Stokes system has a mathematical analogy with the usual Navier–Stokes system. The analogy is useful to analyze the Ohm–Navier–Stokes system. We introduce, in Sec. VI, another method which is analogous to that in Refs. 7–9, and construct solutions for less regular initial data.

We refer the reader to Ref. 10 for the pioneering mathematical contributions to the Ohm–Navier–Stokes system. The present theory improves on the following points. The nonlinear semigroup approach in Sec. V gives a simple proof of the existence. The second method (Sec. VI) improves on the regularity assumption for the initial data. We allow the existence of an externally applied stationary magnetic field. For other contributions see Refs. 11 and 12 and Remark 7.2.

Our results concerned with the Ohm–Navier–Stokes system read: There exists a unique regular solution at least locally-in-time (Theorems 5.1 and 6.1). If the initial value is sufficiently small, the solution solves the system globally-in-time (Theorems 5.2 and 6.1).

II. PHYSICAL BACKGROUND

Let us consider an incompressible plasma with finite resistivity and viscosity, which is contained in a bounded domain Ω $(\subset \mathbf{R}^3)$ surrounded by a perfectly conducting smooth wall $\partial\Omega$. We describe its magnetohydrodynamic evolution by the self-consistent macroscopic model

$$\mathbf{E} = \eta\mathbf{j} - \mathbf{v}\times\mathbf{B}, \tag{2.1}$$

$$\partial_t \mathbf{v} = \nu\triangle \mathbf{v} - (\mathbf{v}\cdot\nabla)\mathbf{v} + (\mathbf{j}\times\mathbf{B} - \nabla p)/\rho, \quad \nabla\cdot\mathbf{v} = 0. \tag{2.2}$$

Here, $\mathbf{E}$ is the electric field, $\mathbf{B}$ is the magnetic flux density with $\nabla\cdot\mathbf{B} = 0$, $\mathbf{j}$ is the current density. We assume η, ν, and ρ are constants. Equations (2.1) and (2.2) are, respectively, Ohm's law and the Navier–Stokes system with the $\mathbf{j}\times\mathbf{B}$ force.

When we combine Ohm's law (2.1) with Faraday's law,

$$\partial_t \mathbf{B} = -\nabla\times\mathbf{E}, \tag{2.3}$$

we get

$$\partial_t \mathbf{B} = (\eta/\mu_0)\triangle \mathbf{B} + \nabla\times(\mathbf{v}\times\mathbf{B}). \tag{2.4}$$

Here we use $\mathbf{j} = \mu_0^{-1}\nabla\times\mathbf{B}$, Ampere's law with the displacement current neglected.

We will now discuss boundary conditions. The perfect conductivity of the wall implies

$$\mathbf{E}\times\mathbf{n} = \mathbf{0} \quad \text{on } \partial\Omega, \tag{2.5}$$

where $\mathbf{n}$ is the normal vector to the wall $\partial\Omega$. For $\mathbf{v}$, we have the adherence boundary condition

$$\mathbf{v} = \mathbf{0} \quad \text{on } \partial\Omega. \tag{2.6}$$

Let us derive boundary conditions for $\mathbf{B}$. Equation (2.1) with (2.5) and (2.6) implies

$$(\nabla\times\mathbf{B})\times\mathbf{n} = \mathbf{0} \quad \text{on } \partial\Omega. \tag{2.7}$$

[a] On leave from Department of Nuclear Engineering, University of Tokyo, 7-3-1 Hongo, Tokyo 113, Japan.
[b] On leave from Department of Mathematics, Nagoya University, Chikusa-ku, Nagoya 464, Japan.

 0022-2488/83/122860-05$02.50

We get from (2.3) and (2.5)

$$\mathbf{B}\cdot\mathbf{n} = g(\mathbf{x}) \quad \text{on } \partial\Omega, \tag{2.8}$$

where $g(\mathbf{x})$ is a certain function independent of time.

To homogenize the boundary condition (2.8) we write

$$\mathbf{B} = \mathbf{B}_p + \mathbf{B}_{ex},$$

where $\mathbf{B}_p$ and $\mathbf{B}_{ex}$ are uniquely defined by

$$\mathbf{B}_p = \nabla\times\mathbf{A} \quad \text{in } \Omega, \quad \mathbf{A}\times\mathbf{n} = \mathbf{0} \quad \text{and } \mathbf{B}_p\cdot\mathbf{n} = 0 \quad \text{on } \partial\Omega,$$

$$\nabla\times\mathbf{B}_{ex} = \mathbf{0} \quad \text{in } \Omega, \quad \mathbf{B}_{ex}\cdot\mathbf{n} = g(\mathbf{x}) \quad \text{on } \partial\Omega.$$

This expression is called the Hodge–Kodaira decomposition for divergence-free fields; see Theorem 7.8.1 (b) in Ref. 13. Physically, $\mathbf{B}_{ex}$ is the externally applied stationary flux density which is interpreted as a given function, and $\mathbf{B}_p$ is the remaining part, the flux density related to the plasma current density ($\mathbf{j} = \mu_0^{-1}\nabla\times\mathbf{B}_p$).

Remark 2.1: We postulate the boundary value $g(\mathbf{x})$ is smooth (C^∞ -class). It then follows that $\mathbf{B}_{ex}$ is smooth (cf. Sec. 6.3 in Ref. 13).

In summary, we consider the initial-boundary-value problem for the system

$$\partial_t\mathbf{B}_p = (\eta/\mu_0)\triangle\mathbf{B}_p + \nabla\times[\mathbf{v}\times(\mathbf{B}_p + \mathbf{B}_{ex})], \quad \nabla\cdot\mathbf{B}_p = 0, \tag{2.9}$$

$$\partial_t\mathbf{v} = \nu\triangle\mathbf{v} - (\mathbf{v}\cdot\nabla)\mathbf{v} + [(\nabla\times\mathbf{B}_p)\times(\mathbf{B}_p + \mathbf{B}_{ex})/\mu_0 - \nabla p]/\rho, \tag{2.10}$$

$$\nabla\cdot\mathbf{v} = 0,$$

with boundary conditions

$$\mathbf{B}_p\cdot\mathbf{n} = 0, \quad (\nabla\times\mathbf{B}_p)\times\mathbf{n} = \mathbf{0} \quad \text{on } \partial\Omega, \tag{2.11}$$

$$\mathbf{v} = \mathbf{0} \quad \text{on } \partial\Omega.$$

Let us address the system (2.9), (2.10) as the Ohm–Navier–Stokes system (ONS for short).

III. MATHEMATICAL FORMULATION

We interpret ONS as an evolution equation that is an ordinary differential equation in a function space. We begin with recalling some fundamental function spaces and operators.

We first give an operator-theoretic interpretation of the dissipative terms in ONS. Let $L^2(\Omega)$ denote the Hilbert space of square-integrable vector functions on Ω. We denote by $H^m(\Omega)$ the Sobolev space of vector functions which are in $L^2(\Omega)$ together with all their derivatives of order $\leqslant m$. We define the Laplace operator $\mathscr{L}_1$ in $L^2(\Omega)$ with boundary condition (2.11) by

$$\mathscr{L}_1 = -(\eta/\mu_0)\triangle,$$

with domain

$$D(\mathscr{L}_1) = \{\mathbf{B}_p\in H^2(\Omega);\ (\nabla\times\mathbf{B}_p)\times\mathbf{n} = \mathbf{0},\ \mathbf{B}_p\cdot\mathbf{n} = 0 \text{ on } \partial\Omega\}.$$

The operator $-\mathscr{L}_1$ describes the dissipation of the magnetic field energy. We next define the Stokes operator associated with the dissipation of the kinetic energy of fluid motion. Let H_σ be the closure in $L^2(\Omega)$ of smooth divergence-free vector fields with compact support in Ω. Obviously, H_σ is a closed subspace of $L^2(\Omega)$. Let $\mathscr{P}$ denote the orthogonal projector in $L^2(\Omega)$ onto H_σ. The Stokes operator $\mathscr{L}_2$ is defined by

$$\mathscr{L}_2 = -\nu\mathscr{P}\triangle,$$

with domain

$$D(\mathscr{L}_2) = \{\mathbf{v}\in H^2(\Omega);\ \mathbf{v} = \mathbf{0} \text{ on } \partial\Omega\}\cap H_\sigma.$$

To eliminate ∇p in the Navier–Stokes system[14] we apply $\mathscr{P}$ to (2.10), and we have

$$\partial_t\mathbf{v} = -\mathscr{L}_2\mathbf{v} - \mathscr{P}(\mathbf{v}\cdot\nabla)\mathbf{v} + \mathscr{P}[(\nabla\times\mathbf{B}_p)\times(\mathbf{B}_p + \mathbf{B}_{ex})/\mu_0\rho]. \tag{3.1}$$

We consider ONS in a Hilbert space $H = L^2\times H_\sigma$ endowed with an inner product

$$\langle u,u'\rangle := \int_\Omega(\mu_0^{-1}\mathbf{B}_p\cdot\mathbf{B}_p' + \rho\mathbf{v}\cdot\mathbf{v}')d\mathbf{x},$$

$$u = \begin{pmatrix}\mathbf{B}_p\\ \mathbf{v}\end{pmatrix}, \quad u' = \begin{pmatrix}\mathbf{B}_p'\\ \mathbf{v}'\end{pmatrix}\in H.$$

We denote by $\|u\|$ the norm[15] of u in H. To simplify the notation we introduce a bilinear operator

$$b\left(\begin{pmatrix}\mathbf{X}\\ \mathbf{Y}\end{pmatrix},\begin{pmatrix}\mathbf{X}'\\ \mathbf{Y}'\end{pmatrix}\right) = \begin{pmatrix}\nabla\times(\mathbf{Y}'\times\mathbf{X})\\ -\mathscr{P}(\mathbf{Y}\cdot\nabla)\mathbf{Y}' + \mathscr{P}[(\nabla\times\mathbf{X}')\times\mathbf{X}]/\mu_0\rho\end{pmatrix}$$

for vector fields $\mathbf{X}, \mathbf{Y}, \mathbf{X}', \mathbf{Y}'$. Now ONS [(2.9) and (3.1)] reads an evolution equation[14] in H

$$\frac{du}{dt} = -\mathscr{L}u + \mathscr{N}u \quad (t\geqslant 0), \quad u(0) = u_0 = \begin{pmatrix}\mathbf{B}_{p0}\\ \mathbf{v}_0\end{pmatrix}, \tag{3.2}$$

$$\mathscr{L}u := \begin{pmatrix}\mathscr{L}_1 & 0\\ 0 & \mathscr{L}_2\end{pmatrix}\begin{pmatrix}\mathbf{B}_p\\ \mathbf{v}\end{pmatrix},$$

$$\mathscr{N}u := b(u,u) + b(u_{ex},u),$$

where

$$u = \begin{pmatrix}\mathbf{B}_p\\ \mathbf{v}\end{pmatrix}, \quad u_{ex} = \begin{pmatrix}\mathbf{B}_{ex}\\ \mathbf{0}\end{pmatrix},$$

and du/dt denotes the strong derivative of u. The condition $\nabla\cdot\mathbf{v} = 0$ in (2.10) is implicitly included in (3.2), while $\nabla\cdot\mathbf{B} = 0$ in (2.9) is dropped in (3.2). However, if initial value $\mathbf{B}_{p0}$ satisfies a compatibility condition $\nabla\cdot\mathbf{B}_{p0} = 0$, the first equation in (2.9) implies $\nabla\cdot\mathbf{B}_p = 0$.

IV. PRELIMINARY LEMMAS

We prepare some fundamental lemmas concerning operators $\mathscr{L}_1$, $\mathscr{L}_2$ and the nonlinear term in (3.2).

Lemma 4.1: The linear operators $\mathscr{L}_1$ and $\mathscr{L}_2$ are positive self-adjoint operators in $L^2(\Omega)$ and H_σ, respectively. Consequently, $\mathscr{L}$ is a positive self-adjoint operator in H, and we can define the power $\mathscr{L}^\alpha$ of $\mathscr{L}$ for every $\alpha\in\mathbf{R}$.

This can be easily proved by the standard argument (cf. Chap. 7 in Ref. 16 and Lemma 1 in Ref. 7).

Lemma 4.2: We have

$$\langle b(u,w_1), w_2\rangle = -\langle b(u,w_2), w_1\rangle$$

for $u, w_1, w_2\in H$ as far as the both sides are well defined. In particular we have $\langle\mathscr{N}u,u\rangle = 0$.

See, for example, Sec. 6.2 in Ref. 11 and Chap. II, Lemma 1.3 in Ref. 17.

We have the following estimates for the nonlinear term (cf. Ref. 7 and Ref. 9).

Lemma 4.3: For δ, $0 \leqslant \delta < \frac{5}{4}$ the estimate

$$\|\mathscr{L}^{-\delta} b(u,w)\| \leqslant C(\|\mathscr{L}^{\theta} u\| \, \|\mathscr{L}^{\rho} w\| + \|\mathscr{L}^{\rho'} u\| \, \|\mathscr{L}^{\theta'} w\|)$$

holds with a constant C independent of u and w, if

$$\delta + \theta + \rho \geqslant \tfrac{5}{4}, \quad \delta + \theta' + \rho' \geqslant \tfrac{5}{4}, \quad \theta, \theta' > 0, \quad \rho, \rho' > 0,$$
$$\rho + \delta > \tfrac{1}{2}, \quad \rho' + \delta > \tfrac{1}{2}.$$

This follows from Hölder's inequality, Sobolev's inequality, and the regularity property of $\mathscr{L}^{\alpha}$.

An energy estimate follows from Lemmas 4.1 and 4.2.

Lemma 4.4: For a regular solution of (3.2) we have the energy equality

$$\frac{d}{dt}\|u\|^2 + 2\|\mathscr{L}^{1/2} u\|^2 = 0.$$

The above-mentioned lemmas show mathematical analogy of ONS with the usual Navier–Stokes system.

V. NONLINEAR SEMIGROUP SOLUTIONS

We will construct a regular solution of ONS, using the nonlinear semigroup theory established by Kōmura.[4,5] Most of our argument is similar to that in Ref. 6, so we omit the detail. To apply the general theory of nonlinear semigroups (see the Appendix) we first truncate the nonlinear operator $\mathscr{N}$. The general theory gives a unique and global-in-time regular solution of the truncated system. Then we check whether this solution satisfies the original ONS.

We consider a truncated operator as follows. Let $\psi_M(s)$ $(s \geqslant 0)$ denote a cut function such that

$$\psi_M(s) = \begin{cases} 1, & 0 \leqslant s \leqslant M' = M/2, \\ 2 - s/M', & M' < s < M, \\ 0, & M \leqslant s. \end{cases}$$

Let $\Psi_M(u)$ denote $\psi_M(\|\mathscr{L}^{1/2} u\|)$. We define the truncated operator $\mathscr{B}_M$ by

$$\mathscr{B}_M u = -\mathscr{L} u + \Psi_M(u)\mathscr{N} u, \quad D(\mathscr{B}_M) = D(\mathscr{L}).$$

This $\mathscr{B}_M$ is well defined because Lemma 4.3 implies that $\mathscr{N} u$ belongs to H if u is in $D(\mathscr{L})$.

Because we have truncated the nonlinear term, we can prove, just like Lemma 2.1, in Ref. 6, that

$$\langle (\mathscr{B}_M - \omega I)u - (\mathscr{B}_M - \omega I)u', u - u' \rangle \leqslant 0$$

for all $u, u' \in D(\mathscr{B}_M)$ with $\omega = \omega(M) := c(M + \|\mathscr{L}^{1/2} u_{\mathrm{ex}}\|)^4$ for some constant $c = c(\Omega)$; here we use Lemmas 4.1–4.3. In other words, $\mathscr{B}_M - \omega I$ is *dissipative* in H.

We see the stationary problem

$$u - \lambda(\mathscr{B}_M - \omega I)u = f$$

is solvable for all $f \in H$, $\lambda > 0$; see Lemma 2.2 in Ref. 6. Such a dissipative operator $\mathscr{B}_M - \omega I$ is called *hyperdissipative*. Summarizing our results, we have

Lemma 5.1: The operator $\mathscr{B}_M - \omega(M)I$ is hyperdissipative in H.

Applying the nonlinear semigroup theory, we solve the evolution equation $du/dt = \mathscr{B}_M u$.

Proposition 5.1: The operator $\mathscr{B}_M$ generates the nonlinear semigroup $T_M(t)$ in H. The function $u(t) = T_M(t)u_0$ for $u_0 \in D(\mathscr{B}_M)$ is absolutely continuous in $t \geqslant 0$ with value in H. Moreover $u(t)$ uniquely and globally solves the truncated system

$$\frac{du}{dt} = \mathscr{B}_M u \quad (\text{a.e. } t \geqslant 0), \quad u(0) = u_0.$$

For $T_M(t)u_0$ we have an estimate

$$\|\mathscr{B}_M T_M(t)u_0\| \leqslant e^{\omega(M)(t-s)} \|\mathscr{B}_M T_M(s)u_0\|, \quad \text{a.e. } t \geqslant s \geqslant 0. \tag{5.1}$$

This $u(t)$ solves the Ohm–Navier–Stokes system (3.2) at least locally-in-time, provided that M is sufficiently large.

Theorem 5.1: We can choose M' $(= M/2)$ such that $M'^2 > \|\mathscr{B}_M u_0\| \, \|u_0\|$. There is a constant $T \geqslant T_0 = \omega(M)^{-1} \times \log(M'^2/\|\mathscr{B}_M u_0\| \, \|u_0\|)$ such that, for $t \in [0,T]$, $u(t) = T_M(t)u_0$ solves the Ohm–Navier–Stokes system

$$\frac{du}{dt} = -\mathscr{L} u + \mathscr{N} u \quad (\text{a.e. } t \geqslant 0), \quad u(0) = u_0 \in D(\mathscr{L}). \tag{5.2}$$

The proof is similar to the proof of Theorem 4.1 in Ref. 6. Since this proof is so easy, we give it here for the reader's convenience.

Proof: It suffices to prove $\|\mathscr{L}^{1/2} u(t)\| \leqslant M'$ (a.e. $t, 0 \leqslant t \leqslant T_0$). Since $\langle \mathscr{N} u, u \rangle = 0$ (Lemma 4.2), we see

$$\|\mathscr{L}^{1/2} u\|^2 = \langle \mathscr{L} u, u \rangle = |\langle \mathscr{B}_M u, u \rangle| \leqslant \|\mathscr{B}_M u\| \, \|u\|. \tag{5.3}$$

The estimate (5.1) gives

$$\|\mathscr{B}_M u(t)\| \leqslant e^{\omega(M)t} \|\mathscr{B}_M u_0\| \quad (\text{a.e. } t \geqslant 0).$$

On the other hand, the energy equality Lemma 4.4 yields

$$\|u(t)\| \leqslant \|u_0\| \quad (t \geqslant 0). \tag{5.4}$$

Combine the above three estimates to get

$$\|\mathscr{L}^{1/2} u(t)\|^2 \leqslant e^{\omega(M)t} \|\mathscr{B}_M u_0\| \, \|u_0\| \quad (\text{a.e. } t \geqslant 0).$$

Therefore we have obtained the desired result.

We now discuss the global existence for small initial data.

Theorem 5.2: Choose M' $(= M/2)$ as in Theorem 5.1. Then there is a positive constant $\epsilon = \epsilon(M,\Omega)$ such that, if $\|\mathscr{L}^{1/2} u_0\| < \epsilon$, $u(t) = T_M(t)u_0$ uniquely and globally solves the Ohm–Navier–Stokes system (5.2).

Proof: To show this theorem we use an *a priori* estimate

$$\int_0^t \|\mathscr{B}_M u(s)\|^2 ds \leqslant \|\mathscr{L}^{1/2} u_0\|^2 + K\|u_0\|^2, \quad K = K(M,\Omega),$$

which can be proved similarly to Lemma 4.1 in Ref. 6. To derive an upper bound for $\|\mathscr{B}_M u(s)\|$ this estimate is not enough because there may appear narrow spikes in the graph of $\|\mathscr{B}_M u(t)\|$ vs t. The growth rate estimate (5.1), however, prohibits such spikes, so we get an upper bound for $\|\mathscr{B}_M u\|$. This together with (5.3) and (5.4) eventually gives an upper bound for $\|\mathscr{L}^{1/2} u\|$. If the upper bound is smaller than M', $u(t)$ solves ONS globally-in-time. This situation is actually realized if the initial data is sufficiently small. For the detailed proof we refer the reader to Theorem 4.2 in Ref. 6.

VI. ALTERNATIVE METHOD

The system (3.2) is parabolic, so it is natural to ask whether a unique solution exists for less regular initial data. The method in Sec. V requires the assumption that the initial value u_0 is in $D(\mathscr{L})$. We give here another method essentially due to Kato and Fujita[7,8]; see also Giga and Miyakawa.[9]

Let us sketch their idea. We rewrite (3.2) in an integral form

$$u(t) = S(t)u_0 + \int_0^t S(t-s)\mathscr{N}u(s)ds, \tag{6.1}$$

where $S(t)$ is the linear semigroup generated by $-\mathscr{L}$ and $U(t) = S(t)a$ solves the linear equation $dU/dt = -\mathscr{L}U$, $U(0) = a$. We want to construct the solution of (6.1) by the successive approximation

$$u_0(t) = S(t)u_0,$$

$$u_{m+1}(t) = u_0(t) + \int_0^t S(t-s)\mathscr{N}u_m(s)ds, \quad m \geqslant 0. \tag{6.2}$$

If $\mathbf{B}_{ex}$ is equal to zero, we can directly apply the method of Kato and Fujita since we have Lemmas 4.1 and 4.3. Ladyzhenskaya and Solonnikov[10] consider (6.2) and get weaker results for $\mathbf{B}_{ex} = \mathbf{0}$. We do not assume $\mathbf{B}_{ex} = \mathbf{0}$ to get results.

Theorem 6.1: Suppose $\|\mathscr{L}^{1/4}u_0\|$ is finite. Then there is a unique regular local solution of (3.2). If $\|\mathscr{L}^{1/4}u_0\|$ is sufficiently small, then solution can be extended globally-in-time.[19]

Proof: We prove here that the estimate $\|\mathscr{L}^{\alpha}u_m\| \leqslant Kt^{\sigma-\alpha}$, $\sigma \leqslant \alpha < 1-\delta$ $(\sigma = \frac{1}{4}, \delta = \frac{1}{2})$ holds with constant K for all time if $\|\mathscr{L}^{\sigma}u_0\|$ is sufficiently small. This is a crucial step of our argument. The rest of the proof is similar to those in Refs. 7–9, and therefore we omit it.

Suppose $\|\mathscr{L}^{\alpha}u_m\| \leqslant K_{\alpha,m}t^{\sigma-\alpha}$, $\sigma \leqslant \alpha < 1-\delta$ for all time. We will estimate $K_{\alpha,m+1}$. Apply Lemma 4.3 to get

$$\begin{aligned}\|\mathscr{L}^{-\delta}b(u_m,u_m)\| &\leqslant C\|\mathscr{L}^{\sigma}u_m\|\,\|\mathscr{L}^{\rho}u_m\| \quad (\rho = \tfrac{1}{2}) \\ &\leqslant CK_{\sigma,m}K_{\rho,m}t^{\sigma+\delta-1}.\end{aligned}$$

For technical reasons we need a trick for u_{ex}. Since u_{ex} is in $C^{\infty}(\bar{\Omega})$ from Remark 2.1, $\|u_{ex}\|_{2(\sigma+\epsilon)} \leqslant C'$ for $\epsilon, 0 \leqslant \epsilon < 1$. We have $\|u_{ex}\|_{2(\sigma+\epsilon)} \leqslant C'/2t^{\epsilon}$ for t, $0 \leqslant t \leqslant T(\epsilon) = 2^{1/\epsilon}$, since $\|u_{ex}\|_{2(\sigma+\epsilon)}$ is independent of time. Using Lemma 4.3, we get

$$\begin{aligned}\|\mathscr{L}^{-\delta}b(u_{ex},u_m)\| &\leqslant C\|u_{ex}\|_{2(\sigma+\epsilon)}\,\|\mathscr{L}^{\rho-\epsilon}u_m\| \\ &\leqslant C_1C'K_{\rho-\epsilon,m}t^{\sigma+\delta-1}, \quad 0 \leqslant t \leqslant T(\epsilon)\end{aligned}$$

for all $0 < \epsilon < \frac{1}{4}$. Apply $\mathscr{L}^{\alpha}$ to (6.2) to get

$$\mathscr{L}^{\alpha}u_{m+1}(t) = \mathscr{L}^{\alpha}u_0(t) + \int_0^t \mathscr{L}^{\alpha+\delta}S(t-s)\mathscr{L}^{-\delta}\mathscr{N}u_m\,ds. \tag{6.3}$$

On the other hand, Lemma 4.1 yields $\|\mathscr{L}^{\alpha+\delta}S(t-s)f\| \leqslant C_{\alpha}(t-s)^{-\alpha-\delta}\|f\|$. Estimating (6.3) by this and the above two estimates eventually imply

$$K_{\sigma,m+1} \leqslant K_{\alpha,0} + C_2(K_{\sigma,m}K_{\rho,m} + C'K_{\rho-\epsilon,m}),$$
$$\sigma \leqslant \alpha < 1-\delta,$$

where C_2 is independent of m and ϵ. This successive inequality yields

$$K_{\alpha,m} \leqslant K, \quad \sigma \leqslant \alpha < 1-\delta \quad \text{for all } m,$$

provided that $K_{\alpha,0}$ and C' are sufficiently small. In other words, if $\|\mathscr{L}^{\sigma}u_0\|$ is sufficiently small, we get

$$\|\mathscr{L}^{\alpha}u_m\| \leqslant Kt^{\sigma-\epsilon}, \quad 0 \leqslant t \leqslant T(\epsilon).$$

Since K is independent of ϵ, this implies

$$\|\mathscr{L}^{\alpha}u_m\| \leqslant Kt^{\sigma-\epsilon}, \quad t \geqslant 0, \ \sigma \leqslant \alpha < 1-\delta,$$

which is the desired result.

Remark 6.1: If we use L^p-theory in Ref. 9, $\|\mathscr{L}^{1/4}u_0\|$ in Theorem 6.1 is replaced by L^3-norm of u_0.

VII. FINAL REMARKS

Below we list some additional remarks.

Remark 7.1 (Regularity): Since the solution $u(t) = T_M(t)u_0$ constructed in Sec. V is in $L^{\infty}(0,T;D(\mathscr{L}))$, $u(t)$ is the strong solution of the Ohm–Navier–Stokes system. Eventually $u(t)$ is in $C^{\infty}((0,T)\times\bar{\Omega})$; in other words $u(t)$ is a classical solution,[9] and the statement (a.e., $t \geqslant 0$) in (5.2) can be replaced by $(t \geqslant 0)$.

Remark 7.2 (Weak solutions): Duvaut and Lions[11] and Sanchez-Palencia[12] construct a global but weak solution of ONS. It is not known whether their weak solutions are regular.

Remark 7.3 (ONS with higher order dissipation): We consider a dissipative operator $-\tilde{\mathscr{L}}_{\epsilon}$ with higher-order perturbation

$$-\tilde{\mathscr{L}}_{\epsilon} = -\mathscr{L} - \epsilon\mathscr{L}^{\alpha}, \quad 0 < \epsilon \ll 1, \quad \alpha \geqslant \tfrac{3}{2}.$$

For a perturbed ONS

$$\frac{du}{dt} = -\tilde{\mathscr{L}}_{\epsilon}u + \cdot\,\mathscr{N}u \quad (t \geqslant 0), \quad u(0) = u_0 \in D(\tilde{\mathscr{L}}_{\epsilon}),$$

we can prove the existence of a unique and global-in-time solution without assuming that the initial data is small. This is because the higher-order dissipation gives an *a priori* bound for $\|\mathscr{L}^{1/2}u\|$.

Remark 7.4 (ONS with an ignorable coordinate): Here we suppose that there is an ignorable spatial coordinate, viz., we consider a two-dimensional system. Then, without assuming that the initial data is small, we can prove the existence of a unique and global-in-time regular solution of the two-dimensional ONS. We have an *a priori* bound for $\|\mathscr{L}^{1/4}u\|$, where u is a solution of ONS. This give the above-mentioned global existence; see Ref. 6.

APPENDIX: NONLINEAR SEMIGROUP THEORY

We briefly review the theory of nonlinear semigroups in Hilbert space, which is due to Kōmura[4,5]; see also Ref. 18.

Let $\mathscr{B}$ be a (nonlinear) operator from a subset $D(\mathscr{B})$ of a Hilbert space H into H. We say $\mathscr{B}$ is *dissipative* if $\langle \xi - \eta, u - v\rangle \leqslant 0$ for all $u, v \in D(\mathscr{B})$, $\xi \in \mathscr{B}u$, $\eta \in \mathscr{B}v$. A dissipative operator $\mathscr{B}$ is called *hyperdissipative* if the range of $I - \lambda\mathscr{B}$ is equal to H for a certain $\lambda > 0$, where I is the identity operator. For simplicity let us assume here that $\mathscr{B}$ is single valued. We have the fundamental theorem.

Theorem: Suppose $\mathscr{B} - \omega I$ is hyperdissipative for some $\omega > 0$. Then $\mathscr{B}$ generates a nonlinear semigroup $T(t)$. Moreover $u(t) = T(t)u_0$ for $u_0 \in D(\mathscr{B})$ is absolutely continuous in $t \geqslant 0$ with value in H. And $u(t)$ uniquely and globally solves

$$\frac{du}{dt} = \mathscr{B}u \quad \text{(a.e. } t \geqslant 0\text{)}, \quad u(0) = u_0,$$

where du/dt denotes the strong derivative of u. The function $u(t)$ also solves

$$\frac{d^+u}{dt} = \mathscr{B}u \quad (t \geqslant 0), \quad u(0) = u_0,$$

where d^+u/dt denotes the right derivative of u. The semigroup $T(t)$ satisfied estimates

$$\|T(t)u_0 - T(t)w_0\| \leqslant e^{\omega t}\|u_0 - w_0\|, \quad t \geqslant 0, \quad u_0, w_0 \in D(\mathscr{B}),$$

$$\|\mathscr{B}T(t)u_0\| \leqslant e^{\omega(t-s)}\|\mathscr{B}T(s)u_0\|, \quad \text{a.e. } t \geqslant s \geqslant 0.$$

ACKNOWLEDGMENTS

The authors are grateful to Professor P. Laurence who kindly pointed out some useful references. The authors are also grateful to Professor H. Grad, Professor W. Grossmann, Professor N. Inoue, Professor Y. Konishi, Professor T. Uchida, and Professor H. Weitzner for their valuable comments and encouragement. Research of the second author was partially supported by The Sakkokai Foundation.

[1]K. Miyamoto, *Plasma Physics for Nuclear Fusion* (MIT, Cambridge, MA, 1981).
[2]N. G. Van Kampen and B. U. Felderhof, *Theoretical Methods in Plasma Physics* (North-Holland, Amsterdam, 1967).
[3]G. Bateman, *MHD Instabilities* (MIT, Cambridge, MA, 1978).
[4]Y. Kōmura, J. Math. Soc. Japan **19**, 493 (1967).
[5]Y. Kōmura, J. Math. Soc. Japan **21**, 375 (1969).
[6]Z. Yoshida and Y. Giga, "A nonlinear semigroup approach to the Navier–Stokes system" (to appear in Comm. Partial Differential Equations).
[7]H. Fujita and T. Kato, Arch. Rational Mech. Anal. **16**, 269 (1964).
[8]T. Kato and H. Fujita, Rend. Sem. Mat. Univ. Padova **32**, 243 (1962).
[9]Y. Giga and T. Miyakawa, "Solutions in L_r to the Navier–Stokes initial value problem" (to appear in Arch. Rational Mech. Anal.).
[10]O. A. Ladyzhenskaya and V. A. Solonnikov, J. Soviet Math. **8**, 384 (1977).
[11]G. Duvaut and J. L. Lions, Arch. Rational Mech. Anal. **46**, 241 (1972).
[12]E. Sanchez-Palencia, Journal de Mécanique **8**, 509 (1969).
[13]C. B. Morrey, Jr., *Multiple Integrals in the Calculus of Variations* (Springer-Verlag, New York, 1966).
[14]When we find the solution u of (3.2), ∇p is implicitly determined (cf. Ref. 17).
[15]Physically $\|u\|^2/2$ is the total energy of the system, viz., the sum of the magnetic field energy and the kinetic energy.
[16]G. Duvaut and J. L. Lions, *Les inéquations en mécanique et en physique* (Dunod, Paris, 1972).
[17]R. Teman, *Navier–Stokes Equations* (North-Holland, Amsterdam, 1977).
[18]H. Brezis, *Opérateurs Maximaux Monotones et Semi-groups de Contractions dan les Espace de Hilbert* (North-Holland, Amsterdam, 1973).
[19]The proof given below requires the smallness of $\|\mathbf{B}_{ex}\|_{1/2+2\epsilon}$, where $\|f\|_s$ deontes the norm of f in $H_s(\Omega)$.

Structural phase transitions in crystals: Broken-symmetry (isotropy) groups[a)]

Marko V. Jarić
Department of Physics, Montana State University, Bozeman, Montana 59717
Institut des Hautes Etudes Scientifiques, 91440 Bures-sur-Yvette, France
Freie Universität Berlin, Institute for Theoretical Physics, Arnimallee 3, D-1000 Berlin 33, F. R. Germany

(Received 21 January 1983; accepted for publication 20 May 1983)

A systematic procedure ensuring the determination of all isotropy groups of a given representation of a space group is presented for the first time. Isotropy groups play an important role in various areas of theoretical solid state physics. For example, only isotropy groups may occur in a structural phase transition driven by an order parameter belonging to a given representation. The method uses the chain criterion directly on the image of the representation, employing a labeling of the matrices by space group elements. A notion of a substar of a wave vector associated with the representation is central to the method. Finally, as a detailed illustration the method is applied to a structural phase transition in A15 systems driven by an X-point order parameter. The result agrees with previously reported ones.

PACS numbers: 64.60. – i, 61.50.Em, 02.20. + b

I. INTRODUCTION

In calculating various physical quantities we often deal not with the full symmetry of the physical system, but with a specific representation of this symmetry. For example, if we are calculating a perturbation of a certain energy level then the wave functions, perhaps degenerate at that particular level, span a representation of the full symmetry group of the physical system under consideration. In such a case it is useful to know some details about the action of matrices of the representation on vectors in the space spanned by the wave functions. This is even more apparent in another example coming from solid state physics, namely, in the Landau[1] and renormalization-group[2] theories of structural phase transitions in crystals. Since the present work was mainly motivated by various applications of these theories, we will outline their content here.

In the Landau theory symmetry of a crystal at a given temperature (pressure, concentration, etc.) is determined as the symmetry of an order parameter which minimizes the Landau free energy. The Landau theory may be considered as an approximate solution of a more general theory. In such a theory the order parameter is given as an expectation value over a certain distribution. The Landau theory replaces this expectation value by the most probable one. This more general approach is incorporated in the renormalization-group method. However, symmetry results of the Landau theory are also relevant to the renormalization-group method. Thus, we will concentrate on the Landau theory.[3]

We will begin with a general outline of the Landau, phenomenological approach.

The order parameter is a vector from a space C over which a representation $\mathbf{R}$ of the high symmetry group G is spanned. On the other hand, the free energy is a family (as temperature is changed) of G-invariant polynomials on C. At each given temperature, an absolute minimum of the free energy is at a vector $\mathbf{c}$ in C. This vector is identified as the order parameter at the given temperature. The associated (low-) symmetry group at that temperature is then the largest subgroup L of G which leaves $\mathbf{c}$ (the order parameter) invariant. That is, L is the isotropy group (little group, stabilizer) of $\mathbf{c}$. Consequently, it would be beneficial to know the isotropy groups L of G (for given $\mathbf{R}$).

A representation $\mathbf{R}$ must first be chosen. $\mathbf{R}$ may be inferred from some theoretical arguments, or partially determined from experiments or determined *a posteriori*, from the requirement that theoretical predictions agree with experiments. For example, an action of the group G on the atomic displacements of a crystal, with some experimentally imposed periodicity, may be analyzed. The resulting representation can then be reduced and some of its irreducible components $\mathbf{R}$ may be used as an order parameter. Conversely, if $\mathbf{R}$ is given, associated displacement patterns may be determined using appropriate projection operators.

The next step is to construct a free-energy polynomial on C (usually quartic) observing the requirement of G-invariance. Then, in order to find the order parameter, the free energy has to be minimized.[4] This step is highly nontrivial since it involves solving a system of c at least cubic equations (c is the dimension of C). There is no general procedure for solving such systems of equations exactly. However, after sufficiently many trials and errors scientists usually find the solutions analytically! This hints that there has to be a systematic procedure for solving those equations. Such a procedure could only be efficient, of course, for the cases of nontrivial G-invariance of the free energy.

This procedure we outline here for the first time.[5] All the isotropy groups L of G and their subduction frequencies $i(L)$ are found first. The subduction frequency $i(L)$ is equal to the number of times the identity representation of L is contained in the restriction of $\mathbf{R}$ to L. Geometrically, $i(L)$ is the dimensionality of the subspace $\mathrm{Fix}(L)$ of all vectors $\mathbf{c}$ left invariant by L. Due to the fact that the free energy is always extremal at $\mathrm{Fix}(L)$ in all directions perpendicular to $\mathrm{Fix}(L)$, it is sufficient to minimize the free energy only within the subspaces $\mathrm{Fix}(L)$. Moreover, due to the maximality conjecture,

[a)] This paper is based on an earlier version which was available as a preprint of the Institut des Hautes Etudes Scientifiques in March 1982 (Preprint No. IHES/P/82/17).

 0022-2488/83/122865-18$02.50

an absolute minimum of a quartic free energy is associated with a maximal isotropy group (among strict subgroups of G).[6,7] Consequently, once the isotropy groups are determined, we choose the maximal ones. Then, using appropriate projection operators, we determine associated subspaces Fix(L) where an absolute minimum should be found. Due to the maximality, these subspaces are often one-dimensional and the problem is reduced to solving a single quadratic equation (having removed the origin, as the trivial solution whose isotropy group is G itself). Therefore, we establish a clear relationship between maximality and the solvability of the Landau theory.

Even if the maximality conjecture is not to be found valid, a knowledge of isotropy groups helps in systematizing a search for absolute minima. Namely, if we know all isotropy groups L, then we can, using projection operators, partition the space C into the subspaces Fix(L). Searching for a minimum first in one-, then in two-, etc., dimensional subspaces Fix(L), i(L) = 1,2,...,c, always removing previously found solutions, the problem will again be often reduced to solving a single quadratic equation.

After an absolute minimum has been found, and the associated low-symmetry group has been identified, it only remains to apply the associated projection operator to the atomic displacements in order to determine the actual crystal structure. This would complete within the above mean-field approach the analysis of a particular transition.

In order to include fluctuations a renormalization-group calculation would suffice. Fluctuations have an additional influence on low-symmetry phases emerging in a second-order transition. However, such questions will not be addressed here. Instead, we will concentrate only on the problem of determining all isotropy groups L of a group G, given a representation **R**, since, irrespective of whether fluctuations are included or not, only such groups may occur in a transition.

The above mentioned problem has been answered by the chain criterion in general for discrete (countable) groups.[8] The criterion states[9]: A subgroup L of G is an isotropy group (given **R**) if and only if for every $L' \supset L$, $L' \subseteq G$ it follows $i(L') < i(L)$ (where $\subset$ denotes a strict subgroup relationship and $<$ is a strict inequality). Thus, in order to apply the chain criterion it is necessary to calculate and compare subduction frequencies for all subgroups of G. A technical interpretation of the chain criterion is that $i(L)$ is calculated for all subgroups L of G. Then, for each $i(L) = 0,1,2,...,c$, the maximal subgroups are the isotropy groups.[10] However, even when G is finite, the above proposed plan may be a very elaborate task and several lemmas come to help (some of them do not hold for continuous groups)[4]:

(i) If L is an isotropy group, so is every L' conjugate to L in G. Consequently, only subgroups up to a conjugation in G need to be considered;

(ii) Kernel of G is the minimal isotropy group (see below for a definition of the kernel);

(iii) If L' has a strict supergroup L, $L' \subset L \subseteq G$ and $i(L) = i(L')$, then L' is not an isotropy group. Moreover, none of the supergroups of L', which are strictly contained in L, is an isotropy group (thus the name *chain* criterion);

(iv) If L and L' are isotropy groups, then their intersection $L \cap L'$ is also an isotropy group. Moreover, $i(L \cap L') \geqslant i(L) + i(L') - i(LL')$, ($LL'$ is the minimal subgroup of G which contains both L and L');

(v) It suffices to consider only irreducible **R** for if **R** is reducible, then its isotropy groups are just intersections between isotropy groups of its irreducible components;

(vi) If **R** is reducible, but physically irreducible, $\mathbf{R} = \Delta \oplus \Delta^*$, then its isotropy groups are the same as those of (either of) its irreducible components, and $i(L; \mathbf{R}) = 2i(L; \Delta) = 2i(L; \Delta^*)$.

In the following sections we will present a systematic procedure designed to determine all isotropy groups of *crystallographic space groups*. The presentation will be self-contained. However, due to interests of most physicists some of the mathematical rigor will be avoided. The rigorous proofs of the theory require a separate paper which will be published elsewhere.[11]

The plan of the paper is the following. In Sec. II a general method and notation will be established. The notation related specifically to space groups and a review of some elementary theory will be given in Sec. IIA. Section IIB will deal with the image of a space group in general, while in Sec. IIC some details of the representation theory of space groups will be reviewed. Section III will be more concretely related to the determination of the isotropy groups of space groups. In Sec. IIIA a notion of a substar and its associated translation group will be introduced. In Sec. IIIB a point group of a substar will be defined and subsequent simplifications in calculating isotropy groups will be also elaborated. In Sec. IIIC some additional technical details will be given. A step by step algorithm for calculating all isotropy groups of a space group will be given in Sec. IIID. In Sec. IV a detailed illustration of an actual application of the algorithm will be given. The example will concern a nonsymmorphic space group O_h^3 (space group symmetry of A15 compounds) and its X-point irreducible representations. The last section of the paper will be devoted to the summary and discussions.

II. METHOD AND NOTATION

For a finite group L, $i(L)$ may be calculated by the formula

$$i(L) = \frac{1}{|L|} \sum_{g \in L} \chi(g), \tag{1}$$

where $|L|$ is the order of the group L and $\chi(g)$ are the characters in **R**. In the case of crystallographic space groups, most often (except for incommensurate transitions, that is, for representations associated with wave vectors having at least one irrational component) a relevant group L contains nontrivial translations. Thus, L is discrete (countable) but infinite and the above formula cannot be applied.[12] In such cases we must consider the group of matrices of the representation, also called the image, Im(G; **R**). When dealing with a particular representation **R**, **R** is often left out of the notation for the image. We will also often use **R** to name interchangeably a particular representation, its actual matrix group (or

any of its isomorphic groups; thus Im $G = \mathbf{R}$), or the homomorphism defined by the representation.

The homomorphism

$$g \longrightarrow \rho(g), \tag{2}$$

where $\rho(g)$ is a matrix in $\mathbf{R}$ representing an element g of G, defines a normal subgroup of G, the kernel $\mathrm{Ker}(G;\mathbf{R})$. Again, we will leave out $\mathbf{R}$ in the notation for the kernel.[13] Ker G contains all the elements of G which are represented by the identity matrix in $\mathbf{R}$. The associated quotient (factor) group $G/\mathrm{Ker}\ G$ is equal (isomorphic) to Im $G = \mathbf{R}$. This means, in particular, that a coset decomposition of G with respect to Ker G,

$$G = g_1 \circ \mathrm{Ker}\ G + g_2 \circ \mathrm{Ker}\ G + \cdots = \mathbf{R} \circ \mathrm{Ker}\ G, \tag{3}$$

provides a useful labeling of matrices in $\mathbf{R}$ by associated coset representatives in G. That is, every matrix ρ in $\mathbf{R}$ may be identified (labeled) by a coset representative g_i in $G/\mathrm{Ker}\ G$:

$$g_i \simeq \rho(g_i). \tag{4}$$

It is important, however, to stress that the coset representatives g_i form a group (isomorphic to $\mathbf{R}$) only under a composition law which is the same as in G (i.e., $\circ$), modulo Ker G. Therefore, given a matrix group $\mathbf{L} \subseteq \mathbf{R}$ it uniquely defines a group $L \subseteq G$

$$L = \mathbf{L} \circ \mathrm{Ker}\ G, \tag{5}$$

where elements in $\mathbf{L}$ are replaced by their coset representative labels according to Eq. (4).

Since we talk of isotropy groups of G relative to a given representation $\mathbf{R}$, it is clear that in order to find isotropy groups of G one should first find isotropy groups in $\mathbf{R}$. Associated isotropy groups in G can then simply be found using Eq. (5). Consequently, it suffices to apply the chain criterion to subgroups $\mathbf{L}$ of $\mathbf{R}$. Accordingly, Eq. (1) is replaced by

$$i(L) = i(\mathbf{L}) = \frac{1}{|\mathbf{L}|} \sum_{g \in \mathbf{L}} \chi(g), \tag{6}$$

where L and $\mathbf{L}$ are related via Eq. (5), g is a coset representative (matrix) [cf. Eq. (4)] and χ is its character (trace).

In all the relevant cases either all isotropy groups $\mathbf{L}$ of $\mathbf{R}$, except $\mathbf{R}$ itself, are finite (incommensurate transitions, irrational wave vectors), or $\mathbf{R}$ itself is finite (commensurate transitions, rational wave vectors). Thus, every isotropy group of $\mathbf{R}$, except $\mathbf{R}$ itself, is finite, and Eq. (6) is applicable. On the other hand, since $\mathbf{R}$ is irreducible,

$$i(\mathbf{R}) = 1 \tag{7}$$

if $\mathbf{R}$ is the identity representation (then $\mathbf{R}$ is the only one isotropy group), or

$$i(\mathbf{R}) = 0 \tag{8}$$

in all other cases ($\mathbf{R}$ is then the isotropy group of the origin $\mathbf{c} = 0$). In what follows we will exclude the trivial case when $\mathbf{R}$ is the identity representation.

In order to further motivate the use of coset representative labels, we note that even though Eq. (6) appears rather simple a selection of subgroups of $\mathbf{R}$ may be very complex. The matrix group $\mathbf{R}$ is sometimes a large subgroup of $O(c)$ or $U(c)$. In addition, c is often greater than three. Hence, it is unlikely to find $\mathbf{R}$ and its subgroups listed somewhere in the literature. On the other hand, direct manipulation of matrices in $\mathbf{R}$ requires first their explicit construction and then a construction of their multiplication table.

All this can be avoided by the use of coset representatives whose composition law is simple ($\circ$, modulo Ker G). Additionally, instead of the complete matrices in $\mathbf{R}$ we need only their traces (the characters). Finally, even if we had known $\mathbf{R}$ and its subgroups $\mathbf{L}$, the identification of the associated subgroups L in G is most easily found using the coset representatives via Eqs. (4) and (5).

We will develop below a systematic procedure (algorithm) for calculating all isotropy groups of a crystallographic space group. The algorithm will be based on the use of the chain criterion and the approach will be designed for the use of the full-group representations.

In the next sections we will establish the notation and we will review some elementary facts regarding crystallographic space groups and their representations. We will also introduce a notion of a substar for an irreducible representation of a space group defining associated translation and point groups.

A. Space group

A crystallographic space group G is an extension of a group of lattice translations T (Bravais lattice) by a point group P (crystallographic) class.[14,15] The abelian group T is a normal subgroup of G, and P is the associated quotient group,

$$P = G/T. \tag{9}$$

We will adopt the ("inverted") Seitz notation for the elements of G, T, and P. Furthermore, an action of the symmetry elements on the configuration space will be assumed active. Therefore, a general element g of G will be denoted by

$$g = \{\vec{\theta}\,|\overleftrightarrow{\phi}\}, \tag{10}$$

where $\overleftrightarrow{\phi}$ is a rotation-reflection from the point group P and $\vec{\theta}$ is a translation, not necessarily from T. If $\vec{\theta}$ is from T, that is, $\vec{\theta}$ is a lattice translation, then we will denote it by $\vec{n}$ or $\vec{t}$.

The action of g on a vector $\vec{r}$ in the configuration space is given by

$$g\cdot\vec{r} = \{\vec{\theta}\,|\overleftrightarrow{\phi}\}\cdot r = \vec{\theta} + \overleftrightarrow{\phi}\cdot\vec{r}, \tag{11}$$

which means that a point $\vec{r}$ is first rotated by $\overleftrightarrow{\phi}$ and then translated by $\vec{\theta}$. For the composition of group elements we will use a symbol $\circ$, to be distinguished from $\cdot$. Thus,

$$g' \circ g'' = \{\vec{\theta}\,'|\overleftrightarrow{\phi}\,'\} \circ \{\vec{\theta}\,''|\overleftrightarrow{\phi}\,''\} = \{\vec{\theta}\,' + \overleftrightarrow{\phi}\,'\cdot\vec{\theta}\,''|\overleftrightarrow{\phi}\,' \circ \overleftrightarrow{\phi}\,''\} \tag{12}$$

and

$$g^{-1} = \{\vec{\theta}\,|\overleftrightarrow{\phi}\}^{-1} = \{-\overleftrightarrow{\phi}^{-1}\cdot\vec{\theta}\,|\overleftrightarrow{\phi}^{-1}\}. \tag{13}$$

The identity element ϵ, for which we will usually reserve a subscript 1, is

$$\epsilon = g_1 = \{\vec{0}|\overleftrightarrow{\varepsilon}\}, \tag{14}$$

where $\vec{0}$ is the null vector ("no translation") and $\overleftrightarrow{\varepsilon}$ is the identity rotation ("no rotation"). Hence, a general pure translation $\vec{\theta}$, which may ($\vec{\theta} = \vec{n}$) but need not be in G, is

$$\{\vec{\theta}\,|\overleftrightarrow{\varepsilon}\} \equiv \vec{\theta} \tag{15}$$

and

$$\vec{\theta}'\circ\vec{\theta}'' = \vec{\theta}' + \vec{\theta}''. \tag{16}$$

Similarly, a general rotation-reflection by $\overleftrightarrow{\phi}$ which may (symmorphic groups) but need not (nonsymmorphic groups) be in G, is

$$\{\vec{0}|\overleftrightarrow{\phi}\} \equiv \overleftrightarrow{\phi}. \tag{17}$$

Note the distinction between $\vec{\theta}$ rotated by $\overleftrightarrow{\phi}$, $\overleftrightarrow{\phi}\cdot\vec{\theta}$, and followed by $\overleftrightarrow{\phi}$, $\overleftrightarrow{\phi}\circ\vec{\theta}$,

$$\overleftrightarrow{\phi}\circ\vec{\theta} = (\overleftrightarrow{\phi}\cdot\vec{\theta})\circ\overleftrightarrow{\phi}, \tag{18}$$

where we employed the short notation of Eqs. (15) and (17). In the same notation, Eq. (10) reads

$$g = \vec{\theta}\circ\overleftrightarrow{\phi}. \tag{19}$$

In order to associate the elements of P with particular elements *in* G it is convenient, according to Eq. (9), to make a coset decomposition of G with respect to T:

$$G = T\circ\{\vec{0}|\mathfrak{E}\} + T\circ\{\vec{\tau}_2|\overleftrightarrow{\phi}_2\} + \cdots. \tag{20}$$

Pure translations $\vec{\tau}$ are not, in general, lattice translations ($\vec{\tau}$ is called a fractional translation). Consequently, every pure translation $\vec{\theta}$, occurring in Eq. (19), can be decomposed into a lattice and a fractional translation,

$$\vec{\theta} = \vec{n} + \vec{\tau}. \tag{21}$$

Clearly, fractional translations are not uniquely defined. At the beginning of any calculation, it is advisable to fix the choice of fractional translations, for example as given in Ref. 16.

The point group P is conventionally given by [cf. Eq. (20)]

$$P = \{\mathfrak{E}, \overleftrightarrow{\phi}_2, \overleftrightarrow{\phi}_3, \ldots\} \tag{22}$$

with the same composition law as in G. However, we will identify P by the complete coset representatives (including fractional translations). Quite generally, coset representatives in a coset decomposition of a group G with respect to a normal subgroup T will not form a group with respect to the original law of composition (e.g., nonsymmorphic space groups). However, since T is a normal subgroup, it may happen that an appropriate choice of coset representatives can be made in such a way that they indeed form a group ($\simeq P$) under the original composition law (e.g., symmorphic space groups). Whenever such a special choice of coset representatives is possible, it is usually advantageous from a point of view of practical applications to adopt that particular choice.

We will identify P, generally, by the coset representatives

$$P = \{\{\vec{0}|\mathfrak{E}\}, \{\vec{\tau}_2|\overleftrightarrow{\phi}_2\}, \ldots\}, \tag{23}$$

with the composition law the same as in G, modulo T. A general element of P we will denote by p,

$$p = \{\vec{\tau}|\overleftrightarrow{\phi}\} = \vec{\tau}\circ\overleftrightarrow{\phi}. \tag{24}$$

Thus,

$$p_i\circ p_j = \vec{n}\circ p_k, \tag{25}$$

where $p_i, p_j, p_k \in P$ and $\vec{n}\in T$. With this particular choice of P, Eq. (23), in mind, we rewrite the coset decomposition of Eq. (20) simply as

$$G = T\circ P. \tag{26}$$

Similarly, every element of G can now be written as

$$g = \vec{n}\circ p = \vec{n}\circ\{\vec{\tau}|\overleftrightarrow{\phi}\} = \vec{n}\circ\vec{\tau}\circ\overleftrightarrow{\phi}. \tag{27}$$

This concludes our description of the space group elements. In the following sections we will introduce further definitions associated with a particular choice for the representation $\mathbf{R}$ of G.

B. Image

It was shown in the introductory part of Sec. II that the factor group of G with respect to Ker G, the image, is equal to $\mathbf{R}$. Also, a coset decomposition of G with respect to Ker G was introduced in Eq. (3), where in the case of a space group g_i is of the general form Eq. (10). If we fix a choice of coset representatives in Eq. (3), we can identify $\mathbf{R}$ as

$$\mathbf{R} = \{g_1, g_2, \ldots\}, \tag{28}$$

with the same composition law as in G, modulo Ker G. That is, for every g_i and g_j from $\mathbf{R}$ we can find an unique g_k from $\mathbf{R}$ and a g_0 from Ker G such that

$$g_i\circ g_j = g_k\circ g_0. \tag{29}$$

Therefore, we write the coset decomposition, Eq. (3), as

$$G = \mathbf{R}\circ\text{Ker } G. \tag{30}$$

At this stage we identified matrices of $\mathbf{R}$ by the associated coset representatives, Eq. (28). Thus, in order to find subgroups of $\mathbf{R}$, we do not have to multiply matrices of $\mathbf{R}$ explicitly. These matrices we need not even know. Instead, we can multiply the associated coset representatives with the appropriate multiplication rule, cf. Eq. (29).

Until this point we only particularized a general theory to a space group. Further simplifications in determining isotropy groups of space groups are strictly due to a special structure of a space group (e.g., existence of an abelian normal subgroup T, etc.).

The group Ker G is a space group (which degenerates into a point group for irrational wave vectors). Its translation group (Bravais lattice) is the kernel of T under $\mathbf{R}$,

$$\text{Ker}(T; R) \equiv \text{Ker } T = \{\vec{0}, \vec{n}_2, \vec{n}_3, \ldots\}; \tag{31}$$

Ker T consists of all translations in T which are represented by the identity matrix in $\mathbf{R}$. Ker T is a normal subgroup of G as well as of Ker G and, trivially, of T. Therefore, we make the following coset decompositions:

$$G = \text{Ker } T\circ\epsilon + \text{Ker } T\circ g_2' + \cdots, \tag{32}$$

$$\text{Ker } G = \text{Ker } T\circ\epsilon + \text{Ker } T\circ p_2' + \cdots, \tag{33}$$

and

$$T = \text{Ker } T\circ\vec{0} + \text{Ker } T\circ\vec{t}_2 + \cdots, \tag{34}$$

where g' and p' are of the general form Eq. (10) and $\vec{t}$ is a pure lattice translation. The associated quotient groups, with the composition law the same as in G, modulo Ker T, are [we assume that a particular choice of the coset representatives, Eqs. (32), (33), and (34), has been made]

$$G/\text{Ker } T = \{\epsilon, g_2', \ldots\}, \tag{35}$$

$$\overline{\text{Ker}}\ P \equiv \text{Ker } G/\text{Ker } T = \{\epsilon, p_2', \ldots\}, \tag{36}$$

and

$$\mathbf{T} \equiv T/\text{Ker } T = \{\vec{0}, \vec{t}_2, \ldots\}. \tag{37}$$

Lattice translations in Eq. (37) can be chosen as all the lattice translations of T which belong to a single unit cell of Ker T. (Incidentally, we note that it would suffice to use Born–Kármán boundary conditions at a single unit cell of Ker T; there is no need to take a limit of infinite periodicity—all the results associated with **R** are preserved.) Therefore, $\vec{0}, \vec{t}_2, \vec{t}_3, \ldots$ are fractional translations for the Bravais lattice Ker T. Similarly, translations entering g_i' and p_i' [cf. Eqs. (35) and (36)] can be chosen to be sums of translations from **T** and appropriate fractional translations $\vec{\tau}$ from Eq. (23).

The quotient group Ker G/Ker T is denoted by $\overline{\text{Ker}}\ P$ because it is isomorphic (equal, modulo **T**) to a normal subgroup of P and, furthermore, all of its elements are represented by the identity matrix in **R**. However, $\overline{\text{Ker}}\ P$ is not in general the kernel since **R** is not in general a representation of P.

In addition to the decompositions, Eqs. (32), (33), and (34), we can make now a coset decomposition of P,

$$P = \epsilon \circ \overline{\text{Ker}}\ P + \pi_2 \circ \overline{\text{Ker}}\ P + \cdots, \tag{38}$$

noting that the equality is modulo **T**; that is, once the fractionals in $\overline{\text{Ker}}\ P$ are chosen we can choose fractionals in P so that the equality is exact (this is, of course, not necessary). Elements π are of the general form Eq. (23).

We again construct an associated quotient group

$$\mathbf{P} \equiv P/\overline{\text{Ker}}\ P = \{\epsilon, \pi_2, \ldots\}, \tag{39}$$

with the composition law the same as in P, modulo $\overline{\text{Ker}}\ P$.

With the above identifications we can rewrite Eqs. (33), (34), and (38) as

$$\text{Ker } G = \text{Ker } T \circ \overline{\text{Ker}}\ P, \tag{40}$$

$$T = \text{Ker } T \circ \mathbf{T}, \tag{41}$$

and

$$P = \mathbf{P} \circ \overline{\text{Ker}}\ P. \tag{42}$$

Equations (41) and (42) are used to rewrite Eq. (26),

$$G = T \circ P = \text{Ker } T \circ \mathbf{T} \circ \mathbf{P} \circ \overline{\text{Ker}}\ P. \tag{43}$$

Similarly, Eq. (40) and the fact that Ker T is a normal subgroup of G are used to rewrite Eq. (30),

$$G = \mathbf{R} \circ \text{Ker } G = \text{Ker } T \circ \mathbf{R} \circ \overline{\text{Ker}}\ P. \tag{44}$$

Comparing Eqs. (43) and (44) we arrive at a simple formula for **R**,

$$\mathbf{R} = \mathbf{T} \circ \mathbf{P}. \tag{45}$$

Clearly, the resulting Eq. (45) is independent of whether we made previously left or right coset decompositions (with respect to a normal subgroup).

Thus, every element ρ of **R** can be identified as a product of two elements in G:

$$\rho = \vec{t} \circ \pi = \{\vec{t} \,|\vec{e}\} \circ \{\vec{\tau}|\vec{\phi}\}, \tag{46}$$

where $\vec{t}$ runs through all elements of **T** and π runs through all elements of **P**. Furthermore, **T** is a normal, abelian subgroup of **R**, and **P** is the associated factor group (it should be stressed again, however, that the composition law in **R** is the same as in G, modulo Ker G).

Since **P** is finite and, for representations **R** associated with a rational wave vector, **T** is finite, it follows that **R** is finite and Eq. (6) is applicable. On the other hand, if **R** is associated with an irrational wave vector then **T**, and thus **R**, are infinite. However, it is then clear that the only lattice translation in T which leaves a nontrivial vector in C invariant is the identity (that is, "no translation"). This is just another way of saying that the intersection of two incommensurate translation groups is the identity. Hence, in such a case an isotropy group **L**, other than **R** itself, is a subgroup of **P** only, thus finite, and Eq. (6) is again applicable.

Since **T** is a subgroup of **R** it follows that every subgroup **L** of **R** will have the following structure:

$$\mathbf{L} = \mathbf{T}_L \circ \mathbf{P}_L, \tag{47}$$

where

$$\mathbf{T}_L \subseteq \mathbf{T} \tag{48}$$

and

$$\mathbf{P}_L \subseteq \mathbf{P} \quad (\text{modulo } \mathbf{F}_L). \tag{49}$$

The quotient group $\mathbf{F}_L$,

$$\mathbf{F}_L = \mathbf{T}/\mathbf{T}_L, \tag{50}$$

is identified by the coset representatives in the coset decomposition of **T** with respect to $\mathbf{T}_L$. That is

$$\mathbf{T} = \mathbf{T}_L \circ \mathbf{F}_L. \tag{51}$$

Equation (47) implies that every element g_L of **L** can be expressed as

$$g_L = \vec{t}_L \circ \pi_L, \tag{52}$$

where $\vec{t}_L$ is from $\mathbf{T}_L$ and π_L is from $\mathbf{P}_L$. On the other hand, Eq. (49) implies

$$\pi_L = \vec{f}_L \circ \pi_L' = \vec{f}_L \circ \{\vec{\tau}_L|\vec{\phi}_L\} = \{\vec{f}_L + \vec{\tau}_L|\vec{\phi}_L\}, \tag{53}$$

where $\vec{f}_L$ is from $\mathbf{F}_L$ (that is, $\vec{f}_L$ becomes an additional fractional translation in $\mathbf{T}_L$ associated with $\vec{\phi}_L$ in $\mathbf{P}_L$) and where π_L' is an element of $\mathbf{P}_L'$, $\mathbf{P}_L' = \mathbf{P}_L$ (modulo $\mathbf{F}_L$),

$$\mathbf{P}_L' \subseteq \mathbf{P}. \tag{54}$$

Consequently, a systematic way of finding all subgroups of **R** would be first to find all subgroups $\mathbf{T}_L$ of **T** and all compatible subgroups $\mathbf{P}_L'$ of **P** (for example, P_L' must be such as to leave the "lattice" $\mathbf{T}_L$ invariant). Possible groups $\mathbf{L} = \mathbf{T}_L \circ \mathbf{P}_L$ are then identified by making appropriate selections of additional fractional translations $\vec{f}_L$ (one $\vec{f}_L$ to each π_L'!). Therefore, at no point is it necessary to consider explicitly matrices of **R** and their multiplication table.

Further simplifications will come from the fact that we are looking for the isotropy (sub) groups only. These simplifications will follow from the representation theory of space groups. However, before indulging in these intricacies, we can state two immediate results regarding isotropy groups: **R** (or G) and **E** (or Ker G), **E** consists of the identity element only, are isotropy groups having the subduction frequencies (**R** is assumed irreducible, different from the identity representation)

$$i(\mathbf{R}) = i(G) = 0, \tag{55}$$

and

$$i(\mathbf{E}) = i(\mathrm{Ker}\ G) = c, \tag{56}$$

where c, the dimension of the space C, is the degeneracy of the representation $\mathbf{R}$.

We will now briefly introduce the necessary notation of the theory of space group irreducible representations. We will, when possible, adopt the approach and the notation of Ref. 14.

C. Irreducible representations

Irreducible representations of space groups are induced from particular representations of their subgroups $G(\vec{k}_1)$. $G(\vec{k}_1)$ is the little group of a wave vector $\vec{k}_1$ from the first Brillouin zone. That is, $G(\vec{k}_1)$ contains all those elements in G whose rotation-reflections $\vec{\phi}$ leave $\vec{k}_1$ invariant, modulo a reciprocal lattice vector $\vec{m}$. Under the action of all elements $\vec{\phi}$ of P on $\vec{k}_1$, a set of different wave vectors $\{\vec{k}_1, \vec{k}_2, ..., \vec{k}_s\}$, called the star of $\vec{k}_1$ and denoted shortly $\{1,2,...,s\}$, will be generated. These wave vectors are in one-to-one correspondence with the coset representatives $\{\vec{\tau}^\sigma|\vec{\phi}^\sigma\}, \sigma = 1,2,...,s$, in the coset decomposition of G with respect to $G(\vec{k}_1)$:

$$\vec{k}_\sigma = \vec{\phi}^\sigma \vec{k}_1 + \vec{m}, \tag{57}$$

where $\vec{m}$ is a reciprocal lattice vector.

A little irreducible representation $D^{\vec{k}_1}$ of $G(\vec{k}_1)$, with characters $\chi^{\vec{k}_1}$, satisfies

$$D^{\vec{k}_1}(\vec{n}) = e^{i\vec{k}_1\cdot\vec{n}} I, \tag{58}$$

where $\vec{n}$ is a pure lattice translation and I is the c/s-dimensional identity matrix (note that we use a plus sign convention in the exponential, like in Ref. 17, instead of a minus sign in Ref. 14). The representation $\mathbf{R}$ induced from $D^{\vec{k}_1}$ is given in a block-matrix form,

$$\rho_{\alpha\beta}(g) = \dot{D}(\{\vec{\tau}^\alpha|\vec{\phi}^\alpha\}^{-1} \circ g \circ \{\vec{\tau}^\beta|\vec{\phi}^\beta\}), \quad \alpha,\beta = 1, 2, ..., s. \tag{59}$$

The dotted matrix $\dot{D}(g)$ is

$$\dot{D}(g) = \begin{cases} D^{\vec{k}_1}(g), & g \in G(\vec{k}_1), \\ 0, & g \notin G(\vec{k}_1). \end{cases} \tag{60}$$

Therefore, the full-group characters of $\mathbf{R}$ may be written as

$$\chi(g) = \sum_{\sigma=1}^{s} \chi_\sigma(g), \tag{61}$$

where

$$\chi_\sigma(g) = \dot{\chi}(\{\vec{\tau}^\sigma|\vec{\phi}^\sigma\}^{-1} \circ g \circ \{\vec{\tau}^\sigma|\vec{\phi}^\sigma\}), \tag{62}$$

and the dotted character $\dot{\chi}(g)$ is

$$\dot{\chi}(g) = \begin{cases} \chi^{\vec{k}_1}(g), & g \in G(\vec{k}_1), \\ 0, & g \notin G(\vec{k}_1). \end{cases} \tag{63}$$

Lattice translations $\vec{n}$ are represented by diagonal matrices in $\mathbf{R}$,

$$\rho_{\alpha\beta}(\vec{n}) = \delta_{\alpha\beta} e^{i\vec{k}_\alpha\cdot\vec{n}} I. \tag{64}$$

The associated character is

$$\chi(\vec{n}) = \frac{c}{s} \sum_{\sigma=1}^{s} e^{i\vec{k}_\sigma\cdot\vec{n}}. \tag{65}$$

Similarly, the character of any element $\vec{n} \circ g$ is

$$\chi(\vec{n} \circ g) = \sum_{\sigma=1}^{s} e^{i\vec{k}_\sigma\cdot\vec{n}} \chi_\sigma(g). \tag{66}$$

Consequently, in order to know characters of all elements in G, it suffices to know only the partial characters χ_σ for the coset representatives, Eq. (20), forming the group P.

These are all the results of the representation theory of space groups that are required for further simplifications in a calculation of the isotropy groups.

III. ISOTROPY GROUPS

Using the results established in previous sections, we can simplify Eq. (6), which gives the subduction frequency. By means of Eq. (61) we first rewrite Eq. (6) as

$$i(\mathbf{L}) = \frac{1}{|\mathbf{L}|} \sum_{g \in \mathbf{L}} \sum_{\sigma \in \{1,2,...,s\}} \chi_\sigma(g). \tag{67}$$

This can be further developed using Eqs. (47), (52), and (53):

$$i(\mathbf{L}) = \sum_{\sigma \in \{1,2,...,s\}} \frac{1}{|\mathbf{T}_L||\mathbf{P}_L|} \sum_{t_L \in \mathbf{T}_L} \sum_{\pi_L \in \mathbf{P}_L} \chi_\sigma(\vec{t}_L \circ \vec{f}_L \circ \pi'_L). \tag{68}$$

The last equation, in conjunction with Eq. (66) gives

$$i(\mathbf{L}) = \sum_{\sigma \in \{1,...,s\}} \left[\frac{1}{|\mathbf{T}_L|} \sum_{t_L \in \mathbf{T}_L} e^{i\vec{k}_\sigma\cdot\vec{t}_L}\right]\left[\frac{1}{|\mathbf{P}_L|} \sum_{\pi_L \in \mathbf{P}_L} e^{i\vec{k}_\sigma\cdot\vec{f}_L} \chi_\sigma(\pi'_L)\right]. \tag{69}$$

If we introduce the notation

$$i_\sigma(\mathbf{T}_L) \equiv \frac{1}{|\mathbf{T}_L|} \sum_{t_L \in \mathbf{T}_L} e^{i\vec{k}_\sigma\cdot\vec{t}_L} \tag{70}$$

and

$$\tilde{i}_\sigma(\mathbf{P}_L) \equiv \frac{1}{|\mathbf{P}_L|} \sum_{\pi_L \in \mathbf{P}_L} e^{i\vec{k}_\sigma\cdot\vec{f}_L} \chi_\sigma(\pi'_L), \tag{71}$$

then $i(\mathbf{L})$ can be written in a simple, condensed form:

$$i(\mathbf{L}) = \sum_{\sigma \in \{1,2,...,s\}} i_\sigma(\mathbf{T}_L)\tilde{i}_\sigma(\mathbf{P}_L). \tag{72}$$

Particularly important is the meaning of $i_\sigma(\mathbf{T}_L)$. It is immediately seen that $i_\sigma(\mathbf{T}_L)$ is the subduction frequency of $\mathbf{T}_L$ for the irreducible representation $\vec{k}_\sigma$ of $\mathbf{T}_L$ (or, more precisely, of $T_L = \mathrm{Ker}\ T \circ \mathbf{T}_L$). Thus, $i_\sigma(\mathbf{T}_L)$ is equal to one if $\vec{k}_\sigma$ belongs to the reciprocal lattice of T_L and it is equal to zero otherwise:

$$i_\sigma(\mathbf{T}_L) = \begin{cases} 1, & \vec{k}_\sigma \in \tilde{T}_L, \\ 0, & \vec{k}_\sigma \notin \tilde{T}_L, \end{cases} \tag{73}$$

where we denote the reciprocal lattice of T_L by $\tilde{T}_L$.

The above result implies that the only $\mathbf{L} \subset \mathbf{R}$ with $i(\mathbf{L}) \neq 0$ [for $i(\mathbf{L}) = 0$ there is only one isotropy group, $\mathbf{L} = \mathbf{R}$] are such that their $\tilde{T}_L$ contain a nonempty subset $\{\vec{k}_\alpha, \vec{k}_\beta, ...\}$ of the star $\{1,2,...,s\}$ (we will also use the shortened notation $\{\alpha,\beta, ...\}$ for these subsets). The same conclusion may be reached in another way if we first realize that $i(\mathbf{L}) > 0$ implies $i(\mathbf{T}_L) > 0$. Then since

$$i(\mathbf{T}_L) = \sum_{\sigma \in \{1,2,...,s\}} i_\sigma(\mathbf{T}_L) \frac{c}{s}, \tag{74}$$

we arrive at the same restrictions on $\mathbf{T}_L$. Therefore, it appears that "allowed" translation subgroups $\mathbf{T}_L$, that is, the only translation subgroups which may occur as the transla-

tion subgroups of isotropy groups, can conveniently be labeled by subsets $\{\alpha, \beta, ...\}$ of the star $\{1, 2, ..., s\}$.

The above derivation hints that $\tilde{T}_L$ (rather than T_L) is to be determined in practice directly from nonequivalent subsets $\{\alpha, \beta, ...\}$ of the star $\{1,2,...,s\}$. Here we refer to P-equivalence: $\{\alpha', \beta', ...\}$ is P-equivalent to $\{\alpha'', \beta'', ...\}$ if, by definition, there is a rotation-reflection $\overleftrightarrow{\phi}$ in some $p \in P$ (it suffices to look for $p \in \mathbf{P}$) such that

$$\{\overleftrightarrow{\phi}\cdot\vec{k}_{\alpha'}, \overleftrightarrow{\phi}\cdot\vec{k}_{\beta'}, ...\} = \{\vec{k}_{\alpha''} + \vec{m}_{\alpha''}, \vec{k}_{\beta''} + \vec{m}_{\beta''}, ...\}, \quad (75)$$

where $\vec{m}_{\alpha''}, \vec{m}_{\beta''}, ... \in \tilde{T}$ and the equality is the equality of two sets regardless of the ordering of elements. An immediate consequence of this definition is that the two subsets can only be P-equivalent if they contain the same number of wave vectors from the star. It should be noted that this P-equivalence is to be distinguished from a T-equivalence which will be introduced in the next section.

A. Substar. Translation group

If we take a subset $\{\alpha', \beta', ...\}$ then we can generate (it suffices to stay in the first Brillouin zone of T) a reciprocal lattice $\tilde{T}_L$, in the following way. A general element $\vec{m}_L$ of $\tilde{T}_L$ is given by

$$\vec{m}_L = \vec{m} + l_{\alpha'}\vec{k}_{\alpha'} + l_{\beta'}\vec{k}_{\beta'} + \cdots, \quad (76)$$

where $\vec{m}$ is a reciprocal lattice translation, $\vec{m} \in \tilde{T}$, and $l_{\alpha'}, l_{\beta'}, ...$ are integers. Therefore, we will denote $\tilde{\mathbf{T}}_L$, the intersection of $\tilde{T}_L$ and the first Brillouin zone of T, symbolically as

$$\tilde{\mathbf{T}}_L = \tilde{\mathbf{T}} + \{\alpha', \beta', ...\}. \quad (77)$$

The lattice $\tilde{\mathbf{T}}_L$ obtained in the above manner may contain additional vectors $\vec{k}_{\alpha''}, \vec{k}_{\beta''}, ...$ which are from the star but which are not in the original subset $\{\alpha', \beta', ...\}$. In such a case the subset $\{\alpha', \beta', ...\}$ may be enlarged by all such vectors $\vec{k}_{\alpha''}, \vec{k}_{\beta''}, ...$. The enlarged subset $\{\alpha', \beta', ..., \alpha'', \beta'', ...\} \equiv \{\alpha, \beta, ...\}$, which is the intersection of $\tilde{\mathbf{T}}_L$ and the star $\{1,2,...,s\}$, may be used to uniquely label the allowed lattices $\mathbf{T}_L$.[11] Such a subset, $\{\alpha, \beta, ...\}$, we will call a *substar*. In order to distinguish a substar from a subset we will denote a substar by $[\alpha, \beta, ...]$. By definition

$$[\alpha, \beta, ...] = \tilde{\mathbf{T}}_L \cap \{1,2,...,s\} = (\tilde{\mathbf{T}} + [\alpha, \beta, ...]) \cap \{1,2,...,s\}, \quad (78)$$

which may in turn be taken as the defining relation for a substar $[\alpha, \beta, ...]$.

The existence of a substar leads naturally to a notion of T-equivalence among subsets of a star: two subsets $\{\alpha', \beta', ...\}$ and $\{\alpha'', \beta'', ...\}$ are T-equivalent if, by definition, they generate the same reciprocal lattice $\tilde{\mathbf{T}}_L$. Therefore, a substar may be defined as the maximal subset among T-equivalent ones. That is, a substar is the union of all T-equivalent subsets. In contrast to P-equivalent subsets, two T-equivalent subsets can have a different number of elements.

Each substar $[\alpha, \beta, ...]$, determines an allowed reciprocal lattice $\tilde{\mathbf{T}}_L \equiv \tilde{\mathbf{T}}[\alpha, \beta, ...]$,

$$\tilde{\mathbf{T}}[\alpha, \beta, ...] = \tilde{\mathbf{T}} + [\alpha, \beta, ...], \quad (79)$$

whose reciprocal lattice $\mathbf{T}[\alpha, \beta, ...]$, equal to an allowed translation group $\mathbf{T}_L$, is unique. It should be remembered that an allowed translation group $T[\alpha, \beta, ...]$ is always given by

$$T[\alpha, \beta, ...] = \text{Ker } T \circ \mathbf{T}[\alpha, \beta, ...]. \quad (80)$$

The following remarks should help us to choose a strategy for determining all possible substars and associated lattices.

If a subset $\{\alpha', \beta', ...\}$ generates a substar $[\alpha, \beta, ...]$, then a subset $\{\alpha'', \beta'', ...\}$ which satisfies

$$\{\alpha', \beta', ...\} \subset \{\alpha'', \beta'', ...\} \subseteq [\alpha, \beta, ...] \quad (81)$$

needs not to be considered since it would generate the same substar. Furthermore, if $\vec{k}_\alpha$ is in a substar then $-\vec{k}_\alpha$ belongs to the associated $\tilde{\mathbf{T}}[\alpha, \beta, ...]$. Thus, whenever $\vec{k}_\alpha$ and $-\vec{k}_\alpha$ belong to the star, then $\vec{k}_\alpha$ and $-\vec{k}_\alpha$ occur in the same substars. Note also that the star $\{1,2,...,s\}$ is trivially a substar and

$$\mathbf{T}[1, 2, ..., s] = \mathbf{E}. \quad (82)$$

We can also consider an "empty" substar, denoted [0], in order to obtain

$$\mathbf{T}[0] = \mathbf{T}, \quad (83)$$

which leads to $\mathbf{L} = \mathbf{R}$.

An effective procedure for determining substars and associated translation subgroups is to work in an ascending fashion from one vector subsets to the full star. Working in an ascending fashion means that a subset should be examined before any other subset containing it has been examined. In this process the above-mentioned remarks, and in particular Eq. (81), should be observed. Furthermore, one needs only to consider non-P-equivalent subsets [see Lemma (i) in Sec. I]. Following this procedure a set of all substars $[\alpha, \beta, ...]$ and a set of associated translation subgroups $\mathbf{T}[\alpha, \beta, ...]$ will be found. We emphasize that these will be the only possible translation subgroups of the isotropy groups.

Utilizing the above results the subduction frequency of $\mathbf{T}[\alpha, \beta, ...]$, Eq. (73), is

$$i_\sigma(\mathbf{T}[\alpha, \beta, ...]) = \begin{cases} 1, & \sigma \in [\alpha, \beta, ...], \\ 0, & \sigma \notin [\alpha, \beta, ...]. \end{cases} \quad (84)$$

As a consequence the subduction frequency of an isotropy group, Eq. (72), is simplified to

$$i(\mathbf{L}) = \sum_{\sigma \in [\alpha, \beta, ...]} \tilde{i}_\sigma(\mathbf{P}_L), \quad (85)$$

where

$$\mathbf{L} = \mathbf{T}[\alpha, \beta, ...] \circ \mathbf{P}_L \quad (86)$$

and $[\alpha, \beta, ...]$ is a substar.

The quantity $\tilde{i}_\sigma(\mathbf{P}_L)$ needs to be calculated next. Thus, any further simplifications will stem from an analysis of $\mathbf{P}_L$.

B. Substar. Point group

In the following analysis we will assume a substar $[\alpha, \beta, ...]$ and the associated translation group $\mathbf{T}[\alpha, \beta, ...]$ to be determined. This automatically restricts possible additional fractional translations $\vec{f}_L$ associated with the elements in $\mathbf{P}_L$. They must, cf. Eq. (50), belong to the quotient group $\mathbf{F}[\alpha, \beta, ...]$,

$$\mathbf{F}[\alpha, \beta, ...] \equiv \mathbf{T}/\mathbf{T}[\alpha, \beta, ...]. \quad (87)$$

An obvious restriction on rotation-reflections $\overleftrightarrow{\phi}_L$ of $\mathbf{P}_L$ is that they must leave the lattice $\mathbf{T}[\alpha, \beta, ...]$ invariant. Due to

the maximality property of substars (with respect to T-equivalence), this restriction on $\overleftrightarrow{\phi}_L$ is equivalent to a requirement that $\overleftrightarrow{\phi}_L$ should leave the substar $[\alpha, \beta, ...]$ invariant (modulo $\tilde{T}$). Therefore, we are naturally led to generalizing the notion of a little group: the little group $\mathbf{R}[\alpha, \beta, ...]$ of a substar $[\alpha, \beta, ...]$ is a subgroup of $\mathbf{R}$ which consists of all elements whose rotation-reflections leave the substar invariant. The associated little space group is given after Eq. (5) as

$$G[\alpha, \beta, ...] = \mathbf{R}[\alpha, \beta, ...] \circ \mathrm{Ker}\, G. \tag{88}$$

Similarly, the little point-group $\mathbf{P}[\alpha, \beta, ...]$ of a substar $[\alpha, \beta, ...]$ is a subgroup of $\mathbf{P}$ which consists of all elements whose rotation-reflections leave the substar invariant. Note that

$$P[\alpha, \beta, ...] = \mathbf{P}[\alpha, \beta, ...] \circ \overline{\mathrm{Ker}}\, P. \tag{89}$$

Hence,

$$\mathbf{R}[\alpha, \beta, ...] = \mathbf{T} \circ \mathbf{P}[\alpha, \beta, ...] \tag{90}$$

(the *set* $\overline{\mathbf{R}}[\alpha, \beta, ...]$,

$$\overline{\mathbf{R}}[\alpha, \beta, ...] = \mathbf{T}[\alpha, \beta, ...] \circ \mathbf{P}[\alpha, \beta, ...], \tag{91}$$

which is included in $\mathbf{R}[\alpha, \beta, ...]$, is not in general a group).

With the above definitions established we see that an isotropy group $\mathbf{L}$ must be a subgroup of $\mathbf{R}[\alpha, \beta, ...]$,

$$\mathbf{L} \subseteq \mathbf{R}[\alpha, \beta, ...], \tag{92}$$

and $\mathbf{P}'_L$, Eq. (54), must be a subgroup of $\mathbf{P}[\alpha, \beta, ...]$,

$$\mathbf{P}'_L \subseteq \mathbf{P}[\alpha, \beta, ...]. \tag{93}$$

The difference between $\mathbf{P}'_L$ and $\mathbf{P}_L$, Eqs. (49) and (54), is essential due to the possible additional fractional translations coming from $\mathbf{F}[\alpha, \beta, ...]$.

The construction of a subgroup $\mathbf{L}$ proceeds by choosing a subgroup $\mathbf{P}'_L$ of $\mathbf{P}[\alpha, \beta, ...]$. $\mathbf{P}'_L$ is then amended by additional fractional translations from $\mathbf{F}[\alpha, \beta, ...]$ in order to construct $\mathbf{P}_L$. A choice of additional fractional translations (several different choices may be possible) must be made in such a way that $\mathbf{L} = \mathbf{T}[\alpha, \beta, ...] \circ \mathbf{P}_L$ is indeed a group. If such a choice is not possible, then another $\mathbf{P}'_L \subseteq \mathbf{P}[\alpha, \beta, ...]$ must be chosen. Consequently, a general element π_L in $\mathbf{P}_L$ will be given by Eq. (53) with

$$\vec{f}_L \in \mathbf{F}[\alpha, \beta, ...] \tag{94}$$

and

$$\pi'_L \in \mathbf{P}'_L \subseteq \mathbf{P}[\alpha, \beta, ...]. \tag{95}$$

With each π'_L only one additional fractional translation is associated.[18]

Equation (85), with Eq. (71) and Eqs. (86), (94), and (95), represent a significant simplification in determining potential isotropy groups and in calculating their subduction frequencies. These equations are the key to a systematic and effective procedure for a determination of all isotropy groups of $\mathbf{R}$.

Before describing the aforementioned procedure we will make several useful technical remarks.

C. Technical remarks

It was already mentioned that a choice of new fractional translations $\vec{f}_L$ must be made in such a way that the group property is satisfied. That is, a composition of two elements $\pi_{L1} = \vec{f}_{L1} \circ \pi'_{L1}$ and $\pi_{L2} = \vec{f}_{L2} \circ \pi'_{L2}$ from $\mathbf{P}_L$ must give a third element $\pi_{L3} = \vec{f}_{L3} \circ \pi'_{L3}$ from $\mathbf{P}_L$. More explicitly,

$$\pi_{L1} \circ \pi_{L2} = \vec{n} \circ (\vec{t}_L \circ \vec{\pi}_{L3}) \circ p, \tag{96}$$

where $\vec{t}_L$ is a lattice translation from $\mathbf{T}[\alpha, \beta, ...]$, $\vec{n}$ is a lattice translation from Ker T and p is from $\overline{\mathrm{Ker}}\, P$. This places a severe restriction on possible choices of the additional fractional translations. It is due to this restriction, for example, that $\overline{\mathbf{R}}[\alpha, \beta, ...]$, Eq. (91), is not a group in general (this is a particular choice, $\mathbf{P}'_L = \mathbf{P}[\alpha, \beta, ...]$ and all $\vec{f}_L = \vec{0}$). A direct consequence of Eq. (96) is that if $\{\vec{\tau}_L | \overleftrightarrow{\phi}_L\}$ is in $\mathbf{P}'_L$ and $(\overleftrightarrow{\phi}_L)^q = \vec{\varepsilon}$ (q is either 1,2,3,4 or 6), then the associated additional fractional translations $\vec{f}_L$ must satisfy

$$(\vec{f}_L \circ \{\vec{\tau}_L | \overleftrightarrow{\phi}_L\})^q = \vec{n} \circ \vec{t}_L, \tag{97}$$

where $\vec{t}_L \in \mathbf{T}[\alpha, \beta, ...]$ and $\vec{n} \in \mathrm{Ker}\, T$. This simpler restriction often suffices to make consistent choices of the additional fractional translations.

Another, indirect check of the group property for a particular choice of the additional fractional translations comes from a requirement that the subduction frequency shall be a nonnegative integer. This requirement also eliminates some of the choices for the point group $\mathbf{P}'_L$.

The next remark regards the actual applications of the chain criterion. Namely, from the formulation of the criterion it is clear that subgroups $\mathbf{L}$ should be examined in a descending fashion. That is, a group $\mathbf{L}$ should be considered only after all of its supergroups have been considered. In this fashion the subduction frequency $i(\mathbf{L})$ of a subgroup $\mathbf{L}$ needs to be compared with the subduction frequencies $i(\mathbf{L}')$ of previously examined subgroups $\mathbf{L}'$. If it then turns out that there is a supergroup $\mathbf{L}'$ of $\mathbf{L}$ such that $i(\mathbf{L}') = i(\mathbf{L})$, then $\mathbf{L}$ is *not* an isotropy group; otherwise, $\mathbf{L}$ is an isotropy group.

This principle, proceeding in a descending fashion, is useful to apply along two lines. One line is making a selection of translation subgroups $\mathbf{T}_L = \mathbf{T}[\alpha, \beta, ...]$ in a descending fashion (which is equivalent to selecting substars $[\alpha, \beta, ...]$ in an ascending fashion). The other line is that at any given $\mathbf{T}[\alpha, \beta, ...]$ a selection of point groups $\mathbf{P}'_L \subseteq \mathbf{P}[\alpha, \beta, ...]$ is made in a descending fashion. This is convenient for the following reason. If $\mathbf{L}'$ is a supergroup of $\mathbf{L}$ [$\mathbf{L}'$ and $\mathbf{L}$ are assumed to be of the general form Eq. (86)], then there are two possibilities,

$$\mathbf{T}_L = \mathbf{T}_{L'} = \mathbf{T}[\alpha, \beta, ...] \tag{98}$$

or

$$\mathbf{T}_L = \mathbf{T}[\alpha, \beta, ...] \subset \mathbf{T}_{L'} = \mathbf{T}[\alpha', \beta', ...]. \tag{99}$$

In the first case it is necessary that

$$\mathbf{P}'_L \subset \mathbf{P}'_{L'} \subseteq \mathbf{P}[\alpha, \beta, ...], \tag{100}$$

whereas in the second case

$$\mathbf{P}'_L \subseteq \mathbf{P}'_{L'} \subseteq \mathbf{P}[\alpha', \beta', ...]. \tag{101}$$

Therefore, when the subduction frequency $i(\mathbf{L})$ is calculated it will be compared first with all the supergroups $\mathbf{L}'$ of $\mathbf{L}$ (which will have all been already calculated due to the "descending" principle) with the same translation group $\mathbf{T}[\alpha, \beta, ...] = \mathbf{T}_L = \mathbf{T}_{L'}$. Then, if there is no such supergroup $\mathbf{L}'$ with $i(\mathbf{L}') = i(\mathbf{L})$, $\mathbf{L}$ *may be* an isotropy group. Otherwise, $\mathbf{L}$ is not an isotropy group. In case $\mathbf{L}$ passes this test, $i(\mathbf{L})$ has

to be compared with $i(\mathbf{L}')$ of its supergroups $\mathbf{L}'$ with larger translation groups $\mathbf{T}[\alpha', \beta', ...] = \mathbf{T}_{L'}$ (that is, smaller substars $[\alpha', \beta', ...]$) satisfying Eq. (101). These will have also been already calculated due to the "descending" principle. Then, if there is no such supergroup $\mathbf{L}'$ with $i(\mathbf{L}') = i(\mathbf{L})$, $\mathbf{L}$ is an isotropy group. Otherwise, $\mathbf{L}$ is not an isotropy group.

Suppose that in the second case discussed above there is an $\mathbf{L}'$ with $i(\mathbf{L}) = i(\mathbf{L}')$. Then we can always construct a group $\mathbf{L}''$,

$$\mathbf{T}_{L''} = \mathbf{T}_{L'} = \mathbf{T}[\alpha', \beta', ...] \tag{102}$$

and

$$\mathbf{P}_{L''} = \mathbf{P}_{L}, \tag{103}$$

such that $\mathbf{L} \subset \mathbf{L}'' \subseteq \mathbf{L}'$ (that is, we break $\mathbf{L} \subset \mathbf{L}'$ into its equitranslational, $\mathbf{L}'' \subseteq \mathbf{L}'$, and equirotational, $\mathbf{L} \subset \mathbf{L}''$, parts) and

$$i(\mathbf{L}) = i(\mathbf{L}'') = i(\mathbf{L}'). \tag{104}$$

On the other hand, Eqs. (85) and (71) imply

$$i(\mathbf{L}) = i(\mathbf{L}'') + \sum_{\sigma \in [\alpha, \beta, ...] \setminus [\alpha', \beta', ...]} \tilde{i}_{\sigma}(\mathbf{P}_L), \tag{105}$$

where $A \setminus B$ denotes elements in A but not in B. Consequently, $i(\mathbf{L}) = i(\mathbf{L}')$ implies

$$\sum_{\sigma \in [\alpha, \beta, ...] \setminus [\alpha', \beta', ...]} \tilde{i}_{\sigma}(\mathbf{P}_L) = 0. \tag{106}$$

Conversely, Eq. (106) implies $i(\mathbf{L}) = i(\mathbf{L}'')$ and $\mathbf{L}$ will not be an isotropy group. Therefore, the second case is reformulated: if there is no such supergroup $\mathbf{L}'$ [satisfying Eqs. (99) and (101)] with Eq. (106) fulfilled $\mathbf{L}$ is an isotropy group. Otherwise $\mathbf{L}$ is not an isotropy group.

As a final remark we mention another merit of the "descending" principle. Due to the chain property of the chain criterion, we do not need to compare $i(\mathbf{L}')$ and $i(\mathbf{L})$ for all supergroups $\mathbf{L}'$ of $\mathbf{L}$. It suffices to compare $i(\mathbf{L}')$ and $i(\mathbf{L})$ of only those supergroups $\mathbf{L}'$ of $\mathbf{L}$ which have already been found to be isotropy groups. Furthermore, due to a similar argument it suffices to check only the immediate isotropy supergroups $\mathbf{L}'$ of $\mathbf{L}$. That is, an isotropy supergroup $\mathbf{L}''$ of $\mathbf{L}$ needs not to be compared if there is an isotropy supergroup $\mathbf{L}'$ of $\mathbf{L}$ which is a subgroup of $\mathbf{L}''$. This general remark applies equally to equitranslational and equirotational examinations of supergroups of $\mathbf{L}$. We also note that a similar type of an argument led to a conclusion after Eq. (106).

D. Algorithm

We are now in a position to formulate a systematic step by step procedure for finding *all* isotropy groups associated with a particular representation $\mathbf{R}$ of G.

Step 0: As a preliminary step, establish a particular choice of P and find partial characters χ_σ, Eq. (62), for elements of P. Then, from the character table, find all elements $\vec{n} \circ p$ of Ker G which obey

$$\chi_\sigma(\vec{n} \circ p) = c/s, \quad \sigma = 1, 2, ..., s. \tag{107}$$

This automatically gives the elements of Ker T and $\overline{\text{Ker}}\ P$. Next, construct groups $\mathbf{T}$ and $\mathbf{P}$, Eqs. (37) and (39).

Step 1: Proceeding in an ascending fashion determine all substars $[\alpha, \beta, ...]$. In the process construct all $\mathbf{T}[\alpha, \beta, ...]$, $\mathbf{F}[\alpha, \beta, ...]$, and $\mathbf{P}[\alpha, \beta, ...]$ (Secs. IIIA and IIIB).

Step 2: Select a $\mathbf{T}[\alpha, \beta, ...]$ in a descending fashion.

Step 3: Select a $\mathbf{P}'_L \subseteq \mathbf{P}[\alpha, \beta, ...]$ in a descending fashion.

Step 4: Select a set of compatible additional fractional translations from $\mathbf{F}[\alpha, \beta, ...]$ and form $\mathbf{L}$. If no such set exists proceed to Step 8. [In Steps 3 and 4 a group property must be observed and, for $\mathbf{L} \neq \mathbf{R}$, $i(\mathbf{L})$ must be a positive integer].

Step 5: Calculate $i(\mathbf{L})$.

Step 6: Check, at the same substar, whether there is an immediate isotropy supergroup $\mathbf{L}'$ of $\mathbf{L}$ such that $i(\mathbf{L}') = i(\mathbf{L})$ [equitranslational examination, Eqs. (98) and (100)]. If the answer is no, proceed to the next step. If the answer is yes, $\mathbf{L}$ is not an isotropy group; proceed to Step 8.

Step 7: Check if there is an immediate isotropy supergroup $\mathbf{L}'$ of $\mathbf{L}$ at a substar $[\alpha', \beta', ...]$ contained in $[\alpha, \beta, ...]$ such that Eq. (106) is fulfilled [equirotational examination, Eqs. (99) and (101)]. If the answer is no, $\mathbf{L}$ is an isotropy group; proceed to the next step. If it is yes, $\mathbf{L}$ is not an isotropy group.

Step 8: Make the next selection at Step 4. If all selections have been tested, make the next selection at Step 3. If all selections have been tested, make the next selection at Step 2. If all selections have been tested, all of the isotropy groups $\mathbf{L}$ of $\mathbf{R}$ have been determined.

Step 9: Determine all isotropy space groups L, each one associated with each one isotropy group $\mathbf{L}$ via Eq. (5).

The above procedure completely solves the problem of determining isotropy groups L. However, several points should be emphasized.

At various stages of calculations Lemmas (i) to (vi) of Sec. I may be used as a help and/or as an independent check.

It is clear that only isotropy groups which are not equivalent under conjugation in $\mathbf{R}$ need to be considered. In particular, only non-P-equivalent substars need to be considered. In the case of a phase transition conjugated isotropy groups are associated with different domains of the same phase.[19] However, in the case that the conjugation element is an improper rotation, the two isotropy space groups may be physically different (since an improper rotation changes the handedness of the coordinate frame), having different space group labels, and both should be listed.

In the case where $\mathbf{R}$ is small or its structure (the lattice of subgroups, etc.) known, a direct determination of isotropy subgroups may be feasible. Then Steps 1 to 8 are omitted but an identification of complete matrices in $\mathbf{R}$ and their identification with elements in G must be made in Step 0.

When an isotropy space group L has been found its elements will be expressed in the same coordinate system in which the elements of G have been defined. In order to identify L as a particular space group a different coordinate system, such as given in Ref. 16, may be required. For example, if the original coordinate system is translated by $\vec{\tau}_0$ and then rotated by $\vec{\phi}_0$ the elements of L expressed in the new coordinate system are exactly the same as those of a conjugate space group L_0,

$$L_0 = \{\vec{\tau}_0 | \vec{\phi}_0\}^{-1} \circ L \circ \{\vec{\tau}_0 | \vec{\phi}_0\}, \tag{108}$$

expressed in the original coordinate system.

The isotropy groups L should be listed by their conventional space-group name with an identification of their Bravais lattices (in terms of the Bravais lattice T) and with an identification of the shift of the origin $\vec{\tau}_0$, and of relative orientation of the axes $\overleftrightarrow{\phi}_0$ ($\{\vec{\tau}_0|\overleftrightarrow{\phi}_0\}$ brings L into a standard setting L_0 while G may be taken from the start in such a standard setting). A standard setting may be taken, for example, as given in Ref. 16. Additional useful informations to be listed are the subduction frequency $i(L)$ and the number of domains different but equivalent under conjugation in G, excluding those with different handedness. The equivalent space groups which are not conjugate in G should be clearly distinguished (they would have $\{\vec{\tau}_0|\overleftrightarrow{\phi}_0\}$ differing by an element not in G).

In the next section we will demonstrate on a concrete example an effective application of the algorithm presented above.

IV. APPLICATION

In this section we will present a detailed application of the algorithm developed above. We will consider a space-group O_h^3. This space group is interesting for several reasons. For example, many compounds have a crystallographic structure compatible with this space group. Among such compounds are much studied high temperature superconductors, like V_3Si, which share A15 structure. Furthermore, O_h^3 as a nonsymmorphic space group illustrates peculiarities associated with such a more general case.

The group theory of O_h^3 has been discussed by various authors[20] and the irreducible representations are known. These representations may be obtained from several tables of space-group representations.[15,17,21]

We choose to consider the irreducible representation $X(3)$ (in the notation of Ref. 17). This representation is associated with a nonzero wave vector and it is six-dimensional. Thus, it offers generality and complexity sufficient for an appreciation of the algorithm. On the other hand, this representation is sufficiently simple that it can be presented in some detail here. X-point representations have also been a focus of interest in some theories of electronic structure of A15 high-temperature superconductors.[22,23]

The group O_h^3 has a simple cubic Bravais lattice T.[24] Therefore, we fix the coordinate frame parallel to the axes of the lattice and denote a pure lattice translation $\vec{n}$ by

$$\vec{n} = (n_1, n_2, n_3), \tag{109}$$

where n_1, n_2, n_3 are integer components of $\vec{n}$ in the above-mentioned frame. Generally, we will denote a vector in direct space by its components relative to this frame in which unit vectors correspond to a single lattice spacing.

The rotation-reflections,

$$\overleftrightarrow{\phi}_i, \quad i = 1, 2, ..., 48, \tag{110}$$

occurring in O_h^3 belong to the point group O_h. Therefore, P is isomorphic to O_h. We will use the following convention for $\overleftrightarrow{\phi}_i$:

$$\overleftrightarrow{\phi}_{i+12} = \overleftrightarrow{\phi}_{13} \circ \overleftrightarrow{\phi}_i, \quad i = 1, 2, ..., 12, \tag{111}$$

and

$$\overleftrightarrow{\phi}_{i+24} = \overleftrightarrow{\phi}_{25} \circ \overleftrightarrow{\phi}_i, \quad i = 1, 2, ..., 24, \tag{112}$$

where $\overleftrightarrow{\phi}_1$ is the identity, $\overleftrightarrow{\phi}_{13}$ is the inversion, and $\overleftrightarrow{\phi}_{25}$ is the twofold rotation around the $(1,\bar{1},0)$ direction (note $\bar{1} \equiv -1$). This is equivalent to the following decomposition of O_h:

$$O_h = \{\overleftrightarrow{\phi}_1, \overleftrightarrow{\phi}_{25}\} \circ T_h = \{\overleftrightarrow{\phi}_1, \overleftrightarrow{\phi}_{25}\} \circ \{\overleftrightarrow{\phi}_1, \overleftrightarrow{\phi}_{13}\} \circ T, \tag{113}$$

where T is the tetrahedral group (not to be confused with the Bravais lattice).

In order to determine coset representatives

$$p_i, \quad i = 1, 2, ..., 48, \tag{114}$$

forming P, we must make a choice of fractional translations $\vec{\tau}_i$. The conventional choice (Refs. 16 and 17) is

$$\vec{\tau}_i = \begin{cases} \vec{0}, & i = 1, 2, ..., 24, \\ \vec{\tau} \equiv (\frac{1}{2}, \frac{1}{2}, \frac{1}{2}), & i = 25, 26, ..., 48. \end{cases} \tag{115}$$

Therefore

$$p_i = \begin{cases} \{\vec{0}|\overleftrightarrow{\phi}_i\}, & i = 1, 2, ..., 24, \\ \{\vec{\tau}|\overleftrightarrow{\phi}_i\}, & i = 25, 26, ..., 48. \end{cases} \tag{116}$$

For the coset representatives p_i hold relationships similar to Eqs. (111) and (112):

$$p_{i+12} = p_{13} \circ p_i \quad i = 1, 2, ..., 12, \tag{117}$$

and

$$p_{i+24} = p_{25} \circ p_i \quad i = 1, 2, ..., 24. \tag{118}$$

For the sake of completeness explicit forms of the rotation-reflections $\overleftrightarrow{\phi}_i$ and of coset representatives p_i are given in Table I.

Since the Bravais lattice of O_h^3 is simple cubic, its reciprocal lattice will be also simple cubic. Therefore, the first Brillouin zone is a cube. We will consider a representation associated with the face centers (X-point). Therefore, let us choose[17]

$$\vec{k}_1 = (0, \tfrac{1}{2}, 0). \tag{119}$$

(We will always express the vectors in the direct space in units of the primitive lattice vectors, while the vectors in the reciprocal space, which includes a factor 2π, we will define in units of primitive reciprocal lattice vectors.) Rotations $\overleftrightarrow{\phi}_2$ and $\overleftrightarrow{\phi}_3$ generate the star [1,2,3] of $\vec{k}_1$:

$$\vec{k}_2 \equiv \overleftrightarrow{\phi}_2 \cdot \vec{k}_1 = (0, 0, \tfrac{1}{2}), \tag{120}$$

$$\vec{k}_3 \equiv \overleftrightarrow{\phi}_3 \cdot \vec{k}_1 = (\tfrac{1}{2}, 0, 0). \tag{121}$$

Using Eq. (66) we have immediately

$$\chi(\vec{n} \circ p) = (-1)^{n_1} \chi_1(p) + (-1)^{n_1} \chi_2(p) + (-1)^{n_1} \chi_3(p), \tag{122}$$

where $\vec{n} = (n_1, n_2, n_3)$ and p is a coset representative from Table I. Therefore, we need only to determine partial characters χ_σ of the coset representatives p.

From Ref. 17 we determine partial characters associated with a particular representation $X(3)$. These partial characters are listed in our Table II.

We note that the number of vectors in the star of $\vec{k}_1$ is

$$s = 3 \tag{123}$$

and the representation $X(3)$ is six-dimensional:

$$c = \sum_{\sigma=1}^{3} \chi_\sigma(p_1) = 6. \tag{124}$$

TABLE I. Rotation-reflections $\overleftrightarrow{\phi}_i$ of O_h and the point group coset representatives p_i of O_h^3. We denote the identity by $\vec{\varepsilon}$. By $c_m(\vec{n})$ we denote counterclockwise rotation for $2\pi/m$ around $\vec{n}$ (as seen from the top of $\vec{n}$). $\sigma(\vec{n})$ is a reflection in a plane perpendicular to $\vec{n}$. $s_m(\vec{n}) \equiv \sigma(\vec{n}) \circ c_m(\vec{n})$. The inversion we denote by i. Superscripts denote powers. $\hat{x}$, $\hat{y}$, $\hat{z}$ denote unit vectors in positive directions x, y, z, respectively. A bar over a number (a letter) denotes minus the number (the letter). All operations $\overleftrightarrow{\phi}_i$ leave the origin invariant. For convenience we show also the result of actions of $\overleftrightarrow{\phi}_i$, $\overleftrightarrow{\phi}_i^{-1}$, and p_i on a general vector (x, y, z). We also include notation h_i of Refs. 17 and 21 corresponding to $\overleftrightarrow{\phi}_i$. Finally, we explicitly show fractional translations $\vec{\tau}_i$ and the inverse p_i^{-1} of a coset representative p_i. The fractional translation $\vec{\tau}$ is equal to $(\frac{1}{2},\frac{1}{2},\frac{1}{2})$.

i	$\overleftrightarrow{\varphi}_i$	$\overleftrightarrow{\varphi}_i \circ (x,y,z)$	h	$\overleftrightarrow{\varphi}_i^{-1} \circ (x,y,z)$	$\vec{\tau}_i$	$p_i = \{\vec{\tau}_i \mid \overleftrightarrow{\varphi}_i\}$	$p_i \circ (x,y,z)$	p_i^{-1}
1	$\overleftrightarrow{\varepsilon}$	(x,y,z)	h_1	(x,y,z)	$\vec{0}$	$\{\vec{0} \mid \overleftrightarrow{\varepsilon}\}$	(x,y,z)	p_1
2	$c_3(1,1,1)$	(z,x,y)	h_9	(y,z,x)	$\vec{0}$	$\{\vec{0} \mid c_3(1,1,1)\}$	(z,x,y)	p_3
3	$c_3^2(1,1,1)$	(y,z,x)	h_5	(z,x,y)	$\vec{0}$	$\{\vec{0} \mid c_3^2(1,1,1)\}$	(y,z,x)	p_2
4	$c_3(1,\bar{1},\bar{1})$	$(\bar{z},\bar{x},y)$	h_{12}	$(\bar{y},z,\bar{x})$	$\vec{0}$	$\{\vec{0} \mid c_3(1,\bar{1},\bar{1})\}$	$(\bar{z},\bar{x},y)$	p_5
5	$c_3^2(1,\bar{1},\bar{1})$	$(\bar{y},z,\bar{x})$	h_7	$(\bar{z},\bar{x},y)$	$\vec{0}$	$\{\vec{0} \mid c_3^2(1,\bar{1},\bar{1})\}$	$(\bar{y},z,\bar{x})$	p_4
6	$c_3(1,1,\bar{1})$	$(y,\bar{z},\bar{x})$	h_6	$(\bar{z},x,\bar{y})$	$\vec{0}$	$\{\vec{0} \mid c_3(1,1,\bar{1})\}$	$(y,\bar{z},\bar{x})$	p_7
7	$c_3^2(1,1,\bar{1})$	$(\bar{z},x,\bar{y})$	h_{11}	$(y,\bar{z},\bar{x})$	$\vec{0}$	$\{\vec{0} \mid c_3^2(1,1,\bar{1})\}$	$(\bar{z},x,\bar{y})$	p_6
8	$c_3(1,\bar{1},1)$	$(\bar{y},\bar{z},x)$	h_8	$(z,\bar{x},\bar{y})$	$\vec{0}$	$\{\vec{0} \mid c_3(1,\bar{1},1)\}$	$(\bar{y},\bar{z},x)$	p_9
9	$c_3^2(1,\bar{1},1)$	$(z,\bar{x},\bar{y})$	h_{10}	$(\bar{y},\bar{z},x)$	$\vec{0}$	$\{\vec{0} \mid c_3^2(1,\bar{1},1)\}$	$(z,\bar{x},\bar{y})$	p_8
10	$c_2(\hat{x})$	$(x,\bar{y},\bar{z})$	h_2	$(x,\bar{y},\bar{z})$	$\vec{0}$	$\{\vec{0} \mid c_2(\hat{x})\}$	$(x,\bar{y},\bar{z})$	p_{10}
11	$c_2(\hat{y})$	$(\bar{x},y,\bar{z})$	h_3	$(\bar{x},y,\bar{z})$	$\vec{0}$	$\{\vec{0} \mid c_2(\hat{y})\}$	$(\bar{x},y,\bar{z})$	p_{11}
12	$c_2(\hat{z})$	$(\bar{x},\bar{y},z)$	h_4	$(\bar{x},\bar{y},z)$	$\vec{0}$	$\{\vec{0} \mid c_2(\hat{z})\}$	$(\bar{x},\bar{y},z)$	p_{12}
13	i	$(\bar{x},\bar{y},\bar{z})$	h_{25}	$(\bar{x},\bar{y},\bar{z})$	$\vec{0}$	$\{\vec{0} \mid i\}$	$(\bar{x},\bar{y},\bar{z})$	p_{13}
14	$s_6^5(1,1,1)$	$(\bar{z},\bar{x},\bar{y})$	h_{33}	$(\bar{y},\bar{z},\bar{x})$	$\vec{0}$	$\{\vec{0} \mid s_6^5(1,1,1)\}$	$(\bar{z},\bar{x},\bar{y})$	p_{15}
15	$s_6(1,1,1)$	$(\bar{y},\bar{z},\bar{x})$	h_{29}	$(\bar{z},\bar{x},\bar{y})$	$\vec{0}$	$\{\vec{0} \mid s_6(1,1,1)\}$	$(\bar{y},\bar{z},\bar{x})$	p_{14}
16	$s_6^5(1,\bar{1},\bar{1})$	$(z,x,\bar{y})$	h_{36}	$(y,\bar{z},x)$	$\vec{0}$	$\{\vec{0} \mid s_6^5(1,\bar{1},\bar{1})\}$	$(z,x,\bar{y})$	p_{17}
17	$s_6(1,\bar{1},\bar{1})$	$(y,\bar{z},x)$	h_{31}	$(z,x,\bar{y})$	$\vec{0}$	$\{\vec{0} \mid s_6(1,\bar{1},\bar{1})\}$	$(y,\bar{z},x)$	p_{16}
18	$s_6^5(1,1,\bar{1})$	$(\bar{y},z,x)$	h_{30}	$(z,\bar{x},y)$	$\vec{0}$	$\{\vec{0} \mid s_6^5(1,1,\bar{1})\}$	$(\bar{y},z,x)$	p_{19}
19	$s_6(1,1,\bar{1})$	$(z,\bar{x},y)$	h_{35}	$(\bar{y},z,x)$	$\vec{0}$	$\{\vec{0} \mid s_6(1,1,\bar{1})\}$	$(z,\bar{x},y)$	p_{18}
20	$s_6^5(1,\bar{1},1)$	$(y,z,\bar{x})$	h_{32}	$(\bar{z},x,y)$	$\vec{0}$	$\{\vec{0} \mid s_6^5(1,\bar{1},1)\}$	$(y,z,\bar{x})$	p_{21}
21	$s_6(1,\bar{1},1)$	$(\bar{z},x,y)$	h_{34}	$(y,z,\bar{x})$	$\vec{0}$	$\{\vec{0} \mid s_6(1,\bar{1},1)\}$	$(\bar{z},x,y)$	p_{20}
22	$\sigma(\hat{x})$	$(\bar{x},y,z)$	h_{26}	$(\bar{x},y,z)$	$\vec{0}$	$\{\vec{0} \mid \sigma(\hat{x})\}$	$(\bar{x},y,z)$	p_{22}
23	$\sigma(\hat{y})$	$(x,\bar{y},z)$	h_{27}	$(x,\bar{y},z)$	$\vec{0}$	$\{\vec{0} \mid \sigma(\hat{y})\}$	$(x,\bar{y},z)$	p_{23}
24	$\sigma(\hat{z})$	$(x,y,\bar{z})$	h_{28}	$(x,y,\bar{z})$	$\vec{0}$	$\{\vec{0} \mid \sigma(\hat{z})\}$	$(x,y,\bar{z})$	p_{24}
25	$c_2(1,\bar{1},0)$	$(\bar{y},\bar{x},\bar{z})$	h_{13}	$(\bar{y},\bar{x},\bar{z})$	$\vec{\tau}$	$\{\vec{\tau} \mid c_2(1,\bar{1},0)\}$	$(\frac{1}{2}-y,\frac{1}{2}-x,\frac{1}{2}-z)$	p_{25}
26	$c_2(0,1,\bar{1})$	$(\bar{x},\bar{z},\bar{y})$	h_{17}	$(\bar{x},\bar{z},\bar{y})$	$\vec{\tau}$	$\{\vec{\tau} \mid c_2(0,1,\bar{1})\}$	$(\frac{1}{2}-x,\frac{1}{2}-z,\frac{1}{2}-y)$	p_{26}
27	$c_2(1,0,\bar{1})$	$(\bar{z},\bar{y},\bar{x})$	h_{21}	$(\bar{z},\bar{y},\bar{x})$	$\vec{\tau}$	$\{\vec{\tau} \mid c_2(1,0,\bar{1})\}$	$(\frac{1}{2}-z,\frac{1}{2}-y,\frac{1}{2}-x)$	p_{27}
28	$c_4^3(\hat{x})$	$(x,z,\bar{y})$	h_{20}	$(x,\bar{z},y)$	$\vec{\tau}$	$\{\vec{\tau} \mid c_4^3(\hat{x})\}$	$(\frac{1}{2}+x,\frac{1}{2}+z,\frac{1}{2}-y)$	$(\bar{1},0,\bar{1}) \circ p_{33}$
29	$c_4^3(\hat{y})$	$(\bar{z},y,x)$	h_{22}	$(z,y,\bar{x})$	$\vec{\tau}$	$\{\vec{\tau} \mid c_4^3(\hat{y})\}$	$(\frac{1}{2}-z,\frac{1}{2}+y,\frac{1}{2}+x)$	$(\bar{1},\bar{1},0) \circ p_{32}$
30	$c_2(1,0,1)$	$(z,\bar{y},x)$	h_{23}	$(z,\bar{y},x)$	$\vec{\tau}$	$\{\vec{\tau} \mid c_2(1,0,1)\}$	$(\frac{1}{2}+z,\frac{1}{2}-y,\frac{1}{2}+x)$	$(\bar{1},0,\bar{1}) \circ p_{30}$
31	$c_2(0,1,1)$	$(\bar{x},z,y)$	h_{18}	$(\bar{x},z,y)$	$\vec{\tau}$	$\{\vec{\tau} \mid c_2(0,1,1)\}$	$(\frac{1}{2}-x,\frac{1}{2}+z,\frac{1}{2}+y)$	$(0,\bar{1},\bar{1}) \circ p_{31}$
32	$c_4(\hat{y})$	$(z,y,\bar{x})$	h_{24}	$(\bar{z},y,x)$	$\vec{\tau}$	$\{\vec{\tau} \mid c_4(\hat{y})\}$	$(\frac{1}{2}+z,\frac{1}{2}+y,\frac{1}{2}-x)$	$(0,\bar{1},\bar{1}) \circ p_{29}$
33	$c_4(\hat{x})$	$(x,\bar{z},y)$	h_{19}	$(x,z,\bar{y})$	$\vec{\tau}$	$\{\vec{\tau} \mid c_4(\hat{x})\}$	$(\frac{1}{2}+x,\frac{1}{2}-z,\frac{1}{2}+y)$	$(\bar{1},\bar{1},0) \circ p_{28}$
34	$c_4^3(\hat{z})$	$(y,\bar{x},z)$	h_{15}	$(\bar{y},x,z)$	$\vec{\tau}$	$\{\vec{\tau} \mid c_4^3(\hat{z})\}$	$(\frac{1}{2}+y,\frac{1}{2}-x,\frac{1}{2}+z)$	$(0,\bar{1},\bar{1}) \circ p_{35}$
35	$c_4(\hat{z})$	$(\bar{y},x,z)$	h_{14}	$(y,\bar{x},z)$	$\vec{\tau}$	$\{\vec{\tau} \mid c_4(\hat{z})\}$	$(\frac{1}{2}-y,\frac{1}{2}+x,\frac{1}{2}+z)$	$(\bar{1},0,\bar{1}) \circ p_{34}$
36	$c_2(1,1,0)$	$(y,x,\bar{z})$	h_{16}	$(y,x,\bar{z})$	$\vec{\tau}$	$\{\vec{\tau} \mid c_2(1,1,0)\}$	$(\frac{1}{2}+y,\frac{1}{2}+x,\frac{1}{2}-z)$	$(\bar{1},\bar{1},0) \circ p_{36}$
37	$\sigma(1,\bar{1},0)$	(y,x,z)	h_{37}	(y,x,z)	$\vec{\tau}$	$\{\vec{\tau} \mid \sigma(1,\bar{1},0)\}$	$(\frac{1}{2}+y,\frac{1}{2}+x,\frac{1}{2}+z)$	$(\bar{1},\bar{1},\bar{1}) \circ p_{37}$
38	$\sigma(0,1,\bar{1})$	(x,z,y)	h_{41}	(x,z,y)	$\vec{\tau}$	$\{\vec{\tau} \mid \sigma(0,1,\bar{1})\}$	$(\frac{1}{2}+x,\frac{1}{2}+z,\frac{1}{2}+y)$	$(\bar{1},\bar{1},\bar{1}) \circ p_{38}$
39	$\sigma(1,0,\bar{1})$	(z,y,x)	h_{45}	(z,y,x)	$\vec{\tau}$	$\{\vec{\tau} \mid \sigma(1,0,\bar{1})\}$	$(\frac{1}{2}+z,\frac{1}{2}+y,\frac{1}{2}+x)$	$(\bar{1},\bar{1},\bar{1}) \circ p_{39}$
40	$s_4(\hat{x})$	$(\bar{x},\bar{z},y)$	h_{44}	$(\bar{x},z,\bar{y})$	$\vec{\tau}$	$\{\vec{\tau} \mid s_4(\hat{x})\}$	$(\frac{1}{2}-x,\frac{1}{2}-z,\frac{1}{2}+y)$	$(0,\bar{1},0) \circ p_{45}$
41	$s_4(\hat{y})$	$(z,\bar{y},\bar{x})$	h_{46}	$(\bar{z},y,x)$	$\vec{\tau}$	$\{\vec{\tau} \mid s_4(\hat{y})\}$	$(\frac{1}{2}+z,\frac{1}{2}-y,\frac{1}{2}-x)$	$(0,0,\bar{1}) \circ p_{44}$
42	$\sigma(1,0,1)$	$(\bar{z},y,\bar{x})$	h_{47}	$(\bar{z},y,\bar{x})$	$\vec{\tau}$	$\{\vec{\tau} \mid \sigma(1,0,1)\}$	$(\frac{1}{2}-z,\frac{1}{2}+y,\frac{1}{2}-x)$	$(0,\bar{1},0) \circ p_{42}$
43	$\sigma(0,1,1)$	$(x,\bar{z},\bar{y})$	h_{42}	$(x,\bar{z},\bar{y})$	$\vec{\tau}$	$\{\vec{\tau} \mid \sigma(0,1,1)\}$	$(\frac{1}{2}+x,\frac{1}{2}-z,\frac{1}{2}-y)$	$(\bar{1},0,0) \circ p_{43}$
44	$s_4^3(\hat{y})$	$(\bar{z},\bar{y},x)$	h_{48}	$(z,\bar{y},\bar{x})$	$\vec{\tau}$	$\{\vec{\tau} \mid s_4^3(\hat{y})\}$	$(\frac{1}{2}-z,\frac{1}{2}-y,\frac{1}{2}+x)$	$(\bar{1},0,0) \circ p_{41}$
45	$s_4^3(\hat{x})$	$(\bar{x},z,\bar{y})$	h_{43}	$(\bar{x},\bar{z},y)$	$\vec{\tau}$	$\{\vec{\tau} \mid s_4^3(\hat{x})\}$	$(\frac{1}{2}-x,\frac{1}{2}+z,\frac{1}{2}-y)$	$(0,0,\bar{1}) \circ p_{40}$
46	$s_4(\hat{z})$	$(\bar{y},x,\bar{z})$	h_{39}	$(y,\bar{x},\bar{z})$	$\vec{\tau}$	$\{\vec{\tau} \mid s_4(\hat{z})\}$	$(\frac{1}{2}-y,\frac{1}{2}+x,\frac{1}{2}-z)$	$(\bar{1},0,0) \circ p_{47}$
47	$s_4^3(\hat{z})$	$(y,\bar{x},\bar{z})$	h_{38}	$(\bar{y},x,\bar{z})$	$\vec{\tau}$	$\{\vec{\tau} \mid s_4^3(\hat{z})\}$	$(\frac{1}{2}+y,\frac{1}{2}-x,\frac{1}{2}-z)$	$(0,\bar{1},0) \circ p_{46}$
48	$\sigma(1,1,0)$	$(\bar{y},\bar{x},z)$	h_{40}	$(\bar{y},\bar{x},z)$	$\vec{\tau}$	$\{\vec{\tau} \mid \sigma(1,1,0)\}$	$(\frac{1}{2}-y,\frac{1}{2}-x,\frac{1}{2}+z)$	$(0,0,\bar{1}) \circ p_{48}$

TABLE II. Partial characters[17] of the coset representatives of Table I for the $X(3)$ irreducible representation of O_h^3.

p_i	p_1	p_2 to p_9	p_{10}	p_{11}	p_{12}	p_{13} to p_{48}
$\chi_1(p_i)$	2	0	-2	-2	2	0
$\chi_2(p_i)$	2	0	2	-2	-2	0
$\chi_3(p_i)$	2	0	-2	2	-2	0

Therefore, elements of Ker O_h^3 must have partial characters equal to 2. By inspection we find from Eq. (122) and Table II that

$$\text{Ker } T = \{(2n_1, 2n_2, 2n_3)\}, \quad n_i = 0, \pm 1, \pm 2, \ldots, \tag{125}$$

which corresponds to doubling of the unit cell in all three directions, and

$$\overline{\text{Ker}}\ O_h = \{p_1, (1,1,0)\circ p_{10}, (0,1,1,)\circ p_{11}, (1,0,1)\circ p_{12}\}, \tag{126}$$

which is isomorphic to the point group D_2. It is easily verified that $\overline{\text{Ker}}\ O_h = D_2$, as given in Eq. (126), is indeed a group with the composition law $\circ$, modulo Ker T. For example,

$$\begin{aligned}[(1,1,0)\circ p_{10}]^2 &= (1,1,0)\circ(1,\bar{1},0)\circ p_{10}^2 \\ &= (2,0,0)\circ p_1\end{aligned} \tag{127}$$

which is "equal" to p_1, since $(2,0,0) \in \text{Ker } T$.

Elements of **T**, cf. Eq. (37), can be obtained either from the coset decomposition of T with respect to Ker T or more readily as those vectors of T which belong to a single cell of Ker T. We take this cell defined by the origin and the three basis vectors

$$(2,0,0),\ (0,2,0),\ (0,0,2) \tag{128}$$

of Ker T. Therefore,

$$\mathbf{T} = \{\vec{0},(1,0,0),(0,1,0),(0,0,1),(1,1,0),(1,0,1),(0,1,1),(1,1,1)\} \tag{129}$$

and

$$|\mathbf{T}| = 8. \tag{130}$$

In order to find the group **P**, cf. Eq. (39), it is easier to find the quotient group O_h/D_2 in terms of $\overleftrightarrow{\phi}_i$ and then simply replace each $\overleftrightarrow{\phi}_i$ by the associated p_i. The coset decomposition of O_h with respect to D_2 thus gives

$$\mathbf{P} = \{p_1, p_2, p_3, p_{13}, p_{14}, p_{15}, p_{25}, p_{26}, p_{27}, p_{37}, p_{38}, p_{39}\} \tag{131}$$

and

$$|\mathbf{P}| = 12. \tag{132}$$

P is isomorphic to D_{3d}. In fact we have chosen coset representatives $\overleftrightarrow{\phi}_i$ in such a way that they indeed form a group with the same law of composition as in P ($\circ$, modulo T) and the requirement modulo $\overline{\text{Ker}}\ P$ is automatically satisfied. We could have chosen, for example, p_5 in place of p_3. Then we would have had $p_2 \circ p_2 = p_3 = p_5 \circ p_{12}$ and the condition modulo $\overline{\text{Ker}}\ P$ would have had to be applied in order to remove p_{12}.

TABLE III. Multiplication table for coset representatives p_i which are in **P**, Eq. (131). The result is expressed modulo Ker T, Eq. (125). In this table $\vec{t} = (1,1,1)$.

p_i \ p_j	p_1	p_2	p_3	p_{13}	p_{14}	p_{15}	p_{25}	p_{26}	p_{27}	p_{37}	p_{38}	p_{39}
p_1	p_1	p_2	p_3	p_{13}	p_{14}	p_{15}	p_{25}	p_{26}	p_{27}	p_{37}	p_{38}	p_{39}
p_2	p_2	p_3	p_1	p_{14}	p_{15}	p_{13}	p_{27}	p_{25}	p_{26}	p_{39}	p_{37}	p_{38}
p_3	p_3	p_1	p_2	p_{15}	p_{13}	p_{14}	p_{26}	p_{27}	p_{25}	p_{38}	p_{39}	p_{37}
p_{13}	p_{13}	p_{14}	p_{15}	p_1	p_2	p_3	$\vec{t}\circ p_{37}$	$\vec{t}\circ p_{38}$	$\vec{t}\circ p_{39}$	$\vec{t}\circ p_{25}$	$\vec{t}\circ p_{26}$	$\vec{t}\circ p_{27}$
p_{14}	p_{14}	p_{15}	p_{13}	p_2	p_3	p_1	$\vec{t}\circ p_{39}$	$\vec{t}\circ p_{37}$	$\vec{t}\circ p_{38}$	$\vec{t}\circ p_{27}$	$\vec{t}\circ p_{25}$	$\vec{t}\circ p_{26}$
p_{15}	p_{15}	p_{13}	p_{14}	p_3	p_1	p_2	$\vec{t}\circ p_{38}$	$\vec{t}\circ p_{39}$	$\vec{t}\circ p_{37}$	$\vec{t}\circ p_{26}$	$\vec{t}\circ p_{27}$	$\vec{t}\circ p_{25}$
p_{25}	p_{25}	p_{26}	p_{27}	p_{37}	p_{38}	p_{39}	p_1	p_2	p_3	p_{13}	p_{14}	p_{15}
p_{26}	p_{26}	p_{27}	p_{25}	p_{38}	p_{39}	p_{37}	p_3	p_1	p_2	p_{15}	p_{13}	p_{14}
p_{27}	p_{27}	p_{25}	p_{26}	p_{39}	p_{37}	p_{38}	p_2	p_3	p_1	p_{14}	p_{15}	p_{13}
p_{37}	p_{37}	p_{38}	p_{39}	p_{25}	p_{26}	p_{27}	$\vec{t}\circ p_{13}$	$\vec{t}\circ p_{14}$	$\vec{t}\circ p_{15}$	$\vec{t}\circ p_1$	$\vec{t}\circ p_2$	$\vec{t}\circ p_3$
p_{38}	p_{38}	p_{39}	p_{37}	p_{26}	p_{27}	p_{25}	$\vec{t}\circ p_{15}$	$\vec{t}\circ p_{13}$	$\vec{t}\circ p_{14}$	$\vec{t}\circ p_3$	$\vec{t}\circ p_1$	$\vec{t}\circ p_2$
p_{39}	p_{39}	p_{37}	p_{38}	p_{27}	p_{25}	p_{26}	$\vec{t}\circ p_{14}$	$\vec{t}\circ p_{15}$	$\vec{t}\circ p_{13}$	$\vec{t}\circ p_2$	$\vec{t}\circ p_3$	$\vec{t}\circ p_1$

We note at this point that the matrix group **R** is a group of $|\mathbf{T}|\ |\mathbf{P}| = 96$ different 6×6 matrices.[25] Therefore, a direct handling of this group may be quite elaborate.

Before proceeding to the next steps it is useful to establish a multiplication table (multiplication $\circ$) for elements of **P**. This is done in Table III. A look at the table immediately verifies that **P** is a group with multiplication $\circ$, modulo T.

Next we have to construct in an ascending fashion all the substars. We start with the subset $\{\vec{k}_1\}$. We generate the reciprocal lattice, cf. Eqs. (76) and (119),

$$\{(m_1, m_2, m_3) + l_1(0, \tfrac{1}{2}, 0)\} = \{(m_1, m_2/2, m_3)\}, \quad (133)$$

where m_1, m_2, m_3, and l_1 can be any integers. Taking the intersection between Eq. (133) and $\{\vec{k}_1, \vec{k}_2, \vec{k}_3\}$ we obtain [cf. Eq. (78)] the associated substar [1]. Clearly the direct lattice is

$$T[1] = \{(n_1, 2n_2, n_3)\}, \quad (134)$$

which corresponds to the doubling of the unit cell in one (y) direction (n_1, n_2, and n_3 are integers). The group **T**[1] is obtained by factoring out Ker T from $T[1]$ (or, equivalently, by selecting all lattice translations of $T[1]$ from a single cell of Ker T),

$$\mathbf{T}[1] = \{\vec{0}, (1,0,0), (0,0,1), (1,0,1)\}. \quad (135)$$

Similarly, cf. Eq. (87),

$$\mathbf{F}[1] = \{\vec{0}, (0,1,0)\}. \quad (136)$$

The group **P**[1] is obtained by selecting those elements of **P** which leave $\vec{k}_1$ invariant, modulo a reciprocal lattice vector. These elements are easily identified using Table I. We find, for example,

$$\overleftrightarrow{\phi}_{13}\cdot\vec{k}_1 = -(0, \tfrac{1}{2}, 0) = \vec{m} + (0, \tfrac{1}{2}, 0) = \vec{k}_1 \quad (137)$$

since $\vec{m} = (0,\bar{1},0)$ is a reciprocal lattice vector. The group **P**[1] is

$$\mathbf{P}[1] = \{p_1, p_{13}, p_{27}, p_{39}\}, \quad (138)$$

which is isomorphic to C_{2h}.

Subsets $\{2\}$ and $\{3\}$ need not be examined since they are P-equivalent to $\{1\}$. Hence, the next subset to consider is $\{\vec{k}_1, \vec{k}_2\}$.

Similarly to the previous case we find the associated substar to be [1,2] and

$$T[1,2] = \{(n_1, 2n_2, 2n_3)\}, \quad (139)$$

which is the doubling of the unit cell in two (y and z) directions (n_1, n_2, n_3 are integers). Furthermore,

$$\mathbf{T}[1,2] = \{\vec{0}, (1,0,0)\}, \quad (140)$$

$$\mathbf{F}[1,2] = \{\vec{0}, (0,1,0), (0,0,1), (0,1,1)\}, \quad (141)$$

and

$$\mathbf{P}[1,2] = \{p_1, p_{13}, p_{26}, p_{38}\}, \quad (142)$$

which is isomorphic to C_{2h}. It is important to note that rotation-reflections entering in **P**[1,2] only need to leave the *set* [1,2] invariant, modulo the reciprocal lattice, independent of ordering. For example,

$$\overleftrightarrow{\phi}_{26}\cdot\vec{k}_1 = -(0, 0, \tfrac{1}{2}) = (0,0,\bar{1}) + \vec{k}_2 = \vec{k}_2 \quad (143)$$

and

$$\overleftrightarrow{\phi}_{26}\cdot\vec{k}_2 = -(0, \tfrac{1}{2}, 0) = (0, \bar{1}, 0) + \vec{k}_1 = \vec{k}_1; \quad (144)$$

hence,

$$\overleftrightarrow{\phi}_{26}\cdot[1,2] = [2,1] = [1,2] \quad (145)$$

and p_{26} belongs to $P[1,2]$.

Subsets $\{1,3\}$ and $\{2,3\}$ are clearly P-equivalent to $\{1,2\}$ and need not be examined. Thus, the only remaining subset is the whole star [1,2,3] which is trivially a substar. We have

$$T[1,2,3] = \text{Ker } T, \quad (146)$$

$$\mathbf{T}[1,2,3] = \{\vec{0}\}, \quad (147)$$

$$\mathbf{F}[1,2,3] = \mathbf{T}, \quad (148)$$

and

$$\mathbf{P}[1,2,3] = \mathbf{P}. \quad (149)$$

TABLE IV. Substars $[\alpha,\beta,\ldots]$ and the associated groups $\mathbf{T}[\alpha,\beta,\ldots]$, $\mathbf{F}[\alpha,\beta,\ldots]$, $\mathbf{P}[\alpha,\beta,\ldots]$, $T[\alpha,\beta,\ldots]$, $P[\alpha,\beta,\ldots]$ for irreducible representation $X(3)$ of O_h^3. Groups $P[\alpha,\beta,\ldots]=\mathbf{P}[\alpha,\beta,\ldots]\circ\overline{\mathrm{Ker}}\ O_h$ are denoted as the corresponding point groups. The point groups are also indicated for $\mathbf{P}[\alpha,\beta,\ldots]$. n_1, n_2, and n_3 are arbitrary integers. $\overline{\mathrm{Ker}}\ O_h$, modulo $\mathbf{T}[\alpha,\beta,\ldots]$, is also given.

$[\alpha,\beta,\ldots]$	[1]	[1,2]	[1,2,3]
$T[\alpha,\beta,\ldots]$	$\{\vec{0},(1,0,0),(0,0,1),(1,0,1)\}$	$\{\vec{0},(1,0,0)\}$	$\{\vec{0}\}$
$F[\alpha,\beta,\ldots]$	$\{\vec{0},(0,1,0)\}$	$\{\vec{0},(0,1,0),(0,0,1),(0,1,1)\}$	$\{\vec{0},(1,0,0),(0,1,0),(0,0,1),(1,1,0),(1,0,1),(0,1,1),(1,1,1)\}=T$
$P[\alpha,\beta,\ldots]$	$\{p_1,p_{13},p_{27},p_{39}\}=C_{2h}$	$\{p_1,p_{13},p_{26},p_{38}\}=C_{2h}$	$\{p_1,p_2,p_3,p_{13},p_{14},p_{15},p_{25},p_{26},p_{27},p_{37},p_{38},p_{39}\}=D_{3d}=P$
$T[\alpha,\beta,\ldots]$	$\{(n_1,2n_2,n_3)\}$	$\{(n_1,2n_2,2n_3)\}$	$\mathrm{Ker}\ T=\{(2n_1,2n_2,2n_3)\}$
$P[\alpha,\beta,\ldots]$	D_{4h}	D_{4h}	O_h
$\overline{\mathrm{Ker}}\ O_h$	$\{p_1,(0,1,0)\circ p_{10},(0,1,0)\circ p_{11},p_{12}\}$	$\{p_1,(0,1,0)\circ p_{10},(0,1,1)\circ p_{11},(0,0,1)\circ p_{12}\}$	$\{p_1,(1,1,0)\circ p_{10},(0,1,1)\circ p_{11},(1,0,1)\circ p_{12}\}$

The results of the first two steps are collected in Table IV. The first substar to be examined is [1]. The point and translation groups of this substar are given in Table IV. In the same table we find the group of additional fractional translations.

Next, we have to select subgroups $\mathbf{P}'_L$ of $\mathbf{P}[1]$. The possible subgroups are easily identified since $\mathbf{P}[1]$ is isomorphic to C_{2h}. These subgroups are

$$C_{2h}=\{p_1,p_{13},p_{27},p_{39}\}, \tag{150}$$

$$C_s=\{p_1,p_{39}\}, \tag{151}$$

$$C_i=\{p_1,p_{13}\}, \tag{152}$$

$$C_2=\{p_1,p_{27}\}, \tag{153}$$

and

$$C_1=\{p_1\}. \tag{154}$$

We have already arranged these groups in a descending fashion (note that C_2, C_i, and C_s may be freely permuted without violating the descending ordering), and they will be examined in this order.

Although all groups, Eqs. (150)–(154), are groups with the multiplication $\circ$ modulo $\mathbf{T}$ they need not be groups with respect to $\circ$ modulo $\mathbf{T}[1]$. For the later to hold a proper choice of additional fractional translations from $\mathbf{F}[1]$ must be made.

We start with $\mathbf{P}'_L=C_{2h}$. Then from the characters, Table II, we see that in $\mathbf{P}[1]$ only p_1 has nonzero characters $\chi_\sigma(p_1)=2$. Therefore, regardless of the fractionals chosen, cf. Eqs. (71) and (85),

$$i(\mathbf{L})=\tilde{i}_1(\mathbf{P}_L)=2/|\mathbf{P}'_L|. \tag{155}$$

Since $i(\mathbf{L})$ must be a positive integer and $|C_{2h}|=4\Rightarrow i(\mathbf{L})=\frac{1}{2}$, we conclude that it is impossible to choose additional fractional translations compatible with C_{2h}. Now we can proceed with the next $\mathbf{P}'_L=C_s$. However, we want at this point to demonstrate explicitly the above-mentioned incompatibility.

From Table III we see in particular that (modulo $T[1]$)

$$p_{39}^2=(0,1,0)\circ p_1. \tag{156}$$

The group property requires that an additional fractional translation $\vec{f}_{39}$ from $\mathbf{F}[1]$ be added to p_{39} in order that (modulo $T[1]$)

$$(\vec{f}_{39}\circ p_{39})^2=p_1. \tag{157}$$

Equations (156) and (157) imply

$$\vec{f}_{39}+\vec{\phi}_{39}\vec{f}_{39}+(0,1,0)\in T[1], \tag{158}$$

but $\vec{f}_{39}\in\mathbf{F}[1]$ implies $f_{39}=\vec{0}$ or $(0,1,0)$ which gives the left-hand side of Eq. (158) as $(0,1,0)$ or $(0,3,0)$ which are not in $T[1]$. This shows explicitly, that in accordance with a simple argument based on integer valuedness of $i(\mathbf{L})$, $\vec{f}_{39}\in\mathbf{F}[1]$ cannot be chosen in such a way that Eq. (157) is satisfied.

Now we proceed to the next $\mathbf{P}'_L$, $\mathbf{P}'_L=C_s$, Eq. (151). However, the previous analysis demonstrated that a consistent $\vec{f}_{39}$ cannot be found. This eliminates all groups containing p_{39}, eliminating in particular C_s.

Next, we analyze the group C_i, Eq. (152). Since p_1 cannot have a fractional translation it is sufficient to determine compatible additional fractional translations $\vec{f}_{13}\in\mathbf{F}[1]$ associated with p_{13}. The only group requirement is

$$(\vec{f}_{13}\circ p_{13})^2\in T[1], \tag{159}$$

which is satisfied for both choices $\vec{f}_{13}\in\mathbf{F}[1]$. Therefore, there are two possible isotropy groups whose point groups are

$$\mathbf{P}_L^{(1)}=\{p_1,p_{13}\} \tag{160}$$

and

$$\mathbf{P}_L^{(2)}=\{p_1,(0,1,0)\circ p_{13}\}, \tag{161}$$

both with the same subduction frequency $i=1$ [cf. Eq. (155)]. The full space-group point groups which determine the crystallographic class are

$$P_L^{(1)}=\{p_1,p_{13}\}\circ\{p_1,(0,1,0)\circ p_{10},(0,1,0)\circ p_{11},p_{12}\} \tag{162}$$

and

$$P_L^{(2)} = \{ p_1, (0,1,0)\circ p_{13}\}\circ\{ p_1,(0,1,0)\circ p_{10}(0,1,0)\circ p_{11},p_{12}\}, \tag{163}$$

where $\overline{\text{Ker}}\ O_h$, given in Eq. (126), is written modulo $T[1]$. The crystallographic class is immediately determined as $C_i \circ D_2 = D_{2h}$. Therefore, the two point groups Eqs. (162) and (163), and also the two associated space groups, we denote $D_{2h}^{(1)}$ and $D_{2h}^{(2)}$, respectively.

At this point it is useful to check if $D_{2h}^{(1)}$ and $D_{2h}^{(2)}$ are equivalent in O_h^3. For this it suffices to look for the equivalence of $\mathbf{P}_L^{(1)}$ and $\mathbf{P}_L^{(2)}$, since Ker O_h^3 is an invariant subgroup of O_h^3. In other words, we are looking for

$$g\in\mathbf{F}[1]\circ\mathbf{P}[1] \tag{164}$$

such that

$$\mathbf{P}_L^{(1)} = g^{-1}\circ\mathbf{P}_L^{(2)}\circ g, \tag{165}$$

modulo Ker O_h^3. In Eq. (164) a requirement that g should leave $\mathbf{T}[1]$ invariant was utilized together with a fact that elements of $\mathbf{P}_L$ necessarily leave $\mathbf{T}[1]$ invariant (for a general substar the requirement Eq. (164) is replaced by

$$g\in\mathbf{F}[\alpha,\beta,\ldots]\circ\mathbf{P}[\alpha,\beta,\ldots]). \tag{166}$$

Additionally, if it is found that g must be an improper operation, then the two space groups $\mathbf{P}_L^{(1)}$ and $\mathbf{P}_L^{(2)}$ will have a different handedness.

The condition of Eq. (165) when applied to Eqs. (160) and (161) reduces to

$$p_{13} = g^{-1}\circ(0,1,0)\circ p_{13}\circ g, \tag{167}$$

and a possible g is p_{27} which is a proper operation. Indeed, using Tables I and III, we find

$$p_{27}^{-1}\circ(0,1,0)\circ p_{13}\circ p_{27} = p_{27}\circ(0,2,0)\circ p_{39} = p_{13}. \tag{168}$$

Therefore, the isotropy space group $D_{2h}^{(2)}$ is equivalent to $D_{2h}^{(1)}$ and need not be separately analyzed.

The next possible choice for $\mathbf{P}_L'$ is $\mathbf{P}_L' = C_2$, Eq. (153). The associated crystallographic class is D_4 and the subduction frequency is 1. Similarly, like in the previous case, we find that both additional fractional translations are allowed leading to

$$D_4^{(1)} = \{ p_1, p_{27}\}\circ\{ p_1, (0,1,0)\circ p_{10}, (0,1,0)\circ p_{11}, p_{12}\} \tag{169}$$

and

$$D_4^{(2)} = \{ p_1(0,1,0)\circ p_{27}\}\circ\{ p_1, (0,1,0)\circ p_{10}, (0,1,0)\circ p_{11}, p_{12}\}. \tag{170}$$

This example offers an illustration of a point made earlier in the text: it is easily checked that $D_4^{(2)}$ and $D_4^{(1)}$ are equivalent in O_h^3,

$$D_4^{(1)} = p_{13}^{-1}\circ D_4^{(2)}\circ p_{13}, \tag{171}$$

but since the conjugation element (e.g., p_{13}) must be an improper rotation, the two space groups will be different and both must be considered.

The last possible choice at this substar is $\mathbf{P}_L' = C_1$, Eq. (154). The associated crystallographic class is D_2 and the subduction frequency is 2. There is only one possibility:

TABLE V. Substar [1]. Point groups $\mathbf{P}_L'$ ($\mathbf{P}_L' \subseteq \mathbf{P}[1]$), corresponding crystallographic class ($\mathbf{P}_L' \circ \overline{\text{Ker}}\ O_h$),[a] point groups of isotropy groups P_L ($P_L = \mathbf{P}_L \circ \overline{\text{Ker}}\ O_h$), associated subduction frequencies i, and numbers of domains δ (halves indicate conjugated groups but with different handedness). The translation group, Bravais lattice, is $T[1] = \{(n_1, 2n_2, n_3)\}$ where n_1, n_2, n_3 are integers.

$\mathbf{P}_L'$	class	P_L	i	δ
$C_{2h} = \{p_1, p_{13}, p_{27}, p_{39}\}$	D_{4h}	——	—	—
$C_s = \{p_1, p_{39}\}$	D_{2d}	——	—	—
$C_i = \{p_1, p_{13}\}$	D_{2h}	$D_{2h}^{(1)} = \{p_1, p_{13}\} \circ \overline{\text{Ker}}\ O_h$	1	12
$C_2 = \{p_1, p_{27}\}$	D_4	$D_4^{(1)} = \{p_1, p_{27}\} \circ \overline{\text{Ker}}\ O_h$	1	12/2
		$D_4^{(2)} = \{p_1(0,1,0) \circ p_{27}\} \circ \overline{\text{Ker}}\ O_h$	1	12/2
$C_1 = \{p_1\}$	D_2	$D_2^{(1)} = \overline{\text{Ker}}\ O_h$	2	24

[a] $\overline{\text{Ker}}\ O_h = \{ p_1, (0,1,0) \circ p_{10}, (0,1,0) \circ p_{11}, p_{12}\}$

$$D_2^{(1)} = \overline{\text{Ker}}\ O_h = \{ p_1, (0,1,0)\circ p_{10}, (0,1,0)\circ p_{11}, p_{12}\}. \tag{172}$$

The same conclusion could have been reached immediately using Lemma (iv) of the introductory section.

In the above cases the checks of Steps 6 and 7 are trivially passed due to Eq. (155).

Isotropy groups so far determined are $D_{2h}^{(1)}$, $D_4^{(1)}$, $D_4^{(2)}$, $D_2^{(1)}$ whose point groups are given in Eqs. (162), (169), (170), and (172), respectively. The Bravais lattice is $T[1]$, given in Eq. (134). These results are summarized in Table V.

Since we exhausted substar [1] we proceed (Step 2, etc.) to analyze substar [1,2]. The point and translation groups of this substar are given in Table IV. In the same table we find the group of additional fractional translations.

Possible subgroups $\mathbf{P}_L'$ and associated crystallographic classes at substar [1,2] are shown in 1st and 2nd columns of Table VI.

We start with $\mathbf{P}_L' = C_{2h}$, the class D_{4h}, Table VI. An analysis similar to the one previously described gives as the only possible fractionals

$$\vec{f}_{38} = (0,1,0) \text{ or } (0,0,1) \tag{173}$$

and

$$\vec{f}_{26} = (0,0,0) \text{ or } (0,1,1). \tag{174}$$

Equations (173) and (174) provide four possible point groups

$$D_{4h}^{(1)} = \{ p_1, (0,1,0)\circ p_{13}, p_{26}, (0,0,1)\circ p_{38}\}\circ \overline{\text{Ker}}\ O_h, \tag{175}$$

$$D_{4h}^{(2)} = \{ p_1, (0,0,1)\circ p_{13}, p_{26}, (0,1,0)\circ p_{38}\}\circ \overline{\text{Ker}}\ O_h, \tag{176}$$

$$D_{4h}^{(3)} = \{ p_1,(0,1,0)\circ p_{13},(0,1,1)\circ p_{26},(0,1,0)\circ p_{38}\}\circ \overline{\text{Ker}}\ O_h, \tag{177}$$

$$D_{4h}^{(4)} = \{ p_1,(0,0,1)\circ p_{13},(0,1,1)\circ p_{26},(0,0,1)\circ p_{38}\}\circ \overline{\text{Ker}}\ O_h. \tag{178}$$

Groups in Eqs. (175)–(178) can be checked for equivalence in $\mathbf{F}[1,2]\circ\mathbf{P}[1,2]$. For example, we look for $g\in\mathbf{F}[1,2]\circ\mathbf{P}[1,2]$ such that

TABLE VI. Substar [1,2]. Point groups $\mathbf{P}'_L$ ($\mathbf{P}'_L \subseteq \mathbf{P}[1,2]$), corresponding crystallographic class ($\mathbf{P}'_L \circ \overline{\mathrm{Ker}}\ O_h$),[a] point groups of isotropy groups P_L ($P_L = \mathbf{P}_L \circ \overline{\mathrm{Ker}}\ O_h$), associated subduction frequencies i, and numbers of domains δ. The translation group, Bravais lattice, is $T[1,2] = \{(n_1, 2n_2, 2n_3)\}$ where n_1, n_2, n_3 are integers.

P'_L	class	P_L	i	δ
$C_{2h} = \{p_1, p_{13}, p_{26}, p_{38}\}$	D_{4h}	$D_{4h}^{(1)} = \{p_1, (0,1,0)\circ p_{13}, p_{26}, (0,0,1)\circ p_{38}\} \circ \overline{\mathrm{Ker}}\ O_h$	1	12
		$D_{4h}^{(2)} = \{p_1, (0,0,1)\circ p_{13}, p_{26}, (0,1,0)\circ p_{38}\} \circ \overline{\mathrm{Ker}}\ O_h$	1	12
$C_s = \{p_1, p_{38}\}$	D_{2d}	$D_{2d}^{(1)} = \{p_1, (0,1,0)\circ p_{38}\} \circ \overline{\mathrm{Ker}}\ O_h$	2	24
$C_i = \{p_1, p_{13}\}$	D_{2h}	$D_{2h}^{(1)} = \{p_1, p_{13}\} \circ \overline{\mathrm{Ker}}\ O_h$	2	24
		$D_{2h}^{(2)} = \{p_1, (0,1,0)\circ p_{13}\} \circ \overline{\mathrm{Ker}}\ O_h$	2	24
		$D_{2h}^{(3)} = \{p_1, (0,0,1)\circ p_{13}\} \circ \overline{\mathrm{Ker}}\ O_h$	2	24
$C_2 = \{p_1, p_{26}\}$	D_4	$D_4^{(1)} = \{p_1, p_{26}\} \circ \overline{\mathrm{Ker}}\ O_h$	2	24
$C_1 = \{p_1\}$	D_2	$D_2^{(1)} = \overline{\mathrm{Ker}}\ O_h$	4	48

[a] $\overline{\mathrm{Ker}}\ O_h = \{p_1, (0,1,0)\circ p_{10}, (0,1,1)\circ p_{11}, (0,0,1)\circ p_{12}\}$

$$D_{4h}^{(1)} = g^{-1} \circ D_{4h}^{(3)} \circ g. \tag{179}$$

We find here $g = (0,0,1) \in \mathbf{F}[1,2]$ which illustrates that g need not be strictly from $\mathbf{P}[1,2]$, as might have been inferred from another example, Eq. (168).

Similarly, we find $D_{4h}^{(4)}$ to be equivalent to $D_{4h}^{(2)}$.

The "subduction frequencies" $\tilde{i}_\sigma(D_{4h}^{(1)})$ are found, using Table II, to be

$$\tilde{i}_1(D_{4h}^{(1)}) = \tilde{i}_2(D_{4h}^{(1)}) = \tfrac{1}{2}, \tag{180}$$

implying

$$i(D_{4h}^{(1)}) = 1. \tag{181}$$

Since $D_{4h}^{(1)}$ has no supergroups at substar [1,2], the condition Step 6 is trivially satisfied. On the other hand the conditions Step 7 and Eq. (106),

$$\tilde{i}_2(D_{4h}^{(1)}) = 0, \tag{182}$$

are, according to Eq. (180), not satisfied. Consequently, $D_{4h}^{(1)}$ passes both Steps 6 and 7, implying that $D_{4h}^{(1)}$ is an isotropy group.

The examination of other subgroups at the substar [1,2] offers no new insights into the method; hence we only summarize the result in Table VI.

We proceed to the last substar [1,2,3], the complete star. The point and translation groups of this substar are given in Table IV. In the same table we find the group of additional fractional translations.

Like at the previous substars, we first determine all possible subgroups $\mathbf{P}'_L$ of $\mathbf{P}[1,2,3]$. At the same time we determine associated crystallographic classes ($\mathbf{P}'_L \circ \overline{\mathrm{Ker}}\ O_h$). These are summarized in columns 1 and 2 of Table VII.

From the characters, Table II, we see again that the only contribution to the subduction frequency comes from p_1. Consequently,

$$i(\mathbf{L}) = 6/|\mathbf{P}'_L|, \tag{183}$$

each partial character giving a contribution $2/|\mathbf{P}'_L|$. Since $i(\mathbf{L})$ must be a positive integer, groups $\mathbf{P}'_L = D_{3d}$ and $\mathbf{P}'_L = C_{2h}$ are automatically eliminated ($|\mathbf{P}'_L| = 12$ and 4, respectively).

The above groups are also eliminated by virtue of the fact that no additional fractional translation $\vec{f}_{37}$ can be chosen so that the square of $\vec{f}_{37} \circ p_{37}$ is equal to p_1, modulo Ker T. A similar conclusion can be made for p_{38} and p_{39}. Consequently, all groups $\mathbf{P}'_L$ containing these elements are eliminated. These are, in addition to D_{3d} and C_{2h}, C_{3v} and C_s.

Using the multiplication table, Table III, and considering squares of elements p_{25}, p_{26}, and p_{27}, we can find possible additional fractional translations:

$$\vec{f}_{25} = \vec{0}, (0,0,1), (1,1,0), (1,1,1), \tag{184}$$

$$\vec{f}_{26} = \vec{0}, (1,0,0), (0,1,1), (1,1,1), \tag{185}$$

and

$$\vec{f}_{27} = \vec{0}, (0,1,0), (1,0,1), (1,1,1). \tag{186}$$

Similarly, the cube of $\vec{f}_2 \circ p_2$ must give p_1, modulo Ker T. This implies possible $\vec{f}_2$,

$$\vec{f}_2 = \vec{0}, (0,1,1), (1,0,1), (1,1,0). \tag{187}$$

Analogous considerations of p_{13}, p_{14}, and p_{15} bring no restrictions on their additional fractional translations.

Further restrictions arise from multiplications of different group elements. For example, in the case $\mathbf{P}'_L = S_6 = C_{3i}$ the following relation must be satisfied:

$$(\vec{f}_{13} \circ p_{13}) \circ (\vec{f}_2 \circ p_2) = (\vec{f}_2 \circ p_2) \circ (\vec{f}_{13} \circ p_{13}). \tag{188}$$

This restricts $\vec{f}_{13}$ to

TABLE VII. Substar [1,2,3]. Point groups $\mathbf{P}'_L$ ($\mathbf{P}'_L \subseteq \mathbf{P}[1,2,3]$), corresponding crystallographic class ($\mathbf{P}'_L \circ \overline{\mathrm{Ker}}\ O_h$),[a] point groups of the isotropy groups P_L ($P_L = \mathbf{P}_L \circ \overline{\mathrm{Ker}}\ O_h$), associated subduction frequencies i, and numbers of domains δ (halves indicate conjugated groups but with different handedness). The translation group, Bravais lattice, is $T[1,2,3] = \{(2n_1, 2n_2, 2n_3)\}$ where n_1, n_2, n_3 are integers.

P'_L	class	P_L	i	δ
$D_{3d} = \{p_1, p_2, p_3, p_{13}, p_{14}, p_{15}, p_{25}, p_{26}, p_{27}, p_{37}, p_{38}, p_{39}\} = P$	O_h	—	—	—
$D_3 = \{p_1, p_2, p_3, p_{25}, p_{26}, p_{27}\}$	O	$O^{(1)} = \{p_1, p_2, p_3, p_{25}, p_{26}, p_{27}\} \circ \overline{\mathrm{Ker}}\ O_h$	1	16/2
		$O^{(2)} = \{p_1, p_2, p_3, (1,1,1) \circ p_{25}, (1,1,1) \circ p_{26}, (1,1,1) \circ p_{27}\} \circ \overline{\mathrm{Ker}}\ O_h$	1	16/2
$C_{3v} = \{p_1, p_2, p_3, p_{37}, p_{38}, p_{39}\}$	T_d	—	—	—
$S_6 = \{p_1, p_2, p_3, p_{13}, p_{14}, p_{15}\}$	T_h	$T_h^{(1)} = \{p_1, p_2, p_3, p_{13}, p_{14}, p_{15}\} \circ \overline{\mathrm{Ker}}\ O_h$	1	16
$C_{2h} = \{p_1, p_{13}, p_{25}, p_{37}\}$	D_{4h}	—	—	—
$C_3 = \{p_1, p_2, p_3\}$	T	$T^{(1)} = \{p_1, p_2, p_3\} \circ \overline{\mathrm{Ker}}\ O_h$	2	32
$C_s = \{p_1, p_{37}\}$	D_{2d}	—	—	—
$C_1 = \{p_1, p_{13}\}$	D_{2h}	$D_{2h}^{(1)} = \{p_1, p_{13}\} \circ \overline{\mathrm{Ker}}\ O_h$	3	48
		$D_{2h}^{(2)} = \{p_1, (0,0,1) \circ p_{13}\} \circ \overline{\mathrm{Ker}}\ O_h$	3	48
$C_2 = \{p_1, p_{25}\}$	D_4	$D_4^{(1)} = \{p_1, p_{25}\} \circ \overline{\mathrm{Ker}}\ O_h$	3	48/2
		$D_4^{(2)} = \{p_1, (0,0,1) \circ p_{25}\} \circ \overline{\mathrm{Ker}}\ O_h$	3	48/2
$C_1 = \{p_1\}$	D_2	$D_2^{(1)} = \overline{\mathrm{Ker}}\ O_h$	6	96

[a] $\overline{\mathrm{Ker}}\ O_h = \{p_1, (1,1,0) \circ p_{10}, (0,1,1) \circ p_{11}, (1,0,1) \circ p_{12}\}$

$$\vec{f}_{13} = \vec{0}, (1,1,1). \tag{189}$$

A similar treatment gives all possible choices of additional fractional translations and of associated point groups $\mathbf{P}_L$. Since the actual manipulation does not bring any new insights into the technique we omit it and we only give the results in Table VII.

However, we note that all subgroups so obtained are isotropy groups; that is, they pass Steps 6 and 7. This can be seen from Eq. (183). For example, at the same substar, $\mathbf{P}_{L_1} \subset \mathbf{P}_{L_2}$ implies $|\mathbf{P}_{L_1}| < |\mathbf{P}_{L_2}|$ and $i(\mathbf{L}_1) > i(\mathbf{L}_2)$, and Step 6 is passed. Similarly,

$$\sum_{\sigma \in [1,2,3] \setminus [\alpha', \ldots]} \tilde{i}_\sigma(\mathbf{P}_L) = \frac{2}{|\mathbf{P}_L|} \text{ or } \frac{4}{|\mathbf{P}_L|} \neq 0, \tag{190}$$

and Step 7 is passed.

Also we emphasize that again groups equivalent in O_h^3 but different space groups occur due to equivalence by an improper operation (only). These are $O^{(1)} \sim O^{(2)}$ and $D_4^{(1)} \sim D_4^{(2)}$ which are in both cases equivalent under conjugation by the inversion p_{13}. Note also that in these cases the number of the domains δ is actually a half of its usual value.[19]

The last step which remains to be performed is an actual identification of conventional space-group labels for all isotropy space groups associated with point group coset representatives given in Tables V–VII.

Given $\mathbf{T}_L$ and $\mathbf{P}_L$ we have already given full translation T_L and point P_L groups in Tables IV and V–VII, respectively. Then the first step in the identification is to express all fractional translations in terms of the new Bravais lattice. After this is done, a translation of the origin and renaming of the axes may be required before the identification can be made. These steps are best illustrated in an example.

Let us take the group $\mathbf{L}$ given by $\mathbf{T}_L = \mathbf{T}[1]$ and $\mathbf{P}'_L = C_2$. In Table IV we find

$$T_L = T[1] = \{(n_1, 2n_2, n_3)\}, \quad n_i = 0, \pm 1, \pm 2, \ldots, \tag{191}$$

which determines the new Bravais lattice. In Table V we find

$$P_L = D_4^{(1)} = \{p_1, p_{27}\} \circ \{p_1, (0,1,0) \circ p_{10}, (0,1,0) \circ p_{11}, p_{12}\}, \tag{192}$$

where we have chosen $D_4^{(1)}$ as an example. Now, we have to reexpress *all* fractional translations in terms of the new Bravais lattice. Therefore,

$$p_{27} = \{(\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}) | \overleftrightarrow{\phi}_{27}\} \tag{193}$$

in the old lattice, becomes

$$P_{27} = \{(\tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{2}) | \overleftrightarrow{\phi}_{27}\} \tag{194}$$

in the new lattice. Similarly, $(0,1,0) \circ p_{10}$ becomes $(0,\tfrac{1}{2},0) \circ p_{10}$, etc.

Next, we choose to make an identification according to Ref. 16. There we find that for the crystallographic class D_4 the unique tetragonal axis is chosen to be the $\hat{z}$ axis. Therefore, we must take an equivalent group to $D_4^{(1)}$, namely, we rotate the coordinate frame:

$$D_4^{(1)} \sim p_3^{-1} \circ D_4^{(1)} \circ p_3 = \{p_1, p_{25}\} \circ \{p_1, p_{10}, (0,0,\tfrac{1}{2}) \circ (0,0,\tfrac{1}{2}) \circ p_{12}\} \tag{195}$$

and

$$p_{25} = \{(\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{4})|\vec{\phi}_{25}\}. \tag{196}$$

At this point it is important to note that the new translation group is now also rotated,

$$T_L = \{(n_1, n_2, 2n_3)\}, \quad n_i = 0, \pm 1, \pm 2, \ldots . \tag{197}$$

Finally, we make the shift of the origin in the old lattice to $(\frac{1}{2},0,\frac{1}{4})$ [in the new lattice, to $(\frac{1}{2},0,\frac{1}{4})$]. This results in

$$\begin{aligned} D_4^{(1)} = \{&(x, y, z), (x, \bar{y}, \bar{z} + \tfrac{1}{2}), (\bar{x}, y, \bar{z}), (\bar{x}, \bar{y}, z + \tfrac{1}{2}), \\ &(\bar{y}, \bar{x}, \bar{z} + \tfrac{3}{4}), (y, \bar{x}, z + \tfrac{1}{4}), (\bar{y}, x, z + \tfrac{3}{4}), \\ &(y, x, \bar{z} + \tfrac{1}{4})\}, \end{aligned} \tag{198}$$

where we expressed the elements of $D_4^{(1)}$ as the coordinates of a vector into which a general vector (x, y, z) is transformed. These are nothing else but the coordinates of equivalent positions of a general vector as given in Ref. 16 for the space group D_4^7. Therefore, we complete our identification

$$D_4^{(1)} \sim D_4^7. \tag{199}$$

Identifications of other space groups follow the same pattern. The result is summarized in Table VIII. With this we have completed a determination of all isotropy space groups of O_h^3 associated with its irreducible representation $X(3)$. The results here agree with those of Ref. 23.

We conclude that in a phase transition driven by an $X(3)$ order parameter the symmetry may change from O_h^3 only to one of its isotropy groups given in Table VIII. Further restrictions on simple continuous transitions come from the maximality conjecture[6,7] and/or the actual minimization which also require the utilization of the results presented in Table VIII. This additional analysis will be explained elsewhere.[5]

We also note that an experimentally observed low-symmetry group C_{4h}^4 (doubling in x and y directions)[26] in A15 superconductors is not among those listed in Table VIII implying that O_h^3 to C_{4h}^4 transitions are either not driven by an $X(3)$ order parameter, or an additional order parameter plays an important role in the transitions.

TABLE VIII. All isotropy space groups L of the space group O_h^3 for the irreducible representation[17] $X(3)$. Associated Bravais lattices T_L are given in terms of primitive translations of the Bravais lattice of O_h^3 (n_1, n_2, n_3 are integers). Vector $\vec{\tau}_0$, expressed in terms of primitive translations in O_h^3, denotes the shift of the origin necessary to transform elements of L, Table I, into a frame in which they coincide with the corresponding ones of Ref. 16. The last two columns give the subduction frequency $i(L)$ and the number of domains which occur in a corresponding transition (halves indicate equivalent space groups but with different handedness).

L	T_L	$\vec{\tau}_0$	$i(L)$	δ
O_h^3	$\{(n_1,n_2,n_3)\}$	—	0	1
D_{2h}^5	$\{(n_1,2n_2,n_3)\}$	—	1	12
D_4^7	$\{(n_1,n_2,2n_3)\}$	$(\frac{1}{2},0,\frac{1}{2})$	1	$12/2$
D_4^3	$\{(n_1,n_2,2n_3)\}$	$(\frac{1}{2},0,\frac{1}{2})$	1	$12/2$
D_{4h}^{13}	$\{(2n_1,2n_2,n_3)\}$	$(\frac{1}{2},0,\frac{1}{2})$	1	12
D_{4h}^{15}	$\{(2n_1,2n_2,n_3)\}$	$(\frac{1}{2},0,\frac{1}{4})$	1	12
O^6	$\{(2n_1,2n_2,2n_3)\}$	—	1	$16/2$
O^7	$\{(2n_1,2n_2,2n_3)\}$	—	1	$16/2$
T_h^6	$\{(2n_1,2n_2,2n_3)\}$	—	1	16
D_2^2	$\{(n_1,n_2,2n_3)\}$	—	2	24
D_{2d}^4	$\{(2n_1,2n_2,n_3)\}$	$(\frac{1}{2},0,\frac{1}{2})$	2	24
D_{2h}^{11}	$\{(n_1,2n_2,2n_3)\}$	—	2	24
D_{2h}^9	$\{(2n_1,2n_2,n_3)\}$	$(\frac{1}{2},0,0)$	2	24
D_{2h}^{13}	$\{(2n_1,2n_2,n_3)\}$	$(\frac{1}{2},0,0)$	2	24
D_4^6	$\{(2n_1,2n_2,n_3)\}$	$(\frac{1}{2},0,\frac{1}{4})$	2	24
T^4	$\{(2n_1,2n_2,2n_3)\}$	—	2	32
D_{2h}^{15}	$\{(2n_1,2n_2,2n_3)\}$	—	3	48
D_{2h}^{16}	$\{(2n_1,2n_2,2n_3)\}$	$(0,0,\frac{1}{2})$	3	48
D_4^8	$\{(2n_1,2n_2,2n_3)\}$	$(\frac{1}{2},0,\frac{3}{4})$	3	$48/2$
D_4^4	$\{(2n_1,2n_2,2n_3)\}$	$(\frac{1}{2},0,\frac{1}{4})$	3	$48/2$
D_2^3	$\{(2n_1,2n_2,n_3)\}$	$(\frac{1}{2},0,0)$	4	48
D_2^4	$\{(2n_1,2n_2,2n_3)\}$	—	6	96

V. DISCUSSION

In the preceding sections we have developed an algorithm for systematic calculation of isotropy groups for irreducible representations of space groups. Such calculations are necessary in the Landau theory of structural phase transitions in crystals. Therefore, both from a physical as well as mathematical point of view, it is needed to calculate isotropy groups for all irreducible representations of all space groups. Our algorithm has been designed for such a task: we have formulated the algorithm in such a way that it may be adopted for use on computers. At the present stage such a general calculation is not attempted. However, success of various computer applications in group theory is encouraging[27] and we believe that our algorithm will prove useful. In the meantime some improvements of the algorithm may be sought. Firstly, it might be advantageous to reformulate the algorithm in such a way that it uses only generators and relations. Secondly, the last step of the algorithm, the identification of space group labels has to be systematized.[28,11]

Until the analysis of all space groups is made the algorithm can be readily used for the analysis of a particular group and the representation at hand. The merit of the algorithm is that it requires only knowledge of partial characters of the representation. Furthermore, it gives a systematic procedure ensuring that *all* isotropy groups will be determined.

In conclusion we would like to emphasize again the relevance of isotropy groups in the theory of phase transitions: Isotropy groups are the only possible low-symmetry groups occurring in a transition driven by an order parameter belonging to a given representation of the high-symmetry group. The transition may be continuous, multicritical or even a "symmetry induced" discontinuous one.[29] This is a group theoretical conclusion independent of a particular theory used (e.g., Landau theory or renormalization group theory). Consequently, group theory alone cannot decide on the type of a transition. For this we must resort to a particular physical theory.

In the case of the Landau theory it has been conjectured that only maximal isotropy groups (other than G) may occur in a simple continuous transition.[6,7] Still, in order to know exactly in which of the maximal isotropy

groups the symmetry will actually break it is necessary to minimize the Landau free energy.

Minimization of the free energy reduces to a solution of a system of nonlinear equations. However, these equations transform as a vector under **R**. Hence, the knowledge of the isotropy groups may be used via projection operator techniques to solve the equations.[5] Therefore, the knowledge of isotropy groups does not only give all the possible low-symmetry groups but it also facilitates a systematic calculation of actual minima of the Landau free energy.

Note added in proof: Counterexamples to the maximality conjecture have been found recently[30] adding to the importance of the general method developed here.

After the submittal of this paper, a paper by Deonarine and Birman[31] has appeared. The results of the present paper and its preprint contain and go beyond the results of Ref. 31.

ACKNOWLEDGMENTS

I am thankful to Professor L. Michel for reading this manuscript and the hospitality at the Institut des Hautes Etudes Scientifiques. I acknowledge an Alexander von Humboldt research fellowship and partial support from the Deutsche Forschungsgemeinschaft.

[1]L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Pergamon, New York, 1958).

[2]See, for example, A. Aharony in *Phase Transitions and Critical Phenomena*, edited by C. Domb and M. S. Green (Academic, New York, 1976), Vol. 6, p. 358 and references therein; M. V. Jarić and J. L. Birman, Phys. Rev. B **17**, 4368 (1978).

[3]A group theoretical formulation of Landau theory may be found in G. Ya. Lyubarskii, *The Application of Group Theory in Physics* (Pergamon, London, 1960).

[4]The work on group theoretical aspects of minimization of the Landau free energy has been recently reviewed by M. V. Jarić, Physica A **114**, 550 (1982).

[5]M. V. Jarić, Phys. Rev. Lett. **48**, 1641 (1982).

[6]E. Ascher, Helv. Phys. Acta **39**, 40 and 466 (1966); E. Ascher, Phys. Lett. **20**, 352 (1962).

[7]L. Michel, Preprint TH. 2716-CERN (1979), gave a rigorous statement of the maximality conjecture.

[8]Development of the chain criterion has been traced in Ref. 4.

[9]M. V. Jarić, Phys. Rev. B **23**, 3460 (1981).

[10]However, this should not be confused with the maximality conjecture, Refs. 6 and 7, which selects the maximal isotropy groups among all isotropy groups excluding G itself, i.e., for all $i(L) = 1,2,...,c$.

[11]M. V. Jarić and M. Senechal (to be published).

[12]The space group may be finitized using Born-Kármán periodic boundary conditions, Eq. (1) may be evaluated, and the limit of infinite period may be taken. This elaborate and nonrigorous procedure is, as we will show, unnecessary.

[13]Use of images and kernels for space groups was emphasized by L. Michel and J. Mozrzymas, Lecture Notes in Physics **79**, 447 (1977).

[14]J. L. Birman, *Theory of Crystal Space Groups and Infrared and Raman Lattice Processes of Insulating Crystals*, Handbuch der Physik 25/2 b, (Springer-Verlag, Heidelberg, 1974).

[15]C. J. Bradley and A. P. Cracknell, *The Mathematical Theory of Symmetry in Solids* (Clarendon, Oxford, 1972).

[16]N. F. M. Henry and K. Lonsdale, *International Tables for X-ray Crystallography* (Kynoch, Birmingham, 1969).

[17]S. C. Miller and W. F. Love, *Irreducible Representations of Space Groups and Co-Representations of Magnetic Space Groups* (Pruett, Boulder, 1967).

[18]Consequently, $|\mathbf{P}_L| = |\mathbf{P}'_L|$.

[19]Note that two different domains may have the same isotropy groups. The number of domains (number of equivalent absolute minima of the free energy) is $\delta = |\mathbf{R}|/|\mathbf{L}| = |\mathbf{T}||\mathbf{P}|/|\mathbf{T}_L||\mathbf{P}_L|$.

[20]See, for example, W. Gorzkowski, Phys. Status Solidi **3**, 910 (1963); L. F. Mattheiss, Progress Report, Solid State and Molecular Theory Group, M.I.T. QPR 51, 54 (1964) (unpublished); M. Weger and I. B. Goldberg, in *Solid State Physics*, edited by H. Ehrenreich, F. Seitz, and D. Turnbull (Academic, New York, 1973), Vol. 28; L. F. Mattheiss, Phys. Rev. B **12**, 2161 (1975).

[21]O. V. Kovalev, *Irreducible Representations of the Space Groups* (Gordon and Breach, New York, 1965).

[22]L. P. Gor'kov, Zh. Eksp. Teor. Fiz. **65**, 1658 (1973) [Sov. Phys. JETP **38**, 830 (1974)]; L. P. Gor'kov *Progress in Low Temperature Physics*, Vol. VIIIb, edited by D. F. Brewer (North-Holland, Amsterdam, 1978), pp. 517–589.

[23]All isotropy groups for all X- and R-point irreducible representations have been calculated by M. V. Jarić, Phys. Rev. B **25**, 2015 (1982).

[24]The notation here follows M. V. Jarić, Ph.D. thesis, CUNY, 1977 (unpublished), where some of the relevant group theory of O_h^3 is given.

[25]In fact, **R** is isomorphic to the space group D_{3d}^6 with Born–Kármán periodic boundary conditions, period being two lattice spacings in all three directions.

[26]G. R. Johnson and D. H. Douglass, J. Low Temp. Phys. **14**, 576 (1974); V. M. Pan *et al.*, Zh. Eksp. Teor. Fiz. Pis'ma Red. **21**, 494 (1975) [JETP Lett. **21**, 228 (1975)].

[27]For a review see B. L. Davies, Physica A **114**, 507 (1982).

[28]From a practical point of view special care must be taken when identifying a space group with nonprimitive (conventional) Bravais lattice.

[29]P. Tolédano and G. Pascoli, Ferroelectrics **25**, 427 (1980).

[30]M. V. Jarić (to be published); D. Mukamel and M. V. Jarić, Preprint No. WIS-83/16 April-Ph (to be published).

[31]S. Deonarine and J. L. Birman, Phys. Rev. B **27**, 4261 (1983).