Structural analysis and elementary representations of SL(4,ℝ) and GL(4,ℝ) and their covering groups[a]

V. K. Dobrev[b] and O. Ts. Stoytchev
Institute of Nuclear Research and Nuclear Energy, Bulgarian Academy of Sciences, Sofia 1184, Bulgaria

(Received 31 October 1984; accepted for publication 6 November 1985)

The structure of the groups SL(4,R) and GL(4,R), their universal covering groups $\overline{\mathrm{SL}(4,\mathrm{R})}$ and $\overline{\mathrm{GL}(4,\mathrm{R})}$, respectively, and Lie algebras sl(4,R) and gl(4,R), respectively, are studied. The parabolic subgroups and subalgebras are identified and the cuspidal parabolic subgroups singled out. The Iwasawa and Bruhat decompositions are given explicitly. All elementary representations (ER) of SL(4,R) are explicitly given in two equivalent realizations. Using the preceding detailed structural analysis the SL(4,R) constructions are used for the explicit realization of all ER of $\overline{\mathrm{SL}(4,\mathrm{R})}$, GL(4,R), and $\overline{\mathrm{GL}(4,\mathrm{R})}$. The results shall be applied (among other things) elsewhere for the construction of all irreducible representations of the above groups.

I. INTRODUCTION

The groups SL(4,R) and GL(4,R), and their universal (double) covering groups $\overline{\mathrm{SL}(4,\mathrm{R})}$ and $\overline{\mathrm{GL}(4,\mathrm{R})}$, respectively, are of physical interest mainly because of their possible applications to gravity.[1–6] A nice review of these and other applications, e.g., in hadron physics, is contained in Ref. 6. (See also a recent proposal for a $\overline{\mathrm{SL}(4,\mathrm{R})}$ classification of hadrons.[7])

Despite this interest there is no constructive description of the irreducible representation of these groups and of their properties. This paper is the first in a program aiming to give the explicit constructions of the *elementary representations* of the above groups and to study their properties and applications. We recall that the elementary representations of a semisimple (or reductive) Lie group G are those induced from the cuspidal parabolic subgroups $P = MAN$ of G, the induction being from discrete series representations of the subgroup M (such representations exist by cuspidality), from arbitrary characters of the Abelian group A, and trivially from the nilpotent group N (see below for more details). The elementary representations (ER) are important since they *exhaust all irreducible representations.* The exact statement is (Langlands,[8] Knapp and Zuckerman[9]) that every admissible irreducible representation of a connected real semisimple Lie group G is equivalent either to an irreducible ER of G or to an irreducible component of a reducible ER of G. The condition for admissibility is purely technical since there is no known example of nonadmissible irreducible representations for such groups.

There are few papers that work with the ER of these groups. (In the mathematical literature the ER are usually called generalized principal series representations.) In Refs. 10 and 11 the reducibility of the principal series of unitary representations of SL(n,R), GL(n,R), respectively, was studied. In Ref. 12 the correspondence between the ER of GL(n,R) and GL(n,C) is discussed in another context. Most relevant for our considerations is the paper of Speh.[13] This paper gives the classification of all unitary irreducible representations of GL(4,R). However, it does not give the explicit construction of the ER and of the intertwining operators between them. Such explicit constructions are suitable for analytic computations needed in the physical applications, as we know from earlier experience.[14–16] For instance, the kernels of the integral intertwining operators provide (in the mathematical aspect) the scalar products in the cases of unitary irreducible representations. In the physical applications these operators are the Green's functions for the G-invariant wave equations or the physical propagators (two-point functions) in a G-invariant quantum field theory. They are also useful in the building of G-invariant actions.

In the mathematical physics literature (cf. Refs. 17–19 and references therein), representations of $\overline{\mathrm{SL}(4,\mathrm{R})}$ and of the connected component of $\overline{\mathrm{GL}(4,\mathrm{R})}$ are usually induced from the representations of the maximal compact subgroup $\overline{\mathrm{SO}(4)}$. Such induced representations are only a small subset of the ER. In Ref. 17 some series of (reducible) unitary representations of $\overline{\mathrm{GL}(4,\mathrm{R})}_c$ are given. Reference 18 is concentrated on the classification only of the unitary representations of $\overline{\mathrm{SL}(4,\mathrm{R})}$, which contain each irreducible representation of $\overline{\mathrm{SO}(4)}$ at most once. In Ref. 19 some partial cases of ER are used and (what amounts to) induction from the Lorentz group SO(3,1) suitably imbedded in SL(4,R) is also used. However, the emphasis in this paper is on the $\overline{\mathrm{GL}(4,\mathrm{R})}$-covariant extensions of the Dirac equation rather than the systematic exposition of the representation theory.

We give now a brief outline of our program. This paper studies the structure of the groups and Lie algebras and gives the construction of the elementary representations. Subsequent papers will deal with construction of the integral and differential intertwining operators between (partially) equivalent representations; identification of the known representations with some ER (see also the end of this paper); invariant sesquilinear forms on pairs of ER and construction of the unitary irreducible representations; study of the invariant subspaces of the reducible ER; physical applications as outlined above as well as to exactly integrable systems, etc.

---
[a] Previous title: Structural analysis of SL (4,R), GL (4,R), their covering groups and Lie algebras.
[b] Address from 1 June 1984 till 15 August 1985: Institute for Theoretical Physics, Technical University Clausthal, 3392 Clausthal-Zellerfeld, Federal Republic of Germany.

 0022-2488/86/040883-17$02.50

Analogous programs have been carried out successfully for the groups $SO(n,1)$ (see Refs. 14 and 15), Spin(5,1) (see Ref. 16), and SU(2,2) (see Refs. 20 and 21). Despite some similarities there are considerable differences. For instance, the group $SL(4,\mathbb{R})$ has split-rank 3 and is maximally split while $SO(n,1)$ and Spin(5,1) have split-rank 1 and SU(2,2) has split-rank 2. This makes the structure analysis and the representation theory much more difficult, as we shall see.

Our basic mathematical reference is Warner.[22] All notions not referred to other sources can be found there or in Refs. 15, 16, and 21.

The organization of the present paper is as follows: Section II is devoted to the study of the Lie algebras of $SL(4,\mathbb{R})$ and $GL(4,\mathbb{R})$. We introduce notation for the basis of the Lie algebra $\mathfrak{g} \equiv sl(4,\mathbb{R})$ and display the three nonconjugate Cartan subalgebras (they are all noncompact). We make explicit the well-known isomorphism with the Lie algebra so(3,3). Then (Sec. II B) we give the root system of the pair $(\mathfrak{g},\mathfrak{a})$ (where $\mathfrak{a}$ in this case is the most noncompact Cartan subalgebra), which is isomorphic to the root system of $(\mathfrak{g}^{\mathbb{C}},\mathfrak{h}^{\mathbb{C}})$ ($\mathfrak{g}^{\mathbb{C}},\mathfrak{h}^{\mathbb{C}}$ are the complexifications of $\mathfrak{g}$ and any of the real Cartan subalgebras, respectively) because $\mathfrak{g}$ is maximally split. The Weyl group $W(\mathfrak{g},\mathfrak{a}) = W(\mathfrak{g}^{\mathbb{C}},\mathfrak{h}^{\mathbb{C}})$ is presented in Sec. II C. In Sec. II D we study the parabolic subalgebras of $\mathfrak{g}$. Although we know how many they should be and how they should be found, we do not know them in general. The identification of the parabolic subalgebras is the main objective of Sec. II. Then in Sec. II E we introduce some nonparabolic subalgebras suitable for the comparison with SO(4) and SO(3,1) inductions in literature. In Sec. II F we give the structure of the Lie algebra $\mathfrak{g}^e = gl(4,\mathbb{R})$. This is not done independently from the $\mathfrak{g} = sl(4,\mathbb{R})$ case but rather uses the fact that $\mathfrak{g}^e = \mathfrak{g} \oplus \mathfrak{z}$ ($\mathfrak{z}$ is a one-dimensional center).

Section III deals with the structural analyses of $SL(4,\mathbb{R})$, $GL(4,\mathbb{R})$, and their universal covering groups. We first study in detail the group $G = SL(4,\mathbb{R})$ and then give the analogous analyses for the other groups. In Sec. III A we introduce by explicit parametrization the important (for ER) subgroups of G: the maximal compact subgroup $K \cong SO(4)$, the Abelian noncompact subgroup $A = \exp(\mathfrak{a})$, the nilpotent subgroups $\widetilde{N}$ and N, which exponentiate the positive and negative root spaces, respectively, in $\mathfrak{g}$ with respect to $\mathfrak{a}$, the centralizer M of A in K, and the minimal parabolic subgroup $P_0 = MAN$. Then we construct explicitly the other parabolic subgroups $P' = M'A'N'$ ($A' \subset A, N' \subset N, M' \supset M$ is the centralizer of A' in G) and prove in Proposition 1 that the only cuspidal parabolic subgroup is the minimal one. Thus the P_0-induced representations shall give all elementary representations. In Sec. III B we introduce the analogs of the above subgroups for $\overline{SL(4,\mathbb{R})}$, $GL(4,\mathbb{R})$, and $\overline{GL(4,\mathbb{R})}$ and prove in Proposition 2 that for each of these groups the only cuspidal parabolic subgroup is the respective minimal one. In Sec. III C we give matrix realizations of the elements of the Weyl group $W(\mathfrak{g},\mathfrak{a})$, which are needed in the explicit construction of the ER and of the intertwining operators between them.[23]

In Sec. IV we consider the Iwasawa[24] and the (Gel'fand–Naimark[25]–) Bruhat[26] decompositions of our groups. The Iwasawa decomposition (Sec. IV A) is more difficult; however, we shall only need the Iwasawa decomposition of the group $\widetilde{N}$, which we give explicitly. In Sec. IV B we construct explicitly the Bruhat decomposition for $SL(4,\mathbb{R})$ in Propositions 3 and 4 and for $GL(4,\mathbb{R})$ in Proposition 5. In Sec. IV C we discuss the Haar measure on G and its subgroups. We make explicit an important connection between the Haar measures on K and on $\widetilde{N}$.

Section V is devoted to the explicit construction of the elementary representations of our groups. We first give the constructions for $SL(4,\mathbb{R})$ in Secs. V A–V C. In Sec. V A the general picture of the ER with representation space in $C^\infty(G,\mathbb{C})$ is introduced. We briefly discuss the lowest weight module (over $\mathfrak{g}^{\mathbb{C}}$) structure of the ER (introduced in general in Ref. 27). In Sec. V B we construct the noncompact picture of the ER. The representation space is comprised from functions in $C^\infty(\widetilde{N},\mathbb{C})$ with special asymptotic properties that are given explicitly. These properties ensure the C^∞ action of the ER in the noncompact picture (Proposition 6) and its equivalence to the general picture (Proposition 7). The principal series of unitary representations is identified. In Sec. V C the infinitesimal generators in the noncompact picture are given explicitly. The values of the Casimir operators for the ER are given, the second-order Casimir being evaluated in two different ways. In Sec. V D the ER of $\overline{SL(4,\mathbb{R})}$, $GL(4,\mathbb{R})$, and $\overline{GL(4,\mathbb{R})}$ are introduced. Due to the detailed structural analysis of the groups and their Lie algebras we are enabled to use the $SL(4,\mathbb{R})$ constructions with very few changes the not so obvious of which are given. In particular, the principal series of unitary representations are identified in the most general setup, which is also compared with the literature. For $GL(4,\mathbb{R})$ and $\overline{GL(4,\mathbb{R})}$ the first-order Casimir operator connected with their one-dimensional center is given.

II. STRUCTURE OF THE LIE ALGEBRAS OF SL(4,R) AND GL(4,R)

A. Realization of SL(4,R) and of its Lie algebra sl(4,R)

The group $SL(4,\mathbb{R})$ is defined standardly by

$$G \equiv SL(4,\mathbb{R}) = \{ g \in GL(4,\mathbb{R}) \mid \det g = 1 \}. \tag{2.1}$$

The Lie algebra $\mathfrak{g} = sl(4,\mathbb{R})$ of G is comprised by the real 4×4 traceless matrices X,

$$\operatorname{Tr} X = 0. \tag{2.2}$$

The Cartan involution θ,

$$\theta X \equiv -{}^tX \tag{2.3}$$

(${}^tX$ is the transpose of X), provides the Cartan decomposition of $\mathfrak{g}$:

$$\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}, \tag{2.4}$$

where $\mathfrak{k}$ is the maximal compact subalgebra of $\mathfrak{g}$ and $\mathfrak{p}$ is a vector subspace of $\mathfrak{g}$ such that

$$X \in \mathfrak{k} \Rightarrow \theta X = X, \quad X \in p \Rightarrow \theta X = -X. \tag{2.5}$$

Explicitly we have (l.s. stands for linear span)

$$\mathfrak{k} = \text{l.s.}\{X_{ij} \equiv e_{ij} - e_{ji}, \quad 1 \leqslant i < j \leqslant 4\} \cong so(4), \tag{2.6}$$

where e_{ij}, $i,j = 1,\dots,4$ are the standard matrices with only one nonzero entry on the ith row and jth column

$$(e_{ij})_{kl} = \delta_{ik}\delta_{jl}, \tag{2.7}$$

$$\mathfrak{p} = \text{l.s.}\{Y_{ij} \equiv e_{ij} + e_{ji}, \quad 1 < i < j \leqslant 4; \quad \hat{e}_1,\hat{e}_2,\hat{e}_3\}, \tag{2.8a}$$

where

$$\hat{e}_1 = e_{11} - e_{33}, \quad \hat{e}_2 = e_{22} - e_{44},$$

$$\hat{e}_3 = \tfrac{1}{2}(e_{11} - e_{22} + e_{33} - e_{44}). \tag{2.8b}$$

Further we list the other important subalgebras of $\mathfrak{g}$. Let $\mathfrak{a}$ be the subspace of $\mathfrak{p}$, which is a maximal Abelian subalgebra. The dimension of this algebra is called the split rank of $\mathfrak{g}$. In our case $\dim \mathfrak{a} = 3$, so the split rank of $\mathfrak{g}$ is equal to its rank and $\mathfrak{g}$ is said to be maximally split. It is natural to choose $\mathfrak{a}$ to represent the most noncompact Cartan subalgebra $\mathfrak{h}_0$ of $\mathfrak{g}$. For the basis of $\mathfrak{a}$ we choose $\hat{e}_1,\hat{e}_2,\hat{e}_3$ [(2.8b)], so

$$\mathfrak{h}_0 = \mathfrak{a} = \text{l.s.}\{\hat{e}_1,\hat{e}_2,\hat{e}_3\}. \tag{2.9}$$

The other two nonconjugate Cartan subalgebras of $\mathfrak{g}$, which are also noncompact, can be chosen as

$$\mathfrak{h}_1 = \text{l.s.}\{\hat{e}_1 + \hat{e}_2, Y_{12}, X_{34}\}, \tag{2.10}$$

$$\mathfrak{h}_2 = \text{l.s.}\{\hat{e}_1 + \hat{e}_2, X_{12}, X_{34}\}. \tag{2.11}$$

Note that $\mathfrak{h}_k$ $(k = 0,1,2)$ has k compact generators.

It is known that $\mathfrak{g}$ is isomorphic to so(3,3). The expressions for the generators Z_{AB} $(A,B = 1,2,\ldots,6)$ of so(3,3) are

$$\begin{aligned} &Z_{12} = \tfrac{1}{2}(X_{13} - X_{24}), \quad Z_{13} = \tfrac{1}{2}(X_{14} - X_{32}),\\ &Z_{23} = \tfrac{1}{2}(X_{12} + X_{34}), \quad Z_{45} = \tfrac{1}{2}(X_{13} + X_{24}),\\ &Z_{46} = \tfrac{1}{2}(X_{14} + X_{32}), \quad Z_{56} = \tfrac{1}{2}(X_{12} - X_{34}), \end{aligned} \tag{2.12a}$$

$$Z_{14} = \tfrac{1}{2}(\hat{e}_1 + \hat{e}_2), \quad Z_{25} = \tfrac{1}{2}(\hat{e}_1 - \hat{e}_2), \quad Z_{36} = \hat{e}_3, \tag{2.12b}$$

$$\begin{aligned} &Z_{15} = \tfrac{1}{2}(-Y_{13} - Y_{24}), \quad Z_{16} = \tfrac{1}{2}(-Y_{14} + Y_{23}),\\ &Z_{24} = \tfrac{1}{2}(-Y_{24} + Y_{13}), \quad Z_{26} = \tfrac{1}{2}(-Y_{12} - Y_{34}),\\ &Z_{34} = \tfrac{1}{2}(Y_{23} + Y_{14}), \quad Z_{35} = \tfrac{1}{2}(-Y_{34} + Y_{12}). \end{aligned} \tag{2.12c}$$

Indeed

$$[Z_{AB},Z_{CD}] = \eta_{AC}Z_{BD} + \eta_{BD}Z_{AC} - \eta_{AD}Z_{BC} - \eta_{BC}Z_{AD},$$

where $A,B,C,D = 1,2,\ldots,6$, $\eta_{11} = \eta_{22} = \eta_{33} = -\eta_{44} = -\eta_{55} = -\eta_{66} = 1$, and $\eta_{AB} = 0$ for $A \neq B$. Note that the (2.12a) span so(4) = so(3) $\oplus$ so(3), Z_{ab}, $a,b = 1,2,3$, $a < b$ or $a,b = 4,5,6$, $a < b$, spanning the two so(3) subalgebras.

B. Root systems and the Iwasawa decomposition

Denote by $\mathfrak{a}^*$ the space of real linear functionals over $\mathfrak{a}$. Define for $\lambda \in \mathfrak{a}^*$, $\lambda \neq 0$,

$$\mathfrak{g}_\lambda \equiv \{X \in \mathfrak{g} \mid [\hat{e}_a,X] = \lambda(\hat{e}_a)X, \quad a = 1,2,3\}, \tag{2.13a}$$

$$\Lambda \equiv \{\lambda \in \mathfrak{a}^* \mid \lambda \neq 0, \quad \mathfrak{g}_\lambda \neq \{0\}\}. \tag{2.13b}$$

We easily obtain

$$\Lambda = \{\pm\lambda_k, \ k = 1,2,\ldots,6\}, \tag{2.14a}$$

and we choose the set Λ^+ of positive roots to be (enumerated in the order of largeness with this ordering)

$$\begin{aligned} &\lambda_1(\hat{e}_1,\hat{e}_2,\hat{e}_3) = (0,2,0), \quad \lambda_2 = (1,-1,-1),\\ &\lambda_3 = (1,-1,1), \quad \lambda_4 = (1,1,-1),\\ &\lambda_5 = (1,1,1), \quad \lambda_6 = (2,0,0),\\ &\lambda_4 = \lambda_1 + \lambda_2, \quad \lambda_5 = \lambda_1 + \lambda_3, \quad \lambda_6 = \lambda_1 + \lambda_2 + \lambda_3. \end{aligned} \tag{2.14b}$$

(The simple roots are $\lambda_1,\lambda_2,\lambda_3$.) The corresponding root spaces $\mathfrak{g}_k^\pm \equiv \mathfrak{g}_{\pm\lambda_k}$ are spanned by the root vectors $E_k^\pm$:

$$\begin{aligned} &E_1^+ = e_{24}, \quad E_2^+ = e_{43}, \quad E_3^+ = e_{12}, \quad E_4^+ = e_{23},\\ &E_5^+ = e_{14}, \quad E_6^+ = e_{13}, \quad E_k^- = {}^tE_k^+. \end{aligned} \tag{2.15}$$

We define the positive and negative root spaces

$$\tilde{\mathfrak{n}} \equiv \bigoplus_k \mathfrak{g}_k^+, \quad \mathfrak{n} \equiv \bigoplus_k \mathfrak{g}_k^-. \tag{2.16}$$

Of course $\tilde{\mathfrak{n}} = \theta\mathfrak{n}$ and we can write the standard decomposition

$$\mathfrak{g} = \tilde{\mathfrak{n}} \oplus \mathfrak{a} \oplus \mathfrak{n}. \tag{2.17}$$

Then we note that the map

$$J: \tilde{\mathfrak{n}} \to \mathfrak{k},$$

$$J(X) \equiv X + \theta X, \quad X \in \tilde{\mathfrak{n}}, \tag{2.18}$$

is bijective. So

$$\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{a} \oplus \mathfrak{n} \tag{2.19}$$

is the *Iwasawa decomposition* of $\mathfrak{g}$ (see Ref. 24).

We now turn to the root system of the complexified Lie algebra $\mathfrak{g}^{\mathbb{C}} \cong$ sl(4,$\mathbb{C}$). Its Cartan subalgebra $\mathfrak{h}^{\mathbb{C}}$ is unique (up to conjugation) and is the complexification of any of the Cartan subalgebras $\mathfrak{h}_a$ $(a = 0,1,2)$ of $\mathfrak{g}$. Since sl(4,$\mathbb{R}$) is the normal real form of sl(4,$\mathbb{C}$) (see Ref. 28) we can (and it is useful to) choose in $\mathfrak{h}^{\mathbb{C}}$ the same basis as in $\mathfrak{a}$ and not to use the standard basis of $\mathfrak{h}^{\mathbb{C}}$, consisting of $e_{kk} - e_{k+1\,k+1}$ $(k = 1,2,3)$. Then of course the root system of $(\mathfrak{g}^{\mathbb{C}},\mathfrak{h}^{\mathbb{C}})$ coincides with that of $(\mathfrak{g},\mathfrak{a})$. The corresponding root spaces $\tilde{\mathfrak{g}}_k^\pm \equiv \mathfrak{g}^{\mathbb{C}}_{\pm\lambda_k}$ are complexly spanned by the same root vectors $E_k^\pm$ (2.15). Then we have for the analogs of (2.16) and (2.17):

$$\tilde{\mathfrak{n}}^{\mathbb{C}} \equiv \bigoplus_k \tilde{\mathfrak{g}}_k^+, \quad \mathfrak{n}^{\mathbb{C}} \equiv \bigoplus_k \tilde{\mathfrak{g}}_k^-, \tag{2.20a}$$

$$\mathfrak{g}^{\mathbb{C}} = \tilde{\mathfrak{n}}^{\mathbb{C}} \oplus \mathfrak{h}^{\mathbb{C}} \oplus \mathfrak{n}^{\mathbb{C}}. \tag{2.20b}$$

C. The Weyl group W($\mathfrak{g}$,$\mathfrak{a}^0$)

We define for every $\lambda_k \in \Lambda^+$ a vector $H_k \in \mathfrak{a}^0$ by

$$B(H_k,\hat{e}_a) = \lambda_k(\hat{e}_a) \quad (a = 1,2,3), \tag{2.21}$$

where B is the Killing form on $\mathfrak{g}$ with normalization $B(X,Y) = \text{Tr}(XY)$. So we obtain

$$\begin{aligned} &H_1 = \hat{e}_2, \quad H_2 = \tfrac{1}{2}(\hat{e}_1 - \hat{e}_2) - \hat{e}_3,\\ &H_3 = \tfrac{1}{2}(\hat{e}_1 - \hat{e}_2) + \hat{e}_3,\\ &H_4 = H_1 + H_2, \quad H_5 = H_1 + H_3,\\ &H_6 = H_1 + H_2 + H_3. \end{aligned} \tag{2.22}$$

Note also

$$\lambda_k(H_k) = 2, \quad H_k = [E_k^+,E_k^-], \quad k = 1,\ldots,6, \tag{2.23}$$

which is an equivalent definition [instead of (2.21)]. The Weyl reflections in $\mathfrak{a}$, corresponding to the positive roots, are standardly defined as

$$w_k(X) \equiv X - \frac{2\lambda_k(X)}{\lambda_k(H_k)} H_k = X - \lambda_k(X)H_k, \quad X \in \mathfrak{a}, \tag{2.24a}$$

with the explicit actions on the basis of $\mathfrak{a}$ given by

$$\begin{aligned} &w_1(\hat{e}_1,\hat{e}_2,\hat{e}_3) = (\hat{e}_1, -e_2,\hat{e}_3),\\ &w_2(\hat{e}_1,\hat{e}_2,\hat{e}_3) = (H_5,H_4,\tfrac{1}{2}(\hat{e}_1-\hat{e}_2)),\\ &w_3(\hat{e}_1,\hat{e}_2,\hat{e}_3) = (H_4,H_5,-\tfrac{1}{2}(\hat{e}_1-\hat{e}_2)),\\ &w_4(\hat{e}_1,\hat{e}_2,\hat{e}_3) = (H_3,-H_2,\tfrac{1}{2}(\hat{e}_1+\hat{e}_2)),\\ &w_5(\hat{e}_1,\hat{e}_2,\hat{e}_3) = (H_2,-H_3,-\tfrac{1}{2}(e_1+\hat{e}_2)),\\ &w_6(\hat{e}_1,\hat{e}_2,\hat{e}_3) = (-\hat{e}_1,\hat{e}_2,\hat{e}_3), \quad w_k(H_k) = -H_k. \end{aligned} \tag{2.24b}$$

We note from (2.24),

$$\begin{aligned} &w_k^2 = \mathrm{id}, \quad k = 1,\ldots,6,\\ &w_4 = w_1w_2w_1 = w_2w_1w_2, \quad w_5 = w_1w_3w_1 = w_3w_1w_3,\\ &w_6 = w_3w_4w_3 = w_2w_5w_2, \quad w_2w_3 = w_3w_2. \end{aligned} \tag{2.25}$$

We choose for the generating elements of the Weyl group $W(\mathfrak{g},\mathfrak{a})$, w_1, w_2, and w_3, which correspond to the simple roots. Then we have for the 24 elements of W:

$$\begin{aligned} W(\mathfrak{g},\mathfrak{a}) = \{&\mathrm{id}, w_k \ (k=1,\ldots,6), w_{12}, w_{13}, w_{31}, w_{23}, w_{21},\\ &w_{123}, w_{231}, w_{213}, w_{312}, w_{1231}, w_{1213}, w_{1312}, w_{2131},\\ &w_{3121}, w_{12131}, w_{13121}, w_{131213}\}; \end{aligned} \tag{2.26a}$$

$$w_{i_1 i_2 \cdots i_n} \equiv w_{i_1} w_{i_2} \cdots w_{i_n}. \tag{2.26b}$$

When $w \in W(\mathfrak{g},\mathfrak{a})$ is expressed as in (2.26), we say that it is given in a reduced form. The induced action on the roots is defined as

$$w_k^* \lambda_j \equiv \lambda_j \circ w_k, \tag{2.27a}$$

from which follows

$$w_k^* \lambda_j = \lambda_j - \lambda_j(H_k)\lambda_k. \tag{2.27b}$$

For the generating elements, (2.27) gives

$$\begin{aligned} &w_1^*(\lambda_1,\lambda_2,\lambda_3,\lambda_4,\lambda_5,\lambda_6) = (-\lambda_1,\lambda_4,\lambda_5,\lambda_2,\lambda_3,\lambda_6),\\ &w_2^*(\lambda_1,\lambda_2,\lambda_3,\lambda_4,\lambda_5,\lambda_6) = (\lambda_4,-\lambda_2,\lambda_3,\lambda_1,\lambda_6,\lambda_5),\\ &w_3^*(\lambda_1,\lambda_2,\lambda_3,\lambda_4,\lambda_5,\lambda_6) = (\lambda_5,\lambda_2,-\lambda_3,\lambda_6,\lambda_1,\lambda_4). \end{aligned} \tag{2.28}$$

Obviously $w_k^*\lambda_k = -\lambda_k$ and w_k^* obey relations (2.25).

Finally we note that since $\mathrm{sl}(4,\mathbb{R})$ is the normal real form of $\mathrm{sl}(4,\mathbb{C})$, the Weyl group of the complexified pair $W(\mathfrak{g}^{\mathbb{C}},\mathfrak{h}^{\mathbb{C}})$ coincides with $W(\mathfrak{g},\mathfrak{a})$.

D. The parabolic subalgebras of $\mathfrak{g}$

We recall the definition of the *minimal parabolic subalgebra* $\mathfrak{p}_0$ (see Ref. 22),

$$\mathfrak{p}_0 \equiv \mathfrak{m} \oplus \mathfrak{a} \oplus \mathfrak{n}, \tag{2.29a}$$

where $\mathfrak{m}$ is the centralizer of $\mathfrak{a}$ in $\mathfrak{k}$. In our case (as always when $\mathfrak{g}$ is maximally split) $\mathfrak{m}$ is trivial, $\mathfrak{m} = \{0\}$. Thus we have

$$\mathfrak{p}_0 \equiv \mathfrak{a} \oplus \mathfrak{n}. \tag{2.29b}$$

A *standard parabolic subalgebra*[22] of $\mathfrak{g}$ is any subalgebra of $\mathfrak{g}$ containing $\mathfrak{p}_0$. The number of standard parabolic subalgebras is $2^l = 8$, where $l = \dim \mathfrak{a} = 3$. One is $\mathfrak{g}$ itself for which one can formally write

$$\mathfrak{p}_\mathfrak{g} = \mathfrak{m}_\mathfrak{g} \oplus \mathfrak{a}_\mathfrak{g} \oplus \mathfrak{n}_\mathfrak{g}, \tag{2.30a}$$

with

$$\mathfrak{m}_\mathfrak{g} = \mathfrak{p}_\mathfrak{g} = \mathfrak{g}, \quad \mathfrak{a}_\mathfrak{g} = \mathfrak{n}_\mathfrak{g} = \{0\}. \tag{2.30b}$$

The remaining six are also given in the form (2.29a)

$$\begin{aligned} &\mathfrak{p}_{ab} = \mathfrak{m}_{ab} \oplus \mathfrak{a}_{ab} \oplus \mathfrak{n}_{ab}\\ &(a = 0,1,2, \quad a < b = 1,2,3), \end{aligned} \tag{2.31}$$

where $\mathfrak{m}_{ab}$ is the centralizer of $\mathfrak{a}_{ab}$ in $\mathfrak{g}$ (not in $\mathfrak{k}$ since $\mathfrak{m}_{ab}$ contains noncompact elements) and the $\mathfrak{a}_{ab}$ are defined as follows:

$$\mathfrak{a}_{0b} \equiv \{X \in \mathfrak{a} \mid w_b(X) = -X\} \quad (b = 1,2,3), \tag{2.32a}$$

$$\mathfrak{a}_{ab} \equiv \mathfrak{a}_{0a} \oplus \mathfrak{a}_{0b} \quad (a \neq 0). \tag{2.32b}$$

The $\mathfrak{n}_{ab}$ (resp. $\tilde{\mathfrak{n}}_{ab}$) are the negative (resp. positive) root spaces of the system $(\mathfrak{g},\mathfrak{a}_{ab})$. Explicitly we have

$$\begin{aligned} &\mathfrak{a}_{01} = \mathrm{l.s.}\ \hat{e}_2, \quad \mathfrak{a}_{02} = \mathrm{l.s.}(\hat{e}_1 - \hat{e}_2 - 2\hat{e}_3),\\ &\mathfrak{a}_{03} = \mathrm{l.s.}(\hat{e}_1 - \hat{e}_2 + 2\hat{e}_3), \end{aligned} \tag{2.33}$$

$$\begin{aligned} &\mathfrak{m}_{01} = \mathrm{l.s.}\{\hat{e}_3\} \oplus \mathrm{l.s.}\left\{\hat{e}_1, \begin{pmatrix} 0 & e_1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ e_1 & 0 \end{pmatrix}\right\},\\ &\mathfrak{m}_{02} = \mathrm{l.s.}\{\hat{e}_1 + \hat{e}_2\} \oplus \mathrm{l.s.}\left\{\frac{1}{2}(\hat{e}_1 - \hat{e}_2) + \hat{e}_3, \begin{pmatrix} \sigma_+ & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} \sigma_- & 0 \\ 0 & 0 \end{pmatrix}\right\}, \end{aligned} \tag{2.34a}$$

$$\mathfrak{m}_{03} = \mathrm{l.s.}\{\hat{e}_1 + \hat{e}_2\} \oplus \mathrm{l.s.}\left\{\frac{1}{2}(\hat{e}_1 - \hat{e}_2) - \hat{e}_3, \begin{pmatrix} 0 & 0 \\ 0 & \sigma_+ \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & \sigma_- \end{pmatrix}\right\},$$

$$\begin{aligned} &\mathfrak{m}_{12} = \mathrm{l.s.}\{2\hat{e}_1 + \hat{e}_3\}, \quad \mathfrak{m}_{13} = \mathrm{l.s.}\{2\hat{e}_1 - \hat{e}_3\},\\ &\mathfrak{m}_{23} = \mathrm{l.s.}\{\hat{e}_1 + \hat{e}_2\}, \end{aligned} \tag{2.34b}$$

$$\begin{aligned} &e_1 \equiv (\mathbb{1}_2 + \sigma_3)/2, \quad e_2 \equiv (\mathbb{1}_2 - \sigma_3)/2,\\ &\sigma_\pm \equiv \tfrac{1}{2}(\sigma_1 \pm i\sigma_2). \end{aligned} \tag{2.34c}$$

It is easy to see that

$$\mathfrak{m}_{0b} \cong \mathrm{so}(1,1) \oplus \mathrm{sl}(2,\mathbb{R}), \tag{2.35a}$$

$$\mathfrak{m}_{ab} \cong \mathrm{so}(1,1) \quad (a \neq 0). \tag{2.35b}$$

(Of course, the so(1,1) and $\mathrm{sl}(2,\mathbb{R})$ factors are imbedded differently in $\mathrm{sl}(4,\mathbb{R})$ for different a,b.)

We shall see in Sec. III A that only the minimal parabolic subgroup, corresponding to $\mathfrak{p}_0$, shall be relevant for the construction of the elementary representations of $G = \mathrm{SL}(4,\mathbb{R})$. For that reason we shall not carry out to the end the structure analysis for $\mathfrak{p}_{ab}$ as for $\mathfrak{p}_0$: root systems, restricted Weyl groups $W(\mathfrak{g},\mathfrak{a}_{ab})$, etc.

E. Nonparabolic subalgebras

Usually in the mathematical physics literature (cf. Refs. 17–19 and references therein) representations are induced from various nonparabolic subalgebras. Most often representations are induced from the representations of the maximal compact subgroup or subalgebra. Often $\mathfrak{k} \cong so(4)$ is interpreted as the Euclidean counterpart of the Lorentz subalgebra so(3,1). Our exposition below is trying to incorporate these approaches into our scheme of structure analysis. This would help us in subsequent papers to identify the known representations as some of the elementary representations.

Let $\mathfrak{a}^k$ be a subalgebra of $\mathfrak{a}$ that commutes with the Cartan subalgebra $\mathfrak{h}^k$ of $\mathfrak{k}$. We obtain [cf. (2.11.)]

$$\mathfrak{a}^k = \text{l.s.}\, D, \quad D = \tfrac{1}{2}(\hat{e}_1 + \hat{e}_2), \qquad \mathfrak{h}^k = \text{l.s.}(X_{12}, X_{34}), \quad \mathfrak{h}_2 = \mathfrak{a}^k \oplus \mathfrak{h}^k. \tag{2.36}$$

We can introduce a restricted root system of $\mathfrak{g}$ related to $\mathfrak{a}^k$. We define for $\mu \in (\mathfrak{a}^k)^*, \mu \neq 0$,

$$\mathfrak{g}_\mu \equiv \{X \in \mathfrak{g} \mid [D, X] = \mu(D) X\}, \tag{2.37}$$

$$M^k \equiv \{\mu \in (\mathfrak{a}^k)^* \mid \mu \neq 0, \quad \mathfrak{g}_\mu \neq \{0\}\} = \{\mu_+, \mu_-\}, \quad \mu_\pm(D) = \pm 1, \tag{2.38}$$

$$\tilde{\mathfrak{n}}^k \equiv \mathfrak{g}_{\mu_+} = \text{l.s.}(E_1^+, E_4^+, E_5^+, E_6^+), \tag{2.39a}$$

$$\mathfrak{n}^k \equiv \mathfrak{g}_{\mu_-} = \text{l.s.}(E_1^-, E_4^-, E_5^-, E_6^-). \tag{2.39b}$$

We also note that the $\mu_\pm$ are the restrictions of part of Λ to D:

$$\mu_\pm = \lambda_k^\pm|_D \quad (k = 1,4,5,6). \tag{2.40}$$

The restricted Weyl reflection s corresponding to μ_+ is defined as

$$s(D) = D - 2\mu_+(D) D = -D, \tag{2.41}$$

and the corresponding restricted Weyl group is $W(\mathfrak{g}, \mathfrak{a}^k) = \{\text{id}, s\}$.

Another type of nonparabolic induction (see, e.g., Mickelsson[14]) is from the Lorentz subalgebra so(3,1), itself imbedded suitably in sl(4,$\mathbb{R}$):

$$\mathfrak{m}^l \cong so(3,1) = \text{l.s.}\{X_{ij}, 1 \leqslant i < j \leqslant 3, Y_{k4}, k = 1,2,3\}. \tag{2.42}$$

Define $\mathfrak{a}^l$ to be the subalgebra of $\mathfrak{a}$ that commutes with the Cartan subalgebra $\mathfrak{h}^l$ of $\mathfrak{m}^l$. We obtain [cf. (2.10) and (2.36)]

$$\mathfrak{a}^l = \mathfrak{a}^k = \text{l.s.}\, D, \quad \mathfrak{h}^l = \text{l.s.}(X_{12}, Y_{34}), \qquad \mathfrak{h}_1 = \mathfrak{a}^l \oplus \mathfrak{h}^l. \tag{2.43}$$

F. Structure of the Lie algebra of GL(4,$\mathbb{R}$)

The Lie algebra gl(4,$\mathbb{R}$) of GL(4,$\mathbb{R}$) consists of all 4×4 real matrices. The structure analysis of gl(4,$\mathbb{R}$) is very easy after one has done the analysis of sl(4,$\mathbb{R}$) since

$$\mathfrak{g}^e \equiv gl(4,\mathbb{R}) = sl(4,\mathbb{R}) \oplus \mathfrak{z}, \qquad \mathfrak{z} = \text{l.s.}\, \hat{e}_4, \quad \hat{e}_4 = \mathbf{1}_4. \tag{2.44}$$

Thus we shall list without comment the analogous notions and formulas: Cartan decomposition [cf. (2.4) and (2.8)],

$$\mathfrak{g}^e = \mathfrak{k} \oplus \mathfrak{p}^e, \quad \mathfrak{p}^e = \mathfrak{p} \oplus \mathfrak{z}; \tag{2.45}$$

Cartan subalgebras [cf. (2.9)–(2.11)],

$$\mathfrak{h}_a^e = \mathfrak{h}_a \oplus \mathfrak{z} \quad (a = 0,1,2); \tag{2.46}$$

root system [cf. (2.13)–(2.19)] of $(\mathfrak{g}^e, \mathfrak{a}^e \equiv \mathfrak{h}_0^e)$ (note that this is not the standard root system and root vectors),

$$\Lambda^e = \{\lambda_k^e, k = 1,\dots,6\}, \tag{2.47}$$

$$\lambda_k^e(\hat{e}_1, \hat{e}_2, \hat{e}_3) = \lambda_k(\hat{e}_1, \hat{e}_2, \hat{e}_3), \quad \lambda_k^e(\hat{e}_4) = 0, \tag{2.48}$$

$$\mathfrak{g}_k^{e\pm} = \mathfrak{g}_k^\pm, \quad E_k^{e\pm} = E_k^\pm, \tag{2.49}$$

$$\tilde{\mathfrak{n}}^e = \tilde{\mathfrak{n}}, \quad \mathfrak{n}^e = \mathfrak{n}, \tag{2.50}$$

$$\mathfrak{g}^e = \tilde{\mathfrak{n}} \oplus \mathfrak{a}^e \oplus \mathfrak{n}, \tag{2.51}$$

$$\mathfrak{g}^e = \mathfrak{k} \oplus \mathfrak{a}^e \oplus \mathfrak{n}; \tag{2.52}$$

Weyl reflections and Weyl group [cf. (2.24)–(2.28)],

$$w_k^e(\hat{e}_1, \hat{e}_2, \hat{e}_3) = w_k(\hat{e}_1, \hat{e}_2, \hat{e}_3), \quad w_k^e(\hat{e}_4) = \hat{e}_4, \tag{2.53}$$

$$W^e(\mathfrak{g}^e, \mathfrak{a}^e) = W(\mathfrak{g}, \mathfrak{a}); \tag{2.54}$$

parabolic subalgebras [cf. (2.29), (2.32), and (2.33)],

$$\mathfrak{p}_0^e = \mathfrak{p}_0 \oplus \mathfrak{z} = \mathfrak{a}^e \oplus \mathfrak{n}, \tag{2.55}$$

$$\mathfrak{p}_{ab}^e = \mathfrak{p}_{ab} \oplus \mathfrak{z} = \mathfrak{m}_{ab} \oplus \mathfrak{a}_{ab}^e \oplus \mathfrak{n}_{ab}, \tag{2.56}$$

$$\mathfrak{a}_{ab}^e = \mathfrak{a}_{ab} \oplus \mathfrak{z}; \tag{2.57}$$

and nonparabolic subalgebras [cf. (2.36) and (2.43)],

$$(\mathfrak{a}^k)^e = \mathfrak{a}^k \oplus \mathfrak{z}, \quad (\mathfrak{a}^l)^e = \mathfrak{a}^l. \tag{2.58}$$

III. STRUCTURE OF SL(4,$\mathbb{R}$), $\overline{\text{SL(4,}\mathbb{R}\text{)}}$, GL(4,$\mathbb{R}$), AND $\overline{\text{GL(4,}\mathbb{R}\text{)}}$

A. Important subgroups of SL(4,$\mathbb{R}$)

We shall most often write the elements of $G = SL(4,\mathbb{R})$ as

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

where $\alpha, \beta, \gamma, \delta$ are 2×2 real matrices constrained by the condition $\det g = 1$. Additionally for each 2×2 matrix α we shall use the following decomposition [cf. (2.34c)]:

$$\alpha \equiv \alpha_1 e_1 + \alpha_2 e_2 + \alpha_+ \sigma_+ + \alpha_- \sigma_-. \tag{3.1}$$

The maximal compact subgroup K of G is given by

$$K \equiv \{g \in G \mid g^{-1} = {}^t g\} = \left\{ g = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \,\middle|\, {}^t\alpha\alpha + {}^t\gamma\gamma = \mathbf{1}_2, \quad {}^t\beta\beta + {}^t\delta\delta = \mathbf{1}_2, \quad {}^t\alpha\beta + {}^t\gamma\delta = 0, \quad \det g = 1 \right\} \cong SO(4), \tag{3.2}$$

and its Lie algebra is $\mathfrak{k}$. Further we introduce the subgroups corresponding to the subalgebras $\mathfrak{a}$, $\tilde{\mathfrak{n}}$, and $\mathfrak{n}$. We introduce parametrization of $\mathfrak{a}$, $\tilde{\mathfrak{n}}$, and $\mathfrak{n}$ [thus making explicit formulas (2.9) and (2.16)]:

$$\mathfrak{a}=\{s\hat{e}_1+t\hat{e}_2+r\hat{e}_3|s,t,r\in\mathbb{R}\}=\left\{\operatorname{diag}\left(s+\frac{r}{2},t-\frac{r}{2},-s+\frac{r}{2},-t-\frac{r}{2}\right)\right\},$$
$$\tilde{\mathfrak{n}}=\{y_+E_3^+ +y_-E_2^+ +x_1'E_6^+ +x_2'E_1^+ +x'_+E_5^+ +x'_-E_4^+ \,|\, y_\pm,x_1',x_2',x'_\pm\in\mathbb{R}\}$$
$$=\left\{\begin{pmatrix} y_+\sigma_+ & \hat{x}' \\ 0 & y_-\sigma_-\end{pmatrix},\hat{x}'\equiv x_1'e_1+x_2'e_2+x'_+\sigma_+ +x'_-\sigma_-\right\},$$
$$\mathfrak{n}=\left\{\begin{pmatrix} z_-\sigma_- & 0 \\ \hat{b}' & z_+\sigma_+\end{pmatrix},\quad \hat{b}'\text{as } \hat{x}',z_\pm,b_1',b_2',b'_\pm\in\mathbb{R}\right\}.$$

(Parameters $\hat{x}'$ and $\hat{b}'$ are primed for convenience, see below.)

Thus we can define

$$A\equiv\exp\mathfrak{a}=\left\{a=\begin{pmatrix} a_+ & 0 \\ 0 & a_-\end{pmatrix}\middle|\, a_\pm\equiv e_1e^{\pm s+r/2}+e_2e^{\pm t-r/2};s,t,r\in\mathbb{R}\right\}\cong(\mathbb{R}^+)^3, \tag{3.3}$$

$$\widetilde{N}\equiv\exp\tilde{\mathfrak{n}}=\left\{\tilde{n}=\exp\begin{pmatrix} y_+\sigma_+ & \hat{x}' \\ 0 & y_-\sigma_-\end{pmatrix}\middle|\, y_\pm,x_1',x_2',x'_\pm\in\mathbb{R}\right\}$$
$$=\left\{\left(\begin{array}{c|c} 1+y_+\sigma_+ & \widehat{X}' \\ \hline 0 & 1+y_-\sigma_-\end{array}\right),\quad \widehat{X}'=\hat{x}'+\tfrac{1}{2}[\,y+x_2'\sigma_+ +(y_+x'_- +x'_+\,y_-)e_1+x_2'\,y_-\sigma_-]+\tfrac{1}{6}y_+y_-x_2'e_1\right\} \tag{3.4a}$$

$$=\left\{\left(\begin{array}{c|c} 1+y_+\sigma_+ & \hat{x}(1+y_-\sigma_-) \\ \hline 0 & 1+y_-\sigma_-\end{array}\right),\ \hat{x}=\widehat{X}'(1-y_-\sigma_-)\right\}, \tag{3.4b}$$

$$N\equiv\exp\mathfrak{n}=\left\{n=\exp\left(\begin{array}{c|c} z_-\sigma_- & 0 \\ \hline \hat{b}' & z_+\sigma_+\end{array}\right),\ z_\pm,b_1',b_2',b'_\pm\in\mathbb{R}\right\} \tag{3.5a}$$

$$=\left\{\left(\begin{array}{c|c} 1+z_-\sigma_- & 0 \\ \hline \hat{b}(1+z_-\sigma_-) & 1+z_+\sigma_+\end{array}\right)\right\}. \tag{3.5b}$$

In (3.4b) and (3.5b) we have introduced the more convenient parameters $\hat{x}$ and $\hat{b}$ instead of $\hat{x}'$ and $\hat{b}'$ and we shall use them from now on.

Let M be the centralizer of A in K, i.e.,

$$M\equiv\{m\in K\,|m^{-1}am=a,\ \forall a\in A\}=\left\{m=\operatorname{diag}(\nu_1,\nu_2,\nu_3,\nu_4)\,|\nu_i=\pm 1,\ 1\leqslant i\leqslant 4,\prod_{i=1}^{4}\nu_i=1\right\}$$
$$=\{m_1^{N_1}m_2^{N_2}m_3^{N_3}|m_1\equiv\operatorname{diag}(-1,1,1,-1),\ m_2\equiv\operatorname{diag}(1,-1,1,-1),$$
$$m_3=\operatorname{diag}(1,1,-1,-1);N_1,N_2,N_3=0,1\}\cong(\mathbb{Z}_2)^3. \tag{3.6}$$

Thus M consists of the diagonal elements of K.

Analogously to the Lie algebra considerations $P_0\equiv MAN$ is a *minimal parabolic subgroup* of G and a *standard parabolic subgroup* is any closed subgroup of G containing P_0. We recall the standard construction of the parabolic subgroups[22]: Let $\Psi=\{w_1,w_2,w_3\}$ be the set of generating elements of $W(\mathfrak{g},\mathfrak{a})$. Then to each subset $\psi\in\Psi$ corresponds a standard parabolic subgroup of G:

$$P_\psi\equiv\bigcup_{w\in\psi}P_0\omega(w)P_0, \tag{3.7}$$

where $\omega(w)$ is some matrix representation of the elements of $W(\mathfrak{g},\mathfrak{a}_0)$ to be given explicitly in Sec. III C. Then we have

$$P_\varnothing=P_0,\quad P_{0a}=P_0\omega(w_a)P_0\quad(a=1,2,3),$$
$$P_{ab}=P_{0a}\cup P_{0b},\quad a,b=1,2,3,\quad a<b, \tag{3.8}$$
$$P_\Psi=G=M_\Psi,\quad A_\Psi=N_\Psi=\{1\},$$

or more explicitly [cf. (2.32)–(2.35) and (3.6)]:

$$P_{ab}=M_{ab}A_{ab}N_{ab},\quad a,b=0,1,2,3,\quad a<b, \tag{3.9}$$
$$A_{ab}=\exp\mathfrak{a}_{ab}\subset A, \tag{3.10a}$$
$$N_{ab}=\exp\mathfrak{n}_{ab}\subset N, \tag{3.10b}$$
$$M_{0b}\cong SO(1,1)\times((M/m_b')\circledS SL(2,\mathbb{R})),\quad b=1,2,3, \tag{3.10c}$$
$$m_1'\equiv m_1m_3,\quad m_2'\equiv m_1m_2,\quad m_3'=m_3,$$
$$M_{ab}\cong M\times SO(1,1),\quad 0\neq a<b, \tag{3.10d}$$

where $G'\circledS G''$ denotes the semidirect product with G' acting on G''. Note that SO(1,1) and SL(2,R) are differently imbedded in SL(4,R) for different pairs of indices of a,b [cf. (2.34) and (2.35)]. The factorization in (3.10c) is needed since in each case the element m_b' is contained also in the corresponding imbedding of SL(2,R).

Further we shall investigate the cuspidality of the parabolic subgroups. A parabolic subgroup $P'=M'A'N'$ of an arbitrary semisimple (or reductive) Lie group is said to be *cuspidal* iff the subgroup M' has discrete series representations. The minimal parabolic subgroup P_0 is cuspidal for any group. For the case $P'=G$ it is known that SL(4,R) does not have discrete series representations. Indeed recall the Harish-Chandra criterion: a semisimple Lie group G has discrete series representations iff rank $G = $ rank K, while for $G = \mathrm{SL}(4,\mathbf{R})$, rank $G = 3 \neq$ rank $K = 2$ [$K = \mathrm{SO}(4)$]. In the six cases $P' = P_{ab}$, M_{ab} is either isomorphic to SO(1,1) or contains SO(1,1) as a direct product factor. Since SO(1,1) does not have discrete series representations [rank $\mathrm{SO}(1,1) = 1 \neq \mathrm{rank}\, K(\mathrm{SO}(1,1)) = 0$] neither does M_{ab} have. Thus we have proved the following proposition.

Proposition 1: The only cuspidal parabolic subgroup of SL(4,R) is the minimal parabolic subgroup P_0.

Remark: By the Langlands[8]–Knapp–Zuckerman[9] theorem only P_0 shall be needed to construct the elementary representations of G. That is why we do not give in so much detail the explicit parametrization in the case of the other parabolics.

Further we introduce the subgroup, corresponding to the nonparabolic subalgebra $\mathfrak{a}^k$ (2.36):

$$A^k = \exp(\mathfrak{a}^k) = \left\{ a^k \in G \,|\, a^k = \begin{pmatrix} \sqrt{|a|}\mathbf{1} & 0 \\ 0 & \sqrt{|a|}^{-1}\mathbf{1} \end{pmatrix}, \quad |a| \in \mathbf{R}^+ \right\}. \tag{3.11}$$

Later we shall give the exact imbedding of SO(3,1) in SL(4,R) after we display the Iwasawa decomposition (Sec. IV A).

B. Important subgroups of $\overline{\mathrm{SL}(4,\mathbf{R})}$, GL(4,R), and $\overline{\mathrm{GL}(4,\mathbf{R})}$

We turn now to the universal (double) covering $\overline{G} = \overline{\mathrm{SL}(4,\mathbf{R})}$ of SL(4,R). The maximal compact subgroup $\overline{K}$ of $\overline{G}$ is the double covering of K,

$$\overline{K} = \mathrm{SU}(2)\times\mathrm{SU}(2) = \overline{\mathrm{SO}(4)}. \tag{3.12}$$

Explicitly the double covering map $\overline{K} \mapsto K$ is given by

$$\mathrm{SU}(2)\times\mathrm{SU}(2) \ni (u,v) \mapsto (k_{ij}(u,v)) \in \mathrm{SO}(4), \tag{3.13a}$$

$$k_{ij}(u,v) \equiv \tfrac{1}{2}\mathrm{tr}(q_i^+ u q_j v^+), \tag{3.13b}$$

where q_j represent the quaternion units

$$q_4 = \mathbf{1}_2, \quad q_j = -i\sigma_j = -q_j^+ \quad (j = 1,2,3), \tag{3.13c}$$

through which u,v are also conveniently expressed as

$$u = u_1 q_1 + u_2 q_2 + u_3 q_3 + u_4 q_4, \quad u_k \in \mathbf{R}, \qquad \det u = u_1^2 + u_2^2 + u_3^2 + u_4^2 = 1. \tag{3.13d}$$

The groups A, $\widetilde{N}$, and N and the nonparabolic A^k are simply connected and can be thought of as subgroups of $\overline{G}$ also.

The centralizer $\overline{M}$ of A in $\overline{K}$ is

$$\overline{M} \equiv \{(u,v) \in \overline{K} \,|\, k(u,v)^{-1} a k(u,v) = a, \quad \forall a \in A\}$$

$$= \bigcup_{\substack{k=1 \\ \epsilon = \pm 1}}^{4} \overline{M}_k^\epsilon, \tag{3.14a}$$

$$\overline{M}_k^\epsilon \equiv \{(\pm q_k, \pm \epsilon q_k)\}. \tag{3.14b}$$

The 16 elements of $\overline{M}$ may be generated by

$$\overline{m}_k \equiv (q_k, -q_k) \quad (k = 1,2,3), \tag{3.14c}$$

i.e.,

$$\overline{M} = \{(\mathbf{1}_2,\mathbf{1}_2), \overline{m}_1, \overline{m}_2, \overline{m}_3, \overline{m}_1\overline{m}_2, \overline{m}_2\overline{m}_1, \overline{m}_1\overline{m}_3, \overline{m}_3\overline{m}_1, \overline{m}_2\overline{m}_3, \overline{m}_3\overline{m}_2, \overline{m}_1\overline{m}_2\overline{m}_3, \overline{m}_3\overline{m}_2\overline{m}_1, (\overline{m}_1)^2, (\overline{m}_1)^3, (\overline{m}_2)^3, (\overline{m}_3)^3\}. \tag{3.14d}$$

However, we shall use a more convenient parametrization of $\overline{M}$,

$$\overline{M} = \{\overline{m}_1^{N_1}\overline{m}_2^{N_2}\overline{m}_3^{N_3}\overline{m}_4^{N_4} \,|\, \overline{m}_4 \equiv (-\mathbf{1}_2, -\mathbf{1}_2), \quad N_k = 0,1 \quad (k = 1,2,3,4)\}, \tag{3.14e}$$

which displays the fact that

$$\overline{M} \cong (\mathbf{Z}_2)^3 \otimes \mathbf{Z}_2 \cong M \otimes \mathbf{Z}_2. \tag{3.14f}$$

The double covering of M by $\overline{M}$ is given by

$$\overline{M}_k^\epsilon \to m_k^\epsilon, \quad k = 1,2,3,4,$$

$$m_1^\epsilon \equiv \mathrm{diag}(\epsilon, -\epsilon, -\epsilon, \epsilon), \quad m_2^\epsilon \equiv \mathrm{diag}(-\epsilon, \epsilon, -\epsilon, \epsilon), \tag{3.14g}$$

$$m_3^\epsilon \equiv \mathrm{diag}(-\epsilon, -\epsilon, \epsilon, \epsilon), \quad m_4^\epsilon \equiv \epsilon\mathbf{1}_4.$$

Note that $\overline{m}_k \to m_k$ [cf. (3.6)], $\overline{m}_4 \to \mathbf{1}_4$.

Analogously to the SL(4,R) case $\overline{P}_0 = \overline{M}AN$ is the *minimal parabolic subgroup* of $\overline{G}$. The number of standard parabolic subgroups is the same and they are explicitly given as follows [cf. (3.8) and (3.9)]:

$$\overline{P}_\varnothing = \overline{P}_0, \quad \overline{P}_\Psi = \overline{G} = \overline{M}_\Psi, \quad \overline{A}_\Psi = \overline{N}_\Psi = \{1\}, \tag{3.15a}$$

$$\overline{P}_{ab} = \overline{M}_{ab} A_{ab} N_{ab}, \quad a,b = 0,1,2,3, \quad a < b,$$

$$\overline{M}_{ab} \cong \overline{M} \times \mathrm{SO}(1,1) \cong M_{ab} \otimes \mathbf{Z}_2 \quad (a \neq 0), \tag{3.15b}$$

$$\overline{M}_{0a} \cong \mathrm{SO}(1,1) \times \{(\overline{M}/\overline{M}'_a) \otimes \overline{\mathrm{SL}(2,\mathbf{R})}\} \cong M_{0a} \otimes \mathbf{Z}_2,$$

$$\overline{M}'_1 \equiv \overline{M}_2^{+1}, \quad \overline{M}'_2 \equiv \overline{M}_3^{+1}, \quad \overline{M}'_3 \equiv \overline{M}_3^{-1},$$

where $\overline{\mathrm{SL}(2,\mathbf{R})}$ is a double (not the infinite universal) covering of SL(2,R). We shall explain this double covering at the end of Sec. IV A, after we introduce the Iwasawa decomposition.

Next we consider the group $G^e \equiv \mathrm{GL}(4,\mathbf{R})$. Analogously to the algebraic discussion in Sec. II F, we note that $K^e \cong O(4)$, $\widetilde{N}^e = \widetilde{N}$, $N^e = N$. Now instead of (3.2a), we have [cf. (2.44)–(2.46)]

$$\mathfrak{a}^e = \{s\hat{e}_1 + t\hat{e}_2 + r\hat{e}_3 + q\hat{e}_4 \,|\, s,t,r,q \in \mathbf{R}\}$$

$$= \left\{\mathrm{diag}\left(s + \frac{r}{2} + q, t - \frac{r}{2} + q, -s + \frac{r}{2} + q, -t - \frac{r}{2} + q\right)\right\}, \tag{3.16a}$$

and, instead of (3.3), we have

$$A^e \equiv \exp \mathfrak{a}^e = \left\{ a^e = e^q \begin{pmatrix} a_+ & 0 \\ 0 & a_- \end{pmatrix} \middle| \, s,t,r,q \in \mathbf{R} \right\} \cong \mathbf{R}^+ \times A. \tag{3.16b}$$

Next, instead of (3.6), we have, for the centralizer of A^e (or equivalently A) in K^e [cf. (3.6)],

$$M^e = \{m = \mathrm{diag}(\nu_1,\nu_2,\nu_3,\nu_4)\,|\,\nu_i = \pm 1, \quad 1\leqslant i\leqslant 4\}$$
$$= \{(m_1^e)^{N_1}\cdots(m_4^e)^{N_4}|m_k^e = m_k, \quad k=1,2,3;$$
$$m_4^e = (1,1,1,-1); \quad N_k = 0,1\} \cong M\times\mathbb{Z}_2. \tag{3.17}$$

Recall that each element k^e of K^e can be written in the form $k^e = (m_4^e)^{N_4}k$, where $k\in K$, $N_4 = 0$, if $k^e\in K$ and $N_4 = 1$ if $k^e\notin K$. Thus we can write

$$K^e \cong O(4) \cong (M^e/M) \otimes \mathrm{SO}(4) \cong \mathbb{Z}_2 \otimes K. \tag{3.18}$$

The *minimal parabolic subgroup* P_0^e of G^e is

$$P_0^e = M^eA^eN \cong \mathbb{R}^+\times\mathbb{Z}_2\times MAN \cong \mathbb{R}^+\times\mathbb{Z}_2\times P_0, \tag{3.19a}$$

and the other seven parabolic subgroups are [cf. (2.33) and (3.10a)]

$$P_\Psi^e = G^e = M_\Psi^e, \quad A_\Psi^e = N_\Psi^e = \{1\},$$
$$P_{ab}^e = M_{ab}^e A_{ab}^e N_{ab} \quad (a,b = 0,1,2,3, \quad a<b), \tag{3.19b}$$
$$A_{ab}^e = \exp(\mathfrak{a}_{ab}^e), \quad M_{ab}^e \cong \mathbb{Z}_2 \otimes M_{ab};$$

the M_{ab}^e are given by (3.10c) and (3.10d) with the change $M \to M^e$.

For the nonparabolic subalgebra analogous to A^k we have [cf. (2.58) and (3.11)]:

$$(A^k)^e = \exp((\mathfrak{a}^k)^e) = \{a^{ke} = e^q a^k\,|\,q\in\mathbb{R}, \ a^k\in A^k\}. \tag{3.20}$$

Finally we consider the group $\overline{G}^e \equiv \overline{\mathrm{GL}(4,\mathbb{R})}$. For the maximal compact subgroup $\overline{K^e}$ we use (3.18) and obtain

$$\overline{K^e} \cong \overline{O(4)} \cong \overline{\mathbb{Z}_2 \otimes K} = \mathbb{Z}_2 \otimes \overline{K}. \tag{3.21a}$$

For the simply connected subgroups we have $\overline{A^e} = A^e$, $\widetilde{N}^e = \widetilde{N}$, $\overline{N^e} = N$, and $\overline{(A^k)^e} = (A^k)^e$. The centralizer $\overline{M^e}$ of $\overline{A^e} = A^e$ (or equivalently A) in $\overline{K^e}$] obviously is

$$\overline{M^e} \cong \mathbb{Z}_2 \otimes \overline{M} \cong \mathbb{Z}_2 \otimes M \otimes \mathbb{Z}_2. \tag{3.21b}$$

For the parabolic subgroups we obtain $\overline{P}_0^e = \overline{M}^eA^eN \cong \mathbb{R}^+\times\mathbb{Z}_2\otimes\overline{P}_0$, $\overline{P_\Psi^e} = \overline{G^e} = \overline{M_\Psi^e}$, $\overline{A_\Psi^e} = \overline{N_\Psi^e} = \{1\}$, and $\overline{P_{ab}^e} = \overline{M_{ab}^e A_{ab}^e} N_{ab}$, where

$$\overline{M}_{ab}^e \cong \mathbb{Z}_2 \otimes \overline{M}_{ab} \cong \mathbb{Z}_2 \otimes M_{ab} \otimes \mathbb{Z}_2. \tag{3.21c}$$

The $\overline{M}_{ab}^e$ are given by (3.15b) with $\overline{M}$ replaced by $\overline{M}^e$.

Now we are ready to prove an analog of Proposition 1.

Proposition 2: The only cuspidal parabolic subgroup of $\overline{G} \equiv \overline{\mathrm{SL}(4,\mathbb{R})}$ [$G^e \equiv \mathrm{GL}(4,\mathbb{R})$, $\overline{G^e} \equiv \overline{\mathrm{GL}(4,\mathbb{R})}$] is the minimal parabolic subgroup $\overline{P}_0$ (P_0^e, $\overline{P_0^e}$, respectively) given above.

The *proof* follows the reasoning for $G = \mathrm{SL}(4,\mathbb{R})$. Indeed rank $\overline{G} = 3$, rank G^e = rank $\overline{G^e} = 4$, rank $\overline{K}$ = rank K^e = rank $\overline{K^e} = 2$. Thus $\overline{P}_\Psi, P_\Psi^e, \overline{P_\Psi^e}$ are not cuspidal. For the other parabolic subgroups the subgroup M always contains the factor SO(1,1) [cf. (3.15b), (3.19), and (3.21c)], which does not have discrete series representations. Thus $\overline{P}_{ab}$, P_{ab}^e, and $\overline{P_{ab}^e}$ are not cuspidal.

C. Explicit construction of the Weyl groups

Let M' be the normalizer of A in K, i.e.,

$$M' \equiv \{k\in K\,|\,k^{-1}ak\in A, \quad \forall a\in A\}. \tag{3.22}$$

It is known that M'/M is isomorphic to the Weyl group $W(\mathfrak{g},\mathfrak{a})$. Let ω' be the homomorphism from W to M' satisfying

$$\omega'(w)^{-1}\hat{e}_a\omega'(w) = w(\hat{e}_a), \quad a = 1,2,3, \quad \forall w\in W. \tag{3.23}$$

Of course it is enough to impose (3.23) for the generating elements of $W - w_1,w_2,w_3$. For these we obtain

$$\omega'(w_1) = \left\{ \left(\begin{array}{cc|cc} \alpha_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \beta_2 \\ \hline 0 & 0 & \delta_1 & 0 \\ 0 & \gamma_2 & 0 & 0 \end{array}\right); \alpha_1, \beta_2,\gamma_2,\delta_1 = \pm 1; \quad \alpha_1\beta_2\gamma_2\delta_1 = -1 \right\}, \tag{3.24a}$$

$$\omega'(w_2) = \left\{ \left(\begin{array}{cc|cc} \alpha_1 & 0 & \multicolumn{2}{c}{\multirow{2}{*}{0}} \\ 0 & \alpha_2 & & \\ \hline \multicolumn{2}{c|}{\multirow{2}{*}{0}} & 0 & \delta_+ \\ & & \delta_- & 0 \end{array}\right); \quad \alpha_1,\alpha_2,\delta_+,\delta_- = \pm 1; \quad \alpha_1\alpha_2\delta_+\delta_- = -1 \right\}, \tag{3.24b}$$

$$\omega'(w_3) = \left\{ \left(\begin{array}{cc|cc} 0 & \alpha_+ & \multicolumn{2}{c}{\multirow{2}{*}{0}} \\ \alpha_- & 0 & & \\ \hline \multicolumn{2}{c|}{\multirow{2}{*}{0}} & \delta_1 & 0 \\ & & 0 & \delta_2 \end{array}\right); \quad \alpha_+,\alpha_-,\delta_1,\delta_2 = \pm 1; \quad \alpha_+\alpha_-\delta_1\delta_2 = -1 \right\}. \tag{3.24c}$$

We note that each $\omega'(w_k)$ is isomorphic to M, as it should be as an element of M'. For the remaining elements of W we have [cf. (2.26)]

$$\omega'(w_{i_1\cdots i_n}) \equiv \omega'(w_{i_1}\cdots w_{i_n}) = \omega'(w_1)\cdots\omega'(w_{i_n}). \tag{3.25}$$

In the explicit construction of the intertwining operators it is convenient to work with fixed elements of M', obtaining thus explicit matrix representation of W. For this we should fix the parameters in (3.24). We shall usually work with the following particular choice, which gives the isomorphism between W and M'/M:

$$\omega(w_1)=\left(\begin{array}{cc|cc}1&0&0&0\\0&0&0&1\\\hline 0&0&-1&0\\0&1&0&0\end{array}\right), \tag{3.26a}$$

$$\omega(w_2)=\left(\begin{array}{cc|cc}1&0&&\\0&-1&\multicolumn{2}{c}{0}\\\hline &&0&1\\\multicolumn{2}{c|}{0}&1&0\end{array}\right), \tag{3.26b}$$

$$\omega(w_3)=\left(\begin{array}{cc|cc}0&1&&\\1&0&\multicolumn{2}{c}{0}\\\hline &&1&0\\\multicolumn{2}{c|}{0}&0&-1\end{array}\right). \tag{3.26c}$$

Analogously let $\overline{M'}$ be the normalizer of A in $\overline{K}$,

$$\overline{M'}\equiv\{\overline{k}=(u,v)\in\overline{K}\,|\,k^{-1}ak\in A,$$
$$k=k(\overline{k})=k(u,v),\quad \forall a\in A\}, \tag{3.27}$$

and note that $\overline{M'}/M=W(\mathfrak{g},\mathfrak{a})=M'/M$. Let $\overline{\omega}$ be the homomorphism from W to $\overline{M'}$, satisfying

$$k(\overline{\omega}(w))^{-1}\hat{e}_a k(\overline{\omega}(w))=w(\hat{e}_a)\quad (a=1,2,3),\quad \forall w\in W. \tag{3.28}$$

It is an easy (but tedious) calculation to find $\overline{\omega}$ from (3.28). However, we shall not reproduce it here since in the applications we need only the isomorphism between W and M'/M displayed in (3.26).

In the nonparabolic case (2.36) the representatives $\sigma(s)$ of the restricted Weyl reflection (2.41) must satisfy

$$\sigma(s)^{-1}D\sigma(s)=-D. \tag{3.29}$$

Thus we have a seven-parameter family of representatives,

$$\sigma(s)=\begin{pmatrix}0&\beta\\\gamma&0\end{pmatrix}\in G, \tag{3.30}$$

and the most frequent choices shall be

$$\sigma(s)=\begin{pmatrix}0&-\mathbf{1}_2\\\mathbf{1}_2&0\end{pmatrix}\quad\text{or}\quad \sigma(s)=\begin{pmatrix}0&\mathbf{1}_2\\\mathbf{1}_2&0\end{pmatrix}. \tag{3.31}$$

IV. IWASAWA AND BRUHAT DECOMPOSITIONS. HAAR MEASURES

A. The Iwasawa decomposition

It is well known that every element of a semisimple Lie group G may be represented uniquely as a product,[24] which we shall write, in the case $G=\mathrm{SL}(4,\mathbb{R})$, as

$$g=\begin{pmatrix}\alpha&\beta\\\gamma&\delta\end{pmatrix}=k(g)n_I(g)a_I(g) \tag{4.1}$$

where

$$k(g)=\begin{pmatrix}\alpha'&\beta'\\\gamma'&\delta'\end{pmatrix}\in K,$$
$$^t\alpha'\alpha'+{}^t\gamma'\gamma'=\mathbf{1}_2,\quad {}^t\beta'\beta'+{}^t\delta'\delta'=\mathbf{1}_2,$$
$$^t\alpha'\beta'+{}^t\gamma'\delta'=0$$

[cf. (3.2)], $n_I(g)\in N$, $a_I(g)\in A$ shall be parametrized as in (3.5b) [resp. (3.3)]. The explicit expressions of the parameters of k,n_I,a_I through the parameters of g is very involved and requires the consideration of many particular cases. Fortunately in the application to the elementary representations we shall need only the Iwasawa decomposition of the elements of the subgroup $\widetilde{N}$ [cf. (3.4b)]. In this case the explicit expressions are (we drop the primes on $\alpha',\beta',\gamma',\delta'$)

$$\alpha_1=1/\Delta_+,$$
$$\alpha_2=[1+x_1^2+x_+^2-y_+(x_1x_-+x_2x_+)]/\Delta\Delta_+, \tag{4.2a}$$
$$\alpha_+=[y_+(1+x_2^2+x_-^2)-x_1x_--x_2x_+]/\Delta\Delta_+,$$
$$\alpha_-=-y_+/\Delta_+,$$

$\delta_1,\delta_2,\delta_\pm$ as $\alpha_1,\alpha_2,\alpha_\pm$, respectively, with the changes

$$x_+\leftrightarrow x_-, \tag{4.2b}$$
$$\Delta_+\to\Delta_-,\quad y_+\to -y_-,$$
$$\Delta_\pm\equiv[1+x_1^2+x_\pm^2\mp 2y_\pm(x_1x_\mp+x_2x_\pm)$$
$$+y_\pm^2(1+x_2^2+x_\mp^2)]^{1/2},$$
$$\Delta\equiv[1+x_1^2+x_2^2+x_+^2$$
$$+x_-^2+(x_1x_2-x_+x_-)^2]^{1/2}, \tag{4.2c}$$
$$z_{I\pm}=[\pm(x_1x_\pm+x_2x_\mp)+y_\mp(1+x_2^2+x_\pm^2)]/\Delta, \tag{4.2d}$$

$$e^{s_I}=1/(\Delta_+\Delta_-)^{1/2},$$
$$e^{t_I}=(\Delta_+\Delta_-)^{1/2}/\Delta,\quad e^{r_I}=\Delta_-/\Delta_+, \tag{4.2e}$$
$$\beta=\hat{x}\delta,\quad \gamma=-{}^t\hat{x}\alpha,\quad \hat{b}_I=\delta^{-1}\,{}^t\hat{x}\alpha, \tag{4.2f}$$

where in (4.2f), α and δ are substituted from (4.2a) and (4.2b).

In the case $G^e=\mathrm{GL}(4,\mathbb{R})$ we have, for the analog of (4.1),

$$g^e=k^e(g^e)n_I(g^e)a_I^e(g^e), \tag{4.3a}$$

where

$$k^e=\begin{pmatrix}\alpha^e&\beta^e\\\gamma^e&\delta^e\end{pmatrix}\in K^e\cong O(4);$$
$$^t\alpha^e\alpha^e+{}^t\gamma^e\gamma^e=\mathbf{1}_2,\quad {}^t\beta^e\beta^e+{}^t\delta^e\delta^e=\mathbf{1}_2,$$
$$^t\alpha^e\beta^e+{}^t\gamma^e\delta^e=0,\quad \det k^e=\pm 1,$$

$n_I\in N$, $a_I^e\in A^e$ will be parametrized as in (3.5b), (3.16b), respectively. The expressions for the parameters of k^e,n,a^e through the parameters of g^e are as in the case of $\mathrm{SL}(4,\mathbb{R})$, the additional parameters q and $\det k^e$ being [cf. text after (3.17)]

$$e^q=|\det g^e|^{1/4},\quad \det k^e=\mathrm{sgn}(\det g^e)=(-1)^{1/4}. \tag{4.3b}$$

In particular, for the Iwasawa decomposition of $\tilde{n}$, formulas (4.2) hold without changes, while, instead of (4.3b), we have

$$e^q=|\det\tilde{n}|^{1/4}=1,\quad \det k^e=\mathrm{sgn}(\det\tilde{n})=1. \tag{4.3c}$$

Formulas (4.1) and (4.3) can be written globally also:

$$G=KNA,\quad G^e=K^eNA^e\cong\mathbb{R}^+\times(\mathbb{Z}_2\otimes KNA)$$
$$\cong\mathbb{R}^+\times(\mathbb{Z}_2\otimes G). \tag{4.4a}$$

Analogously we shall write in the cases $\overline{G}=\overline{\mathrm{SL}(4,\mathbb{R})}$ and

$$
\begin{aligned}
\overline{G}^{\,e} &= \overline{\mathrm{GL}(4,\mathbb{R})}\\
\overline{G} &= \overline{K}NA,\\
\overline{G}^{\,e} &= \overline{K}^{\,e}NA^{\,e} \qquad (4.4\mathrm{b})\\
&\cong \mathbb{R}^+\times(Z_2 \ltimes \overline{K}NA) \cong \mathbb{R}^+\times(Z_2 \ltimes \overline{G}).
\end{aligned}
$$

We shall write also the analogs of (4.1), (4.3)

$$\bar{g} = \bar{k}(\bar{g})n_I(g(\bar{g}))a_I(g(\bar{g})), \qquad (4.1')$$

$$\bar{g}^e = \bar{k}^{\,e}(\bar{g}^e)n_I(g^e(\bar{g}^e))a_I^e(g^e(\bar{g}^e)), \qquad (4.3')$$

where $g(\bar{g}),g^e(\bar{g}^e)$ is the image of $\bar{g},\bar{g}^e$, respectively, under the projection $\overline{G}\to G$, $\overline{G}^{\,e}\to G^{\,e}$, respectively. Formulas (4.1′), (4.3′) do not have a matrix representation in general; however, they have it for $\bar{g}$ (respectively $\bar{g}^e$) belonging to the 15-dimensional as $\overline{G}$ (resp. 16-dimensional as $\overline{G}^{\,e}$) manifold $\widetilde{N}NA$ (resp. $\widetilde{N}NA^{\,e}$) (the order of the factors is not essential). (The meaning of this will become clear in the next subsection.) In particular, $\bar{k}(\tilde{n}) = \bar{k}^{\,e}(\tilde{n}) = k(\tilde{n})$.

As we mentioned at the end of Sec. III A, we shall use the Iwasawa decomposition to give the exact imbedding of the group $G_e \equiv \mathrm{SO}_e(3,1)$ in $\mathrm{SL}(4,\mathbb{R})$. Namely let $G_e = K_eA_eN_e$ be the Iwasawa decomposition of G_e. Then we have [cf. (2.6), (2.8), (2.42), (3.5b)]:

$$K_e = \left\{ k_e \in K \,|\, k_e = \left(\begin{array}{ccc|c} & \hat{k}_e & & 0\\ & & & 0\\ & & & 0\\ \hline 0 & 0 & 0 & 1\end{array}\right),\quad \hat{k}_e \in \mathrm{SO}(3)\right\}, \qquad (4.5\mathrm{a})$$

$$A_e = \left\{ a_e = \exp hY_{34} = \left(\begin{array}{c|cc} \mathbf{1}_2 & \multicolumn{2}{c}{0}\\ \hline 0 & \cosh h & \sinh h\\ & \sinh h & \cosh h\end{array}\right),\quad h\in\mathbb{R}\right\}, \qquad (4.5\mathrm{b})$$

$$N_e = \{ n_e = \exp(b_1(Y_{14}+X_{13}) + b_2(Y_{24}+X_{23})) = \begin{pmatrix} \mathbf{1}_{12} & {}^t\underline{b} & {}^t\underline{b}\\ -\underline{b} & 1-\underline{b}^2/2 & -\underline{b}^2/2\\ \underline{b} & \underline{b}^2/2 & 1+\underline{b}^2/2\end{pmatrix},\quad \begin{array}{l}\underline{b}\equiv(b_1,b_2),\\ \underline{b}^2 \equiv b_1^2+b_2^2,\end{array} \qquad (4.5\mathrm{c})$$

$$\widetilde{N}_e = \{ \tilde{n}_e = \exp[x_1(Y_{14}-X_{13}) + x_2(Y_{24}-X_{23})] = \begin{pmatrix} \mathbf{1}_2 & -{}^t\underline{x} & {}^t\underline{x}\\ \underline{x} & 1-\underline{x}^2/2 & \underline{x}^2/2\\ \underline{x} & -\underline{x}^2/2 & 1+\underline{x}^2/2\end{pmatrix},\quad \begin{array}{l}\underline{x}\equiv(x_1,x_2),\\ \underline{x}^2 \equiv x_1^2+x_2^2.\end{array} \qquad (4.5\mathrm{d})$$

Note that $Y_{14}-X_{13}$, $Y_{24}-X_{23}$ and $Y_{14}+X_{13}$, $Y_{24}+X_{23}$ span the positive and negative, respectively, root spaces with respect to the algebra, spanned by Y_{34}. For more details on the structure of $\mathrm{SO}_e(3,1)$ [and $\mathrm{SO}_e(n,1)$], see Ref. 15.

As we mentioned after formula (3.15) we shall explain in more detail the double cover of $G' \equiv \mathrm{SL}(2,\mathbb{R})$. Let $G' = K'A'N'$ be the Iwasawa decomposition. Then $K' \cong \mathrm{SO}(2)$ and its double covering group is the group Spin 2, which is a one-parameter subgroup of the double covering $\overline{K} = \mathrm{SU}(2)\times\mathrm{SU}(2)$ of $K = \mathrm{SO}(4)$. For the three different imbeddings of $\overline{G}{}' = \overline{\mathrm{SL}(2,\mathbb{R})}$ in $\overline{G} = \overline{\mathrm{SL}(4,\mathbb{R})}$ in formula (3.15b) the group Spin 2 is the one that contains the set $\overline{M}{}'_a$ (which for that reason is factorized out from $\overline{M}$).

B. The (Gel'fand–Naimark-) Bruhat decomposition

We recall that almost every element of a semisimple Lie group may be written in a unique way as a product[26]

$$g = \tilde{n}(g)n_B(g)a_B(g)m(g)$$
$$(\tilde{n}\in\widetilde{N},\quad n_B\in N,\quad a_B\in A,\quad m\in M). \qquad (4.6)$$

The exact statement in the case of $G = \mathrm{SL}(4,\mathbb{R})$ is the following proposition.

Proposition 3: Let

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

and let $\kappa_2\delta_1 \det\delta \neq 0$ $(\kappa \equiv \alpha - \beta\delta^{-1}\gamma)$. Then formula (4.6) holds. Also let $\tilde{n},n_B,a_B,m$ be parametrized as in (3.3)–(3.6). Then these parameters are expressed through the parameters of g as follows:

$$e^{s_B} = \left|\frac{\det\kappa}{\delta_1\kappa_2}\right|^{1/2},\quad e^{t_B} = \left|\frac{\delta_1\kappa_2}{\det\delta}\right|^{1/2},$$

$$e^{r_B/2} = \left|\frac{\delta_1}{\kappa_2}\right|^{1/2}\left|\frac{\det\kappa}{\det\delta}\right|^{1/4}, \qquad (4.7\mathrm{a})$$

$$\hat{x} = \beta\delta^{-1},\quad y_+ = \frac{\kappa_+}{\kappa_2},\quad y_- = \frac{\delta_-}{\delta_1}, \qquad (4.7\mathrm{b})$$

$$\hat{b}_B = \left(1 - \frac{\delta_-}{\delta_1}\sigma_-\right)\gamma\kappa^{-1}\left(1 + \frac{\kappa_+}{\kappa_2}\sigma_+\right),$$

$$z_{B+} = \delta_+\delta_1/\det\delta,\quad z_{B-} = \kappa_-\kappa_2/\det\kappa; \qquad (4.7\mathrm{c})$$

$$\nu_1 = \mathrm{sgn}(\kappa_2 \det \kappa) = (-1)^{N_1},$$
$$\nu_2 = \mathrm{sgn}\, \kappa_2 = (-1)^{N_2},$$
$$\nu_3 = \mathrm{sgn}\, \delta_1 = (-1)^{N_3}, \qquad (4.7\mathrm{d})$$
$$\nu_4 = \mathrm{sgn}\, \delta_1 \det \delta = (-1)^{N_1+N_2+N_3}.$$

Proof: By straightforward matrix multiplication. (Note that $\det \kappa \neq 0$ always.)

When the condition $\kappa_2 \delta_1 \det \delta \neq 0$ is not fulfilled the decompositions of g are of the form

$$g = \omega(w)\tilde{n}^w nam, \quad w \in W(\mathfrak{g},\mathfrak{a}), \qquad (4.8)$$

where

$$\tilde{n}^w \in \tilde{N}^w \equiv \omega(w)^{-1}\tilde{N}\omega(w) \cap \tilde{N}. \qquad (4.9)$$

Denote

$$G^w \equiv \omega(w)\tilde{N}^w NAM,$$
$$d_{nw} \equiv \dim \tilde{N}^w, \quad d_w \equiv \dim G^w, \qquad (4.10)$$

$$d_w = d_{nw} + \dim(NAM) = d_{nw} + 9. \qquad (4.11)$$

Then we can state the following proposition.

Proposition 4: Let

$$g = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in G = \mathrm{SL}(4,\mathbb{R}).$$

Let $\kappa_2 \delta_1 \det \delta = 0$. Then there exists $w \in W(\mathfrak{g},\mathfrak{a}_0)$, such that formula (4.8) holds and g belongs to the lower-dimensional submanifold G^w. Further if $w_1, w_2 \in W(\mathfrak{g},\mathfrak{a})$, $w_1 \neq w_2$, then $\tilde{N}^{w_1} \neq \tilde{N}^{w_2}$, $G^{w_1} \neq G^{w_2}$. Let $w = w_{i_1} w_{i_2} \cdots w_{i_n}$ be given in the reduced form (2.26). Then $\dim \tilde{N}^w = 6 - n$.

We shall give only a sketch of the *proof.* First one determines $\tilde{N}^w$ for every w. For instance,

$$\tilde{N}^{w_k} = \tilde{N} \setminus \exp(\text{l.s. } E_k^+), \quad k = 1,2,3;$$
$$\tilde{N}^{w_5} = \exp(\text{l.s. } E_2^+, E_4^+, E_6^+),$$
$$\tilde{N}^{w_4} = \exp(\text{l.s. } E_3^+, E_5^+, E_6^+),$$
$$\tilde{N}^{w_6} = \exp(\text{l.s. } E_1^+), \quad \tilde{N}^{w_0} = \{1\}, \quad w_0 \equiv w_{131213},$$

etc. As a by-product of this we get $\dim \tilde{N}^w = 6 - n$. Then we build G^w explicitly and show by exhaustion that all cases, when $\kappa_2 \delta_1 \det \delta = 0$ holds, are accounted for.

In the case $G^e = \mathrm{GL}(4,\mathbb{R})$ we have the following proposition.

Proposition 5: Let

$$g^e = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in G^e.$$

Let also $\kappa_2 \delta_1 \det \delta \neq 0$. Then

$$g^e = \tilde{n} n a^e m^e \quad (\tilde{n} \in \tilde{N}, \ n \in N, \ a^e \in A^e, \ m \in M^e) \qquad (4.12)$$

holds. Let $\tilde{n}, n$ be parametrized as in (3.4)–(3.6) and a^e, m^e be parametrized as in (3.16) and (3.17). Then these parameters (except ν_4 and e^q) are expressed through the parameters of g^e as in formulas (4.7) and

$$\nu_4 = \mathrm{sgn}\, \delta_1 \det \delta = (-1)^{N_1+N_2+N_3+N_4},$$
$$e^q = |\det g^e|^{1/4}. \qquad (4.13)$$

Proof: By straightforward matrix multiplication. [Note $\nu_1 \nu_2 \nu_3 \nu_4 = \mathrm{sgn}(\det \kappa \cdot \det \delta) = \mathrm{sgn} \det g^e = (-1)^{N_4}$.]

When $\kappa_2 \delta_1 \det \delta = 0$ the analogs of (4.8) hold with $a \to a^e$, $m \to m^e$. Note that $(\tilde{N}^e)^w = \tilde{N}^w$,

$$(G^e)^w = \omega(w)\tilde{N}^w NA^e M^e, \qquad (4.14\mathrm{a})$$

$$d_w^e = \dim(G^e)^w = d_w + 1 = d_{nw} + 10. \qquad (4.14\mathrm{b})$$

Then Proposition 4 is true after the change $G \to G^e$, $G^w \to (G^w)^e$.

If we set $w = \mathrm{id} \in W(\mathfrak{g},\mathfrak{a})$ [cf. (2.26a)] in (4.10) and (4.14), then $\omega(\mathrm{id}) = \mathbf{1}_4$, $\tilde{N}^{\mathrm{id}} = \tilde{N}$, and

$$G^0 \equiv G^{\mathrm{id}} = \tilde{N}NAM, \quad G^{e0} \equiv (G^e)^{\mathrm{id}} = \tilde{N}NA^eM^e \qquad (4.15\mathrm{a})$$

are dense submanifolds of G, G^e, respectively. Analogously

$$\bar{G}^0 \equiv \tilde{N}NA\bar{M}, \quad \bar{G}^{e0} \equiv \tilde{N}NA^e\bar{M}^e \qquad (4.15\mathrm{b})$$

are dense submanifolds of $\bar{G}, \bar{G}^e$, respectively.

C. Haar measures

It is useful to note that almost every element of K can be decomposed in the form

$$k = k(\tilde{n}(k))m(k), \qquad (4.16)$$

where $k(\tilde{n})$ is from the Iwasawa decomposition of $\tilde{n} \in \tilde{N}$ [cf. (4.1), (4.2a), and (4.2b)], and $\tilde{n}(k)$ and $m(k)$ are from the Bruhat decomposition of k [cf. (3.2), (4.7b), and (4.7d)]. To prove (4.16) we use the Bruhat decomposition of $k(\tilde{n}(k))$. Note that (4.16) is the group structure parallel of the algebraic map (2.18).

Further we summarize some facts (see also Ref. 15) on the invariant measure on G and its subgroups. The group G is unimodular, its Haar measure is both left and right invariant; the measures of the nonunimodular factors are chosen to be left invariant.

The Haar measures on K, N, $\tilde{N}$, and A are given by

$$\int_K dk = 1, \qquad (4.17\mathrm{a})$$

$$dn = db_1\, db_2\, db_+\, db_-\, dz_+\, dz_-,$$
$$d\tilde{n} = dx_1\, dx_2\, dx_+\, dx_-\, dy_+\, dy_-, \qquad (4.17\mathrm{b})$$
$$da = ds\, dt\, dr.$$

Then the Haar measure on G has the form

$$dg = dk\, dn\, da. \qquad (4.18)$$

Using (4.16) we can express dk in terms of $d\tilde{n}$:

$$dk = \frac{1}{2\pi^4} e^{2\rho(\log a_I(\tilde{n}(k)))} d\tilde{n}(k) = \frac{1}{2\pi^4} \frac{d\tilde{n}}{\Delta_+^2 \Delta_-^2 \Delta^2}, \qquad (4.19)$$

where log is the inverse map of exp: $\mathfrak{a} \to A$,

$$\log: \ A \to \mathfrak{a}; \qquad (4.20)$$

ρ is half the sum of the positive roots [cf. (2.14)],

$$\rho = \frac{1}{2} \sum_{k=1}^{6} \lambda_k, \quad \rho(\hat{e}_1,\hat{e}_2,\hat{e}_3) = (3,1,0); \qquad (4.21)$$

and the normalization is fixed to satisfy (4.17a).

For the Haar measures on $\bar{G}$, G^e, $\bar{G}^e$, we have, respectively,

$$d\bar{g} = d\bar{k}\, dn\, da, \quad d\bar{k} = \tfrac{1}{2}dk, \quad \int_{\bar{K}} d\bar{k} = 1; \tag{4.22}$$

$$dg^e = dk^e\, dn\, da^e, \quad dk^e = \tfrac{1}{2}dk, \quad da^e = da\, dq,$$

$$\int_{K^e} dk^e = 1; \tag{4.23}$$

$$d\bar{g}^e = d\bar{k}^e\, dn\, da^e, \quad d\bar{k}^e = \tfrac{1}{4}dk, \quad \int_{\bar{K}^e} d\bar{k}^e = 1. \tag{4.24}$$

V. THE ELEMENTARY REPRESENTATIONS OF SL (4 R) AND GL (4, R) AND THEIR COVERING GROUPS

A. The elementary representations of SL (4, R)

As we establish in Sec. III A (Proposition 1) the only cuspidal parabolic subgroup of $G = \mathrm{SL}(4,\mathbb{R})$ is the minimal parabolic subgroup $P_0 = MAN$. Thus by the results of Langlands[8]–Knapp–Zuckerman[9] the elementary representations will be P_0-induced representations of G. They will be parametrized by

$$\chi \equiv [c_1,c_2,c_3;\epsilon_1,\epsilon_2,\epsilon_3], \tag{5.1}$$

where $c_k \in \mathbb{C}$ $(k = 1,2,3)$, will characterize the representations of A, and are the values of a linear functional λ for the basis elements of $\mathfrak{a}$ [cf. (2.9)]:

$$\lambda(\hat{e}_1,\hat{e}_2,\hat{e}_3) = (c_1,c_2,c_3), \tag{5.2}$$

where $\epsilon_k = 0,1$ $(k = 1,2,3)$, index the characters of M.

Let D^χ be the one-dimensional representation of MA given by [cf. (3.3), and (3.6)]

$$D^\chi(ma) \equiv (-1)^{\epsilon_1 N_1 + \epsilon_2 N_2 + \epsilon_3 N_3} e^{-s(3+c_1) - t(1+c_2) - (r/2)c_3}. \tag{5.3}$$

(Here, as usual, we have added to the linear functional λ the half sum of the positive roots ρ [cf. (4.21)].) Now we are ready to introduce the *representation space* for the elementary representations of G:

$$\mathscr{C}_\chi \equiv \{\mathscr{f} \in C^\infty(G,\mathbb{C}) \mid \mathscr{f}(gman) = D^\chi(ma)^{-1}\mathscr{f}(g), \quad g\in G,\ m\in M,\ a\in A,\ n\in N\}. \tag{5.4}$$

The *elementary representation* (ER) $\mathscr{T}^\chi$ induced by the representation D^χ of $P_0 = MAN$ (N is represented trivially) is given by the left regular action of G

$$(\mathscr{T}^\chi(g)\mathscr{f})(g') \equiv \mathscr{f}(g^{-1}g'), \quad g,g' \in G, \quad \mathscr{f} \in \mathscr{C}_\chi. \tag{5.5}$$

We shall write down some useful properties of the ER following Ref. 27. We introduce the standard right action of $\mathfrak{g}^{\mathbb{C}}$ in $C^\infty(G,\mathbb{C})$ (see Ref. 28):

$$(X\cdot\mathscr{f})(g) \equiv \frac{d}{dt}\mathscr{f}(g \exp tX)|_{t=0}, \quad X \in \mathfrak{g}^{\mathbb{C}}. \tag{5.6}$$

It is easy to see that [cf. (2.15) and (2.20)]

$$E_k^- \cdot \mathscr{f} = 0, \quad k = 1,\dots,6; \tag{5.7a}$$

$$\hat{e}_1\cdot\mathscr{f} = (3+c_1)\mathscr{f}, \quad \hat{e}_2\cdot\mathscr{f} = (1+c_2)\mathscr{f}, \quad \hat{e}_3\cdot\mathscr{f} = (c_3/2)\mathscr{f}. \tag{5.7b}$$

We notice that (5.7) is equivalent to the covariance property in (5.4) restricted to AN; since M is a discrete subgroup the covariance property with respect to it cannot be translated into algebraic information. We also notice that every element $\mathscr{f} \in \mathscr{C}_\chi$ may play the role of a lowest weight vector of a lowest weight module over $\mathfrak{g}^{\mathbb{C}}$. Indeed $\mathscr{f}$ is annihilated by $n^{\mathbb{C}}$ and the Cartan subalgebra $\mathfrak{h}^{\mathbb{C}}$ (of which the $\hat{e}_k$ form a basis) acts on $\mathscr{f}$ by scalars. This becomes more apparent if we rewrite (4.7b) in the standard form[29]

$$\hat{e}_k\cdot\mathscr{f} = (\lambda+\rho)(\hat{e}_k)\mathscr{f}, \quad k = 1,2,3. \tag{5.7c}$$

Properties (5.7) will be used in a sequel of this paper for the construction of the invariant differential operators between reducible elementary representations (see also Ref. 27).

We introduce a K-invariant scalar product in $\mathscr{C}_\chi$ by

$$(\mathscr{f}_1,\mathscr{f}_2)_{\mathscr{C}_\chi} \equiv 2\pi^4 \int_K dk\ \overline{\mathscr{f}_1(k)}\mathscr{f}_2(k) \tag{5.8}$$

[cf. (4.17a)]; the factor $2\pi^4$ is introduced for convenience. The representation $\mathscr{T}^\chi$ is continuous with respect to the topology defined by (5.8). For some other properties of $\mathscr{T}^\chi$ see the end of the next subsection.

B. Noncompact picture of the elementary representations

Here we introduce the so-called noncompact picture of the elementary representations. [There is also a compact picture that we shall not consider here (cf. Refs. 15 and 17).] The representation space in the noncompact picture C_χ consists of C^∞-functions over the subgroup $\tilde{N}$ [cf. (3.4)]. These functions have special asymptotic behavior, namely,

$$C_\chi \equiv \Bigg\{ f \in C^\infty(\tilde{N},\mathbb{C});\ |\hat{x}| \equiv (x_1^2 + x_2^2 + x_+^2 + x_-^2)^{1/2};$$

$$f(\hat{x},y_\pm) \underset{\substack{|x_{1,2}|\sim|x_\pm|\sim|\hat{x}|\to\infty \\ |\det\hat{x}|/|\hat{x}|<\infty}}{\approx} \frac{(\operatorname{sgn} x_2)^{\epsilon_2}}{|\hat{x}|^{1+c_2}} \sum_{k=0}^\infty \varphi^\chi_{1k}(w_1\hat{x},y_\pm); \tag{5.9a}$$

$$f(\hat{x},y_\pm) \underset{|y_-|\to\infty}{\approx} \frac{(\operatorname{sgn} y_-)^{\epsilon_3}}{|y_-|^{1+(1/2)(c_1-c_2-c_3)}} \sum_{k=0}^\infty \varphi^\chi_{2k}(\hat{x},y_+)\left(\frac{1}{y_-}\right)^k; \tag{5.9b}$$

$$f(\hat{x},y_\pm) \underset{|y_+|\to\infty}{\approx} \frac{(\operatorname{sgn} y_+)^{\epsilon_1+\epsilon_2}}{|y_+|^{1+(1/2)(c_1-c_2+c_3)}} \sum_{k=0}^\infty \varphi^\chi_{3k}(\hat{x},y_-)\left(\frac{1}{y_+}\right)^k \Bigg\}. \tag{5.9c}$$

The functions φ^{χ}_{1k} are homogeneous polynomials of degree k in the first variable and C^{∞}-functions with respect to $y_{\pm}$; the functions φ^{χ}_{2k}, φ^{χ}_{3k} are C^{∞}-functions of their arguments. The action of the generating elements of the Weyl group $W(\mathfrak{g},\mathfrak{a})$ [cf. (2.24)–(2.26)] on $\tilde{N}$, i.e., $w_k\hat{x}$, $w_k y_{\pm}$, is obtained by the Bruhat decomposition [(4.6) and (4.7)] of $\omega(w_k)^{-1}\tilde{n}(\hat{x},y_{\pm})$, where $\omega(w_k)$ are the matrix representatives of w_k in (3.26):

$$\omega(w_k)^{-1}\tilde{n}(\hat{x},y_{\pm}) = \tilde{n}(w_k\hat{x},\omega_k y_{\pm})n(\tilde{n},\omega)^{-1}a(\tilde{n},\omega)^{-1}m(\tilde{n},\omega)^{-1}. \tag{5.10}$$

Explicitly we have

$$w_1\hat{x} = \frac{1}{x_2}\begin{vmatrix} -\det\hat{x} & x_+ \\ x_- & 1\end{vmatrix},$$

$$w_1 y_{\pm} = \pm x_{\pm} - x_2 y_{\pm}, \quad x_2 \neq 0; \tag{5.11a}$$

$$w_2\hat{x} = \begin{vmatrix} x_+ & x_1 \\ -x_2 & -x_-\end{vmatrix},$$

$$w_2 y_+ = -y_+, \quad w_2 y_- = 1/y_-, \quad y_- \neq 0; \tag{5.11b}$$

$$w_3\hat{x} = \begin{vmatrix} x_- & -x_2 \\ x_1 & -x_+\end{vmatrix},$$

$$w_3 y_+ = 1/y_+, \quad w_3 y_- = -y_-, \quad y_+ \neq 0. \tag{5.11c}$$

The elementary representation in the noncompact picture is defined in the space C_{χ} by

$$(T^{\chi}(g)f)(\hat{x},y_{\pm}) \equiv D^{\chi}(m_g a_g)f(\hat{x}_g,y_{g\pm}), \tag{5.12a}$$

where x_g, $y_{g\pm}$, m_g, and a_g are obtained from the Bruhat decomposition

$$g^{-1}\tilde{n}(\hat{x},y_{\pm}) = \tilde{n}(\hat{x}_g,y_{g\pm})n_g^{-1}a_g^{-1}m_g^{-1}. \tag{5.12b}$$

In the cases when (5.12b) does not exist, formula (5.12a) will be defined by the appropriate limit as we shall explain below.

Explicitly we obtain in the generic cases (a) $g = \tilde{n}' \in \tilde{N}$ ("translations" of $\tilde{N}$),

$$(T^{\chi}(\tilde{n}')f)(\hat{x},y_{\pm}) = f(\hat{x}'',y''_{\pm}), \tag{5.13a}$$

$$\hat{x}'' = (1 - y'_+\sigma_+)(\hat{x} - \hat{x}')(1 + y'_-\sigma_-),$$

$$y''_{\pm} = y_{\pm} - y'_{\pm};$$

(b) $g = m \in M$ (reflections of $\tilde{N}$),

$$(T^{\chi}(m)f)(\hat{x},y_{\pm}) = (-1)^{\Sigma_k \epsilon_k N_k} f(\hat{x}',y'_{\pm}), \tag{5.13b}$$

$$x'_{1,2} = (-1)^{N_1+N_3}x_{1,2}, \quad x'_{\pm} = (-1)^{N_2+N_3}x_{\pm},$$

$$y'_{\pm} = (-1)^{N_1+N_2}y_{\pm};$$

(c) $g = a \in A$ (dilatations of $\tilde{N}$),

$$(T^{\chi}(a)f)(\hat{x},y_{\pm}) = e^{-s(3+c_1)-t(1+c_2)-(r/2)c_3} f(\hat{x}',y'_{\pm}), \tag{5.13c}$$

$$x'_1 = x_1 e^{-2s}, \quad x'_2 = x_2 e^{-2t},$$

$$x'_{\pm} = x_{\pm}e^{-s-t\mp r}, \quad y'_{\pm} = y_{\pm}e^{-s+t\mp r};$$

(d) $g = \omega(w_1)$ (Weyl inversion of $\hat{x}$),

$$(T^{\chi}(\omega(w_1))f)(\hat{x},y_{\pm})\big|_{x_2\neq 0} = \frac{(-1)^{\epsilon_3}(\operatorname{sgn} -x_2)^{\epsilon_2}}{|x_2|^{1+c_2}} f(w_1\hat{x},w_1 y_{\pm}), \tag{5.13d}$$

$$(T^{\chi}(\omega(w_1))f)(\hat{x},y_{\pm})\big|_{x_2=0} \equiv \frac{(-1)^{\epsilon_2+\epsilon_3}}{|1 + x_+^2 + x_-^2 + x_+x_-^2|^{1+c_2}} \times \sum_{k=0}^{\infty} \varphi^{\chi}_{1k}(\hat{x}(x_2=0), \pm x_{\pm}); \tag{5.13d$'$}$$

(e) $g = \omega(w_2)$ (Weyl inversion of y_-),

$$(T^{\chi}(\omega(w_2))f)(\hat{x},y_{\pm})\big|_{y_-\neq 0} = \frac{(-1)^{\epsilon_2}(\operatorname{sgn} y_-)^{\epsilon_3}}{|y_-|^{1+(1/2)(c_1-c_2-c_3)}} f(w_2\hat{x},w_2 y_{\pm}), \tag{5.13e}$$

$$T^{\chi}(\omega(w_2))f(\hat{x},y_{\pm})\big|_{y_-=0} \equiv (-1)^{\epsilon_2}\varphi^{\chi}_{2,0}(w_2\hat{x},-y_+); \tag{5.13e$'$}$$

and (f) $g = \omega(w_3)$ (Weyl inversion of y_+),

$$(T^{\chi}(\omega(w_3))f)(\hat{x},y_{\pm})\big|_{y_+\neq 0} = \frac{(-1)^{\epsilon_1}(\operatorname{sgn} y_+)^{\epsilon_1+\epsilon_2}}{|y_+|^{1+(1/2)(c_1-c_2+c_3)}} f(w_3\hat{x},w_3 y_{\pm}), \tag{5.13f}$$

$$(T^{\chi}(\omega(w_3))f)(\hat{x},y_{\pm})\big|_{y_+=0} \equiv (-1)^{\epsilon_1}\varphi^{\chi}_{3,0}(w_3\hat{x},-y_-). \tag{5.13f$'$}$$

Obviously formulas (5.13d′), (5.13e′), and (5.13f′) are supplementing (5.12a). Now we can prove the following proposition.

Proposition 6: The C^{∞} action of the elementary representation is provided by formulas (5.13) together with the asymptotic conditions (5.9).

Proof: First we notice that the asymptotic conditions (5.9a), (5.9b), and (5.9c), respectively, ensure the C^{∞} passage from formulas (5.13d), (5.13e), and (5.13f) to (5.13d′), (5.13e′), and (5.13f′), respectively, in the limit $x_2 \to 0$, $y_- \to 0$, $y_+ \to 0$, respectively. Further we recall (cf. Propositions 3 and 4) that every element of G can be uniquely decomposed in the form $g = \tilde{n}nam$ [cf. (4.6)] or in the form $g = \omega(w)\tilde{n}^w n'a'm'$ [cf. (4.8)]. Now every element $n \in N$ can be obtained from some element of $\tilde{N}$ by Weyl conjugation. Explicitly let $n(b_1) \in N$ denote an element in the parametrization (3.5b) with all parameters except b_1 being set to zero; analogously for $n(b_2)$, $n(b_{\pm})$, $n(z_{\pm})$, and for the elements $\tilde{n} \in \tilde{N}$ in the parametrization (3.4b). Then we have

$$\begin{aligned}
n(b_1) &= \omega(w_6)^{-1}\tilde{n}(x_1 = b_1)\omega(w_6),\\
n(b_2) &= \omega(w_1)^{-1}\tilde{n}(x_2 = b_2)\omega(w_1),\\
n(b_+) &= \omega(w_4)^{-1}\tilde{n}(x_- = b_+)\omega(w_4),\\
n(b_-) &= \omega(w_5)^{-1}\tilde{n}(x_+ = b_-)\omega(w_5),\\
n(z_+) &= \omega(w_2)^{-1}\tilde{n}(y_- = z_+)\omega(w_2),\\
n(z_-) &= \omega(w_3)^{-1}\tilde{n}(y_+ = z_-)\omega(w_3),
\end{aligned} \tag{5.14}$$

where w_4, w_5, w_6 are given in (2.25) through w_1, w_2, w_3 [cf. also (3.25) and (3.26)]. Thus T^{χ} is C^{∞} defined for every element of G that concludes the proof.

Further we introduce a K-invariant scalar product in C_{χ} by

$$(f_1,f_2)_{C_{\chi}} \equiv \int_{\tilde{N}} \overline{f_1(\hat{x},y_{\pm})}\, f_2(\hat{x},y_{\pm})\mu^{\chi}(\hat{x},y_{\pm})\,d\tilde{n}, \tag{5.15a}$$

$$\mu^\chi(\tilde{n}) \equiv \mu^\chi(\hat{x}, y_\pm)$$
$$\equiv (\Delta_+)^{1/2(c_1+\bar{c}_1-c_2-\bar{c}_2+c_3+\bar{c}_3)}$$
$$\times(\Delta_-)^{1/2(c_1+\bar{c}_1-c_2-\bar{c}_2-c_3-\bar{c}_3)}\Delta^{c_2+\bar{c}_2}, \qquad (5.15\text{b})$$

and $\Delta_\pm$, Δ are given in (4.2c). The representation T^χ is continuous with respect to the topology defined by (5.15). On general grounds[23] we know that the representations $\mathcal{T}^\chi$ and T^χ should be equivalent. In more detail we have, in our case, the following proposition.

Proposition 7: The equivalence of the representations $\mathcal{T}^\chi$, formula (5.5), and T^χ, formulas (5.12) and (5.13), is given by the operator B and its inverse, which are defined as follows:

$$B: \mathscr{C}_\chi \to C_\chi, \quad (B\not{f})(\hat{x}, y_\pm) \equiv \not{f}(\tilde{n}(\hat{x}, y_\pm)); \qquad (5.16)$$

$$(B^{-1}f)(g) \equiv D^\chi(ma)^{-1} f(\hat{x}, y_\pm), \quad g = \tilde{n}(\hat{x}, y_\pm)nam, \qquad (5.17\text{a})$$

$$(B^{-1}f)(g) \equiv D^\chi(ma)^{-1} \lim (T^\chi(\omega(w))f)(\hat{x}, y_\pm),$$
$$g = \omega(w)\tilde{n}(\hat{x}^w, y_\pm^w)nam, \quad \tilde{n} \to \tilde{n}^w \equiv \tilde{n}(\hat{x}^w, y_\pm^w). \qquad (5.17\text{b})$$

The operator B is isometric.

Proof: We must show that

$$B\mathcal{T}^\chi = T^\chi B, \quad B^{-1}T^\chi = \mathcal{T}^\chi B^{-1}, \qquad (5.18)$$

which is straightforward and requires some care only when (5.17b) is involved (its analog is missing in Ref. 23). [The interested reader may find explicit expressions for the asymptotic functions φ^χ_{ak} in (5.9) in terms of limits of functions from $\mathscr{C}_\chi$.] For the isometricity proof we shall show

$$(\not{f}_1, \not{f}_2)_{\mathscr{C}_\chi} = (B\not{f}_1, B\not{f}_2)_{C_\chi}. \qquad (5.19\text{a})$$

Indeed using (4.16), (5.4), (4.19), and (4.2) we have

$$(\not{f}_1, \not{f}_2)_{\mathscr{C}_\chi}$$
$$= 2\pi^4 \int_k \overline{\not{f}_1(k)}\not{f}_2(k)dk$$
$$= 2\pi^4 \int_k \overline{\not{f}_1(k(\tilde{n}(k))m(k))}\not{f}_2(k(\tilde{n}(k))m(k))dk$$
$$= 2\pi^4 \int_k \overline{\not{f}(k(\tilde{n}))}\,\not{f}_2(k(\tilde{n}))dk(\tilde{n})$$
$$= \int_{\tilde{N}} \overline{\not{f}_1(k(\tilde{n}))}\not{f}_2(k(\tilde{n}))\frac{d\tilde{n}}{\Delta_+^2\,\Delta_-^2\,\Delta^2}$$
$$= \int_{\tilde{N}} \overline{\not{f}_1(k(\tilde{n})n(\tilde{n})a(\tilde{n}))}$$
$$\times \not{f}_2(k(\tilde{n})n(\tilde{n})a(\tilde{n}))\mu^\chi(\tilde{n})d\tilde{n}$$
$$= \int_{\tilde{N}} \overline{\not{f}_1(\tilde{n})}\not{f}_2(\tilde{n})\mu^\chi(\tilde{n})d\tilde{n} = (B\not{f}_1, B\not{f}_2)_{C_\chi}. \qquad (5.19\text{b})$$

Further we recall the general fact that $\mathscr{C}_\chi$ and C_χ are not complete as normed linear spaces with respect to the topology defined by their scalar products. [They are complete with respect to some Fréchet space topology that we shall not introduce here (cf. Ref. 22)] Their completion with respect to the scalar products (5.8) and (5.15), respectively, will be denoted by $\mathscr{H}_\chi$ and H_χ, respectively.

Finally we note that both scalar products are also G-invariant if and only if $c_k = i\tau_k$, $\tau_k \in \mathbb{R}$, $k = 1,2,3$. [Note that then $\mu^\chi(\tilde{n}) = 1$.] These elementary representations form the *principal series of unitary representations* of $SL(4,\mathbb{R})$ acting in the spaces $\mathscr{H}_\chi$ or H_χ (see Refs. 9 and 12).

C. Infinitesimal generators and Casimir operators of the elementary representations

We write down the expressions for the infinitesimal generators in the noncompact picture for the generic (continuous) subgroups $\tilde{N}, A, N$ using formulas (5.13) and (5.14).

(a) The subgroup $\tilde{N}$: let $k = 1,2,+,-$, $e_\pm \equiv \sigma_\pm$, then we have

$$T_k f(\hat{x}, y_\pm)$$
$$\equiv \frac{\partial}{\partial x'_k}(T^\chi(\tilde{n}x'_k))f)(\hat{x}, y_\pm)\big|_{x'_k=0}$$
$$= -\frac{\partial}{\partial x_k} f(\hat{x}, y_\pm); \qquad (5.20\text{a})$$

$$T_{y_\pm} f(\hat{x}, y_\pm)$$
$$\equiv \frac{\partial}{\partial y'_\pm}(T^\chi(\tilde{n}y'_\pm))f)(\hat{x}, y_\pm)\big|_{y'_\pm=0}$$
$$= \left(\mp x_\pm \frac{\partial}{\partial x_1} \mp x_2 \frac{\partial}{\partial x_+} - \frac{\partial}{\partial y_\pm}\right) f(\hat{x}, y_\pm). \qquad (5.20\text{b})$$

In (5.20) we have used the notational convention of (5.14) to write in the arguments of $\tilde{n}(\cdot)$ only the nonzero parameters. This convention will be used also below for $n(\cdot)$ as in (5.14) and for $a(\cdot) \in A$.

(b) The subgroup A:

$$D_1 f(\hat{x}, y_\pm) \equiv \frac{\partial}{\partial s}(T^\chi(a(s))f)(\hat{x}, y_\pm)\big|_{s=0}$$
$$= -\left[3 + c_1 + x_+\frac{\partial}{\partial x_+} + x_-\frac{\partial}{\partial x_-} + 2x_1\frac{\partial}{\partial x_1} + y_+\frac{\partial}{\partial y_+} + y_-\frac{\partial}{\partial y_-}\right] f(\hat{x}, y_\pm); \qquad (5.21\text{a})$$

$$D_2 f(\hat{x}, y_\pm) \equiv \frac{\partial}{\partial t}(T^\chi(a(t))f)(\hat{x}, y_\pm)\big|_{t=0}$$
$$= -\left[1 + c_2 + x_+\frac{\partial}{\partial x_+} + x_-\frac{\partial}{\partial x_-} + 2x_2\frac{\partial}{\partial x_2} - y_+\frac{\partial}{\partial y_+} - y_-\frac{\partial}{\partial y_-}\right] f(\hat{x}, y_\pm); \qquad (5.21\text{b})$$

$$D_3 f(\hat{x}, y_\pm) \equiv \frac{\partial}{\partial r}(T^\chi(a(r))f)(x, y_\pm)\big|_{r=0}$$
$$= -\left[\frac{c_3}{2} + x_+\frac{\partial}{\partial x_+} - x_-\frac{\partial}{\partial x_-} + y_+\frac{\partial}{\partial y_+} - y_-\frac{\partial}{\partial y_-}\right] f(\hat{x}, y_\pm). \qquad (5.21\text{c})$$

(c) The subgroup N: defining

$$C_k f(\hat{x},y_\pm) \equiv \frac{\partial}{\partial b_k}(T^\chi(n(b_k))f)(\hat{x},y_\pm)|_{b_k=0},$$

$$k=1,2,+,-, \tag{5.22}$$

$$C_{z_\pm} f(\hat{x},y_\pm) \equiv \frac{\partial}{\partial z_\pm}(T^\chi(n(z_\pm))f)(\hat{x},y_\pm)\,|_{z_\pm=0},$$

we obtain, using (5.13) and (5.14),

$$\begin{aligned} C_1 = {} & x_1^2\frac{\partial}{\partial x_1} + x_1x_+\frac{\partial}{\partial x_+} + x_1x_-\frac{\partial}{\partial x_-} + x_+x_-\frac{\partial}{\partial x_2} \\ & + (x_1 + x_+y_-)y_-\frac{\partial}{\partial y_-} + (x_1 - x_-y_+)y_+\frac{\partial}{\partial y_+} \\ & + (1 + \tfrac{1}{2}(c_1 - c_2 + c_3))(x_1 - x_-y_+) \\ & + (1 + \tfrac{1}{2}(c_1 - c_2 - c_3))(x_1 + x_+y_-) + (1 + c_2)x_1; \end{aligned} \tag{5.23a}$$

$$\begin{aligned} C_2 = {} & x_+x_-\frac{\partial}{\partial x_1} + x_2x_+\frac{\partial}{\partial x_+} + x_2x_-\frac{\partial}{\partial x_-} \\ & + x_2^2\frac{\partial}{\partial x_2} + (x_+ - y_+x_2)\frac{\partial}{\partial y_+} \\ & - (x_- + y_-x_2)\frac{\partial}{\partial y_-} + (1 + c_2)x_2; \end{aligned} \tag{5.23b}$$

$$\begin{aligned} C_+ = {} & x_-x_1\frac{\partial}{\partial x_1} + x_1x_2\frac{\partial}{\partial x_+} + x_-^2\frac{\partial}{\partial x_-} + x_-x_2\frac{\partial}{\partial x_2} \\ & + (x_- + x_2y_-)y_-\frac{\partial}{\partial y_-} + (x_1 - x_-y_+)\frac{\partial}{\partial y_+} \\ & + (2 + \tfrac{1}{2}(c_1 + c_2 - c_3))x_- \\ & + (1 + \tfrac{1}{2}(c_1 - c_2 - c_3))x_2y_-; \end{aligned} \tag{5.23c}$$

$$\begin{aligned} C_- = {} & x_+x_1\frac{\partial}{\partial x_1} + x_+^2\frac{\partial}{\partial x_+} + x_1x_2\frac{\partial}{\partial x_-} + x_+x_2\frac{\partial}{\partial x_2} \\ & + (x_+ - x_2y_+)y_+\frac{\partial}{\partial y_+} - (x_1 + x_+y_-)\frac{\partial}{\partial y_-} \\ & + (2 + \tfrac{1}{2}(c_1 + c_2 + c_3))x_+ \\ & - (1 + \tfrac{1}{2}(c_1 - c_2 + c_3))x_2y_+; \end{aligned} \tag{5.23d}$$

$$\begin{aligned} C_{z_+} = {} & x_1\frac{\partial}{\partial x_+} + x_-\frac{\partial}{\partial x_2} \\ & + y_-^2\frac{\partial}{\partial y_-} + (1 + \tfrac{1}{2}(c_1 - c_2 - c_3))y_-; \end{aligned} \tag{5.23e}$$

$$\begin{aligned} C_{z_-} = {} & -x_1\frac{\partial}{\partial x_-} - x_+\frac{\partial}{\partial x_2} \\ & + y_+^2\frac{\partial}{\partial y_+} + (1 + \tfrac{1}{2}(c_1 - c_2 + c_3))y_+. \end{aligned} \tag{5.23f}$$

It is well known that the elementary representations are operator irreducible in the sense of Schur's lemma. Thus the Casimir operators are multiples of the unit operator. In particular the second-order Casimir operator

$$\begin{aligned} \mathscr{C}_2(\chi) = {} & \tfrac{1}{2}(D_1^2 + D_2^2) + D_3^2 \\ & + [T,C_1]_+ + [T_2,C_2]_+ + [T_+,C_-]_+ \\ & + [T_-,C_+]_+ + [T_{y_+},C_{z_-}]_+ + [T_{y_-},C_{z_+}]_+ \\ = {} & \tfrac{1}{2}(D_1^2 + D_2^2) + D_3^2 - 3D_1 - D_2 \\ & + 2(T_1C_1 + T_2C_2 + T_+C_- + T_-C_+ + T_{y_+} \\ & \times C_{z_-} + T_{y_-}C_{z_+}) \end{aligned} \tag{5.24b}$$

has the value

$$\mathscr{C}_2(\chi) = \tfrac{1}{2}(c_1^2 + c_2^2 + c_3^2/2) - 5. \tag{5.25}$$

Formula (5.25) is derived by direct substitution of the generators from (5.21)–(5.23). Another way is to use the lowest weight module structure shown in (5.7). We exploit the fact that the Casimir operator has the same expression (5.24) every generator being replaced by its right-acting counterpart according to (5.6). Then in (5.24b) only the terms with D_k (replaced by $\hat{e}_k$) remain since [by (5.7a)] the other terms annihilate the functions of $\mathscr{C}_\chi$; then if we use (5.7b) we obtain (5.25) at once. (Note that the constant terms in D_k coincide with the action of $-\hat{e}_k$.) For a more general discussion we refer to Ref. 27.

Analogously we can obtain the higher-order Casimir operators. The explicit expressions are

$$C_3(\chi) = (c_1^2 - c_2^2)c_3, \tag{5.26a}$$

$$C_4(\chi) = (1 + c_3^2/2)(\tfrac{1}{4}c_3^2 + 1 - c_1^2 - c_2^2) + c_1^2c_2^2 - c_3^2. \tag{5.26b}$$

D. The elementary representations of $\overline{SL(4,\mathbb{R})}$, $GL(4,\mathbb{R})$, and $\overline{GL(4,\mathbb{R})}$

Denote as before $\bar{G} = \overline{SL(4,\mathbb{R})}$, $G^e = GL(4,\mathbb{R})$, and $\bar{G}^e = \overline{GL(4,\mathbb{R})}$ and the corresponding minimal parabolic subgroups by $\bar{P}_0$, P_0^e, and $\bar{P}_0^e$. The elementary representations induced from the minimal parabolic subgroups will be parametrized by the signatures

$$\bar{\chi} \equiv [c_1,c_2,c_3;\epsilon_1,\epsilon_2,\epsilon_3,\epsilon_4], \tag{5.27a}$$

$$\chi^e \equiv [c_1,c_2,c_3,c_4;\epsilon_1,\epsilon_2,\epsilon_3,\epsilon_5], \tag{5.27b}$$

$$\bar{\chi}^e \equiv [c_1,c_2,c_3,c_4;\epsilon_1,\epsilon_2,\epsilon_3,\epsilon_4,\epsilon_5], \tag{5.27c}$$

respectively, where $c_k \in \mathbb{C}$. These elementary representations, which characterize the representation of A in (5.27a) and of $A^e = A \times R^+$ in (5.27b) and (5.27c), are the values of linear functionals λ,λ^e, respectively, for the basis elements of $\mathfrak{a}$, $\mathfrak{a}^e = \mathfrak{a} \oplus \mathfrak{z}$, respectively [cf. (3.16a)]:

$$\lambda^e|_{\mathfrak{a}} = \lambda, \quad \lambda^e(\hat{e}_4) = c_4, \tag{5.28}$$

λ is as in (4.2); $\epsilon_k = 0,1$ index the characters of $\bar{M}$, M^e, $\bar{M}^e$.

The one-dimensional representations analogous to $D^\chi(ma)$ in (5.3) are given by [cf. (3.14e), (3.16b), (3.17), and (3.21b)]:

$$D^{\bar{\chi}}(\bar{m}a) = (-1)^{\epsilon_1N_1 + \epsilon_2N_2 + \epsilon_3N_3 + \epsilon_4N_4} \times e^{-s(3+c_1) - t(1+c_2) - (r/2)c_3}; \tag{5.29a}$$

$$D^{\chi^e}(m^ea^e) = (-1)^{\epsilon_5N_5}e^{-qc_4}D^\lambda(ma), \tag{5.29b}$$

$$m^e = m \cdot (m_4^e)^{N_5}, \quad a^e = e^q a;$$

$$D^{\bar{\chi}^e}(\bar{m}^e a^e) = (-1)^{\epsilon_5 N_5} e^{-qc_4} D^{\bar{\chi}}(\bar{m}a), \quad \bar{m}^e = \bar{m}(m_4^e)^{N_5}. \tag{5.29c}$$

In (5.29a) we have added ρ [cf. (4.21)] to λ as in (5.3); in (5.29b) and (5.29c) we have added to λ^e the half sum of the positive roots of the system $(\mathfrak{g}^e,\mathfrak{a}^e)$ [cf. (2.48)]: $\rho^e|_\mathfrak{a} = \rho$, $\rho^e|_\mathfrak{z} = 0$.

The representation spaces $\mathscr{C}_{\bar{\chi}}$, $\mathscr{C}_{\chi^e}$, $\mathscr{C}_{\bar{\chi}^e}$ are given by (5.4) and the ER $\mathscr{T}^{\bar{\chi}}$, $\mathscr{T}^{\chi^e}$, $\mathscr{T}^{\bar{\chi}^e}$ act by (5.5) with the obvious substitutions.

Further, for G^e and $\bar{G}^e$ one should add to (5.7b) the equation $\hat{e}_4 f = c_4 f$ and (5.7c) holds also for $k=4$.

The $\bar{K}$-, K^e-, $\bar{K}^e$-invariant scalar products are given by (5.8) with the constant $2\pi^4$ being replaced by $4\pi^4, 4\pi^4, 8\pi^4$, respectively [cf. (4.22)–(4.24)] besides the other (obvious) substitutions.

We note that these ER are also ER of G when $c_4 = 0$, $\epsilon_4 = \epsilon_5 = 0$. Analogously the ER of $\bar{G}^e$ with $c_4 = 0 = \epsilon_5$ are ER of $\bar{G}$ and with $\epsilon_4 = 0$ are ER of G^e. The ER of $\bar{G}$ (resp. $\bar{G}^e$) with $\epsilon_4 = 1$ are double-valued representations of G (resp. G^e). [To facilitate the phrasing of the above statements we have enumerated the signatures in (5.27b) a little oddly.]

Bearing the above in mind we can introduce the noncompact picture of the ER for all three groups using the formulas for G. Besides the obvious substitutions, the changes are as follows: in (5.12b) we write

$$g(\bar{g})^{-1}\tilde{n}(\hat{x}, y_\pm) = \tilde{n}(\hat{x}_{\bar{g}}, y_{\bar{g}_\pm}) n_{\bar{g}}^{-1} a_{\bar{g}}^{-1} m(\bar{m}_{\bar{g}})^{-1}, \tag{5.30}$$

where $g(\bar{g})$, $m(\bar{m})$ are the images of $\bar{g}, \bar{m}$ under the covering map (3.14g) [cf. also (4.4) and (4.15)]; in (5.13b) the sum $\Sigma_k \epsilon_k N_k$ involves the ϵ_k of the corresponding signature; and the second line of (5.13b) is replaced by

$$x_1' = (-1)^{N_1+N_3} x_1, \quad x_2' = (-1)^{N_1+N_3+N_5} x_2,$$
$$x_+' = (-1)^{N_2+N_3+N_5} x_+, \quad x_-' = (-1)^{N_2+N_3} x_-, \tag{5.13b'}$$
$$y_+' = (-1)^{N_1+N_2} y_+, \quad y_-' = (-1)^{N_1+N_2+N_5} y_-;$$

in (5.13c) for $a \to a^e = e^q a$ the right-hand side is multiplied by e^{-qc_4}. We stress that $x', y_\pm'$ in (5.13c) remain unchanged as all formulas in (5.13a) and (5.13d)–(5.13f).

There are several technical reasons for the applicability of the SL(4,$\mathbb{R}$) formulas: the Weyl group is the same for all cases, the basis of $\bar{M}$ was chosen as to project on the basis of M plus the unit matrix, and in the G^e and $\bar{G}^e$ cases the representations of the center $\mathbb{R}^+$ do not act on $\tilde{N}$.

Thus all statements including Propositions 6 and 7 remain valid also for $\bar{G}$, G^e, and $\bar{G}^e$. In particular, the *principal series of unitary representations* of $\bar{G}$, G^e, and $\bar{G}^e$ is obtained for $c_k = i\tau_k$, $\tau_k \in \mathbb{R}$, $1 \leqslant k \leqslant 3,4,4$, respectively. Note that no restrictions on the corresponding ϵ_k are made. For G^e we refer to Refs. 11 and 13. For the connected part of $\bar{G}^e$, that is $\bar{G}_c^e = \bar{G}^e/Z_2 \cong R^+ \times \bar{G}$, thus $\epsilon_5 = 0$, the principal series of unitary representations was obtained for $\epsilon_1 = \epsilon_2 = \epsilon_3 = \epsilon_4 = 0$ by algebraic considerations in Ref. 17 and was used for $\epsilon_2 = \epsilon_3 = 0$ in Ref. 19 (the latter is not straightforward).

Since G and $\bar{G}$ have the same Lie algebra the infinitesimal generators and Casimir operators of the ER χ and $\bar{\chi}$ coincide and are given by formulas (5.20)–(5.26). For the Lie subalgebra $\mathfrak{a}^e$ in the cases G^e and $\bar{G}^e$ we can introduce an additional generator D_4,

$$D_4 f(\hat{x}, y_\pm) \equiv \frac{\partial}{\partial q}(T^\chi(a^e(q)f))(\hat{x}, y^\pm)|_{q=0} = -c_4 f, \tag{5.31}$$

its action naturally being equal to the right action of $-\hat{e}_4$ by (5.6). It is obvious that it commutes with all other generators and is the fourth (first-order) Casimir operator in the cases G^e and $\bar{G}^e$:

$$-D_4 = C_1(\chi^e) = C_1(\bar{\chi}^e) = c_4. \tag{5.32}$$

It is, of course, a general feature that reductive Lie groups (as G^e and $\bar{G}^e$) have dim $\mathfrak{z}$ in number first-order Casimir operators, where as here $\mathfrak{z}$ is the Lie algebra of the continuous center (in our cases dim $\mathfrak{z} = 1$).

ACKNOWLEDGMENTS

The authors are thankful to the referee for bringing to their attention Refs. 11 (and through it Ref. 10), 12, and 17. One of the authors (V.K.D.) would like to thank Professor H. D. Doebner for his hospitality at the Institute for Theoretical Physics of Clausthal Technical University.

V.K.D. would also like to thank the Humboldt Foundation for financial support during the work on the revised version.

[1]Y. Dothan, M. Gell-Mann, and Y. Ne'eman, Phys. Lett. **17**, 148 (1965).
[2]C. N. Yang, Phys. Rev. Lett. **33**, 445 (1974).
[3]V. I. Ogievetsky, Lett. Nuovo Cimento **8**, 17 (1973).
[4]C. J. Isham, A. Salam, and J. Strathdee, Ann. Phys. **62**, 98 (1971).
[5]F. Mansouri and L. N. Chang, Phys. Rev. D **13**, 3192 (1976).
[6]Y. Ne'eman, Found. Phys. **13**, 467 (1983).
[7]Y. Ne'eman and Dj. Sijacki, Tel Aviv Univ. preprint TAUP N159, 1985.
[8]R. Langlands, "On the Classification of irreducible representations of real algebraic groups," mimeographic notes, IAS, Princeton, 1973.
[9]A. W. Knapp and G. J. Zuckerman, "Classification theorems for representations of semi-simple groups," in *Non-Commutative Harmonic Analysis, Lecture Notes in Mathematics*, Vol. 587 (Springer, Berlin, 1977), p. 138.
[10]N. R. Wallach, Trans. Am. Math. Soc. **158**, 107 (1971).
[11]H. Kraljevič, Glasnik Mat. **7**, 49 (1972).
[12]L. Clozel, in *Non-Commutative Harmonic Analysis, Lecture Notes in Mathematics*, Vol. 728 (Springer, Berlin, 1979), p. 17.
[13]B. Speh, Math. Ann. **258**, 113 (1981).
[14]V. K. Dobrev, V. B. Petkova, S. G. Petrova, and I. T. Todorov, Phys. Rev. D **13**, 887 (1976).
[15]V. K. Dobrev, G. Mack, V. B. Petkova, S. G. Petrova, and I. T. Todorov, "Harmonic analysis on the *n*-dimensional Lorentz group and its applications to conformal QFT," *Lecture Notes in Physics*, Vol. 63 (Springer, Berlin, 1977).
[16]V. K. Dobrev and V. B. Petkova, Rep. Math. Phys. **13**, 233 (1978).
[17]A. B. Borisov, Rep. Math. Phys. **13**, 141 (1978).
[18]J. L. Friedman and R. D. Sorkin, J. Math. Phys. **21**, 1269 (1980).
[19]J. Mickelsson, Commun. Math. Phys. **88**, 551 (1983).
[20]V. B. Petkova and G. M. Sotkov, Bulg. J. Phys. **10**, 144, 256 (1983); Lett. Math. Phys. **8**, 217 (1984).
[21]V. K. Dobrev, ICTP, Trieste, Internal report IC/83/36, 1983; J. Math. Phys. **26**, 235 (1985).
[22]G. Warner, *Harmonic Analysis on Semi-Simple Lie Groups. I, II* (Springer, Berlin, 1972).

[23] A. W. Knapp and E. M. Stein, Ann. Math. **93**, 489 (1971); Invent. Math. **60**, 9 (1980).
[24] K. Iwasawa, Ann. Math. **50**, 507 (1949).
[25] I. M. Gel'fand and M. A. Naimark, Izv. AN SSSR Math. **11**, 411 (1947).
[26] F. Bruhat, Bull. Soc. Math. France **84**, 97 (1956).
[27] V. K. Dobrev, Congress talk and INRNE Sofia preprint, 1983; Lett. Math. Phys. **9**, 205 (1985); talk at IV Clausthal Workshop, 1984.
[28] S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962).
[29] J. Dixmier, *Enveloping Algebras* (North-Holland, New York, 1977).

The symmetric group: Algebraic formulas for some S_f $6j$ symbols and $S_f \supset S_{f_1} \times S_{f_2}$ $3jm$ symbols

R. W. Haase and R. Dirl
Institut für Theoretische Physik, Technische Universität Wien, Karlsplatz 13, A-1040 Wien, Austria

(Received 27 March 1985; accepted for publication 28 October 1985)

Explicit rank-dependent expressions have been obtained for some symmetric group (S_f) $6j$ symbols and some $S_f \supset S_{f_1} \times S_{f_2}$ $3jm$ symbols using Butler's recursion method. A key point in deriving these results is the use of the reduced notation introduced by Murnaghan to label irreps. Various symmetries of the $6j$ and $3jm$ symbols have been imposed. These include the complex conjugation, permutation, and transpose conjugation. We incorporate a new symmetry that arises from the occurrence of the two isomorphic direct product groups $S_{f_1} \times S_{f_2}$ and $S_{f_2} \times S_{f_1}$ as subgroups of S_f. In relation to the tables of $6j$ and $3jm$ symbols presented, a discussion is given of the symmetric group-unitary group duality.

I. INTRODUCTION

The symmetric group has come to play an important role in many different contexts in mathematics, physics, and chemistry. As a finite group it assumes a central role by Cayley's theorem and exemplifies many of the group-theoretical concepts. Furthermore its close relationship with the general linear group and its subgroups via the methods of tensor analysis imply that the symmetric group lies in the background of numerous problems. Moreover, the symmetric groups occur as a symmetry groups of quantum (and classical) mechanical systems that contain a number of identical particles, and as such have been used extensively in atomic spectroscopy, nuclear physics, and molecular theory. The physical reason for the quantum applications stems from the bosonic and fermionic nature of the particles.

One main consideration to be made, especially in regard to applications, is the determination of the $3jm$ symbols (Wigner, symmetrized coupling, or Clebsch–Gordan coefficients) and $6j$ symbols (Racah or symmetrized recoupling coefficients). From the early 1950's, nuclear spectroscopists, such as Jahn, Elliot, Hope, Horie, Kaplan, Kramer, and Vanagas,[1] developed methods of calculating these coefficients. Hamermesh[2] gives a systematic treatment of the S_f Clebsch–Gordan coefficients by a recursion method employing $S_f \supset S_{f-1}$ coupling (or isoscalar) factors. More recently Schindler and Mirman[3] have produced tables of these coefficients for the groups S_2 to S_6, using projection operator techniques, while Chen and co-workers[4] have presented independently similar tables using a commuting operator method. Both sets of tables give numerical values, the former in floating point and the latter in rational form. However, due to the rapidly increasing rank and dimensions of the irreducible representations (irreps) of the symmetric group, these methods and others based on the explicit construction of matrix representations are seen to be cumbersome and formidable to use. Moreover such tables for $S_f, f > 6$, would be enormous and impractical to present. The unitary groups also share the same problems. These were shown[5] to be largely overcome by choosing an algebraic approach. Butler's building-up[6] method is well suited for such an approach since no explicit matrix representations are required—only a knowledge of the character theory, namely dimensions of irreps, products, and branching rules, is needed. Furthermore, by casting these results in a rank-independent form, the rank dependency of the unitary group $6j$ symbols and $U_{p_1p_2} \supset U_{p_1} \times U_{p_2}$ $3jm$ symbols was able to be obtained.

In this paper the symmetric group is used to illustrate again this rank-independent algebraic approach. We present tables of algebraic formulas of primitive $6j$ symbols for S_f and primitive $3jm$ symbols for $S_f \supset S_{f_1} \times S_{f_2}$ $(f = f_1 + f_2)$ valid for all values of f_1, f_2, and f. The paper is arranged in the following way. In Sec. II we give the necessary character theory of the symmetric group. To cast it in a rank-independent form we have used the reduced notation introduced by Murnaghan[7] for labeling irreps of S_f. Littlewood[8] and Butler and King[9] employed this notation to derive many of the character theory results given here. In Secs. III and IV we give a guide to the tables and an outline of the method of calculation; a detailed account can be found in Ref. 8. In Secs. V–VII various symmetries of the $6j$ and $3jm$ symbols are discussed. The transposition symmetry arises from the occurrence of $S_{f_1} \times S_{f_2}$ and $S_{f_2} \times S_{f_1}$ as subgroups of $S_{f_1+f_2}$, while the transpose conjugate symmetry originates as a consequence of the one-dimensional alternating irrep $[1^f]$ of S_f. Above we mentioned briefly the connection between the symmetric group and the compact continuous groups. By way of example, we consider the unitary groups and the "duality" symmetries that arise. This duality leads to a powerful method of determining unitary group transformation coefficients, such as the $U_{p_1p_2} \supset U_{p_1} \times U_{p_2}$ $3jm$ symbols. Moreover this symmetry is independent of the unitary group ranks.

II. S_f GROUP INFORMATION

In this section we give an outline of the properties of the irreps of the symmetric groups. We shall be using two notations to label the irreps of S_f. The first is the well-known partition label. A partition of the integer f is a set of p integers $\lambda_1, \lambda_2, \ldots, \lambda_p$ with $\lambda_1 + \lambda_2 + \cdots + \lambda_p = f$, and denoted $(\lambda_1\lambda_2\cdots\lambda_p)$ or merely (λ). We include the possibility of negative integers for one or more of the λ_i's. If the parts also satisfy $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_p > 0$, (λ) is said to be regular. There is then a natural one-to-one correspondence between all regu-

 0022-2488/86/040900-14$02.50

lar partitions of the integer f and the S_f irrep labels, which are denoted by enclosing the partition in square brackets $[\lambda]$. Such a labeling is often called natural or standard. The nonstandard labels, that is, those with nonregular parts are character-equivalent to within a sign to standard irrep labels. The prescription for this correspondence is given by the irrep label modification rule

$$[\lambda_1,\ldots,\lambda_i,\lambda_{i+1},\ldots,\lambda_p] = -[\lambda_1,\ldots,\lambda_{i+1}-1,\lambda_i+1,\ldots,\lambda_p]\,. \tag{2.1}$$

Repeated application may be necessary to obtain a regular partition. Note also that if $\lambda_{i+1}=\lambda_i+1$ for any $1\leqslant i<p$, the partition label is inadmissible or null. We shall denote this null label by $\varnothing$.

The second labeling, called the reduced notation, originated from the work of Murnaghan[7] on the group embedding $O_{f-1}\supset S_f$ and exploited to great effect by Littlewood[8] and Butler and King.[9] It is obtained from the standard S_f irrep label $[\lambda]$ by dropping the first part λ_1. The resulting partition (γ) is then a regular partition of $f-\lambda_1$ into $p-1$ parts. We shall denote this labeling of the symmetric group irreps by using angular brackets $\langle\gamma\rangle$. For a given f the standard labeling can be recovered by

$$\langle\gamma\rangle=\langle\gamma_1\gamma_2\cdots\gamma_q\rangle=[f-l,\gamma_1\gamma_2\cdots\gamma_q]\,, \tag{2.2}$$

where $\langle\gamma\rangle$ is a partition of l into q parts. When $f-l<\gamma_1$, this correspondence leads to nonstandard labels, which may not necessarily be discarded and must be modified according to (2.1). In what follows, we shall use λ, μ, and ν to denote standard irrep labels. The corresponding reduced notation for these standard labels will be denoted γ, η, and κ, respectively. The advantage of the reduced notation is that the properties of the irreps can be given independent of the rank. For example the dimension of the irrep $[\lambda]$ is given by Robinson's formula[10] as

$$|[\lambda]|=f!/H_{[\lambda]}\,,$$

where

$$H_{[\lambda]}=\frac{h_1!\cdots h_p!}{\Pi_{i<j}(h_i-h_j)}\,, \tag{2.3}$$

and the hook length $h_i=\lambda_i+p-i$. This gives numerical results only. The rank-independent formula can be obtained from

$$|\langle\gamma\rangle|_f=\frac{f!}{h_0!}\left[\prod_{i=1}^{q}(h_0-h_i)\right]\frac{1}{H_{[\gamma]}}\,, \tag{2.4}$$

where (γ) is a regular partition of l into q parts, $h_0=f+q-l$, and $h_i=\gamma_i+q-i$ $(i=1,\ldots,q)$ are the hook lengths of (γ). The f dependence has now been factored, and the dimension is given as a factored polynomial in f. However, the f dependency can be obtained in a different way via the modification rules. Using (2.1), we find the particular values, say a, for which the irrep label $[f-l,\gamma_1\cdots\gamma_q]$ reduces to the null label $\varnothing$. Since $\varnothing$ is of "dimension" 0, the dimension formula for $[f-l,\gamma_1\cdots\gamma_q]$ must reflect this vanishing by an $(f-a)$ dependency. For example, using (2.1) and (2.2), $\langle 2\rangle$ is null for $f=0$ (i.e., $h_0=-2$, $\langle 2\rangle=[-220]=-[1-10]=[1-10]=\varnothing$) and for $f=3$ (i.e., $h_0=1$, $\langle 2\rangle=[12]=-[12]=\varnothing$). Hence $|\langle 2\rangle|_f\propto f(f-3)$. The numerical part can be obtained by taking a fixed value of f for which $\langle 2\rangle$ is admissible such as $f=4$. We know $|\langle 2\rangle|_4=|[22]|=2$. This implies that $|\langle 2\rangle|_f=f(f-3)/2$.

The Kronecker products of the irreps are usually obtained using the inner multiplication of Schur functions[11,12]

$$[\lambda_1]\times[\lambda_2]=\sum m^{\lambda}_{\lambda_1\circ\lambda_2}[\lambda]\,, \tag{2.5}$$

where (λ_1), (λ_2), and (λ) are partitions of the same integer f and $m^{\lambda}_{\lambda_1\circ\lambda_2}$ is the multiplicity of λ in the Schur function inner multiplication $\lambda_1\circ\lambda_2$. Tables of inner multiplication are to be found in Refs. 12 and 13. Again these results apply for a specified rank. Butler and King[9] give a rank-independent result based on the reduced notation

$$\langle\gamma_1\rangle\times\langle\gamma_2\rangle=\sum_{\xi\xi_1\xi_2}\langle(\gamma_1/\xi\xi_1)\cdot(\gamma_2/\xi\xi_2)\cdot(\xi_1\circ\xi_2)\rangle$$
$$=\sum_{\gamma}m^{\gamma}_{\gamma_1\gamma_2}\langle\gamma\rangle\,, \tag{2.6}$$

where "/", "·", and "∘" are the Schur function operations of division, and outer multiplication and inner multiplication, respectively. Tables of these operations are to be found in Ref. 12. The partitions ξ_1 and ξ_2 are restricted to being partitions of the same integer by the inner multiplication. The result is valid for all f, however, to make an application to any particular S_f, we must use (2.2) and (2.1), by which the product can be given in standard partition form. As an example of the use of (2.6) we give the following product:

$$\begin{aligned}\langle 1^2\rangle\times\langle 2\rangle&=\sum_{\xi_1\xi_2\xi}\langle(1^2/\xi\xi_1)\cdot(2/\xi\xi_2)\cdot(\xi_1\circ\xi_2)\rangle\\&=\sum[\langle(1^2/\xi_1)\cdot(2/\xi_2)\cdot(\xi_1\circ\xi_2)\rangle\\&\quad+\langle(1/\xi_1)\cdot(1/\xi_2)\cdot(\xi_1\circ\xi_2)\rangle]\,,\end{aligned} \tag{2.7}$$

since ξ is restricted to being 0 and 1, with $(1^2/0)=1^2$, $(1^2/1)=1$, $(2/0)=2$, and $(2/1)=1$. The terms $(\xi_1\circ\xi_2)$ can range over only $(0\circ 0)=0$, $(1\circ 1)=1$, and $(1^2\circ 2)=1$. Hence using $(1^2/1^2)=0$, $(2/1^2)=\varnothing$, and the other Schur function divisions given above, we have

$$\begin{aligned}\langle 1^2\rangle\times\langle 2\rangle&=\langle 1^2\cdot 2\cdot 0\rangle+\langle 1\cdot 1\cdot 1\rangle+\langle 0\cdot 0\cdot 1^2\rangle\\&\quad+\langle 1\cdot 1\cdot 0\rangle+\langle 0\cdot 0\cdot 1\rangle\\&=\langle 21^2\rangle+\langle 31\rangle+\langle 1^3\rangle+2\langle 21\rangle+\langle 3\rangle\\&\quad+2\langle 1^2\rangle+\langle 2\rangle+\langle 1\rangle\,.\end{aligned} \tag{2.8}$$

For $f\geqslant 7$ this result needs no modification but for smaller values of f we have

$$\begin{aligned}f=0:&\quad\langle 0\rangle\times\varnothing=+\varnothing+\varnothing-\langle 0\rangle+2\varnothing\\&\qquad+\varnothing+2\langle 0\rangle+\varnothing-\langle 0\rangle=\varnothing\,,\\f=1:&\quad\varnothing\times-\langle 0\rangle=-\langle 0\rangle+\varnothing+\varnothing+2\langle 0\rangle\\&\qquad+\varnothing+2\varnothing-\langle 0\rangle+\varnothing=\varnothing\,,\\f=2:&\quad\varnothing\times-\langle 1\rangle=+\varnothing+\langle 0\rangle+\varnothing+\varnothing\\&\qquad-\langle 0\rangle+2\varnothing-\langle 1\rangle+\langle 1\rangle=\varnothing\,,\\f=3:&\quad\langle 1^2\rangle\times\varnothing=+\varnothing+\varnothing+\varnothing-2\langle 1^2\rangle\\&\qquad-\langle 1\rangle+2\langle 1^2\rangle+\varnothing+\langle 1\rangle=\varnothing\,,\end{aligned} \tag{2.9}$$

$$f=4:\quad \langle 1^2\rangle\times\langle 2\rangle = -\langle 1^3\rangle - \langle 1^2\rangle + \langle 1^3\rangle + 2\varnothing - \langle 2\rangle + 2\langle 1^2\rangle + \langle 2\rangle + \langle 1\rangle\,,$$

$$f=5:\quad \langle 1^2\rangle\times\langle 2\rangle = +\varnothing - \langle 21\rangle + \langle 1^3\rangle + 2\langle 21\rangle + \varnothing + 2\langle 1^2\rangle + \langle 2\rangle + \langle 1\rangle\,,$$

$$f=6:\quad \langle 1^2\rangle\times\langle 2\rangle = +\langle 21^2\rangle + \varnothing + \langle 1^3\rangle + 2\langle 21\rangle + \langle 3\rangle + 2\langle 1^2\rangle + \langle 2\rangle + \langle 1\rangle\,.$$

Note the cancellation of terms, especially the multiplicity cases $\langle 1^2\rangle$ and $\langle 21\rangle$ for $f=4$ and $f=5$, respectively. Such results illustrate the point that the reduced notation and the modification rule (2.1) give a natural f dependence to the multiplicity separation problem.

The symmetric groups contain two one-dimensional irreps, the scalar irrep $\langle 0\rangle = [f]$ and the alternating or pseudoscalar irrep $\langle 1^{f-1}\rangle = [1^f]$. These lead to symmetries within the group. The former is associated with complex conjugation symmetry; the scalar irrep always occurs in the symmetric part of the Kronecker square of any irrep, hence all the irreps are real orthogonal irreps. The latter gives rise to the transpose conjugate symmetry, which relates pairs of irreps $[\lambda]$ and $[\tilde{\lambda}]$. The partition $(\tilde{\lambda})$ is obtained from (λ) by interchanging rows and columns of the Young diagram of (λ). The relationship is expressed by the Kronecker product

$$[\lambda]\times[1^f] = [\tilde{\lambda}]$$

or

$$\langle\gamma\rangle\times\langle 1^{f-1}\rangle = \langle\tilde{\gamma}\rangle\,. \tag{2.10}$$

From the second equation we note two points. The first is that the partition $(\tilde{\gamma})$ is not the transpose conjugate partition of (γ) in the sense that $(\tilde{\lambda})$ is of (λ). The notation used here is to denote that for each $\langle\gamma\rangle$ there is one label $\langle\tilde{\gamma}\rangle$ given by (2.10). Now, $(\tilde{\gamma})$ can be obtained from (γ) by the steps

$$\langle\gamma\rangle \xrightarrow{(2.2)} [\lambda] \xrightarrow{(2.10)} [\tilde{\lambda}] \xrightarrow{(2.2)} \langle\tilde{\gamma}\rangle\,,$$

e.g., (2.11)

$$\langle 1^2\rangle\to[f-2,1^2]\to[3,1^{f-3}]\to\langle 1^{f-3}\rangle\,.$$

The second point is that (2.10) is rank dependent. To express the transpose conjugate symmetry in a rank-independent manner we shall denote the pseudoscalar irrep as

$$\langle\tilde{0}\rangle = \langle 1^{f-1}\rangle = [1^f]\,. \tag{2.12}$$

This emphasizes more clearly the one-dimensional nature of $[1^f]$, and the symmetry derived from it,

$$\langle\gamma\rangle\times\langle\tilde{0}\rangle = \langle\tilde{\gamma}\rangle\,. \tag{2.13}$$

Butler[6] has used a similar notation for the point groups. The transpose conjugate symmetry will be discussed further in Sec. VI.

The final group information that we require is the branching rule for the group–subgroup chain $S_f\supset S_{f_1}\times S_{f_2}$ with $f_1+f_2=f$. In terms of regular partitions this is given by the outer multiplication of Schur functions involving the well-known Littlewood–Richardson rules

$$[\lambda]\downarrow\sum_{\mu}[\mu]\times[\lambda/\mu] = \sum_{\mu,\nu} m^{\lambda}_{\mu\nu}\,[\mu]\times[\nu]\,, \tag{2.14}$$

where $[\mu]$ and $[\nu]$ label S_{f_1} and S_{f_2} irreps. The same branching rule can be expressed in a rank-independent form

$$\langle\gamma\rangle\downarrow\sum_{n\eta}\langle\eta\rangle\cdot\langle\gamma/n\cdot\eta\rangle = \sum_{\eta\kappa} m^{\gamma}_{\eta\kappa}\,\langle\eta\rangle\times\langle\kappa\rangle\,, \tag{2.15}$$

where $\langle\eta\rangle$ and $\langle\kappa\rangle$ are S_{f_1} and S_{f_2} irrep labels, and n is a one row partition whose integer values range between 0 and γ_1. For any particular values f_1 and f_2, the modification rule and (2.2) may be necessary to convert this rank-independent result into standard partition form. For example, we derive the decomposition for irrep $\langle 2\rangle$,

$$\begin{aligned}\langle 2\rangle&\downarrow\sum_{\eta}\langle\eta\rangle\times\langle 2/(0+1+2)\cdot\eta\rangle\\ &\downarrow\sum_{\eta}\langle\eta\rangle\times\langle(2+1+0)/\eta\rangle\\ &\downarrow\langle 0\rangle\times(\langle 2\rangle+\langle 1\rangle+\langle 0\rangle)\\ &\quad+\langle 1\rangle\times(\langle 1\rangle+\langle 0\rangle)+\langle 2\rangle\times\langle 0\rangle\,.\end{aligned} \tag{2.16}$$

TABLE I. Symmetric group properties.

Notation:									
f-independent	0	1	2	1^2	...	$\tilde{1}^2$	$\tilde{2}$	$\tilde{1}$	$\tilde{0}$
reduced	$\langle 0\rangle$	$\langle 1\rangle$	$\langle 2\rangle$	$\langle 1^2\rangle$	...	$\langle 1^{f-3}\rangle$	$\langle 21^{f-4}\rangle$	$\langle 1^{f-2}\rangle$	$\langle 1^{f-1}\rangle$
standard	$[f]$	$[f-1,1]$	$[f-2,2]$	$[f-2,1^2]$	...	$[3,1^{f-3}\rangle$	$[2^2,1^{f-4}]$	$[2,1^{f-2}]$	$[1^f]$
Transpose conjugate	$\tilde{0}$	$\tilde{1}$	$\tilde{2}$	$\tilde{1}^2$	...	1^2	2	1	0
Dimension	1	$f-1$	$f(f-3)/2$	$(f-1)(f-2)/2$	...	$(f-1)(f-2)/2$	$f(f-3)/2$	$f-1$	1
Power of the irrep	0	1	2	2	...	$f-3$	$f-2$	$f-2$	$f-1$

Triads and $3j$ phases

0	0	0	0	+	2	2	1	0	+	1^2	2	2	0	+
1	1	0	0	+	2	2	2	0	+	1^2	1^2	0	0	+
1	1	1	0	+	2	2	2	1	+	1^2	1^2	2	0	+
2	1	1	0	+	1^2	1	1	0	−	1^2	1^2	2	1	+
2	2	0	0	+	1^2	2	1	0	±	1^2	1^2	1^2	0	−

Branching rules and transposition phases

0	↓	+	0×0										
1	↓	−	0×0	+	0×1	+	1×0						
2	↓	+	0×0	+	0×1	+	1×0	+	1×1	+	0×2	+	2×0
1^2	↓	+	0×1	+	1×0	−	1×1	+	0×1^2	+	$1^2\times 0$		

TABLE II. The S_f $6j$ symbols.

1	**1**	**1**		
1	1	1	0000	$+\frac{f-3}{(f-1)(f-2)}$
2	**1**	**1**		
1	1	1	0000	$-\frac{1}{(f-1)(f-2)}$
2	1	1	0000	$+\frac{f^2-3f+4}{f(f-1)(f-2)(f-3)}$
2	**2**	**1**		
1	1	1	0000	$+\frac{1}{(f-2)}\sqrt{\frac{f-4}{f(f-3)}}$
1	1	2	0000	$+\frac{f^2-7f+8}{(f-1)(f-2)}\sqrt{\frac{1}{f(f-3)(f-4)}}$
2	1	1	0000	$+\frac{1}{(f-2)(f-3)}$
2	2	1	0000	$+\frac{2f^3-15f^2+28f-16}{f(f-1)(f-2)(f-3)(f-4)}$
$\mathbf{1^2}$	**1**	**1**		
1	1	1	0000	$+\frac{1}{(f-1)(f-2)}$
2	1	1	0000	$+\frac{1}{(f-1)(f-2)}$
1^2	1	1	0000	$+\frac{1}{(f-1)(f-2)}$
$\mathbf{1^2}$	**2**	**1**		
1	1	1	0000	$+\frac{1}{(f-2)}\sqrt{\frac{1}{f-1}}$
1	1	2	0000	$+\{1^2\,2\,1\,0\}\frac{1}{(f-2)}\sqrt{\frac{f-4}{f(f-3)}}$
1	2	1	0000	$\frac{f-4}{f(f-2)(f-3)}$
2	1	1	0000	$-\frac{1}{(f-2)}\sqrt{\frac{f-4}{f(f-1)(f-3)}}$
2	2	1	0000	$-\frac{1}{(f-1)(f-2)(f-3)}$
1^2	1	1	0000	$+\frac{1}{(f-1)(f-2)}$
1^2	2	1	0000	$+\frac{2f^2-7f+4}{f(f-1)(f-2)(f-3)}$
$\mathbf{1^2}$	**2**	**2**		
1	1	1	0000	$+\frac{1}{(f-2)}\sqrt{\frac{f-4}{f(f-3)}}$
1	2	1	0000	$+\frac{1}{(f-2)(f-3)}\sqrt{\frac{1}{f-1}}$
1	2	2	0000	$-\frac{f^3-9f^2+20f-16}{f(f-1)(f-2)(f-3)(f-4)}$
2	1	1	0000	$-\frac{1}{(f-1)(f-2)}\sqrt{\frac{f}{(f-3)(f-4)}}$

1^2	1	1	0000	$-\{1^2\,2\,1\,0\}\frac{1}{(f-1)(f-2)}\sqrt{\frac{f-4}{f(f-3)}}$
1^2	2	1	0000	$+\frac{1}{(f-1)(f-2)}$
$\mathbf{1^2}$	$\mathbf{1^2}$	**1**		
1	1	1	0000	$-\frac{\sqrt{f-3}}{(f-1)(f-2)}$
1	1	2	0000	$+\{1^2\,2\,1\,0\}\frac{\sqrt{f-3}}{(f-1)(f-2)}$
1	1	1	0000	$+\frac{f-5}{(f-1)(f-2)}\sqrt{\frac{1}{f-3}}$
2	1	1	0000	$-\{1^2\,2\,1\,0\}\frac{1}{f-2}\sqrt{\frac{1}{(f-1)(f-3)}}$
2	1	2	0000	$-\{1^2\,2\,1\,0\}\frac{1}{(f-2)(f-3)}\sqrt{\frac{f-4}{f(f-1)}}$
2	2	1	0000	$+\{1^2\,2\,1\,0\}\frac{2f-3}{(f-1)(f-2)(f-3)}\sqrt{\frac{f-4}{f}}$
2	2	2	0000	$-\frac{f^2-7f+8}{(f-1)(f-2)(f-3)}\sqrt{\frac{1}{f(f-4)}}$
1^2	1	1	0000	$+\frac{1}{(f-1)(f-2)}$
1^2	2	1	0000	$-\frac{1}{(f-1)(f-2)(f-3)}$
1^2	1^2	1	0000	$+\frac{2f-7}{(f-1)(f-2)(f-3)}$
$\mathbf{1^2}$	$\mathbf{1^2}$	$\mathbf{1^2}$		
1	1	1	0000	$+\frac{\sqrt{f-3}}{(f-1)(f-2)}$
2	1	1	0000	$+\{1^2\,2\,1\,0\}\frac{1}{(f-1)(f-2)}\sqrt{\frac{1}{f-3}}$
2	2	1	0000	$+\{1^2\,2\,1\,0\}\frac{1}{(f-2)(f-3)}\sqrt{\frac{f-4}{f}}$
1^2	1	1	0000	$+\frac{1}{(f-1)(f-2)}\sqrt{\frac{1}{f-3}}$
1^2	1^2	1	0000	$-\frac{f-5}{(f-1)(f-2)(f-3)}$

We note two symmetries of the branching rules. If

$$\langle\gamma\rangle \supset m^{\gamma}_{\eta\kappa}\,\langle\eta\rangle\times\langle\kappa\rangle\,, \tag{2.17}$$

that is, it forms a "ket branching," then we have for f_1 and f_2 large enough

$$\langle\gamma\rangle \supset m^{\gamma}_{\kappa\eta}\,\langle\kappa\rangle\times\langle\eta\rangle\,, \tag{2.18}$$

and

$$\langle\tilde{\gamma}\rangle \supset m^{\tilde{\gamma}}_{\tilde{\eta}\tilde{\kappa}}\,\langle\eta\rangle\times\langle\kappa\rangle\,, \tag{2.19}$$

where $m^{\tilde{\gamma}}_{\tilde{\eta}\tilde{\kappa}} = m^{\gamma}_{\kappa\eta} = m^{\gamma}_{\eta\kappa} > 0$, as ket branchings. We must remember also that $(\tilde{\gamma})$, $(\tilde{\eta})$, and $(\tilde{\kappa})$ are not the transpose conjugate partitions of (γ), (η), and (κ), but must be obtained by (2.11). The symmetries originate from the transposition of the subgroups S_{f_1} and S_{f_2} and the pseudoscalar irreps, respectively. We will return to give a fuller discussion of these symmetries with regard to $6j$ and $3jm$ symbols in Secs. V and VI.

III. A GUIDE TO THE TABLES

Table I summarizes all the group information of the symmetric group required for our calculation. The irreps are ordered according to the power of the irrep, defined as the smallest integer $p(\gamma)$ such that the $p(\gamma)$th power of the primitive (or defining) irrep $\langle 1\rangle$ contains $\langle\gamma\rangle$. For those irreps of equal power, the ordering is fixed according to a comparison of the parts of the two partitions, the one with highest part first. Note that in reduced notation $p(\gamma) = l$ if (γ) is a partition of l. A conversion between our rank-independent notation, standard labels, and the reduced notation is given along with the dimension formula, power, and transpose conjugate label for each irrep.

The Kronecker product rules are specified by means of triads. Thus $(\gamma_1\gamma_2\gamma_3 r)$ forms a triad if the scalar irrep occurs at least $r+1$ times in the triple product $\langle\gamma_1\rangle \times \langle\gamma_2\rangle \times \langle\gamma_3\rangle$. We take the range of r to be initialized from zero. The triad is ordered according to the irrep order with highest first. Each ordered triad has an associated phase, the $3j$ phase, which gives the symmetry on reordering coupled products. These $3j$ phases, $\{\gamma_1\gamma_2\gamma_3 r\}$, are given in Table I.

The branching table gives the reduction of those irreps of power less than 3. The reduction of the other listed irreps can be obtained using the transpose conjugate symmetry. No branching multiplicity occurs. The table also includes the sign of the transposition phase (see Sec. V.), which gives the symmetry on transposing the subgroup irrep labels of the associated ket branching $(\gamma\, a\, \eta\, \kappa)$, $a = 0,\ldots,m^{\gamma}_{\eta\kappa} - 1$. This sign is placed before each subgroup irrep label.

The $6j$ symbol is a transformation coefficient related to recouplings between a set of six irreps by means of four triad couplings. The triads occur in the $6j$ symbol

$$\begin{Bmatrix} \gamma_1' & \gamma_2' & \gamma_3' \\ \gamma_1 & \gamma_2 & \gamma_3 \end{Bmatrix}_{r_1r_2r_3r_4}$$

in the order $(\gamma_1'\,\gamma_2\gamma_3 r_1)$, $(\gamma_1\,\gamma_2'\,\gamma_3 r_2)$, $(\gamma_1\gamma_2\,\gamma_3'\,r_3)$, $(\gamma_1'\,\gamma_2'\,\gamma_3'\,r_4)$.

The $6j$ symbol is necessarily zero unless the six irrep and four multiplicity labels fulfill the four triad conditions. Symmetries are used to reduce the size of the table and so we need the following (we have used the reality and orthogonality of the symmetric group irreps): (i) invariance under even permutations on the columns; (ii) column interchange symmetries, such as the (23) column interchange operation

$$\begin{Bmatrix} \gamma_1' & \gamma_2' & \gamma_3' \\ \gamma_1 & \gamma_2 & \gamma_3 \end{Bmatrix}_{r_1r_2r_3r_4} = \begin{Bmatrix} \gamma_1' & \gamma_3' & \gamma_2' \\ \gamma_1 & \gamma_3 & \gamma_2 \end{Bmatrix}_{r_1r_3r_2r_4} \times\{\gamma_1'\,\gamma_2\gamma_3 r_1\}\{\gamma_1\,\gamma_2'\,\gamma_3 r_2\} \times\{\gamma_1\gamma_2\,\gamma_3'\,r_3\}\{\gamma_1'\,\gamma_2'\,\gamma_3'\,r_4\}; \tag{3.1}$$

and (iii) the row-flip symmetries, such as the (23) flip operation

$$\begin{Bmatrix} \gamma_1' & \gamma_2' & \gamma_3' \\ \gamma_1 & \gamma_2 & \gamma_3 \end{Bmatrix}_{r_1r_2r_3r_4} = \begin{Bmatrix} \gamma_1' & \gamma_2 & \gamma_3 \\ \gamma_1 & \gamma_2' & \gamma_3' \end{Bmatrix}_{r_4r_3r_2r_1}. \tag{3.2}$$

The phase in (ii) is the same for all interchanges. The $6j$ symbols of S_f are tabulated in Table II. The boldfaced type headings denote the top line of the $6j$ symbol and each subsequent entry denotes the possible lower line (three irrep and four multiplicity labels) and its corresponding algebraic formula.

The $3jm$ symbols

$$\begin{pmatrix} \gamma_1 & \gamma_2 & \gamma_3 \\ a_1 & a_2 & a_3 \\ \eta_1\kappa_1 & \eta_2\kappa_2 & \eta_3\kappa_3 \end{pmatrix}^{r}_{st}$$

are also zero unless the top and bottom lines form triads of their respective groups, and the columns satisfy the ket branching criterion. Similarly symmetries are used to reduce the size of the table. These are as follows: (i) invariance under even permutations of the columns; and (ii) a possible sign change under odd permutations of the columns, the (23) operation being

$$\begin{pmatrix} \gamma_1 & \gamma_2 & \gamma_3 \\ a_1 & a_2 & a_3 \\ \eta_1\kappa_1 & \eta_2\kappa_2 & \eta_3\kappa_3 \end{pmatrix}^{r}_{st} = \begin{pmatrix} \gamma_1 & \gamma_3 & \gamma_2 \\ a_1 & a_2 & a_3 \\ \eta_1\kappa_1 & \eta_3\kappa_3 & \eta_2\kappa_2 \end{pmatrix}^{r}_{st} \times\{\gamma_1\gamma_2\gamma_3 r\}\{\eta_1\eta_2\eta_3 s\}\{\kappa_1\kappa_2\kappa_3 t\}. \tag{3.3}$$

The sign is the same for all such interchanges. Table III gives the tabulation of the $3jm$ symbols of $S_f \supset S_{f_1} \times S_{f_2}$. The group triad is used as a header. Each subsequent entry gives the three allowed subgroup irrep labels, $\eta\kappa$, and the corresponding algebraic formula.

IV. METHOD OF CALCULATION

The building-up method used for the calculation of the $6j$ and $3jm$ symbols (see Refs. 6, 14, and 15 for detailed accounts) takes advantage of the "phase freedom" that occurs within the Racah–Wigner algebra and that follows from Schur's lemmas. The phase freedom describes transformations in the product and branching multiplicity spaces. In the multiplicity-free case this phase freedom reduces to a phase, hence the origin of the term. In the present calculation the symmetric group properties provide simplifications within the Racah–Wigner algebra. In particular all irreps are real and orthogonal. Therefore the $1j$ phase $\{\gamma\}$ is always $+1$, and the A-matrix and $2jm$ symbol, which describe the complex conjugation symmetry of $6j$ and $3jm$ symbols, can always be chosen to be the unit matrix

$$A(\gamma_1\gamma_2\gamma_3)^{r'r} = \delta^{r'r}, \quad \begin{pmatrix} \gamma & \gamma \\ a' & a \\ \eta\kappa & \eta\kappa \end{pmatrix} = \delta_{a'a}. \tag{4.1}$$

In our calculation no nonsimple phase irreps occur; the first are the irreps $\langle 21\rangle = [f-3,2,1]$, $f \geqslant 6$. As a consequence the permutation matrix appropriate to the reordering of coupled products can be chosen diagonal:

$$\{c\gamma_1\gamma_2\gamma_3\}^{r'}_{r} = \delta^{r'}_{r}, \quad c \text{ an even permutation},$$
$$\{\tau\gamma_1\gamma_2\gamma_3\}^{r'}_{r} = \{\gamma_1\gamma_2\gamma_3 r\}\,\delta^{r'}_{r}, \quad \tau \text{ an odd permutation}, \tag{4.2}$$

where $\{\gamma_1\gamma_2\gamma_3 r\}$ takes the values ± 1. The transposition matrix associated with the transposition of the subgroup irrep labels of S_{f_1} and S_{f_2} can also be chosen to be diagonal

$$T(\gamma,\eta\kappa)^{a'}{}_{a} = (\gamma a \eta\kappa)\delta^{a'}{}_{a}, \tag{4.3}$$

where $(\gamma a\eta\kappa)$ is a sign factor ± 1 (see Sec. V).

The $6j$ and $3jm$ symbols of the symmetric group can now be calculated recursively by building up from the trivial $6j$

TABLE III. The $S_f \supset S_{f_1} \times S_{f_2}$ $3jm$ symbols.

1	**1**	**1**	**0**	
0×0	0×0	0×0	00	$+\sqrt{\frac{1}{(f-1)(f-2)f_1f_2}}\,[f_1-f_2]$
0×1	0×1	0×0	00	$+\sqrt{\frac{f_1(f_2-1)}{(f-1)(f-2)f_2}}$
0×1	0×1	0×1	00	$+\sqrt{\frac{f(f_2-1)(f_2-2)}{(f-1)(f-2)f_2}}$
1×0	1×0	0×0	00	$-\sqrt{\frac{(f_1-1)f_2}{(f-1)(f-2)f_1}}$
1×0	1×0	1×0	00	$+\sqrt{\frac{f(f_1-1)(f_1-2)}{(f-1)(f-2)f_1}}$
2	**1**	**1**	**0**	
0×0	0×0	0×0	00	$+\sqrt{\frac{2f(f_1-1)(f_2-1)}{(f-1)(f-2)(f-3)f_1f_2}}$
0×0	0×1	0×1	00	$-\sqrt{\frac{2f_1(f_1-1)}{f(f-1)(f-2)(f-3)f_2}}$
0×0	1×0	1×0	00	$-\sqrt{\frac{2f_2(f_2-1)}{f(f-1)(f-2)(f-3)f_1}}$
0×1	0×1	0×0	00	$+\sqrt{\frac{(f_2-1)(f_2-2)}{(f-2)(f-3)f_2}}$
0×1	0×1	0×1	00	$-\sqrt{\frac{4f_1(f_2-1)}{f(f-2)(f-3)f_2}}$
1×0	1×0	0×0	00	$-\sqrt{\frac{(f_1-1)(f_1-2)}{(f-2)(f-3)f_1}}$
1×0	1×0	1×0	00	$-\sqrt{\frac{4(f_1-1)f_2}{f(f-2)(f-3)f_1}}$
1×1	1×0	0×1	00	$+\sqrt{\frac{(f_1-1)(f_2-1)}{f(f-3)}}$
0×2	0×1	0×1	00	$+\sqrt{\frac{f_2(f_2-3)}{f(f-3)}}$
2×0	1×0	1×0	00	$+\sqrt{\frac{f_1(f_1-3)}{f(f-3)}}$
2	**2**	**1**	**0**	
0×0	0×0	0×0	00	$-\sqrt{\frac{4(f-1)}{f(f-2)(f-3)(f-4)f_1f_2}}\,[f_1-f_2]$
0×1	0×0	0×1	00	$+\sqrt{\frac{2(f_1-1)(f_2-2)}{(f-2)(f-3)(f-4)f_2}}$
0×1	0×1	0×0	00	$+\sqrt{\frac{f_2-1}{f(f-1)(f-2)(f-3)(f-4)f_1f_2}}\,[2f_1^2+f_1f_2-f_2^2-2f_1+2f_2]$
0×1	0×1	0×1	00	$+\sqrt{\frac{f_2-1}{(f-1)(f-2)(f-3)(f-4)f_2(f_2-2)}}\,[f_1f_2+f_2^2-4f_1-4f_2+4]$
1×0	0×0	1×0	00	$+\sqrt{\frac{2(f_1-2)(f_2-1)}{(f-2)(f-3)(f-4)f_1}}$
1×0	1×0	0×0	00	$-\sqrt{\frac{f_1-1}{f(f-1)(f-2)(f-3)(f-4)f_1f_2}}\,[2f_2^2+f_1f_2-f_1^2+2f_1-2f_2]$
1×0	1×0	1×0	00	$+\sqrt{\frac{f_1-1}{(f-1)(f-2)(f-3)(f-4)f_1(f_1-2)}}\,[f_1f_2+f_1^2-4f_1-4f_2+4]$

1×1	0×1	1×0	00	$-\sqrt{\frac{(f_1-1)(f_2-1)(f_2-2)}{(f-1)(f-3)(f-4)}}$
1×1	1×0	0×1	00	$-\sqrt{\frac{(f_1-1)(f_1-2)(f_2-1)}{(f-1)(f-3)(f-4)}}$
1×1	1×1	0×0	00	$+\sqrt{\frac{(f-2)(f_1-1)(f_2-1)}{f(f-1)(f-3)(f-4)f_1f_2}}\,[f_1-f_2]$
1×1	1×1	0×1	00	$+\sqrt{\frac{(f-2)(f_1-1)(f_2-1)(f_2-2)}{(f-1)(f-3)(f-4)f_2}}$
1×1	1×1	1×0	00	$+\sqrt{\frac{(f-2)(f_1-1)(f_1-2)(f_2-1)}{(f-1)(f-3)(f-4)f_1}}$
0×2	0×1	0×1	00	$+\sqrt{\frac{f_1f_2(f_2-3)}{(f-1)(f-3)(f-4)(f_2-2)}}$
0×2	0×2	0×0	00	$+\sqrt{\frac{2(f-2)f_1(f_2-3)}{f(f-1)(f-3)(f-4)}}$
0×2	0×2	0×1	00	$+\sqrt{\frac{(f-2)(f_2-1)(f_2-3)(f_2-4)}{(f-1)(f-3)(f-4)(f_2-2)}}$
2×0	1×0	1×0	00	$+\sqrt{\frac{f_1(f_1-3)f_2}{(f-1)(f-3)(f-4)(f_1-2)}}$
2×0	2×0	0×0	00	$-\sqrt{\frac{2(f-2)(f_1-3)f_2}{f(f-1)(f-3)(f-4)}}$
2×0	2×0	1×0	00	$+\sqrt{\frac{(f-2)(f_1-1)(f_1-3)(f_1-4)}{(f-1)(f-3)(f-4)(f_1-2)}}$
1^2	**1**	**1**	**0**	
0×1	0×1	0×0	00	$+\sqrt{\frac{f_2-1}{(f-1)(f-2)}}$
0×1	0×1	0×1	00	0
1×0	1×0	0×0	00	$-\sqrt{\frac{f_1-1}{(f-1)(f-2)}}$
1×0	1×0	1×0	00	0
1×1	1×0	0×1	00	$+\sqrt{\frac{(f_1-1)(f_2-1)}{(f-1)(f-2)}}$
0×1	0×1	0×1	00	$+\sqrt{\frac{(f_2-1)(f_2-2)}{(f-1)(f-2)}}$
1×0	1×0	1×0	00	$+\sqrt{\frac{(f_1-1)(f_1-2)}{(f-1)(f-2)}}$
1^2	**2**	**1**	**0**	
0×1	0×0	0×1	00	$+\sqrt{\frac{2(f_1-1)}{f(f-2)(f-3)}}$
0×1	0×1	0×0	00	$+\sqrt{\frac{(f_2-1)(f_2-2)}{(f-1)(f-2)(f-3)f_1}}$
0×1	0×1	0×1	00	$+\sqrt{\frac{(f-2)(f_2-1)}{f(f-1)(f-3)}}$
0×1	1×1	1×0	00	$+\sqrt{\frac{(f_1-1)f_2(f_2-1)}{f(f-1)(f-3)f_1}}$
0×1	0×2	0×1	00	$-\sqrt{\frac{f_1(f_2-3)}{f(f-1)(f-3)}}$
1×0	0×0	1×0	00	$+\sqrt{\frac{2(f_2-1)}{f(f-2)(f-3)}}$

1×0	1×0	0×0	00	$+\sqrt{\frac{(f_1-1)(f_1-2)}{(f-1)(f-2)(f-3)f_2}}$
1×0	1×0	1×0	00	$+\sqrt{\frac{(f-2)(f_1-1)}{f(f-1)(f-3)}}$
1×0	1×1	0×1	00	$+\sqrt{\frac{f_1(f_1-1)(f_2-1)}{f(f-1)(f-3)f_2}}$
1×0	2×0	1×0	00	$-\sqrt{\frac{(f_1-3)f_2}{f(f-1)(f-3)}}$
1×1	0×1	1×0	00	$-\sqrt{\frac{(f_1-1)(f_2-1)(f_2-2)}{(f-1)(f-2)(f-3)f_1}}$
1×1	1×0	0×1	00	$+\sqrt{\frac{(f_1-1)(f_1-2)(f_2-1)}{(f-1)(f-2)(f-3)f_2}}$
1×1	1×1	0×0	00	$-\sqrt{\frac{f(f_1-1)(f_2-1)}{(f-1)(f-3)f_1f_2}}$
1×1	1×1	0×1	00	$-\sqrt{\frac{(f_1-1)(f_2-1)(f_2-2)}{(f-1)(f-3)f_2}}$
1×1	1×1	1×0	00	$+\sqrt{\frac{(f_1-1)(f_1-2)(f_2-2)}{(f-1)(f-3)f_1}}$
0×1^2	0×1	0×1	00	$-\sqrt{\frac{f_1(f_2-1)}{(f-1)(f-2)(f-3)}}$
0×1^2	0×2	0×1	00	$+\sqrt{\frac{(f_2-1)(f_2-3)}{(f-1)(f-3)}}$
$1^2\times 0$	1×0	1×0	00	$-\sqrt{\frac{(f_1-1)f_2}{(f-1)(f-2)(f-3)}}$
$1^2\times 0$	2×0	1×0	00	$+\sqrt{\frac{(f_1-1)(f_1-3)}{(f-1)(f-3)}}$
1^2	**1^2**	**1**	**0**	
0×1	0×1	0×0	00	$-\sqrt{\frac{(f_2-1)}{(f-1)(f-2)(f-3)f_1f_2}}\,[2f_1-f_2]$
0×1	0×1	0×1	00	$-\sqrt{\frac{f(f_2-1)(f_2-2)}{(f-1)(f-2)(f-3)f_2}}$
1×0	1×0	0×0	00	$+\sqrt{\frac{(f_1-1)}{(f-1)(f-2)(f-3)f_1f_2}}\,[2f_2-f_1]$
1×0	1×0	1×0	00	$-\sqrt{\frac{f(f_1-1)(f_1-2)}{(f-1)(f-2)(f-3)f_1}}$
1×1	0×1	1×0	00	$-\sqrt{\frac{(f_1-1)f_2(f_2-1)}{(f-1)(f-2)(f-3)f_1}}$
1×1	1×0	0×1	00	$+\sqrt{\frac{f_1(f_1-1)(f_2-1)}{(f-1)(f-2)(f-3)f_2}}$
1×1	1×1	0×0	00	$-\sqrt{\frac{(f_1-1)(f_2-1)}{(f-1)(f-2)(f-3)}}\,[f_1-f_2]$
1×1	1×1	0×1	00	$-\sqrt{\frac{f(f_1-1)(f_2-1)(f_2-2)}{(f-1)(f-2)(f-3)f_2}}$
1×1	1×1	1×0	00	$-\sqrt{\frac{f(f_1-1)(f_1-2)(f_2-1)}{(f-1)(f-2)(f-3)f_1}}$
0×1^2	0×1	0×1	00	$-\sqrt{\frac{f_1(f_2-1)(f_2-2)}{(f-1)(f-2)(f-3)f_2}}$

TABLE III. (*Continued.*)

0×1^2	0×1^2	0×0	00	$-\sqrt{\dfrac{2f_1(f_2-1)(f_2-2)}{(f-1)(f-2)(f-3)f_2}}$
0×1^2	0×1^2	0×1	00	$+\sqrt{\dfrac{f(f_2-1)(f_2-2)(f_2-3)}{(f-1)(f-2)(f-3)f_2}}$
$1^2\times0$	1×0	1×0	00	$-\sqrt{\dfrac{(f_1-1)(f_1-2)f_2}{(f-1)(f-2)(f-3)f_1}}$
$1^2\times0$	$1^2\times0$	0×0	00	$+\sqrt{\dfrac{2(f_1-1)(f_1-2)f_2}{(f-1)(f-2)(f-3)f_1}}$
$1^2\times0$	$1^2\times0$	1×0	00	$+\sqrt{\dfrac{f(f_1-1)(f_1-2)(f_1-3)}{(f-1)(f-2)(f-3)f_1}}$

and $3jm$ symbols and incorporating the symmetries given above. The basic procedure follows that used in numeric[16–20] and algebraic[5,21] examples. Therefore we give a brief outline only. Phase freedom still exists subject to certain restrictions imposed by the above choices (4.1)–(4.3). Such freedom allows a choice of the sign or phase or even magnitude of certain $6j$ and $3jm$ symbols subject only to the unitarity conditions. Linear equations in the unknown $6j$ and $3jm$ symbols, which have no phase freedom, are then generated by identities in the Racah–Wigner algebra, for example the Racah backcoupling relation, the Biedenharn–Elliott sum rule and the Wigner relation. Since the $6j$ symbols are independent of the subgroup basis labels these are calculated first. A small set of $6j$ symbols for both group and subgroup are required in the Wigner relation to calculate the $3jm$ symbols.

V. THE TRANSPOSITION SYMMETRY

As mentioned earlier, the transposition symmetry arises as a consequence of being able to embed the two isomorphic direct product groups $S_{f_1}\times S_{f_2}$ and $S_{f_2}\times S_{f_1}$ in S_f. This correspondence is an inner automorphism of S_f and an outer automorphism for the subgroups. By requiring that the matrix representations of these two subgroups be identical, the symmetry can be described by a matrix $T(\gamma,\eta\kappa)$, which is indexed by only the branching multiplicity label a. This matrix, which we call the transposition matrix, can be chosen, within the hierarchy of symmetries, to be diagonal.[22] In the particular case of the symmetric groups the diagonal entries, called transposition phases, satisfy

$$(\gamma a\kappa\eta) = (\gamma a\eta\kappa)^* = (\gamma a\eta\kappa) = \pm 1\,, \tag{5.1}$$

and are chosen such that they are rank independent. In the special case when $\eta=\kappa$, the phases are fixed by the character theory. Replacing the reduced notation labels by the standard labels we have

$$(\gamma a\eta\eta) = (\lambda a\mu\mu) = \begin{cases} +1\,, & \text{if } \mu \otimes 2 \text{ contains} \\ & \text{the } a\text{th occurrence of } \lambda, \\ -1\,, & \text{if } \mu \otimes 1^2 \text{ contains} \\ & \text{the } a\text{th occurrence of } \lambda\,, \end{cases} \tag{5.2}$$

where $\otimes$ denotes the Schur function operation of outer plethysm.[23] Even though the branchings in our calculation are multiplicity-free, the values of the transposition phases can have a striking effect on the algebraic form of some $3jm$ symbols as we will show shortly.

The transposition symmetry thus relates $3jm$ symbols of $S_f\supset S_{f_1}\times S_{f_2}$ and those of $S_f\supset S_{f_2}\times S_{f_1}$ (we distinguish the latter by a prime):

$$\begin{pmatrix} \gamma_1 & \gamma_2 & \gamma_3 \\ a_1 & a_2 & a_3 \\ \eta_1\kappa_1 & \eta_2\kappa_2 & \eta_3\kappa_3 \end{pmatrix}^{r}_{st} = \begin{pmatrix} \gamma_1 & \gamma_2 & \gamma_3 \\ a_1 & a_2 & a_3 \\ \kappa_1\eta_1 & \kappa_2\eta_2 & \kappa_3\eta_3 \end{pmatrix}^{\prime r}_{st} \times(\gamma_1 a_1\eta_1\kappa_1)(\gamma_2 a_2\eta_2\kappa_2)(\gamma_3 a_3\eta_3\kappa_3)\,. \tag{5.3}$$

If $f_1=f_2$, the pair of $3jm$ symbols belong to the same group–subgroup chain and with $\eta_i=\kappa_i$, $i=1,2,3$, the two $3jm$ symbols are identical. In such cases the product of transposition phases in (5.3) may be -1. The $3jm$ symbol must then vanish. Such vanishings force the appearance of the factor (f_1-f_2) in the algebraic formula of the $3jm$ symbol. This can be seen with

$$\begin{pmatrix} 1^2 & 1 & 1 \\ 0 & 0 & 0 \\ 1\times1 & 1\times1 & 1\times1 \end{pmatrix}^{0}_{00}, \tag{5.4}$$

where the transposition phases are

$$(1^2\ 0\ 1\times1) = -1 \text{ and } (1\ 0\ 1\times1) = +1\,,$$

and the S_{f_1} and S_{f_2} subgroup labels have been separated by "×."

VI. THE TRANSPOSE CONJUGATE SYMMETRY

As noted by several authors,[2,24,25] the one-dimensional irreps of a group lead to symmetries in the Racah–Wigner algebra. For the symmetric groups, the pseudoscalar irrep $\langle\tilde{0}\rangle=[1^f]$ provides such a symmetry. In this section we follow to a large degree the work of Ford and Butler[25] but note that for the symmetric groups they consider only the embedding $S_f\supset S_{f-1}\times S_1$. Their results are extended to the general case $S_f\supset S_{f_1}\times S_{f_2}$ (see Ref. 26), but for our present purposes we give only results that pertain to nonsimple phase irreps and multiplicity-free products and branchings.

To describe the transpose conjugate (or "tilde") symmetry we define two special matrices, the Λ-matrices,

$$\Lambda(\gamma_1\gamma_2\gamma_3)^{r'r} = |\gamma_1,\gamma_2|^{1/2}\begin{Bmatrix}\tilde{0} & \tilde{\gamma}_3 & \gamma_3\\ \gamma_1 & \gamma_2 & \tilde{\gamma}_2\end{Bmatrix}_{0r'r0},$$
$$\Lambda(\gamma,\eta\kappa)_{a'a} = \left|\frac{\gamma}{\eta\kappa}\right|^{1/2}\begin{pmatrix}\tilde{0} & \tilde{\gamma} & \gamma\\ 0 & a' & a\\ \tilde{0}\,\tilde{0} & \tilde{\eta}\tilde{\kappa} & \eta\kappa\end{pmatrix}^{0}_{00}, \tag{6.1}$$

which can be seen to be similar to the trivial 6*j* and 3*jm* symbol, although the one-dimensional irrep is placed to the far left as opposed to the right for the trivial 6*j* and 3*jm*. This is because the power of $\langle\tilde{0}\rangle, f-1$, is the largest of all the S_f irreps and the ordering of the 6*j* and 3*jm* symbols in the tables is based on the ordering of the triads. For our present calculation the Λ-matrices take a simple form

$$\Lambda(\gamma_1\gamma_2\gamma_3)^{00} = \{\sim,\gamma_1\gamma_2\gamma_30\},$$
$$\Lambda(\gamma,\eta\kappa)_{00} = (\sim,\gamma0\eta\kappa)\,. \tag{6.2}$$

The symmetries of the 6*j* and 3*jm* symbols imply the following constraints on the Λ-phases $\{\sim,\gamma_1\gamma_2\gamma_30\}$ and $(\sim,\gamma0\eta\kappa)$:

$$(\sim,\gamma_1\gamma_2\gamma_30) = \pm 1\,,$$
$$\{\sim,\gamma_1\tilde{\gamma}_2\tilde{\gamma}_30\}\{\sim,\gamma_1\gamma_2\gamma_30\} = \{\gamma_1\tilde{\gamma}_2\tilde{\gamma}_30\}\{\gamma_1\gamma_2\gamma_30\}\{\tilde{0}\tilde{\gamma}_3\gamma_30\}\,,$$
$$\{\sim,\gamma_1\tilde{\gamma}_3\tilde{\gamma}_20\} = \{\sim,\gamma_1\gamma_2\gamma_30\}\,,$$
$$(\sim,\gamma0\eta\kappa) = \pm 1\,, \tag{6.3}$$
$$(\sim,\gamma0\eta\kappa)(\sim,\gamma0\eta\kappa) = \{\tilde{0}\tilde{\gamma}\gamma0\}\{\tilde{0}\tilde{\eta}\eta0\}\{\tilde{0}\tilde{\kappa}\kappa0\}\,,$$
$$(\sim,\gamma0\kappa\eta)(\sim,\gamma0\eta\kappa) = (\tilde{\gamma}0\tilde{\eta}\tilde{\kappa})(\gamma0\eta\kappa)\,.$$

Unlike the 3*j* phase, the Λ-phase $\{\sim\gamma_1\gamma_2\gamma_30\}$ is in general dependent on the order of the irrep arguments. Using Eqs. (4.1) and (5.1) of Ref. 25 and the above, we obtain the transpose conjugate symmetry of the multiplicity-free 6*j* and 3*jm* symbols

$$\begin{Bmatrix}\gamma_1' & \gamma_2' & \gamma_3'\\ \tilde{\gamma}_1 & \tilde{\gamma}_2 & \tilde{\gamma}_3\end{Bmatrix}_{0000} = \begin{Bmatrix}\gamma_1' & \gamma_2' & \gamma_3'\\ \gamma_1 & \gamma_2 & \gamma_3\end{Bmatrix}_{0000}\{\tilde{0}\tilde{\gamma}_1\gamma_10\}\{\tilde{0}\tilde{\gamma}_2\gamma_20\}\{\tilde{0}\tilde{\gamma}_3\gamma_30\}$$
$$\times\{\sim,\gamma_1'\,\gamma_3\gamma_20\}\{\sim,\gamma_2'\,\gamma_1\gamma_30\}\{\sim,\gamma_3'\,\gamma_2\gamma_10\}$$
$$\times\{\gamma_1'\,\gamma_2\gamma_30\}\{\gamma_1\,\gamma_2'\,\gamma_30\}\{\gamma_1\gamma_2\,\gamma_3'0\}\,, \tag{6.4}$$

$$\begin{pmatrix}\tilde{\gamma}_1 & \tilde{\gamma}_2 & \gamma_3\\ 0 & 0 & 0\\ \tilde{\eta}_1\tilde{\kappa}_1 & \tilde{\eta}_2\tilde{\kappa}_2 & \eta_3\kappa_3\end{pmatrix}^{0}_{00} = \begin{pmatrix}\gamma_1 & \gamma_2 & \gamma_3\\ 0 & 0 & 0\\ \eta_1\kappa_1 & \eta_2\kappa_2 & \eta_3\kappa_3\end{pmatrix}^{0}_{00}$$
$$\times(\sim,\gamma_10\eta_1\kappa_1)(\sim,\gamma_20\eta_2\kappa_2)$$
$$\times\{\sim,\gamma_3\gamma_1\gamma_20\}\{\sim,\eta_3\eta_1\eta_20\}\{\sim,\kappa_3\kappa_1\kappa_20\}\,. \tag{6.5}$$

TABLE IV. The Λ-phase constraints.

Phase constraints for $f=4$

$\{\sim,1\,1\,1\,0\} = -\{\tilde{0}\,\tilde{1}\,1\,0\}$

$\{\sim,2\,1\,1\,0\} = -\{\tilde{0}\,\tilde{1}\,1\,0\}$

$\{\sim,1^2\,1\,1\,0\} = -\{\tilde{0}\,\tilde{1}\,1\,0\}$

$\{\sim,1^2\,2\,1\,0\}\{\sim,1\,2\,1\,0\} = -\{1\,2\,1\,0\}$

$\{\sim,1\,1^2\,2\,0\}\{\sim,1\,1\,1^2\,0\}\{\sim,1\,2\,1\,0\} = +\{\tilde{0}\,\tilde{2}\,2\,0\}$

$\{\sim,2\,1^2\,1\,0\}\{\sim,1\,1\,1^2\,0\} = -1$

$\{\sim,2\,1\,1^2\,0\}\{\sim,1\,1\,1^2\,0\} = -1$

$\{\sim,1^2\,1^2\,1\,0\}\{\sim,1\,1\,1^2\,0\} = -\{\tilde{0}\,\tilde{1}\,1\,0\}\{\tilde{0}\,\tilde{1}^2\,1^2\,0\}$

$\{\sim,1\,1^2\,1^2\,0\} = +\{\tilde{0}\,\tilde{1}^2\,1^2\,0\}\{\tilde{1}\,\tilde{1}^2\,1\,0\}$

$\{\sim,1^2\,1\,1^2\,0\}\{\sim,1^2\,1^2\,1\,0\}\{\sim,1\,1^2\,1^2\,0\} = -\{\tilde{0}\,\tilde{1}\,1\,0\}$

$\{\sim,1^2\,1^2\,1^2\,0\} = -\{\tilde{0}\,\tilde{1}^2\,1^2\,0\}\{\tilde{1}^2\,\tilde{2}\,1\,0\}\{1\,1^2\,2\,0\}$

Phase constraint for $f=5$

$\{\sim,1^2\,1^2\,1^2\,0\} = \{\tilde{0}\,\tilde{1}^2\,1^2\,0\}$

Phase constraints for all f

$\{\sim,0\,0\,0\,0\} = +\{\tilde{0}\,\tilde{0}\,0\,0\} = +1$

$\{\sim,1\,1\,0\,0\} = \{\sim,1\,0\,1\,0\} = \{\sim,0\,1\,1\,0\} = +\{\tilde{0}\,\tilde{1}\,1\,0\}$

$\{\sim,2\,2\,0\,0\} = \{\sim,2\,0\,2\,0\} = \{\sim,0\,2\,2\,0\} = +\{\tilde{0}\,\tilde{2}\,2\,0\}$

$\{\sim,1^2\,1^2\,0\,0\} = \{\sim,1^2\,0\,1^2\,0\} = \{\sim,0\,1^2\,1^2\,0\} = +\{\tilde{0}\,\tilde{1}^2\,1^2\,0\}$

$(\sim,0\,0\,0\,0) = +1$ for all (f_1f_2)

$(\sim,1\,0\,0\,1)(\sim,1\,0\,0\,0) = \begin{cases}-1 \text{ for } (f_1f_2) = (1,2)\\ +1 \text{ for } (f_1f_2) = (1,3),(2,2)\end{cases}$

$(\sim,2\,0\,0\,0)(\sim,1\,0\,0\,0) = -\{\sim,1\,2\,1\,0\}_f$ for $(f_1f_2) = (2,2)$

$(\sim,2\,0\,0\,1)(\sim,1\,0\,0\,0) = -\{\sim1\,2\,1\,0\}_f$ for $(f_1f_2) = (1,3)$

$(\sim,2\,0\,1\,1)(\sim,2\,0\,0\,0) = -1$ for $(f_1f_2) = (2,2)$

$(\sim,1^2\,0\,0\,1)(\sim,1\,0\,0\,1) = +\{\sim,1\,1^2\,1\,0\}_f$ for $(f_1f_2) = (1,2),\ (1,3),\ (2,2)$

$(\sim,1^2\,0\,1\,1)(\sim,1\,0\,1\,0) = +\{\sim,1\,1^2\,1\,0\}_f$ for $(f_1f_2) = (2,2)$

$(\sim,1^2\,0\,0\,1^2)_{(1,3)}\,(\sim,1^2\,0\,0\,1)_{(1,2)} = +(\sim,1^2\,0\,0\,1)_{(1,3)}\,(\sim1\,0\,0\,1)_{(1,2)}$

The transposition symmetry gives the following constraint

$(\sim,\gamma\,0\,\kappa\eta)(\sim,\gamma\,0\,\eta\kappa) = (\tilde{0}\,0\,\tilde{0}\,\tilde{0})(\tilde{\gamma}\,0\,\tilde{\eta}\tilde{\kappa})(\gamma\,0\,\eta\kappa)$ for all (f_1f_2).

We call the 6*j* and 3*jm* symbols appearing on the left-hand side as being "tilded." By using the row-flip symmetry of the 6*j* symbol to place one of the four triads at the top, four different tilded 6*j* symbols can be obtained from just one 6*j* symbol. Similarly employing the even permutational symmetry of the 3*jm* symbol, three different tilded 3*jm* symbols can be obtained from the one 3*jm* symbol. Such tilded symbols have the same f dependence though a sign difference may occur. Thus Eqs. (6.4) and (6.5) can be used to expand on the size of Tables II and III, respectively. Since we have considered only irreps up to power 2, the four tilded 6*j* symbols and the three tilded 3*jm* symbols are always distinct for $f \geqslant 6$ from their "untilded" 6*j* or 3*jm* symbols. However for $f<6$, the tilded symbol may be equivalent to a symbol already listed in the table. During the calculation of the listed 6*j* and 3*jm* symbols, phase freedoms were made such that phase consistencies are achieved between such "tilde" and "untilde" 6*j* and 3*jm* symbols. These considerations lead to constraints placed on the sign of the Λ-phases for $f<6$. These phase constraints are presented in Table IV. We note that some Λ-phases are f-dependent, for example,

$$\{\sim,0110\} = \{\tilde{0}\tilde{1}10\},$$

but for

$$f=3, \quad \{\tilde{0}\tilde{1}10\} = \{1^3 21\ 21\ 0\} = -1,$$

and

$$f=2, \quad \{\tilde{0}\tilde{1}10\} = \{1^2 21^2 0\} = +1.$$

In general it is not possible to make phase choices such that the Λ-phases are f independent.

VII. THE SCHUR-WEYL DUALITY SYMMETRY

Many authors have studied the connection between the symmetric groups and the unitary groups. For a review and the terminology used in this section, we refer the reader to Refs. 27 and 28. The important results established from this connection, which we call the Schur–Weyl duality, are relations between certain transformation factors of the symmetric groups and those of the unitary groups. Two relations that concern us here can be summarized as follows: (i) S_f recoupling factor$\simeq U_{p_1p_2p_3}$ resubduction factor, and (ii) $S_{f_1+f_2}$ coupling factor $\simeq U_{p_1p_2}$ coupling factor. The full expression of these relations is contained in Eqs. (4.4) and (4.10) of Ref. 28. In the multiplicity-free cases these relations reduce to the following (here we assume the use of standard partition labels):

$$\begin{aligned}
&\langle \lambda\, 0\lambda_1\lambda_{23}(0\lambda_2\lambda_3)\,|\,\lambda\, 0\lambda_{12}(0\lambda_1\lambda_2)\lambda_3\rangle_{p_1p_2p_3}\\
&\quad = \langle \lambda_1(\lambda_2\lambda_3)0\lambda_{23},0\lambda\,|\,(\lambda_1\lambda_2)0\lambda_{12},\lambda_3,0\lambda\,\rangle_f\\
&\quad\quad \times D_{p_1p_{23}}(\lambda,\lambda_1\lambda_{23})^* D_{p_2p_3}(\lambda_{23},\lambda_2\lambda_3)^*\\
&\quad\quad \times D_{p_{12}p_3}(\lambda,\lambda_{12}\lambda_3) D_{p_1p_2}(\lambda_{12},\lambda_1\lambda_2)\,,
\end{aligned}\tag{7.1}$$

$$\begin{aligned}
&{}^{\mathrm{U}_p}_{\mathrm{U}_{p_1}\mathrm{U}_{p_2}}\left\langle \begin{array}{c|c} \mu\quad\nu & \mu\nu \\ 0\quad 0 & 0 \\ \mu_1\mu_2\nu_1\nu_2 & \lambda \\ 00 & 0 \\ \lambda_1\lambda_2 & \lambda_1\lambda_2 \end{array}\right\rangle
= {}^{\mathrm{S}_f}_{\mathrm{S}_{f_1}\mathrm{S}_{f_2}}\left\langle \begin{array}{c|c} \lambda_1\lambda_2 & \lambda_1\lambda_2 \\ 00 & 0 \\ \mu_1\nu_1\mu_2\nu_2 & \lambda \\ 00 & 0 \\ \mu\nu & \mu\nu \end{array}\right\rangle\\
&\quad \times D_{p_1p_2}(\mu,\mu_1\mu_2)^* D_{p_1p_2}(\nu,\nu_1\nu_2)^* D_{p_1}(\mu_1\nu_1,\lambda_1)^*\\
&\quad \times D_{p_2}(\mu_2\nu_2,\lambda_2)^* D_p(\mu\nu,\lambda) D_{p_1p_2}(\lambda_1\lambda_2,\lambda)\,,
\end{aligned}\tag{7.2}$$

where $D_{.}(\cdots)$, called duality factors, are in general indexed by multiplicity labels and relate symmetric group and unitary group phase freedoms. Moreover the duality factors $D_{p_1p_2}(\lambda,\lambda_1\lambda_2)$ are zero unless $(\lambda_1\lambda_2\lambda r)$ forms a triad for S_f and $(\lambda a\lambda_1\lambda_2)$ forms a ket branching for $U_p \supset U_{p_1}\times U_{p_2}$. Similarly $D_p(\mu\nu,\lambda)$ is zero unless $(\lambda a\mu\nu)$ forms a ket branching for $S_f \supset S_{f_1}\times S_{f_2}$ and $(\mu\nu\lambda\ {}^*r)$ forms a triad in U_p. When these conditions are satisfied, the duality factors are elements of a unitary matrix, which in the multiplicity-free case reduces to a phase. This phase can be chosen to be unity. Inspecting Eqs. (7.1) and (7.2) we see that the left-hand side is independent of the symmetric group while the symmetric group recoupling and coupling factors are clearly independent of the unitary group. This point makes these relations powerful as a means of determining the unitary group transformation factors that appear in (7.1) and (7.2). Using now the relation between the recoupling and coupling factors with the more symmetrical $6j$ and $3jm$ symbols, respectively, the unitary group transformation factors can be written

$$\begin{aligned}
&\langle \lambda\, 0\lambda_1\lambda_{23}(0\lambda_2\lambda_3)\,|\,\lambda\, 0\lambda_{12}(0\lambda_1\lambda_2)\lambda_3\rangle_{p_1p_2p_3}\\
&\quad = [\,|\lambda_{12}\lambda_{23}|_f]^{1/2}\ {}_f\begin{Bmatrix} \lambda & \lambda_{23} & \lambda_1 \\ \lambda_2 & \lambda_{12} & \lambda_3 \end{Bmatrix}_{0000}\\
&\quad\quad \times \{\lambda\lambda_{23}\lambda_1 0\}_f \{\lambda_2\lambda_{23}\lambda_3 0\}_f\,,
\end{aligned}\tag{7.3}$$

$$\begin{aligned}
&{}^{\mathrm{U}_p}_{\mathrm{U}_{p_1}\mathrm{U}_{p_2}}\begin{pmatrix} \lambda^* & \mu & \nu \\ 0 & 0 & 0 \\ \lambda_1^*\lambda_2^* & \mu_1\mu_2 & \nu_1\nu_2 \end{pmatrix}^{0}_{00}\\
&\quad = {}^{\mathrm{S}_f}_{\mathrm{S}_{f_1}\mathrm{S}_{f_2}}\begin{pmatrix} \lambda & \lambda_1 & \lambda_2 \\ 0 & 0 & 0 \\ \mu\nu & \mu_1\nu_1 & \mu_2\nu_2 \end{pmatrix}^{0}_{00}\ {}^{\mathrm{U}_p}_{\mathrm{U}_{p_1}\mathrm{U}_{p_2}}\begin{pmatrix} \lambda \\ 0 \\ \lambda_1\lambda_2 \end{pmatrix}\\
&\quad\quad \times \left[\frac{|\lambda\,|_f}{|\mu|_{f_1}|\nu|_{f_2}}\cdot\frac{|\lambda_1|_{p_1}|\lambda_2|_{p_2}}{|\lambda\,|_p}\right]^{1/2}
\end{aligned}\tag{7.4}$$

(the unity choice for the duality factors is assumed). Combining the algebraic formulas for the symmetric group $6j$ and $3jm$ symbols with Eqs. (7.3) and (7.4), we can then obtain algebraic expressions for certain unitary group resubduction factors and $3jm$ symbols that are labeled by irreps $\{f\}$, $\{f-1,1\}$, $\{f-2,2\}$, and $\{f-2,1^2\}$. In addition the transpose conjugate symmetry extends the list of these unitary group transformation factors to include those with labels $\{1^f\}$, $\{2,1^{f-2}\}$, $\{2,2,1^{f-4}\}$, and $\{3,1^{f-3}\}$. We illustrate this method in the following way:

$$\begin{aligned}
&\langle\{f-2,1^2\}0\{f-1,1\}\{f-2,2\}(0\{f-1,1\}\{f-1,1\})\,|\,\{f-2,1^2\}0\{f-2,2\}(0\{f-1,1\}\{f-1,1\})\{f-1,1\}\rangle_{p_1p_2p_3}\\
&\quad = {}_f\begin{Bmatrix} [f-2,1^2] & [f-2,2] & [f-1,1] \\ [f-1,1] & [f-2,2] & [f-1,1] \end{Bmatrix}_{0000}\\
&\quad\quad \times [\,|[f-2,2]|,|[f-2,2]|\,]^{1/2} \times \{[f-2,1^2][f-2,2][f-1,1]0\}_f \{[f-1,1][f-2,2][f-1,1]0\}_f\\
&\quad = +\frac{f-4}{f(f-2)\,(f-3)}\cdot\frac{f(f-3)}{2}\cdot\{[f-2,1^2][f-2,2][f-1,1]0\}_f = \frac{1}{2}\cdot\frac{f-4}{f-2}\{\langle 1^2\rangle\langle 2\rangle 0\}\,,
\end{aligned}$$

$$\begin{array}{c}U_p\\ \\ U_{p_1}U_{p_2}\end{array}\begin{pmatrix} \{f-2,2\}^* & \{f_1-1,1\} & \{f_2-1,1\}\\ 0 & 0 & 0\\ \{f-2,2\}^*\times\{f-1,1\}^* & \{f_1-1,1\}\times\{f_1\} & \{f_2-1,1\}\times\{f_2\}\end{pmatrix}^{0}_{00}$$

$$= \begin{array}{c}S_f\\ \\ S_{f_1}S_{f_2}\end{array}\begin{pmatrix} [f-2,2] & [f-2,2] & [f-1,1]\\ 0 & 0 & 0\\ [f_1-1,1]\times[f_2-1,1] & [f_1-1,1]\times[f_2-1,1] & [f_1]\times[f_2]\end{pmatrix}^{0}_{00}$$

$$\times \begin{array}{c}U_p\\ \\ U_{p_1}U_{p_2}\end{array}\begin{pmatrix} \{f-2,2\}\\ 0\\ \{f-1,2\}\times\{f-1,1\}\end{pmatrix}\left[\frac{|[f-2,2]|_f}{|[f-1,1]|_{f_1}\,|[f-1,1]|_{f_2}}\times\frac{|\{f-2,2\}|_{p_1}\,|\{f-1,1\}|_{p_2}}{|\{f-2,2\}|_p}\right]^{1/2}$$

$$= +\sqrt{\frac{(f-2)(f_1-1)(f_2-1)}{f(f-1)(f-3)(f-4)f_1f_2}}\,(f_1-f_2)\cdot\sqrt{\frac{f(f-3)}{2\cdot(f_1-1)(f_2-1)}}\;\begin{array}{c}U_p\\ \\ U_{p_1}U_{p_2}\end{array}\begin{pmatrix} \{f-2,2\}\\ 0\\ \{f-1,2\}\{f-1,1\}\end{pmatrix}$$

$$\times\sqrt{\frac{(p_1+f-3)!}{(p_1-2)!}p_1\cdot\frac{f-3}{(f-1)!2!}\cdot\frac{(p_2+f-2)!}{(p_2-2)!}\cdot\frac{f-1}{f!}\cdot\frac{(p-2)!}{(p+f-3)!}\cdot\frac{(f-1)!2!}{(f-3)p}}$$

$$= \begin{array}{c}U_p\\ \\ U_{p_1}U_{p_2}\end{array}\begin{pmatrix} \{f-2,2\}\\ 0\\ \{f-1,2\}\{f-1,1\}\end{pmatrix}(f_1-f_2)\cdot\sqrt{\frac{(f-2)}{(f-4)\,p_2\cdot f_1f_2\cdot 2\cdot f!}\,\frac{(p_1+f-3)!}{(p_1-2)!}\,\frac{(p_2+f-2)!}{(p_2-2)!}\,\frac{(p-2)!}{(p+f-3)!}}\,.$$

VIII. CONCLUDING REMARKS

In presenting the tables of the symmetric group $6j$ and $3jm$ symbols, we have demonstrated that an algebraic approach to their determination can be performed. Clearly such an approach is preferrable not only in saving space in presenting tables but also in showing the explicit rank dependence and vanishing of the $6j$ and $3jm$ symbols. In some circumstances a partial understanding of these vanishings can be given and may be connected to either the transposition symmetry or the transpose conjugate symmetry or even the modification rules of the irrep labels themselves as sug-

TABLE V. The $S_5 \supset A_5$ embedding. We use the f-independent notation throughout.

Branching Rules

$0 \downarrow 0$	$1 \downarrow 1$	$2 \downarrow 2$	$1^2 \downarrow 1^2_+ + 1^2_-$
$\tilde{0} \downarrow 0$	$\tilde{1} \downarrow 1$	$\tilde{2} \downarrow 2$	

Note that 1^2 is self-conjugate in S_5, i.e., $\tilde{1}^2 \simeq 1^2$. Butler's notation for the icosahedral group irreps differs:

f-independent	0	1	2	1^2_+	1^2_-
Butler	0	3	2	1	$\tilde{1}$

$3jm$ Symbols

$\begin{pmatrix}0&0&0\\0&0&0\end{pmatrix}^0_0 = +1$	$\begin{pmatrix}2&2&2\\2&2&2\end{pmatrix}^0_0 = 0$	$\begin{pmatrix}1^2&2&0\\1^2_-&2&2\end{pmatrix}^0_0 = -\frac{1}{2}$
$\begin{pmatrix}1&1&0\\1&1&0\end{pmatrix}^0_0 = +1$	$\begin{pmatrix}2&2&2\\2&2&2\end{pmatrix}^0_1 = +1$	$\begin{pmatrix}1^2&1^2&0\\1^2_+&1^2_+&0\end{pmatrix}^0_0 = +\frac{1}{2}$
$\begin{pmatrix}1&1&1\\1&1&1\end{pmatrix}^0_0 = +1$	$\begin{pmatrix}1^2&1&1\\1^2_+&1&1\end{pmatrix}^0_0 = +\frac{1}{2}$	$\begin{pmatrix}1^2&1^2&0\\1^2_-&1^2_-&0\end{pmatrix}^0_0 = +\frac{1}{2}$
$\begin{pmatrix}2&1&1\\2&1&1\end{pmatrix}^0_0 = +1$	$\begin{pmatrix}1^2&1&1\\1^2_-&1&1\end{pmatrix}^0_0 = +\frac{1}{2}$	$\begin{pmatrix}1^2&1^2&1\\1^2_-&1^2_+&1\end{pmatrix}^0_0 = +\frac{1}{2}$
$\begin{pmatrix}2&2&0\\2&2&0\end{pmatrix}^0_0 = +1$	$\begin{pmatrix}1^2&2&1\\1^2_+&2&1\end{pmatrix}^0_0 = +\frac{1}{2}$	$\begin{pmatrix}1^2&1^2&1^2\\1^2_+&1^2_+&1^2_+\end{pmatrix}^0_0 = -\frac{1}{2}$
$\begin{pmatrix}2&2&1\\2&2&1\end{pmatrix}^0_0 = 0$	$\begin{pmatrix}1^2&2&1\\1^2_-&2&1\end{pmatrix}^0_0 = +\frac{1}{2}$	$\begin{pmatrix}1^2&1^2&1^2\\1^2_-&1^2_-&1^2_-\end{pmatrix}^0_0 = -\frac{1}{2}$
$\begin{pmatrix}2&2&1\\2&2&1\end{pmatrix}^0_1 = +1$	$\begin{pmatrix}1^2&2&2\\1^2_+&2&2\end{pmatrix}^0_0 = -\frac{1}{2}$	

gestive of the determination of the dimension formulas. Such symmetry considerations may give the algebraic form of the symmetric group $6j$ and $3jm$ symbols given the irrep label constituents. However, more study is required to fully understand and draw out these connections.

Although our calculation has only considered nonsimple phase irreps and multiplicity-free $6j$ and $3jm$ symbols, the method employed can be applied to the more general cases as exemplified by the numerical examples[17–20] and the algebraic example of the unitary groups.[21,22] With a table of such symmetric group $6j$ and $3jm$ symbols, the relationship via the Schur–Weyl duality with the unitary groups can be explored further, as can the relationship with other compact continuous groups.[29]

The tables have been checked with numerical tables produced by Butler.[8] The symmetric group S_4 is isomorphic to the tetrahedral group T_d, while S_5 contains the alternating group A_5, which is isomorphic to the icosahedral group. Our $6j$ formulas agree with the corresponding $6j$ symbols to within phase choices. In the S_5 case, the $6j$ symbols of S_5 and those of A_5 are related by the Wigner relation (see Butler[6], Eq. 3.3.29), which requires some $S_5 \supset A_5$ $3jm$ symbols. These are given in Table V. The group reduction $T_d \supset C_{3v}$, which is isomorphic to $S_4 \supset S_3 \times S_1$, was used to give a check of the $3jm$ formulas.

ACKNOWLEDGMENT

We wish to acknowledge financial support from the Austrian FWF (Fonds zur Förderung der wissenschaftlichen Forschung) Project No. P5170.

[1]H. A. Jahn, Proc. R. Soc. London Ser. **201**, 516 (1950); Phys. Rev. **96**, 989 (1954); J. P. Elliott, J. Hope, and H. A. Jahn, Philos. Collect. R. Soc. London A **246**, 241 (1953); I. G. Kaplan, Sov. Phys. JETP **14**, 401, 568 (1962); H. Horie, J. Phys. Soc. Jpn. **19**, 1783 (1964); P. Kramer, Z. Phys. **205**, 181 (1967); **216**, 68 (1967); V. V. Vanagas and A. K. Petrauskas, Yad. Fiz. **5**, 555 (1967); Sov. J. Nuc. Phys. **5**, 393 (1967); V. V. Vanagas, *Algebraic Methods in Nuclear Theory* (Vilnius, Mintis, 1971) (in Russian); and P. Kramer, G. John, and D. Schenzle, "Group Theory and the Interaction of Composite Nuclear Systems," in *Clustering Phenomena in Nuclei*, Vol. 2 (Vieweg, Braunschweig, 1981).

[2]M. Hamermesh, *Group Theory and its Applications to Physical Problems* (Addison–Wesley, Reading, MA 1962).

[3]S. Schindler and R. Mirman, J. Math. Phys. **18**, 1678, 1697 (1977).

[4]J-Q. Chen, F. Wang, and M. J. Gao, Acta Phys. Sin. Abstr. **26**, 307 (1977) [Chin. J. Phys. Peking Engl. Transl. **1**, 533 (1981)].

[5]R. W. Haase and P. H. Butler, J. Math. Phys. **26**, 1493 (1985).

[6]P. H. Butler, *Point Group Symmetry Applications; Methods and Tables* (Plenum, New York, 1981).

[7]F. D. Murnaghan, Proc. Natl. Acad. Sci. USA **37**, 55 (1951); An. Acad. Bras. Cienc. **23**, 141 (1951); Proc. Natl. Acad. Sci. USA **41**, 396, 514, 1096 (1955).

[8]D. E. Littlewood, Can. J. Phys. **10**, 1, 17 (1958).

[9]P. H. Butler and R. C. King, J. Math. Phys. **14**, 1176 (1973).

[10]G. de B. Robinson, *Representation Theory of the Symmetric Group* (University of Toronto, Toronto, 1961).

[11]D. E. Littlewood, *The Theory of Group Characters* (Oxford University, London, 1940).

[12]B. G. Wybourne, *Symmetry Principles and Atomic Spectroscopy With an Appendix of Tables by P. H. Butler* (Wiley, New York, 1970).

[13]C. Itzykson and M. Nauenberg, Rev. Mod. Phys. **38**, 95 (1966).

[14]P. H. Butler, Philos. Collect. R. Soc. London **277**, 545 (1975).

[15]R. P. Bickerstaff and B. G. Wybourne, J. Phys. G **7**, 275, 995 (E) (1981).

[16]P. H. Butler and B. G. Wybourne, Int. J. Quant. Chem. **X**, 615 (1976).

[17]R. P. Bickerstaff, P. H. Butler, M. B. Butts, R. W. Haase, and M. F. Reid, J. Phys. A **15**, 1087 (1983).

[18]P. H. Butler, R. W. Haase, and B. G. Wybourne, Aust. J. Phys. **31**, 131 (1978); **32**, 137 (1979).

[19]M. F. Reid and P. H. Butler, J. Phys. A **13**, 2887 (1980).

[20]M. F. Reid, J. Phys. A **17**, 1755 (1984).

[21]P. H. Butler, Int. J. Quant. Chem. **10**, 599 (1976).

[22]R. W. Haase and P. H. Butler, J. Phys. A: Math. Gen. **19**, 1 (1986).

[23]P. H. Butler and B. G. Wybourne, At. Data **3**, 133 (1971).

[24]J. S. Griffith, *The Irreducible Tensor Method for Molecular Symmetry Groups* (Prentice–Hall, Englewood Cliffs, NJ, 1962).

[25]P. H. Butler and A. M. Ford, J. Phys. A **12**, 1357 (1979).

[26]R. W. Haase and R. Dirl, "The symmetric group," Technische Universität, Wien preprint.

[27]R. W. Haase and P. H. Butler, J. Phys. A **17**, 47 (1984).

[28]R. W. Haase and P. H. Butler, J. Phys. A. **17**, 61 (1984).

[29]R. W. Haase, P. H. Butler, and R. Dirl, in *Proceedings of the 13th International Colloquium on Group Theoretical Methods in Physics*, Maryland, May 1984, edited by W. W. Zachary (World Scientific, Singapore 1984).

Extension and replacement bases for semisimple Lie algebras

Richard H. Capps
Department of Physics, Purdue University, West Lafayette, Indiana 47907

(Received 5 August 1985; accepted for publication 13 November 1985)

Two simple prescriptions are given for obtaining sets of orthogonal bases for semisimple Lie algebras. The first method allows one to obtain the irreducible representations of all the simple Lie algebras, starting from SU(2) and Dynkin's method for constructing representations from simple roots and highest weights. The second method relates algebras of the same rank. Several examples are discussed. A method is given for listing the bases that may be obtained from the prescriptions.

I. INTRODUCTION

The purpose of this paper is to present two straightforward prescriptions for constructing orthogonal bases for complex semisimple Lie algebras. One of the goals is to provide tools useful for treating the exceptional groups E_8, E_7, E_6, F_4, and G_2; recent developments in string theory imply that some of these groups may be important in particle theory. Different bases are useful for different models, or different symmetry-breaking schemes. Therefore, our aim is not to exhibit two or three bases for each group, but rather to present prescriptions that may be used to construct bases with different desired properties.

The most successful general procedure for constructing irreducible representations of simple Lie algebras is based on techniques introduced by Dynkin in the 1950's.[1] If these techniques are used with a specific orthogonal basis, the ease of calculation depends not only on the general nature of the basis but also on the choice of simple roots, which depends on the definition of positivity. One advantage of our procedure is that the simple root sets are easy to handle.

The first prescription is developed and illustrated in Secs. II–V. The method is constructive; one may obtain the roots and representations of all simple Lie algebras with no prior knowledge except the properties of SU(2) and Dynkin's method for constructing representations from simple roots. The second prescription is developed and illustrated in Sec. VI. In Sec. VII methods are given for listing all the bases that may be constructed from these prescriptions. Of course some of the resulting bases are familiar. However, some are not familiar.

Dynkin introduced two nonorthogonal bases. These are not needed to apply the prescriptions of this paper. However, they are used to prove the validity of various parts of the construction procedures.

II. THE EXTENSION PRESCRIPTION

We give here a prescription for extending an orthogonal basis of a semisimple Lie algebra of rank n to a basis for an algebra of rank $n+1$. The terms and concepts used may be found in standard references.[2–5] One starts with any orthogonal basis for the original algebra H. Positivity is defined by the criterion that a weight vector is positive if and only if its first (starting from the left) nonzero component is positive. The roots are the weights of the adjoint representation. A simple root is a positive root that cannot be written as a sum of positive roots. One determines the n linearly independent simple roots of H.

Next one considers some irreducible representation of H other than the adjoint, and finds the most negative weight vector W_-. If the ratio of the length of W_- to the relevant root length L satisfies the inequality

$$|W_-/L| < (1/M)^{1/2}, \tag{2.1}$$

where M is a positive integer, then root systems for at least M different algebras of rank $n+1$ may be constructed. If H is simple the relevant root length is the length of the shortest simple root of H that is not perpendicular to W_-. The definition of L for semisimple (but not simple) algebras is postponed to Sec. IV. A new orthogonal dimension is added to the original n dimensions; the component of the new dimension is placed on the extreme left. The roots of H have zero components in the new dimension. A new root is formed by placing the components of W_- in positions 2 to $n+1$, and adding to this vector a positive first component of magnitude $[(1/M)L^2 - W_-^2]^{1/2}$. If $M > 1$, one may follow this prescription for M and also for all smaller positive integers. The new set of $n+1$ vectors are the simple roots for a Lie algebra of rank $n+1$, using the same definition of positivity.

If $M > 1$, the above procedure is called the "short-root procedure." An alternate, "long-root procedure" is to determine the new root as discussed above, and then multiply it by M. This procedure also generates a Lie algebra, generally different from that obtained by the short-root procedure.

Before demonstrating the validity of these rules, we illustrate the method for the case where H is SU(2). In order to be consistent throughout the paper we always normalize so that the longest root of H is of length $\sqrt{2}$, even though this is not the natural normalization for the SU(2) case. The simple root (state of $j = j_z = 1$) is $\sqrt{2}$. Only spin-0 and spin-$\frac{1}{2}$ states are short enough to be used in the extension. Clearly, the spin-0 representation may be used to generate SU(2) $\otimes$ SU(2). If the spin-$\frac{1}{2}$ representation is used, $W_- = \sqrt{2}(-\frac{1}{2})$. This satisfies Eq. (2.1) with $M = 1, 2$, or 3. Since there are only two roots in the extended system, the short- and long-root procedures are equivalent, so we consider only the short-root procedure. The extended root systems of $M = 1,2$, and 3 are, respectively,

$$\begin{array}{lll} \sqrt{2}\{0\ 1\}, & \sqrt{2}\{0\ 1\}, & \sqrt{2}\{0\ 1\}, \\ \sqrt{2}\{(\frac{3}{4})^{1/2}\ -\frac{1}{2}\}, & \sqrt{2}\{\frac{1}{2}\ -\frac{1}{2}\}, & \sqrt{2}\{(\frac{1}{12})^{1/2}\ -\frac{1}{2}\}. \end{array}$$

 0022-2488/86/040914-10$02.50

If one classifies these three pairs of vectors by the symbol $(M^{1/2},\theta)$, where $M^{1/2}$ is the length ratio and θ the angle between the vectors, the results are $(1, 120°)$, $(\sqrt{2}, 135°)$, and $(\sqrt{3}, 150°)$. These are the magnitude and angle relations required for simple root sets of the algebras SU(3), B_2, and G_2, respectively.[4,5]

Next, we will prove that the short-root construction procedure is always valid. (Extending the proof to the long-root case is not difficult.) We call two roots "connected" if they are not perpendicular. Dynkin's classification theorem may be stated in the following way: a set of vectors R_i may be taken as the simple roots for a semisimple Lie algebra if and only if the following three conditions are satisfied.

(i) The vectors are linearly independent.

If two vectors R_1 and R_2 are connected, and if R_2 is not longer than R_1, then

(ii) $(R_2^2/R_1^2) = 1/M$,

where M is a positive integer, and

(iii) $R_1 \cdot R_2 = -R_1^2/2$.

In practice, $M<4$, because if $M>4$, condition (iii) is impossible, and if $M=4$ the vectors are linearly dependent.[6] The phrase "may be taken as the simple roots" needs clarification. One may have a criterion for positivity such that the set of vectors are not all positive. One could still use these vectors to generate a Lie algebra, but the simple roots would not be the original roots. However, one can always redefine positivity in such a way that the original vectors are all positive, in which case they are the simple roots.

The construction is such that condition (ii) is satisfied explicitly. Furthermore, since the added root is the only one with a component in the added dimension, linear independence of the original roots guarantees the linear independence of the extended root set. Thus the validity demonstration reduces to showing that condition (iii) is satisfied by the new root. The two nonorthogonal bases of Dynkin will be used for this purpose. We list below some properties of these bases.[7]

The components $\bar{a}_i$ of a vector (a) in the dual basis are defined by writing (a) as a linear combination of the simple roots R_i, i.e.,

$$a = \sum_i \bar{a}_i \frac{2}{R_i^2} R_i. \tag{2.2}$$

It is seen from this equation that the dual-basis components of a simple root R_j are

$$(\bar{R}_j)_i = \delta_{ij}(R_i^2/2). \tag{2.3}$$

The components of the vector (a) in the Dynkin basis are denoted by a_j. They may be determined from the dual components by

$$a_j = \sum_i \bar{a}_i \frac{2}{R_i^2} A_{ij}, \tag{2.4}$$

where A is the Cartan matrix, with elements

$$A_{ij} = 2R_i \cdot R_j / R_j^2. \tag{2.5}$$

Since the two bases are dual to each other, a dot product is written most simply by expanding one vector in each basis, i.e.,

$$a \cdot b = \sum_i a_i \bar{b}_i. \tag{2.6}$$

Dynkin has shown that the indices a_i for all weights in all irreducible representations (irreps) are integers. The Dynkin indices of the most positive weight W_+ of an irrep are non-negative integers; these are used to denote the representation. Thus, from Eqs. (2.6) and (2.3),

$$W_+ \cdot R_j = \sum_i a_i (\bar{R}_j)_i = a_j \frac{R_j^2}{2}. \tag{2.7}$$

The basic irreps ψ_k are those with Dynkin indices,

$$a_i(W_{+,k}) = \delta_{ik}, \tag{2.8}$$

where $W_{+,k}$, the highest weight of ψ_k, is called a basic weight. For the basic weights, Eq. (2.7) reduces to

$$W_{+,k} \cdot R_j = \delta_{jk}(R_j^2/2). \tag{2.9}$$

This is called the basic weight equation in later sections.

The conjugate of a weight vector is the negative of the vector. The irrep conjugate to ψ (denoted by ψ^c) is such that

$$W_-(\psi) = -W_+(\psi^c), \tag{2.10}$$

where W_- is the most negative weight of the representation.

We now return to the justification of the extension procedure. If the algebra H is simple it can be shown that the only irreps with sufficiently short W_- to satisfy Eq. (2.1) are the basic irreps; if H is semisimple (Sec. IV) the only irreps with sufficiently short W_- are direct products of basic irreps. In either case the Dynkin indices are all zero or one. The Dynkin indices of the conjugate representation must also be zero or one. It follows from Eq. (2.10) that the Dynkin components of W_- are all zero or minus one. The W_- is connected to a root R_j only if the index a_j of W_- is minus one, in which case Eqs. (2.6) and (2.3) lead to the result

$$W_- \cdot R_j = -R_j^2/2.$$

If W_- is extended by the short-root procedure to an orthogonal dimension, the dot product is unchanged and the new root is not longer than the connecting old root, so condition (iii) is satisfied.

The simple roots of H are positive. By construction, the new root is positive. Hence, the new root set are the simple roots of a new algebra, with an unchanged definition of positivity. This completes the demonstration.

III. OUTLINE OF SOME EXTENSIONS OF SIMPLE ALGEBRAS

In this section we outline the construction of some extension bases. Two specific examples are discussed in detail in Sec. V. Many authors have listed sets of useful orthogonal bases for Lie algebras.[8] If positivity is defined appropriately, some of these coincide with bases obtainable from the prescriptions of this paper.

For convenience we list the Dynkin diagrams for the four infinite Cartan classes and the five exceptional simple algebras in Fig. 1. The root-numbering conventions of this paper are shown in the figure.

We introduce some convenient terminology. The length of an irrep is defined as the length of the longest weights in the representation. (This is the length of W_+ or W_-.) A

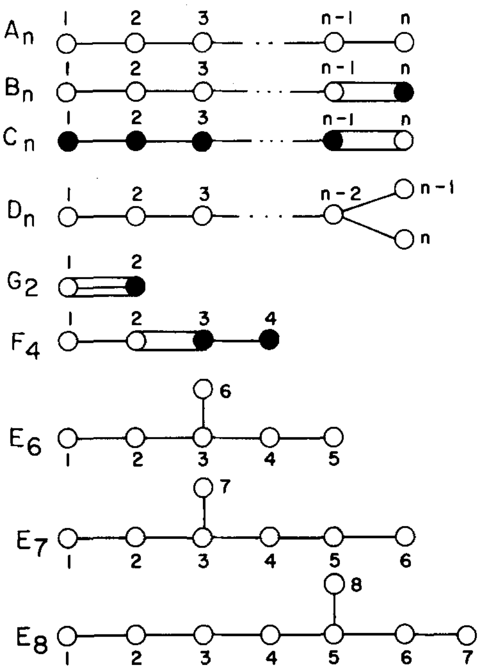

FIG. 1. Dynkin diagrams for simple Lie algebras. Black circles represent shorter roots.

short irrep is one that is shorter than the relevant root length. An equal-magnitude representation is one in which all weights have the same length. An equal-magnitude algebra is one for which all nonzero roots have the same length.

A. Extending SU*(n)*

Since every simple Lie algebra except F_4 may be obtained from a one-root extension of some $SU(n)$, $SU(n)$ bases are particularly useful. We list here some basic properties of $SU(n)$ quarks. We label specific quarks by the early letters *a,b,c*, etc., and use *q,r,s*, etc. to refer to general quarks. $SU(n)$ quarks may be defined as a set of n vectors of equal length in an $(n-1)$-dimensional Euclidean space subject to the conditions that the angles between all quark pairs are the same, and that the sum of the n different quark vectors is zero. These conditions determine that the ratio $(q\cdot r)/q^2$ is equal to $-1/(n-1)$, where q and r are any two different quarks. If the vector set is normalized so that the nonzero roots $(q\bar{r})$ are of length $\sqrt{2}$, then the quark and antiquark lengths and dot products are given by

$$q^2=(n-1)/n,\quad q\cdot r=-1/n,\quad \bar{q}=-q. \qquad (3.1)$$

Although the quarks are not orthogonal, Eq. (3.1) is so simple that it is convenient to write representation weights in terms of the quarks, using a specific orthogonal basis only when necessary. From Eq. (3.1), the dot products of quarks and roots are

$$q\cdot(q\bar{r})=1,\quad r\cdot(q\bar{r})=-1,\quad s\cdot(q\bar{r})=0, \qquad (3.2)$$

where *q,r*, and *s* are any different quarks.

We label the quarks *a,b*, etc. according to positivity, with a the most negative quark. The positive roots are those whose antiquark label is an earlier letter than the quark label, i.e., $(d\bar{b})$. The simple roots are those with adjacent letters. The Dynkin diagram is shown in Fig. 2.

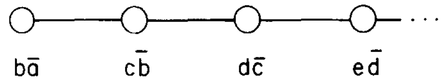

FIG. 2. Some simple $SU(n)$ roots and their connections.

A few comments concerning this result are in order. Positivity may be defined in many different ways. However, it would not be sufficient to assign the quarks different real numbers, arbitrary except for the requirement that the sum of the numbers is zero. If this were done it might occur that two different nonzero roots had the same positivity, in which case the definition would not be legitimate. The point of the result of Fig. 2 is that if the positivity definition is legitimate, all one needs to know is the ranking of the quarks in order to determine the simple roots and construct the irreps. The basic reason for this apparent paradox is that in the construction procedure, there is no significance to the relative positivity of two weights, unless one may be obtained by adding one or more simple roots to the other.[9]

Any irrep with a Dynkin index larger than one has states containing two or more identical quarks and so is not a short representation. We may limit attention to the basic irreps, totally antisymmetric states of quarks, or of antiquarks. These are all equal-magnitude irreps. The length-squared of an antisymmetric j-quark state is $j(q^2)+j(j-1)(q\cdot r)$. If we use Eqs. (3.1) we obtain the result

$$|\psi_j|=|\psi_{n-j}|=[j(n-j)/n]^{1/2}. \qquad (3.3)$$

If the antisymmetric quark state $(rs\cdots)$ contains the r and s quarks but not the q quark, its dot product with the roots is

$$(rs\cdots)\cdot(q\bar{r})=-1,\quad (rs\cdots)\cdot(r\bar{s})=0,\text{ etc.} \qquad (3.4)$$

We number the roots from the left in Fig. 2 and illustrate this convention by considering the antisymmetric two-quark and two-antiquark states of $SU(5)$. The highest weights in these two irreps are (de) and $(\overline{ab})$. It may be seen from the basic weight equation, Eq. (2.9), and from Eqs. (3.4) that these are the highest states of ψ_{n-2} and ψ_2, respectively.

We now list the possible extensions of $SU(n)$ for $n>3$. The relevant root length is $\sqrt{2}$. Clearly, extending $SU(n)$ by using the identity representation yields $SU(n)\otimes SU(2)$. In future examples we will not consider this obvious type of extension. If one uses the quark representation ψ_{n-1}, the lowest state is (a), which connects to the first root of Fig. 2. Since the quark length is $[(n-1)/n]^{1/2}$, the M values allowed by Eq. (2.1) are 1 and 2. If one uses the $M=1$ procedure, the resulting algebra is $SU(n+1)$. The $M=2$ short- and long-root procedures yield, respectively, B_n and C_n (see Ref. 10).

Next we consider the antisymmetric states ψ_j and ψ_{n-j}, where $j>1$. We avoid duplication by requiring that $n>2j$. In the case $j=2$, the lowest state of ψ_{n-2} is (ab). It is seen from Eq. (3.4) that this connects only with the second root $(c\bar{b})$ of Fig. 2. From Eqs. (3.3) and (2.1), only $M=1$ is possible. The algebra generated is $D_n\,[SO(2n)]$.

The lowest state of ψ_{n-3} is (abc), which connects only to the third root $(d\bar{c})$ in Fig. 2. The ratio M must be one. From Eq. (3.3), the lengths of the W_- for $SU(6)$, $SU(7)$, $SU(8)$, and $SU(9)$ are, respectively,

$(\frac{3}{2})^{1/2}$, $(\frac{12}{7})^{1/2}$, $(\frac{15}{8})^{1/2}$, and $\sqrt{2}$. Thus, this type extension of SU(9) cannot be made; the other three cases yield E_6, E_7, and E_8.

If $j > 3$ and $n \geqslant 2j$, the length given by Eq. (3.3) is never less than $\sqrt{2}$, so no extension is possible.

B. Extending D_n

We use a standard orthogonal basis for D_n [SO($2n$)]. The nonzero roots have two nonzero components, each of unit magnitude, and the weights of the vector representation ψ_1 have one nonzero element, of unit magnitude. We use the symbols 1_+3_+ and 1_+3_- to denote the roots with components $1\ 0\ 1 \cdots$ and $1\ 0\ -1 \cdots$, and the symbols 1_+ and 1_- to denote the vector states $1\ 0\ 0 \cdots$ and $-1\ 0\ 0 \cdots$. In the basic spinor representations ψ_{n-1} and ψ_n, each component is either $\frac{1}{2}$ or $-\frac{1}{2}$; we use only the $+$ and $-$ signs in the weight symbols. These two irreps are distinguished by the sign of the product of all the components. We use SO(10) as an example. It may be shown that the simple roots are 1_+2_-, 2_+3_-, 3_+4_-, 4_+5_-, and 4_+5_+. The lowest weights of the two spinor irreps are $(-\ -\ -\ -\ +)$ and $(-\ -\ -\ -\ -)$.

The smallest D_n that is different from an SU(l) or product of SU(l)'s is D_4. For any D_n with $n \geqslant 4$, only $M = 1$ extensions are possible. Extending with the vector representation yields D_{n+1} in the standard basis and so is not very useful. For D_5, D_6, and D_7, the length-squared of a spinor representation is $\frac{5}{4}$, $\frac{3}{2}$, and $\frac{7}{4}$, respectively. Extending with these representations yields the algebras E_6, E_7, and E_8.

C. Extending B_n, with $n > 2$

For B_n [SO($2n+1$)] there is a standard orthogonal basis that is similar to that of D_n. We illustrate with the specific case of B_3. There are twelve long roots $(\pm 1\ \pm 1\ 0)$, $(\pm 1\ 0\ \pm 1)$, and $(0\ \pm 1\ \pm 1)$, and six nonzero short roots $(\pm 1\ 0\ 0)$, $(0\ \pm 1\ 0)$, and $(0\ 0\ \pm 1)$. The vector representation ψ_1 has one zero weight, and six other weights that are equal to the six short roots. The eight spinor states $(\pm\frac{1}{2}, \pm\frac{1}{2}, \pm\frac{1}{2})$ are all in the same irreducible representation. In the notation of the preceding subsection the simple roots of B_3 are 1_+2_-, 2_+3_-, and 3_+.

For any B_n the lowest state of the vector representation ψ_1 is 1_-. Since this connects only with a long root (1_+2_-), the relevant root length in Eq. (2.1) is $\sqrt{2}$, so an $M = 1$ extension may be made. This yields B_{n+1} in the standard basis and so is not very useful. However, this example is interesting because it turns out that the $B_n \to B_{n+1}$ and $C_3 \to F_4$ cases are the only extensions of simple algebras that may be made with non-equal-magnitude irreps.

The length-squared of the spinor representation ψ_n is $n/4$. In the lowest state all components are $-\frac{1}{2}$. Since this connects with the short root, the relevant root length is 1. Therefore, a ψ_n extension is possible only with $M = 1$, and only when n is 2 or 3. The resulting algebras are C_3 and F_4.

D. Extending exceptional algebras

The lengths of the basic irreps ψ_1 and ψ_5 of E_6 are $(\frac{4}{3})^{1/2}$, where the roots are labeled as in Fig. 1. Extending with one of these representations yields E_7. The length of the representation ψ_6 of E_7 is $(\frac{3}{2})^{1/2}$. Extending with this representation yields E_8. There are no short representations (other than the identity) for E_8, F_4, and G_2. These are not extendable.

IV. EXTENSIONS OF SEMISIMPLE ALGEBRAS

In this section we consider some cases where the original algebra H is not simple, but is the direct product of simple constituents. The final algebra G is simple. The extending (or connecting) representation is a direct product of basic irreps of each constituent. We must make some rules concerning the relative normalizations of the roots of the different constituents, and generalize the definition of Sec. II of the relevant root length.

No more than one of the constituents may have nonzero roots of different lengths. If such a constituent is present, the relevant root length L is the length of the root of this constituent that is connected to the extending representation. The roots of the other constituents must be taken equal to L. In this case only the $M = 1$ procedure is possible; the connecting representation is extended to length L.

If all the constituents of H are equal-magnitude algebras, the $M = 1$ procedure is to set all root lengths equal to $\sqrt{2}$ and extend the connecting representation to length $\sqrt{2}$.

If there are only two constituents, and both are equal-magnitude algebras, $M = 2$ procedures may be used in some cases. One normalizes the roots of one constituent to length $\sqrt{2}$, and the roots of the other to unit length. The short-root procedure may be used if a connecting irrep exists such that $|W_-| < 1$. One extends W_- to length one. In the long-root procedure one considers only irreps with Dynkin indices $a_i = 2\delta_{ik}$ in the constituent chosen to be short. If the length of W_- is less than $\sqrt{2}$, one extends it to length $\sqrt{2}$. The short- and long-root procedures are not applicable to exactly the same set of cases. For example, suppose that one wants to connect an A_1 to an end of an A_3 root chain. If A_3 is chosen as the short-root constituent, the length inequality is satisfied only for the short-root procedure; the final algebra is C_5. If A_3 is the long-root constituent, the length inequality is satisfied only for the long-root procedure; the final algebra is B_5.

We illustrate with two examples of extending semisimple algebras.

(A) Extending $A_4 \otimes A_n$ to an exceptional algebra. This type of extension is illustrated in Fig. 3(a), where X is the extension root. The extending representation is the direct product of the antisymmetric two-quark representation of A_4 and the one-quark representation of A_n. It is seen from Eq. (3.3) that an $M = 1$ extension is possible if $\frac{6}{5} + [n/(n+1)] < 2$. The $n = 1,2,$and 3 extensions yield the algebras E_6, E_7, and E_8.

(B) Extending $A(n_1) \otimes A(n_2) \otimes A(n_3)$ by a direct product of quark representation. This extension is illustrated in Fig. 3(b). The $M = 1$ extension is possible if

$$\frac{n_1}{n_1+1} + \frac{n_2}{n_2+1} + \frac{n_3}{n_3+1} < 2.$$

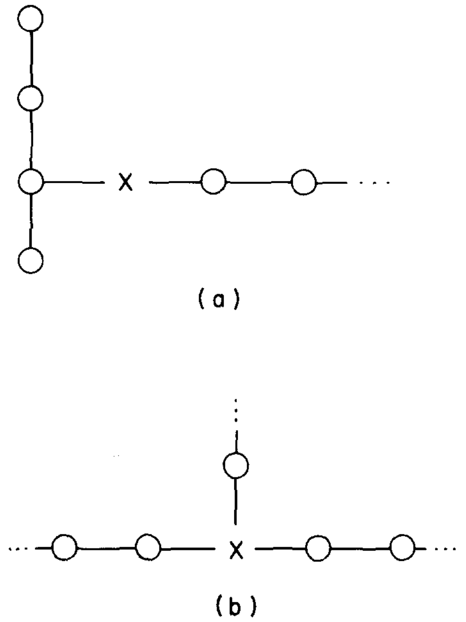

FIG. 3. Extending some semisimple algebras to simple algebras.

If the range of the n_i includes zero, all equal-magnitude simple algebras may be obtained in this way.

V. EXAMPLES OF EXTENSION BASES

Before we construct some specific orthogonal bases, it is useful to discuss further Dynkin's nonorthogonal bases, since these are used to justify some of the construction rules. If the Dynkin and dual indices (a_i and $\bar{a}_i$) are arranged in row vectors, then the square matrix G is defined by[11]

$$a = \bar{a}G^{-1}, \tag{5.1}$$

$$\bar{a} = aG. \tag{5.2}$$

If the diagonal matrix R^2 is defined by $(R^2)_{ij} = \delta_{ij}R_i^2$, where R_i is a simple root, then it is seen from Eqs. (2.4) and (5.1) that

$$G^{-1} = 2(R^2)^{-1}A, \tag{5.3}$$

so that

$$G = A^{-1}(R^2/2). \tag{5.4}$$

It follows from Eqs. (2.5) and (5.3) that G^{-1} (and hence G) is symmetric.

Since the Dynkin indices of an irrep refer to the highest weight, it is seen from Eqs. (2.6) and (5.2) that the length-squared of an irrep is

$$W_+^2 = \sum_i \bar{a}_i a_i = \sum_{ij} a_j G_{ji} a_i. \tag{5.5}$$

For the basic representation ψ_k, $a_j = \delta_{jk}$, so the length-squared is

$$|\psi_k|^2 = W^2_{+,k} = G_{kk}. \tag{5.6}$$

We use the symbols $\dot{a}_i$ to denote the components of a vector when written in terms of the simple roots, i.e.,

$$a = \sum_i \dot{a}_i R_i. \tag{5.7}$$

It is seen from Eq. (2.2) that the relation between this root basis and the dual basis is

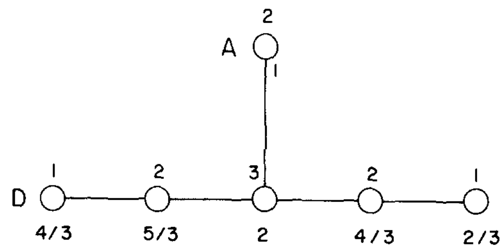

FIG. 4. Two rows of the A^{-1} matrix for E_6.

$$\dot{a}_i = \bar{a}_i(2/R_i^2). \tag{5.8}$$

It is seen from Eqs. (2.4) and (5.8) that the Dynkin basis is related to the root basis by $a_j = \Sigma_i \dot{a}_i A_{ij}$ or, in matrix notation,

$$a = \dot{a}A, \tag{5.9}$$

$$\dot{a} = aA^{-1}. \tag{5.10}$$

From the symmetry of G and Eqs. (5.4), conjugate elements of A^{-1} are related by

$$(A^{-1})_{ij}/(A^{-1})_{ji} = R_i^2/R_j^2. \tag{5.11}$$

A row of A^{-1} may be calculated quickly, because the allowed root dot products are such that A is very simple. It is seen from Eqs. (5.9) and (2.5) that one may calculate the $\dot{a}_i$ of the basic weight $W_{+,k}$ from the following prescription. One sets some convenient $\dot{a}_j$ equal to an arbitrary positive value. One then assigns values to the other root basis components so that each $\dot{a}_i$ (for $i \neq k$) is equal to its stability value.[12] This is half the sum of the $\dot{a}_j$ of all roots connected to it, with one exception. If the root i is connected to a longer root j, one multiplies $\dot{a}_j$ by M (the square of the ratio of root lengths) when computing the stability value of $\dot{a}_i$. One then renormalizes all the $\dot{a}_i$ so that $\dot{a}_k$ exceeds its stability value by $\frac{1}{2}$. This is illustrated in Fig. 4. The upper and lower numbers refer to the basic weights associated with the roots A and D, respectively.

Since the construction yields the basic weight $W_{+,k}$, it is seen from Eq. (5.10) that the components are the k row of the A^{-1} matrix. Usually it is easier to carry out this construction than to determine A^{-1} from tables. Furthermore, the construction leads to a generalization (used in Sec. VII) that is not obtainable from tables.

We demonstrate the simplicity of the extension procedure by showing that if one knows the basic weights of the unextended algebra H, very little extra effort is needed to obtain the basic weights of the extended algebra.

We take as examples the A_5[SU(6)] and D_5 extensions to E_6, shown in Fig. 5. The roots A and D are the new roots in these two cases. The first expressions by each circle are the extended A_5 root set, discussed in Sec. III A. The length of the extension vector x is determined so that the length of the extension root (x,abc) is $\sqrt{2}$. The second expressions are the basic E_6 weights in the extended A_5 basis. There are two parts to these basic weights, an A_5 part and an extension part Cx, where C is a numerical coefficient. The A_5 part may be determined from Eq. (2.9) and the A_5 roots. The A_5 part of $W_{+,x}$ (the basic weight associated with the extension root) is a singlet. The extension coefficient C_k for $W_{+,k}$ could be de-

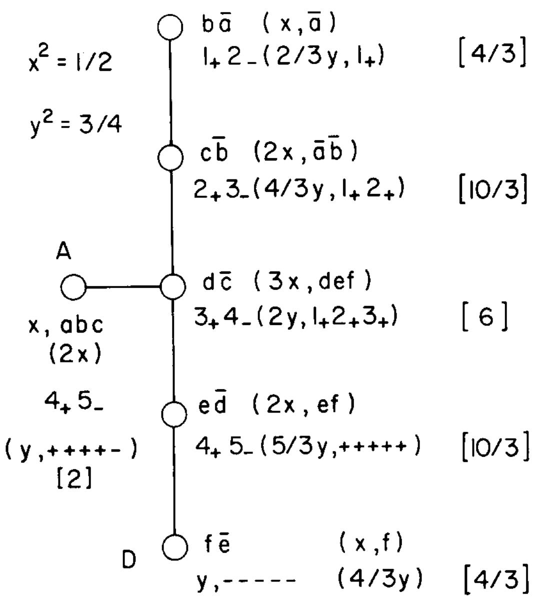

FIG. 5. The $A_5 \to E_6$ and $D_5 \to E_6$ extensions. The symbols not in parentheses are simple roots. The symbols in parentheses are basic weights. The numbers in square brackets are the squares of the lengths of basic weights.

termined by taking the dot product with the extension root, using Eq. (2.9). However, it is simpler to use the formula (to be justified later)

$$C_k = (A^{-1})kx. \tag{5.12}$$

Thus the C_k are a column of the A^{-1} matrix. Since E_6 is an equal-magnitude algebra, the A^{-1} matrix is symmetric and we may use the row elements shown in Fig. 4.

The third and fourth expressions by a circle in Fig. 5 are the simple roots and basic weights in the extended D_5 basis, determined by the procedure described above. The notation for D_5 is that of Sec. III B. The length of the extension vector y is determined by the condition that the length of the extension root $(y, - \; - \; - \; - \; -)$ is $\sqrt{2}$. The symbol $2/3y$ denotes $\frac{2}{3}y$, etc. The numbers in square brackets are squares of the lengths of the basic weights. These may be determined easily in either basis, or taken from Eq. (5.6).

We now prove the column rule, Eq. (5.12). Since the extension root is the only simple root with a component in the added dimension, the coefficient $\dot{a}_x$ in the root basis for any weight is the coefficient of the extension vector. For the basic weight $W_{+,k}$ it follows from Eqs. (2.8) and (5.10) that $\dot{a}_x(W_{+,k}) = C_k = (A^{-1})_{kx}$.

These bases simplify many calculations. In the standard Dynkin method, one computes all the positive roots level by level by adding simple roots.[9] If we identify the zero roots and simple roots with levels zero and one, respectively, the level number of the highest root is the sum of all the components in the root basis. This number is not small for many important groups. Similarly, for an arbitrary irreducible representation, one may start with the highest weight and subtract simple roots until one obtains the lowest weight. The height T, or level difference between the highest and lowest weights, is given by

$$T = \sum_i (\dot{a}_{+i} - \dot{a}_{-i}), \tag{5.13}$$

where the $+$ and $-$ refer to the highest and lowest weights. The a_{+i} may be determined from the Dynkin indices and Eq. (5.10); for the a_{-i} one uses Eq. (2.10) also.

On the other hand, if one uses an extension of an A_n or D_n algebra, the A and D simple roots are so familiar that they may be added or subtracted almost automatically. Care is needed only when adding or subtracting the extension root R_x. Thus the effective level of the highest root is just the component $\dot{a}_x$ of this root, and the effective height of an irrep may be obtained by considering only the term $i = x$ in Eq. (5.13). For example, consider the problem of constructing the zero and positive weights of the 2925-dimensional self-conjugate irrep ψ_3 of E_6. Half the height, determined from the 3 column (or row) of the A^{-1} matrix, is 21. However, it is seen from Fig. 5 that the effective half-height is only 3 in the extended A_5 basis, or 2 in the extended D_5 basis. With these bases, construction of some large representations is tractable.

Instead of constructing a large representation we illustrate the utility of the method by showing how quickly some small representations may be constructed. We concentrate on two simple types of irreps, namely the adjoint and short irreps of equal-magnitude algebras. These short irreps are necessarily equal-magnitude irreps. For these representations one has the following simple rules.

Root Rule: The dot product of two different roots that are not conjugate to each other is either 1, 0, or -1. The two roots may be added to form another root if and only if the product is -1. The nonzero roots are nondegenerate.

Weight Rule: The dot product of a root and a weight is 1, 0, or -1. The root may be added to the weight to form a new weight if and only if the product is -1. The weights are nondegenerate.

We construct first the positive roots of E_6 using the extended A_5 basis. Various sums of the simple A_5 roots lead to the 15 positive roots of A_5. Since (abc) is the lowest weight of the A_5 representation ψ_3, the sum of the extension root with various sums of simple A_5 roots yields all the 20 states of the $U(1) \otimes A_5$ representation containing (x,abc). We next determine the states of the $2x$ level. In order to do this we find the dot produce of the simple root (x,abc) with states of the x level containing different numbers of the generator quarks a, b, and c. By using $x^2 = \frac{1}{2}$ and Eq. (3.1), we obtain

$$(x,abc) \cdot (x,abd) = 1, \tag{5.14a}$$

$$(x,abc) \cdot (x,ade) = 0, \tag{5.14b}$$

$$(x,abc) \cdot (x,def) = -1. \tag{5.14c}$$

Only the last dot product is -1, so we may add these roots, i.e.,

$$(x,abc) + (x,def) = (2x).$$

The dot product of (x,abc) and $(2x)$ is positive, so the construction is complete. The positive roots are the fifteen positive A_5 roots, the twenty (x,abc) roots and the one $(2x)$ root.

Next we construct the 27-dimensional short basic representation ψ_1 by subtracting simple roots from the highest state (x,f). However, instead of subtracting the extension root (x,abc) we will add its conjugate $(-x, def)$. Subtracting A_5 roots from (x,f) leads to the six states of the

$U(1) \times A_5$ representation characterized by (x,a). It is easy to see that the dot product of $(-x,def)$ and a state (x,q) is -1 only if q is not d, e, or f. Hence we generate states of the type, $(-x,def) + (x,c) = (\overline{ab})$. There are 15 states of the type $(\overline{ab})$. The dot product of $(-x,def)$ with a $(\overline{qr})$ state is -1 only if both q and r are in the set (def). Thus we may form $(-x,def) + (\overline{de}) = (-x,f)$. There are six of these states. The smallest dot product of $(-x,def)$ with a $(-x,q)$ state is zero, so the construction is complete.

Briefly, we summarize the results of the same procedure in the extended D_5 basis, starting with the root construction. There are 20 positive D_5 roots, and 16 y-level spinor roots of the type $(y,-----)$. (These include all spinors where the product of the signs is negative.) There are no spinor–spinor dot products smaller that $(y,-----)\cdot(y,++++-)$, which is zero. Thus there are no roots at the $2y$ level; the construction is complete.

The extended D_5 construction of ψ_1 of E_6 yields the result

$$
\begin{aligned}
&(\tfrac{4}{3}y) \quad (1),\\
&(\tfrac{1}{3}y,+++++) \quad (16),\\
&(-\tfrac{2}{3}y,1_+) \quad (10),
\end{aligned}
\tag{5.15}
$$

where the most positive state of each $U(1) \otimes D_5$ representation is given, and the number at the right is the number of states.

We emphasize that the characteristics of a particular basis depend not only on the subalgebra used to label states, but also on the choice of simple roots. For example, in Ref. 4 standard methods are used to specify the $D_5 \otimes U(1)$ properties of the roots of E_6. It may be seen from Table 20 of this reference that only two of the simple roots of E_6 are D_5 roots. As a consequence, in the level diagram for the 27 of E_6 (Table 11b) the D_5 representations 10 and 16 are interlaced, while the D_5 singlet is at the center. This is in sharp contrast to the hierarchy arrangement of our Eq. (5.10). The point is that these two bases are different; each is convenient for certain purposes.

If one or more of the coefficients of the extension vector in the basic weights is nonintegral, the basis exhibits a congruence class of the group.[13] This follows because the coefficients of the extension vector in the simple roots are integral. It is seen from Eq. (5.12) that the possible congruence classes of a simple Lie algebra may be determined by inspecting the A^{-1} matrix. Since any root of the final algebra G may be used as the extension root, any A^{-1} column with one or more nonintegral elements leads to a nontrivial congruence relation. One multiplies the elements n_i in the column by the smallest integer N that leads to integral values of all $n_i' = Nn_i$. The congruence class of an irrep with Dynkin indices a_i is then

$$\sum_i n_i' a_i \quad (\mathrm{mod}\ N).$$

For example, the coefficients of y in Fig. 5 exhibit the E_6 triality relation $a_1 + 2a_2 + a_4 + 2a_5$ (mod 3).

VI. REPLACEMENT BASES

The extension technique is based on finding an irrep of the algebra H such that W_- is shorter than $(1/M)^{1/2}$ times the relevant root length, where M is a positive integer. In many cases $|W_-|$ is equal to $(1/M)^{1/2}$ times the relevant root length. In these cases if W_- is added to the simple root set of H, the argument of Sec. II shows that the extended root set will satisfy criteria (ii) and (iii) for a simple root system. However, since the added root is not extended, it is linearly dependent on the other roots. On the other hand, if one discards one of the original roots, the remaining roots will be linearly independent, and may be taken as the simple roots for an algebra of the same rank as H.

In all cases one representation that satisfies the above length equality is the adjoint. Replacing a simple root by the most negative root is the basis of Dynkin's extended-diagram technique for listing maximal, regular, semisimple subalgebras.[14] We will not consider this type of replacement, but will limit our attention to cases in which some irrep other than the adjoint satisfies the length equality.

In this procedure it is clear that the new root is negative. Therefore, it is convenient to redefine positivity so that all members of the new root set are positive, in which case they will be the simple roots of the new algebra. Usually, it is easy to find the appropriate redefinition; often all one has to do is to make the discarded root as negative as possible.

We illustrate the method by considering the $A_8 \to E_8$ and $D_8 \to E_8$ bases, shown in Fig. 6. In the A_8 case the weight ordering of the quarks is $i, h, \ldots b, a$, (with i the highest); the sum of the quark weights is zero. The notation of Fig. 6 is similar to that of Fig. 5. The A_8 simple roots, determined as in Sec. III A, are the first expressions next to the circles on the vertical line. The replacement root (abc) is of length $\sqrt{2}$. The root $9(i\bar{h})$ is discarded, so the new algebra is E_8. The

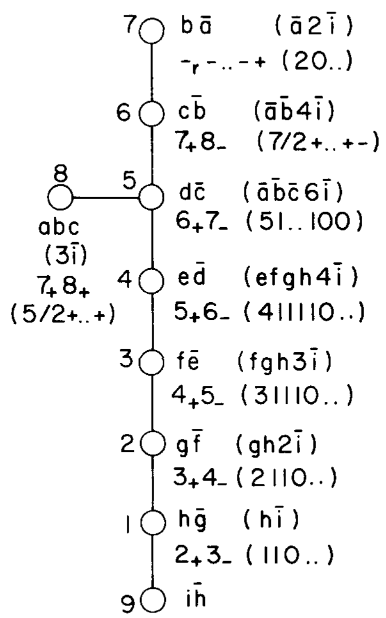

FIG. 6. Simple roots and basic weights in the $A_8 \to E_8$ and $D_8 \to E_8$ replacement bases.

quark and antiquark (i) and $(\bar i)$ do not appear in any of the E_8 simple roots. Hence, we redefine the quark weight order as $(h,g,...,a,i)$, choosing (i) so negative that all other quarks are positive. In an orthogonal basis the redefinition may be obtained by choosing (i) along the negative first axis. All roots in the new set of eight are positive. The second expression by each circle is the basic weight in the $A_8 \to E_8$ basis obtained by using Eq. (2.9) and the quark properties of Eq. (3.1).

The third expression by a circle in Fig. 6 is the D_8 root. Dots in a D_8 weight symbol denote a set of components equal to the adjacent component. In the $D_8 \to E_8$ basis the replacement root (no. 7) is the spinor $(- - - - - - - +)$. The 9 root $(1_+ 2_-)$ is discarded. Of the remaining roots only the spinor has a component in the first direction. Therefore, we redefine positivity by reflecting the first axis, so that all members of the new root set are positive. The subscript r refers to a component taken before reflection. The fourth expression by a circle in Fig. 6 is the E_8 basic weight in the D_8 basis, written in the new axis system.

We may construct the positive root set in the $A_8 \to E_8$ basis from the simple roots, using the root rule that precedes Eq. (5.14a). In this scheme the level number depends on the nature of the A_8 representation and whether or not the antiquark $\bar i$ is present. The level number is defined to be $\frac{1}{3}$ for each of the quarks $(a,b,...,h)$ and $-\frac{8}{3}$ for quark (i). This is proportional to the first component in the orthogonal basis defined above, but normalized so that the transition (replacement) root (abc) has level number one. We list in Table I the results of the positive-root construction. The complete set of roots includes the 80 SU(9) roots, the 84 antisymmetric three-quark states, and the 84 antisymmetric three-antiquark states.

In the $D_8 \to E_8$ basis the complete set of roots includes the 120 D_8 roots and the 128 spinors with positive products of signs.

We give one more example, the $A_7 \to E_7$ replacement shown in Fig. 7. The antisymmetric four-quark state $(abcd)$ is not short enough to be extendable for any A_n, but may be used as a replacement in the case of A_7, since the length is $\sqrt{2}$. The $(h\bar g)$ root is discarded and positivity redefined so that the h quark weight is along the negative first axis. As before, the second expression by a root is the basic weight. The complete root set consists of the 63 SU(8) roots and the 70 states of the irrep $(abcd)$. The 56-dimensional short irrep ψ_6 (top circle in Fig. 7) consists of the 28 (ab) and 28 $(\overline{ab})$ states.

VII. LISTING EXTENSION AND REPLACEMENT BASES

If one wants to list all possible extension bases for a simple algebra G, the fastest way is to work backward. One

TABLE I. The results of the positive-root construction.

Level	Highest state	Number of states
0	$(h\bar a)$	28
1	(fgh)	56
2	$(\overline{abi})$	28
3	$(h\bar i)$	8

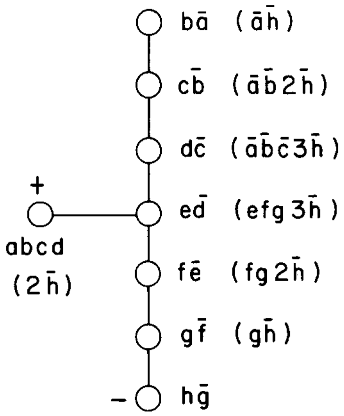

FIG. 7. The $A_7 \to E_7$ replacement.

writes the Dynkin diagram for G and labels any root with a plus sign. This corresponds to the basis obtained by starting with all roots other than the plus root and extending with the plus root. This listing procedure corresponds to Dynkin's procedure for listing the maximal regular nonsemisimple subalgebras of G (see Ref. 15). In a certain sense our extension method is the inverse of Dynkin's method for finding these subalgebras.

We present here a method for listing all possible replacement bases. We do not prove every assertion that is made. First, we define a candidate Dynkin diagram to be any indecomposable diagram in which the connected roots satisfy the conditions (ii) and (iii) of Sec. II. It is not required that the diagram may represent actual vectors. One then chooses a key root k and follows a procedure similar to that used to obtain A^{-1} rows in Sec. V. One sets some convenient hypothetical root basis component $\dot a_j$ equal to a positive value and chooses all the other components so that every $\dot a_i$ (for $i \neq k$) is equal to its stability value. If any of the resulting $\dot a_i$ are zero or negative, the scheme is classified as negative. If the $\dot a_i$ are all positive one classifies the scheme as positive, zero, or negative if the key component $\dot a_k$ is greater than, equal to, or less than its stability value, respectively. We assert that the classification is independent of the choice of key root, and so is a property of the candidate diagram. The positive diagrams correspond to simple Lie algebras. If a diagram is negative, it is impossible to find vectors satisfying the specified angle and length relations. For each zero diagram a set of appropriate vectors may be found, but they are linearly dependent. (The stability values for all $\dot a_i$ correspond to the zero vector.)

An alternate way to make the classification is to calculate A from Eq. (2.5) and then find the eigenvalues of A or of the symmetrized matrix $2(R^2)^{-1}A$ [see Eq. (5.3)]. A positive diagram corresponds to a positive definite A, while a zero diagram corresponds to an A with zero and positive eigenvalues. (If one considers only diagrams formed by adding one root to a positive diagram, then A has no more than one nonpositive eigenvalue, so one may make the classification in terms of the determinant of A.) This procedure is more illuminating than the key-root procedure, but takes longer to apply.

The zero diagrams are the basic tool used for listing replacement bases. They all may be obtained from a simple

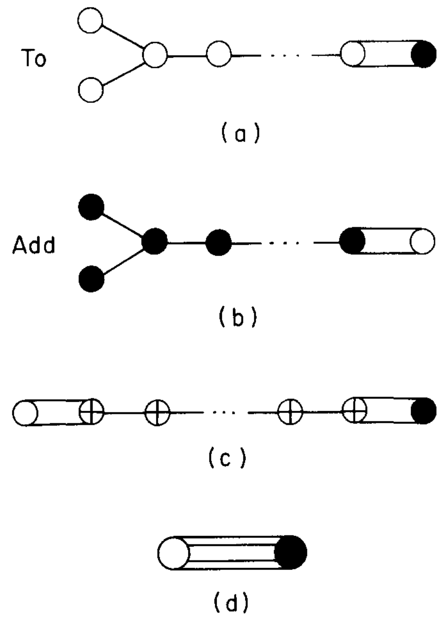

FIG. 8. Some zero diagrams.

three-step procedure. One starts with the four infinite classes and five specific extended Dynkin diagrams.[14] For each diagram with two root lengths one adds the diagram with root lengths reversed. An example is shown in Fig. 8. Diagram (a) is the extended Dynkin diagram associated with the algebras B_n. Diagram (b) is the added "length-exchange" diagram. Finally, to this list one adds the two diagrams (c) and (d) of Fig. 8. In diagram (c), the roots $\oplus$ are of intermediate length, a factor $\sqrt{2}$ longer than the shaded root and a factor $\sqrt{2}$ shorter than the clear root. Diagram (d) denotes two roots related as in conditions (ii) and (iii) of Sec. II, with $M = 4$.

We do not prove that the resulting set of diagrams is complete, but list some facts that may be used in a proof. First, if a new root is connected to an indecomposable zero or negative diagram, the result is negative. If a root and its connecting lines are removed from any indecomposable zero or positive diagram, the result is a set of one or more positive indecomposable diagrams. It follows that every zero diagram may be obtained by adding a root somewhere to a diagram for a simple Lie algebra; all one needs to do is to consider all the possibilities.[16]

For each zero diagram, any assignment of a plus sign to one root and a minus sign to another root corresponds to a replacement basis. The plus root is the replacement and the minus root is the discarded root.

In order to make clear the relation between this listing procedure and Dynkin's extended-diagram procedure for listing maximal regular semisimple subalgebras,[15] we consider a particular zero diagram, the nine-root diagram of Fig. 6. In Dynkin's procedure the diagram is obtained by adding root 9 to the diagram for E_8, so the diagram is associated with E_8. The plus sign is assigned to root 9, and the eight possible assignments of the minus sign correspond to eight subalgebras of E_8. Although Dynkin did not construct orthogonal bases, it is clear that any E_8 basis could be used as a basis for the subalgebra.

In the $A_8 \to E_8$ and $D_8 \to E_8$ replacement bases illustrated in Fig. 6, the minus sign is attached to root 9 and the plus sign to root 8 and root 7, respectively. However, it is not necessary that either sign be assigned to root 9. When one lists replacement bases, the diagram is not associated particularly with one root, but equally with all nine roots. In the general replacement $H \to G$, it is not required that H is a subalgebra of G, and it is not required that G is a subalgebra of H.

VIII. CONCLUDING REMARKS

There is an alternate, obvious method of obtaining orthogonal bases. If G is a simple algebra, and if either H, $H \otimes \mathrm{U}(1)$, or $H \otimes \mathrm{SU}(2)$ is a subalgebra of the same rank, one consults a table to find the structure of the adjoint of G in terms of representations of the subalgebra. One then uses a basis appropriate to the subalgebra to find the roots of G, defines positivity, finds the simple roots of G, and proceeds.

The extension and replacement procedures have several advantages over the subalgebra procedure. First, in the subalgebra procedure one must have considerable knowledge of G, *a priori*. The extension and replacement procedures require no such knowledge, and so are more illuminating. Second, in the subalgebra procedure it takes effort to determine the simple roots. Furthermore, as discussed in Sec. V, there is no guarantee that the simple roots will be a convenient root set. Third, the subalgebra procedure cannot be used to generate those replacement bases for which H is not a subalgebra of G.

ACKNOWLEDGMENT

This work was supported in part by the U. S. Department of Energy.

[1] E. B. Dynkin, Am. Math. Soc. Transl. **1**, 1 (1950); Ser. 2 **6**, 111 (1957).

[2] R. Gilmore, *Lie Groups, Lie Algebras, and Some of their Applications* (Wiley–Interscience, New York, 1974).

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

[4] R. Slansky, Phys. Rep. **79**, 1 (1981).

[5] Howard Georgi, *Lie Algebras in Particle Physics* (Benjamin/Cummings, Reading, MA, 1982).

[6] This is not the usual statement of the criteria. However, if one takes the conditions of page 177 of Ref. 5, and combines them with the results of Eqs. (VIII.2) and (VIII.3) (page 68), the results may be stated in the form used here. See also Ref. 4, page 27.

[7] See Ref. 4, pp. 27 and 28.

[8] Several orthogonal bases are listed in Refs. 2 and 3. A recent paper containing useful orthogonal bases is G. Feldman, T. Fulton, and P. T. Matthews, J. Math. Phys. **25**, 1222 (1984).

[9] The rules for constructing arbitrary irreps are discussed in Ref. 4, pp. 31 and 32.

[10] In the case of the symplectic algebras C_n, extension from $\mathrm{SU}(n)$ is a standard method of obtaining convenient bases. See Ref. 5, Chap. XXVI.

[11] Slansky, Ref. 4, p. 28. The G matrices for all the simple algebras are given in Table 7 of this reference.

[12] This term is used because if $V = \Sigma_i \dot{a}_i R_i$, then $\partial V^2/\partial \dot{a}_i = 0$, when $\dot{a}_i$ is

equal to its stability value.

[13]The congruence relations for the simple Lie algebras are listed by F. Lemire and J. Patera, J. Math. Phys. **21**, 2026 (1980).

[14]The extended Dynkin diagrams are discussed in Refs. 4 and 5. They are listed in Table 16 of Ref. 4 and p. 190 of Ref. 5.

[15]See Ref. 4, p. 41.

[16]An alternate method for obtaining the set of zero diagrams is given by D. Olive and Neil Turok, Nucl. Phys. B **215**, 470 (1983).

Boson realization of sp(4, R). II. The generating kernel formulation

O. Castaños
Centro de Estudios Nucleares, Universidad Nacional Autonoma de México, Apartado Postal 70-543, México, D. F. 04510 Mexico

P. Kramer
Institut für Theoretische Physik, Universitat Tübingen, Federal Republic of Germany

M. Moshinsky[a)]
Instituto de Física, Universidad Nacional Autonoma de México, Apartado Postal 20-364, México, D. F. 01000 Mexico

(Received 3 October 1985; accepted for publication 13 November 1985)

In a previous paper of this series, the matrix elements were discussed with respect to boson states of an operator K^2 required for the boson realization of the sp(4, R) Lie algebra. In the present paper, it is shown that these matrix elements can be obtained from a generating kernel given by the overlap of sp(4, R) coherent states. The results have relevance for the determination of the matrix elements of the generators of the sp(4, R) Lie algebra with respect to the basis of irreps of the positive discrete series for the corresponding group, and are, in principle, generalizable to symplectic algebras of higher dimensions.

I. INTRODUCTION AND SUMMARY

The determination of matrix elements of the generators of a symplectic Lie algebra with respect to the basis of irreducible representations (irreps) of the corresponding group is of considerable interest in problems of collective motions,[1] nuclear spectroscopy,[2] generalized atomic Hamiltonians,[3] etc. An important technique for the determination of these matrix elements is the boson realization of symplectic algebras.[4–7] In a previous paper of this series[8] (to be denoted by I with its equations quoted by their number followed by I) we discussed this realization for the case of sp(4, R), as it already shows the problems that arise in the general case sp($2d$,R), where d is any integer. In paper I we stressed that a Dyson type of boson realization can be obtained straightforwardly, but what is required is the one with appropriate Hermitian properties, which is known as the Holstein–Primakoff realization. To pass from the first to the second type of realization we need the matrix elements with respect to a complete set of boson states of an appropriate operator K, as originally discussed by Deenen and Quesne[4] and Rowe *et al.*[5] These matrix elements satisfy recursion relations,[5,9] whose solution was discussed in detail in paper I for the case sp(4, R). The main objective of the present paper is to show that with the help of the coherent states associated with a symplectic Lie algebra, one can determine a generating kernel from which an explicit analytic expression can be found for the matrix elements of K^2 in the case of sp(4, R) and that, in principle, the method can also be extended to sp($2d$,R) when $d>2$.

We proceed now to summarize the contents of the paper. In Sec. II we review briefly the results of paper I, to indicate that a boson realization of sp(4, R) requires not only the boson creation and annihilation operators $b_i^\dagger$, b_i, $i=1,2,3$ satisfying $[b_i, b_j^\dagger]=\delta_{ij}$, which are the generators of a Weyl Lie algebra in three dimensions $\mathscr{w}(3)$, but also extra degrees of freedom associated with an independent su(2) Lie algebra to whose generators S_i, $i=1,2,3$ we gave the name of spin. Our realization then expresses the generators of sp(4, R) in terms of the generators of the direct sum Lie algebra $\mathscr{w}(3)\oplus$su(2) and of the operator K mentioned. The Hermiticity properties $(b_i^\dagger)^\dagger=b_i$, $S_i^\dagger=S_i$, as well as those of the generators of sp(4, R), allow us then to derive the operator equations satisfied by K^2.

In Sec. III we introduce the coherent states for $\mathscr{w}(3)\oplus$su(2) and find the differential equations satisfied by the generating kernel, which is the matrix element of K^2 with respect to these coherent states.

In Sec. IV we discuss the coherent states of sp(4, R) and find out that their overlap satisfies the same differential equations as the matrix elements of K^2 with respect to the coherent states of $\mathscr{w}(3)\oplus$su(2). Thus the generating kernel can be identified with the overlap.

In Sec. V we determine explicitly the overlap mentioned, and in Sec. VI we use it to obtain the matrix elements of K^2 with respect to the boson states that are the basis for irreps of $\mathscr{w}(3)\oplus$su(2). In Sec. VII we indicate how the latter matrix elements can be used for the determination of those of the generators of sp(4, R) Lie algebra with respect to the basis of the positive discrete series irreps of the corresponding group. Finally in the concluding section we indicate possible generalizations to symplectic algebras of a larger number of dimensions and in particular to sp(6,R).

II. OPERATOR EQUATIONS FOR K^2

As discussed in paper I, the ten generators of the sp(4, R) Lie algebra can be expressed in terms of a scalar and three vectors

$$\mathscr{N},\ B_i^\dagger, J_i, B_i,\quad i=1,2,3, \tag{2.1}$$

with the commutation rules given in (3.2I), i.e.,

$$[B_i^\dagger, B_j^\dagger]=0, \tag{2.2a}$$

$$[B_i, B_j]=0, \tag{2.2b}$$

$$[B_i, B_j^\dagger]=-2i\epsilon_{ijk}J_k+2\delta_{ij}\mathscr{N}, \tag{2.2c}$$

[a)] Member of El Colegio Nacional and Instituto Nacional de Investigaciones Nucleares.

 0022-2488/86/040924-12$02.50

$$[J_i, B_j^\dagger] = i\epsilon_{ijk}B_k^\dagger, \tag{2.2d}$$

$$[J_i, B_j] = i\epsilon_{ijk}B_k, \tag{2.2e}$$

$$[J_i, J_j] = i\epsilon_{ijk}J_k, \tag{2.2f}$$

$$[\mathcal{N}, B_i^\dagger] = B_i^\dagger, \tag{2.2g}$$

$$[\mathcal{N}, B_i] = -B_i, \tag{2.2h}$$

$$[\mathcal{N}, J_i] = 0, \tag{2.2i}$$

where ϵ_{ijk} is the antisymmetric tensor and repeated indices are summed from 1 to 3.

Particularizing a discussion by Gilmore[10] to the case of sp(4, R), the defining representation of these generators, in terms of 4×4 matrices, takes the form

$$\mathcal{N} = \frac{1}{2}\begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}, \tag{2.3a}$$

$$B_i^\dagger = \begin{pmatrix} 0 & -i\sigma_i\sigma_2 \\ 0 & 0 \end{pmatrix}, \tag{2.3b}$$

$$J_i = \frac{1}{2}\begin{pmatrix} \sigma_i & 0 \\ 0 & -\bar{\sigma}_i \end{pmatrix}, \tag{2.3c}$$

$$B_i = \begin{pmatrix} 0 & 0 \\ i\bar{\sigma}_2\sigma_i & 0 \end{pmatrix}, \tag{2.3d}$$

where all submatrices are 2×2, I is the unit matrix while σ_i, $\bar{\sigma}_i$, $i = 1,2,3$ are, respectively, the Pauli matrices and their conjugates. It is easy to check, with the help of

$$\sigma_i\sigma_j = i\epsilon_{ijk}\sigma_k + \delta_{ij}I, \tag{2.4a}$$

$$\bar{\sigma}_2\sigma_i\sigma_2 = \bar{\sigma}_i, \tag{2.4b}$$

that all the commutation relations (2.2) are satisfied. We shall use this realization in Sec. V to get the overlap of the sp(4, R) coherent states.

The set (2.1) of ten generators of sp(4, R) can be divided into three subsets of raising-, weight-, and lowering-type separated below by semicolons, i.e.,

$$B_i^\dagger,\ J_+;\quad \mathcal{N},\ J_0;\quad B_i,\ J_-,\quad i = 1,2,3, \tag{2.5}$$

where $J_\pm = J_1 \pm iJ_2$, and $J_0 = J_3$.

The lowest weight state is then given by the solution of the equations

$$B_i|ws\rangle = 0, \quad i = 1,2,3, \tag{2.6a}$$

$$J_-|ws\rangle = 0, \tag{2.6b}$$

$$\mathcal{N}|ws\rangle = w|ws\rangle, \tag{2.6c}$$

$$J_0|ws\rangle = -s|ws\rangle, \tag{2.6d}$$

where we differ from the notation of paper I by calling w the eigenvalue of $\mathcal{N}$ instead of $\omega + n/2$, as we are dealing abstractly with the Lie algebra of sp(4, R) instead of its realization in terms of creation and annihilation operators of an n-body problem in two-dimensional space.

The irrep of sp(4, R) is now characterized by w, s or, if we follow the analysis leading to (3.8I), by

$$[w+s, w-s], \tag{2.7}$$

and a complete but nonorthonormalized set of states corresponding to this irrep can be written as

$$|m\mathbf{n}\rangle = P_\mathbf{n}(\mathbf{B}^\dagger)J_+^m|ws\rangle, \tag{2.8}$$

where we can take for $P_\mathbf{n}(\mathbf{B}^\dagger)$ the monomial

$$P_\mathbf{n}(\mathbf{B}^\dagger) = (B_1^\dagger)^{n_1}(B_2^\dagger)^{n_2}(B_3^\dagger)^{n_3}, \tag{2.9}$$

where (n_1, n_2, n_3), and $(B_1^\dagger, B_2^\dagger, B_3^\dagger)$ are the vectors denoted, respectively, by $\mathbf{n}$ and $\mathbf{B}^\dagger$.

The application of the generators (2.1) of sp(4, R) to the states (2.8) leads then to the same operator expressions (4.3I) for these generators only with $\omega + n/2$ replaced by w, but now we derive them through the use of the commutation rules (2.2) and not by the explicit realization of these generators in terms of the creation and annihilation operators of an n-particle system in two-dimensional space. Thus, for example, using only (2.2d) we arrive at

$$\begin{aligned} J_i\{P_\mathbf{n}(\mathbf{B}^\dagger)(J_+^m|ws\rangle)\} &= [J_i, P_\mathbf{n}(\mathbf{B}^\dagger)](J_+^m|ws\rangle) + P_\mathbf{n}(\mathbf{B}^\dagger)J_i(J_+^m|ws\rangle) \\ &= \left\{-i\epsilon_{ijk}B_j^\dagger\frac{\partial P_\mathbf{n}}{\partial B_k^\dagger} + P_\mathbf{n}J_i\right\}(J_+^m|ws\rangle), \end{aligned} \tag{2.10}$$

which agrees with (4.3bI).

From this point, the derivation of the Holstein–Primakoff relatization of the generators of sp(4, R) follows the same steps as those presented in Secs. IV and V of paper I. Besides the boson operators $b_i^\dagger$, b_i, $i = 1,2,3$, we need those of an independent spin S_i, which satisfy the commutation relations

$$[b_i^\dagger, b_j^\dagger] = [b_i, b_j] = [b_i^\dagger, S_j] = [b_i, S_j] = 0, \tag{2.11a}$$

$$[b_i, b_j^\dagger] = \delta_{ij}, \tag{2.11b}$$

$$[S_i, S_j] = i\epsilon_{ijk}S_k, \tag{2.11c}$$

which indicate that they are the generators of a Lie algebra corresponding to a direct sum of a Weyl algebra in three dimensions and an independent unitary unimodular algebra in two dimensions, i.e., $\omega(3)\oplus\mathrm{su}(2)$. These generators have the standard Hermiticity properties

$$(b_i^\dagger)^\dagger = b_i, \tag{2.12a}$$

$$S_i^\dagger = S_i. \tag{2.12b}$$

The generators $\mathcal{N}$, $B_i^\dagger$, J_i, and B_i of sp(4, R) can now be expressed in terms of $b_i^\dagger$, b_i, and S_i through the relations (4.6I) and (4.2I), where the latter involves the operator K we wish to determine. Thus we obtain in vector notation

$$\mathcal{N} = K(N + w)K^{-1}, \tag{2.13a}$$

$$\mathbf{B}^\dagger = K\mathbf{b}^\dagger K^{-1}, \tag{2.13b}$$

$$\mathbf{J} = K(\mathbf{L} + \mathbf{S})K^{-1}, \tag{2.13c}$$

$$\mathbf{B} = K[-\mathbf{b}^\dagger(\mathbf{b}\cdot\mathbf{b}) + (2N + 2w)\mathbf{b} - 2i(\mathbf{b}\times\mathbf{S})]K^{-1}, \tag{2.13d}$$

where

$$\mathbf{L} = -i(\mathbf{b}^\dagger\times\mathbf{b}), \tag{2.14a}$$

$$N = \mathbf{b}^\dagger\cdot\mathbf{b}. \tag{2.14b}$$

We can easily check that from the commutation relations (2.11) the generators of sp(4, R) satisfy the commutation relations (2.2). Note that the realization (2.13) differs from (5.4I) not only by the fact that $\omega + n/2$ is replaced by w, but also by the appearance of K, K^{-1} in (2.13a) and (2.13c) as we do not make here the explicit assumption that K is an invariant of u(2) as indicated in (5.3I).

The generators of sp(4, R) have themselves the Hermitian properties

$$(B_i^\dagger)^\dagger = B_i, \tag{2.15a}$$

$$J_i^\dagger = J_i, \tag{2.15b}$$

$$\mathscr{N}^\dagger = \mathscr{N}, \tag{2.15c}$$

as can be clearly seen from (1.1I) and (3.1I). Thus assuming, as we can do without loss of generality, that K is Hermitian, we get from the Hermitian conjugate of both sides of (2.13) that

$$\mathscr{N} = K^{-1}(N+w)K, \tag{2.16a}$$

$$\mathbf{B} = K^{-1}\mathbf{b}K, \tag{2.16b}$$

$$\mathbf{J} = K^{-1}(\mathbf{L}+\mathbf{S})K, \tag{2.16c}$$

$$\mathbf{B}^\dagger = K^{-1}[-(\mathbf{b}^\dagger\cdot\mathbf{b}^\dagger)\mathbf{b} + \mathbf{b}^\dagger(2N+2w) - 2i(\mathbf{b}^\dagger\times\mathbf{S})]K. \tag{2.16d}$$

Equating corresponding expressions in (2.13) and (2.16) we get the following operator equations for K^2:

$$[N, K^2] = 0, \tag{2.17a}$$

$$[L_i + S_i, K^2] = 0, \tag{2.17b}$$

$$b_i K^2 = K^2[-b_i^\dagger(\mathbf{b}\cdot\mathbf{b}) + (2N+2w)b_i - 2i(\mathbf{b}\times\mathbf{S})_i]. \tag{2.17c}$$

We do not write the equation coming from (2.13b) and (2.16d) as it is the Hermitian conjugate of (2.17c).

In paper I, we saw that the boson states are the elementary ones associated with a particle of spin s in a three-dimensional harmonic oscillator, which in (5.10I) were denoted by the ket

$$|\nu[ls]jm), \tag{2.18}$$

where ν is the total number of quanta, l is the "orbital" angular momentum while j and m correspond, respectively, to the total angular momentum and its projection. These states are orthonormal and from (2.17a) and (2.17b) only the matrix elements

$$(\nu[l's]jm|K^2|\nu[ls]jm) \tag{2.19}$$

are different from 0; besides, they are independent of m. Equation (2.17c) leads to a recursion relation for the matrix elements (2.19) and the procedure for its solution was discussed in the appendix of paper I.

In Sec. VI of the present paper we shall derive an explicit and closed analytic expression for (2.19) using coherent states of both the $\omega(3)\oplus$su(2) and sp(4, R) Lie algebras. To achieve this objective we first discuss in the following section the matrix elements of K^2 with respect to coherent states of $\omega(3)\oplus$su(2), as well as the differential equations that they satisfy.

III. EQUATIONS SATISFIED BY THE GENERATING KERNEL

As the Lie algebra $\omega(3)\oplus$su(2) is a direct sum, its coherent states are the direct product of those of $\omega(3)$ and su(2). For the Weyl group $\omega(3)$ the coherent states are the standard ones associated with the three-dimensional oscillator,[11] while for su(2) they have been discussed, among others, by Kramer and Saraceno.[12] Thus we can write them as the ket

$$|y\mathbf{z}) = \exp(\bar{\mathbf{z}}\cdot\mathbf{b}^\dagger)\exp(\bar{y}S_+)|s), \tag{3.1}$$

where $b_i^\dagger$, b_i, and S_i satisfy the commutation rules (2.11), $S_\pm = S_1 \pm iS_2$, and y and z_i, $i=1,2,3$, are complex numbers, with $\bar{y}$ and $\bar{z}_i$ being their conjugates. The ket $|s)$ is the direct product

$$|s) = |0\}\times|s,-s\}, \tag{3.2}$$

where $|0\}$ is the zero quantum state associated with the bosons $b_i^\dagger$ and b_i, while $|s,-s\}$ is the lowest weight state of the independent spin operator S_i. The corresponding bra is the Hermitian conjugate and thus takes the form

$$(y\mathbf{z}| = (s|\exp(yS_-)\exp(\mathbf{z}\cdot\mathbf{b}). \tag{3.3}$$

The measure $\rho(y,\bar{y},z,\bar{z})$ required for scalar products in the four-dimensional complex space of y, z_1, z_2, z_3 is clearly the product of the measures[11,12] associated with the $\omega(3)$ and su(2) Lie algebras, where in the latter case it depends on the spin s, and it takes the form[11,12]

$$\rho(y,\bar{y},\mathbf{z},\bar{\mathbf{z}}) = \pi^{-3}\exp(-\mathbf{z}\cdot\bar{\mathbf{z}})\pi(2s+1)^{-1}(1+\bar{y}y)^{-(2s+2)}. \tag{3.4}$$

The reproducing kernel is also a product of those associated with the $\omega(3)$ and su(2) Lie algebras and thus is given by[11,12]

$$\exp(\mathbf{z}'\cdot\bar{\mathbf{z}})(1+y'\bar{y})^{2s}. \tag{3.5}$$

Our objective is now to find from the operator equations (2.17) the differential ones satisfied by the generating kernel, i.e., by the matrix elements

$$(y'\mathbf{z}'|K^2|y\mathbf{z}) \tag{3.6}$$

of the operator K^2 with respect to the coherent states (3.1) and (3.3). For this purpose we require first the differential form with respect to the variables y', $\mathbf{z}'$, or $\bar{y}$, $\bar{\mathbf{z}}$ of the operators $b_i^\dagger$, b_i, and S_i when acting on the bra (3.3) or ket (3.1). We start with the operator $b_i^\dagger$ acting on the bra and thus have

$$\begin{aligned}(y'\mathbf{z}'|b_i^\dagger &= (s|\exp(y'S_-)\exp(\mathbf{z}'\cdot\mathbf{b})b_i^\dagger \\ &= (s|\{\exp(\mathbf{z}'\cdot\mathbf{b})b_i^\dagger\exp(-\mathbf{z}'\cdot\mathbf{b})\} \\ &\quad\times\exp(y'S_-)\exp(\mathbf{z}'\cdot\mathbf{b}).\end{aligned} \tag{3.7}$$

We will use, in some of the following discussion, expansions of expressions like the curly bracket in (3.7) in terms of multiple commutators, i.e.,

$$\{\exp A\}B\{\exp(-A)\} = B + [A,B] + \tfrac{1}{2}[A,[A,B]] + \cdots, \tag{3.8}$$

which from the commutation rule $[b_i, b_j^\dagger] = \delta_{ij}$ imply that

$$\exp(\mathbf{z}'\cdot\mathbf{b})b_i^\dagger\exp(-\mathbf{z}'\cdot\mathbf{b}) = b_i^\dagger + z_i'. \tag{3.9}$$

As, furthermore, $(s|b_i^\dagger = \{b_i|s)\}^\dagger = 0$, we then conclude that

$$(y'\mathbf{z}'|b_i^\dagger = z_i'(y'\mathbf{z}'|, \tag{3.10}$$

as we could expect from the analysis in Bargmann Hilbert space.[11] Applying now b_i to the bra we see more directly that

$$(y'\mathbf{z}'|b_i = (s|\exp(y'S_-)\exp(\mathbf{z}'\cdot\mathbf{b})b_i = \frac{\partial}{\partial z_i'}(y'\mathbf{z}'|. \tag{3.11}$$

We turn now our attention to the operator S_q, $q = \pm,0$ where $S_\pm = S_1 \pm iS_2$, $S_0 = S_3$ and apply it to the bra to get

$$\begin{aligned}(y'\mathbf{z}'|S_q &= (s|\exp(y'S'_-)\exp(\mathbf{z}'\cdot\mathbf{b})S_q \\ &= (s|\{\exp(y'S_-)S_q \exp(-y'S_-)\} \\ &\quad\times\exp(y'S_-)\exp(\mathbf{z}'\cdot\mathbf{b}).\end{aligned} \tag{3.12}$$

Again, from the expansion (3.8) and the commutation relations

$$[S_-,S_+] = -2S_0, \quad [S_0,S_\pm] = \pm S_\pm, \tag{3.13}$$

we obtain

$$\begin{aligned}(y'\mathbf{z}'|S_q = (s|\{S_q &+ y'[S_-,S_q] \\ &+ \tfrac{1}{2}y'^2[S_-,[S_-,S_q]]\}\exp(y'S_-)\exp(\mathbf{z}'\cdot\mathbf{b}),\end{aligned} \tag{3.14}$$

and thus

$$\begin{aligned}(y'\mathbf{z}'|S_+ &= (s|\{S_+ - 2y'S_0 - y'^2S_-\} \\ &\quad\times\exp(y'S_-)\exp(\mathbf{z}'\cdot\mathbf{b}) \\ &= -y'\Big(y'\frac{\partial}{\partial y'} - 2s\Big)(y'\mathbf{z}'|\equiv S'_+(y'\mathbf{z}'|,\end{aligned} \tag{3.15a}$$

$$\begin{aligned}(y'\mathbf{z}'|S_0 &= (s|\{S_0 + y'S_-\}\exp(y'S_-)\exp(\mathbf{z}'\cdot\mathbf{b}) \\ &= \Big(y'\frac{\partial}{\partial y'} - s\Big)(y'\mathbf{z}'|\equiv S_0'(y'\mathbf{z}'|,\end{aligned} \tag{3.15b}$$

$$\begin{aligned}(y'\mathbf{z}'| &= (s|S_- \exp(y'S_-)\exp(\mathbf{z}'\cdot\mathbf{b}) \\ &= \frac{\partial}{\partial y'}(y'\mathbf{z}'|\equiv S'_-(y'\mathbf{z}'|,\end{aligned} \tag{3.15c}$$

where we made use of

$$(s|S_+ = \{S_-|s)\}^\dagger = 0, \tag{3.16a}$$

$$(s|S_0 = \{S_0|s)\}^\dagger = -s(s|. \tag{3.16b}$$

Thus the operators $S_\pm$ and S_0, when acting on the bra $(y'\mathbf{z}'|$, become the differential operators $S'_\pm$ and S_0' defined in (3.15).

A similar analysis, in which $b_i^\dagger$, b_i, and S_i now act on the ket $|y,\mathbf{z})$, gives us

$$b_i^\dagger|y\mathbf{z}) = \frac{\partial}{\partial \bar{z}_i}|y\mathbf{z}), \tag{3.17}$$

$$b_i|y\mathbf{z}) = \bar{z}_i|y\mathbf{z}), \tag{3.18}$$

$$S_+|y\mathbf{z}) = \frac{\partial}{\partial \bar{y}}|y\mathbf{z})\equiv -\bar{S}_+|y\mathbf{z}), \tag{3.19a}$$

$$S_0|y\mathbf{z}) = \Big(\bar{y}\frac{\partial}{\partial \bar{y}} - s\Big)|y\mathbf{z})\equiv -\bar{S}_0|y\mathbf{z}), \tag{3.19b}$$

$$S_-|y\mathbf{z}\rangle = \Big(2\bar{y}s - \bar{y}^2\frac{\partial}{\partial \bar{y}}\Big)|y\mathbf{z})\equiv -\bar{S}_-|y\mathbf{z}), \tag{3.19c}$$

as would follow also from Hermitian conjugation of (3.10), (3.11), and (3.15). Note that $\bar{S}_\pm$ and $\bar{S}_0$ defined by (3.19), which are differential operators in the variable $\bar{y}$, have here a $-$ sign as compared with $S'_\pm$ and S_0' of (3.15). This is to be expected, and guarantees that both S_q' and $\bar{S}_q$, $q = \pm,0$ satisfy the commutation relations (3.13). In fact the orbital angular momentum

$$L_i = -i\epsilon_{ijk}b_j^\dagger b_k \tag{3.20}$$

shows a similar behavior, as from (3.10) and (3.11) we have

$$(y'\mathbf{z}'|L_i = -i\epsilon_{ijk}z_j'\frac{\partial}{\partial z_k'}(y'\mathbf{z}'|\equiv L_i'(y\mathbf{z}'|, \tag{3.21a}$$

while from (3.17) and (3.18) we have

$$L_i|y\mathbf{z}) = -i\epsilon_{ijk}\bar{z}_k\frac{\partial}{\partial \bar{z}_j}|y\mathbf{z})\equiv -\bar{L}_i|y\mathbf{z}). \tag{3.21b}$$

We note also that the number operator

$$N = b_i^\dagger b_i, \tag{3.22}$$

when acting on bra and ket, becomes

$$(y'\mathbf{z}'|N = z_i'\frac{\partial}{\partial z_i'}(y'\mathbf{z}'|\equiv N'(y'\mathbf{z}'|, \tag{3.23a}$$

$$N|y\mathbf{z}) = \bar{z}_i\frac{\partial}{\partial \bar{z}_i}|y\mathbf{z})\equiv\bar{N}|y\mathbf{z}). \tag{3.23b}$$

We are now in a position to derive the differential equations satisfied by $(y'\mathbf{z}'|K^2|y\mathbf{z})$ by taking both sides of the equations (2.17) between a bra $(y'\mathbf{z}'|$ and a ket $|y\mathbf{z})$. Making use of the differential expressions for the operators $b_i^\dagger$, b_i, and S_i, $i = 1,2,3$, when acting on a bra and a ket, we then get

$$(N' - \bar{N})(y'\mathbf{z}'|K^2|y\mathbf{z}) = 0, \tag{3.24a}$$

$$(L_i' + S_i' + \bar{L}_i + \bar{S}_i)(y'\mathbf{z}'|K^2|y\mathbf{z}) = 0, \tag{3.24b}$$

$$\begin{aligned}&\frac{\partial}{\partial z_i'}(y'\mathbf{z}'|K^2|y\mathbf{z}) \\ &\quad= \Big\{-(\bar{\mathbf{z}}\cdot\bar{\mathbf{z}})\frac{\partial}{\partial \bar{z}_i} + \bar{z}_i(2\bar{N} + 2w) + 2i(\bar{\mathbf{z}}\times\bar{\mathbf{S}})_i\Big\} \\ &\qquad\times(y'\mathbf{z}'|K^2|y\mathbf{z}),\end{aligned} \tag{3.24c}$$

where N', L_i', and S_i' are defined, respectively, in (3.23a), (3.21a), and (3.15), while $\bar{N}$, $\bar{L}_i$, and $\bar{S}_i$ are given in (3.23b), (3.21b), and (3.19), and where we note that $S_1 = \frac{1}{2}(S_+ + S_-)$, $S_2 = (1/2i)(S_+ - S_-)$, and $S_3 = S_0$.

We proceed now to show that the overlap of coherent states associated with sp(4, R) satisfies the same equations (3.24), and that thus we only need to find this overlap to evaluate the matrix elements (3.6).

IV. OVERLAP FOR COHERENT STATES OF sp(4, R) AND THE EQUATIONS THEY SATISFY

As in the case of $\omega(3)\oplus$su(2), the coherent states of sp(4, R) can be written[13,14] in terms of exponentials of linear combinations of the raising generators of sp(4, R) applied to the lowest weight state. From (2.5) and (2.6) we have, then,

$$|y\mathbf{z}\rangle = \exp(\bar{\mathbf{z}}\cdot\mathbf{B}^\dagger)\exp(\bar{y}J_+)|ws\rangle, \tag{4.1}$$

where we now distinguish these sp(4, R) coherent states from the corresponding ones for $\omega(3)\oplus$su(2) given in (3.1) by using angular rather than round kets. As before y, z_1, z_2, and z_3 are complex numbers and the bar above indicates the conjugate. The corresponding bra is given by the

Hermitian conjugate and thus for new parameters y', $\mathbf{z}'$ it becomes

$$\langle y'\mathbf{z}'| = \langle ws|\exp(y'J_-)\exp(\mathbf{z}'\cdot\mathbf{B}). \tag{4.2}$$

We wish now to find the differential equations satisfied by the overlap

$$\langle y'\mathbf{z}'|y\mathbf{z}\rangle, \tag{4.3}$$

for which we will take the matrix elements of the generators $\mathscr{N}, J_i, B_i$, and $B_i^\dagger$ of sp(4, R) between the bra (4.2) and the ket (4.1). Applying these generators to the left and right we get, with the help of the commutation relations (2.2) and the expansions (3.8), the differential equations for the overlap, which will turn out to be identical to those satisfied by $(y'\mathbf{z}'|K^2|y\mathbf{z})$, which are given by (3.24).

We start with $\mathscr{N}$ and get

$$\begin{aligned}\langle y'\mathbf{z}'|\mathscr{N}|y\mathbf{z}\rangle &= \langle ws|\exp(y'J_-)\exp(\mathbf{z}'\cdot\mathbf{B})\mathscr{N}|y\mathbf{z}\rangle \\ &= \langle ws|\exp(y'J_-)\{\exp(\mathbf{z}'\cdot\mathbf{B})\mathscr{N} \\ &\quad\times\exp(-\mathbf{z}'\cdot\mathbf{B})\}\exp(\mathbf{z}'\cdot\mathbf{B})|y\mathbf{z}\rangle.\end{aligned} \tag{4.4}$$

From (2.2) and (3.8) the curly bracket in (4.4) becomes

$$\{\mathscr{N} + z_j'[B_j, \mathscr{N}]\} = \mathscr{N} + z_j' B_j. \tag{4.5}$$

As $\mathscr{N}$ commutes with J_i, and when applied to $|ws\rangle$ gives w, we have that

$$\langle y'\mathbf{z}'|\mathscr{N}|y\mathbf{z}\rangle = \left(z_j'\frac{\partial}{\partial z_j'} + w\right)\langle y'\mathbf{z}'|y\mathbf{z}\rangle. \tag{4.6}$$

Applying now in (4.4) the generator $\mathscr{N}$ to the right we get in a similar fashion

$$\langle y'\mathbf{z}'|\mathscr{N}|y\mathbf{z}\rangle = \left(\bar{z}_j\frac{\partial}{\partial \bar{z}_j} + w\right)\langle y'\mathbf{z}'|y\mathbf{z}\rangle. \tag{4.7}$$

Equating then (4.6) and (4.7), we obtain the expression

$$(N' - \bar{N})\langle y'\mathbf{z}'|y\mathbf{z}\rangle = 0, \tag{4.8}$$

where N' and $\bar{N}$ are given in (3.23).

Turning our attention now to J_i we have

$$\begin{aligned}\langle y'\mathbf{z}'|J_i|y\mathbf{z}\rangle &= \langle ws|\exp(y'J_-)\exp(\mathbf{z}'\cdot\mathbf{B})J_i|y\mathbf{z}\rangle \\ &= \langle ws|\exp(y'J_-)\{\exp(\mathbf{z}'\cdot\mathbf{B})J_i \\ &\quad\times\exp(-\mathbf{z}'\cdot\mathbf{B})\}\exp(\mathbf{z}'\cdot\mathbf{B})|y\mathbf{z}\rangle.\end{aligned} \tag{4.9}$$

Again we use (2.2) and (3.8) and see that the curly bracket becomes

$$\{J_i + z_j'[B_j', J_i]\} = J_i - i\epsilon_{ijk}z_j'B_k, \tag{4.10}$$

and so the matrix elements can be written as

$$\begin{aligned}&\langle ws|\exp(y'J_-)\left(J_i - i\epsilon_{ijk}z_j'\frac{\partial}{\partial z_k'}\right)\exp(\mathbf{z}'\cdot\mathbf{B})|y\mathbf{z}\rangle \\ &\quad= \langle ws|\{\exp(y'J_-)J_i\exp(-y'J_-)\} \\ &\qquad\times\exp(y'J_-)\exp(\mathbf{z}'\cdot\mathbf{B})|y\mathbf{z}\rangle \\ &\qquad- i\epsilon_{ijk}z_j'\frac{\partial}{\partial z_k'}\langle y'\mathbf{z}'|y\mathbf{z}\rangle \\ &\quad= (S_i' + L_i')\langle y'\mathbf{z}'|y\mathbf{z}\rangle,\end{aligned} \tag{4.11}$$

where S_i' is defined as in (3.15), as the curly bracket in (4.11) can be evaluated in exactly the same fashion as the one appearing in (3.12), while L_i' is given by (3.21a).

Applying now J_i to the right we get in a similar fashion that

$$\langle y\mathbf{z}'|J_i|y\mathbf{z}\rangle = -(\bar{S}_i + \bar{L}_i)\langle y'\mathbf{z}'|y\mathbf{z}\rangle, \tag{4.12}$$

where $\bar{S}_i$ and $\bar{L}_i$ are given, respectively, by (3.19) and (3.21b). Thus from (4.11) and (4.12) we see that the overlap satisfies the equation

$$(L_i' + S_i' + \bar{L}_i + \bar{S}_i)\langle y'\mathbf{z}'|y\mathbf{z}\rangle = 0. \tag{4.13}$$

Considering now B_i we immediately see that its matrix element, when the operator is applied to the left, is given by

$$\begin{aligned}\langle y'\mathbf{z}'|B_i|y\mathbf{z}\rangle &= \langle ws|\exp(y'J_-)\exp(\mathbf{z}'\cdot\mathbf{B}')B_i|y\mathbf{z}\rangle \\ &= \frac{\partial}{\partial z_i'}\langle y'\mathbf{z}'|y\mathbf{z}\rangle.\end{aligned} \tag{4.14}$$

On the other hand, when B_i acts on the right we can write

$$\begin{aligned}\langle y'\mathbf{z}'|B_i|y\mathbf{z}\rangle &= \langle y'\mathbf{z}'|B_i\exp(\bar{\mathbf{z}}\cdot\mathbf{B}^\dagger)\exp(\bar{y}J_+)|ws\rangle \\ &= \langle y'\mathbf{z}'|\exp(\bar{\mathbf{z}}\cdot\mathbf{B}^\dagger)\{\exp(-\bar{\mathbf{z}}\cdot\mathbf{B}^\dagger)B_i \\ &\quad\times\exp(\bar{\mathbf{z}}\cdot\mathbf{B}^\dagger)\}\exp(\bar{y}J_+)|ws\rangle.\end{aligned} \tag{4.15}$$

Again we use (2.2) and (3.8) and see that the curly bracket becomes

$$\begin{aligned}&\{B_i - \bar{z}_j[B_j^\dagger, B_i] + \tfrac{1}{2}\bar{z}_j\bar{z}_k[B_k^\dagger,[B_j^\dagger, B_i]]\} \\ &\quad= \{B_i - 2i\epsilon_{ijk}\bar{z}_jJ_k + 2\bar{z}_i\mathscr{N} \\ &\qquad- (\bar{\mathbf{z}}\cdot\bar{\mathbf{z}})B_i^\dagger + 2\bar{z}_i(\bar{\mathbf{z}}\cdot\mathbf{B}^\dagger)\},\end{aligned} \tag{4.16}$$

so that introducing it in (4.15) we obtain

$$\begin{aligned}\langle y'\mathbf{z}'|B_i|y\mathbf{z}\rangle &= \langle y'\mathbf{z}'|\exp(\bar{\mathbf{z}}\cdot\mathbf{B}^\dagger) \\ &\quad\times\{B_i - 2i\epsilon_{ijk}\bar{z}_jJ_k + 2\bar{z}_i\mathscr{N}\}\exp(\bar{y}J_+)|ws\rangle \\ &\quad+ \langle y'\mathbf{z}'|\left\{-(\bar{\mathbf{z}}\cdot\bar{\mathbf{z}})\frac{\partial}{\partial\bar{z}_i} + 2\bar{z}_i\left(\bar{z}_j\frac{\partial}{\partial\bar{z}_j}\right)\right\} \\ &\quad\times\exp(\bar{\mathbf{z}}\cdot\mathbf{B}^\dagger)\exp(\bar{y}J_+)|ws\rangle.\end{aligned} \tag{4.17}$$

In the first term on the right-hand side of (4.17) we see that

$$B_i\exp(\bar{y}J_+)|ws\rangle = 0, \tag{4.18a}$$

as the commutators of power of J_+ with B_i gives some B_j on the right-hand side and from (2.6), $B_j|ws\rangle = 0$. Furthermore

$$\begin{aligned}&-2i\epsilon_{ijk}\bar{z}_jJ_k(\exp\bar{y}J_+)|ws\rangle \\ &\quad= -2i\epsilon_{ijk}\bar{z}_j\exp(\bar{y}J_+) \\ &\qquad\times\{\exp(-\bar{y}J_+)J_k\exp(\bar{y}J_+)\}|ws\rangle \\ &\quad= 2i\epsilon_{ijk}\bar{z}_j\bar{S}_k(\exp(\bar{y}J_+)|ws\rangle),\end{aligned} \tag{4.18b}$$

where $\bar{S}_k$ is given by (3.19), as follows from the evaluation of the curly bracket in (4.18b). Finally

$$2\bar{z}_i\mathscr{N}\exp(\bar{y}J_+)|ws\rangle = 2\bar{z}_iw\exp(\bar{y}J_+)|ws\rangle, \tag{4.18c}$$

as $\mathscr{N}$ commutes with J_i and $\mathscr{N}|ws\rangle = w|ws\rangle$.

Combining (4.17) and (4.18) with (4.14), we then get the differential equation

$$\frac{\partial}{\partial z_i'}\langle y'\mathbf{z}'|y\mathbf{z}\rangle = \{-(\bar{\mathbf{z}}\cdot\bar{\mathbf{z}})\frac{\partial}{\partial \bar{z}_i} + \bar{z}_i(2\overline{N}+2w) + 2i\epsilon_{ijk}\bar{z}_j\overline{S}_k\}\langle y'\mathbf{z}'|y\mathbf{z}\rangle. \quad (4.19)$$

There remains to evaluate the matrix element of $B_i^\dagger$ but it will lead to no new result, as

$$\langle y'\mathbf{z}'|B_i^\dagger|y\mathbf{z}\rangle = \overline{\langle y\mathbf{z}|B_i|y'\mathbf{z}'\rangle}, \quad (4.20)$$

where the bar above means the conjugate of the matrix element.

We have thus shown that the overlap $\langle y'\mathbf{z}'|y\mathbf{z}\rangle$ satisfies the equations (4.8), (4.13), and (4.19), which are identical to the equations (3.24) satisfied by $(y'\mathbf{z}'|K^2|y\mathbf{z})$, so that we can take

$$(y'\mathbf{z}'|K^2|y\mathbf{z}) = \langle y'\mathbf{z}'|y\mathbf{z}\rangle. \quad (4.21)$$

In the next section we shall evaluate explicitly the overlap $\langle y'\mathbf{z}'|y\mathbf{z}\rangle$ and in Sec. VI determine, with its help, the matrix elements of K^2 with respect to the boson states (2.18).

V. DETERMINATION OF THE OVERLAPS FOR COHERENT STATES OF THE sp(4, R) LIE ALGEBRA

The overlap of coherent states of sp(4, R) is given by

$$\langle y'\mathbf{z}'|y\mathbf{z}\rangle = \langle ws|\exp(y'J_-)\{\exp(\mathbf{z}'\cdot\mathbf{B})\exp(\bar{\mathbf{z}}\cdot\mathbf{B}^\dagger)\} \times\exp(\bar{y}J_+)|ws\rangle, \quad (5.1)$$

and hence it is a matrix element of an operator associated with the symplectic group. To evaluate it we try to write the operator inside the curly bracket of (5.1) in a canonical form where we first have exponentials of linear combinations of raising generators $B_i^\dagger$, then of the operators J_i and $\mathscr{N}$ of a u(2) subalgebra of sp(4, R), and finally of B_i, i.e.,

$$\begin{aligned}&\langle ws|\exp(y'J_-)\{\exp(\mathbf{z}'\cdot\mathbf{B})\exp(\bar{\mathbf{z}}\cdot\mathbf{B}^\dagger)\}\exp(\bar{y}J_+)|ws\rangle\\ &\quad= \langle ws|\exp(y'J_-)\{\exp(\mathbf{d}\cdot\mathbf{B}^\dagger)\exp(e\mathscr{N})\\ &\qquad\times\exp(\mathbf{f}\cdot\mathbf{J})\exp(\mathbf{c}\cdot\mathbf{B})\}\exp(\bar{y}J_+)|ws\rangle,\end{aligned} \quad (5.2)$$

where e is a scalar and $\mathbf{c}$, $\mathbf{d}$, and $\mathbf{f}$ are vectors, all of them functions of $\mathbf{z}'$ and $\mathbf{z}$, which we shall proceed to determine below. The advantage of the form (5.2) is that the application of B_i to the state $\exp(\bar{y}J_+)|ws\rangle$ translates, through the commutation relations (2.2e), in having an operator B_k acting on $|ws\rangle$, which from (2.6) gives zero. Through the Hermiticity property we also get

$$\langle ws|\exp(y'J_-)B_i^\dagger = (B_i\exp(\bar{y}'J_+)|ws\rangle)^\dagger = 0, \quad (5.3)$$

so we would conclude that

$$\langle y'\mathbf{z}'|y\mathbf{z}\rangle = \langle ws|\exp(y'J_-)\exp(e\mathscr{N}) \times\exp(\mathbf{f}\cdot\mathbf{J})\exp(\bar{y}J_+)|ws\rangle, \quad (5.4)$$

and thus the overlap is a matrix element of a representation of a GL(2,C) as y', e, f_1, f_2, f_3, and $\bar{y}$ are arbitrary complex numbers. This representation can be obtained by procedures discussed by Louck.[15]

To obtain e, $\mathbf{c}$, $\mathbf{d}$, and $\mathbf{f}$ we now consider a matrix representation of $\mathscr{N}$, J_i, $B_i^\dagger$, and B_i and we shall use the defining one given in (2.3). From (2.3b) and (2.3d) we have that, *as matrices*, $B_i^\dagger B_j^\dagger$ and $B_i B_j$ vanish and thus

$$\exp(\bar{\mathbf{z}}\cdot\mathbf{B}^\dagger) = \begin{pmatrix} I & \overline{Z}\\ 0 & I\end{pmatrix}, \quad (5.5a)$$

$$\exp(\mathbf{z}'\cdot\mathbf{B}) = \begin{pmatrix} I & 0\\ Z' & I\end{pmatrix}, \quad (5.5b)$$

where all submatrices are 2×2 and

$$\overline{Z} = -i(\bar{\mathbf{z}}\cdot\boldsymbol{\sigma})\sigma_2 = \begin{pmatrix}\bar{z}_1 - i\bar{z}_2 & -\bar{z}_3\\ -\bar{z}_3 & -\bar{z}_1 - i\bar{z}_2\end{pmatrix}, \quad (5.6a)$$

$$Z' = i\sigma_2(\mathbf{z}'\cdot\boldsymbol{\sigma}) = \begin{pmatrix}-z_1' - iz_2' & z_3'\\ z_3' & z_1' - iz_2'\end{pmatrix}. \quad (5.6b)$$

Similarly we have

$$\exp(\mathbf{d}\cdot\mathbf{B}^\dagger) = \begin{pmatrix} I & D\\ 0 & I\end{pmatrix}, \quad (5.7a)$$

$$\exp(\mathbf{c}\cdot\mathbf{B}) = \begin{pmatrix} I & 0\\ C & I\end{pmatrix}, \quad (5.7b)$$

where

$$D = -i(\mathbf{d}\cdot\boldsymbol{\sigma})\sigma_2, \quad (5.8a)$$

$$C = i\sigma_2(\mathbf{c}\cdot\boldsymbol{\sigma}). \quad (5.8b)$$

Finally we can write

$$\exp(e\mathscr{N}) = \begin{pmatrix}\exp eI & 0\\ 0 & \exp(-eI)\end{pmatrix}, \quad (5.9a)$$

$$\exp(\mathbf{f}\cdot\mathbf{J}) = \begin{pmatrix}\exp F & 0\\ 0 & \exp(-\overline{F}^\dagger)\end{pmatrix}, \quad (5.9b)$$

where † indicates Hermitian conjugate and

$$F = \boldsymbol{\sigma}\cdot\mathbf{f}, \quad (5.10a)$$

$$\overline{F^\dagger} = \overline{(\boldsymbol{\sigma}\cdot\mathbf{f})^\dagger} = \overline{(\boldsymbol{\sigma}\cdot\bar{\mathbf{f}})} = \bar{\boldsymbol{\sigma}}\cdot\mathbf{f}, \quad (5.10b)$$

as the Pauli matrices are Hermitian.

Introducing the notation

$$G \equiv eI + F, \quad (5.11a)$$

$$\overline{G}^\dagger = eI + \overline{F}^\dagger, \quad (5.11b)$$

we see that Eq. (5.2) becomes

$$\begin{pmatrix} I & \overline{Z}\\ Z' & I + Z'\overline{Z}\end{pmatrix} = \begin{pmatrix}\exp G + D\exp(-\overline{G}^\dagger)C & D\exp(-\overline{G}^\dagger)\\ \exp(-\overline{G}^\dagger)C & \exp(-\overline{G}^\dagger)\end{pmatrix}, \quad (5.12)$$

from which we obtain

$$\exp(-\overline{G}^\dagger) = (I + Z'\overline{Z}), \quad (5.13a)$$

$$C = (I + Z'\overline{Z})^{-1}Z', \quad (5.13b)$$

$$D = \overline{Z}(I + Z'\overline{Z})^{-1}, \quad (5.13c)$$

$$\begin{aligned}\exp(G) &= I - \overline{Z}(I + Z'\overline{Z})^{-1}Z'\\ &= I - (I + \overline{Z}Z')^{-1}\overline{Z}Z' = (I + \overline{Z}Z')^{-1},\end{aligned} \quad (5.13d)$$

where we used the series expansion of $(I + \overline{Z}Z')^{-1}$ to obtain the result on the right-hand side of (5.13d).

Now we see that the reciprocal of the 2×2 matrix

$$X \equiv I + \overline{Z}Z' = I - (\boldsymbol{\sigma}\cdot\bar{\mathbf{z}})(\boldsymbol{\sigma}\cdot\mathbf{z}') \quad (5.14)$$

is given by

$$X^{-1}=\frac{1}{\det X}\begin{pmatrix} x_{22} & -x_{12} \\ -x_{21} & x_{11} \end{pmatrix}, \tag{5.15}$$

where we denote by $x_{\alpha\beta}$; $\alpha, \beta = 1,2$, the components of X. From (5.14) we have

$$\begin{aligned}\det X &= 1-\operatorname{tr}[(\boldsymbol{\sigma}\cdot\bar{\mathbf{z}})(\boldsymbol{\sigma}\cdot\mathbf{z}')]+\det[(\boldsymbol{\sigma}\cdot\bar{\mathbf{z}})(\boldsymbol{\sigma}\cdot\mathbf{z}')] \\ &= 1-2(\mathbf{z}'\cdot\bar{\mathbf{z}})+(\mathbf{z}'\cdot\mathbf{z}')(\bar{\mathbf{z}}\cdot\bar{\mathbf{z}}),\end{aligned} \tag{5.16}$$

while

$$x_{11}=1-\bar{z}_0 z'_0+2\bar{z}_{-1}z'_1, \tag{5.17a}$$

$$-x_{12}=\sqrt{2}\bar{z}_0 z'_{-1}-\sqrt{2}\bar{z}_{-1}z'_0, \tag{5.17b}$$

$$-x_{21}=-\sqrt{2}\bar{z}_1 z'_0+\sqrt{2}\bar{z}_0 z'_1, \tag{5.17c}$$

$$x_{22}=1-\bar{z}_0 z'_0+2\bar{z}_1 z'_{-1}, \tag{5.17d}$$

where in (5.17) we use the spherical components of the vectors, i.e.,

$$\begin{aligned} z'_{\pm 1} &= \mp(1/\sqrt{2})(z'_1 \pm i z'_2), \quad z'_0 = z'_3; \\ \bar{z}_{\pm 1} &= \mp(1/\sqrt{2})(\bar{z}_1 \pm i\bar{z}_2), \quad \bar{z}_0=\bar{z}_3. \end{aligned} \tag{5.18}$$

From (5.13) we have then that

$$\exp(G)=X^{-1}, \quad \exp(-\bar{G}^{\dagger})=\bar{X}^{\dagger}=\tilde{X}, \tag{5.19}$$

where $\sim$ stands for the transpose of the matrix.

We turn now our attention to

$$\exp(y'J_-), \quad \exp(\bar{y}J_+), \tag{5.20}$$

when we consider J_i, $i=1,2,3$, as the matrices given by (2.3c) so that

$$J_+=\begin{pmatrix}\sigma_+ & 0\\ 0 & \sigma_-\end{pmatrix}, \tag{5.21a}$$

$$J_-=\begin{pmatrix}\sigma_- & 0\\ 0 & -\sigma_+\end{pmatrix}, \tag{5.21b}$$

where $\sigma_\pm=\sigma_1\pm i\sigma_2$ are the 2×2 matrices

$$\sigma_+=\begin{pmatrix}0&1\\0&0\end{pmatrix}, \tag{5.22a}$$

$$\sigma_-=\begin{pmatrix}0&0\\1&0\end{pmatrix}. \tag{5.22b}$$

As $\sigma_\pm^2=0$, we have that

$$\begin{aligned}\exp(y'J_-)&=\begin{pmatrix}I&0\\0&I\end{pmatrix}+\begin{pmatrix}y'\sigma_-&0\\0&-y'\sigma_+\end{pmatrix}\\ &=\begin{pmatrix}Y'&0\\0&(\tilde{Y}')^{-1}\end{pmatrix},\end{aligned} \tag{5.23a}$$

$$\begin{aligned}\exp(\bar{y}J_+)&=\begin{pmatrix}I&0\\0&I\end{pmatrix}+\begin{pmatrix}\bar{y}\sigma_+&0\\0&-\bar{y}\sigma_-\end{pmatrix}\\ &=\begin{pmatrix}\bar{Y}&0\\0&(\tilde{\bar{Y}})^{-1}\end{pmatrix},\end{aligned} \tag{5.23b}$$

where

$$Y'=\begin{pmatrix}1&0\\y'&1\end{pmatrix}, \tag{5.24a}$$

$$\bar{Y}\equiv\begin{pmatrix}1&\bar{y}\\0&1\end{pmatrix}. \tag{5.24b}$$

The matrix that corresponds to the operator appearing in (5.4) becomes

$$\begin{pmatrix}Y'&0\\0&(\tilde{Y}')^{-1}\end{pmatrix}\begin{pmatrix}X^{-1}&0\\0&\tilde{X}\end{pmatrix}\begin{pmatrix}\bar{Y}&0\\0&(\tilde{\bar{Y}})^{-1}\end{pmatrix}=\begin{pmatrix}Y'X^{-1}\bar{Y}&0\\0&(\widetilde{Y'X^{-1}\bar{Y}})^{-1}\end{pmatrix}, \tag{5.25}$$

and thus we see, that the transformation associated with the operator appearing in (5.4) is characterized by the 2×2 matrix

$$\begin{aligned}Q&\equiv\begin{pmatrix}q_{11}&q_{12}\\q_{21}&q_{22}\end{pmatrix}\equiv Y'X^{-1}\bar{Y}=(\det X)^{-1}\begin{pmatrix}1&0\\y'&1\end{pmatrix}\begin{pmatrix}x_{22}&-x_{12}\\-x_{21}&x_{11}\end{pmatrix}\begin{pmatrix}1&\bar{y}\\0&1\end{pmatrix}\\ &=(\det X)^{-1}\begin{pmatrix}x_{22}&x_{22}\bar{y}-x_{12}\\ y'x_{22}-x_{21}&y'\bar{y}x_{22}-x_{21}\bar{y}-y'x_{12}+x_{11}\end{pmatrix}.\end{aligned} \tag{5.26}$$

The discussion carried between Eqs. (5.5) and (5.26) concerned the *matrices* associated with the defining representation of the four-dimensional symplectic Lie group. The matrix element of the *operator* appearing in (5.4) is then an element of the representation for the lowest weight state $|ws\rangle$ of (2.6) of a two-dimensional general complex linear group associated with the matrix Q of (5.26) for a partition $[h_1, h_2]$ which, from (2.7) and (3.8I), is given by

$$h_1=w+s, \quad h_2=w-s. \tag{5.27}$$

In the analysis of Louck[15] the lowest weight matrix element of this representation is given by a diamond pattern whose value is

$$\begin{aligned}\left\langle\begin{matrix} & h_2 & \\ h_1 & & h_2 \\ & h_2 & \end{matrix}\right\rangle &= (q_{22})^{h_1-h_2}(\det Q)^{h_2}\\ &=(q_{22})^{2s}(\det X)^{-(w-s)}.\end{aligned} \tag{5.28}$$

Substituting (5.17) in (5.26), we then conclude that the overlap becomes

$$\langle y'\mathbf{z}'|y\mathbf{z}\rangle=\mathscr{M}^{2s}\Delta^{-(w+s)}, \tag{5.29}$$

where $\mathscr{M}$ is given by

$$\mathscr{M}=(1+y'\bar{y})(1-\mathbf{z}'\cdot\bar{\mathbf{z}})-\sqrt{2}\epsilon^{qq'q''}v_q z'_{q'}\bar{z}_{q''}, \tag{5.30}$$

with ϵ being the antisymmetric tensor in indices

$q = 1,0,-1$, which is $1\ (-1)$ for an even (odd) permutation of $1,0,-1$, and zero for repeated indices. The three components $q = 1,0,-1$ of v_q are given by

$$v_1 = y', \quad v_0 = (1/\sqrt{2})(1 - y'\bar{y}), \quad v_{-1} = -\bar{y}, \qquad (5.31)$$

while

$$\Delta = \det X = [1 - 2\mathbf{z}'\cdot\bar{\mathbf{z}} + (\mathbf{z}'\cdot\mathbf{z}')(\bar{\mathbf{z}}\cdot\bar{\mathbf{z}})]. \qquad (5.32)$$

To construct the operator relations underlying Eq. (5.2), we used the defining matrix representation. In view of the parameters $y'\mathbf{z}'$ and $y\mathbf{z}$ we are dealing on the matrix level with the complex extension Sp(4,C) of the symplectic group. This offers no problem except for the assumption made in Eqs. (5.13b)–(5.13d) that $(I - Z'\bar{Z})$ be invertible. In an approach based on the analytic parametrization of cosets[16] it is shown that both Z' and Z must be restricted to a Siegel domain, that is, we require

$$(I - Z'Z'^{\dagger}) > 0, \quad (I - ZZ^{\dagger}) > 0. \qquad (5.33)$$

Then it can be shown that $(I - Z'\bar{Z})$ is invertible.

On the operator level we are dealing with an infinite-dimensional unitary representation of the real symplectic group. This representation in general cannot be extended to the complex symplectic group so that Eq. (5.2) cannot be justified by operator multiplication. This difficulty is resolved by noting that the operators are not applied to a general state in the representation space but rather to the lowest weight state.

Finally we note that Eq. (5.28) contains a half-integer power of the determinant of the complex matrix X. The prescription for this power is obtained from Bargmann's[17] analysis of the metaplectic representation.

In the next section we make use of the explicit expression (5.29) of the overlap and the relation (4.21) to determine the matrix elements of K^2 with respect to the boson states (2.18).

VI. MATRIX ELEMENTS OF K^2 WITH RESPECT TO THE BOSON STATES

We start by expressing the coherent state (3.1) in terms of the boson states (2.18). For this purpose we recall that the coherent state for the Weyl Lie algebra $\omega(1)$ can be developed in terms of one-dimensional harmonic oscillator states $|\nu)$ as

$$\exp(\bar{z}b^{\dagger})|0) = \sum_{\nu=0}^{\infty} \frac{\bar{z}^{\nu}}{\nu!}(b^{\dagger})^{\nu}|0) = \sum_{\nu=0}^{\infty}|\nu)(\nu|z), \qquad (6.1)$$

where $(\nu|z) = (\nu!)^{-1/2}\bar{z}^{\nu}$. A similar result holds for $\omega(3)$, but in this case we characterize the three-dimensional harmonic oscillator states by total number of quanta ν, angular momentum l, and projection μ, i.e.,[18]

$$\exp(\bar{\mathbf{z}}\cdot\mathbf{b}^{\dagger})|0) = \sum_{\nu l\mu} \overline{P_{\nu l\mu}(\mathbf{z})}P_{\nu l\mu}(\mathbf{b}^{\dagger})|0)$$
$$= \sum_{\nu l\mu}|\nu l\mu)(\nu l\mu|\mathbf{z}), \qquad (6.2)$$

where

$$|\nu l\mu) = P_{\nu l\mu}(\mathbf{b}^{\dagger})|0)$$
$$= A_{\nu l}(\mathbf{b}^{\dagger}\cdot\mathbf{b}^{\dagger})^{(\nu-l)/2}\mathscr{Y}_{l\mu}(\mathbf{b}^{\dagger})|0), \qquad (6.3)$$

with $\mathscr{Y}_{l\mu}(\nu)$ being the solid harmonic (3.10I) and $A_{\nu l}$ the normalization coefficient (5.9I). The bracket $(\mathbf{z}|\nu l\mu) = P_{\nu l\mu}(\mathbf{z})$ and $(\nu l\mu|\mathbf{z})$ is its conjugate.

We now note that the state $|s)$ of (3.2) is the lowest weight state of spin s and so we cannot apply to it powers of S_+ larger than $2s$. Thus[19]

$$\exp(\bar{y}S_+)|s) = \sum_{k=0}^{2s}(k!)^{-1}\bar{y}^k(S_+)^k|s)$$
$$= \sum_{\sigma=-s}^{s}\left\{\left[\frac{(s-\sigma)!}{(2s)!(s+\sigma)!}\right]^{1/2}(S_+)^{s+\sigma}|s)\right\}$$
$$\times\left\{\left[\frac{(2s)!}{(s+\sigma)!(s-\sigma)!}\right]^{1/2}\bar{y}^{s+\sigma}\right\}$$
$$= \sum_{\sigma=-s}^{s}|s\sigma)(s\sigma|y), \qquad (6.4)$$

where $|s\sigma)$, given by the first curly bracket, is now a spin state with projection $\sigma = -s, -s+1, \ldots, s$ while $(s\sigma|y) = \overline{(y|s\sigma)}$ is given by the second curly bracket.

Using the orthonormality property of Clebsch–Gordan coefficients when summed over jm we can now write

$$|y\mathbf{z}) = \sum_{\nu ljm}|\nu[ls]jm)(\nu[ls]jm|y\mathbf{z}), \qquad (6.5)$$

where, as in (5.10I), we have

$$|\nu[ls]jm) = \sum_{\mu,\sigma}\langle l\mu,s\sigma|jm\rangle P_{\nu l\mu}(\mathbf{b}^{\dagger})|s\sigma)$$
$$\equiv [P_{\nu l}(\mathbf{b}^{\dagger})x|s)]_{jm}, \qquad (6.6)$$

while

$$(y\mathbf{z}|\nu[ls]jm) = \overline{(\nu[ls]jm|y\mathbf{z})} = [(\mathbf{z}|\nu l)\times(y|s)]_{jm}. \qquad (6.7)$$

Note that if we substituted $y,\mathbf{z}$ by $y',\mathbf{z}'$ in (6.7) we can immediately check that $(\mathbf{z}'|\nu l\mu)$ and $(y'|s\sigma)$ are irreducible tensors with respect to the L_i' of (3.21a) and S_i' of (3.15), characterized, respectively, by l and s, and thus the state $(y'\mathbf{z}'|\nu[ls]jm)$ is associated with an irrep j of a SU(2) group whose generators are $J_i' = L_i' + S_i'$; $i = 1,2,3$.

As $(\mathbf{z}|\nu l\mu)$ and $(y|s\sigma)$ are orthonormal states[11,12] in the complex Hilbert space of measure (3.4) this holds also for $(y\mathbf{z}|\nu[ls]jm)$ of (6.7) and thus we can write

$$(\nu[l's]j|K^2|\nu[ls]j)$$
$$= \iint(\nu[l's]jm|y'\mathbf{z}')$$
$$\times d\mu(y',\mathbf{z}')(y',\mathbf{z}'|K^2|y\mathbf{z})d\mu(y,\mathbf{z})$$
$$\times(y\mathbf{z}|\nu[ls]jm), \qquad (6.8)$$

in which the matrix element is diagonal in ν, j and independent of m, as from (3.24a) and (3.24b), $(y'\mathbf{z}'|K^2|y\mathbf{z})$ is a scalar of the u(2) Lie subalgebra of sp(4, R). The volume element $d\mu(y,\mathbf{z})$ is given by

$$d\mu(y,\mathbf{z}) = \rho(y,\bar{y},\mathbf{z},\bar{\mathbf{z}})d\operatorname{Re}y\, d\operatorname{Im}y \times \prod_{i=1}^{3}(d\operatorname{Re}z_i\, d\operatorname{Im}z_i), \tag{6.9}$$

with ρ taking the value (3.4).

To evaluate (6.8) explicitly we first develop separately the two factors $\mathscr{M}^{2s}$ and $\Delta^{-(w+s)}$ appearing in $(y'\mathbf{z}'|K^2|y\mathbf{z})$ of (5.29) in terms of the complete set of orthonormal states (6.7), i.e.,

$$(y'\mathbf{z}'|K^2|y\mathbf{z}) = \sum_{\substack{\nu NL'L\\ JM\lambda\alpha}}\{(y'\mathbf{z}'|N[L's]JM)(N[L's]J|\mathscr{M}^{2s}|N[Ls]J)(N[Ls]JM|y\mathbf{z})\}$$
$$\times\{(\mathbf{z}'|\nu-N\lambda\alpha)(\nu-N,\lambda|\Delta^{-(w+s)}|\nu-N,\lambda)(\nu-N,\lambda\alpha|\mathbf{z})\}, \tag{6.10}$$

where again we use the fact that $\mathscr{M}^{2s}$ and $\Delta^{-(w+s)}$ separately are also scalars of the u(2) Lie subalgebra.

Using the orthonormality property of Clebsch–Gordan coefficients we can then write (6.10) as

$$(y'\mathbf{z}'|K^2|y\mathbf{z}) = \sum_{\nu jm}\sum_{NL'L}\sum_{\lambda J}[(\mathbf{z}'|\nu-N\lambda)\times[(\mathbf{z}'|NL')\times(y'|s)]_J]_{jm}$$
$$\times(N[L's]J|\mathscr{M}^{2s}|N[Ls]J)(\nu-N\lambda|\Delta^{-(w+s)}|\nu-N\lambda)\ \overline{[(\mathbf{z}|\nu-N\lambda)\times[(\mathbf{z}|NL)\times(y|s)]_J]_{jm}}\,. \tag{6.11}$$

We now carry a recoupling of the states and a reduction of the products $[(\mathbf{z}|\nu-N\lambda)\times(\mathbf{z}|NL)]_{l\mu}$ to get

$$[(\mathbf{z}|\nu-N\lambda)\times[(\mathbf{z}|NL)\times(y|s)]_J]_{jm}$$
$$=\sum_l[(2l+1)(2J+1)]^{1/2}(-1)^{l+J-\lambda-s}W(ljLJ;s\lambda)H(\lambda,L,l)$$
$$\times A_{\nu-N\lambda}A_{NL}(A_{\nu l})^{-1}(y\mathbf{z}|\nu[ls]jm), \tag{6.12}$$

where

$$A_{\nu l} = (-1)^{(\nu-l)/2}(4\pi)^{1/2}[(\nu+l+1)!!(\nu-l)!!]^{-1/2}, \tag{6.13a}$$
$$H(l'l''l) = [(2l'+1)(2l''+1)]^{1/2}[4\pi(2l+1)]^{-1/2}\langle l'0,l''0|l0\rangle, \tag{6.13b}$$

and W is a Racah coefficient.

Substituting (6.12) in the left- and right-hand side of (6.11) and going back to expressions (6.6) for the matrix elements of K^2 with respect to the boson states, we obtain

$$(\nu[l's]j|K^2|\nu[ls]j)$$
$$=\sum_{NL'L\lambda J}\{(-1)^{l'+J-\lambda-s}[(2l'+1)(2J+1)]^{1/2}W(l'jL'J;s\lambda)A_{\nu-N\lambda}A_{NL'}(A_{\nu l'})^{-1}(-1)^{l+J-\lambda-s}$$
$$\times[(2l+1)(2J+1)]^{1/2}W(ljLJ;s\lambda)A_{\nu-N\lambda}A_{NL}(A_{\nu l})^{-1}H(\lambda,L',l')H(\lambda,L,l)$$
$$\times(N[L's]J|\mathscr{M}^{2s}|N[Ls]J)(\nu-N\lambda|\Delta^{-(w+s)}|\nu-N\lambda)\}. \tag{6.14}$$

We now have to obtain the matrix elements of $\mathscr{M}^{2s}$ and $\Delta^{-(w+s)}$ with respect to the states of type (6.6). We start by introducing what we could call the cylindrical states in the space y, z_i through the definition

$$(y\mathbf{z}|N\tau\mu s\sigma)\equiv\frac{(z_1)^{(N-\tau+\mu)/2}(z_0)^{\tau}(z_{-1})^{(N-\tau-\mu)/2}(y|s\sigma)}{[((N-\tau+\mu)/2)!\tau!((N-\tau-\mu)/2)!]^{1/2}}\,. \tag{6.15}$$

The states (6.6) can now be expanded in terms of (6.15) as

$$(y\mathbf{z}|N[Ls]JM) = \sum_{\mu\sigma}\sum_{\tau}\{(y\mathbf{z}|N\tau\mu s\sigma)\langle L\mu,s\sigma|JM\rangle(N\tau\mu|NL\mu)\}, \tag{6.16}$$

where the angular bracket is a Clebsch–Gordan coefficients, while the last round one is the transformation that relates the cylindrical and spherical states of the three-dimensional oscillator[20]

$$(N\tau\mu|NL\mu) = (-1)^{(N-L)/2}2^{(N-\tau)/2-L}[(2L+1)(L-\mu)!\tau!((N+\mu-\tau)/2)!]^{1/2}$$
$$\times[(N+L+1)!!(N-L)!!(L+\mu)!((N-\mu-\tau)/2)!]^{-1/2}$$
$$\times\sum_{k=0}^{L}\frac{(-1)^k\left(\frac{N-L}{2}+k\right)!(2L-2k)!}{k!(L-k)!(L-2k-\mu)!}\left[\left(\frac{\mu+\tau-L}{2}+k\right)!\right]^{-1}. \tag{6.17}$$

We then have that

$$(N[L's]J|\mathscr{M}^{2s}|N(Ls)J) = \iint (N[L's]JM|y'\mathbf{z}')d\mu(y',\mathbf{z}')(y'\mathbf{z}'|\mathscr{M}^{2s}|y\mathbf{z})d\mu(y,\mathbf{z})(y\mathbf{z}|N[Ls]JM)$$

$$= \sum_{\tau'\tau}\sum_{\mu'\mu}\sum_{\sigma'\sigma}\{\langle L'\mu',s\sigma'|JM\rangle(NL'\mu'|N\tau'\mu')\langle L\mu,s\sigma|JM\rangle(N\tau\mu|NL\mu)(N\tau'\mu's\sigma'|\mathscr{M}^{2s}|N\tau\mu s\sigma)\}, \tag{6.18}$$

where

$$(N\tau'\mu's\sigma'|\mathscr{M}^{2s}|N\tau\mu s\sigma) = \iint \{(N\tau'\mu's\sigma'|y'\mathbf{z}')d\mu(y',\mathbf{z}')(y'\mathbf{z}'|\mathscr{M}^{2s}|y\mathbf{z})d\mu(y,\mathbf{z})(y\mathbf{z}|N\tau\mu s\sigma)\}. \tag{6.19}$$

The matrix element (6.19) is easy to evaluate because of the orthonormality of powers of y and z_q, $q = 1,0,-1$, with appropriate coefficients[11,12] with those of the corresponding $\bar{y}$ and $\bar{z}_q$. Thus developing $\mathscr{M}^{2s}$ in terms of $y', z_q', \bar{y}$, and $\bar{z}_q$ with the help of the binomial theorem, we get

$$\begin{aligned}
&(N\tau'\mu's\sigma'|\mathscr{M}^{2s}|N\tau\mu s\sigma)\\
&\quad = (-1)^{s+\mu+\mu'+\sigma}2^{N-(\tau+\tau')/2}[(s+\sigma)!(s-\sigma)!(s+\sigma')!(s-\sigma')!]^{1/2}\\
&\quad\times\left[\left(\frac{N-\tau'-\mu'}{2}\right)!\tau'!\left(\frac{N-\tau'+\mu'}{2}\right)!\left(\frac{N-\tau-\mu}{2}\right)!\tau!\left(\frac{N-\tau+\mu}{2}\right)!\right]^{1/2}\\
&\quad\times\sum_{\gamma}\left(\frac{\tau+\tau'-\sigma-\sigma'}{2}-N+s+\gamma\right)!\left[(N-2s-\sigma-\sigma'+2\gamma)!\left(\frac{\tau+\tau'-\sigma-\sigma'}{2}-s+\gamma\right)!\right]^{-1}D_{\gamma},
\end{aligned} \tag{6.20}$$

where

$$\begin{aligned}
D_{\gamma} = \sum_{d}&\left[\left(s-\sigma-\frac{N-\tau+\mu}{2}-d\right)!d!\left(-s+\gamma-\sigma'+d+\frac{N-\tau+\mu}{2}\right)!\right.\\
&\times\left(\gamma-\frac{N-\tau'-\mu'}{2}+d\right)!\left(\frac{N-\tau'-\mu'}{2}-d\right)!(s+\sigma'-\gamma-d)!\\
&\left.\times\left(\frac{\tau'-\tau-\mu+\mu'}{2}+d\right)!\left(s-\gamma+\frac{\sigma'+\sigma+\tau-\tau'}{2}-d\right)!\right]^{-1}.
\end{aligned} \tag{6.21}$$

Turning our attention now to $\Delta^{-(w+s)}$, we show in the Appendix that

$$\langle\nu-N\lambda|\Delta^{-(w+s)}|\nu-N\lambda\rangle = \frac{(\nu-N+2w+2s-\lambda-3)!!(\nu-N+2w+2s+\lambda-2)!!}{(2w+2s-2)!!(2w+2s-3)!!}. \tag{6.22}$$

Substituting (6.20) in (6.18), and the latter together with (6.22) in (6.14) we get the explicit form of the matrix elements of K^2 with respect to the boson states (6.6).

As indicated in paper I, once we have the matrix elements of K^2 with respect to the boson states (6.6) our problem is to find those of K. As K^2 is diagonal in ν, j, and m, we have to deal only with the finite Hermitian matrices in l' and l for fixed values of ν and j, where l' and l are restricted by

$$|j-s|\leqslant l',\ l\leqslant j+s;\quad \nu-l,\ \nu-l'\ \text{are even}. \tag{6.23}$$

To obtain the matrix K we have to diagonalize K^2, take its square root (which is feasible as all the eigenvalues of K^2 are non-negative), and then return K to the original basis. All of this is possible numerically but unfortunately not analytically if the secular equation is of order higher than 4 and thus an explicit analytic boson realization of sp(4, R) cannot be carried out as we stressed in paper I. The expression (6.14) for the matrix elements of K^2 with respect to the states (6.6) is, though, fundamental for the calculation of the matrix elements of the generators of a sp(4, R) Lie algebra with respect to the basis of irreps of the corresponding group, as we proceed to show in the next section.

VII. MATRIX ELEMENTS OF THE GENERATORS OF sp(4, R) WITH RESPECT TO THE IRREPS OF THE CORRESPONDING GROUP

By exactly the same analysis as that of Sec. VI, we see that the coherent state (4.1) associated with a sp(4, R) Lie algebra can be expanded as

$$|y\mathbf{z}\rangle = \sum_{\nu ljm}|\nu[ls]jm\rangle(\nu[ls]jm|y\mathbf{z}), \tag{7.1}$$

where the round bracket is still given by (6.7), while the angular ket now takes the form

$$|\nu[ls]jm\rangle \equiv \sum_{\mu\sigma}\langle l\mu,s\sigma|jm\rangle P_{\nu l\mu}(\mathbf{B}^{\dagger})|ws\sigma\rangle, \tag{7.2}$$

where

$$|ws\sigma\rangle = \left[\frac{(s-\sigma)!}{(2s)!(s+\sigma)!}\right]^{1/2}J_{+}^{s+\sigma}|ws\rangle. \tag{7.3}$$

As we indicated in paper I, the states (7.2) are basis for the irreps of the chain of groups

$$\underset{(w,s)}{\mathrm{Sp}(4,R)}\supset\mathrm{U}(2) = \underset{\nu}{\mathrm{U}(1)}\times\underset{j}{\mathrm{SU}(2)};\quad \underset{j}{\mathrm{SU}(2)}\supset\underset{m}{\mathrm{O}(2)}, \tag{7.4}$$

where underneath each group we give the quantum number characterizing its irrep. The number l is a multiplicity index that distinguishes the different states in a given irrep (ws) of $sp(4,R)$ that have the same irrep ν, j of $U(1)\times SU(2)$.

The basis (7.2) is not orthonormal[8] and thus to be able to use it we need first to calculate the overlap of its states, i.e.,

$$\langle \nu[l's]j|\nu[ls]j\rangle, \tag{7.5}$$

where as ν, j, and m are associated with irreps of the groups U(1), SU(2), and O(2), they must be the same in bra and ket and the overlap is independent of m.

From the orthonormality properties of the $(\mathbf{yz}|\nu[ls]jm)$ discussed in Sec. VI, we immediately conclude that

$$\langle \nu[l's]j|\nu[ls]j\rangle = \int\int\{(\nu[l's]jm|y'\mathbf{z}')d\mu(y',\mathbf{z}')\langle y'\mathbf{z}'|y\mathbf{z}\rangle \times d\mu(y,\mathbf{z})(y\mathbf{z}|\nu[ls]jm)\}, \tag{7.6}$$

so that from (4.21) and (6.8) we get

$$\langle \nu[l's]j|\nu[ls]j\rangle = (\nu[l's]j|K^2|\nu[ls]j), \tag{7.7}$$

where the right-hand side of (7.7) is given by (6.14).

Thus we have an explicit, analytic, and closed expression for the overlap (7.5), which for fixed ν, j, and m gives the elements of a finite Hermitian matrix, as l' and l are restricted by (6.23). The unitary matrix that diagonalizes the overlap matrix, together with the square roots of the eigenvalues of the overlap matrix, will allow us to transform the nonorthonormal basis (7.2) into an orthonormal one.

It is sufficient now to calculate the matrix elements of the generators of the $sp(4,R)$ Lie algebra with respect to the states (7.1), as the explicit form of the overlap matrix will allow us to evaluate them and, if we wish, to transform them to an orthonormal basis by the procedure indicated in the previous paragraph.

We have then, using spherical components with index $q=1,0,-1$ for the vectors, that

$$\langle \nu[l's]jm|\mathcal{N}|\nu[ls]jm\rangle = \nu\langle \nu[l's]j|\nu[ls]j\rangle, \tag{7.8a}$$

$$\langle \nu[l's]jm'|Jq|\nu[ls]jm\rangle = [j(j+1)]^{1/2}\langle jm,|q|j'm'\rangle\langle \nu[l's]j|\nu[ls]j\rangle, \tag{7.8b}$$

$$\langle \nu+1[l's]j'm'|B_q^\dagger|\nu[ls]jm\rangle = \overline{\langle \nu[ls]jm|B^q|\nu+1[l's]jm\rangle} = \langle jm,1q|j'm'\rangle\langle \nu+1[l's]j'\|B^\dagger\|\nu[ls]j\rangle, \tag{7.8c}$$

where

$$\begin{aligned}\langle \nu+1[l's]j\|B^\dagger\|\nu[ls]j\rangle &= (-1)^{l+j-s}(2j+1)^{1/2}\\ &\times\{W(l+1,l,j',j;1s)[(\nu+l+3)(l+1)]^{1/2}\\ &\times\langle \nu+1[l's]j'|\nu+1[l+1,s]j'\rangle\\ &+W(l-1,l,j',j;1s)[(\nu-l+2)l]^{1/2}\\ &\times\langle \nu+1[l's]j'|\nu+1[l-1,s]j'\rangle\}.\end{aligned} \tag{7.8d}$$

The values (7.8a) and (7.8b) are obvious, while (7.8c) and (7.8d) can be calculated by the same recoupling techniques as the matrix element of a creation operators applied to states of a particle with spin s in a three-dimensional harmonic oscillator potential.[18] Note that the matrix element of B_q can be obtained by Hermitian conjugation, if we lower the index in (7.8c), i.e., $B^q=(-1)^q B_{-q}$.

With the matrix elements (7.8) we can carry out calculations for spectra and shape of many body systems in a physical space of two dimensions.

VIII. CONCLUSION

We showed in the present paper that the matrix elements of the operator K^2 with respect to boson states, which are essential for the boson realization of the $sp(4,R)$ Lie algebra, can be obtained in closed form from the overlap of $sp(4,R)$ coherent states. Furthermore we used these matrix elements for the explicit determination of those of the generators of the $sp(4,R)$ Lie algebra with respect to the basis of irreps of the positive discrete series for the corresponding group.

Our analysis can be extended to $sp(2d,R)$ where the overlap of coherent states can be obtained by reasoning similar to that presented in Sec. V. In fact this overlap has already been determined for the case of $sp(6,R)$ by Kramer and Papadopolos[13] and by Quesne.[14] The determination of the matrix elements of K^2 with respect to the boson states corresponding to $sp(6,R)$ can be obtained by a procedure similar to that developed in Sec. VI, though it will be more complex, as now the expression corresponding to (5.29) will have three instead of two factors and the recoupling techniques involve the SU(3) instead of the SU(2) group.

ACKNOWLEDGMENT

Two of the authors (O. C. and M. M.) are indebted to C. Quesne for many helpful discussions on the subject of this paper.

APPENDIX: DETERMINATION OF $(\mathbf{z}'|\Delta^{-(w+s)}|\mathbf{z})$

The function $(\mathbf{z}'|\Delta^{-(w+s)}|\mathbf{z})$ can be expanded as follows:

$$(\mathbf{z}'|\Delta^{-(w+s)}|\mathbf{z}) = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\binom{-w-s}{k}\binom{-w-s-k}{l}(-2)^l u^l v^k, \tag{A1}$$

where the binomial coefficient $\binom{p}{n}\equiv[p(p-1)\cdots(p-n+1)]/n!$, while $u\equiv\mathbf{z}'\cdot\bar{\mathbf{z}}$, and $v\equiv(\mathbf{z}'\cdot\mathbf{z}')(\bar{\mathbf{z}}\cdot\bar{\mathbf{z}})$.

Using the addition theorem of spherical harmonics it is straightforward to show that

$$u^l = l!4\pi{\sum_\lambda}'\sum_{\alpha=-\lambda}^{\lambda}\frac{\mathcal{Y}_{\lambda\alpha}(\mathbf{z}')\overline{\mathcal{Y}_{\lambda\alpha}(\mathbf{z})}}{(l-\lambda)!!(l+\lambda+1)!!}, \tag{A2}$$

where the prime in the symbol of sum indicates that $\lambda=l, l-2,\ldots,0$ or 1 depending if l is even or odd.

Substituting Eq. (A2) in (A1), simplifying the binomial coefficients, and with the change of index $l=\gamma-2k$, we get

$$(\mathbf{z}'|\Delta^{-(w+s)}|\mathbf{z}) = \sum_{\gamma=0}^{\infty}\sum_{\lambda}{}' \sum_{\alpha=-\lambda}^{\lambda}\left\{\sum_{k=0}^{\infty}\frac{(-1)^k(2\gamma+2w+2s-2k-2)!!}{k!2^k(\gamma-2k-\lambda)!!}\frac{1}{(\gamma-2k+\lambda+1)!!}\right\}$$

$$\times\frac{4\pi}{|A_{\gamma\lambda}|^2(2w+2s-2)!!}(\mathbf{z}'|\gamma\lambda\alpha)(\gamma\lambda\alpha|\mathbf{z}), \tag{A3}$$

where the coefficient $A_{\gamma\lambda}$ is given by (6.13a) and $(z|\gamma\lambda\alpha)$ by

$$(\mathbf{z}|\gamma\lambda\alpha) = A_{\gamma\lambda}(\mathbf{z}\cdot\mathbf{z})^{(\gamma-\lambda)/2}\mathscr{Y}_{\lambda\alpha}(\mathbf{z}). \tag{A4}$$

The curly bracket in Eq. (A3) can be identified with an hypergeometric function ${}_2F_1(a,b;c;1)$, i.e.,

$$\{\ \} = \frac{(2\gamma+2w+2s-2)!!}{(\gamma+\lambda+1)!!(\gamma-\lambda)!!}\times{}_2F_1\left[-\frac{\gamma+\lambda+1}{2},-\frac{\gamma-\lambda}{2};-\gamma-w+1;1\right]. \tag{A5}$$

This hypergeometric function can be summed[21]

$${}_2F_1(a,-m,c;1) = \frac{(c-a)_m}{(c)_m}, \tag{A6}$$

where the symbol $(a)_m \equiv a(a+1)\cdots(a+m-1)$.

From the equations (A3) and (A5) we find that

$$(\mathbf{z}'|\Delta^{-(w+s)}|\mathbf{z}) = \sum_{\gamma,\lambda,\alpha}\frac{(\gamma+2w+2s-\lambda-3)!!(\gamma+2w+2s+\lambda-2)!!}{(2w+2s-2)!!(2w+2s-3)!!}\times(\mathbf{z}'|\gamma\lambda\alpha)(\gamma\lambda\alpha|\mathbf{z}). \tag{A7}$$

Taking the scalar product with $(\gamma\lambda\alpha|\mathbf{z}')(\mathbf{z}|\gamma\lambda\alpha)$, we arrive at the result given in Eq. (6.22).

[1]G. Rosensteel and D. J. Rowe, Ann. Phys. NY **96**, 1 (1976); **123**, 36 (1978); **126**, 198 (1981).

[2]K. T. Hecht and J. P. Elliott, Nucl. Phys. A **438**, 29 (1985); K. T. Hecht, *Proceedings of the VIIIth Symposium on Nuclear Physics*, Oaxtepec, Mexico (Instituto de Física, UNAM, Mexico City, 1985); Nucl. Phys. A **444**, 189 (1985).

[3]L. D. Mlodinow and N. Papanicolau, Ann. Phys. NY **128**, 314 (1980); **131**, 1 (1981).

[4]J. Deenen and C. Quesne, J. Math. Phys. **23**, 878, 2004 (1982); **25**, 1638, 2354 (1984); J. Phys. A: Math. Gen. **16**, 2095 (1983).

[5]D. J. Rowe, J. Math. Phys. **25**, 2662 (1984); D. J. Rowe, G. Rosensteel, and R. Carr, J. Phys. A: Math. Gen. **17**, L399 (1984).

[6]M. Moshinsky, J. Phys. A: Math. Gen. **18**, L1 (1985).

[7]O. Castaños, P. Kramer, and M. Moshinsky, J. Phys. A: Math. Gen. **18**, L493 (1985).

[8]O. Castaños, E. Chacón, M. Moshinsky, and C. Quesne, J. Math. Phys. **26**, 2107 (1985).

[9]M. Moshinsky, E. Chacón, and O. Castaños, in *Proceedings of the XIII Colloquium on Group Theoretical Methods in Physics* (World Scientific, Singapore, 1984).

[10]R. Gilmore, *Lie Groups, Lie Algebras and Some of Their Applications* (Wiley–Interscience, New York, 1974), p. 188.

[11]V. Bargmann, Commun. Pure Appl. Math. **14**, 187 (1961); **20**, 1 (1967).

[12]P. Kramer and M. Saraceno, *Geometry of the Time-Dependent Variational Principle in Quantum Mechanics* (Springer, New York, 1981), p. 50.

[13]P. Kramer and Z. Papadopolos, "Hilbert Spaces of Analytic Functions and Representations of the Positive Discrete Series of Sp(6, R)," preprint, University of Tübingen, 1985.

[14]C. Quesne, "Coherent states of the real symplectic group in a complex analytic parametrization. I. Perelomov coherent states," preprint PTM 85-04, Universite Libre de Bruxelles, Belgium.

[15]J. D. Louck, Am. J. Phys. **38**, 3 (1970).

[16]P. Kramer, Ann. Phys. NY **141**, 254 (1982).

[17]V. Bargmann, "Group representations in Hilbert spaces of analytic functions," in *Analytic Methods in Mathematical Physics*, edited by R. P. Gilbert and R. G. Newton (Gordon and Breach, New York, 1968).

[18]M. Moshinsky, *The Harmonic Oscillator in Modern Physics: From Atoms to Quarks* (Gordon and Breach, New York, 1969).

[19]E. M. Rose, *Elementary Theory of Angular Momentum* (Wiley, New York, 1957).

[20]E. Chacón and M. de Llano, Rev. Mex. Fis. **12**, 57 (1963).

[21]L. J. Slater, *Generalized Hypergeometric Functions* (Cambridge U. P., New York, 1966), p. 243.

The Fourier transform of confining potentials

W. D. Heiss and G. M. Welke
Nuclear Theory Research Group of the Department of Physics, University of the Witwatersrand, Johannesburg, South Africa

(Received 21 August 1985; accepted for publication 26 November 1985)

The precise meaning of the Fourier transform of $|x|^\nu$ is examined. A general expression is given for real positive ν. For odd ν, derivatives of principal value integrals are obtained, while even ν gives rise to derivatives of the delta function.

Within the framework of quantum chromodynamics (QCD) the basic quark–quark interaction is the one gluon exchange. In addition, a widely accepted feature of QCD is that it leads to confinement. As a consequence, the static part of the effective quark–quark interaction is expected to increase at large distances. A simplistic argument leads to a linear power law[1,2] as this is the one-dimensional Fourier transform of the gluon propagator q^{-2}; the one-dimensional transform is taken to mimic confinement. The actual asymptotic behavior may well be different from a linear power law.[3,4] For a one-dimensional quark gas, sensible mean field solutions can only be obtained for $|x|^\nu$, $\nu<1$; in this case a clustering solution has been found to exist with an energy lower than that for the uniform solution.[5]

In this paper we address ourselves to the mathematical problem with regard to potentials that increase at large distances. In particular, the precise mathematical meaning of their Fourier transform is scrutinized. This is of importance since many practical calculations are performed in the momentum representation. We restrict ourselves to one dimension and consider

$$V(x)=|x|^\nu,\quad \nu\geqslant 0. \tag{1}$$

The Fourier transform does not exist *a priori*. A tentative definition that is physically attractive is to consider the limit of a screened potential, viz.

$$V(q)=\lim_{\mu\to0}\int_{-\infty}^{\infty} e^{-iqx}|x|^\nu e^{-\mu|x|}\,dx,\quad \mu>0, \tag{2}$$

which yields

$$V(q)=2\Gamma(\nu+1)\lim_{\mu\to0}\frac{\cos((\nu+1)\arctan(q/\mu))}{(q^2+\mu^2)^{(\nu+1)/2}} = -\frac{2\Gamma(\nu+1)\sin(\pi\nu/2)}{|q|^{\nu+1}}. \tag{3}$$

Strictly speaking the limit is valid only for $q\neq0$. It is exactly at $q=0$, where a precise meaning has to be given to $V(q)$.

In actual applications one is usually faced with expressions of the nature

$$f(k)=\int_{-\infty}^{\infty} dq\; V(k-q)g(q), \tag{4}$$

where we now assume $g(q)$ to be smooth and sufficiently decreasing for large q. We thus consider

$$f(k)=\Gamma(\nu+1)\lim_{\mu\to0}\int_{-\infty}^{\infty} dq\left[\frac{1}{(\mu-i(k-q))^{\nu+1}}+\frac{1}{(\mu+i(k-q))^{\nu+1}}\right]g(q). \tag{5}$$

Integration by parts yields

$$f(k)=\frac{\Gamma(\nu+1)}{\nu}\lim_{\mu\to0}\int_{-\infty}^{\infty} dq\left[\frac{(-i)^{\nu+1}}{(q-k-i\mu)^\nu}+\frac{i^{\nu+1}}{(q-k+i\mu)^\nu}\right]g'(q). \tag{6}$$

To facilitate the discussion we restrict ourselves for the moment to $0<\nu<1$ and set $k=0$. Under the assumption that $g'(q)$ is analytic in a μ neighborhood of $q=0$, the limit can be rewritten as

$$\frac{\Gamma(\nu+1)}{\nu}\lim_{\mu\to0}\Bigg(\int_{\mu}^{\infty} dq\frac{2\cos\frac12\pi(\nu+1)}{q^\nu}g'(q)+\int_{-\infty}^{-\mu} dq\left[\frac{(-i)^{\nu+1}}{|q|^\nu e^{-i\pi\nu}}+\frac{i^{\nu+1}}{|q|^\nu e^{i\pi\nu}}\right]g'(q)+\int_{C_1} dq\frac{(-i)^{\nu+1}}{q^\nu}g'(q)+\int_{C_2} dq\frac{i^{\nu+1}}{q^\nu}g'(q)\Bigg), \tag{7}$$

where the contours C_1 and C_2 are semicircles of radius μ around $q=0$ in the lower and upper plane, respectively. On the same footing, the phases for negative values for q in the second integral are $e^{-i\pi}$ and $e^{i\pi}$, respectively. Since, for $\nu<1$, $q^{-\nu}$ is integrable at zero, the limit can be taken for the first two integrals, while each contour integral is of order $\mu^{1-\nu}$. We eventually obtain

$$f(k)=\frac{2\Gamma(\nu+1)}{\nu}\sin\frac{1}{2}\pi\nu\int_{-\infty}^{\infty} dq\,\frac{\operatorname{sgn}(k-q)}{|k-q|^\nu}g'(q), \tag{8}$$

which is well defined if the derivative of g is well behaved. The cases $\nu=0$ and $\nu=1$ follow at once:

$$f(k)=2\pi g(k),\quad \text{for } \nu=0,$$

as anticipated from $V(q)=2\pi\delta(q)$, and

$$f(k)=2\fint_{-\infty}^{\infty}\frac{g'(q)}{k-q}dq,\quad \text{for } \nu=1.$$

Note that in the latter case the two contour integrals in Eq. (7) cancel in the limit $\mu\to0$. For the principal value integral to be well defined, the derivative of $g(k)$ must actually be Hölder continuous.[6] In this case we can also use ($\nu=1$)

$$f(k)=2\frac{d}{dk}\fint_{-\infty}^{\infty}\frac{g(q)}{k-q}dq.$$

It is now a simple matter to apply the same principle for $\nu>1$. Integrating in Eq. (5) by parts n times for $n-1<\nu<n$, we obtain

 0022-2488/86/040936-02$02.50

$$f(k) = \frac{-2\Gamma(\nu+1)}{\nu(\nu-1)\cdots(\nu-n+1)} \sin\frac{1}{2}\pi\nu \times \int_{-\infty}^{\infty} dq \frac{[\mathrm{sgn}(q-k)]^n}{|k-q|^{\nu-n+1}} g^{(n)}(q). \quad (9)$$

For even ν it is known that

$$V(q) = (-1)^{\nu/2} 2\pi \delta^{(\nu)}(q). \quad (10)$$

This is in accordance with Eq. (9), where

$$f(k) = (-1)^{\nu/2} 2\pi g^{(\nu)}(k) \quad (11)$$

is obtained for $\nu \to n-1$ when Eq. (9) is used for odd n. It is instructive to compare Eq. (3) with Eq. (10). Likewise, for $\nu \to n$, n odd, in Eq. (9), we find

$$f(k) = 2(-1)^{(\nu-1)/2} \left(\frac{d}{dk}\right)^{\nu} ⨍_{-\infty}^{\infty} \frac{g(q)}{k-q} dq. \quad (12)$$

Thus we have found a precise meaning of the distribution $V(q)$ being the Fourier transform of $|x|^{\nu}$. The physically interesting case is probably $\nu < 1$, but larger values of ν have also been used.[3]

ACKNOWLEDGMENT

One of us (G. M. W.) was supported by a bursary from C. S. I. R., South Africa.

[1] F. E. Close, *An Introduction to Quarks and Partons* (Academic, New York, 1979).
[2] T. D. Lee, *Particle Physics and Introduction to Field Theory* (Harwood, New York, 1981).
[3] V. Bernard, R. Brockman, M. Schaden, W. Weise, and E. Werner, Nucl. Phys. A **412**, 349 (1984).
[4] C. Quigg, *Proceedings of the Fourth Advanced Course in Theoretical Physics*, Stellenbosch 1985, to appear in *Lecture Notes in Physics* (Springer, Berlin).
[5] G. M. Welke and W. D. Heiss, to be published in Phys. Rev. D.
[6] N. F. Mushkhelishvili, *Singular Integral Equations* (Noordhoff, Groningen, 1953) (English translation).

A class of sums of Gegenbauer functions: Twenty-four sums in closed form

H. van Haeringen
Department of Mathematics and Informatics, Delft University of Technology, Delft, The Netherlands

(Received 30 May 1985; accepted for publication 28 October 1985)

Series of the type $\Sigma_{n=0}^{\infty}(n+\lambda)\,[C_n^\lambda(1)]^{2-k-l-m}\Pi_{h=1}^{k} C_n^\lambda(x_h)\cdot\Pi_{i=1}^{l} D_n^\lambda(z_i)\cdot\Pi_{j=1}^{m}\mathfrak{D}_n^\lambda(u_j)$ are studied. Here C_n^λ is Gegenbauer's polynomial, also called the "ultraspherical polynomial," and D_n^λ and $\mathfrak{D}_n^\lambda$ are Gegenbauer functions of the second kind. Interrelationships and analytic properties are discussed, and closed forms for 24 of these sums are given, more than half of which are new.

I. INTRODUCTION, DEFINITIONS, AND SUMMARY

A. Introduction and summary

Series[1] and integrals[2–5] of Gegenbauer[6,7] (or ultraspherical[8,9]) functions and associated Legendre[10] functions have been studied for a long time. Closed formulas for such series and integrals have proved useful in physical applications. On the other hand, sometimes a new "purely mathematical" closed formula follows from physical considerations.[11]

Recently there has been much interest in series and integrals of products of Gegenbauer and Legendre functions; see, e.g., Askey,[12] Durand,[13,14] Rahman,[15,16] Din,[17] van Haeringen,[18–20] Askey *et al.*,[21] and Rahman and Shah.[22,23]

In this paper we study a particular class S_{klm}^λ (to be defined shortly) of sums of products of Gegenbauer functions. We obtain a large number of these sums in closed form and discuss various interesting aspects. Especially when the variables are situated on the interval $(-1,1)$ the evaluation of certain sums can be tricky. In particular, S_{310}^λ turns out to have a closed form that is quite complicated, even more than claimed by Rahman and Shah[23]; we correct the closed form for S_{310}^λ (in our notation) given by these authors.

Let us define the series

$$
\begin{aligned}
S_{klm}^\lambda &:= S_{klm}^\lambda(x_1,\dots,x_k,z_1,\dots,z_l,u_1,\dots,u_m)\\
&:= a_\lambda^{-1}\sum_{n=0}^{\infty}(n+\lambda)[C_n^\lambda(1)]^{2-k-l-m}\\
&\quad\times\prod_{h=1}^{k}C_n^\lambda(x_h)\cdot\prod_{i=1}^{l}D_n^\lambda(z_i)\cdot\prod_{j=1}^{m}\mathfrak{D}_n^\lambda(u_j). \qquad (1.1)
\end{aligned}
$$

Here $\Pi_{i=1}^{0}:=1$,

$$a_\lambda^{-1}:=1/a_\lambda:=2\pi^{1/2}\Gamma(\lambda)/\Gamma(\lambda+\tfrac{1}{2})$$

is included for convenience, and C_n^λ, D_n^λ, and $\mathfrak{D}_n^\lambda$ are Gegenbauer functions to be defined below. In particular,

$$C_\nu^\lambda(1)=(2\lambda)_\nu/\Gamma(\nu+1),$$

where the shifted factorial $(2\lambda)_\nu$ is defined by

$$(2\lambda)_\nu:=\Gamma(2\lambda+\nu)/\Gamma(2\lambda).$$

We shall identify a convergent series with its sum. In order to avoid the use of subscripts we shall not employ the variables $x_1,\dots$, as in Eq. (1.1), but x, y, z, and u instead, and write these explicitly only if necessary. Thereby we assign as many variables as are needed to the Gegenbauer functions in the standard order C_n^λ, D_n^λ, $\mathfrak{D}_n^\lambda$. Thus, for example,

$$S_{201}^\lambda(x,z,u):=a_\lambda^{-1}\sum_{n=0}^{\infty}(n+\lambda)\frac{C_n^\lambda(x)C_n^\lambda(z)\mathfrak{D}_n^\lambda(u)}{C_n^\lambda(1)}$$

and

$$
\begin{aligned}
S_{310}^\lambda(x,y,z,u):=a_\lambda^{-1}\sum_{n=0}^{\infty}(n+\lambda)C_n^\lambda(x)C_n^\lambda(y)\\
\times C_n^\lambda(z)D_n^\lambda(u)[C_n^\lambda(1)]^{-2}.
\end{aligned}
$$

The aim of this paper is sixfold.

(1) We wish to present in a systematic and compact way closed expressions—in an elegant and optimal, maximal-reduced form—for all the sums S_{klm}^λ that can be expressed in terms of a single hypergeometric function.

(2) We wish to correct the expression recently given in the literature[23] for S_{310}^λ.

(3) We wish to discuss several interesting relations that connect the different sums S_{klm}^λ, with emphasis on analytic continuation.

(4) We wish to show how the closed forms under (1) can be obtained in a simple way, and explicitly give a few proofs.

(5) We wish to work out a few particular cases in which the Gegenbauer functions reduce to more elementary functions like sines and cosines.

(6) We wish to give some new formulas involving the Gegenbauer functions.

We wish to give a systematic account of the sums S_{klm}^λ that can be evaluated in a relatively simple closed form. For the more general case of $S_{a_1,\dots,a_p}^\lambda$ with $a_i\in\mathbb{N}:=\{0,1,\dots\}$ and $\sigma:=\Sigma a_i$ there are $\binom{\sigma+p-1}{\sigma}$ different sums to be distinguished. Thus for $0\leqslant\sigma\leqslant s$ there are

$$\sum_{\sigma=0}^{s}\binom{\sigma+p-1}{\sigma}=\binom{s+p}{s}$$

different sums. In the present case $p=3$; taking $s=4$ we see that there are $\binom{4+3}{4}=35$ different sums S_{klm}^λ.

We shall give closed forms for 23 of these 35; in addition we shall evaluate S_{500}^λ for the special case in which all the variables are equal to zero; see Table I. We conjecture that the remaining sums S_{klm}^λ can be reduced at best to a double integral.

We have evaluated many more series of products of Gegenbauer functions of a related type; these will be published in the near future.[24]

 0022-2488/86/040938-15$02.50

TABLE I. The 24 sums of the type S^{λ}_{klm} whose closed forms are given in the text. Here klm stands for $S^{\lambda}_{klm}\,(x,y,z,u)$, $\sigma := k+l+m$, and $d := l+m$ ($k, l, m = 0, 1, \ldots$). The members of the triads are closely related; for example, 100 and 010 follow from 001 by analytical continuation, by application of Eqs. (1.9) – (1.11). The square brackets indicate a restricted case: 211 and 220 are evaluated only for $\lambda = 1$, whereas 500 and 030 are evaluated only for $x = y = z = u = 0$.

$\sigma =$	0	1	2	3	4	5
	000	001	101	201	301	
$d\leqslant 1$		100	200	300	400	[500]
		010	110	210	310	
			002	102	202	
$d\leqslant 2$			011	111	[211]	
			020	120	[220]	
$d = 3$				[030]		

The sum S^{λ}_{300} was evaluated by Dougall,[25] and S^{λ}_{210} by Rahman and Shah.[22] Further, we obtained[19] $S^{1/2}_{301}$, which was generalized to S^{λ}_{301} by Rahman and Shah.[23] Sums that follow from these by a simple reduction may be considered as known, too, except for the conditions to be imposed for the convergence of the series (see Sec. II). The results for the remaining sums (13 in total) of Table I are new, to the best of our knowledge. Moreover, their interrelationships, especially by means of analytic continuation (Sec. II), have presumably not been discussed in this form before. Not unexpectedly, the most complicated sums of Table I are those with $\sigma \equiv k + l + m = 4$.

In general, λ may be complex, with the exception of the nonpositive half integers. We assume that $-2\lambda \notin \mathbb{N}$ unless there is evidence to the contrary; the limits for $\lambda \to -\frac{1}{2}n$, $n \in \mathbb{N}$, are easily evaluated (cf. Sec. VII). In some derivations λ must be restricted to $0 < \operatorname{Re} \lambda < 1$; the final closed expressions are then to be analytically continued in λ.

In the closed expressions for S^{λ}_{klm} to be given in Secs. IV and V we shall restrict the variables of the C^{λ}_n's and D^{λ}_n's to $(-1, 1)$, and those of the $\mathfrak{D}^{\lambda}_n$'s to $(1, \infty)$. These expressions are then to be analytically continued in these variables. Although straightforward in principle, analytic continuation may be quite tricky: see, for example, the explicit expressions given in Sec. IV for S^{λ}_{301} and S^{λ}_{310}. The symmetry relations (1.17)–(1.19) are especially helpful here.

In Sec. VI we shall give closed forms for sums S^{λ}_{klm}, in which all the variables $x, y, \ldots$ are equal to zero. In Sec. VII we shall briefly consider the special cases $\lambda = 1$ and $\lambda \to -N$. Section VIII contains a few elementary derivations. In Sec. IX we shall give an elementary derivation of S^{λ}_{301} that is similar to the derivation of $S^{1/2}_{301}$ given in Ref. 19; Rahman and Shah[23] have derived S^{λ}_{301} in a slightly more complicated way.

In Secs. X-XII we shall derive S^{λ}_{400} and S^{λ}_{310} in closed form. Despite the complexity of these derivations we have been able to express S^{λ}_{400} in a simple and elegant way, in terms of the Legendre functions $\mathfrak{Q}_{\lambda-1}$ and $Q_{\lambda-1}$. Further, we express S^{λ}_{310} in terms of the Legendre functions $\mathfrak{Q}_{\lambda-1}$, $P_{\lambda-1}$, and $\mathfrak{P}_{\lambda-1}$; in this case we have to distinguish many different cases (see Table II, Sec. IV).

B. The Gegenbauer functions $C^{\lambda}_{\nu}(z)$, $D^{\lambda}_{\nu}(z)$, and $\mathfrak{D}^{\lambda}_{\nu}(z)$

The Gegenbauer function of the first kind, $C^{\lambda}_{\nu}(z)$, and those of the second kind, $D^{\lambda}_{\nu}(z)$ and $\mathfrak{D}^{\lambda}_{\nu}(z)$, are solutions of the differential equation

$$(1-z^2) f''(z) - (2\lambda+1) z f'(z) + \nu(\nu+2\lambda) f(z) = 0. \tag{1.2}$$

They are also called ultraspherical functions. When $\nu = n$, C^{λ}_n is Gegenbauer's polynomial, also called the ultraspherical polynomial and sometimes denoted by $P^{(\lambda)}_n$ (see, e.g., Szegö[8]). We shall employ Durand's definitions[13,14]:

$$\begin{aligned} C^{\lambda}_{\nu}(z) :&= [\Gamma(\nu+2\lambda)/\Gamma(2\lambda)\Gamma(\nu+1)] \\ &\quad \times {}_2F_1(-\nu, \nu+2\lambda; \lambda+\tfrac{1}{2}; \tfrac{1}{2}-\tfrac{1}{2}z) &(1.3)\\ &= \pi^{1/2} 2^{1/2-\lambda} [\Gamma(\nu+2\lambda)/\Gamma(\lambda)\Gamma(\nu+1)] \\ &\quad \times (z^2-1)^{(1/2)(1/2-\lambda)} \mathfrak{P}^{1/2-\lambda}_{\nu+\lambda-1/2}(z) &(1.4)\end{aligned}$$

and

$$\begin{aligned} e^{-i\pi\lambda}\mathfrak{D}^{\lambda}_{\nu}(z) :& \\ &= [\Gamma(\nu+2\lambda)/\Gamma(\lambda)\Gamma(\nu+\lambda+1)](2z)^{-\nu-2\lambda} \\ &\quad \times {}_2F_1(\tfrac{1}{2}\nu+\lambda, \tfrac{1}{2}\nu+\lambda+\tfrac{1}{2}; \nu+\lambda+1; z^{-2}) &(1.5)\\ &= \pi^{-1/2} e^{i\pi(\lambda-1/2)} 2^{1/2-\lambda} [\Gamma(\nu+2\lambda)/\Gamma(\lambda)\Gamma(\nu+1)] \\ &\quad \times (z^2-1)^{1/2(1/2-\lambda)} \mathfrak{Q}^{1/2-\lambda}_{\nu+\lambda-1/2}(z) &(1.6)\\ &= i\pi^{-1/2} e^{-i\pi\lambda} [\Gamma(\lambda)]^{-1} 2^{1/2-\lambda} \\ &\quad \times (z^2-1)^{1/2(1/2-\lambda)} \mathfrak{Q}^{\lambda-1/2}_{\nu+\lambda-1/2}(z). &(1.7)\end{aligned}$$

Here $\operatorname{Re} z > 0$ is needed because of that part of the branch cut of $(z^2-1)^{\mu}$ that is associated with $z^2 \leqslant 0$. In the following we shall mostly adopt Erdélyi's convention[2] that $(z^2-1)^{\mu}$ is *defined* as $(z-1)^{\mu}(z+1)^{\mu}$, so that the only branch cut is $(-\infty, 1]$, and $\operatorname{Re} z > 0$ may be omitted.

It should be noted that $\mathfrak{D}^{\lambda}_{\nu}(z)$ is often differently defined in the literature, viz., as[1,2]

$$e^{-i\pi\lambda} 2^{2\lambda-1} \Gamma^2(\lambda) \mathfrak{D}^{\lambda}_{\nu}(z). \tag{1.8}$$

The functions $C^{\lambda}_{\nu}(z)$ and $\mathfrak{D}^{\lambda}_{\nu}(z)$ are analytic in the complex z plane cut, respectively, from $-\infty$ to -1, and from $-\infty$ to $+1$; $C^{\lambda}_{\nu}(z)$ and $e^{-i\pi\lambda}\,\mathfrak{D}^{\lambda}_{\nu}(z)$ are real and increase (decrease) monotonically with z for z real, $z > 1$, and ν and λ real. The functions $C^{\lambda}_{\nu}(z)$ and $\mathfrak{D}^{\lambda}_{\nu}(z)$ as defined above satisfy the same recurrence relations.[13]

On $-1 < x < 1$, $C^{\lambda}_{\nu}(x)$ is defined as the restriction of $C^{\lambda}_{\nu}(z)$ to $(-1, 1)$; note that $C^{\lambda}_{\nu}(z)$ is analytic there. It follows that

$$C^{\lambda}_{\nu}(x) = \mathfrak{D}^{\lambda}_{\nu}(x+i0) + e^{-2i\pi\lambda}\mathfrak{D}^{\lambda}_{\nu}(x-i0), \quad -1 < x < 1, \tag{1.9}$$

where $\pm i0$ means that the limit for $\epsilon \downarrow 0$ of the function with the variable $x \pm i\epsilon$ has to be taken.

The so-called Gegenbauer function of the second kind "on the cut" $(-1, 1)$ is defined by

$$iD^{\lambda}_{\nu}(x) := \mathfrak{D}^{\lambda}_{\nu}(x+i0) - e^{-2i\pi\lambda}\mathfrak{D}^{\lambda}_{\nu}(x-i0), \quad -1 < x < 1; \tag{1.10}$$

hence

$$e^{-i\pi\lambda}\mathfrak{D}_\nu^\lambda(x\pm i0)=\tfrac{1}{2}e^{\mp i\pi\lambda}[C_\nu^\lambda(x)\pm iD_\nu^\lambda(x)],$$
$$-1<x<1. \tag{1.11}$$

Note that

$$[\mathfrak{D}_\nu^\lambda(z)]^*=e^{-2i\pi\lambda}\mathfrak{D}_\nu^\lambda(z^*),\quad \nu \text{ and } \lambda \text{ real}. \tag{1.12}$$

Then we have

$$D_\nu^\lambda(x)=2\pi^{-1/2}2^{1/2-\lambda}[\Gamma(\nu+2\lambda)/\Gamma(\lambda)\Gamma(\nu+1)]$$
$$\times(1-x^2)^{1/2(1/2-\lambda)}Q^{1/2-\lambda}_{\nu+\lambda-1/2}(x),$$
$$-1<x<1, \tag{1.13}$$

where Q_ν^μ is the usual Legendre function on the cut. Note the difference between

$$\mathfrak{D}_\nu^{1/2}(z)=i\pi^{-1}\mathfrak{Q}_\nu(z), \tag{1.14}$$

$$D_\nu^{1/2}(z)=2\pi^{-1}Q_\nu(z), \tag{1.15}$$

and

$$C_\nu^{1/2}(z)=P_\nu(z)=\mathfrak{P}_\nu(z). \tag{1.16}$$

We emphasize that $D_\nu^\lambda(x)$ can be analytically continued to the complex x plane cut along $(-\infty,-1]$ and $[+1,\infty)$. Indeed we shall need this function for *complex x*.

Further we have, for $n\in\mathbb{N}$

$$C_n^\lambda(-z)=(-1)^nC_n^\lambda(z), \tag{1.17}$$

$$D_n^\lambda(-z)=(-1)^{n+1}D_n^\lambda(z), \tag{1.18}$$

$$\mathfrak{D}_n^\lambda(-z)=(-1)^ne^{\pm2i\pi\lambda}\mathfrak{D}_n^\lambda(z),\quad \operatorname{Im}z\gtrless0. \tag{1.19}$$

These symmetry relations are extremely useful for checking closed formulas for the sums S^λ_{klm}, and for extending (whenever appropriate) the region of validity of these closed formulas. For example, a direct consequence is that

$$S^\lambda_{210}(-x,-z,u)=S^\lambda_{210}(x,z,u), \tag{1.20}$$

$$S^\lambda_{111}(-x,-z,u)=-S^\lambda_{111}(x,z,u), \tag{1.21}$$

and

$$S^\lambda_{120}(x,-z,-u)=S^\lambda_{120}(x,z,u). \tag{1.22}$$

II. CONVERGENCE, ANALYTIC CONTINUATION, AND CONTINUOUS EXTENSION

A. Convergence

The conditions for the convergence of the series S^λ_{klm} follow from the asymptotic expansions of the Gegenbauer functions, which in turn follow directly from the well-known asymptotic expansions[2] of the Legendre functions $\mathfrak{P}_\nu^\mu$, P_ν^μ, $\mathfrak{Q}_\nu^\mu$, and Q_ν^μ. We find

$$C_\nu^\lambda(z)\Gamma(\lambda)=2^{-\lambda}\nu^{\lambda-1}(z^2-1)^{-(1/2)\lambda}$$
$$\times[z+(z^2-1)^{1/2}]^{\nu+\lambda}[1+O(1/\nu)],$$
$$\nu\to\infty,\quad \operatorname{Re}z>0, \tag{2.1}$$

$$\mathfrak{D}_\nu^\lambda(z)\Gamma(\lambda)=e^{i\pi\lambda}\pi^{-1}2^{1-\lambda}\nu^{\lambda-1}(z^2-1)^{-(1/2)\lambda}$$
$$\times[z+(z^2-1)^{1/2}]^{-\nu-\lambda}[1+O(1/\nu)],$$
$$\nu\to\infty,\quad \operatorname{Re}z>0. \tag{2.2}$$

[Note that

$$(z+(z^2-1)^{1/2})(z-(z^2-1)^{1/2})=1.]$$

The condition $\operatorname{Re}z>0$ is needed only for the correct definition of $(z^2-1)^\mu$ $[\mu=\frac{1}{2}$ or $-\frac{1}{2}\lambda]$. It may be omitted if $(z^2-1)^\mu$ is replaced everywhere by $(z-1)^\mu\ (z+1)^\mu$. [See the remark following Eq. (1.7) and cf. the symmetry relations (1.17)–(1.19).] The above expansions are uniform in z for z outside any fixed closed contour enclosing the branch cut $-\infty<z\leqslant1$.

Further, let ϵ be fixed and $0<\epsilon\leqslant\theta\leqslant\pi-\epsilon$. Then[2]

$$\mathfrak{D}_\nu^\lambda(\cos\theta\pm i0)\Gamma(\lambda)$$
$$=\nu^{\lambda-1}(2\sin\theta)^{-\lambda}\exp[i\pi\lambda\mp i\{\nu\theta+\lambda(\theta+\tfrac{1}{2}\pi)\}]$$
$$\times[1+O(1/\nu)],\quad \nu\to\infty, \tag{2.3}$$

$$C_\nu^\lambda(\cos\theta)\Gamma(\lambda)$$
$$=\nu^{\lambda-1}(2\sin\theta)^{-\lambda}2\cos[\nu\theta+\lambda(\theta-\tfrac{1}{2}\pi)]$$
$$\times[1+O(1/\nu)],\quad \nu\to\infty, \tag{2.4}$$

$$D_\nu^\lambda(\cos\theta)\Gamma(\lambda)$$
$$=-\nu^{\lambda-1}(2\sin\theta)^{-\lambda}2\sin[\nu\theta+\lambda(\theta-\tfrac{1}{2}\pi)]$$
$$\times[1+O(1/\nu)],\quad \nu\to\infty, \tag{2.5}$$

and these expansions are uniform in θ on $[\epsilon,\pi-\epsilon]$.

Let us define the functions on the left-hand sides of Eqs. (2.1)–(2.5) at the apparent singularities $\lambda=0,-1,\dots$ by means of analytic continuation. Then the asymptotic expansions on the right-hand sides of Eqs. (2.1)–(2.5) are uniform in λ for $\operatorname{Re}\lambda\geqslant-L$, for arbitrarily large real L. Thus the series S^λ_{klm} in Eq. (1.1) is uniformly convergent whenever

$$\prod_{h=1}^{k}|x_h+(x_h^2-1)^{1/2}|\cdot\prod_{i=1}^{l}|z_i+(z_i^2-1)^{1/2}|$$
$$\times\prod_{j=1}^{m}|u_j-(u_j^2-1)^{1/2}|\leqslant\eta<1, \tag{2.6}$$

where η is fixed, $(z^2-1)^{1/2}$ is to be interpreted as $(z-1)^{1/2}(z+1)^{1/2}$ [see the remark following Eq. (1.7)], and $|z+(z^2-1)^{1/2}|$ may be defined as 1 when $-1<z<1$. We point out that the equation

$$|z+(z-1)^{1/2}(z+1)^{1/2}|=R,\quad R>1 \tag{2.7}$$

represents, in the complex z plane, an ellipse with foci at $+1$ and -1, with the major axis equal to $R+1/R$ and the minor axis equal to $R-1/R$. When the left-most side of (2.6) equals one, the associated conditions for convergence follow also from Eqs. (2.1)–(2.5); we shall not consider this case.

Let us now consider in particular S^λ_{kl0} and set for convenience

$$x_h=\cos\alpha_h,\quad z_i=\cos\beta_i;\quad \alpha_h,\beta_i\in[\epsilon,\pi-\epsilon],\quad \epsilon>0.$$

We find from Eqs. (2.3)–(2.5) that the series S^λ_{kl0} is

(i) divergent for

$$(4-k-l)\operatorname{Re}\lambda\geqslant1, \tag{2.8}$$

(ii) absolutely and hence uniformly convergent for

$$(4-k-l)\operatorname{Re}\lambda<0, \tag{2.9}$$

(iii) conditionally convergent for

$$(4-k-l)\operatorname{Re}\lambda<1, \tag{2.10}$$

provided that

$$(\alpha_1 \pm \alpha_2 \pm \cdots \pm \alpha_k \pm \beta_1 \pm \beta_2 \pm \cdots \pm \beta_l)/2\pi \notin \mathbb{Z} \tag{2.11}$$

for all the 2^{k+l-1} possible combinations of the signs $\pm$ taken independently from one another; outside fixed neighborhoods of these "points of conditional divergence" the convergence is uniform.

If any such combination equals an integer multiple of 2π, the series S^{λ}_{kl0} is in general divergent when

$$0 \leqslant (4-k-l)\operatorname{Re}\lambda < 1, \tag{2.12}$$

save in exceptional cases. For example, if one of the x_h's and one of the z_i's are both equal to zero, the sum in Eq. (1.1) equals zero since $C^{\lambda}_n(0)\, D^{\lambda}_n(0) = 0$ for $n = 0, 1,\ldots$, because of Eqs. (1.17) and (1.18). Further, keeping in mind that $\lambda \neq 0, -\frac{1}{2}, -1,\ldots$, we find that S^{λ}_{000} converges for $\operatorname{Re}\lambda < 0$, $S^{\lambda}_{100}(0)$ and $S^{\lambda}_{010}(0)$ for $\operatorname{Re}\lambda < \frac{1}{3}$, $S^{\lambda}_{200}(0,0)$ and $S^{\lambda}_{020}(0,0)$ for $\operatorname{Re}\lambda < 0$, $S^{\lambda}_{300}(0,0,0)$ and $S^{\lambda}_{030}(0,0,0)$ for $\operatorname{Re}\lambda < 1$, and $S^{\lambda}_{500}(0,0,0,0,0)$ for $\operatorname{Re}\lambda > -1$, whereas $S^{\lambda}_{400}(0,0,0,0)$ and $S^{\lambda}_{040}(0,0,0,0)$ are divergent for all λ.

As stated in Sec. I, we assume that the arguments of C^{λ}_n and D^{λ}_n occurring in the sums S^{λ}_{klm} are restricted to $(-1, 1)$, and the argument of $\mathfrak{D}^{\lambda}_n$ to $(1, \infty)$, unless there is evidence to the contrary. The closed forms to be given in Secs. IV and V are valid on these intervals, as is indicated there. Analytic continuation is always possible, provided the conditions for the convergence of the series S^{λ}_{klm} are satisfied. Thus an exception occurs for $m = 0$: The series $S^{\lambda}_{klm}(x, y,\ldots)$ are divergent for general complex $x, y,\ldots$ outside $[-1, 1]$, and hence analytic continuation of the associated closed forms makes no sense. When $m = 0$ we always assume that Eqs. (2.10) and (2.11) are satisfied, so that the series S^{λ}_{kl0} is conditionally and uniformly convergent.

According to Eqs (2.8)–(2.11), the convergence of S^{λ}_{400} is independent of λ provided $x, y, z, u \in (-1, 1)$ and the special "points of conditional divergence" are avoided; S^{λ}_{400} is conditionally (but not absolutely) convergent for *all* λ. Further, S^{λ}_{500} is conditionally convergent for $\lambda > -1$ and absolutely convergent for $\lambda > 0$. Note that the inequality sign changes from $<$ to $>$ when k changes from 3 to 5 (for $k = 4$ the sign is irrelevant).

It is interesting to note that taking the limit for $z\uparrow 1$ or $z\downarrow 1$ of a sum S^{λ}_{klm} leads to another sum for which $k + l + m$ is decreased by 1, as follows from Eq. (1.1). This latter sum requires for its convergence that λ be further restricted. Let us illustrate this with a simple example. We have

$$\begin{aligned} S^{\lambda}_{100}(x) &= a_{\lambda}^{-1}\sum_{n=0}^{\infty}(n+\lambda)C^{\lambda}_n(1)C^{\lambda}_n(x) \\ &= 2\cos\pi\lambda(1-x)^{-2\lambda}, \quad -1<x<1, \quad \operatorname{Re}\lambda<\tfrac{1}{3}. \end{aligned} \tag{2.13}$$

The limit for $x\uparrow 1$ yields 0 provided that $\operatorname{Re}\lambda < 0$; this is precisely the condition under which $S^{\lambda}_{100}(1) = S^{\lambda}_{000}$ converges. Indeed,

$$\begin{aligned} S^{\lambda}_{000} &= a_{\lambda}^{-1}\sum_{n=0}^{\infty}(n+\lambda)[C^{\lambda}_n(1)]^2 \\ &= a_{\lambda}^{-1}\lambda\, {}_3F_2(2\lambda,2\lambda,\lambda+1;\lambda,1;1), \quad \operatorname{Re}\lambda<0. \end{aligned} \tag{2.14}$$

This ${}_3F_2$ is well defined if $\operatorname{Re}\lambda < 0$ and it vanishes, as follows from

$$\begin{aligned} &{}_3F_2(2\lambda,2\lambda,\lambda+1;\lambda,1;z) \\ &= \lambda^{-1}z^{1-\lambda}\frac{\partial}{\partial z}z^{\lambda}\, {}_2F_1(2\lambda,2\lambda;1;z) \\ &= {}_2F_1(2\lambda,2\lambda;1;z) + 4\lambda z\, {}_2F_1(2\lambda+1,2\lambda+1;2;z). \end{aligned} \tag{2.15}$$

Moreover, by taking $x\downarrow -1$ in Eq. (2.13) we get

$$\lim_{x\downarrow -1} S^{\lambda}_{100}(x) = 2^{1-2\lambda}\cos\pi\lambda, \quad \operatorname{Re}\lambda<\tfrac{1}{3}. \tag{2.16}$$

The right-hand side agrees with $S^{\lambda}_{100}(-1)$ except for the condition on λ:

$$\begin{aligned} S^{\lambda}_{100}(-1) &= a_{\lambda}^{-1}\lambda\, {}_3F_2(2\lambda,2\lambda,\lambda+1;\lambda,1;-1) \\ &= 2^{1-2\lambda}\cos\pi\lambda, \quad \operatorname{Re}\lambda<\tfrac{1}{4}. \end{aligned} \tag{2.17}$$

[This follows by using Eq. (2.15) again.] The series $S^{\lambda}_{100}(-1)$ converges (conditionally) for $\operatorname{Re}\lambda < \frac{1}{4}$ and diverges for $\frac{1}{4} \leqslant \operatorname{Re}\lambda$.

B. Analytic continuation and continuous extension

The sums S^{λ}_{klm} are analytic in the parameter λ and in the variables x_h, z_i, and u_j in certain regions in the associated complex planes that are interrelated. These regions can be determined, for λ and each one of the variables separately, with the help of the uniform convergence of the series S^{λ}_{klm} (see below) and the following theorems (see, e.g., Titchmarsh,[26] pp. 3–8 and 95).

Theorem 1: Absolute convergence of a series implies its uniform convergence.

Theorem 2: If $a_n \downarrow 0$ for $n \to \infty$ and x is real, then the series $\Sigma_n\, a_n \exp(inx)$ is uniformly convergent for x in any closed interval not containing an integer multiple of 2π.

Theorem 3: The sum of a uniformly convergent series of continuous functions is a continuous function.

Theorem 4: The sum of a uniformly convergent series of analytic functions is an analytic function.

It follows that the series S^{λ}_{klm} $(k, l, m = 0, 1,\ldots)$ and their sums are closely interrelated. Many of these sums can be easily and conveniently evaluated in closed form by means of *analytic continuation* and *continuous extension*. In order to show how this is done in general we shall work out a particular case in detail.

We start with the series $S^{\lambda}_{001}(u)$, whose sum will be derived in Sec. VIII,

$$S^{\lambda}_{001}(u) = e^{i\pi\lambda}(u-1)^{-2\lambda}, \quad 1<u, \quad \text{all } \lambda. \tag{2.18}$$

The series $S^{\lambda}_{001}(u)$ is uniformly convergent and hence, because of Theorem 4, analytic in u outside the branch cut $(-\infty, 1]$, i.e., on the open set $\mathbb{C}\backslash(-\infty, 1]$. The right-hand side of Eq. (2.18) can be analytically extended directly to this open set.

Now we restrict λ to $\operatorname{Re}\lambda < \frac{1}{2}$ and keep λ fixed; then $S^{\lambda}_{100}(x)$ and $S^{\lambda}_{010}(x)$ are uniformly convergent series for $x\in[\delta - 1, 1-\delta]$, for any fixed $\delta\in(0, 1)$. (We take δ arbitrarily small.) Further we introduce, for this occasion only,

$$S^{\lambda}_{\pm}(u) := S^{\lambda}_{001}(u) := a_{\lambda}^{-1}\sum_{n=0}^{\infty}(n+\lambda)C^{\lambda}_{n}(1)\mathfrak{D}^{\lambda}_{n}(u),$$

for $u\in\mathbb{C}\setminus(-\infty, 1]$,

$$S^{\lambda}_{\pm}(x) := a_{\lambda}^{-1}\sum_{n=0}^{\infty}(n+\lambda)C^{\lambda}_{n}(1)\mathfrak{D}^{\lambda}_{n}(x\pm i0),$$

$$\text{for } x\in(-1, 1), \tag{2.19}$$

where, as before,

$$\mathfrak{D}^{\lambda}_{n}(x\pm i0) := \lim_{\epsilon\downarrow 0}\mathfrak{D}^{\lambda}_{n}(x\pm i\epsilon).$$

[One should be careful with this notation: Clearly $x\pm i0$ may not be replaced by x; also, $\mathfrak{D}^{\lambda}_{n}(x)$ is not defined for $x\leqslant 1$.] Thus both $S^{\lambda}_{+}(u)$ and $S^{\lambda}_{-}(u)$ are defined for u outside the point $+1$ and the interval $(-\infty, -1]$, while the interval $(-\infty, +1]$ is the line of discontinuity of both functions. Now $S^{\lambda}_{+}(u)$ represents a uniformly convergent series of functions that are continuous in particular on the set

$$\{u|\delta - 1\leqslant\operatorname{Re}u\leqslant 1-\delta \text{ and } \operatorname{Im}u\geqslant 0\},\quad \delta>0.$$

Thus, according to Theorem 3, $S^{\lambda}_{+}(u)$ is *upper continuous* on $(-1, 1)$, i.e., continuous from $(-1, 1)$ into the upper part of the complex plane, and similarly $S^{\lambda}_{-}(u)$ is *lower continuous* on $(-1, 1)$ [$\operatorname{Im}u\leqslant 0$]. This may be compared with the right and left continuity of functions defined on the real line.

Because of this upper and lower continuity of S^{λ}_{+} and S^{λ}_{-}, respectively, on $(-1, 1)$, we can derive the following closed forms from Eq. (2.18):

$$S^{\lambda}_{\pm}(x) = e^{i\pi\lambda}\lim_{\epsilon\downarrow 0}(x\pm i\epsilon - 1)^{-2\lambda} = e^{i\pi\lambda}e^{\mp 2i\pi\lambda}(1-x)^{-2\lambda},$$

$$-1<x<1,\quad \operatorname{Re}\lambda<\tfrac{1}{2}. \tag{2.20}$$

By using the relations (1.9)–(1.11) between C^{λ}_{n}, D^{λ}_{n}, and $\mathfrak{D}^{\lambda}_{n}$ we now get directly

$$S^{\lambda}_{100}(x) = S^{\lambda}_{+}(x) + e^{-2i\pi\lambda}S^{\lambda}_{-}(x) = 2\cos\pi\lambda(1-x)^{-2\lambda}, \tag{2.21}$$

$$S^{\lambda}_{010}(x) = -i[S^{\lambda}_{+}(x) - e^{-2i\pi\lambda}S^{\lambda}_{-}(x)] = -2\sin\pi\lambda(1-x)^{-2\lambda},$$

$$-1<x<1,\ \operatorname{Re}\lambda<\tfrac{1}{2}. \tag{2.22}$$

The same procedure works in principle for all the series S^{λ}_{klm}. However, it becomes more and more complicated as the number of variables increases. Let us consider another example. The series

$$S^{\lambda}_{011}(x,z) := a_{\lambda}^{-1}\sum_{n=0}^{\infty}(n+\lambda)D^{\lambda}_{n}(x)\mathfrak{D}^{\lambda}_{n}(z) \tag{2.23}$$

is convergent for x inside the ellipse that has the foci $+1$ and -1, and that goes through $z\in\mathbb{C}\setminus(-\infty, 1]$. As a function of x it is analytic inside this ellipse with the exception of $(-\rho, -1]$ and $[1, \rho)$ [$\rho := \frac{1}{2}(R+1/R)>1$; see Eq. (2.7)], where it has branch cuts originating from the branch cuts $(-\infty, -1]$ and $[1, \infty)$ of $D^{\lambda}_{n}(x)$.

When $\sigma\equiv k+l+m$ equals 3 or 4 the analytic continuation is much more difficult because the closed forms contain apparent singularities (branch cuts) that we have to remove in order to get the "right" analytical structure. For example, it is anything but a trivial matter to continue the closed form to be given in Sec. IV for

$$S^{\lambda}_{301}(x,y,z,u),\quad x,y,z\in(-1,1),\quad 1<u,$$

in the variable u analytically to the open set $\mathbb{C}\setminus(-\infty, 1]$; nevertheless we know from the above discussion that it is actually analytic there. We have carried out this procedure for $S^{\lambda}_{301}(x, y, z, u)$ only for the special case in which $z=0$, and derived in this way $S_{400}(x, y, 0, u)$ and $S^{\lambda}_{310}(x, y, 0, u)$ for $x, y, u\in(-1, 1)$. The closed forms for these sums in the general case [$z\in(-1, 1)$] will be derived in Secs. X and XI with the help of Eq. (3.1).

In general, care is needed with analytic continuation. Let us consider, for instance, the two expressions to be given in Sec. IV for S^{λ}_{301}. In the first, we require $u>\sqrt{2}$ in order to guarantee that $W>0$ (see Sec. III C). The second expression possesses a branch cut for $1<u<\infty$, which arises from $T^{-(1/2)\lambda}$ and $T^{-1/2}$, since $T<0$ for $1<u<\infty$. However, this branch cut is easily removed with the help of the analytic properties of $\mathfrak{Q}_{\lambda-1}$. Another example is found in the expresion for S^{λ}_{310} (Sec. IV). The Legendre function $\mathfrak{P}_{\lambda-1}(z)$ [or $P_{\lambda-1}(z)$] has a convenient representation in terms of ${}_2F_1(\cdots; 1-z^2)$, which is valid only for $\operatorname{Re}z>0$, and especially not for $-1<z<0$ (see Sec. III B).

III. AUXILIARY FORMULAS

A. Introduction

In Secs. IV–VII we shall give closed formulas for a large number of sums. The derivations of these formulas are sometimes quite lengthy. We have chosen not to give all these derivations in detail, but to give instead a list of auxiliary formulas that are most suitable for working out these derivations. Especially since the final results to be obtained are given, the actual derivations should not be too difficult.

In Secs. VIII and IX we shall work out a few sums to give the general idea, and in Secs. X and XI we shall briefly derive the more complicated sums S^{λ}_{400} and S^{λ}_{310}, respectively.

Some formulas of the following list are presumably new, but most of them follow easily from, e.g., Ref. 2.

B. Formulas involving Gegenbauer or ultraspherical functions

First we give a few elementary relations that will play a role:

$$\operatorname{arcsinh}(iz) = i\arcsin z = \ln[iz + (1-z^2)^{1/2}],$$

$$\operatorname{arctanh}(iz) = i\arctan z = \tfrac{1}{2}\ln[(1+iz)/(1-iz)],$$

$$\Gamma(2\lambda) = \pi^{-1/2}2^{2\lambda-1}\Gamma(\lambda)\Gamma(\lambda+\tfrac{1}{2}),$$

$$\Gamma(\lambda)\Gamma(1-\lambda) = \pi/\sin\pi\lambda,$$

$$\Gamma(\tfrac12+\lambda)\Gamma(\tfrac12-\lambda)=\pi/\cos\pi\lambda,$$
$$a_\lambda := 2^{-2\lambda}\Gamma(2\lambda)/\Gamma^2(\lambda),$$
$$a_\lambda^{-1}=2\pi^{1/2}\Gamma(\lambda)/\Gamma(\lambda+\tfrac12)$$
$$=2\pi^{-1/2}\cos\pi\lambda\,\Gamma(\lambda)\Gamma(\tfrac12-\lambda).$$

For definitions of the Gegenbauer functions, see Eqs. (1.3)–(1.10), for symmetry relations, Eqs. (1.14)–(1.16), and for the special case $\lambda=\frac12$, Eqs. (1.11)–(1.13).

Next, let $\operatorname{Re}\lambda>0$, $\operatorname{Re}\nu>-1$, $\nu+2\lambda\neq 0$,
$$c_\nu^\lambda := 2a_\lambda(2\lambda)_\nu/\Gamma(\nu+1)=2a_\lambda C_\nu^\lambda(1),$$
$$\bar x := (1-x^2)^{1/2},\quad \bar y := (1-y^2)^{1/2},\quad x,y\in(-1,1),$$
and (note the difference)
$$\bar z := (z^2-1)^{1/2},\quad \bar u := (u^2-1)^{1/2},\quad z,u\in(1,\infty).$$
Then
$$c_\nu^\lambda\int_{-1}^1 C_\nu^\lambda(xy+\bar x\bar y t)(1-t^2)^{\lambda-1}\,dt=\begin{cases}C_\nu^\lambda(x)C_\nu^\lambda(y), & \text{if } x+y>0,\\ C_\nu^\lambda(-x)C_\nu^\lambda(-y), & \text{if } x+y<0,\end{cases}\tag{3.1}$$
$$c_\nu^\lambda\int_{-1}^1 C_\nu^\lambda(zu+\bar z\bar u t)(1-t^2)^{\lambda-1}\,dt=C_\nu^\lambda(z)C_\nu^\lambda(u),\tag{3.2}$$
$$c_\nu^\lambda\int_{-1}^1 D_\nu^\lambda(xy+\bar x\bar y t)(1-t^2)^{\lambda-1}\,dt=\begin{cases}C_\nu^\lambda(x)D_\nu^\lambda(y), & \text{if } x>\pm y,\\ C_\nu^\lambda(y)D_\nu^\lambda(x), & \text{if } y>\pm x,\\ C_\nu^\lambda(-y)D_\nu^\lambda(-x), & \text{if } y<\pm x,\\ C_\nu^\lambda(-x)D_\nu^\lambda(-y), & \text{if } x<\pm y.\end{cases}\tag{3.3}$$

(Each condition should hold for both signs $+$ and $-$. For $x=\pm y$ take the appropriate limits; this procedure is consistent.) For complex ν these relations seem to be new. For $\nu=n$ they can be simplified with the help of the symmetry relations (1.17)–(1.19).

Further[13]
$$c_\nu^\lambda\int_{-1}^1 \mathfrak{D}_\nu^\lambda(zu+\bar z\bar u t)(1-t^2)^{\lambda-1}\,dt=C_\nu^\lambda(z)\mathfrak{D}_\nu^\lambda(u),\quad \text{if } 1<z<u,\tag{3.4}$$
$$e^{i\pi\lambda}c_\nu^\lambda\int_1^\infty \mathfrak{D}_\nu^\lambda(zu+\bar z\bar u t)(t^2-1)^{\lambda-1}\,dt=\mathfrak{D}_\nu^\lambda(z)\mathfrak{D}_\nu^\lambda(u),\tag{3.5}$$
$$4e^{-i\pi\lambda}c_\nu^\lambda\int_1^\infty \mathfrak{D}_\nu^\lambda[x^2+(1-x^2)t](t^2-1)^{\lambda-1}\,dt=[C_\nu^\lambda(x)]^2+[D_\nu^\lambda(x)]^2=4e^{-2i\pi\lambda}\lim_{\epsilon\downarrow 0}\mathfrak{D}_\nu^\lambda(x+i\epsilon)\mathfrak{D}_\nu^\lambda(x-i\epsilon).\tag{3.6}$$

The relations (3.1)–(3.6) are useful for deriving S_{klm}^λ with increasing k, l, or m. Note that (3.1) for $\nu=0$ directly yields
$$1=2a_\lambda\int_{-1}^1(1-t^2)^{\lambda-1}\,dt.$$

Next
$$\mathfrak{D}_\nu^\lambda(\cosh\alpha)=e^{i\pi\lambda}2^{-\lambda}\pi^{-1}\sin\pi\lambda\times\int_\alpha^\infty e^{-(\nu+\lambda)t}(\cosh t-\cosh\alpha)^{-\lambda}\,dt,$$
$$\alpha>0,\quad \operatorname{Re}\lambda<1,\quad \operatorname{Re}(\nu+2\lambda)>0,\tag{3.7}$$
$$=e^{i\pi\lambda}2^{-\lambda}\Gamma(\nu+2\lambda)[\Gamma^2(\lambda)\Gamma(\nu+1)]^{-1}\times(\sinh\alpha)^{1-2\lambda}\int_\alpha^\infty e^{-(\nu+\lambda)t}\times(\cosh t-\cosh\alpha)^{\lambda-1}\,dt,$$
$$\alpha>0,\quad \operatorname{Re}\lambda>0,\quad \operatorname{Re}\nu>-1,\quad \nu+2\lambda\neq 0.\tag{3.8}$$

These two integral representations follow from the analogous ones for $\mathfrak{Q}_\nu^\mu(z)$; note that
$$\Gamma(\nu-\mu+1)e^{-i\pi\mu}\mathfrak{Q}_\nu^\mu(z)=\Gamma(\nu+\mu+1)e^{i\pi\mu}\mathfrak{Q}_\nu^{-\mu}(z).$$
Other useful integral representations valid for $\nu\in\mathbb{C}$ are
$$e^{-i\pi\lambda}\mathfrak{D}_\nu^\lambda(z)\pi^{1/2}2^{\nu+2\lambda}\Gamma(\lambda)\Gamma(\nu+\lambda+\tfrac12)/\Gamma(\nu+2\lambda)=(z-1)^{1/2-\lambda}(z+1)^{1/2-\lambda}\times\int_{-1}^1(z\pm t)^{-\nu-1}(1-t^2)^{\nu+\lambda-1/2}\,dt$$
$$=\int_{-1}^1(z\pm t)^{-\nu-2\lambda}(1-t^2)^{\nu+\lambda-1/2}\,dt,$$
$$\operatorname{Re}(\nu+\lambda)>-\tfrac12.\tag{3.9}$$
We have found for $\nu\to n\in\mathbb{N}$
$$\mathfrak{D}_n^\lambda(z)d_\lambda(z)=\int_{-1}^1 C_n^\lambda(t)(1-t^2)^{\lambda-1/2}(z-t)^{-1}\,dt,$$
$$\operatorname{Re}\lambda>-\tfrac12,\tag{3.10}$$
$$d_\lambda(z) := 2\pi e^{-i\pi\lambda}(z-1)^{\lambda-1/2}(z+1)^{\lambda-1/2}.$$
This relation does not hold for general n; in particular for $n=-1$ it turns out that $C_{-1}^\lambda(t)$ in the integrand should be replaced by $(z-t)/(1-t)$ and thus
$$\mathfrak{D}_{-1}^\lambda(z)d_\lambda(z)=\int_{-1}^1(1-t^2)^{\lambda-3/2}\,dt=B(\tfrac12,\lambda-\tfrac12);$$
hence
$$d_\lambda(z)=2\pi a_\lambda(\lambda-\tfrac12)^{-1}/\mathfrak{D}_{-1}^\lambda(z).$$
By applying Eq. (1.10) to Eq. (3.10) we get for $D_n^\lambda(x)$ the principal-value integral
$$D_n^\lambda(x)=\pi^{-1}(1-x^2)^{1/2-\lambda}\int_{-1}^1 C_n^\lambda(t)\times(1-t^2)^{\lambda-1/2}(x-t)^{-1}\,dt,$$
$$-1<x<1,\quad \operatorname{Re}\lambda>-\tfrac12,\quad n=0,1,\ldots.\tag{3.11}$$

Equations (3.10) and (3.11) are very convenient for converting C_n^λ into $\mathfrak{D}_n^\lambda$ and D_n^λ, respectively, in series of Gegenbauer functions. Near the branch cut $-\infty<z\leqslant-1$ of $C_\nu^\lambda(z)$, the branch cut $-\infty<z\leqslant 1$ common to $\mathfrak{D}_\nu^\lambda(z)$, $\mathfrak{P}_\nu^\mu(z)$, and $\mathfrak{Q}_\nu^\mu(z)$, and the branch cuts $-\infty<z\leqslant-1$ and $1\leqslant z<\infty$ of $D_\nu^\lambda(z)$, $P_\nu^\mu(z)$, and $Q_\nu^\mu(z)$, the following relations are useful:

$$D_\nu^\lambda(z\pm i0) = \pm iC_\nu^\lambda(z) \mp 2ie^{\pm i\pi\lambda}e^{-i\pi\lambda}\mathfrak{D}_\nu^\lambda(z), \quad z>1,$$

$$Q_\nu^\mu(z\pm i0) = \pm \tfrac{1}{2}i\pi e^{\pm(1/2)i\pi\mu}\mathfrak{P}_\nu^\mu(z) + e^{\mp(1/2)i\pi\mu}e^{-i\pi\mu}\mathfrak{Q}_\nu^\mu(z), \quad z>1,$$

$$\mathfrak{Q}_\nu(x\pm i0) = Q_\nu(x) \mp \tfrac{1}{2}i\pi P_\nu(x), \quad -1<x<1,$$

$$\mathfrak{Q}_\nu(-z\mp i0) = -e^{\pm i\pi\nu}\mathfrak{Q}_\nu(z), \quad z>1,$$

$$\mathfrak{P}_\nu(-z\mp i0) = e^{\mp i\pi\nu}\mathfrak{P}_\nu(z) - 2\pi^{-1}\sin\pi\nu\mathfrak{Q}_\nu(z), \quad z>1.$$

Further we mention the quadratic transformation

$${}_2F_1(a,1-a;c;z) = (1-z)^{c-1}\,{}_2F_1(\tfrac{1}{2}(c-a), \tfrac{1}{2}(c+a-1);c;4z-4z^2), \quad \operatorname{Re} z<\tfrac{1}{2}$$

valid for $z=\frac{1}{2}$ and for $\operatorname{Re} z<\frac{1}{2}$. This is a *corrected version* of Ref. 4, p. 50. Related to this is

$$P_{\lambda-1}(z) = \mathfrak{P}_{\lambda-1}(z) = {}_2F_1(\tfrac{1}{2}\lambda,\tfrac{1}{2}-\tfrac{1}{2}\lambda;1;1-z^2), \quad \operatorname{Re} z>0.$$

This relation is not generally valid for $\operatorname{Re} z<0$. In particular, for $-1<x<1$, we have

$$P_{\lambda-1}(-x) = -\cos\pi\lambda P_{\lambda-1}(x) + 2\pi^{-1}\sin\pi\lambda Q_{\lambda-1}(x).$$

Also

$$Q_{\lambda-1}(-x) = \cos\pi\lambda Q_{\lambda-1}(x) + \tfrac{1}{2}\pi\sin\pi\lambda P_{\lambda-1}(x), \quad -1<x<1,$$

$$\pi P_{\lambda-1}(-x) = \sin\pi\lambda\,[Q_{\lambda-1}(x) + Q_{-\lambda}(x)], \quad -1<x<1,$$

$$\pi P_{\lambda-1}(x) = \tan\pi\lambda\,[Q_{\lambda-1}(x) - Q_{-\lambda}(x)], \quad -1<x<1,$$

and

$$\mathfrak{Q}_{-\lambda}(z) = \mathfrak{Q}_{\lambda-1}(z) - \pi\cot\pi\lambda\mathfrak{P}_{\lambda-1}(z), \quad z\notin(-\infty,1].$$

C. Some special functions and relations involving x, y, z, and u

Unless there is evidence to the contrary we assume that $-2\lambda\notin\mathbb{N}$ and $\operatorname{Re}\nu>-1$. Further, $\bar{x}:=(1-x^2)^{1/2}$, and $\bar{y},\bar{z}$, and $\bar{u}$ are similarly defined when $x,y,z,u\in(-1,1)$. The quantity T is always defined by

$$T:=4(1-x^2)(1-y^2)(1-z^2)(1-u^2).$$

If $\sigma:=k+l+m=4$, we use

$$W:=x^2+y^2+z^2+u^2-2-2xyzu.$$

For $y\to1$ this reduces to $W=x^2+z^2+u^2-1-2xzu$, which we use when $\sigma=3$. It is sometimes convenient to use the notation with sines and cosines; we shall occasionally set

$$x=\cos\alpha, \quad y=\cos\beta, \quad z=\cos\gamma, \quad u=\cos\delta, \quad \alpha,\beta,\gamma,\delta\in(0,\pi).$$

There exist some interesting nontrivial (although elementary) relations that are useful especially when $\sigma=4$; direct proofs follow from

$$\cos\alpha-\cos\beta = 2\sin\tfrac{1}{2}(\beta-\alpha)\sin\tfrac{1}{2}(\beta+\alpha), \quad \cos\alpha+\cos\beta = 2\cos\tfrac{1}{2}(\beta-\alpha)\cos\tfrac{1}{2}(\beta+\alpha).$$

In particular for S_{400}^λ, S_{310}^λ, and S_{202}^λ the following equalities are useful:

$$[\cos(\alpha+\beta)-\cos(\gamma+\delta)]\,[\cos(\alpha+\beta)-\cos(\gamma-\delta)] = [\cos(\alpha+\beta+\gamma)-\cos\delta]\,[\cos(\alpha+\beta-\gamma)-\cos\delta],$$

$$[\cos(\alpha+\beta)-\cos(\gamma-\delta)]\,[\cos(\alpha-\beta)-\cos(\gamma+\delta)] = W+T^{1/2},$$

$$[\cos(\alpha+\beta)-\cos(\gamma+\delta)]\,[\cos(\alpha-\beta)-\cos(\gamma-\delta)] = W-T^{1/2}.$$

Finally, for S_{202}^λ the following relations also play a role:

$$\frac{\cos(\alpha-\beta)-\cos(\gamma-\delta)}{\cos(\alpha+\beta)-\cos(\gamma-\delta)} = \frac{V-W+T^{1/2}}{V-W-T^{1/2}}, \quad \frac{\cos(\alpha-\beta)-\cos(\gamma+\delta)}{\cos(\alpha+\beta)-\cos(\gamma+\delta)} = \frac{V_--W-T^{1/2}}{V_--W+T^{1/2}}.$$

Here

$$T^{1/2} = 2\sin\alpha\sin\beta\sin\gamma\sin\delta, \quad W:=x^2+y^2+z^2+u^2-2-2xyzu,$$

$$V:=(zu+\bar{z}\bar{u})(zu+\bar{z}\bar{u}-2xy)+x^2+y^2-1, \quad V_-:=(zu-\bar{z}\bar{u})(zu-\bar{z}\bar{u}-2xy)+x^2+y^2-1,$$

from which we obtain

$$V=W+2\bar{z}\bar{u}(\bar{z}\bar{u}+zu-xy), \quad V_-=W+2\bar{z}\bar{u}(\bar{z}\bar{u}-zu+xy), \quad VV_-=W^2-T,$$

$$V+V_-=2W+4\bar{z}^2\bar{u}^2, \quad V-V_-=4\bar{z}\bar{u}(zu-xy).$$

When dealing with functions of four variables we find, not unexpectedly, that even elementary derivations can be quite complicated. For example, let us consider the first expression to be given for S_{301}^λ. Here we must avoid the branch cut associated with $W<0$, which arises from $W^{-\lambda}$; this is the reason for which $u>\sqrt{2}$ is required there. This condition is necessary and sufficient. We shall prove that

$$u^2>2 \Leftrightarrow W>0, \quad \forall x,y,z\in(-1,1), \quad u^2\geqslant2\Leftrightarrow W\geqslant0, \quad \forall x,y,z\in(-1,1);$$

the equality $W=0$ occurs only for $u^2=2$ and $x=y=z=0$. Since $W=u^2-2$ for $x=y=z=0$ it is obvious that the above inequalities are optimal. The proof of the above statement follows from

$$W(u)=u^2-2xyzu+x^2+y^2+z^2-2=t_1+t_2+t_3,\ t_1:=(u\pm\sqrt{2})(u\mp\sqrt{2}-2xyz),$$

$$t_2:=(x\pm yz\sqrt{2})^2\geqslant 0,\quad t_3:=\tfrac{1}{2}-\tfrac{1}{2}(1-2y^2)(1-2z^2)\geqslant 0.$$

It follows that $t_1>0$ for $u>\sqrt{2}$ (take the lower signs), and for $u<-\sqrt{2}$ (take the upper signs). Further, $t_2=t_3=0$ occurs only for $x=y=z=0$, which completes the proof. It is interesting to note that the zeros $u_\pm$ of $W(u)$ are real:

$$u_\pm=xyz\pm[(1-y^2)(1-z^2)+(1-x^2)(1-y^2z^2)]^{1/2},$$

and that $|u_\pm|\leqslant\sqrt{2}$ is equivalent to

$$x^2+y^2+z^2\pm 2\sqrt{2}xyz\equiv t_2+t_3\geqslant 0.$$

IV. CLOSED FORMULAS FOR S^λ_{klm} WITH $l+m<1$

We have the following:

$$S^\lambda_{001}(u)=e^{i\pi\lambda}(u-1)^{-2\lambda},\quad 1<u,\quad \text{all }\lambda,\quad S^\lambda_{100}(x)=2\cos\pi\lambda(1-x)^{-2\lambda},\quad -1<x<1,\quad \operatorname{Re}\lambda<\tfrac{1}{2},$$

$$S^\lambda_{010}(x)=-2\sin\pi\lambda(1-x)^{-2\lambda},\quad -1<x<1,\quad \operatorname{Re}\lambda<\tfrac{1}{2};$$

$$S^\lambda_{101}(x,u)=e^{i\pi\lambda}(u-x)^{-2\lambda},\quad -1<x<1,\quad 1<u,\quad \text{all }\lambda,$$

$$S^\lambda_{200}(x,z)=2\cos\pi\lambda\,|z-x|^{-2\lambda},\quad x,z\in(-1,1),\quad \operatorname{Re}\lambda<\tfrac{1}{2},$$

$$S^\lambda_{110}(x,z)=\pm 2\sin\pi\lambda\,|z-x|^{-2\lambda},\quad \text{if } z\gtrless x,\quad x,z\in(-1,1),\quad \operatorname{Re}\lambda<\tfrac{1}{2};$$

$$S^\lambda_{201}(x,z,u)=e^{i\pi\lambda}W^{-\lambda},\quad x,z\in(-1,1),\quad 1<u,\quad \text{all }\lambda,$$

$$W:=x^2+z^2+u^2-1-2xzu=(xz-u)^2-(1-x^2)(1-z^2),$$

$$S^\lambda_{300}(x,z,u)=\begin{cases}2(-W)^{-\lambda}, & \text{if } W<0,\\ 2\cos\pi\lambda W^{-\lambda}, & \text{if } W>0,\end{cases}\quad x,z,u\in(-1,1),\quad W:=x^2+z^2+u^2-1-2xzu,$$

$$S^\lambda_{210}(x,z,u)=\begin{cases}0,\ \text{if } W<0 & [xz-\overline{xz}<u<xz+\overline{xz}],\\ 2\sin\pi\lambda W^{-\lambda}, & \text{if } xz+\overline{xz}<u\quad [W>0],\\ -2\sin\pi\lambda W^{-\lambda}, & \text{if } u<xz-\overline{xz}\quad [W>0],\end{cases}\quad x,z,u\in(-1,1),\quad W:=x^2+z^2+u^2-1-2xzu;$$

$$S^\lambda_{301}(x,y,z,u)=e^{i\pi\lambda}W^{-\lambda}\,{}_2F_1(\tfrac{1}{2}\lambda,\tfrac{1}{2}\lambda+\tfrac{1}{2};\lambda+\tfrac{1}{2};TW^{-2}),\quad 2^{1/2}<u,$$

$$=e^{i\pi\lambda}a_\lambda 2^{\lambda+1}T^{-(1/2)\lambda}\mathfrak{Q}_{\lambda-1}(WT^{-1/2}),\quad \operatorname{Re}u>1,\quad \operatorname{Im}u>0,$$

$$T:=4(1-x^2)(1-y^2)(1-z^2)(1-u^2),\quad W:=x^2+y^2+z^2+u^2-2-2xyzu,$$

$$x,y,z\in(-1,1),\quad \text{all }\lambda\quad (2^{1/2}<u\Rightarrow W>0).$$

For convenience we rewrite this as

$$S^\lambda_{301}(x,y,z,u)=e^{i\pi\lambda}\mathfrak{Q}_{\lambda-1}(\omega)U_\lambda,\quad \omega:=WT^{-1/2},\qquad U_\lambda:=a_\lambda 2^{\lambda+1}T^{-(1/2)\lambda}=T^{-(1/2)\lambda}2^{1-\lambda}\Gamma(2\lambda)/\Gamma^2(\lambda).$$

Then

$$S^\lambda_{400}(x,y,z,u)=\begin{cases}2\cos\pi\lambda\mathfrak{Q}_{\lambda-1}(\omega)U_\lambda, & \text{if } 1<\omega,\\ 2Q_{\lambda-1}(-\omega)U_\lambda, & \text{if } -1<\omega<1,\\ 2\mathfrak{Q}_{\lambda-1}(-\omega)U_\lambda, & \text{if } \omega<-1,\end{cases}\quad x,y,z,u\in(-1,1),\quad \text{all }\lambda,$$

$$S^\lambda_{310}(x,y,z,u)=\begin{cases}\pm 2\sin\pi\lambda\mathfrak{Q}_{\lambda-1}(\omega)U_\lambda, & \text{if } 1<\omega,\\ \pm\pi P_{\lambda-1}(-\omega)U_\lambda, & \text{if } -1<\omega<1,\\ 0\ \text{or}\ \pm 2\pi\mathfrak{P}_{\lambda-1}(-\omega)U_\lambda, & \text{if } \omega<-1,\end{cases}\quad x,y,z,u\in(-1,1),\quad \text{all }\lambda,$$

where the different cases (0 or $\pm\cdots$; 24 in total) are given in Table II. In this table, $\mathfrak{Q}$ stands for $2\sin\pi\lambda\mathfrak{Q}_{\lambda-1}(\omega)U_\lambda$, P stands for $\pi P_{\lambda-1}(-\omega)U_\lambda$, and $\mathfrak{P}$ stands for $\pi\mathfrak{P}_{\lambda-1}(-\omega)U_\lambda$. It should be noted that $P_{\lambda-1}$ and $\mathfrak{P}_{\lambda-1}$ denote one and the same Legendre function (whereas $Q_{\lambda-1}$ differs from $\mathfrak{Q}_{\lambda-1}$). Out of the many representations in terms of ${}_2F_1$'s for this function we mention here

$$P_{\lambda-1}(-\omega)=\mathfrak{P}_{\lambda-1}(-\omega)={}_2F_1(\tfrac{1}{2}\lambda,\tfrac{1}{2}-\tfrac{1}{2}\lambda;1;1-\omega^2),$$

which holds only for $\operatorname{Re}\omega<0$. In particular for $-1<\omega<1$ we have

$$P_{\lambda-1}(-\omega)=-\cos\pi\lambda P_{\lambda-1}(\omega)+2\pi^{-1}\sin\pi\lambda Q_{\lambda-1}(\omega),\quad -1<\omega<1.$$

TABLE II. Twenty-four cases for the sum $S_{310}^{\lambda}(x,y,z,u)$, $x,y,z,u\in(-1,1)$, $\bar{x}:=(1-x^2)^{1/2}$, $\bar{y}:=(1-y^2)^{1/2}$, $\bar{z}:=(1-z^2)^{1/2}$, $\bar{u}:=(1-u^2)^{1/2}$, and $\bar{x}$, $\bar{y}$, $\bar{z}$, $\bar{u}\in(0,1)$. The third column defines the six cases indicated by 1–6. The second column shows in which interval $\omega:=WT^{-1/2}$ is situated. The signs of $u\bar{z}\pm\bar{u}z$ (last two rows) define the four different cases denoted by A–D. The entries $\pm\mathfrak{Q}$, $\pm P$, $\pm 2\mathfrak{P}$, and 0 in the 6×4 matrix give the value of $S_{310}^{\lambda}(x,y,z,u)$. Here $\mathfrak{Q}$ stands for $2\sin\pi\lambda\,\mathfrak{Q}_{\lambda-1}(\omega)U_\lambda$, P stands for $\pi P_{\lambda-1}(-\omega)U_\lambda$, and $\mathfrak{P}$ for $\pi\mathfrak{P}_{\lambda-1}(-\omega)U_\lambda$.

case	ω	Definition of the cases 1–6	A	B	C	D
1	$1<\omega$	$zu+\bar{z}\bar{u}<xy-\bar{x}\bar{y}$	$\mathfrak{Q}$	$-\mathfrak{Q}$	$\mathfrak{Q}$	$-\mathfrak{Q}$
2	$-1<\omega<1$	$zu-\bar{z}\bar{u}<xy-\bar{x}\bar{y}<zu+\bar{z}\bar{u}<xy+\bar{x}\bar{y}$	P	$-P$	P	$-P$
3	$\omega<-1$	$xy-\bar{x}\bar{y}<zu\pm\bar{z}\bar{u}<xy+\bar{x}\bar{y}$	$2\mathfrak{P}$	0	0	$-2\mathfrak{P}$
4	$\omega<-1$	$zu-\bar{z}\bar{u}<xy\pm\bar{x}\bar{y}<zu+\bar{z}\bar{u}$	0	0	0	0
5	$-1<\omega<1$	$xy-\bar{x}\bar{y}<zu-\bar{z}\bar{u}<xy+\bar{x}\bar{y}<zu+\bar{z}\bar{u}$	P	P	$-P$	$-P$
6	$1<\omega$	$xy+\bar{x}\bar{y}<zu-\bar{z}\bar{u}$	$\mathfrak{Q}$	$\mathfrak{Q}$	$-\mathfrak{Q}$	$-\mathfrak{Q}$
Definition of the cases A–D		Sign of $u+z$ or $u\bar{z}+\bar{u}z$:	+	+	−	−
		Sign of $u-z$ or $u\bar{z}-\bar{u}z$:	+	−	+	−

Note on Table II: For $z,u\in(-1,1)$ we have

$$\operatorname{sgn}(u+z)=\operatorname{sgn}(u\bar{z}+\bar{u}z)\,,\quad \operatorname{sgn}(u-z)=\operatorname{sgn}(u\bar{z}-\bar{u}z)\,,$$

where the signum function is defined by

$$\operatorname{sgn}(x):=\begin{cases}1, & x>0,\\ 0, & x=0,\\ -1, & x<0.\end{cases}$$

Proof follows directly with the help of, e.g.,

$$\cos\gamma+\cos\delta=2\cos\tfrac{1}{2}(\delta-\gamma)\cos\tfrac{1}{2}(\delta+\gamma)\,,\quad \cos\gamma-\cos\delta=2\sin\tfrac{1}{2}(\delta-\gamma)\sin\tfrac{1}{2}(\delta+\gamma)\,.$$

The cases $u=\pm z$ ($u\bar{z}\pm\bar{u}z=0$) can be included in Table II by taking the appropriate limits. That this procedure gives consistent results follows by making the following observations.

(i) $u+z=0$ is possible only in the cases 1, 2, and 4. Then clearly A and C give the same result, as do B and D. Thus for $u+z=0$ we get $\pm\mathfrak{Q}$ in case 1 ($\pm P$ in case 2) according to $\operatorname{sgn}(u-z)=\pm 1$.

(ii) $u-z=0$ is possible only in the cases 4, 5, and 6. Then A and B give the same result, as do C and D. Thus for $u-z=0$ we get $\pm P$ in case 5 ($\pm\mathfrak{Q}$ in case 6) according to $\operatorname{sgn}(u+z)=\pm 1$.

(iii) $u\pm z=0$ (i.e., $u=z=0$) is possible only in case 4; then A, B, C, and D give the same result 0.

We emphasize that the above results for S_{400}^{λ} and S_{310}^{λ} are in agreement with the symmetry relations in Eqs. (1.17) and (1.18).

The result for S_{310}^{λ} given in Ref. 23, Eqs. (3.7)–(3.9), comes down in essence to column B of Table II. The six cases 1–6 are equivalent with those in Eq. (3.9) of Ref. 23 (corrected for a misprint); however, Eq. (3.8) of Ref. 23 is correct only for $W>0$ [see the representations for $P_{\lambda-1}(\pm x)$ in Sec. III].

V. CLOSED FORMULAS FOR S_{klm}^{λ} WITH $l+m=2$

We have the following:

$$S_{002}^{\lambda}(z,u)=e^{i\pi\lambda}(u-z)^{-2\lambda}\mathfrak{D}_0^{\lambda}(\zeta),\quad \zeta:=(uz-1)/(u-z)\,,\quad 1<z<u\quad[1<\zeta]\,,\quad \text{all }\lambda,$$

$$S_{011}^{\lambda}(z,u)=e^{i\pi\lambda}(u-z)^{-2\lambda}D_0^{\lambda}(\zeta),\quad \zeta:=(uz-1)/(u-z)\,,\quad -1<z<1,\quad 1<u\quad[-1<\zeta<1],\quad \text{all }\lambda\,,$$

$$S_{020}^{\lambda}(z,u)=\begin{cases}(u-z)^{-2\lambda}[2\cos\pi\lambda-4e^{-i\pi\lambda}\mathfrak{D}_0^{\lambda}(-\zeta)]\,, & \text{if } -1<z<u<1\quad[\zeta<-1],\\ (z-u)^{-2\lambda}[2\cos\pi\lambda-4e^{-i\pi\lambda}\mathfrak{D}_0^{\lambda}(\zeta)]\,, & \text{if } -1<u<z<1\quad[\zeta>1]\,,\end{cases}$$

$$\zeta:=(uz-1)/(u-z),\quad \operatorname{Re}\lambda<\tfrac{1}{2}\,,$$

$$S_{102}^{\lambda}(x,z,u)=e^{i\pi\lambda}W^{-\lambda}\mathfrak{D}_0^{\lambda}(\zeta),\quad \zeta:=(uz-x)W^{-1/2}\,,\quad -1<x<1,\quad 1<z,\quad 1<u\quad[\zeta>1],\quad \text{all }\lambda\,,$$

$$W:=x^2+z^2+u^2-1-2xzu=(u-xz)^2+(z^2-1)(1-x^2)>0\,,$$

$$S_{111}^{\lambda}(x,z,u)=e^{i\pi\lambda}W^{-\lambda}D_0^{\lambda}(\zeta),\quad \zeta:=(uz-x)W^{-1/2}\,,\quad x,z\in(-1,1),\quad 1<u\quad[-1<\zeta<1],\quad \text{all }\lambda\,,$$

$$W:=x^2+z^2+u^2-1-2xzu=(u-xz+\bar{x}\bar{z})(u-xz-\bar{x}\bar{z})>0\,,$$

$$S^{\lambda}_{120}(x,z,u) = \begin{cases} 2i(-W)^{-\lambda} D^{\lambda}_{0}[i(uz-x)(-W)^{-1/2}], & \text{if } W<0 \quad [zu-\bar{z}\bar{u}<x<zu+\bar{z}\bar{u}], \\ W^{-\lambda}[2\cos\pi\lambda - 4e^{-i\pi\lambda}\mathfrak{D}^{\lambda}_{0}(-\zeta)], & \text{if } zu+\bar{z}\bar{u}<x \quad [W>0,\ \zeta<-1], \\ -W^{-\lambda}[2\cos\pi\lambda - 4e^{-i\pi\lambda}\mathfrak{D}^{\lambda}_{0}(\zeta)], & \text{if } x<zu-\bar{z}\bar{u} \quad [W>0,\ \zeta>1], \end{cases}$$

$$x,z,u\in(-1,1), \quad \operatorname{Re}\lambda<1,$$

$$W := x^2+z^2+u^2-1-2xzu = (x-zu+\bar{z}\bar{u})(x-zu-\bar{z}\bar{u}), \quad \zeta := (uz-x)W^{-1/2};$$

$$S^{\lambda}_{202}(x,y,z,u) = 2a_{\lambda}e^{2i\pi\lambda}\int_0^1 (1-t^2)^{\lambda-1}\xi^{-\lambda}\,dt$$

$$= e^{2i\pi\lambda}2^{2\lambda}a_{\lambda}\lambda^{-1}V^{-\lambda}F_1\left(\lambda,\lambda,\lambda,\lambda+1;\frac{W+T^{1/2}}{V},\frac{W-T^{1/2}}{V}\right),$$

$$\xi := t^2[W+(\bar{z}\bar{u})^2]+2t\bar{z}\bar{u}(zu-xy)+(\bar{z}\bar{u})^2, \quad W := (xy-zu)^2-(\bar{x}\bar{y})^2-(\bar{z}\bar{u})^2,$$

$$V := W+2\bar{z}\bar{u}(\bar{z}\bar{u}+zu-xy), \quad T^{1/2}=2\bar{x}\bar{y}\bar{z}\bar{u}, \quad \bar{x} := (1-x^2)^{1/2}, \quad \bar{y} := (1-y^2)^{1/2},$$

$$x,y\in(-1,1), \quad \bar{z} := (z^2-1)^{1/2}, \quad \bar{u} := (u^2-1)^{1/2}, \quad 1<z, \quad 1<u; \quad \text{all } \lambda.$$

Now S^{λ}_{211} and S^{λ}_{220} can be expressed in terms of the same Appell function[2,3] F_1. Finally, S^{λ}_{003}, and hence S^{λ}_{012}, S^{λ}_{021}, and S^{λ}_{030}, can be expressed in terms of a double integral in various ways.

VI. SPECIAL CASES FOR ZERO VARIABLES

When we take the variables x, y,... equal to zero we get from the definition of S^{λ}_{kl0} a hypergeometric function ${}_{q+1}F_q(\cdots;\pm 1)$, by using

$$C^{\lambda}_{2m+1}(0)=0, \quad C^{\lambda}_{2m}(0)=(-1)^m(\lambda)_m/m!,$$

$$D^{\lambda}_{2m}(0)=0, \quad D^{\lambda}_{2m+1}(0)=(-1)^{m+1}\frac{\Gamma(\lambda+\frac{1}{2})}{\Gamma(\lambda)\Gamma(\frac{3}{2})}\frac{(\lambda+\frac{1}{2})_m}{(\frac{3}{2})_m}.$$

Note that S^{λ}_{k00} $(0,\dots,0)$ and S^{λ}_{0k0} $(0,\dots,0)$ with $-2\lambda\notin\mathbb{N}$ are convergent iff

$$(4-k)\operatorname{Re}\lambda<0, \quad \text{for } k \text{ even}; \quad (4-k)\operatorname{Re}\lambda<1, \quad \text{for } k \text{ odd}.$$

From the closed formulas of Sec. V we thus obtain simple expressions for these ${}_{q+1}F_q(\cdots;\pm 1)$. The results are as follows:

$$S^{\lambda}_{000} = a_{\lambda}^{-1}\lambda\,{}_3F_2(2\lambda,2\lambda,\lambda+1;\lambda,1;1)=0, \quad \operatorname{Re}\lambda<0,$$

$$S^{\lambda}_{100}(0) = 2\cos\pi\lambda = a_{\lambda}^{-1}\lambda\,{}_4F_3(\lambda,\lambda,\lambda+\tfrac{1}{2},1+\tfrac{1}{2}\lambda;\tfrac{1}{2}\lambda,\tfrac{1}{2},1;-1), \quad \operatorname{Re}\lambda<\tfrac{1}{3},$$

$$S^{\lambda}_{200}(0,0) = a_{\lambda}^{-1}\lambda\,{}_3F_2(\lambda,\lambda,1+\tfrac{1}{2}\lambda;\tfrac{1}{2}\lambda,1;1)=0, \quad \operatorname{Re}\lambda<0 \quad (\text{for } \lambda\to 2\lambda \text{ same } {}_3F_2 \text{ as for } S^{\lambda}_{000}),$$

$$S^{\lambda}_{300}(0,0,0) = 2 = a_{\lambda}^{-1}\lambda\,{}_4F_3(\lambda,\lambda,\tfrac{1}{2},1+\tfrac{1}{2}\lambda;\tfrac{1}{2}\lambda,\lambda+\tfrac{1}{2},1;-1), \quad \operatorname{Re}\lambda<1,$$

$S^{\lambda}_{400}(0,0,0,0)$ diverges for all λ,

$$S^{\lambda}_{500}(0,0,0,0,0) = 2\,{}_3F_2(\lambda,\lambda,\tfrac{1}{2};\lambda+\tfrac{1}{2},\lambda+\tfrac{1}{2};1)$$

$$= 2\pi^{1/2}\Gamma(\lambda+\tfrac{1}{2})[\Gamma(\lambda)]^{-1}{}_3F_2(\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2};1,\lambda+\tfrac{1}{2};1)$$

$$= a_{\lambda}^{-1}\lambda\,{}_6F_5(\lambda,\lambda,\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2},1+\tfrac{1}{2}\lambda;\tfrac{1}{2}\lambda,\lambda+\tfrac{1}{2},\lambda+\tfrac{1}{2},\lambda+\tfrac{1}{2},1;-1), \quad \operatorname{Re}\lambda>0;$$

$$\to 2, \quad \text{for } \lambda\downarrow 0,$$

$$= 4G, \quad \text{for } \lambda=1 \quad (G\approx 0.916 \text{ is Catalan's constant})$$

$$= \tfrac{1}{2}\pi^{-3}\Gamma^4(\tfrac{1}{4}), \quad \text{for } \lambda=\tfrac{1}{2};$$

$$S^{\lambda}_{010}(0) = -2\sin\pi\lambda = -8\lambda(\lambda+1)\,{}_4F_3(\lambda+1,\lambda+\tfrac{1}{2},\lambda+\tfrac{1}{2},\tfrac{1}{2}\lambda+\tfrac{3}{2};\tfrac{1}{2}\lambda+\tfrac{1}{2},\tfrac{3}{2},\tfrac{3}{2};-1), \quad \operatorname{Re}\lambda<\tfrac{1}{3},$$

$$S^{\lambda}_{020}(0,0) = -4a_{\lambda}\lambda^{-1} = 16a_{\lambda}(\lambda+1)\,{}_4F_3(1,\lambda+\tfrac{1}{2},\lambda+\tfrac{1}{2},\tfrac{1}{2}\lambda+\tfrac{3}{2};\tfrac{1}{2}\lambda+\tfrac{1}{2},\tfrac{3}{2},\tfrac{3}{2};1), \quad \operatorname{Re}\lambda<0,$$

$$-S^{\lambda}_{030}(0,0,0)\lambda(4a_{\lambda})^{-2} = 2(\lambda+1)\,{}_5F_4(1,1,\lambda+\tfrac{1}{2},\lambda+\tfrac{1}{2},\tfrac{1}{2}\lambda+\tfrac{3}{2};\tfrac{1}{2}\lambda+\tfrac{1}{2},\lambda+1,\tfrac{3}{2},\tfrac{3}{2};-1)$$

$$= {}_3F_2(1,\lambda+\tfrac{1}{2},\tfrac{1}{2};\tfrac{3}{2},\lambda+1;1) = \tfrac{1}{2}\pi\,{}_3F_2(\tfrac{1}{2},\tfrac{1}{2},\lambda;\lambda+1,1;1)$$

$$= \lambda\int_0^1 t^{\lambda-1}K(t^{1/2})\,dt, \quad \operatorname{Re}\lambda<1 \quad (K \text{ is a complete elliptic integral}^{2-4}),$$

$$\to 2, \quad \text{for } \lambda\uparrow 1; \quad =1, \quad \text{for } \lambda=-\tfrac{1}{2}; \quad =\tfrac{5}{3}, \quad \text{for } \lambda=-\tfrac{3}{2};$$

S^{λ}_{040} $(0,0,0,0)$ diverges for all λ.

VII. THE SPECIAL CASES $\lambda = 1$ AND $\lambda \to -N$

There exist many interesting special cases of the sums S^{λ}_{klm}. Here we shall briefly consider the cases $\lambda = 1$ and $\lambda \to -N$, $N = 0,1,\ldots$.

The case $\lambda = 1$: Here it is convenient to set $x = \cos\theta$, $0<\theta<\pi$, and $z = \cosh\theta$, $\theta > 0$. Then we obtain, for $n \in \mathrm{N}$,

$$C^1_n(\cos\theta) = \sin[(n+1)\theta]/\sin\theta, \quad C^1_n(\cosh\theta) = \sinh[(n+1)\theta]/\sinh\theta,$$

$$D^1_n(\cos\theta) = \cos[(n+1)\theta]/\sin\theta, \quad \mathfrak{D}^1_n(\cosh\theta) = -\exp[-(n+1)\theta]/(2\sinh\theta).$$

By inserting these expressions into S^{λ}_{klm} we obtain sums that can be regarded as generalizations of the elementary trigonometric sums

$$\sum_{n=1}^{\infty} n^{-1}\sin\theta n = \tfrac{1}{2}(\pi-\theta), \quad 0<\theta<2\pi, \quad \sum_{n=1}^{\infty} n^{-1}\cos\theta n = -\ln(\sin\tfrac{1}{2}\theta), \quad 0<\theta<2\pi.$$

We shall give closed forms for a few sums S^1_{klm} for which $k+l+m=4$.

Let T, V, and W be defined as in Secs. IV and V, and

$$x = \cos\alpha, \quad \bar{x} = \sin\alpha, \quad y = \cos\beta, \quad \bar{y} = \sin\beta, \quad \alpha,\beta \in (0,\pi).$$

Further we take

$$z = \cos\gamma, \quad \bar{z} = \sin\gamma, \quad u = \cos\delta, \quad \bar{u} = \sin\delta, \quad \gamma,\delta \in (0,\pi),$$

unless otherwise indicated. Then

$$S^1_{301}(x,y,z,u) = \tfrac{1}{2} i(-T)^{-1/2} \ln\left[\frac{W + i(-T)^{1/2}}{W - i(-T)^{1/2}}\right], \quad 1<u,$$

$$S^1_{202}(x,y,z,u) = \tfrac{1}{2} T^{-1/2} \ln\left[\frac{V - W + T^{1/2}}{V - W - T^{1/2}}\right] = \tfrac{1}{2} T^{-1/2} \ln\left[\frac{zu + (z^2-1)^{1/2}(u^2-1)^{1/2} - xy + \bar{x}\bar{y}}{zu + (z^2-1)^{1/2}(u^2-1)^{1/2} - xy - \bar{x}\bar{y}}\right],$$
$$1<z, \quad 1<u,$$

$$S^1_{211}(x,y,z,u) = \tfrac{1}{2}(-T)^{-1/2} \ln\left[\frac{u^2 + z^2 - 1 - 2uz\cos(\alpha-\beta) + \cos^2(\alpha-\beta)}{u^2 + z^2 - 1 - 2uz\cos(\alpha+\beta) + \cos^2(\alpha+\beta)}\right]$$
$$= \tfrac{1}{2}(-T)^{-1/2} \ln\left[\frac{u - \cos(\alpha-\beta+\gamma)}{u - \cos(\alpha+\beta+\gamma)}\,\frac{u - \cos(\alpha-\beta-\gamma)}{u - \cos(\alpha+\beta-\gamma)}\right], \quad 1<u,$$

$$S^1_{220}(x,y,z,u) = T^{-1/2} \ln\left|\frac{\cos\delta - \cos(\alpha+\beta+\gamma)}{\cos\delta - \cos(\alpha-\beta+\gamma)}\,\frac{\cos\delta - \cos(\alpha+\beta-\gamma)}{\cos\delta - \cos(\alpha-\beta-\gamma)}\right|$$
$$= T^{-1/2} \ln\left|\frac{\cos(\alpha+\beta) - \cos(\gamma+\delta)}{\cos(\alpha-\beta) - \cos(\gamma+\delta)}\,\frac{\cos(\alpha+\beta) - \cos(\gamma-\delta)}{\cos(\alpha-\beta) - \cos(\gamma-\delta)}\right|,$$

$$S^1_{211}(0,0,z,u) = (-T)^{-1/2} \ln[(u-z)/(u+z)], \quad -T = 4(1-z^2)(u^2-1), \quad -1<z<1, \quad 1<u,$$

$$S^1_{310}(0,0,u,z) = (2i\bar{z}\bar{u})^{-1}\left[\ln\frac{u+i0-z}{u+i0+z} - \ln\frac{u-i0-z}{u-i0+z}\right], \quad u,z \in (-1,1),$$

$$S^1_{220}(0,0,z,u) = (\bar{z}\bar{u})^{-1} \ln|(u+z)/(u-z)|, \quad z,u \in (-1,1).$$

The case $\lambda \to 0$: We get, for $n = 0$,

$$\lim_{\lambda\to 0} C^{\lambda}_0(z) = 1, \quad \lim_{\lambda\to 0} \mathfrak{D}^{\lambda}_0(z) = \tfrac{1}{2},$$

$$\lim_{\lambda\to 0} \lambda^{-1} D^{\lambda}_0(x) = 2\arcsin x, \quad -1<x<1.$$

For $n>0$ it is convenient to set, as in the case $\lambda = 1$, $x = \cos\theta$, $0<\theta<\pi$, and $z = \cosh\theta$, $\theta>0$, respectively. Then we find, for $n = 1,2,\ldots$,

$$\lim_{\lambda\to 0} \lambda^{-1} C^{\lambda}_n(\cos\theta) = 2n^{-1}\cos\theta n,$$

$$\lim_{\lambda\to 0} \lambda^{-1} C^{\lambda}_n(\cosh\theta) = 2n^{-1}\cosh\theta n,$$

$$\lim_{\lambda\to 0} \lambda^{-1} D^{\lambda}_n(\cos\theta) = -2n^{-1}\sin\theta n,$$

$$\lim_{\lambda\to 0} \lambda^{-1} \mathfrak{D}^{\lambda}_n(\cosh\theta) = n^{-1}e^{-\theta n}.$$

Clearly, insertion of these expressions into S^{λ}_{klm} yields sums that are similar to those that apply to the case $\lambda = 1$. Let us give a few examples:

$$\lim_{\lambda\to 0} S^{\lambda}_{102}(x,z,u) = \tfrac{1}{2},$$

$$\lim_{\lambda\to 0} \lambda^{-1} S^{\lambda}_{111}(x,z,u)$$
$$= 2\arcsin\zeta$$
$$= 2\arcsin z - 4\sum_{n=1}^{\infty} n^{-1}\cos\alpha n \sin\gamma n e^{-\delta n},$$
$$\zeta := (uz - x)W^{-1/2}, \quad x = \cos\alpha,$$
$$z = \cos\gamma, \quad u = \cosh\delta.$$

$$\lim_{\lambda\to 0} \lambda^{-1} S^{\lambda}_{120}(x,z,u)$$
$$= 4i\arcsin[i(uz - x)(-W)^{-1/2}]$$

$$= 8 \sum_{n=1}^{\infty} n^{-1} \cos \alpha n \sin \gamma n \sin \delta n,$$

$$W<0 \quad (u = \cos \delta) .$$

Because of the identity

$$i \arcsin z = \ln[iz + (1 - z^2)^{1/2}] ,$$

this leads to

$$\sum_{n=1}^{\infty} n^{-1} \cos \alpha n \sin \gamma n \sin \delta n$$

$$= \tfrac{1}{2} i \arcsin[i(uz - x)(-W)^{-1/2}]$$

$$= \frac{1}{2} \ln\left(\frac{x - uz + \bar{u}\bar{z}}{(-W)^{1/2}}\right) = \frac{1}{4} \ln\left(\frac{\cos\alpha - \cos(\gamma+\delta)}{\cos(\gamma-\delta) - \cos\alpha}\right),$$

if $W<0$ or, equivalently, if both the numerator and the denominator of the last fraction are positive.

The case $\lambda \to -1$: After analytic continuation with respect to λ we obtain

$$C_0^{-1}(z) = 1, \quad C_1^{-1}(z) = -2z,$$

$$C_2^{-1}(z) = 1, \quad C_n^{-1}(z) = 0, \quad n = 3,4,\ldots,$$

$$\mathfrak{D}_0^{-1}(z) = \tfrac{1}{2}, \quad \mathfrak{D}_1^{-1}(z) = -z,$$

$$\mathfrak{D}_2^{-1}(z) = \tfrac{1}{2}, \quad \mathfrak{D}_n^{-1}(z) = 0, \quad n = 3,4,\ldots,$$

$$D_n^{-1}(z) = 0, \quad n = 0,1,\ldots .$$

Substitution of these expressions in the sums yields in many cases the equality $0 = 0$. However, new interesting sums are generated as follows.

The case $\lambda \to -N$: It turns out that the limits

$$\lim_{\lambda \to -N} (\lambda + N)^{-1} D_{m+N}^{\lambda}(z),$$

$$N = 0,1,\ldots; \quad m = 0, \pm 1, \pm 2, \ldots, \pm N,$$

exist and that the resulting expressions are *even* in m. Hence

$$\lim_{\lambda \to -N} \sum_{n=0}^{2N} (n+\lambda)(\lambda+N)^{-l} \prod_{i=1}^{l} D_n^{\lambda}(z_i) = 0 .$$

Further we find, for $n \geqslant 2N+1$,

$$\lim_{\lambda \to -N} (\lambda+N)^{-1} D_n^{\lambda}(z)$$

$$= -2^{2N+1}(N!)^2[(2N+1)!]^{-1}(1-z^2)^{N+1/2}$$

$$\times C_{n-2N-1}^{N+1}(z)/C_{n-2N-1}^{N+1}(1),$$

$$n - 2N - 1 = 0,1,\ldots .$$

Clearly we obtain in this way from S_{010}^{λ}, $\lambda \to -N$, sums of another, related family of sums of products of Gegenbauer functions.[24]

VIII. A FEW ELEMENTARY DERIVATIONS

We can easily verify that Eqs. (3.1)–(3.6) (with $\nu = n$) are convenient for generating closed forms for the series S_{klm}^{λ} with increasing k, l, or m. Further, Eqs. (1.9)–(1.11) are suitable for transforming (i) $\mathfrak{D}_n^{\lambda}$ into either C_n^{λ} or D_n^{λ}, and (ii) C_n^{λ} and D_n^{λ} into $\mathfrak{D}_n^{\lambda}$. Finally, we can convert C_n^{λ} into $\mathfrak{D}_n^{\lambda}$ or D_n^{λ} by applying Eqs. (3.11) and (3.12), respectively. However, in order to be able to start this process of generating closed forms, we must know at least *one* of the sums S_{klm}^{λ} in closed form, where $k + l + m > 0$. We choose S_{001}^{λ} and we shall now give a short derivation of this sum. Further we shall briefly evaluate S_{101}^{λ} and S_{201}^{λ}.

For evaluating S_{001}^{λ} in closed form, Eq. (3.7) is most convenient. We get

$$S_{001}^{\lambda}(\cosh\alpha)$$

$$:= a_{\lambda}^{-1} \sum_{n=0}^{\infty} (n+\lambda)(2\lambda)_n (n!)^{-1} \mathfrak{D}_n^{\lambda}(\cosh\alpha)$$

$$= a_{\lambda}^{-1} e^{i\pi\lambda} 2^{-\lambda} \pi^{-1}$$

$$\times \sin\pi\lambda \int_{\alpha}^{\infty} e^{-\lambda t} (\cosh t - \cosh\alpha)^{-\lambda}$$

$$\times \lambda\, {}_2F_1(2\lambda, \lambda+1; \lambda; e^{-t}) dt .$$

This is reduced by substituting

$${}_2F_1(2\lambda, \lambda+1; \lambda; z) = (1+z)(1-z)^{-2\lambda-1} .$$

Inserting $\cosh\alpha = 1 + 2\sinh^2 \tfrac{1}{2}\alpha$ and introducing $\tau := \sinh \tfrac{1}{2}t / \sinh \tfrac{1}{2}\alpha$ we reduce the above integral to

$$(\sinh \tfrac{1}{2}\alpha)^{-4\lambda} \lambda\, 2^{1-3\lambda} \int_1^{\infty} (\tau^2 - 1)^{-\lambda} \tau^{-2\lambda-1} d\tau$$

$$= (\cosh\alpha - 1)^{-2\lambda} \lambda\, 2^{-\lambda} \int_1^{\infty} (x-1)^{-\lambda} x^{-\lambda-1} dx .$$

The latter integral equals $B(1-\lambda, 2\lambda) = \Gamma(1-\lambda)\Gamma(2\lambda)/\Gamma(1+\lambda)$ and thus we get directly

$$S_{001}^{\lambda}(\cosh\alpha) = e^{i\pi\lambda}(\cosh\alpha - 1)^{-2\lambda} .$$

In order to derive S_{101}^{λ} we use Eq. (3.4) and obtain

$$S_{101}^{\lambda}(z,u) = 2a_{\lambda} \int_{-1}^{1} S_{001}^{\lambda}(zu + \bar{z}\bar{u}t)(1-t^2)^{\lambda-1} dt$$

$$= e^{i\pi\lambda}(u-z)^{-2\lambda},$$

$$1 < z < u, \quad \bar{z} := (z^2-1)^{1/2}, \quad \bar{u} := (u^2-1)^{1/2},$$

where we used

$$2a_{\lambda} \int_{-1}^{1} (\zeta \pm t)^{-2\lambda}(1-t^2)^{\lambda-1} dt = (\zeta^2 - 1)^{-\lambda},$$

$$\zeta > 1 .$$

In exactly the same way we find

$$S_{201}^{\lambda}(x,z,u) = 2a_{\lambda} \int_{-1}^{1} S_{101}^{\lambda}(x, zu + \bar{z}\bar{u}t)(1-t^2)^{\lambda-1} dt$$

$$= e^{i\pi\lambda} W^{-\lambda},$$

$$W := x^2 + z^2 + u^2 - 1 - 2xzu$$

$$= (u - xz)^2 - (1-x^2)(1-z^2) .$$

IX. ELEMENTARY DERIVATION OF S_{301}^{λ}

Let now $x, y \in (-1,1)$, $z,u \in (1,\infty)$, and

$$\bar{x} := (1-x^2)^{1/2}, \quad \bar{y} := (1-y^2)^{1/2},$$

$$\bar{z} := (z^2-1)^{1/2}, \quad \bar{u} := (u^2-1)^{1/2} .$$

Then

$$S_{301}^{\lambda}(x,y,z,u)$$

$$= 2a_{\lambda} \int_{-1}^{1} S_{201}^{\lambda}(x, y, zu + \bar{z}\bar{u}t)(1-t^2)^{\lambda-1} dt$$

$$= 2a_{\lambda} e^{i\pi\lambda} \int_{-1}^{1} [(zu + \bar{z}\bar{u}t - xy)^2 - (\bar{x}\bar{y})^2]^{-\lambda}$$

$$\times (1-t^2)^{\lambda-1}\,dt$$
$$= e^{i\pi\lambda} a_\lambda 2^{\lambda+1} T^{-(1/2)\lambda} \mathfrak{Q}_{\lambda-1}(WT^{-1/2}),$$
$$T^{1/2} := 2\bar{x}\bar{y}\bar{z}\bar{u} > 0,$$
$$W := (zu-xy)^2 - (\bar{z}\bar{u})^2 - (\bar{x}\bar{y})^2$$
$$= (zx-uy)^2 + (z^2-1)(1-x^2) + (u^2-1)(1-y^2) > 0.$$

The last integral here can be evaluated in an elementary way as follows.

With the help of

$$\tau := (a+c-2b)^{1/2}(a+c+2b)^{-1/2}(1-t)(1+t)^{-1}$$

and

$$\int_0^\infty (1+2\beta\tau+\tau^2)^{-\lambda}\tau^{\lambda-1}\,d\tau$$
$$= 2^{1-\lambda}(\beta^2-1)^{-(1/2)\lambda}\mathfrak{Q}_{\lambda-1}[\beta(\beta^2-1)^{-1/2}],$$
$$\operatorname{Re}\beta > 0, \quad \operatorname{Re}\lambda > 0,$$

we obtain

$$\int_{-1}^{1} (a \pm 2bt + ct^2)^{-\lambda}(1-t^2)^{\lambda-1}dt$$
$$= 2^{2\lambda-1}[(a+c)^2-4b^2]^{-(1/2)\lambda} \times \int_0^\infty (1+2\beta\tau+\tau^2)^{-\lambda}\tau^{\lambda-1}\,d\tau$$
$$= 2^\lambda T^{-(1/2)\lambda}\mathfrak{Q}_{\lambda-1}(WT^{-1/2}),$$
$$\beta := (a-c)[(a+c)^2-4b^2]^{-1/2};$$

in the present case a, b, and c are to be identified as

$$a := (zu-xy)^2 - (\bar{x}\bar{y})^2,$$
$$b := \bar{z}\bar{u}(zu-xy),$$
$$c := (\bar{z}\bar{u})^2,$$

and thus

$$b^2 - ac = T/4, \quad a - c = W,$$
$$(a+c)^2 - 4b^2 = (a-c)^2 - 4(b^2-ac) = W^2 - T,$$
$$\beta = W(W^2-T)^{-1}.$$

X. DERIVATION OF S^λ_{400}

In this section we shall derive S^λ_{400} from S^λ_{300}. By employing Eq. (3.1) again, we obtain

$$S^\lambda_{400}(x,y,z,u)$$
$$= 2a_\lambda \int_{-1}^{1} S^\lambda_{300}(xy+\bar{x}\bar{y}t,z,u)(1-t^2)^{\lambda-1}dt,$$
$$x,y,z,u \in (-1,1), \quad \bar{x} := (1-x^2)^{1/2}, \quad \bar{y} := (1-y^2)^{1/2}.$$

Let further $\bar{z} := (1-z^2)^{1/2}$, $\bar{u} := (1-u^2)^{1/2}$,

$$t_\pm := (zu \pm \bar{z}\bar{u} - xy)/\bar{x}\bar{y} \quad [t_- < t_+],$$
$$p(t) := 4a_\lambda(\bar{x}\bar{y})^{-2\lambda}[(t-t_-)\times(t-t_+)]^{-\lambda}(1-t^2)^{\lambda-1},$$
$$n(t) := 4a_\lambda(\bar{x}\bar{y})^{-2\lambda}[-(t-t_-)\times(t-t_+)]^{-\lambda}(1-t^2)^{\lambda-1}.$$

We express the six cases 1–6 defined in Sec. IV, Table II, in terms of t_+ and t_-. Thus we obtain from the closed expression for S^λ_{300} (Sec. IV) for S^λ_{400} the expressions given in Table III.

In Table III,

$$p_1 := \int_{-1}^{1} p(t)dt, \quad t_\pm < -1,$$
$$p_6 := \int_{-1}^{1} p(t)dt, \quad 1 < t_\pm,$$
$$p_2 := \int_{t_+}^{1} p(t)dt, \quad t_- < -1 < t_+ < 1,$$
$$p_5 := \int_{-1}^{t_-} p(t)dt, \quad -1 < t_- < 1 < t_+,$$
$$p_3 := \int_{-1}^{t_-} p(t)dt, \quad -1 < t_\pm < 1,$$
$$p_3' := \int_{t_+}^{1} p(t)dt, \quad -1 < t_\pm < 1,$$

and

$$n_2 := \int_{-1}^{t_+} n(t)dt, \quad t_- < -1 < t_+ < 1,$$
$$n_5 := \int_{t_-}^{1} n(t)dt, \quad -1 < t_- < 1 < t_+,$$
$$n_3 := \int_{t_-}^{t_+} n(t)dt, \quad -1 < t_\pm < 1,$$
$$n_4 := \int_{-1}^{1} n(t)dt, \quad t_- < -1, \quad 1 < t_+.$$

These integrals are invariant under the transformation $x \to -x$ and $z \to -z$, as follows from the closed expressions to be given shortly. Under this transformation we get $t_+ \to -t_-$, $t_- \to -t_+$, and hence $p(t) \to p(-t)$ and $n(t) \to n(-t)$, so that

$$p_1 = p_6, \quad p_2 = p_5, \quad p_3 = p_3', \quad \text{and } n_2 = n_5.$$

The integrals p_i and n_i reduce to hypergeometric-function integrals after elementary transformations. The resulting hypergeometric functions ${}_2F_1$ are most conveniently expressed in terms of the Legendre functions $P_{\lambda-1}$, $\mathfrak{P}_{\lambda-1}$, $\mathfrak{Q}_{\lambda-1}$, and $\mathfrak{Q}_{-\lambda}$. Using, as in Sec. IV,

$$\omega := WT^{-1/2}, \quad U_\lambda := a_\lambda 2^{\lambda+1}T^{-(1/2)\lambda},$$

we obtain

TABLE III. Expressions for S^λ_{400} in the same six cases as defined in Table II.

case	ω	condition on $t_\pm$	S^λ_{400}
1	$1<\omega$	$t_\pm < -1$	$p_1 \cos \pi\lambda$
2	$-1<\omega<1$	$t_- < -1 < t_+ < 1$	$n_2 + p_2 \cos \pi\lambda$
3	$\omega < -1$	$-1 < t_\pm < 1$	$p_3 \cos \pi\lambda + n_3 + p_3' \cos \pi\lambda$
4	$\omega < -1$	$t_- < -1, \quad 1 < t_+$	n_4
5	$-1<\omega<1$	$-1 < t_- < 1 < t_+$	$p_5 \cos \pi\lambda + n_5$
6	$1<\omega$	$1 < t_\pm$	$p_6 \cos \pi\lambda$

$$p_1 = 2\mathfrak{Q}_{\lambda-1}(\omega)U_\lambda, \quad 1<\omega,$$
$$p_2 = (\pi/\sin\pi\lambda)P_{\lambda-1}(-\omega)U_\lambda, \quad -1<\omega<1,$$
$$p_3 = (\pi/\sin\pi\lambda)\mathfrak{P}_{\lambda-1}(-\omega)U_\lambda, \quad \omega<-1,$$
$$n_2 = (\pi/\sin\pi\lambda)P_{\lambda-1}(\omega)U_\lambda, \quad -1<\omega<1,$$
$$n_3 = 2\mathfrak{Q}_{-\lambda}(-\omega)U_\lambda, \quad \omega<-1,$$
$$n_4 = 2\mathfrak{Q}_{\lambda-1}(-\omega)U_\lambda, \quad \omega<-1.$$

The actual derivations of these expressions are quite lengthy and are therefore omitted. We note that p_1, p_2, and p_3 have been derived by Rahman and Shah[23] (in a different notation). Our result for p_1 agrees with theirs, but for p_2 and p_3 there is a slight disagreement. To be specific, Eq. (A13) of Ref. 23 is valid only for $\operatorname{Re} z<\frac{1}{2}$ and consequently Eq. (3.8) of Ref. 23 is correct only for $W>0$; the ${}_2F_1$ given there equals $P_{\lambda-1}$ or $\mathfrak{P}_{\lambda-1}$ only if $W>0$.

By inserting the above closed forms for p_i and n_i into Table III we get closed forms for S^λ_{400}. These are recast into a more attractive form by using the relations[2–5]

$$P_{\lambda-1}(x) + \cos\pi\lambda\,\mathrm{P}_{\lambda-1}(-x) = 2\pi^{-1}\sin\pi\lambda Q_{\lambda-1}(-x), \quad -1<x<1,$$

and

$$\mathfrak{Q}_{-\lambda}(z) = \mathfrak{Q}_{\lambda-1}(z) - \pi\cot\pi\lambda\mathfrak{P}_{\lambda-1}(z), \quad z\notin(-\infty,1].$$

Thus we obtain in cases:

1,6: $p_1\cos\pi\lambda = 2\cos\pi\lambda\mathfrak{Q}_{\lambda-1}(\omega)U_\lambda, \quad 1<\omega,$

2,5: $p_2\cos\pi\lambda + n_2 = 2Q_{\lambda-1}(-\omega)U_\lambda, \quad -1<\omega<1,$

3 : $2p_3\cos\pi\lambda + n_3 = 2\mathfrak{Q}_{\lambda-1}(-\omega)U_\lambda, \quad \omega<-1,$

4 : $n_4 = 2\mathfrak{Q}_{\lambda-1}(-\omega)U_\lambda, \quad \omega<-1.$

This completes the derivation of $S^\lambda_{400}(x,y,z,u)$.

XI. DERIVATION OF S^λ_{310}

Let us now derive S^λ_{310} from S^λ_{210}. By employing Eq. (3.1) again we obtain

$$S^\lambda_{310}(x,y,z,u) = 2a_\lambda\int_{-1}^{1} S^\lambda_{210}(xy+\bar{x}\bar{y}t,z,u)\times(1-t^2)^{\lambda-1}\,dt. \tag{11.1}$$

We use the same notation as in Sec. X. Let us first recall the closed expression for S^λ_{210}:

$$S^\lambda_{210}(x,z,u) = 0, \quad \text{if } W_3<0, \tag{11.2}$$
$$= 2\sin\pi\lambda W_3^{-\lambda}, \quad \text{if } xz+\bar{x}\bar{z}<u \quad [W_3>0], \tag{11.3}$$
$$= -2\sin\pi\lambda W_3^{-\lambda}, \quad \text{if } xz-\bar{x}\bar{z}>u \quad [W_3>0], \tag{11.4}$$

where

$$W_3 := x^2+z^2+u^2-1-2xzu.$$

First of all, $W_3<0$ is equivalent to

$$zu-\bar{z}\bar{u}<x<zu+\bar{z}\bar{u}.$$

Replacing x by $xy+\bar{x}\bar{y}t$ we see that this is equivalent to

$$t_-<t<t_+,$$

where, as before,

$$t_\pm := (zu\pm\bar{z}\bar{u}-xy)/\bar{x}\bar{y}.$$

Thus we see from (11.2) that $\int_{t_-}^{t_+}\cdots = 0$ in Eq. (11.1). We distinguish again the six different cases 1–6 of Tables II and III. Moreover, the conditions $xz+\bar{x}\bar{z}<u$ and $xz-\bar{x}\bar{z}>u$ occurring in S^λ_{210} are not suitable for direct application in Eq. (11.1). Therefore we convert these conditions as follows:

$$xz+\bar{x}\bar{z}<u \Leftrightarrow \text{either } zu+\bar{z}\bar{u}<x \text{ and } u\bar{z}-\bar{u}z>0 \text{ or } zu-\bar{z}\bar{u}>x \text{ and } u\bar{z}+\bar{u}z>0,$$
$$xz-\bar{x}\bar{z}>u \Leftrightarrow \text{either } zu+\bar{z}\bar{u}<x \text{ and } u\bar{z}-\bar{u}z<0 \text{ or } zu-\bar{z}\bar{u}>x \text{ and } u\bar{z}+\bar{u}z<0.$$

Thus we get the positive sign from Eq. (11.3) if either

$$t_+<t \quad\text{and}\quad u\bar{z}-\bar{u}z>0$$

or

$$t_->t \quad\text{and}\quad u\bar{z}+\bar{u}z>0,$$

and the negative sign from Eq. (11.4) if either

$$t_+<t \quad\text{and } u\bar{z}-\bar{u}z<0$$

or

$$t_->t \quad\text{and } u\bar{z}+\bar{u}z<0.$$

In this way we find for $S^\lambda_{310}(x,y,z,u)/\sin\pi\lambda$ the expressions under the associated conditions as displayed in Table IV.

In Sec. X we have given closed forms for p_1, p_2, and p_3. By substituting these expressions we find that the above results lead to the closed forms for S^λ_{310} given in Table II, Sec. IV. Thus Tables II and IV give the same information, presented in a different way.

Apparently our result for S^λ_{310} differs from that obtained in Ref. 23. In our opinion, the statement directly following Eq. (3.2) of Ref. 23 is incorrect and this leads to unjustifiable simplifications.

XII. PARTIAL SECOND DERIVATION OF S^λ_{310}

Let us finally employ Eq. (3.3) for deriving S^λ_{310} from S^λ_{210}. Using the symmetry relations (1.15)–(1.17) we obtain

$$c^\lambda_n\int_{-1}^{1} D^\lambda_n(xy+\bar{x}\bar{y}t)(1-t^2)^{\lambda-1}\,dt = \begin{cases} C^\lambda_n(x)D^\lambda_n(y), & \text{if } x>\pm y, \\ -C^\lambda_n(x)D^\lambda_n(y), & \text{if } x<\pm y, \end{cases}$$

where each condition should hold for both signs + and −. By interchanging x and y we get two other expressions that give, however, no new information. Defining

$$S := 2a_\lambda\int_{-1}^{1} S^\lambda_{210}(z,u,xy+\bar{x}\bar{y}t)(1-t^2)^{\lambda-1}\,dt$$

we obtain

$$S = \begin{cases} S^\lambda_{310}(x,y,z,u), & \text{if } [x>\pm y \Leftrightarrow x\bar{y}\pm\bar{x}y>0], \\ -S^\lambda_{310}(x,y,z,u), & \text{if } [x<\pm y \Leftrightarrow x\bar{y}\pm\bar{x}y<0]. \end{cases}$$

TABLE IV. Expressions for $S_{310}^{\lambda}/\sin \pi\lambda$ in the same six cases as in Tables II and III.

case	conditions on $t_{\pm}$	$S_{310}^{\lambda}/\sin \pi\lambda$
1	$t_{\pm} < -1$	$\pm \int_{-1}^{1} p(t)dt = \pm p_1, \quad u\bar{z} - \bar{u}z \gtrless 0$
2	$t_- < -1 < t_+ < 1$	$\pm \int_{t_+}^{1} p(t)dt = \pm p_2, \quad u\bar{z} - \bar{u}z \gtrless 0$
3	$-1 < t_- < t_+ < 1$	$\pm \int_{t_+}^{1} p(t)dt = \pm p_3, \quad u\bar{z} - \bar{u}z \gtrless 0$
		plus $\pm \int_{-1}^{t_-} p(t)dt = \pm p_3, \quad u\bar{z} + \bar{u}z \gtrless 0$
4	$t_- < -1, \quad 1 < t_+$	Zero
5	$-1 < t_- < 1 < t_+$	$\pm \int_{-1}^{t_-} p(t)dt = \pm p_2, \quad u\bar{z} + \bar{u}z \gtrless 0$
6	$1 < t_{\pm}$	$\pm \int_{-1}^{1} p(t)dt = \pm p_1, \quad u\bar{z} + \bar{u}z \gtrless 0$

Now S is easily evaluated by inserting the closed form for S_{210}^{λ} (see Sec. IV). With exactly the same six cases 1–6 considered before we find that S corresponds precisely to the column indicated by C in Table II, Sec. IV. Hence S_{310}^{λ} (x, y, z, u) is given by

$$\text{column } C, \quad \text{if } x\bar{y} \pm \bar{x}y > 0 ,$$
$$\text{column } B, \quad \text{if } x\bar{y} \pm \bar{x}y < 0 .$$

In order to be able to compare this with the results given in Table II, we interchange $x \leftrightarrow z$ and $y \leftrightarrow u$. Then the cases 1–6 are interchanged as follows: $1 \leftrightarrow 6$, $2 \leftrightarrow 5$, and $3 \leftrightarrow 4$; apparently this effectively comes down to interchanging the columns B and C. Thus we find that S_{310}^{λ} (x, y, z, u) is given by

$$\text{column } B, \quad \text{if } z\bar{u} \pm \bar{z}u > 0 ,$$
$$\text{column } C, \quad \text{if } z\bar{u} \pm \bar{z}u < 0 ,$$

which is in complete agreement with Table II. Note, however, that the full result of Table II is not retrieved in this way.

[1]E. R. Hansen, *A Table of Series and Products* (Prentice-Hall, Englewood Cliffs, NJ, 1975).

[2]Bateman Manuscript Project, *Higher Transcendental Functions*, edited by A. Erdélyi (McGraw-Hill, New York, 1953–1955), Vols. I–III; *Tables of Integral Transforms*, edited by A. Erdélyi (McGraw-Hill, New York, 1953–1955), Vols. I and II.

[3]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products, Corrected and Enlarged Edition* (Academic, New York, 1980).

[4]W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer, New York, 1966).

[5]The formula books under 1–4 contain misprints and errors. Roughly a thousand of these have been collected and listed in H. van Haeringen and L. P. Kok, Corrigenda Reports 85 01-85 04, Delft–Groningen, 1985.

[6]L. Gegenbauer, Sitzungsber. Math. Naturwiss. Kl. Akad. Wiss. Wien, **70**, 433 (1874); **75**, 891 (1877); **97**, 259 (1888); **100**, 225 (1891); **102**, 942 (1893); Denkschr. Akad. Wiss. Wien, Math. Naturwiss. Kl. **48**, 293 (1884); **57**, 425 (1890).

[7]E. Heine, *Handbuch der Kugelfunctionen*, Vols. I and II (Reimer, Berlin, 1878, 1881), 2nd ed.

[8]G. Szegö, Math. Ann. **110**, 501 (1934).

[9]G. Szegö, *Orthogonal Polynomials* (Am. Math. Soc., Providence, Rhode Island, 1975), 4th ed.

[10]A. M. Legendre, *Exercices de Calcul Intégral sur Divers Ordres de Transcendantes*, Vol. 2 (Courcier, Paris, 1817).

[11]H. van Haeringen and L. P. Kok, J. Math. Phys. **22**, 2482 (1981).

[12]R. Askey, *Orthogonal Polynomials and Special Functions* (SIAM, Philadelphia, 1975).

[13]L. Durand, in *Theory and Applications of Special Functions*, edited by R. Askey (Academic, New York, 1975), pp. 353–374.

[14]L. Durand, SIAM J. Math. Anal. **9**, 76 (1978).

[15]M. Rahman, SIAM J. Math. Anal. **8**, 1014 (1977).

[16]M. Rahman, Canad. J. Math. **33**, 915 (1981).

[17]A. M. Din, Lett. Math. Phys. **5**, 207 (1981).

[18]H. van Haeringen, Delft Prog. Rep. Ser. A **6**, 236 (1981).

[19]H. van Haeringen, J. Math. Phys. **23**, 964 (1982).

[20]H. van Haeringen, J. Math. Phys. **24**, 1054 (1983).

[21]R. Askey, T. H. Koornwinder, and M. Rahman, "An integral of products of ultraspherical functions and a q-extension," Proc. Lond. Math. Soc., to appear.

[22]M. Rahman and M. J. Shah, SIAM J. Math. Anal. **16**, 859 (1985).

[23]M. Rahman and M. J. Shah, J. Math. Phys. **26**, 627 (1985).

[24]H. van Haeringen, "Sums of products of Gegenbauer and Legendre functions," in preparation.

[25]J. Dougall, Proc. Edinburgh Math. Soc. **37**, 33 (1919).

[26]E. C. Titchmarsh, *The Theory of Functions* (Oxford U. P., London, 1939), 2nd ed.

Partial wave decomposition of the Glauber amplitude for the elastic scattering of structureless charged particles by atomic hydrogen

Brian K. Thomas[a)] and Victor Franco
Department of Physics, Brooklyn College of CUNY, Brooklyn, New York 11210

(Received 12 August 1985; accepted for publication 20 November 1985)

The partial wave amplitudes for the conventional Glauber approximation to the elastic scattering of structureless charged particles by hydrogen atoms are evaluated in closed form. The asymptotic behavior of these amplitudes in various limits is described. The large-r asymptotic behavior of the Glauber effective interaction and the logarithmic divergence of the elastic Glauber amplitude as $\theta \rightarrow 0$ are discussed. Some numerical results for these partial amplitudes are presented.

I. INTRODUCTION

Applications of the Glauber approximation[1] to atomic collisions display two seemingly contradictory features. Present evidence[2,3] indicates that, on the whole, the Glauber approximation to the scattering amplitude yields reasonably reliable predictions of both differential and integrated cross sections for *inelastic* scattering of charged particles by neutral atoms at intermediate and high energies, i.e., at incident particle speeds $v_i \gtrsim 2$ a.u. On the other hand, it is now well established that the Glauber approximation fails to predict the measured absolute angular distributions for *elastic* scattering of electrons by hydrogen[2,4–7] and helium[2,7,8] atoms. The failure of the Glauber approximation is twofold for these elastic collisions. At very small scattering angles θ near the forward direction the Glauber predictions diverge[2,9,10] as $\ln(\sin \theta/2)$, whereas reasonable extrapolations of the data[4] to the forward direction appear finite. At intermediate and wide scattering angles the Glauber approximation appears to predict the shapes of the measured angular distributions, but substantially underestimates the absolute data.[4] For example, in e^-–H(1s) elastic scattering at 140°, the Glauber prediction is too low by $\sim 40\%$ for 100 eV incident electrons and by nearly an order of magnitude for 30 eV incident electrons. Although exchange effects are neglected throughout in these Glauber applications, at intermediate incident electron energies, these effects, even when appreciable, are not expected to be as large as the aforementioned differences between theory and experiment. It is not surprising, however, that below ~ 100 eV the Glauber predictions are inaccurate. It is also not surprising that the Glauber predictions near 140° are inaccurate. The theory is not expected to be valid for such energies and for such scattering angles.[2,9,11]

In view of the truly remarkable success of the Glauber approximation when applied to *inelastic* atomic collisions, the apparent failure of the approximation in these applications to elastic scattering is not clearly understood in detail. Nonetheless, it has been shown[6,9,10] that the $\ln(\sin \theta/2)$ divergence of the Glauber amplitude as $\theta \rightarrow 0$ is traceable to the long-range nature of the Coulomb force, together with the use of the so-called closure approximation in the derivation[2] of the Glauber amplitude formula from the Lippmann–Schwinger integral equation for composite systems. Thus the failure of the Glauber approximation for small angle elastic scatterings (at small momentum transfers) is apparently understandable. On the other hand, Thomas[12] has shown that the failure of the Glauber approximation for wide angle e^-–H elastic scattering at incident electron energies $E_i \geqslant 50$ eV is due to the normalization of those Glauber predictions. In particular it is found that the observed[4] wide angle e^-–H angular distributions are quite reliably consistent with the predictions of point Coulomb scattering of the incident electron by the nuclear proton, and that the Glauber predictions, while reproducing the shape of point Coulomb scattering at these angles, fail to reproduce the normalization of purely point Coulomb scattering. Alternatively, Ishihara and Chen[5] have argued that the failure of the Glauber approximation at all scattering angles stems primarily from an improper treatment of the small angular momentum contributions to the predicted elastic scattering amplitude.

The failure of the Glauber approximation to provide reliable absolute estimates of elastic electron–neutral atom scattering has prompted several attempts to improve upon or provide alternatives to the Glauber approximation for these collisions. These efforts include the two-potential eikonal approximation,[5] the eikonal–Born–series approximation,[6,7] the Glauber angle approximation,[13] the eikonal–optical model,[14] and the modified Glauber approximation of Gien,[15] among others.[2,16–19] Rather than suggest another such alternative to the conventional Glauber approximation, we examine the partial wave decomposition of the full Glauber amplitude for the elastic scattering of structureless charged particles by ground state hydrogen atoms. We are able to obtain the corresponding Glauber partial wave amplitudes in closed form. This result not only reflects another[2] of the remarkable analytic properties of the Glauber amplitude, but more importantly should enable an assessment of the elastic Glauber predictions in terms of general theoretic considerations[20] conventionally describing the properties of the elastic scattering partial wave amplitudes and the corresponding phase shifts. Moreover, we believe this to be one of the few, if not the only, successful attempts to obtain analytic closed form expressions for the partial wave amplitudes associated with an essentially high energy approximate to the

a) Deceased.

 0022-2488/86/040953-13$02.50

full scattering amplitude, other than first Born. Indeed, in atomic collision theory high-energy approximates are usually formulated explicitly to avoid the computation of the partial wave amplitudes.

Although the analysis is necessarily detailed and somewhat complicated, we are able to show clearly from these results that the failure of the full Glauber amplitude at *small* scattering angles stems primarily from the *unphysical* behavior of the Glauber partial wave amplitudes at *large* angular momenta. This unphysical behavior is reflected in the large-r asymptotic behavior of the effective potential scattering interaction for the Glauber approximation; to leading order the interaction is found to be purely *absorptive* and proportional to r^{-3}. The $\ln(\sin\theta/2)$ divergence of the elastic Glauber amplitude occurs in the imaginary part of the amplitude. Since the observed e^-–H total scattering cross sections are finite, the Glauber approximation fails to satisfy the optical theorem.[2] Nevertheless, we are able to clarify somewhat the sense in which the Glauber approximation approximately satisfies other constraints imposed by unitarity.

We stress that the analysis and discussion in this paper are intended not to supplant, but rather to supplement, presently available analyses[2,5–7,12–19] of the deficiencies of the Glauber approximation in charged particle–hydrogen atom elastic scattering. However, we also believe the availability of these Glauber partial wave elastic scattering amplitudes may lead to further insight into the physical content and deficiencies of the Glauber approximation as applied to atomic collisions.

The contents of this paper now may be summarized as follows. In Sec. II the Glauber partial wave amplitudes for the elastic scattering of structureless charged particles by hydrogen atoms is obtained in closed form. In Sec. III we derive the large-r asymptotic behavior of the effective potential scattering interaction for the Glauber approximation from the large-l behavior of the partial amplitudes, while in Sec. IV we show that the large-l behavior of the partial amplitudes leads directly to the $\ln(\sin\theta/2)$ divergence as $\theta\to 0$. In Sec. V we present the results for some numerical computations of the partial wave amplitudes both as functions of angular momentum l at fixed incident energy, and as functions of energy at fixed l. We summarize our conclusions in Sec. VI. For convenience, some of the detailed analysis leading to the results obtained in Secs. III–V has been deferred to the appendices. We obtain the asymptotic behavior of the amplitude at large angular momenta in Appendix A, while in Appendix B we obtain the asymptotic behavior of the Glauber partial wave amplitudes in various limits, including the limit as the incident energy approaches zero, all at fixed angular momenta.

II. REDUCTION OF THE GLAUBER ELASTIC PARTIAL WAVE AMPLITUDE

The Glauber approximation to the scattering amplitude for a direct collision (excluding exchange or rearrangement, but including ionization) of a structureless particle of charge $Z_i e$ with a hydrogen atom which consequently undergoes a transition from an initial state i to a final state f is given by[1,2,9]

$$F^{G}(i\to f;\mathbf{q}) = \frac{-iK_i}{2\pi}\int e^{i\mathbf{q}\cdot\mathbf{b}}\,u_f^*(\mathbf{r})\times\{e^{i\chi(\mathbf{b},\mathbf{r})}-1\}\,u_i(\mathbf{r})\,d^2b\,d\mathbf{r}, \tag{1a}$$

where the phase shift function χ is given by

$$\chi(\mathbf{b},\mathbf{r}) = -\frac{1}{\hbar v_i}\int_{-\infty}^{\infty} dz'\,V_i(\mathbf{r}',\mathbf{r}). \tag{1b}$$

In Eqs. (1a) and (1b) $V_i(\mathbf{r}',\mathbf{r})$ is the interaction potential seen by the incident particle with coordinate $\mathbf{r}'$; $\mathbf{r}$ denotes the coordinate of the initially bound electron relative to the target atom nucleus; and u_i and u_f specify the initial and final state wave functions of the target atom. The momentum transfer is given by $\mathbf{q}=\mathbf{K}_i-\mathbf{K}_f$ with $\hbar\mathbf{K}_i, \hbar\mathbf{K}_f=\mu\mathbf{v}_i, \mu\mathbf{v}_f$, where $\mathbf{v}_i$ and $\mathbf{v}_f$ are the initial and final relative velocities of the scattered particle in the center-of-mass system, and μ is the reduced mass of the incident particle–hydrogen atom pair. It is by now well understood[2,3] that identifying Eqs. (1a) and (1b) as the direct scattering amplitude incorporates the subsumption that the z direction in Eq. (1b) is to be taken along a direction $\hat{\mathbf{v}}$ perpendicular to $\mathbf{q}$, at each $\mathbf{K}_i$ and $\mathbf{K}_f$, and such that $\hat{\mathbf{v}}$ lies in the scattering plane. Thus in Eqs. (1), $\mathbf{b}$ is the projection of $\mathbf{r}$ onto the plane perpendicular to $\hat{\mathbf{v}}$.

When spin-dependent interactions are neglected, V_i in Eq. (1b) is

$$V_i = Z_i e^2/r' - Z_i e^2/|\mathbf{r}'-\mathbf{r}|,$$

and the Glauber amplitudes for direct elastic scattering and excitation from the ground state H(1s) can be evaluated in closed form.[3,10] In particular, the Glauber elastic scattering amplitude is given by

$$F^{G}(1s\to 1s;\mathbf{q}) = -iK_i\,\frac{1}{4}\lambda^3\frac{\partial}{\partial\lambda}I_0(\lambda,q)\bigg|_{\lambda=2/a_0}, \tag{2a}$$

where

$$I_0(\lambda,q) = -4i\eta|\Gamma(1+i\eta)|^2\lambda^{-2-2i\eta}q^{-2+2i\eta}\times{}_2F_1(1-i\eta,1-i\eta;1;-\lambda^2/q^2), \tag{2b}$$

with $\eta\equiv -Z_i e^2/\hbar v_i$ and a_0 the Bohr radius; of course, for elastic scattering $K_i=K_f$ and $q^2=2K_i^2(1-\cos\theta)$, where θ is the center-of-mass scattering angle of the outgoing particle.

Since $F^{G}(1s\to 1s;\mathbf{q})$ is azimuthally symmetric, F^{G} can be expanded in Legendre polynomials $P_l(\cos\theta)$. Thus we write

$$F^{G}(1s\to 1s;\mathbf{q}) = \sum_{l=0}^{\infty}(2l+1)\,\mathscr{F}_l^{G}(1s\to 1s;K_i)\,P_l(\cos\theta) \tag{3}$$

so that the partial wave amplitudes $\mathscr{F}_l^{G}$ are given by

$$\mathscr{F}_l^{G}(1s\to 1s;K_i) = \frac{1}{2}\int_0^{\pi}\sin\theta\,d\theta\,P_l(\cos\theta)\,F^{G}(1s\to 1s;\mathbf{q}). \tag{4a}$$

Comparing Eq. (3) with the usual expansion[21] of the scattering amplitude in terms of the phase shift δ_l, we see that

$$\mathscr{F}_l^{G}(1s\to 1s;K_i) = \frac{1}{2iK_i}(e^{2i\delta_l}-1) = \frac{1}{K_i}e^{i\delta_l}\sin\delta_l, \tag{4b}$$

where δ_l is generally complex.

If we insert Eqs. (2a) and (2b) into (4a), we obtain the analogs of (2a) and (2b) for the partial wave amplitudes; namely,

$$\mathscr{F}_l^{\rm G}(1s\to 1s;K_i) = -iK_i\frac{1}{4}\lambda^3\frac{\partial}{\partial\lambda}\mathscr{I}_l^{\rm G}(\lambda,K_i)\Big|_{\lambda=2/a_0}, \quad (5a)$$

where

$$\mathscr{I}_l^{\rm G}(\lambda,K_i)\equiv -2i\eta\,\lambda^{-2-i\eta}|\Gamma(1+i\eta)|^2 \times\int_0^\pi \sin\theta\,d\theta\,P_l(\cos\theta)\,q^{-2+2i\eta} \times{}_2F_1(1-i\eta,1-i\eta;1;-\lambda^2/q^2). \quad (5b)$$

The integration over the angle θ in Eq. (5b) may be evaluated in closed form as follows. We first write the hypergeometric function ${}_2F_1$ appearing in (5b) in terms of its Mellin–Barnes integral representation[22]

$${}_2F_1(1-i\eta,1-i\eta;1;-\lambda^2/q^2) = \frac{1}{[\Gamma(1-i\eta)]^2}\frac{1}{2\pi i} \times\int_{C_1} ds\,\frac{[\Gamma(1-i\eta+s)]^2}{\Gamma(1+s)}\Gamma(-s)\left(\frac{\lambda^2}{q^2}\right)^s, \quad (6)$$

which is valid provided $|\arg(\lambda^2/q^2)|<\pi$. In Eq. (6) the contour C_1 runs from $-i\infty$ to $+i\infty$ and is chosen so that the poles of $\Gamma(1-i\eta+s)$ lie to the left of C_1 while the poles of $\Gamma(-s)$ lie to the right. Since at the poles of $\Gamma(1-i\eta+s)$ we have $\mathrm{Re}(s)\leqslant -1$ and at the poles of $\Gamma(-s)$ we have $\mathrm{Re}(s)\geqslant 0$, we may explicitly choose C_1 to be the straight line running from $-\epsilon-i\infty$ to $-\epsilon+i\infty$, where $0<\epsilon<1$. When the contour C_1 is closed at infinity in the right half plane, Eq. (6) yields the usual expansion[23] of the hypergeometric function in powers of $(-\lambda^2/q^2)$, whereas, if C_1 is closed at infinity in the left half plane, the double poles of $\Gamma(1+s-i\eta)$ lead to the usual well-defined analytic continuation[24] of the hypergeometric function in terms of $\ln(\lambda^2/q^2)$ and powers of $(-q^2/\lambda^2)$.

Now, using Eq. (6) in (5b) together with $q^2=2K_i^2(1-\cos\theta)$, we obtain

$$\mathscr{I}_l^{\rm G}(\lambda,K_i) = -2i\eta\,\lambda^{-2-2i\eta}\,(2K_i^2)^{i\eta-1}\frac{\Gamma(1+i\eta)}{\Gamma(1-i\eta)} \times\int_0^\pi \sin\theta\,d\theta\,P_l(\cos\theta)(1-\cos\theta)^{-1+i\eta} \times\frac{1}{2\pi i}\int_{-\epsilon-i\infty}^{-\epsilon+i\infty} ds\,\frac{[\Gamma(1-i\eta+s)]^2}{\Gamma(1+s)}\Gamma(-s) \times\left(\frac{\lambda^2}{2K_i^2}\right)^s(1-\cos\theta)^{-s}.$$

On the contour, $\mathrm{Re}(s)<0$; thus the singularity of $(1-\cos\theta)^{-1-s+i\eta}$ at $\theta=0$ is integrable and the θ integration is well defined for all values of s on the contour. Consequently the orders of integration may be interchanged so that

$$\mathscr{I}_l^{\rm G}(\lambda,K_i) = \frac{-i\eta}{2\pi i}\lambda^{-2}K_i^{-2}\left(\frac{\lambda^2}{2K_i^2}\right)^{-i\eta}\frac{\Gamma(1+i\eta)}{\Gamma(1-i\eta)} \times\int_{-\epsilon-i\infty}^{-\epsilon+i\infty} ds\,\frac{[\Gamma(1-i\eta+s)]^2}{\Gamma(1+s)}\Gamma(-s)\left(\frac{\lambda^2}{2K_i^2}\right)^s \times\int_0^\pi \sin\theta\,d\theta\,P_l(\cos\theta)(1-\cos\theta)^{i\eta-1-s}. \quad (7)$$

But

$$\int_0^\pi \sin\theta\,d\theta\,P_l(\cos\theta)(1-\cos\theta)^{i\eta-1-s} = \int_{-1}^1 dx\,P_l(x)(1-x)^{-1-s+i\eta} = (-1)^l\int_{-1}^1 dx\,P_l(x)(1+x)^{i\eta-1-s}$$

since $P_l(-x)=(-1)^lP_l(x)$. Moreover,[25]

$$\int_{-1}^1 dx(1+x)^\sigma P_\nu(x) = \frac{2^{\sigma+1}[\Gamma(1+\sigma)]^2}{\Gamma(\sigma+\nu+2)\Gamma(1+\sigma-\nu)},$$

provided $\mathrm{Re}(\sigma)>-1$. Therefore,

$$\int_0^\pi \sin\theta\,d\theta\,P_l(\cos\theta)(1-\cos\theta)^{i\eta-1-s} = \frac{(-1)^l2^{i\eta-s}[\Gamma(i\eta-s)]^2}{\Gamma(i\eta-s+l+1)\Gamma(i\eta-s-l)} \quad (8)$$

and

$$\mathscr{I}_l^{\rm G}(\lambda,K_i) = -i\eta\,\lambda^{-2}K_i^{-2}\left(\frac{\lambda^2}{4K_i^2}\right)^{-i\eta}\frac{\Gamma(1+i\eta)}{\Gamma(1-i\eta)} \times(-1)^l\frac{1}{2\pi i}\int_{-\epsilon-i\infty}^{-\epsilon+i\infty} ds\,\frac{[\Gamma(1-i\eta+s)]^2}{\Gamma(1+s)}\Gamma(-s) \times\frac{\Gamma(i\eta-s)\Gamma(i\eta-s)}{\Gamma(i\eta-s+l+1)\Gamma(i\eta-s-l)}\left(\frac{\lambda^2}{4K_i^2}\right)^s. \quad (9)$$

Since l is an integer, we employ the relation[26]

$$\frac{\Gamma(1+\alpha)}{\Gamma(1+\alpha-m)} = (-1)^m\frac{\Gamma(m-\alpha)}{\Gamma(-\alpha)},\quad m=1,2,3,..., \quad (10)$$

to infer

$$\frac{\Gamma(i\eta-s)\,\Gamma(i\eta-s)}{\Gamma(i\eta-s-l)\,\Gamma(i\eta-s+l+1)} = -\frac{\Gamma(l+1-i\eta+s)}{\Gamma(1-i\eta+s)}\frac{\Gamma(s-i\eta-l)}{\Gamma(1-i\eta+s)},$$

and thereby obtain

$$\mathscr{I}_l^{\rm G}(\lambda,K_i) = i\eta\,\lambda^{-2}K_i^{-2}\left(\frac{\lambda^2}{4K_i^2}\right)^{-i\eta}\frac{\Gamma(1+i\eta)}{\Gamma(1-i\eta)}(-1)^l \times\frac{1}{2\pi i}\int_{-\epsilon-i\infty}^{-\epsilon+i\infty} ds\,\frac{\Gamma(s-i\eta-l)\Gamma(s-i\eta+l+1)}{\Gamma(1+s)} \times\Gamma(-s)\left(\frac{\lambda^2}{4K_i^2}\right)^s. \quad (11)$$

Equation (11) is subject to the restriction that $|\arg(\lambda^2/$

$4K_i^2)| < \pi$, since $|\arg(\lambda^2/q^2)| < \pi$ in Eq. (6). Although the integrand of Eq. (11) has the proper form for identifying the integral as a Mellin–Barnes representation of a hypergeometric function, the integral cannot be equated with a hypergeometric function because not all of the poles of $\Gamma(s - i\eta - l)$ lie to the left of C_1 (see Ref. 22). In particular, there are $l+1$ poles of $\Gamma(s - i\eta - l)$ at values of $s = l + i\eta - j$ ($j = 0,1,\ldots, l$) to the right of C_1. However, if we let C_2 be a closed contour, which encircles only these $l+1$ poles of $\Gamma(s - i\eta - l)$ in a counterclockwise sense, then

$$\frac{1}{2\pi i}\int_{C_1} ds\cdots = \frac{1}{2\pi i}\int_{C_1+C_2} ds\cdots - \frac{1}{2\pi i}\int_{C_2} ds\cdots. \quad (12)$$

The integration over the contour $C_1 + C_2$ now may be identified with a hypergeometric function, while the integral over C_2 may be evaluated by applying the residue theorem. In particular,[22]

$$\begin{aligned}\frac{1}{2\pi i}\int_{C_1+C_2} ds\,&\frac{\Gamma(s-i\eta-l)\Gamma(s-i\eta+l+1)}{\Gamma(1+s)}\\ &\times\Gamma(-s)\,(\lambda^2/4K_i^2)^s\\ &=\Gamma(-l-i\eta)\Gamma(l+1-i\eta)\\ &\quad\times{}_2F_1(-l-i\eta, l+1-i\eta;1;-\lambda^2/4K_i^2). \quad (13)\end{aligned}$$

Since the only poles of the integrand within the closed contour C_2 are the aforementioned $l+1$ simple poles of $\Gamma(s-i\eta-l)$ we find that

$$\begin{aligned}\frac{1}{2\pi i}\int_{C_2} ds\,&\frac{\Gamma(s-i\eta-l)\Gamma(s-i\eta+l+1)}{\Gamma(1+s)}\\ &\times\Gamma(-s)\,(\lambda^2/4K_i^2)^s\\ &=\sum_{j=0}^{l}\frac{(-1)^j}{j!}\frac{\Gamma(2l+1-j)}{\Gamma(1+l+i\eta-j)}\Gamma(-l-i\eta+j)\\ &\quad\times(\lambda^2/4K_i^2)^{l+i\eta-j}. \quad (14a)\end{aligned}$$

Now let $m = -j$, then

$$\begin{aligned}\frac{1}{2\pi i}\int_{C_2} ds\cdots = \sum_{m=0}^{l}&\frac{(-1)^{l-m}\Gamma(l+1+m)}{\Gamma(l+1-m)\,\Gamma(1+i\eta+m)}\\ &\times\Gamma(-i\eta-m)\,(\lambda^2/4K_i^2)^{m+i\eta}. \quad (14b)\end{aligned}$$

We can use Eq. (10) to simplify (14b) considerably. Noting that

$$\begin{aligned}\frac{\Gamma(l+1+m)}{\Gamma(l+1-m)} &= \frac{\Gamma(l+1+m)}{\Gamma(l+1)}(-1)^m\frac{\Gamma(m-l)}{\Gamma(-l)}\\ &= (-1)^m(l+1)_m(-l)_m\end{aligned}$$

and

$$\begin{aligned}\Gamma(-i\eta-m) &= (-1)^m\Gamma(-i\eta)\frac{\Gamma(1+i\eta)}{\Gamma(1+i\eta+m)}\\ &= (-1)^m\frac{\Gamma(-i\eta)}{(1+i\eta)_m},\end{aligned}$$

where $(a)_m$ is Pochhammer's symbol,[27] we obtain

$$\begin{aligned}\frac{1}{2\pi i}\int_{C_2} ds\,&\frac{\Gamma(s-i\eta-l)\Gamma(s-i\eta+l+1)}{\Gamma(1+s)}\\ &\times\Gamma(-s)\left(\frac{\lambda^2}{4K_i^2}\right)^s\\ &= (-1)^l\left(\frac{\lambda^2}{4K_i^2}\right)^{i\eta}\frac{\Gamma(-i\eta)}{\Gamma(1+i\eta)}\\ &\quad\times\sum_{m=0}^{l}\frac{(l+1)_m(-l)_m}{[(1+i\eta)_m]^2}\left(\frac{-\lambda^2}{4K_i^2}\right)^m. \quad (14c)\end{aligned}$$

Finally employing the results of Eqs. (12)–(14c) in Eq. (11) we find, after some further minor manipulations with the gamma functions, that Eq. (11), and therefore Eq. (5b), can be reduced to the comparatively simple form

$$\begin{aligned}&\mathscr{I}_l^{\rm G}(\lambda, K_i)\\ &= \lambda^{-2}K_i^{-2}\left\{\sum_{m=0}^{l}\frac{(l+1)_m(-1)_m}{[(1+i\eta)_m]^2}\left(-\frac{\lambda^2}{4K_i^2}\right)^m\right.\\ &\quad-|\Gamma(1+i\eta)|^2\left(\frac{\lambda^2}{4K_i^2}\right)^{-i\eta}\frac{(1-i\eta)_l}{(1+i\eta)_l}\\ &\quad\left.\times{}_2F_1\left(-i\eta-l,-i\eta+l+1;1;-\frac{\lambda^2}{4K_i^2}\right)\right\}. \quad (15)\end{aligned}$$

Recalling Eq. (5a), it is necessary to differentiate $\mathscr{I}_l^{\rm G}(\lambda, K_i)$ once with respect to λ in order to obtain the elastic Glauber partial wave scattering amplitude; this is most easily accomplished by noting that if $z = \lambda^2/4K_i^2$ then $\partial/\partial\lambda = (\lambda/2K_i^2)(\partial/\partial z)$ and that[28]

$$\begin{aligned}&\frac{\partial}{\partial z}{}_2F_1(a,b;c;-z)\\ &\quad= -(ab/c)\,{}_2F_1(a+1,b+1;c+1;-z).\end{aligned}$$

In terms of the parameter z, we write

$$\begin{aligned}\mathscr{I}_l^{\rm G}(\lambda, K_i) = \frac{1}{4}K_i^{-4}z^{-1-i\eta}&\left\{\sum_{m=0}^{l}\frac{(l+1)_m(-l)_m}{[(1+i\eta)_m]^2}\right.\\ &\times(-1)^m z^{m+i\eta} - |\Gamma(1+i\eta)|^2\frac{(1-i\eta)_l}{(1+i\eta)_l}\\ &\left.\times{}_2F_1(-i\eta-l, 1+l-i\eta;1;-z)\right\}. \quad (16)\end{aligned}$$

After some elementary manipulations, we obtain from (5a) and (16) the desired result that

$$\begin{aligned}&\mathscr{F}_l^{\rm G}(1s\to 1s;K_i)\\ &=\frac{1}{2iK_i}\left\{-(1+i\eta)\right.\\ &\quad\times\left[\sum_{m=0}^{l}\frac{(l+1)_m(-l)_m}{[(1+i\eta)_m]^2}\left(-\frac{1}{K_i^2a_0^2}\right)^m\right.\\ &\quad-|\Gamma(1+i\eta)|^2\frac{(1-i\eta)_l}{(1+i\eta)_l}\left(\frac{1}{K_i^2a_0^2}\right)^{-i\eta}\\ &\quad\left.\times{}_2F_1\left(-i\eta-l,1+l-i\eta;1;-\frac{1}{K_i^2a_0^2}\right)\right]\\ &\quad+i\eta\left[\sum_{m=0}^{l}\frac{(l+1)_m(-l)_m}{(i\eta)_m(1+i\eta)_m}\left(-\frac{1}{K_i^2a_0^2}\right)^m\right.\\ &\quad-|\Gamma(1+i\eta)|^2\frac{(1-i\eta)_{l+1}}{(i\eta)_l}\left(\frac{1}{K_i^2a_0^2}\right)^{1-i\eta}\\ &\quad\left.\left.\times{}_2F_1\left(-i\eta-l+1,-i\eta+l+2;2;-\frac{1}{K_i^2a_0^2}\right)\right]\right\}. \quad (17)\end{aligned}$$

The first Born elastic partial wave amplitudes may be obtained in a fashion very similar to the foregoing reduction of the elastic Glauber amplitude. One can easily show[2,9] that the fundamental relation between the Born and corresponding Glauber amplitude is

$$F^{\mathrm{B}}(i\rightarrow f;\mathbf{q}) = i\eta \lim_{\eta\rightarrow 0} \{(1/i\eta)\, F^{\mathrm{G}}(i\rightarrow f;\mathbf{q})\} \tag{18a}$$

for fixed K_i and each momentum transfer $\mathbf{q}$. A similar relation holds for the corresponding partial wave elastic scattering amplitudes; at fixed K_i,

$$\mathscr{F}_l^{\mathrm{B}}(1s\rightarrow 1s; K_i) = i\eta \lim_{\eta\rightarrow 0} \{(1/i\eta)\, \mathscr{F}_l^{\mathrm{G}}(1s\rightarrow 1s; K_i\}\,. \tag{18b}$$

Applying the relation (18b) to Eq. (17) is *not* straightforward, but rather, unduly tedious. However, Eq. (18b) can be easily applied to Eqs. (5) to give

$$\mathscr{F}_l^{\mathrm{B}}(1s\rightarrow 1s; K_i) = K_i\,\eta\left(\frac{2}{a_0}\right)^3 \frac{1}{4}\frac{\partial}{\partial\lambda}\mathscr{I}_l^{\mathrm{B}}(\lambda, K_i)\bigg|_{\lambda=2/a_0}, \tag{19a}$$

where

$$\mathscr{I}_l^{\mathrm{B}}(\lambda, K_i) = \lim_{\eta\rightarrow 0}\{(1/i\eta)\,\mathscr{I}_l^{\mathrm{G}}(\lambda, K_i)\}\,. \tag{19b}$$

Applying the relation (19b) to Eq. (11) yields an integral representation for the generating function $\mathscr{I}_l^{\mathrm{B}}$, which may be evaluated easily after some minor manipulations. We find that

$$\mathscr{I}_l^{\mathrm{B}}(\lambda, K_i) = -4\lambda^{-4}\left(\frac{4K_i^2}{\lambda^2}\right)^l \frac{\Gamma(l+1)\Gamma(l+1)}{\Gamma(2l+2)} \times {}_2F_1(l+1, l+1; 2l+2; -4K_i^2/\lambda^2). \tag{20}$$

The hypergeometric function in (20) is related to the Legendre function $Q_l(z)$ of the second kind,[29] with $z = 1+\lambda^2/2K_i^2$. Thus from (19a) and (20) it can be shown that[30]

$$\mathscr{F}_l^{\mathrm{B}}(1s\rightarrow 1s; K_i) = -\frac{1}{2}(K_i\,\eta)K_i^{-2}\lambda^3\frac{\partial}{\partial\lambda} \times\left[\lambda^{-2}Q_l\left(1+\frac{\lambda^2}{2K_i^2}\right)\right]\bigg|_{\lambda=2/a_0} \tag{21a}$$

$$= (K_i\,\eta)\,K_i^{-2}\left\{\left(1+\frac{z(l+1)}{1+z}\right)Q_l(z) - \frac{l+1}{1+z}Q_{l+1}(z)\right\}, \tag{21b}$$

where $z = 1+2/K_i^2a_0^2$. It should be noted that Eq. (21b) for $\mathscr{F}_l^{\mathrm{B}}$ can be shown to be equivalent to the seemingly different expression given by Mott and Massey.[31] Moreover, since[30]

$$(l+1)z\,Q_l(z) - (l+1)Q_{l+1}(z) = -(z^2-1)^{1/2}Q_l^1(z)$$

and[32] $Q_l(z)>0$, while $Q_l^1(z)<0$ when $z>1$, the term in braces on the right side of (21b) is purely positive for $0<K_i<\infty$. Thus $\mathscr{F}_l^{\mathrm{B}}>0$ if $\eta>0$ (i.e., $Z_i<0$), and $\mathscr{F}_l^{\mathrm{B}}<0$ if $\eta<0$ (i.e., $Z_i>0$).

III. EFFECTIVE INTERACTION AT LARGE *r*

In this section we obtain the large-r effective interaction of the Glauber approximation to the elastic scattering of charged particles from hydrogen atoms. By the effective interaction—which we denote by V_{eff}—we generally mean that interaction such that, at each incident energy, the elastic scattering of the incident particle is described exactly by the scattering solutions to the single particle Schrödinger equation with $V = V_{\mathrm{eff}}$. Of course, the exact V_{eff} for the scattering is $V_{\mathrm{eff}} = V_{\mathrm{opt}}$, where V_{opt} is the exact optical potential. To obtain the Glauber effective interaction, $V_{\mathrm{eff}}^{\mathrm{G}}$, at large distances r, we exploit the relation between the large-l partial wave amplitudes for scattering by a potential and the first Born approximate to those amplitudes.

As is well known, the Born partial wave amplitudes for scattering by a potential $V(r)$ may be written[33]

$$\mathscr{F}_l^{\mathrm{B}} = -\frac{\pi\mu}{\hbar^2K_i}\int_0^\infty V(r)\,[\,J_{l+1/2}(K_i\,r)]^2 r\,dr. \tag{22}$$

If $V(r) = Cr^{-\alpha}$, then,[34] provided $2l+2>\mathrm{Re}(\alpha-1)>0$,

$$\mathscr{F}_l^{\mathrm{B}} = -\frac{\mu\pi C}{\hbar^2K_i}\left(\frac{K_i}{2}\right)^{\alpha-2}\frac{1}{2}\,\frac{\Gamma(\alpha-1)\Gamma(l+\frac{3}{2}-\alpha/2)}{[\Gamma(\alpha/2)]^2\Gamma(l+\frac{1}{2}+\alpha/2)}. \tag{23}$$

In particular, for $\alpha = 3$ we obtain

$$\mathscr{F}_l^{\mathrm{B}} = -\mu C/\hbar^2 l(l+1)\,. \tag{24}$$

For large l, we see from Eq. (22) that most of the contribution to the integral will come from large r, i.e., from the asymptotic form of $V(r)$. Hence, if $V(r)\sim C/r^3$ for large r, then

$$\mathscr{F}_l^{\mathrm{B}} \sim -\mu C/\hbar^2 l^2, \tag{25}$$

for large l. Furthermore for such a potential, for large l the exact partial wave amplitude $\mathscr{F}_l$ and the Born partial wave amplitude $\mathscr{F}_l^{\mathrm{B}}$ approach each other:

$$\mathscr{F}_l \approx \mathscr{F}_l^{\mathrm{B}} \sim -\mu C/\hbar^2 l^2. \tag{26}$$

But in Appendix A we obtain the large-l asymptotic behavior of the Glauber partial wave amplitudes. From relation (A11), $\mathscr{F}_l^{\mathrm{G}}$ behaves, to leading order, like

$$\mathscr{F}_l^{\mathrm{G}} \sim -(1/iK_i)\,[(K_i\eta)^2a_0^2/l^2]. \tag{27}$$

The l dependence of Eq. (27) is identical to that of Eq. (26). We may now ask for the asymptotic behavior of the effective interaction $V_{\mathrm{eff}}^{\mathrm{G}}(r)$ such that the *exact* potential scattering solution to the Schrödinger equation yields the same large-l partial wave amplitudes as obtained in the Glauber approximation. From our discussion above we see that $V_{\mathrm{eff}}^{\mathrm{G}}(r) \sim C/r^3$, where the constant C is obtained by equating the right-hand sides of (26) and (27), so that

$$C = -\frac{i\hbar^2K_i\eta^2a_0^2}{\mu} = -i\frac{Z_i^2e^4}{\hbar v_i}a_0^2.$$

Thus, for large r,

$$V_{\mathrm{eff}}^{\mathrm{G}} \sim -i\frac{Z_i^2e^4a_0^2}{\hbar v_i}\frac{1}{r^3}. \tag{28}$$

Note that this is a purely absorptive, energy-dependent, effective interaction regardless of the charge of the incident particle. Equation (28) is to be contrasted with the large-r effective interaction corresponding to the amplitudes $\mathscr{F}_l^{\rm B}$ of Eq. (21b), namely,

$$V_{\rm eff}^{\rm B}(r) \sim Z_i e^2 (1/r + 1/a_0) e^{-2r/a_0},$$

which is real, energy dependent, and exponentially decreasing.

The exact large-r asymptotic interaction for the elastic scattering of a charged particle by atomic hydrogen has been examined classically[35] and from the optical potential.[36] To leading order,

$$V_{\rm eff}(r) \sim -\tfrac{1}{2}\alpha_{\rm H} Z_i^2 e^4/r^4, \tag{29}$$

where $\alpha_{\rm H}$ is the polarizability of the hydrogen atom target and $V_{\rm eff}$ is purely real. Comparing Eqs. (28) and (29), we see that $V_{\rm eff}^{\rm G}$ is unphysical on two accounts: it is absorptive, and of order r^{-3} at large r. It is noteworthy that the dominant unphysical behavior of $V_{\rm eff}^{\rm G}$ as $r \to \infty$ can be traced to the Glauber approximate to the second Born approximation. This is easily seen by expanding $e^{i\chi}$ in Eq. (1a) in powers of $\eta = -Z_i e^2/\hbar v_i$; this corresponds to expanding $F^{\rm G}$ itself (or $\mathscr{F}_l^{\rm G}$) in powers of η. It is well known[2] that the Glauber approximates to the nth term in the Born series for the direct scattering amplitude is given at fixed K_i by the term proportional to η^n in the expansion of $F^{\rm G}$ (or $\mathscr{F}_l^{\rm G}$). Since (27) is proportioned to η^2 it follows that (28) is due to the Glauber approximate to the second Born term in the direct scattering amplitude. Similarly, it is known[14,36] that Eq. (29) stems from the second Born term in the direct Born series for the optical potential $V_{\rm opt}$. Thus the failure of $V_{\rm eff}^{\rm G}$ corresponds to the previously noted failure[6,9] of the Glauber approximate to $F_2^{\rm B}$ stemming from the use of closure in the derivation[2] of the Glauber approximation. We show this explicitly in Sec. IV.

We may employ the arguments of this section and the higher-order terms in l of Eq. (A11) to examine asymptotic corrections to Eq. (28). In particular, we examine the corrections stemming from the Glauber approximates to the second and third Born terms in the direct amplitude Born series. From the discussion of the previous paragraph and (A11), we see that at large l the Glauber partial wave approximate $\mathscr{F}_{l2}^{\rm G}$ to the second Born partial waves are, to order $(l+1)^{-4}$,

$$\mathscr{F}_{l2}^{\rm G} \sim i K_i \eta^2 \frac{a_0^2}{(l+1)^2}\left\{1 + \frac{1}{l+1} + \frac{1}{(l+1)^2} \times\left[\frac{3}{2}K_i^2 a_0^2 + 1\right] + \cdots\right\}. \tag{30}$$

Again employing Eq. (23) to each order of l in (30) we find that, at fixed K_i, the effective interaction $V_{\rm eff}^{\rm G2}$ in the Glauber approximate $F_2^{\rm G}$ to the second Born term is asymptotically

$$V_{\rm eff}^{\rm G2} \sim -i\frac{Z_i^2 e^4 a_0^2}{\hbar v_i}\left\{\frac{1}{r^3} + \frac{a_1}{r^4} + \frac{a_2}{r^5}\right\}, \tag{31}$$

where a_1 and a_2 are purely real. Thus the $V_{\rm eff}^{\rm G2}$ does contain terms of order r^{-4}, but the coefficient is again imaginary. Moreover, $V_{\rm eff}^{\rm G2}$ contains terms proportional to r^{-5}. But it is known[14,35,36] that the exact asymptotic $V_{\rm eff}$ is

$$V_{\rm eff}(r) \sim -\frac{1}{2}\frac{\alpha_{\rm H} Z_i^2 e^4}{r^4} + O(r^{-6}), \tag{32}$$

where the terms of order r^{-6} in (32) stem, again, from the second Born term in $V_{\rm opt}$. Thus, the unphysical nature of $V_{\rm eff}^{\rm G2}$ is manifest not only in the imaginary coefficient, but also in the presence of terms of order r^{-3} and r^{-5}. We note that the imaginary coefficient of $V_{\rm eff}^{\rm G2}$ reflects the fact that the Glauber approximate to the second Born term in the amplitude is purely imaginary.[2,6,7,9] We also point out that $V_{\rm eff}(r)$ generally must contain absorptive terms to represent the removal of flux from the direct elastic channel. In e^-–H(1s) collisions, these absorptive terms must be present even below the first excitation threshold to allow for exchange effects. However, the absorptive terms in $V_{\rm eff}$ are generally expected to be short range.

In a similar fashion, we also may discuss the effective interaction due to the Glauber approximate to the third Born term. From (A11) we find that to leading order in large l, the Glauber partial wave approximate $\mathscr{F}_{l3}^{\rm G}$ (proportional to η^3) to the third Born partial wave is

$$\mathscr{F}_{l3}^{\rm G} \sim 3 K_i^3 a_0^4 \eta^3 \{1/l^4 + O(1/l^5)\}. \tag{33}$$

Thus, the effective interaction $V_{\rm eff}^{\rm G3}$ due to the Glauber approximate to the third Born term is asymptotically

$$V_{\rm eff}^{\rm G3} \sim \frac{9}{2}\frac{(Z_i e^2)^4}{(\hbar v_i)^3}\frac{1}{r^5} + O\left(\frac{1}{r^6}\right). \tag{34}$$

Although $V_{\rm eff}^{\rm G3}$ has the desired property of being real, we see from Eq. (32) that the exact large-r asymptotic form for $V_{\rm eff}(r)$ contains no r^{-5} dependence. Consequently, the continued use of $F_3^{\rm G}$ to approximate the exact third Born term in elastic scattering applications of the eikonal Born-series approximation[6,7] and in the modified Glauber approximation of Gien[15] is probably somewhat suspect, particularly at small scattering angles. Moreover, we suspect that the Glauber approximate to each term in the exact e^-–H(1s) elastic scattering Born series is correspondingly unphysical.

Finally we note that the amplitude formula of the two-potential eikonal approximation of Ishihara and Chen[5] can be written as a finite sum of partial waves plus an integral of the form of Eq. (1a) with a phase shift function χ that differs from (1b) at small values of b. At large values of b, however, the two-potential eikonal χ reduces to (1b). Recalling that (1a) can be derived by exploiting the correspondence $(l+\frac{1}{2})/K_i \to b$ at fixed K_i (see Ref. 2, Sec. 2.1.5; and Ref. 5), we see immediately that at large l, the behavior of $\mathscr{F}_l^{\rm G}$ and of the corresponding large-l partial wave amplitudes of the two-potential eikonal approximation is determined for both approximations by the behavior of the integrand in (1a) at large b. Thus at large l, the two-potential eikonal partial wave amplitudes behave like $\mathscr{F}_l^{\rm G}$. Consequently, the conclusions of this section, stemming as they do from the large-l behavior of $\mathscr{F}_l^{\rm G}$, also apply to the two-potential eikonal approximation. We suspect these conclusions also apply to the Glauber-angle approximation as well.[13]

IV. THE GLAUBER AMPLITUDE NEAR THE FORWARD DIRECTION

In this section we investigate the divergence of $F^{\rm G}(q)$ as $q \to 0$ at fixed K_i (i.e., as the scattering angle θ approaches zero). Although the amplitude is known[2,6,9] to diverge logarithmically, there seems to be some confusion regarding the source of this divergence.

It is well known[37] that potentials which fall off as r^{-3} at large r lead to divergent elastic scattering amplitudes near the forward direction. As was shown in Sec. III, the asymptotic form of the Glauber amplitude $\mathscr{F}_l^{G}$ for large l implies an effective potential proportional to r^{-3} for large r. Consequently it is this large-r dependence of the effective potential or, equivalently, the large-l dependence of the partial wave amplitude that leads to the divergence of the full scattering amplitude F^{G} as $\theta \to 0$.

To see this in more detail, it is instructive to investigate the scattering amplitude F_G near the forward direction both by means of Eq. (2a) and by means of the sum given by Eq. (3) of the partial wave amplitudes $\mathscr{F}_l^{G}$. By explicitly performing the indicated differentiation in Eq. (2a), we may express the full elastic scattering amplitude as

$$\begin{aligned} F^{G}(q) = {} & 2K_i\,\eta|\Gamma(1+i\eta)|^2 q^{-2}\xi^{-i\eta} \\ & \times[(1+i\eta)\,{}_2F_1(1-i\eta,1-i\eta;1;-\xi) \\ & +(1-i\eta)^2\xi\,{}_2F_1(2-i\eta,2-i\eta;2;-\xi)], \end{aligned} \tag{35}$$

where $\xi \equiv 4/q^2a_0^2 = [K_ia_0\sin(\theta/2)]^{-2}$. As $\theta \to 0$, we have $\xi \to \infty$. Now as $\xi \to \infty$ we have,[24] for fixed η (i.e., fixed K_i),

$$\begin{aligned} & {}_2F_1(1-i\eta,1-i\eta;1;-\xi) \\ & \quad \sim \frac{1}{\Gamma(1-i\eta)\Gamma(i\eta)}\xi^{-1+i\eta} \\ & \quad \times[\ln\xi+2\psi(1)-\psi(1-i\eta)-\psi(i\eta)], \end{aligned} \tag{36a}$$

$$\begin{aligned} & {}_2F_1(2-i\eta,2-i\eta;2;-\xi) \\ & \quad \sim \frac{1}{\Gamma(2-i\eta)\Gamma(i\eta)}\xi^{-2+i\eta} \\ & \quad \times[\ln\xi+2\psi(1)-\psi(2-i\eta)-\psi(i\eta)], \end{aligned} \tag{36b}$$

where[38] $\psi(z) = d[\ln\Gamma(z)]/dz$.

Using these asymptotic results in Eq. (35), we obtain, for fixed K_i and η,

$$\begin{aligned} F^{G}(q) & \sim iK_ia_0^2\,\eta^2[-2\ln(\sin\theta/2)-2\ln(K_ia_0) \\ & \quad -\tfrac{1}{2}+2\psi(1)-\psi(1-i\eta)-\psi(i\eta)] \qquad (37) \\ & \sim iK_ia_0^2\,\eta^2[-2\ln(\sin\theta/2)-2\ln(K_ia_0) \\ & \quad -\tfrac{1}{2}-2\gamma-2\,\mathrm{Re}[\psi(1-i\eta)]+(i\eta)^{-1}], \end{aligned} \tag{38}$$

where (38) is obtained from (37) via the properties of the ψ functions.[38] The logarithmic divergence of F^{G} as $\theta \to 0$ is made explicit in either (37) or (38).

Let us next investigate F^{G} in terms of the sum over partial waves. From Eq. (A11), for large l and $l+1 \gg K_ia_0$

$$\mathscr{F}_l^{G}(K_i) \sim iK_ia_0^2\,\eta^2(l+1)^{-2}[1+(l+1)^{-1}] \equiv \tilde{\mathscr{F}}_l^{G}. \tag{39}$$

We may express the full Glauber amplitude as

$$\begin{aligned} F^{G}(q) = {} & iK_ia_0^2\,\eta^2\sum_{l=0}^{\infty}\frac{2l+1}{(l+1)^2}\left[1+\frac{1}{l+1}\right]P_l(\cos\theta) \\ & +\sum_{l=0}^{L}(2l+1)P_l(\cos\theta)[\mathscr{F}_l^{G}-\tilde{\mathscr{F}}_l^{G}] \\ & +\sum_{l=L+1}^{\infty}(2l+1)[\mathscr{F}_l^{G}-\tilde{\mathscr{F}}_l^{G}]\,P_l(\cos\theta). \end{aligned} \tag{40}$$

By choosing L appropriately we may make the last sum in Eq. (40) as small in magnitude as desired. Consequently we shall omit that sum hereafter. Now

$$\frac{2l+1}{(l+1)^2}\left(1+\frac{1}{l+1}\right) = \frac{2}{l+1}+\frac{1}{l(l+1)}-\frac{2l+1}{(l+1)^3 l}.$$

Also[39]

$$\sum_{l=0}^{\infty}\frac{1}{l+1}P_l(\cos\theta) = \ln\left(\frac{1+\sin\theta/2}{\sin\theta/2}\right) \tag{41}$$

and

$$\sum_{l=1}^{\infty}\frac{1}{l(l+1)}P_l(\cos\theta) = 1-2\ln\left(1+\sqrt{\frac{1-\cos\theta}{2}}\right). \tag{42}$$

Therefore

$$\begin{aligned} & \sum_{l=0}^{\infty}\frac{2l+1}{(l+1)^2}\left(1+\frac{1}{l+1}\right)P_l(\cos\theta) \\ & \quad = 2\ln\left(\frac{1+\sin\theta/2}{\sin\theta/2}\right)-2\ln(1+\sin\theta/2)+1 \\ & \qquad -\sum_{l=1}^{\infty}\frac{2l+1}{l(l+1)^3}P_l(\cos\theta). \end{aligned} \tag{43}$$

Consequently, by choosing L appropriately we find that, to any desired accuracy,

$$\begin{aligned} F^{G}(q) = {} & -2iK_i\,a_0^2\,\eta^2\ln(\sin\theta/2) \\ & +\mathscr{F}_0^{G}-\tilde{\mathscr{F}}_0^{G}+iK_i\,a_0^2\,\eta^2 \\ & +\sum_{l=1}^{L}(2l+1)\left\{\mathscr{F}_l^{G}-\tilde{\mathscr{F}}_l^{G}\right. \\ & \left.-iK_i\,a_0^2\,\eta^2\frac{1}{l(l+1)^3}\right\}P_l(\cos\theta). \end{aligned} \tag{44}$$

As $\theta \to 0$ we obtain

$$F^{G}(q) \sim -2iK_i\,a_0^2\,\eta^2\ln(\sin\theta/2) \tag{45}$$

in agreement with Eq. (38). The logarithmic divergence arises from the sum (41) and in particular stems from the large-l behavior of the summand, i.e., from the large-l behavior of the partial wave amplitude. We also point out that (45) holds at small θ even when $v_i \to 0$ (i.e., $K_i \to 0$ at fixed μ), despite the apparent $\ln(K_i)$ divergence in (37) and (38). This can be seen by applying to Eq. (35) the methods of Appendix B, wherein the $v_i \to 0$ behavior of $\mathscr{F}_l^{G}$ is obtained.

As we remarked in Sec. III, the large-l asymptotic behavior of $\mathscr{F}_l^{G}$ is due to the Glauber approximate to the 2nd term in the Born series for the direct scattering amplitude. Thus the results of this section are consistent with, and serve to clarify, previous discussions[2,6,9] of the divergence of F^{G} in the forward direction. Moreover, in Sec. III we found that the partial wave amplitudes of the two-potential eikonal approximation behave at large l like $\mathscr{F}_l^G$. Since the large-l behavior of $\mathscr{F}_l^G$ determines the $\ln(\sin\theta/2)$ divergence of F^G, the full two-potential eikonal amplitude must similarly diverge as $\theta \to 0$.

V. SOME NUMERICAL RESULTS FOR THE PARTIAL WAVE AMPLITUDES

In this section we present some results of numerical computation of the e^-–H(1s) elastic Glauber partial wave amplitudes $\mathscr{F}_l^G(K_i)$, both as a function of $l+1$ for fixed K_i and as a function of K_i for fixed l.

In Fig. 1 we show $K_i\,\mathscr{F}_l^G(K_i)$ as a function of $l+1$ for $K_i\,a_0 = \frac{1}{2}$ and for $K_i\,a_0 = 2$ and $1 \leqslant l+1 \leqslant 20$. The solid and broken curves connect the discrete points (dots), serving merely as a guide to the eye. For the case $K_i\,a_0 = \frac{1}{2}$, the asymptotic expansion (A11) is expected to be valid for $l+1 \gg \frac{1}{2}$ and $l \gg 1$. We see from the figure that the asymptotic behavior (straight line portion) occurs for $l+1 \gtrsim 5$, which is consistent with $l+1 \gg \frac{1}{2}$ and $l \gg 1$. For $K_i\,a_0 = 2$, the conditions for the validity of asymptotic expansion are $l+1 \gg 2$ and $l \gg 1$. From Fig. 1 we see that the asymptotic behavior occurs for $l+1 \gtrsim 10$, which is consistent with $l+1 \gg 2$ and $l \gg 1$. This figure also reveals that the imaginary part of $\mathscr{F}_l^G$ behaves asymptotically like $(l+1)^{-2}$ whereas the real part behaves like $(l+1)^{-4}$, as can be inferred also from Eq. (A11).

In Fig. 2 we show $K_i\,\mathscr{F}_l^G(K_i)$ as a function of $l+1$ for $K_i\,a_0 = 5$ and for $K_i\,a_0 = 10$ and $1 \leqslant l+1 \leqslant 20$. For $l+1 = 20$, $(l+1)/K_i\,a_0$ is 4 and 2 for the two cases shown.

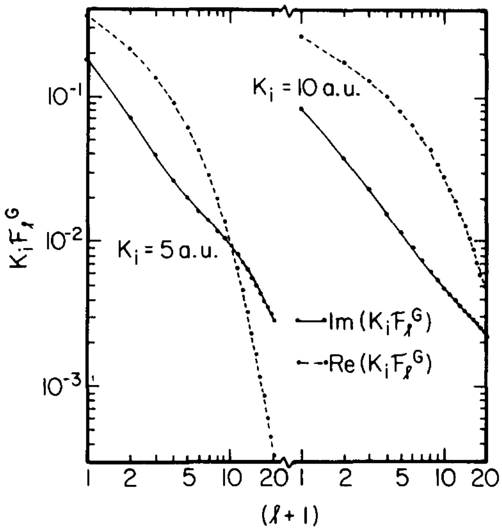

FIG. 2. Same as Fig. 1 for $K_i\,a_0 = 5$ and 10.

These values are not large enough for the asymptotic expansion to be valid, as can be readily seen from the figure.

In Fig. 3 we present $K_i\,\mathscr{F}_l^G(K_i)$ for $l = 0$ and $l = 1$ in Argand diagrams. For $l = 0$ the threshold value of $K_i\,\mathscr{F}_l^G$ is $0.474i$ and for $l = 1$ it is $0.246i$. The results for small K_i were obtained from the small-v_i asymptotic form of $\mathscr{F}_l^G$ [Eq. (B9)], derived in Appendix B. The numbers with the corresponding arrows denote the incident momentum K_i in atomic units. We note that at all energies $K_i\,\mathscr{F}_l^G$ lies within the unitary circle. This is to be contrasted with the Born approximation in which $K_i\,\mathscr{F}_l^G$ lies outside the unitary circle for finite K_i. We also note that at zero energy the partial wave

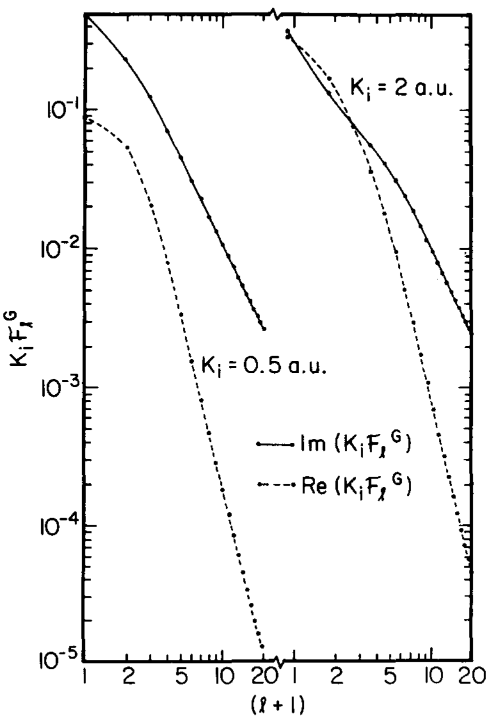

FIG. 1. Real and imaginary parts of Glauber approximation partial wave amplitudes for elastic scattering of electrons by ground state hydrogen, for incident momenta $K_i\,a_0 = \frac{1}{2}$ and 2. The curves connecting the dots serve merely to guide the eye.

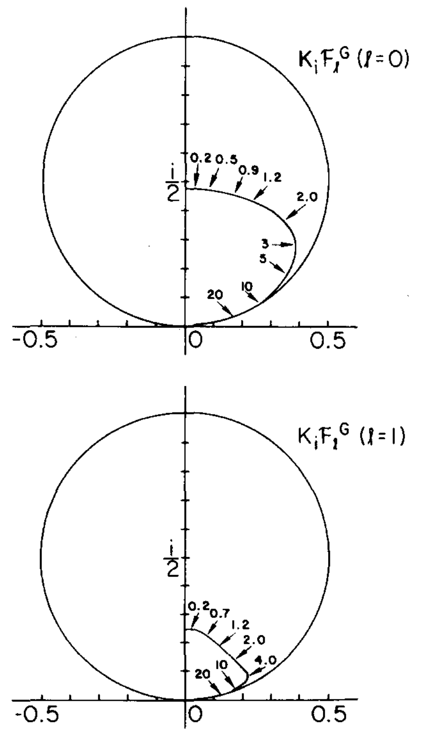

FIG. 3. Argand diagrams for $K_i\,\mathscr{F}_l^G(K_i)$ for $l = 0$ and $l = 1$. The threshold values are $0.474i$ for $l = 0$ and $0.246i$ for $l = 1$. The arrows point to the position of $K_i\,\mathscr{F}_l^G$ for the corresponding momenta, which are given in atomic units.

amplitudes are purely imaginary, indicating an absorptive interaction. Recall from Sec. III that the large-r (large-l) Glauber interaction was also absorptive. The clockwise direction with increasing K_i, of the Argand diagram, indicates that the real part of the Glauber effective potential interaction is attractive for incident electrons.

We point out that for incident positrons, the Glauber amplitude is given at fixed K_i and fixed q by

$$F^{G}\,(\text{positrons}) = -\{F^{G}\,(\text{electrons})\}^{*}.$$

Thus the numerical results of Figs. 1 and 2 hold for incident positrons provided the curves for $\mathrm{Re}(K_i\,\mathscr{F}_l^{G})$ are multiplied by -1. Moreover, the positron curves corresponding to Fig. 3 are obtained simply by reflecting the plotted curves about the imaginary axis. Although we have yet to perform numerical calculations of $\mathscr{F}_l^{G}$ for other than incident electrons (or positrons), some results for incident protons may be inferred from the results of other sections of this paper. In particular, in Appendix B we obtain the $v_i \rightarrow 0$ behavior of $\mathscr{F}_l^{G}$ at fixed l. In the case where the reduced mass μ, or the charge Z_i, is large, we have, from Eq. (B10),

$$\mathscr{F}_l^{G} \sim \frac{i}{K_i}\left[1-(l-1)_4\left(\frac{\hbar^2}{|Z_i|e^2\mu a_0}\right)^4+\cdots\right].$$

Thus for incident protons, the Argand plots of $K_i\,\mathscr{F}_l^{G}$ terminate at $K_i = 0$ on the imaginary axis at $i/2$.

It is known[2] that the Glauber approximation in purely potential scattering satisfies the optical theorem at high energies, and thus satisfies in some sense the constraints imposed by unitarity. In composite atomic collisions applications, and in particular for the elastic scattering of charged particles by hydrogen atoms and helium atoms, however, the Glauber approximation fails to satisfy the optical theorem because the imaginary part of the elastic scattering amplitude diverges in the forward direction, whereas measured total cross sections are found to be finite. On the other hand, the Argand plots of $K_i\,\mathscr{F}_l^{G}$ shown in Fig. 3 for $l=0$ and $l=1$, clearly show that for e^-–H, or e^+–H elastic scattering, $K_i\,\mathscr{F}_l^{G}$ lies within, or on, the unitary circle. Although we do not present Argand plots for values of $l \geqslant 2$, the numerical calculations of $\mathscr{F}_l^{G}$ for $l \leqslant 20$ indicate that at each fixed l, $K_i\,\mathscr{F}_l^{G}$ lies within the unitary circle for $K_i\,a_0 \leqslant 20$. Generally we believe, but have not yet proven, that $K_i\,\mathscr{F}_l^{G}$ lies within the unitary circle for *all* fixed values of l, not only for $e^{\mp}$–H elastic scattering, but also for the elastic scattering of any structureless charged particle by a hydrogen atom. This conclusion is, in part, supported by the asymptotic behavior of $\mathscr{F}_l^{G}$ as $v_i \rightarrow 0$, as $\mu \rightarrow \infty$, and as $|Z_i| \rightarrow \infty$ which we discuss in Appendix B. In contradistinction to behavior of the Glauber $\mathscr{F}_l^{G}$, the conventional Born partial amplitudes $\mathscr{F}_l^{B}$ are found to diverge as $\mu \rightarrow \infty$ or $|Z_i| \rightarrow \infty$ at fixed K_i. More generally, we believe, but cannot yet prove, that $K_i\,\mathscr{F}_l^{G}$ lies within the unitary circle when the Glauber approximation is applied to the elastic scattering of structureless charged particles by arbitrary neutral atoms or molecules. This is to be contrasted with first Born applications to the elastic scattering from many-electron atoms where it is known[40] that the small-l Born partial wave amplitudes violate the unitary bounds[41] on $|\mathscr{F}_l^{B}|^2$ at low energies.

VI. CONCLUSION

In this paper we have shown that the $\theta \rightarrow 0$ divergence of the Glauber charged particle–hydrogen atom elastic scattering amplitude stems directly from the large-l behavior of the Glauber partial wave amplitudes $\mathscr{F}^{G}$. Furthermore this large-l behavior is associated with a purely absorptive, r^{-3} effective potential in the Glauber direct elastic scattering channel, and clearly stems from the Glauber approximate to the second term in the Born series for the direct amplitude. It has been previously remarked that the $\ln(\sin\theta/2)$ divergence, and hence the absorptive Glauber effective interaction, can be traced[2,6,9] to the use of the closure approximation in the derivation[2] of the Glauber amplitude formula for composite collisions. In this derivation, closure is employed, on shell in the direct elastic channel, to eliminate the infinite sum over all states of the target hydrogen atom from the direct channel asymptotic propagator.[2] Thus, even though the approximate Glauber solution to the three-particle Schrödinger equation no longer contains terms that explicitly depend upon excited states of the target atom, the use of closure implicitly admits an expansion of the Glauber scattering solution in *all* states of the target, wherein propagation in *all* channels corresponding to excited states of the target is allowed, even below the first excitation threshold. From Eq. (28) it is clear that the absorptive V_{eff}^{G} diverges as $K_i \rightarrow 0$. In some sense this divergence of V_{eff}^{G} reflects the possibility of propagation in all target channels, both discrete and continuous, at $K_i = 0$. As remarked in Sec. III, the exact direct channel effective interaction must contain absorptive terms in e^-–H collisions, if only to account for the removal of direct channel flux into the exchange channel, even below the first excitation threshold. Since all target channels are open in the Glauber approximation, it is tempting to assert that V_{eff}^{G} crudely (though assuredly incorrectly) accounts for these effects via the use of closure in the elastic channel of the incident particle. Finally, the availability of the analytic expressions for the Glauber elastic partial amplitudes for charged particle–hydrogen atom collisions should enable a more formal examination of the properties of the Glauber amplitudes; such studies will, perhaps, lead to a better understanding of the successes and failures of the approximation.

We have not fully exploited the formal analytic properties of the amplitudes $\mathscr{F}_l^{G}$ in this present paper. However, we have made some studies of these properties. In particular, we have in part examined the continuation of $\mathscr{F}_l^{G}$ of Eq. (17) onto the physical energy sheet defined by $0 < \arg(K_i^2) < 2\pi$. As it stands, Eq. (17) defines $\mathscr{F}_l^{G}$ for $-\pi < \arg(K_i^2) < \pi$. The analytic continuation of (17) into the region $\pi \leqslant \arg(K_i^2) < 2\pi$ (the lower half physical sheet) is necessarily complicated by the K_i dependence of η. Nevertheless it is clear from (17) that the physical cut $0 \leqslant K_i^2 < \infty$ is required to make $(K_i^2 a_0^2)^{i\eta}$ single valued. Preliminary results for the continuation of $\mathscr{F}_l^{G}$ onto the negative real axis and lower physical sheet indicate the possible need for a branch cut running from $-\infty < K_i^2 \leqslant -1$, although this result is not assured. However, for incident particles with $Z_i < 0$, these same preliminary results indicate that at fixed l, $\mathscr{F}_l^{G}$ has discrete simple poles on the negative real axis, stemming from the factors $\Gamma(1-i\eta)(1-i\eta)_l = \Gamma(1+l-i\eta)$ in Eq. (17), at energies E_n given by

$$E_n = -Z_i^2 e^4 \mu / 2\hbar^2 n^2,$$

where $n = l + 1 + r = 0, 1, 2,\ldots$ at each l.

ACKNOWLEDGMENTS

This paper is dedicated to the memory of Brian K. Thomas. This work was supported in part by the National Aeronautics and Space Administration and by the PSC–BHE Research Award Program of the City University of New York.

APPENDIX A: LARGE-l ASYMPTOTIC BEHAVIOR OF $\mathscr{F}_l^G$ and $\mathscr{F}_l^B$ at fixed K_i

In this part of this Appendix we derive asymptotic forms for the Glauber and Born partial wave amplitudes valid for *large* angular momentum quantum numbers l. We shall first consider the large-l behavior of the Glauber partial wave amplitudes.

We begin by returning to Eq. (11) for the generating function $\mathscr{I}_l^G(\lambda, K_i)$, which yields the Glauber partial wave amplitude $\mathscr{F}_l^G$ via Eq. (5a). In Eq. (11) we again employ (10) to rewrite $\Gamma(s-i\eta-l)$ as

$$\Gamma(s-i\eta-l) = (-1)^l \frac{\Gamma(s-i\eta)\Gamma(1-s+i\eta)}{\Gamma(1-s+i\eta+l)}.$$

Using this relation together with

$$\Gamma(s-i\eta)\Gamma(1-s+i\eta) = -\Gamma(1+s-i\eta)\Gamma(i\eta-s)$$

in (11) and after applying (5a) we obtain

$$\begin{aligned}\mathscr{F}_l^G(K_i) &= -(K_i\,\eta)\frac{1}{2K_i^2}\frac{\Gamma(1+i\eta)}{\Gamma(1-i\eta)}\frac{1}{2\pi i}\\ &\quad\times\int_{-\epsilon-i\infty}^{-\epsilon+i\infty} ds\,\frac{\Gamma(-s)\Gamma(i\eta-s)\Gamma(1+s-i\eta)}{\Gamma(1+s)}\\ &\quad\times(s-i\eta-1)\frac{\Gamma(l+1+s-i\eta)}{\Gamma(l+1-s+i\eta)}x^{s-i\eta},\end{aligned} \tag{A1}$$

where $x \equiv \lambda^2/4K_i^2 = (K_i^2 a_0^2)^{-1}$. To obtain the large-$l$ asymptotic behavior of $\mathscr{F}_l^G$, we apply the standard asymptotic expansion for the ratio of two gamma functions[42] to the l-dependent ratio in Eq. (A1); we find that

$$\begin{aligned}&\frac{\Gamma(l+1+s-i\eta)}{\Gamma(l+1-s+i\eta)}\\ &\quad\sim(l+1)^{2s-2i\eta}\left\{1-\frac{s-i\eta}{(l+1)}\right.\\ &\quad\left.-\frac{1}{6}\frac{(s-i\eta)(s-i\eta-1)(2s-2i\eta-1)}{(l+1)^2}+\cdots\right\}.\end{aligned} \tag{A2}$$

With this result $\mathscr{F}_l^G$ is given to second order in $(l+1)^{-1}$ at fixed x and η (i.e., fixed K_i), by

$$\begin{aligned}\mathscr{F}_l^G(K_i) &\sim (K_i\,\eta)\frac{1}{2K_i^2}\frac{\Gamma(1+i\eta)}{\Gamma(1-i\eta)}\frac{1}{2\pi i}\int_{-\epsilon-i\infty}^{-\epsilon+i\infty} ds\\ &\quad\times\frac{\Gamma(-s)\Gamma(i\eta-s)\Gamma(1+s-i\eta)}{\Gamma(1+s)}y^{s-i\eta}\\ &\quad\times\left\{1-(s-i\eta)+\frac{(s-i\eta)(s-i\eta-1)}{(l+1)}\right.\\ &\quad+\frac{1}{6}\frac{(s-i\eta)(s-i\eta-1)}{(l+1)^2}\\ &\quad\times[3+7(s-i\eta-2)\\ &\quad\left.+2(s-i\eta-2)(s-i\eta-3)]\right\},\end{aligned} \tag{A3}$$

where $y = (l+1)^2 x$. When l is large, it can be shown that the use of (A2) in Eq. (A1) involves the neglect of exponentially decreasing terms. Thus (A3) will yield the large-l asymptotic behavior of $\mathscr{F}_l^G$ provided the right-hand side of (A3) dominates decreasing exponential behavior as $l \to \infty$. We now seek to evaluate the integral in (A3) as a function of y. Relation (A3) is rather unwieldly. It may, however, be replaced by the considerably simpler form

$$\begin{aligned}\mathscr{F}_l^G(K_i) &\sim (K_i\,\eta)\frac{1}{2K_i^2}\left\{1-y\frac{\partial}{\partial y}+\frac{y^2}{l+1}\frac{\partial^2}{\partial y^2}\right.\\ &\quad\left.+\frac{1}{6(l+1)^2}\left[3y^2\frac{\partial^2}{\partial y^2}+7y^3\frac{\partial^3}{\partial y^3}+2y^4\frac{\partial^4}{\partial y^4}\right]\right\}\\ &\quad\times A(y),\end{aligned} \tag{A4}$$

where

$$\begin{aligned}A(y) &\equiv \frac{\Gamma(1+i\eta)}{\Gamma(1-i\eta)}\frac{1}{2\pi i}\int_{-\epsilon-i\infty}^{-\epsilon+i\infty} ds\\ &\quad\times\frac{\Gamma(-s)\Gamma(i\eta-s)\Gamma(1+s-i\eta)}{\Gamma(1+s)}y^{s-i\eta}.\end{aligned} \tag{A5}$$

The integral in Eq. (A5) converges if[43] $|\arg(y)| < \pi$; in this region the result may be readily obtained. Since the poles of $\Gamma(-s)$ and $\Gamma(i\eta - s)$ lie to the right of the contour, and the poles of $\Gamma(1+s-i\eta)$ lie to the left, we may immediately identify the integral in (A5) as the Mellin–Barnes representation of a Meijer G-function[43,44] so that

$$\begin{aligned}A(y) &= \frac{\Gamma(1+i\eta)}{\Gamma(1-i\eta)}y^{-i\eta}G_{13}^{21}\left(y\,\middle|\,\begin{matrix}i\eta\\ i\eta,0,0\end{matrix}\right)\\ &= \frac{\Gamma(1+i\eta)}{\Gamma(1-i\eta)}G_{13}^{21}\left(y\,\middle|\,\begin{matrix}0\\ 0,-i\eta,-i\eta\end{matrix}\right).\end{aligned} \tag{A6}$$

It can be readily verified that the G-function in Eqs. (A6) can be expressed in terms of the modified Lommel function $\mathscr{L}_{2i\eta-1}(2i\sqrt{y})$ defined by Thomas and Chan.[8] From (A3) we therefore have

$$\begin{aligned}\mathscr{F}_l^G(K_i) &\sim (K_i\,\eta)\frac{1}{2K_i^2}\left\{1-y\frac{\partial}{\partial y}+\frac{y^2}{l+1}\frac{\partial^2}{\partial y^2}\right.\\ &\quad+\frac{1}{6(l+1)^2}\left[3y^2\frac{\partial^2}{\partial y^2}+7y^3\frac{\partial^3}{\partial y^3}\right.\\ &\quad\left.\left.+2y^4\frac{\partial^4}{\partial y^4}\right]\right\}\frac{\Gamma(1+i\eta)}{\Gamma(1-i\eta)}\\ &\quad\times G_{13}^{21}\left(y\,\middle|\,\begin{matrix}0\\ 0,-i\eta,-i\eta\end{matrix}\right).\end{aligned} \tag{A7}$$

However, from the Mellin–Barnes integral representation of the G^{21}_{13}, one may easily show that

$$\left(\frac{d}{dz}\right)^j G^{21}_{13}\left(z\,\middle|\,\begin{matrix}0\\0,-i\eta,-i\eta\end{matrix}\right) = (-1)^j z^{-j} G^{31}_{13}\left(z\,\middle|\,\begin{matrix}0\\j,-i\eta,-i\eta\end{matrix}\right). \quad \text{(A8)}$$

Consequently

$$\begin{aligned}\mathscr{F}^{\rm G}_l(K_i) \sim (K_i\ \eta)\,\frac{1}{2K_i^2}\,\frac{\Gamma(1+i\eta)}{\Gamma(1-i\eta)} &\times\left\{G^{21}_{13}\left(y\,\middle|\,\begin{matrix}0\\0,-i\eta,-i\eta\end{matrix}\right)\right.\\ &+G^{21}_{13}\left(y\,\middle|\,\begin{matrix}0\\1,-i\eta,-i\eta\end{matrix}\right)\\ &+\frac{1}{l+1}G^{21}_{13}\left(y\,\middle|\,\begin{matrix}0\\2,-i\eta,-i\eta\end{matrix}\right)\\ &+\frac{1}{(l+1)^2}\left[\frac{1}{2}G^{21}_{13}\left(y\,\middle|\,\begin{matrix}0\\2,-i\eta,-i\eta\end{matrix}\right)\right.\\ &-\frac{7}{6}G^{21}_{13}\left(y\,\middle|\,\begin{matrix}0\\3,-i\eta,-i\eta\end{matrix}\right)\\ &\left.\left.+\frac{1}{3}G^{21}_{13}\left(y\,\middle|\,\begin{matrix}0\\4,-i\eta,-i\eta\end{matrix}\right)+\cdots\right]\right\}, \qquad \text{(A9)}\end{aligned}$$

where $y=(l+1)^2/K_i^2 a_0^2$. But when y is large (i.e., $l+1 \gg K_i\,a_0$) the asymptotic behavior of G^{21}_{13} may in turn be determined from its Mellin–Barnes integral representation, namely[43,45]

$$\begin{aligned}\frac{\Gamma(1+i\eta)}{\Gamma(1-i\eta)}G^{21}_{13}\left(y\,\middle|\,\begin{matrix}0\\j,-i\eta,-i\eta\end{matrix}\right) &\sim i\eta\, y^{-1}\left\{\sum_{n=0}^{N}\frac{(n+j)!(1-i\eta)_n(1-i\eta)_n}{n!}y^{-n}\right.\\ &\left.+O(y^{-N-1})\right\}. \qquad \text{(A10)}\end{aligned}$$

Evidently, so long as $\eta\neq 0$, the leading-order behavior of the right side of (A9) dominates the terms neglected when (A2) is employed in (A1). Consequently, the large-l asymptotic behavior of the Glauber partial wave amplitude is given at fixed K_i and $\eta\neq 0$ by

$$\begin{aligned}\mathscr{F}^{\rm G}_l(K_i) \sim i\eta(K_i\ \eta)\,\frac{a_0^2}{(l+1)^2}&\left\{1+\frac{1}{l+1}+\frac{1}{(l+1)^2}\right.\\ &+\frac{3}{2}(1-i\eta)^2\left(\frac{K_i\,a_0}{l+1}\right)^2\\ &\left.\times\left[1+\frac{2}{l+1}+\frac{5}{(l+1)^2}\right]+\cdots\right\}, \qquad \text{(A11)}\end{aligned}$$

provided l is large and $l+1\gg K_i\,a_0$.

The large-l asymptotic behavior of $\mathscr{F}^{\rm B}_l$ cannot be obtained by applying the limiting relation (18b) to (A11) since the limit is zero. Moreover (18b) cannot be applied to (A9) to obtain the asymptotic behavior of $\mathscr{F}^{\rm B}_l$, even though the right side of (A9) is well defined in the limit of (18b). The failure of the foregoing large-l asymptotic analysis when $\eta=0$ in the contour integral in Eq. (A1), stems from the fact that[44,46]

$$G^{21}_{13}\left(y\,\middle|\,\begin{matrix}0\\0,0,0\end{matrix}\right)=G^{20}_{02}\,(y|0,0)=2K_0(2y^{1/2}),$$

where K_0 is the modified Bessel function of the second kind. Consequently, all terms on the right of (A9) decrease as $e^{-2y^{1/2}}$ as $l\to\infty$ at fixed K_i and the right side of (A9) is of the same order as the terms neglected in (A3).

The large-l asymptotic behavior of $\mathscr{F}^{\rm B}_l$ may be obtained from Eq. (21b) by exploiting the relation[47]

$$\begin{aligned}Q_l(z)=\sqrt{\pi}\,[z+(z^2-1)^{1/2}]^{-l-1}\,[\Gamma(l+1)/\Gamma(l+\tfrac{3}{2})]\\ \times{}_2F_1(\tfrac{1}{2},l+\tfrac{1}{2};l+\tfrac{3}{2};w), \qquad \text{(A12)}\end{aligned}$$

where

$$w\equiv\frac{z-(z^2-1)^{1/2}}{z+(z^2-1)^{1/2}}=[z+(z^2-1)^{1/2}]^{-2}.$$

Recalling that in Eq. (21b), $z=1+2/K_i^2 a_0^2$, we see that $0\leqslant w<1$ if $0\leqslant K_i\,a_0<\infty$. From (A12) it can be shown that, for large l,

$$\begin{aligned}Q_l(z)\sim\sqrt{\pi}\,(l+1)^{-1/2}\,w^{(l+1)/2}(1-w)^{-1/2}\\ \times\left\{1+\frac{1}{8(l+1)}\left(1-\frac{2\omega}{1-\omega}\right)+\cdots\right\}. \qquad \text{(A13)}\end{aligned}$$

Since $z=1+2/K_i^2 a_0^2$, the asymptotic relation (A13) provides a good approximate to $Q_l(z)$ only if $l+1\gg K_i\,a_0$. Employing (A13) in Eq. (21b) we therefore find, provided $l+1\gg K_i\,a_0$,

$$\begin{aligned}\mathscr{F}^{\rm B}_l(K_i)\sim(K_i\ \eta)\,K_i^{-2}\sqrt{\pi}(1-w)^{-3/2}&\\ \times(1-\sqrt{w})^2\,w^{(l+1)/2}(l+1)^{1/2}&\\ \times\left\{1+\frac{1}{l+1}\left[\frac{1+2\sqrt{w}+2w}{1-w}\right.\right.&\\ \left.\left.+\frac{1}{8}\left(1-\frac{2w}{1-w}\right)\right]+\cdots\right\},& \qquad \text{(A14)}\end{aligned}$$

since $z=\frac{1}{2}(1+w)/\sqrt{w}$.

That (A11) and (A14) provide useful large-l asymptotic expansions of $\mathscr{F}^{\rm G}_l$ and $\mathscr{F}^{\rm B}_l$ has been verified directly. Indeed, (A11) especially provides a good approximation to $\mathscr{F}^{\rm G}_l$ whenever $(l+1)/K_i\,a_0\gg 1$ and $l+1\gtrsim 5$ hold; in other words, (A11) yields a good approximation (three or four significant figures) to the exact $\mathscr{F}^{\rm G}_l$ computed via Eq. (17) even at relatively small values of l provided K_i is also small. However, one should not infer from these results that (A11) is the asymptotic form of $\mathscr{F}^{\rm G}_l$ when K_i is small. Relation (A14) provides a good approximate (~ 3 significant figures) to the exact $\mathscr{F}^{\rm B}_l$ computed via (21b), whenever $l\gtrsim 10$ and $l+1\gg K_i\,a_0$.

APPENDIX B: ASYMPTOTIC BEHAVIOR OF $\mathscr{F}^{\rm G}_l$ as $|\eta|\to\infty$

We next discuss the asymptotic behavior of $\mathscr{F}^{\rm G}_l$ as $|\eta|\to\infty$ at fixed l. Recall that $\eta=-Z_i e^2/\hbar v_i=-Z_i\mu e^2/\hbar^2 K_i$. As we shall see, this limit of $\mathscr{F}^{\rm G}_l$ corresponds to three possibilities: (i) $|Z_i|\to\infty$; (ii) $\mu\to\infty$ at fixed K_i; and (iii) $v_i\to 0$. We again return to the integral representation of $\mathscr{F}_l^G$ given by Eq. (11). We apply (5a) to Eq. (11); after performing the required differentiation and changing integration variables via $t = -s-l-1+i\eta$, we obtain

$$\mathscr{F}_l^G = \frac{\eta}{2K_i}\frac{\Gamma(1+i\eta)}{\Gamma(1-i\eta)}(-1)^{l+1}\frac{1}{2\pi i}\int_{-l-1+\epsilon-i\infty}^{-l-1+\epsilon+i\infty} dt \times\Gamma(-2l-1-t)\Gamma(-t)(t+l+2) \times\frac{\Gamma(t+l+1-i\eta)}{\Gamma(-t-l+i\eta)}x^{-t-l-1}, \tag{B1}$$

where $x = 1/K_i^2 a_0^2$ and $|\arg(x)| < \pi$. To obtain the asymptotic behavior of $\mathscr{F}_l^G$ as $|\eta| \to \infty$ we again employ the asymptotic formula for the ratio of two gamma functions[42] to write

$$\frac{\Gamma(l+1+t-i\eta)}{\Gamma(1-i\eta)} \sim (-i\eta)^{l+t}\left\{1+\frac{(l+t)(l+t+1)}{2(-i\eta)}+\cdots\right\}$$

and

$$\frac{\Gamma(1+i\eta)}{\Gamma(-l-t+i\eta)} \sim (i\eta)^{l+t+1}\left\{1+\frac{(1+l+t)(-l-t)}{2i\eta}+\cdots\right\}.$$

Employing these expressions in (B1) and collecting terms yields

$$\mathscr{F}_l^G \sim \frac{i}{2K_i}(-1)^{l+1}\frac{1}{2\pi i}\int_{\epsilon-l-1-i\infty}^{\epsilon-l-1+i\infty} dt \times\Gamma(-2l-1-t)\Gamma(-t)\,y^{l+1+t}\{(t+l+2) -(1/i\eta)(t+l+2)(t+l+1)(t+l)+\cdots\}, \tag{B2}$$

where $y = \eta^2/x = \eta^2 K_i^2 a_0^2 = Z_i^2 e^4\mu^2 a_0^2/\hbar^4$. Since y is independent of the incident particle speed v_i, it should be evident that (B2) directly yields the asymptotic behavior of $\mathscr{F}_l^G$ as $v_i \to 0$. On the other hand, if either $|Z_i| \to \infty$ or $\mu \to \infty$, we must first evaluate the right side of (B2) and then examine the asymptotic behavior as $y \to \infty$ as well. For arbitrary y, we may reexpress (B2) as

$$\mathscr{F}_l^G \sim \frac{i}{2K_i}(-1)^{l+1}\left\{\frac{d}{dy}-\frac{1}{i\eta}y^2\frac{\partial^3}{\partial y^3}\right\}y^{l+2} \times\frac{1}{2\pi i}\int_{-l-1+\epsilon-i\infty}^{-l-1+\epsilon+i\infty} dt\,\Gamma(-2l-1-t)\Gamma(-t)\,y^t. \tag{B3}$$

The integral in (B3) converges provided[43] $|\arg(y)| < \pi$. We next note that the integrand in (B3) is of the proper form to identify the Mellin–Barnes integral as the representation of a modified Bessel function of the second kind[43,46]; however, we may not directly do so because $l+1$ poles of $\Gamma(-2l-1-t)$ lie to the left of the contour. Moreover, the contour may not be closed in the left half t plane. Therefore we let C_1 denote the contour indicated in (B3) and let C_2 denote a closed counterclockwise contour that encircles only the $l+1$ poles of $\Gamma(-2l-1-t)$ to the left of C_1. Then

$$\frac{1}{2\pi i}\int_{C_1} dt\,\Gamma(-t-2l-1)\Gamma(-t)\,y^t = \frac{1}{2\pi i}\int_{C_1-C_2} dt\cdots+\frac{1}{2\pi i}\oint_{C_2} dt\cdots,$$

where the contour labeled (C_1-C_2) may be deformed into the straight line contour running from $-2l-1-\epsilon-i\infty$ to $-2l-1-\epsilon+i\infty$ parallel to the imaginary t axis. The integral over the contour C_2 may be evaluated via the residue theorem, which gives

$$\frac{1}{2\pi i}\oint_{C_2} dt\,\Gamma(-2l-1-t)\Gamma(-t)\,y^t = -\sum_{j=0}^{l}\frac{(-1)^j(2l-j)!}{j!}y^{j-2l-1}. \tag{B4}$$

Since all poles of the integrand now lie to the right of the contour (C_1-C_2), we have[43,46]

$$\frac{1}{2\pi i}\int_{C_1-C_2} dt\,\Gamma(-2l-1-t)\Gamma(-t)\,y^t = G_{02}^{20}(y|0,-2l-1) = 2y^{-l-1/2}K_{2l+1}(2y^{1/2}) \tag{B5}$$

provided $|\arg(y)| < \pi$. Thus

$$\frac{1}{2\pi i}\int_{-l-1+\epsilon-i\infty}^{-l-1+\epsilon+i\infty} dt\,\Gamma(-t-2l-1)\Gamma(-t)\,y^t = 2y^{-l-1/2}\left\{K_{2l+1}(2y^{1/2}) -\frac{1}{2}\sum_{j=0}^{l}\frac{(-1)^j(2l-j)!}{j!}y^{j-l-1/2}\right\}. \tag{B6}$$

Since $y = Z_i^2 e^4\mu^2 a_0^2/\hbar^4$, the requirement that $|\arg(y)| < \pi$ dictates the definition $y^{1/2} = |Z_i|e^2\mu a_0/\hbar^2 \equiv \zeta$. Consequently, from (B3) and (B6),

$$\mathscr{F}_l^G \sim \frac{i}{K_i}(-1)^{l+1}\left\{\frac{1}{2\zeta}\frac{\partial}{\partial\zeta}-\frac{1}{i\eta}\zeta^4\left(\frac{1}{2\zeta}\frac{\partial}{\partial\zeta}\right)^3\right\} \times\left\{\zeta^3\left[K_{2l+1}(2\zeta)-\frac{1}{2} \times\sum_{j=0}^{l}\frac{(-1)^j(2l-j)!}{j!}\zeta^{2j-2l-1}\right]\right\}. \tag{B7}$$

From the standard ascending series[48] for K_{2l+1}, it can be easily seen that the finite sum in (B7) exactly removes from $K_{2l+1}(2\zeta)$ all terms that are singular as $\zeta \to 0$. To carry out the differentiations indicated in (B7) we employ the relation[49]

$$\frac{1}{\zeta}\frac{\partial}{\partial\zeta}[\zeta^\mu K_\nu(2\zeta)] = (\mu-\nu)\zeta^{\mu-2}K_\nu(2\zeta)-2\zeta^{\mu-1}K_{\nu-1}(2\zeta), \tag{B8}$$

while

$$\left(\frac{1}{2\zeta}\frac{\partial}{\partial\zeta}\right)^n[\zeta^3\zeta^{2j-2l-1}] = (-1)^n(l-1-j)_n\,\zeta^{2j-2l+2-2n}.$$

Therefore, as $|\eta| \to \infty$,

$$\mathscr{F}_l^G \sim \frac{i}{K_i}(-1)^l\left\{\left[(l-1)\zeta K_{2l+1}(2\zeta)+\zeta^2 K_{2l}(2\zeta)\right.\right.$$

$$-\frac{1}{2}\sum_{j=0}^{l}\frac{(-1)^{j}(2l-j)!}{j!}(l-1-j)\zeta^{2j-2l}\Bigg]$$
$$-\frac{1}{i\eta}\Bigg[(l-1)_3\zeta K_{2l+1}(2\zeta)+3(l-1)_2\zeta^2K_{2l}(2\zeta)$$
$$+3(l-1)\zeta^3K_{2l-1}(2\zeta)+\zeta^4K_{2l-2}(2\zeta)$$
$$-\frac{1}{2}\sum_{j=0}^{\infty}\frac{(-1)^{j}(2l-j)!}{j!}(l-1-j)_3\zeta^{2j-2l}\Bigg]+\cdots\Bigg\}, \tag{B9}$$

where $\zeta=|Z_i|e^2\mu a_0/\hbar^2$ and l is fixed.

Relation (B9) holds as $|\eta|\rightarrow\infty$. In particular, if $|\eta|\rightarrow\infty$ means $v_i\rightarrow 0$, and $\zeta\sim 1$ (e.g., for incident electrons or positrons $\zeta=1$), the Bessel function terms in (B9) must be evaluated exactly. However, if $|\eta|\rightarrow\infty$ and $\zeta\gg 1$, i.e., if $|Z_i|\gg 1$, or $\mu\gg 1$ (e.g., incident protons), the Bessel function terms are exponentially damped and

$$\mathscr{F}_l^{\mathrm{G}}\sim(i/2K_i)[1+(l-1)_4\zeta^{-4}+\cdots]. \tag{B10}$$

In (B10) the term proportional to ζ^{-4} vanishes if $l=0$ or $l=1$.

By comparison with the foregoing analysis of the $|\eta|\rightarrow\infty$ behavior of $\mathscr{F}_l^{\mathrm{G}}$, the corresponding analysis of $\mathscr{F}_l^{\mathrm{B}}$ is much more straightforward. Now, however, the limits $v_i\rightarrow 0,\mu\rightarrow\infty$, and $|Z_i|\rightarrow\infty$ must be considered separately. First from (21b) it is clear that in the limit $|Z_i|\rightarrow\infty$ or $\mu\rightarrow\infty$ at fixed K_i that $\mathscr{F}_l^{\mathrm{B}}$ is unbounded. Of course, K_i fixed as $\mu\rightarrow\infty$ means $v_i\rightarrow 0$. Second, the limit as $v_i\rightarrow 0$ (i.e., $K_i\rightarrow 0$) may be directly obtained from Eq. (20), after differentiating according to (19a) and expanding the resulting hypergeometric functions in powers of $(K_i^2a_0^2)$. We easily find that, as $v_i\rightarrow 0$ at fixed l,

$$\mathscr{F}_l^{\mathrm{B}}\sim\frac{1}{2}(K_i\,\eta\,a_0^2)\frac{(l!)^2}{(2l+1)!}(K_i^2a_0^2)^l$$
$$\times\{(l+2)-\tfrac{1}{2}(l+1)(l+3)K_i^2a_0^2+\cdots\}. \tag{B11}$$

Finally, to obtain the behavior of $\mathscr{F}_l^{\mathrm{B}}$ as $\mu\rightarrow\infty$ at fixed v_i (i.e., $K_i\rightarrow\infty$) we directly examine the result (21b). Since $z=1+2/K_i^2a_0^2$, as $K_i\rightarrow\infty$, $z\rightarrow 1$; but[50] as $z\rightarrow 1$, $Q_l(z)\sim-[\tfrac{1}{2}\ln((z-1)/z)+\gamma+\psi(l+1)]$, where[38] γ is Euler's constant and $\psi(z)=d[\ln\Gamma(z)]/dz$. Thus, as $\mu\rightarrow\infty$ at fixed v_i

$$\mathscr{F}_l^{\mathrm{B}}\sim(K_i\,\eta)(1/K_i^2)[\ln(K_i^2a_0^2)$$
$$-\gamma-\psi(l+1)+\tfrac{1}{2}+\cdots], \tag{B12}$$

where $\eta/K_i=-Z_ie^2/(\mu v_i^2)$ and $\hbar K_i=\mu v_i$, and we have exploited the recurrence formula[38] for $\psi(z)$. Relation (B12) also holds when $K_i\rightarrow\infty$ and μ is fixed.

[1]R. J. Glauber, in *Lectures in Theoretical Physics*, edited by W. E. Brittin and L. G. Duncan (Interscience, New York, 1959), Vol. I, pp. 315–444.
[2]E. Gerjuoy and B. K. Thomas, Rep. Prog. Phys. **37**, 1345 (1974).
[3]See, also, B. K. Thomas and V. Franco, Phys. Rev. A **13**, 2004 (1976); F. T. Chan, M. Lieber, G. Foster, and W. Williamson, Jr., in *Advances in Electronics and Electron Physics*, edited by L. Marton and C. Marton (Academic, New York, 1979), Vol. 49, p. 133.
[4]J. F. Williams, J. Phys. B **8**, 2191 (1975); P. J. O. Teubner, C. R. Lloyd, and E. Weigold, Phys. Rev. A **9**, 2552 (1974); C. R. Lloyd, P. J. O. Teubner, E. Weigold, and B. R. Lewis, Phys. Rev. A **10**, 175 (1974).
[5]T. Ishihara and J. C. Y. Chen, Phys. Rev. A **12**, 370 (1975).
[6]F. W. Byron and C. J. Joachain, Phys. Rev. A **8**, 1267 (1973); J. Phys. B **7**, L212 (1974).
[7]F. W. Byron, Jr. and C. J. Joachain, J. Phys. B **10**, 207 (1977).
[8]V. Franco, Phys. Rev. A **1**, 1705 (1970); B. K. Thomas and F. T. Chan, Phys. Rev. A **8**, 252 (1973).
[9]V. Franco, Phys. Rev. Lett. **20**, 709 (1968).
[10]B. K. Thomas and E. Gerjuoy, J. Math. Phys. **12**, 1567 (1971).
[11]V. Franco and R. J. Glauber, Phys. Rev. **142**, 1195 (1966).
[12]B. K. Thomas, "Point Coulomb scattering in e^-–H(1s) elastic scattering," unpublished.
[13]J. C. Y. Chen and L. Hambro, J. Phys. B **5**, L199 (1972); J. C. Y. Chen, L. Hambro, A. L. Sinfailam, and K. T. Chung, Phys. Rev. A **7**, 2003 (1973); L. Hambro, J. C. Y. Chen, and T. Ishihara, Phys. Rev. A **8**, 1283 (1973).
[14]F. W. Byron Jr. and C. J. Joachain, Phys. Rev. A **9**, 2559 (1974).
[15]T. T. Gien, J. Phys. B **9**, 3203 (1976); Phys. Rev. A **14**, 1918 (1977).
[16]J. N. Gau and J. Macek, Phys. Rev. A **10**, 522 (1974); **12**, 1760 (1975).
[17]R. N. Madan, Phys. Rev. A **11**, 1968 (1975); **12**, 2631 (1975).
[18]G. Foster and W. Williamson, Jr., Phys. Rev. A **13**, 936, 2023 (1976).
[19]A. Birman and S. Rosendorff, J. Phys. B **9**, L189 (1976).
[20]See, e.g., J. R. Taylor, *Scattering Theory* (Wiley, New York, 1972), or any of the by now standard treatments of quantum scattering theory.
[21]N. F. Mott and H. S. W. Massey, *The Theory of Atomic Collisions* (Oxford U. P., Oxford, 1965), 3rd ed., p. 24.
[22]A. Erdelyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. I, p. 62.
[23]Reference 22, p. 56.
[24]Reference 22, p. 63.
[25]I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1965), 4th ed., p. 717.
[26]W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer, New York, 1966), p. 4.
[27]Reference 26, p. 3.
[28]Reference 22, p. 58.
[29]Reference 22, p. 133.
[30]Reference 22, p. 161.
[31]Reference 21, p. 464. Note, however, that these authors do not identify the right-hand side of Eqs. (21) with $\mathscr{F}_l^{\mathrm{B}}=[e^{2i\delta_l^{\mathrm{B}}}-1]/2iK_i$, but rather with $\delta_l^{\mathrm{B}}/K_i$; i.e., the right-hand side of (21) is taken as the usual first Born approximation to the phase shift. If the phase shift is not small for all l, the approximate series $[\sum_{l=0}^{\infty}(2l+1)\delta_l^{\mathrm{B}}P_l(\cos\theta)]/K_i$ does not sum to the conventional first Born approximation to the full scattering amplitude.
[32]Reference 22, p. 122.
[33]See, e.g., Ref. 21, p. 89.
[34]G. N. Watson, *Theory of Bessel Functions* (Cambridge U. P., London, 1944), 2nd ed., p. 403.
[35]M. R. C. McDowell and J. P. Coleman, *Introduction to the Theory of Ion–Atom Collisions* (North-Holland, Amsterdam, 1970), pp. 27–33.
[36]See Ref. 14 and references contained therein.
[37]See, e.g., L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Pergamon, Oxford, 1977), 3rd ed., p. 507.
[38]*Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Natl. Bur. Stand., Washington, D. C., 1964), p. 258.
[39]Reference 26, p. 238.
[40]See, e.g., H. R. J. Walters, J. Phys. B **6**, 1003 (1973).
[41]Reference 21, pp. 324–7.
[42]Reference 38, p. 257.
[43]Reference 22, p. 207.
[44]Reference 22, p. 209.
[45]N. Bleistein and R. A. Handelsman, *Asymptotic Expansions of Integrals* (Holt, Rinehart, and Winston, New York, 1975), Chap. 4.
[46]Reference 22, p. 216.
[47]Reference 22, p. 136.
[48]Reference 38, p. 375.
[49]Equation (B8) may be obtained from the standard differentiation formulas for the modified Bessel functions of the second kind. See, for example, Ref. 38, p. 376.
[50]Reference 22, p. 163.

Systematics of strongly self-dominant higher-order differential equations based on the Painlevé analysis of their singularities

R. B. King
Department of Chemistry, University of Georgia, Athens, Georgia 30602

(Received 23 July 1985; accepted for publication 3 December 1985)

This paper presents a simple way of classifying higher-order differential equations based on the requirements of the Painlevé property, i.e., the presence of no movable critical points. The fundamental building blocks for such equations may be generated by strongly self-dominant differential equations of the type $(\partial/\partial x)^n u = (\partial/\partial x^m)[u^{(m-n+p)/p}]$ in which m and n are positive integers and p is a negative integer. Such differential equations having both a constant degree d and a constant value of the difference $n - m$ form a *Painlevé chain*; however, only three of the many possible Painlevé chains can have the Painlevé property. Among the three Painlevé chains that can have the Painlevé property, one contains the Burgers' equation; another contains the dominant terms of the first Painlevé transcendent, the isospectral Korteweg–de Vries equation, and the isospectral Boussinesq equation; and the third contains the dominant terms of the second Painlevé transcendent and the isospectral modified (cubic) Korteweg–de Vries equation. Differential equations of the same order and having the same value of the quotient $(n-m)/(d-1)$ can be mixed to generate a new hybrid differential equation. In such cases a hybrid can have the Painlevé property even if only one of its components has the Painlevé property. Such hybridization processes can be used to generate the various fifth-order evolution equations of interest, namely the Caudrey–Dodd–Gibbon, Kuperschmidt, and Morris equations.

I. INTRODUCTION

In recent years theoretical physicists have been very interested in a certain class of nonlinear partial differential equations known as evolution equations.[1,2] This interest has arisen from the realization that these equations possess a special type of elementary solution, which takes the form of localized disturbances that act somewhat like particles and are therefore known as solitons.

Solution of these evolution equations involves the so-called inverse scattering transform.[1–3] In this connection it has been noted[4–6] that there is a relation between nonlinear partial differential equations solvable by an inverse scattering transform and nonlinear ordinary differential equations (ODE's) without movable critical points; such ODE's are said to have the *Painlevé property*. An algorithm has been developed[5,6] for the determination of whether a given system of differential equations has the Painlevé property.

In applying the algorithm to test systems of differential equations for the Painlevé property, the question naturally arises to what extent useful information can be obtained on the properties of a given differential equation by simple inspection of the equation without extensive calculation. This paper explores this question and provides a simple approach for identifying higher-order differential equations possessing necessary conditions to have the Painlevé property. More extensive tests are then required in order to determine whether these necessary conditions are sufficient for specific equations to have the Painlevé property.

II. DOMINANCE CLASSES AND FAMILIES

In this work we are interested in evolution equations of the form

$$u_t + f(x,u,u_{1x},\dots,u_{nx}) = 0, \qquad (1)$$

in which

$$u_t = \frac{\partial u}{\partial t} \text{ and } u_{jx} = \left(\frac{\partial}{\partial x}\right)^j u, \quad j = 0,1,\dots,n. \qquad (2)$$

In these equations u may be regarded as an amplitude, x as a distance, and t as time. Of particular interest are time-independent solutions, where $u_t = 0$ and therefore

$$f(x,u,u_{1x},\dots,u_{nx}) = 0. \qquad (3)$$

Let us adjust the distance scale x so that $x = x^*$ is a critical point. The dominant behavior of solutions of Eq. (3) in the neighborhood of the critical point $x = x^*$ can be expressed as the following series:

$$u = a(x - x^*)^p \quad \text{as} \quad x \to x^*. \qquad (4)$$

Substitution of Eq. (4) into Eq. (3) shows that for certain values of the exponent p, two or more terms may balance (possibly depending upon a) and the rest can be ignored as $x \to x^*$. For each choice of p the terms that can balance are called the *dominant terms*. Modification of Eq. (3) by deletion of all nondominant terms in general leads to a new simpler equation called the *dominant truncation* of Eq. (3). All equations giving the same dominant truncation may be considered as forming a *dominance class*. All dominance classes that have identical dominant truncations except for multiplicative constants may be regarded as forming a *dominance family*. A *self-dominant* equation is one in which all of its terms are dominant and is therefore identical to its dominant truncation.

Painlevé[7] has identified 50 canonical forms of second-order differential equations that lack movable critical points and that, therefore, are related to nonlinear differential equations solvable by inverse scattering transforms. The methods and ideas to be used in this paper can be illustrated first with the simplest irreducible Painlevé equation, namely

 0022-2488/86/040966-06$02.50

$$u_{xx} = 6u^2 + x. \tag{5}$$

Expressing u by Eq. (4) gives

$$ap(p-1)(x-x^*)^{p-2} = 6a^2(x-x^*)^{2p} + x. \tag{6}$$

Balancing the $(x-x^*)^{p-2}$ term on the left with the $(x-x^*)^{2p}$ term on the right gives $p-2=2p$ or $p=-2$. The x term on the right of Eqs. (5) and (6) is not involved in the balancing. Such terms are called *recessive terms* and are those dropped from the differential equation to form its dominant truncation. Thus the dominant truncation of Eq. (5) is

$$u_{xx} = 6u^2. \tag{7}$$

Equation (7) is also the Painlevé canonical equation II, which is solvable by elliptic functions.[7]

The next simplest irreducible Painlevé equation is

$$u_{xx} = 2u^3 + xu + b. \tag{8}$$

Analogously expressing u in Eq. (8) by Eq. (4) gives

$$ap(p-1)(x-x^*)^{p-2} = 2a^3(x-x^*)^{3p} + ax(x-x^*)^{p} + b. \tag{9}$$

Balancing the $(x-x^*)^{p-2}$ term on the left with the $(x-x^*)^{3p}$ term on the right gives $p-2=3p$ or $p=-1$. The $(x-x^*)^{p}$ term on the right of Eq. (9) cannot be involved in the balancing since $p-2\neq p$. This term is therefore a recessive term as is the constant term b of Eq. (8). Therefore the dominant truncation of Eq. (8) is

$$u_{xx} = 2u^3. \tag{10}$$

A simple type of second-order differential equation reducible to a first-order differential equation of Riccati type is the Painlevé canonical equation V, namely

$$u_{xx} = -2uu_x + bu_x + b'u. \tag{11}$$

Methods analogous to those used above indicate that $p=-1$ for Eq. (11) and that its dominant truncation is

$$u_{xx} = -2uu_x. \tag{12}$$

Thus Eqs. (8) and (11) lead to the same value for the exponent p when expanded in the neighborhood of a critical point by using Eq. (4) but lead to dominant truncations having very different forms. Thus Eqs. (8) and (11) are in different dominant families.

Second-order differential equations without movable critical points are also possible which have dominant truncations that are linear combinations of equations of the types (10) and (12). This phenomenon, which can be called *hybridization*, is possible because Eqs. (10) and (12) lead to the same value of the exponent p when Eq. (4) is substituted into them. The simplest example of hybridization in the Painlevé canonical equations[7] occurs in Painlevé equation VI, namely

$$u_{xx} = -3uu_x - u^3 + bu_x - bu^2. \tag{13}$$

The dominant truncation of Eq. (13) is

$$u_{xx} = -3uu_x - u^3. \tag{14}$$

This is a linear combination of Eqs. (10) and (12) with appropriate adjustments of the multiplicative constants.

Another phenomenon is observed in the third and higher irreducible Painlevé equations.[7] Thus the third irreducible Painlevé equation (canonical form[7] XIII) is

$$u_{xx} = (u_x^2/u) + bu^3 + cu^2 + d + (e/u). \tag{15}$$

Substituting Eq. (4) into Eq. (10) gives

$$\begin{aligned} ap(p-1)(x-x^*)^{p-2} &= ap^2(x-x^*)^{p-2} + ba^3(x-x^*)^{3p} + ca^2(x-x^*)^{2p} \\ &\quad + d + (e/a)(x-x^*)^{-p}. \end{aligned} \tag{16}$$

The $ap^2(x-x^*)^{p-2}$ and $ba^3(x-x^*)^{3p}$ terms on the right of Eq. (16) are both dominant terms but only the $ba^3(x-x^*)^{3p}$ term can be used to determine the exponent p to be -1. The $ba^3(x-x^*)^{3p}$ term may therefore be considered to be an *active* dominant term. Similarly $ap^2(x-x^*)^{p-2}$ may be regarded as a *passive* dominant term. Self-dominant equations having only active dominant terms may be called *strongly self-dominant* equations. Since passive dominant terms are not found in the evolution equations of interest, only strongly self-dominant equations will be considered in this paper. These will be seen to relate closely to the evolution equations.

III. PAINLEVÉ ANALYSIS OF STRONGLY SELF-DOMINANT EQUATIONS

We will now consider the general features of the so-called Painlevé analysis used to determine whether a given strongly self-dominant differential equation has the Painlevé property. Such equations can be expressed as polynomials of the following type:

$$u_{nx} = g(x,u,u_{1x},\ldots,u_{(n-1)x}) = \sum_i c_i \prod_{j=0}^{n-1} u_{jx}^{q_j}, \tag{17}$$

in which n is thus the order of the differential equation. Let us now assign to Eq. (17) the following integers: n = order of the equation [order of the derivative u_{nx} on the left-hand side of Eq. (17) which is the highest-order derivative in the equation];

$$m_i = \sum_{j=0}^{n-1} jq_j$$

[the weighted sums of the derivatives in the terms on the right-hand side of Eq. (17)]; and

$$d_i = \sum_{j=0}^{n-1} q_j$$

[degrees of the polynomial terms on the right-hand side of Eq. (17)]. In general $m_i \neq m_k$ and $d_i \neq d_k$. However, initially we shall consider *homogeneous* equations (17) in which $m_i = m_k$ and $d_i = d_k$ for all values of i and k. For such a homogeneous equation we can assign unique values of m and d. Let us call m and d the *co-order* and the *degree*, respectively, of the equation.

Let us now apply Painlevé analysis to Eq. (17). Express u as the power series in Eq. (4). Determine the exponent p that balances the terms. By taking appropriate derivatives of Eq. (4) the following relationship can be seen to hold:

$$p = (m-n)/(d-1). \tag{18}$$

If p is not an integer, then Eq. (17) has a movable algebraic branch point implying non-Painlevé behavior. We therefore are interested in self-dominant systems of the type represented by Eq. (17) for which p, as determined by Eq. (18), is a negative integer.

If p [Eq. (18)] is a negative integer, then Eq. (4) may represent the first term in a Laurent series[8] valid in a deleted neighborhood of a movable *pole*. In this case a solution of Eq. (17) is of the following type:

$$u = (x - x^*)^p \sum_{k=0}^{\infty} a_k (x - x^*)^k, \tag{19}$$

where $x - x^* \neq 0$. In this case the position x^* of the singular value of x corresponds to one of the n integration constants. If $n-1$ of the coefficients a_k are also arbitrary, the n integration constants of Eq. (17) are then accounted for and Eq. (19) represents the solution of Eq. (17) in the deleted neighborhood of the singularity x^*. The powers of x at which these arbitrary constants enter are called the *resonances* and will be designated as $r_1, r_2, \ldots, r_n$ so that $r_i < r_k$ for $i < k$.

In order to find the resonances the following equation for u is substituted into Eq. (17):

$$u = a(x - x^*)^p + b(x - x^*)^{p+r}. \tag{20}$$

The coefficient a is obtained by equating the coefficients of the $(x - x^*)^{p-n}$ terms, which are the leading terms in the neighborhood of x^*. For a *homogeneous* equation (17) the coefficient a is uniquely determined. After determining a then the coefficients of the next-higher powers $(x - x^*)^{p+r-n}$ are equated in order to determine the resonances. In this way the resulting equations for the resonances to leading order in b reduce to

$$Q(r) b (x - x^*)^q = 0, \quad q \geqslant p + r - n, \tag{21}$$

in which $Q(r)$ is a polynomial in r of degree n. The roots of $Q(r)$ determine the resonances since $Q(r) = 0$ corresponds to the "indicial equation" used to solve a linear ordinary differential equation near a regular singular point.[9]

Let us now consider some features of this indicial polynomial $Q(r)$. Because of the rules for differentiation, the left-hand side of Eq. (17) will generate the nth-degree polynomial $L(r)$ of the following type:

$$L(r) = (r+p)(r+p-1)\cdots(r+p-n). \tag{22}$$

Since p is a negative integer, $L(r)$ is *not* divisible by $r+1$ and all of its roots are real *positive* integers. However, the polynomial $Q(r)$ must be divisible by $r+1$, reflecting the arbitrariness of the singularity $x = x^*$ corresponding to one of the n integration constants. This leads to an automatic root of -1 for $Q(r)$. Therefore substitution of Eq. (20) into the right-hand side of Eq. (17) must generate a polynomial $R(r)$ so that the difference $L(r) - R(r)$ is divisible by $r+1$ *and* is factorable into linear factors so that all of its roots are real integers. However, the degree of $R(r)$ is $m < n$ so that the terms in $L(r)$ of the type $h_k r^{n-k}$ in which $n - k > m$ will be the same as the corresponding terms of the indicial polynomial $Q(r)$.

Arguments based on the relationships between the coefficients of the highest-degree terms of polynomials and the sums of powers of their roots[10] indicate that the degree of $R(r)$, which corresponds to the co-order m of the original differential equation (17), must be high enough so that the difference $L(r) - R(r)$ with $L(r)$ defined by Eq. (22) becomes divisible by $r+1$, while remaining factorable into linear factors even though $L(r)$ itself is not divisible by $r+1$. Otherwise the corresponding differential equation (17) will not have the Painlevé property. For this reason only differential equations (17) having $p = -1$ or $p = -2$ can be candidates for equations having the Painlevé property.

The next step is to find the roots of $Q(r)$. If all of the roots of $Q(r)$ other than the automatic -1 roots are real integers with at least one root greater than -1, then the system can be free from algebraic branch points. The corresponding homogeneous differential equation (17) is a possible candidate for a system having the Painlevé property. The complete Painlevé analysis requires additional steps involving determination of the integration constants.[5] These additional steps are sufficiently more complicated and tedious so that they cannot be applied readily to the diverse variety of systems considered in this paper. We therefore shall limit the discussion in this paper to the identification of the types of differential equations (17), which can lead to the integral resonances required for the Painlevé property.

IV. RESULTS

Table I lists all of the possible types of strongly self-dominant homogeneous differential equations of order $\leqslant 4$ that have the negative integral balancing exponents p [Eq. (18)] required for the Painlevé property. Of these 16 equation types, nine are shown by the methods outlined above to have the real integral resonances required for the Painlevé property. These nine equations are all members of the three Painlevé chains described in Table II. In this context a *Painlevé chain* consists of differential equations of the following type:

$$u_{nx} = \left(\frac{\partial}{\partial x}\right)^m [u^{(m-n+p)/p}], \quad m = 1,2,\ldots. \tag{23}$$

The Painlevé chains may be characterized by the fraction $(n-m)/(d-1)$, which, by Eq. (18), is the negative of the exponent p. The three Painlevé chains depicted in Table II are the only possible Painlevé chains for which $n - m \leqslant 2$ and therefore, for reasons outlined in the previous section, are the only possible Painlevé chains giving the integral resonances required for the Painlevé property. Note also that each member of a Painlevé chain has all of the resonances of the previous member plus one additional resonance. In ascending a Painlevé chain to higher orders through the successive differentiations implied by Eq. (23), points may be reached where the indicial polynomial $Q(r)$ has a multiple root (i.e., order 4 for the 2/2 chain and order 5 for the 2/1 chain) and a point is reached where the right-hand side of the differential equation splits into more than one term. The algorithm for determining the resonances is independent of the parameter k (see Table I) since products of the type ka^{d-1} (a from Eq. 4 for u) are constant. However, in equations having multiple terms on the right-hand side, the positions of the resonances depend upon the ratios of the coefficients of these terms. However, the process of obtaining the members of a Painlevé chain through Eq. (23) and the implied successive differen-

TABLE I. Self-dominant homogeneous differential equations of order $\leqslant 4$ with negative integral balancing exponent p.

Equation type	Order n	Co-order m	Degree d	Exponent p	Fraction $(n-m)/(d-1)$	Resonances	Equation type[a]
$u_{2x} = ku^3$	2	0	3	-1	2/2	$-1, +4$	DT of Painlevé II
$u_{2x} = kuu_x$	2	1	2	-1	1/1	$-1, +2$	Burgers
$u_{2x} = ku^2$	2	0	2	-2	2/1	$-1, +6$	DT of Painlevé I
$u_{3x} = ku^4$	3	0	4	-1	3/3	complex	
$u_{3x} = ku^2u_x$	3	1	3	-1	2/2	$-1, +3, +4$	Modified KdV
$u_{3x} = k(u_x^2 + uu_{2x})$	3	2	2	-1	1/1	$-1, +2+3$	
$u_{3x} = kuu_x$	3	1	2	-2	2/1	$-1, +4, +6$	Korteweg–de Vries
$u_{3x} = ku^2$	3	0	2	-3	3/1	complex	
$u_{4x} = ku^5$	4	0	5	-1	4/4	complex	
$u_{4x} = ku^3u_x$	4	1	4	-1	3/3	complex	
$u_{4x} = k(2uu_x^2 + u^2u_{2x})$	4	2	3	-1	2/2	$-1, +3, +4, +4$	
$u_{4x} = k(uu_{3x} + 3u_xu_{2x})$	4	3	2	-1	1/1	$-1, +2, +3, +4$	
$u_{4x} = ku^3$	4	0	3	-2	4/2	complex	
$u_{4x} = k(u_x^2 + uu_{2x})$	4	2	2	-2	2/1	$-1, +4, +5, +6$	Boussinesq
$u_{4x} = kuu_x$	4	1	2	-3	3/1	complex	
$u_{4x} = ku^2$	4	0	2	-4	4/1	complex	

[a] DT = dominant truncation.

tiations suggest ratios between the coefficients of the terms on the right-hand side as given in Tables I and II, which may have special significance.

Table II also indicates the relationships of the Painlevé chains to the evolution equations. Each of the three Painlevé chains contains at least one of the evolution equations. The 2/1 chain is the most important one in this connection since it contains both the Korteweg–de Vries and Boussinesq equations. The higher-order equations in the three Painlevé chains of Table II are interesting candidates for detailed future study, since some of them are possibilities for new equations solvable by the inverse scattering transform or related methods.

Note from Tables I and II that the 2/2 and 1/1 chains have the same exponent p, namely -1. The homogeneous equations of a given order in these chains can be mixed to give a hybrid equation that is still self-dominant. Similar mixing of homogeneous differential equations of a given order and fractions 4/2 and 2/1 (i.e., $p = -2$) can also give hybrid self-dominant equations; this latter type of mixing is important in the study of fifth-order evolution equations as discussed below. Note that mixing a 4/2 equation with a 2/1 equation of the same order can give a hybrid equation with the integral resonances required for the Painlevé property even though a pure 4/2 equation cannot have the Painlevé property since it has complex resonances rather than integral resonances.

Let us consider hybridization of the third-order equations $u_{3x} = k(u_x^2 + uu_{2x})$ (fraction 1/1) and $u_{3x} = k'u^2u_x$ (fraction 2/2); the latter is the dominant truncation of the

TABLE II. Painlevé chains of self-dominant homogeneous differential equations with integral resonances.

2/2 chain						
$u_{2x} = ku^3$ $(+4)$ Dominant truncation of Painlevé II	$\rightarrow$	$u_{3x} = ku^2u_x$ $(+3, +4)$ Modified KdV	$\rightarrow$	$u_{4x} = k(2uu_x^2 + u^2u_{2x})$ $(+3, +4, +4)$	$\rightarrow$	$u_{5x} = k(2u_x^3 + 6uu_xu_{2x} + u^2u_{2x})$ $(+3, +4, +4, +5)$
1/1 chain						
$u_{2x} = kuu_x$ $(+2)$ Burgers	$\rightarrow$	$u_{3x} = k(uu_{2x} + u_x^2)$ $(+2, +3)$	$\rightarrow$	$u_{4x} = k(uu_{3x} + 3u_xu_{2x})$ $(+2, +3, +4)$	$\rightarrow$	$u_{5x} = k(uu_{4x} + 4u_xu_{3x} + 3u_{2x}^2)$ $(+2, +3, +4, +5)$
2/1 chain						
$u_{2x} = ku^2$ $(+6)$ Dominant truncation of Painlevé I	$\rightarrow$	$u_{3x} = kuu_x$ $(+4, +6)$ Korteweg–de Vries	$\rightarrow$	$u_{4x} = k(u_x^2 + uu_{2x})$ $(+4, +5, +6)$ Boussinesq	$\rightarrow$	$u_{5x} = k(3u_xu_{2x} + uu_{3x})$ $(+4, +5, +6, +6)$

TABLE III. Painlevé analysis of the third-order hybrid equation $u_{3x} = u_x^2 + uu_{xx} + hu^2u_x$.

h	a	Resonances	Properties
2/3	6	$-1, +3, +10$	good
2/3	$-3/2$	$-1, +3, +5/2$	bad: nonintegral root
3	2	$-1, +3, +6$	good
3	-1	$-1, +3, +3$	bad: double root
9	1	$-1, +3, +5$	good
9	$-2/3$	$-1, +3, +10/3$	bad: nonintegral root
$-1/3$	-3	$-1, +1, +3$	good
$-1/3$	-6	$-2, -1, +3$	satisfactory

modified (cubic) KdV equation. If the 1/1 coefficient ratio of the u_x^2 and uu_{2x} terms is preserved, the resulting hybrid can be expressed in the following form:

$$u_{3x} = u_x^2 + uu_{2x} + hu^2u_x. \tag{24}$$

Balancing the $(x - x^*)^{-4}$ terms according the Painlevé procedure gives the following quadratic equation for a:

$$a^2 - 3a/h - 6/h = 0. \tag{25}$$

Therefore

$$a = 3/2h \pm (1/2h)\sqrt{9 + 24h}. \tag{26}$$

Thus for any value of $h > -\frac{3}{8}$, a has two values indicating two solution branches. Such multiple solution branches are a characteristic of hybrid equations.

Table III illustrates the Painlevé analysis for the hybrid third-order differential equation (24) using as examples several values of h leading to rational values of a. In three of the four cases one of the two solution branches has a "good" set of resonances for the Painlevé property (resonances at -1 and two other distinct integers including at least one positive integer) > -1 and the other branch has a "flaw" in its set of resonances (nonintegral resonance or a double root). Note, however, that regardless of the values of h and a, two of the three resonances in Eq. (24) appear at -1 and $+3$, which are the two resonances possessed by *both* components of the hybrid, namely $u_{3x} = k(u_x^2 + uu_{2x})$.

Table IV summarized the Painlevé analysis for the three homogeneous fifth-order differential equations that are members of the three Painlevé chains in Table II as well as three hybrid fifth-order differential equations that have been studied as evolution equations. All three of the hybrid systems are mixtures of the 2/1 quadratic and 4/2 cubic fifth-order differential equations. The following points about these fifth-order hybrid differential equations are of interest.

(1) In contrast to the third-order differential equations of Table III the ratios between the coefficients of the u_xu_{2x} and uu_{3x} terms of the 2/1 components (designated as r_{12}/r_{03}) are different for the hybrid systems than for the pure homogeneous fifth-order 2/1 equation generated by the successive differentiations implied by Eq. (23). Thus for the homogeneous systems r_{12}/r_{03} is 3 whereas for the hybrid systems r_{12}/r_{03} is 1, 5/2, and 2 for the Caudrey–Dodd–Gibbon, the Kuperschmidt, and the higher-order Korteweg–de Vries equations, respectively. The hybrid systems may therefore be viewed as being generated from the pure 2/1 system by perturbing the ratio r_{12}/r_{03} from that generated for the pure system by differentiation (namely 3) and then mixing in enough of the 4/2 equation (i.e., $u_{5x} = k'u^2u_x$) to restore the integral resonances required for the Painlevé property.

(2) For reasons noted above the fifth-order 2/1 + 4/2 hybrid equations have two solution branches arising from the two roots of a quadratic equation analogous to Eq. (25). In all three cases one solution branch has all integers greater than the mandatory -1 resonance whereas the other solution branch has one resonance below -1 in addition to three distinct integral resonances greater than -1.

V. SUMMARY

This paper demonstrates a simple way of classifying higher-order differential equations based on the requirements of the Painlevé property related to the solubility of the equation by inverse scattering transform methods.[1–3] As expected, the known evolution equations such as the Korteweg–de Vries, Burger, Boussinesq, and Caudrey–

TABLE IV. Some fifth-order strongly self-dominant differential equations with integral resonances.

Equation type	Co-order m	Degree d	Exponent p	Fraction $(n-m)/(d-1)$	Resonances[a]	Reference
(A) Homogeneous equations ($n = 5$)						
$u_{5x} = k(2u_x^3 + 6uu_xu_{2x} + u^2u_{3x})$	3	3	-1	2/2	$-1, +3, +4, +4, +5$	
$u_{5x} = k(uu_{4x} + 4u_xu_{3x} + 3u_{2x}^2)$	4	2	-1	1/1	$-1, +2, +3, +4, +5$	
$u_{5x} = k(3u_xu_{2x} + uu_{3x})$	3	2	-2	2/1	$-1, +4, +5, +6, +6$	
(B) Hybrid equations ($n = 5$)						
$u_{5x} = -30u_xu_{2x} - 30uu_{3x} - 180u^2u_x$	3,1	2,3	-2	2/1 + 4/2	$-1, +2, +3, +6, +10$ $-2, -1, +5, +6, +12$	b
$u_{5x} = -25u_xu_{2x} - 10uu_{3x} - 20u^2u_x$	3,1	2,3	-2	2/1 + 4/2	$-1, +3, +5, +6, +7$ $-7, -1, +6, +10, +12$	c
$u_{5x} = -20u_xu_{2x} - 10uu_{3x} - 30u^2u_x$	3,1	2,3	-2	2/1 + 4/2	$-1, +2, +5, +6, +8$ $-3, -1, +6, +8, +10$	d

[a] The resonance of both solution branches are listed for the hybrid systems.
[b] The Caudrey–Dodd–Gibbon equation: P. J. Caudrey, R. K. Dodd, and J. B. Gibbon, Proc. R. Soc. London, Ser. A **351**, 407 (1976).
[c] The Kuperschmidt equation: A. P. Fordy and J. Gibbons, Phys. Lett. A **75**, 325 (1980).
[d] A higher-order Korteweg–de Vries equation: H. Morris, J. Math. Phys. **18**, 530 (1977).

Dodd–Gibbon equations occupy prominent positions in this classification scheme. This classification scheme also identifies new potential candidates for higher-order differential equations with the Painlevé property and possibly soluble by inverse scattering transform methods. A major objective of this paper is to stimulate further work, which hopefully will relate the ideas in this paper to such important aspects in the study of evolution equations as the generation of Lax pairs,[11,12] conservation laws,[1,2] Bäcklund transformations,[12–16] recursion relations,[6,12] Schwarzian derivatives,[12,14–17] and prolongation structures[18,19] as well as details of the inverse scattering transform procedure.[1–3]

ACKNOWLEDGMENT

I am indebted to the Office of Naval Research for partial support of this research.

[1]G. L. Lamb, Jr., *Elements of Soliton Theory* (Wiley–Interscience, New York, 1980).

[2]F. Calogero and A. Degasperis, *Spectral Transform and Solitons* (North-Holland, Amsterdam, 1982).

[3]M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Stud. Appl. Math. **53**, 249 (1974).

[4]M. J. Ablowitz and H. Segur, Phys. Rev. Lett. **38**, 1103 (1977).

[5]M. J. Ablowitz, A. Ramani, and H. Segur, J. Math. Phys. **21**, 715 (1980).

[6]J. Weiss, M. Tabor, and G. Carnevale, J. Math. Phys. **24**, 522 (1983).

[7]E. L. Ince, *Ordinary Differential Equations* (Longmans, Green, London, 1927), Chap. XIV.

[8]E. D. Rainville, *Intermediate Differential Equations* (Macmillan, New York, 1964).

[9]G. Sansone and J. Gerretsen, *Lectures on the Theory of Functions of a Complex Variable* (Wolters–Noordhoff, Groningen, 1969).

[10]N. B. Conkwright, *Introduction to the Theory of Equations* (Ginn, Boston, 1941), Chap. X.

[11]P. Lax, Commun. Pure Appl. Math. **21**, 467 (1968).

[12]J. Weiss, J. Math. Phys. **24**, 1405 (1983).

[13]R. Sasaki, Nucl. Phys. B **154**, 343 (1979).

[14]J. Weiss, Phys. Lett. A **102**, 329 (1984).

[15]J. Weiss, Phys. Lett. A **105**, 387 (1984).

[16]J. Weiss, J. Math. Phys. **26**, 258 (1985).

[17]J. Weiss, J. Math. Phys. **25**, 13 (1984).

[18]H. D. Wahlquist and F. B. Estabrook, J. Math. Phys. **16**, 1 (1975).

[19]R. K. Dodd and J. D. Gibbon, Proc. R. Soc. London Ser. A **359**, 411 (1978).

Symmetry transformations, isovectors, and conservation laws

Giacomo Caviglia
Istituto Matematico dell'Università, Via Leon Battista Alberti 4, 16132 Genova, Italy

(Received 24 June 1985; accepted for publication 5 December 1985)

A system of second-order partial differential equations is considered. It is shown that a conservation law may be associated with any pair consisting of (i) a symmetry transformation, (ii) a symmetry transformation if the system is self-adjoint, a solution to the adjoint of the equations of variation otherwise. Such conservation laws continue to hold when symmetry transformations are replaced by isovectors. It is also proved that isovectors identify additional conservation laws by deformation of a given one, whence it follows that there exists a natural conserved current associated with every isovector. Also an application is made to find conserved currents for the Navier–Stokes equations.

I. INTRODUCTION

The subject of symmetries and of the associated conservation laws has been extensively analyzed within the framework of continuum mechanics and field theories, in view of its practical as well as theoretical relevance. Actually, Noether's theorem in its various formulations provides the most basic tool for the derivation of conservation laws based on invariance properties of Hamilton's variational principle.[1] However, it appears that Noether's theorem cannot be applied in many practical circumstances, because the equations modeling the evolution of a large variety of physical systems do not admit any natural variational formulation.[2,3] In addition, it is also known that in general the class of symmetry transformations of the differential equations of evolution is larger than the class of symmetry transformations of the action functional.[4] These drawbacks have motivated further investigations on the relationships between invariance properties and conservation laws, aiming at suitable extensions of Noether's theorem.

Although much discussion has already been devoted to this topic, it seems that some points need further clarification. Accordingly, this paper is primarily concerned with the description of systematic procedures for finding conserved currents by making use of the symmetry transformations and the isovectors of the given field equations. In so doing we also will extend to continuous media and field theories some recent results on the existence of invariants for systems in a finite number of degrees of freedom.

The first step in our approach consists in the definition of the concept of symmetry transformation, regarded as an infinitesimal transformation leaving invariant a given set of partial differential equations (Sec. II). The connections between the formal properties of the equations defining symmetry transformations on the one hand, and the solvability of the inverse problem of the calculus of variations for the given set of evolution equations on the other, are then reviewed in some detail. Indeed, besides giving a deeper insight into the concept of symmetry transformation, this analysis constitutes the starting point for the subsequent construction of conservation laws in terms of the Green's formulas (Sec. III).

Next we turn our attention to a geometric formulation of the given field equations. Within this context it is found that symmetry transformations are nothing but particular examples of isovectors; nevertheless, the conservation laws generated by symmetry transformations maintain their validity even when symmetry transformations are replaced by generic isovectors. It is also shown that isovectors give rise to conservation laws by "deformation" of a given one and, in addition, every isovector identifies a canonically associated conserved current (Sec. IV).

As an outstanding application, we make use of recent results on the group properties of the Navier–Stokes equations,[5,6] in order to find corresponding conservation laws (Sec. V). Additional comments on these results can be found in Sec. VI.

II. SYMMETRY TRANSFORMATIONS

Consider a system of m partial differential equations with m unknown functions ϕ^c depending on the variables $x^1,\dots,x^n$, namely

$$F_a(x^\alpha,\phi^c,\phi^c_\alpha,\phi^c_{\alpha\beta})=0\,, \tag{2.1}$$

where Greek (Latin) indices vary from 1 to n (m); $\phi^c_\alpha=\partial\phi^c/\partial x_\alpha$, and $\phi^c_{\alpha\beta}=\partial^2\phi^c/\partial x^\alpha\,\partial x^\beta$. An infinitesimal transformation of the form

$$\bar{x}^\alpha=x^\alpha+\epsilon\tau^\alpha(x^\beta,\phi^b)\,, \tag{2.2}$$

$$\bar{\phi}^c=\phi^c+\epsilon\xi^c(x^\beta,\phi^b) \tag{2.3}$$

is said to be a *symmetry transformation* iff it leaves the evolution equations invariant up to terms of order ϵ^2.

To find the corresponding restrictions on the functions τ^α and ξ^c, it is to be observed that

$$\bar{\phi}^c_\alpha=\phi^c_\alpha+\epsilon(D_\alpha\xi^c-\phi^c_\beta\,D_\alpha\tau^\beta)+o(\epsilon^2)\,, \tag{2.4}$$

$$\bar{\phi}^c_{\alpha\beta}=\phi^c_{\alpha\beta}+\epsilon(D_{\alpha\beta}\xi^c-\phi^c_{\alpha\sigma}\,D_\beta\tau^\sigma-\phi^c_{\beta\sigma}\,D_\alpha\tau^\sigma-\phi^c_\sigma\,D_{\alpha\beta}\tau^\sigma)+o(\epsilon^2)\,, \tag{2.5}$$

with

$$\bar{\phi}^c_\alpha=\frac{\partial\bar{\phi}^c}{\partial\bar{x}^\alpha},\quad \bar{\phi}^c_{\alpha\beta}=\frac{\partial^2\bar{\phi}^c}{\partial\bar{x}^\alpha\,\partial\bar{x}^\beta},\quad D_{\alpha\beta}=D_\alpha\circ D_\beta\,,$$

where D_α denotes the total derivative with respect to x^α; in particular D_α reduces to $\partial/\partial x^\alpha$ whenever it acts on functions of the independent variables x^β. Thus the transforma-

 0022-2488/86/040972-07$02.50

tion (2.2)–(2.5) leaves (2.1) invariant iff the generators τ^α and ξ^c satisfy

$$\tau^\alpha \frac{\partial F_a}{\partial x^\alpha} + \xi^c \frac{\partial F_a}{\partial \phi^c} + (D_\alpha \xi^c - \phi^c_\sigma D_\alpha \tau^\sigma) \frac{\partial F_a}{\partial \phi^c_\alpha}$$
$$+ (D_{\alpha\beta}\xi^c - \phi^c_{\alpha\sigma} D_\beta \tau^\sigma - \phi^c_{\beta\sigma} D_\alpha \tau^\sigma$$
$$- \phi^c_\sigma D_{\alpha\beta}\tau^\sigma) \frac{\partial F_a}{\partial \phi^c_{\alpha\beta}} = 0\,. \qquad (2.6)$$

Symmetry transformations can be used to find similarity solutions for the given equations.[6,7] Presumably, this is the reason why, e.g., the whole set of such symmetry generators for the heat equation and the Navier–Stokes equations have been explicitly computed.[5–8]

Perhaps, more familiar examples of symmetry transformations can be found under the assumption that the set of independent variables is one-dimensional—specifically we let x^1 represent the time variable t—and the ϕ^c's are identified with the Lagrangian coordinates q^c of a mechanical system in n degrees of freedom. More precisely, if the equations of motion are given in the normal form

$$\ddot{q}^a - \Lambda^a(t,q,\dot{q}) = 0\,, \qquad (2.7)$$

where $\cdot$ denotes the total time derivative, then (2.6) may be written as

$$-\left[\tau \frac{\partial}{\partial t} + \xi^c \frac{\partial}{\partial q^c} + (\dot{\xi}^c - \dot{q}^c \dot{\tau}) \frac{\partial}{\partial \dot{q}^c}\right] \Lambda^a$$
$$+ (\dot{\xi}^a - \dot{q}^a \dot{\tau})^{\cdot} - \dot{\tau}\Lambda^a = 0\,. \qquad (2.8)$$

Suppose now that the generators $\tau^1 = \tau$ and ξ^c are allowed to depend on the generalized velocities $\dot{q}^c$. In this connection it is immediately recognized that (2.8) is merely the definition of *dynamical symmetry* and that the operator between square brackets in (2.8) is a dynamical symmetry of the system (2.7).[9,10] We have thus established the existence of a canonical correspondence between velocity dependent symmetry transformations and dynamical symmetries, from which it follows in particular that several examples of such symmetry transformations are already available.[11,12]

III. CONSERVATION LAWS

The approach to conservation laws through symmetry transformations developed in the present section requires a preliminary interpretation of symmetry transformations as generators of solutions to the equations of variation of the system (2.1). To achieve this result, we observe that the total derivative of (2.1) with respect to x^α yields

$$\frac{\partial F_a}{\partial x^\alpha} + \phi^c_\alpha \frac{\partial F_a}{\partial \phi^c} + \phi^c_{\alpha\beta} \frac{\partial F_a}{\partial \phi^c_\beta} + \phi^c_{\alpha\beta\lambda} \frac{\partial F_a}{\partial \phi^c_{\beta\lambda}} = 0\,. \qquad (3.1)$$

Thus, substitution into (2.6) of the expression for $\partial F_a/\partial x^\alpha$, which is obtained from (3.1), leads to

$$- \tau^\alpha(a_{ac}\,\phi^c_\alpha + b^{\,\beta}_{ac}\,\phi^c_{\alpha\beta} + c^{\beta\lambda}_{ac}\,\phi^c_{\alpha\beta\lambda}) + a_{ac}\xi^c$$
$$+ b^{\,\alpha}_{ac}(D_\alpha\xi^c - \phi^c_\sigma D_\alpha\tau^\sigma) + c^{\alpha\beta}_{ac}(D_{\alpha\beta}\xi^c - \phi^c_{\alpha\sigma} D_\beta\tau^\sigma$$
$$- \phi^c_{\beta\sigma} D_\alpha \tau^\sigma - \phi^c_\sigma D_{\alpha\beta}\tau^\sigma) = 0\,, \qquad (3.2)$$

where we have introduced the same notations as in Refs. 3 by letting

$$a_{ac} = \frac{\partial F_a}{\partial \phi^c}\,, \quad b^{\,\beta}_{ac} = \frac{\partial F_a}{\partial \phi^c_\beta}\,, \quad c^{\alpha\beta}_{ac} = \frac{\partial F_a}{\partial \phi^c_{\alpha\beta}}\,. \qquad (3.3)$$

It is straightforward to verify that (3.3) may be written in the equivalent form

$$M_a(\eta) = a_{ac}\eta^c + b^{\,\alpha}_{ac} D_\alpha \eta^c + c^{\alpha\beta}_{ab} D_{\alpha\beta}\eta^c = 0\,, \qquad (3.4)$$

where

$$\eta^c = \xi^c - \tau^\alpha \phi^c_\alpha\,. \qquad (3.5)$$

Equations (3.4) have been referred to as the *equations of variation*[3] for the original system (2.1). In this respect, our analysis has shown that any symmetry transformation gives rise to a solution $\eta^c(x^\alpha)$ of the equations of variation (3.4), for every field $\phi^c(x^\alpha)$ satisfying the given evolution equations. To simplify, symmetry transformations can be looked at as generators of solutions to the equations of variation.

In order to get some feeling as to the meaning of this result, it is somewhat instructive to examine a particular case: namely, suppose that (2.1) are given in the form of the Euler–Lagrange equations, that is

$$F_a(x^\alpha,\phi^c,\phi^c_\alpha,\phi^c_{\alpha\beta}) = D_\alpha\left(\frac{\partial L}{\partial \phi^a_\alpha}\right) - \frac{\partial L}{\partial \phi^a} = 0\,, \qquad (3.6)$$

where $L(x^\alpha,\phi^c,\phi^c_\alpha)$ is a suitable Lagrangian density. Then (3.4) can be shown to read

$$\left[D_\alpha\left(\frac{\partial^2 L}{\partial \phi^a_\alpha\,\partial\phi^c}\right) - \frac{\partial^2 L}{\partial\phi^a\,\partial\phi^c}\right]\eta^c$$
$$+ \left[D_\alpha\left(\frac{\partial^2 L}{\partial\phi^a_\alpha\,\partial\phi^c_\beta}\right) + \frac{\partial^2 L}{\partial\phi^a_\beta\,\partial\phi^c} - \frac{\partial^2 L}{\partial\phi^a\,\partial\phi^c_\beta}\right] D_\beta\eta^c$$
$$+ \frac{\partial^2 L}{\partial\phi^a_\alpha\,\partial\phi^c_\beta} D_{\alpha\beta}\eta^c = 0\,, \qquad (3.7)$$

and usually are known as the Jacobi equations for the given functional. Therefore, any symmetry transformation of (3.6) is related to a Jacobi field by means of (3.5). Indeed, when we consider the case of a single independent variable, Eq. (3.7) describes an already known property of dynamical symmetries—i.e., of symmetry transformations—which is at the very origin of a procedure leading to the construction of constants of motion.[10,11] Needless to say, in the practical search for conservation laws all attempts are worth trying; in particular it is appropriate to ask whether it is possible to find conservation laws arising from Jacobi fields—or symmetry transformations—under the assumption that the evolution equations involve several independent variables. Clearly, the answer to this question should be found by an extension of the approach that is known to work for mechanical systems in a finite number of degrees of freedom,[10] and consequently it should basically depend on the formal properties of (3.4).

More precisely, consider the adjoint system to $M_a(\eta)$, say $\widetilde{M}_a(\tilde{\eta})$, which is defined by the condition

$$\tilde{\eta}^a M_a(\eta) - \eta^a \widetilde{M}_a(\tilde{\eta}) = D_\alpha J^\alpha\,, \qquad (3.8)$$

so that the explicit form of the adjoint system is

$$\widetilde{M}_a(\tilde{\eta}) = \tilde{\eta}^c a_{ca} - D_\alpha(\tilde{\eta}^c\, b^{\,\alpha}_{ca}) + D_{\alpha\beta}(\tilde{\eta}^c\, c^{\alpha\beta}_{ca})\,, \qquad (3.9)$$

and the current J^α reads

$$J^{\alpha} = \tilde{\eta}^a b^{\ \alpha}_{ac}\eta^c + \tilde{\eta}^a c^{\alpha\beta}_{ac} D_\beta \eta^c - [D_\beta(\tilde{\eta}^a c^{\beta\alpha}_{ac})]\eta^c . \tag{3.10}$$

The linear system $M_a(\eta) = 0$ is said to be *self-adjoint* whenever it coincides with its adjoint system $\widetilde{M}_a(\eta) = 0$, that is

$$M_a(\eta) = \widetilde{M}_a(\eta) , \tag{3.11}$$

for all admissible η. In that case, the evolution equations (2.1) are also termed self-adjoint.

Substitution of the expressions (3.4) and (3.9) for M_a and $\widetilde{M}_a$ into (3.11) shows that (3.11) is mathematically equivalent to

$$c^{\alpha\beta}_{ac} = c^{\beta\alpha}_{ac} = c^{\alpha\beta}_{ca} = c^{\beta\alpha}_{ca} , \tag{3.12a}$$

$$b^{\ \alpha}_{ac} + b^{\ \alpha}_{ca} = D_\beta c^{\beta\alpha}_{ca} + D_\beta c^{\alpha\beta}_{ca} = 2 D_\beta c^{\alpha\beta}_{ca} , \tag{3.12b}$$

$$a_{ac} - a_{ca} = D_{\alpha\beta} c^{\alpha\beta}_{ca} - D_\beta b^{\ \beta}_{ca} . \tag{3.12c}$$

A rather long but straightforward calculation shows that the Euler–Lagrange equations (3.6) are self-adjoint[3]; namely, conditions (3.12) hold for the Jacobi equation (3.7), it being understood that the coefficients a_{ac}, $b^{\ \alpha}_{ac}$, and $c^{\alpha\beta}_{ac}$ are obtained by comparison of (3.7) with (3.4).

It is the most fundamental result within the analysis of the inverse problem of the calculus of variations that the converse of the last statement is also true. Specifically, the self-adjointness of the evolution equations (2.1) is also a sufficient condition for the existence of a Lagrangian density L such that the representation of the system (2.1) in terms of Euler–Lagrange equations as in (3.6) is allowed.[2,3] Thus the self-adjointness of the equations for symmetry transformations, written down in the form (3.4), ensures the existence of a Lagrangian formulation for (2.1).

The previous remarks on the formal properties of the equations of variation and on the relationships between symmetry generators and solutions of the equations of variation will now be used to construct conservation laws. To this aim, suppose that (3.4) are self-adjoint and consider any two symmetry transformations generated by $(\tau^\alpha_1, \xi^{\ c}_1)$ and $(\tau^\alpha_2, \xi^{\ c}_2)$, respectively. Defining η^c_1 and η^c_2 as in (3.5), it is found immediately that $M_a(\eta_1) = \widetilde{M}_a(\eta_2) = 0$, provided (3.11) is also taken into account. Then it follows from (3.8) that $D_\alpha J^\alpha = 0$, so that the current density J^α given by (3.10) is conserved. We have thus established the following theorem.

Theorem 3.1: Suppose that the system (2.1) is self-adjoint. Then to every pair of symmetry transformations $(\tau^\alpha_i, \xi^{\ c}_i)$, with $i = 1,2$, there corresponds a conserved current of the form

$$\begin{aligned} J^\alpha &= (\xi^{\ a}_2 - \tau^\beta_2 \phi^a_\beta)\frac{\partial F_a}{\partial \phi^c_\alpha}(\xi^{\ c}_1 - \tau^\gamma_1 \phi^c_\gamma) \\ &\quad + (\xi^{\ a}_2 - \tau^\beta_2 \phi^a_\beta)\frac{\partial F_a}{\partial \phi^c_{\alpha\sigma}} D_\sigma(\xi^{\ c}_1 - \tau^\gamma_1 \phi^c_\gamma) \\ &\quad - D_\sigma\left[(\xi^{\ a}_2 - \tau^\beta_2 \phi^a_\beta)\frac{\partial F_a}{\partial \phi^c_{\sigma\alpha}}\right](\xi^{\ c}_1 - \tau^\gamma_1 \phi^c_\gamma) . \end{aligned} \tag{3.13}$$

Notice that (3.13) is simply a reformulation of the definition (3.10), where, in particular, the coefficients $b^{\ \alpha}_{ac}$ and $c^{\alpha\beta}_{ac}$ have been substituted by their expressions in terms of the given F_a's. Of course, if the Lagrangian density yielding the evolution equations is explicitly given, then the form of the above coefficients can be directly inferred from Eq. (3.7).

It is also to be remarked that (3.8) can be used to generate conserved currents even when the system (2.1) is not self-adjoint. In that case the construction of a conservation law is achieved by means of a pair η^c, $\tilde{\eta}^c$ where η^c is a solution to (3.4) generated by a symmetry transformation through (3.5), whereas $\tilde{\eta}^c$ is any solution to $\widetilde{M}_a(\tilde{\eta}) = 0$. Therefore we have the following result.

Thereom 3.2: Every pair consisting of a symmetry transformation $(\tau^\alpha, \xi^{\ c})$ and of a solution $\tilde{\eta}^a$ to the adjoint of the equations of variation identifies a conserved current of the form

$$\begin{aligned} J^\alpha &= \tilde{\eta}^a \frac{\partial F_a}{\partial \phi^c_\alpha}(\xi^{\ c} - \tau^\gamma \phi^c_\gamma) + \tilde{\eta}^a \frac{\partial F_a}{\partial \phi^c_{\alpha\beta}} D_\beta(\xi^{\ c} - \tau^\gamma \phi^c_\gamma) \\ &\quad - D_\beta\left(\tilde{\eta}^a \frac{\partial F_a}{\partial \phi^c_{\beta\alpha}}\right)(\xi^{\ c} - \tau^\gamma \phi^c_\gamma) . \end{aligned} \tag{3.14}$$

An outstanding application of this result is discussed in Sec. V, where we also determine solutions to the adjoint equation $\widetilde{M}_a(\tilde{\eta}) = 0$ generated by a set $\tilde{\eta}^a(x^\alpha, \phi^c)$, by restriction of ϕ^c to a solution of the evolution equations, specifically the Navier–Stokes equation.

IV. SYMMETRY TRANSFORMATIONS, ISOVECTORS, AND CONSERVATION LAWS

In this section we propose a geometric interpretation of symmetry transformations as isovector fields. Recalling the comments at the end of Sec. II, this implies in particular that the dynamical symmetries of a system of ordinary differential equations of the form (2.7) may be regarded as isovectors. We will also discuss the role of isovectors as generators of conservation laws, thus extending well-known properties of dynamical symmetries.

Our approach is based on a preliminary reformulation of the system (2.1) in terms of exterior forms.[13–15] To this end we consider a formal differentiable manifold M referred to local coordinates $(x^\alpha, \phi^c, Z^{\ c}_\alpha, W^{\ c}_{\alpha\beta})$, where the following differential forms have been defined:

$$F_a(x^\alpha, \phi^c, Z^{\ c}_\alpha, W^{\ c}_{\alpha\beta}) , \tag{4.1}$$

$$d\phi^c - Z^{\ c}_\alpha \, dx^\alpha , \tag{4.2}$$

$$dZ^{\ c}_\alpha - W^{\ c}_{\alpha\beta} \, dx^\beta , \tag{4.3}$$

$$\frac{\partial F_a}{\partial x^\alpha} dx^\alpha + \frac{\partial F_a}{\partial \phi^c} d\phi^c + \frac{\partial F_a}{\partial Z^{\ c}_\alpha} dZ^{\ c}_\alpha + \frac{\partial F_a}{\partial W^{\ c}_{\alpha\beta}} dW^{\ c}_{\alpha\beta} , \tag{4.4}$$

$$dZ^{\ c}_\alpha \wedge dx^\alpha , \tag{4.5}$$

$$dW^{\ c}_{\alpha\beta} \wedge dx^\beta . \tag{4.6}$$

Then we observe that every solution $\phi^c = \phi^c(x^\alpha)$ to (2.1) identifies a submanifold of M given by the map Φ: $(x^\alpha) \to (x^\alpha, \phi^c, Z^{\ c}_\alpha, W^{\ c}_{\alpha\beta})$, where $\phi^c = \phi^c(x^\alpha)$; $Z^{\ c}_\alpha = \phi^c_\alpha$; $W^{\ c}_{\alpha\beta} = \phi^c_{\alpha\beta}$. The map Φ is such that the restriction to the submanifold of any of the forms (4.1)–(4.6) does identically vanish. Equivalently, it may be asserted that the map Φ^* annihilates the forms (4.1)–(4.6), whence it follows that the map Φ yields an n-dimensional integral manifold of the given exterior system.

Conversely, every n-dimensional integral manifold of the system (4.1)–(4.6) gives rise to a solution for Eq. (2.1), provided the x^α's may be taken as independent variables. In such a case (4.2) and (4.3) yield $Z^c_\alpha = \phi^c_\alpha$ and $W^c_{\alpha\beta} = \phi^c_{\alpha\beta}$, respectively, so that (4.1) reduces to the left-hand side of (2.1).

Recalling now that the vanishing of the forms (4.1)–(4.6) is mathematically equivalent to the vanishing of the ideal I of the ring of forms of M generated by (4.1)–(4.6), we obtain the required geometric reformulation of the given system (2.1), which consists in finding integral manifolds of I described in terms of the independent variables x^α. It is to be remarked that I is closed by construction, which means that the exterior derivative of each form of I still belongs to I; therefore, sufficient conditions for the existence of integral manifolds with fixed independent variables are given by the Cartan–Kähler theorem.[13,14] Throughout our analysis it will always be assumed that such requirements are fulfilled. Notice also that simplified versions of this approach are also allowed, provided (2.1) are explicitly given in a suitable form.[7,16,17]

Consider now a vector field Y on M and denote by $\mathscr{L}_Y$ the Lie derivation operator along Y. Then Y is said to be an *isovector* of I iff[7,14,17]

$$\mathscr{L}_Y \omega \in I, \tag{4.7}$$

for every $\omega \in I$. Definition (4.7) implies that the ideal I is invariant under the action of the local one-parameter group of differentiable transformations generated by the vector Y. Then, recalling that the ideal I geometrizes the system of partial differential equations (2.1), it turns out that (4.7) geometrizes the concept of invariance transformations for the given equations.

To prove that Y is an isovector of I it suffices to show that condition (4.7) holds for a set of generators of I, because the Lie derivative acts as a derivation operator on the exterior product of forms. Moreover, from the fact that $\mathscr{L}_Y d\omega = d\mathscr{L}_Y\omega$ for every form ω, it follows that Y is an isovector of I iff the Lie derivatives of (4.1)–(4.3) in the direction of Y belong to I.

The last statement may be used to find a local characterization of isovectors that allows a close examination of the relationships between isovectors and symmetry transformations. Specifically, consider a vector Y locally represented as

$$Y = \tau^\alpha \frac{\partial}{\partial x^\alpha} + \xi^c \frac{\partial}{\partial \phi^c} + \eta^c_\alpha \frac{\partial}{\partial Z^c_\alpha} + \lambda^c_{\alpha\beta} \frac{\partial}{\partial W^c_{\alpha\beta}}. \tag{4.8}$$

Then, on using the well-known identity $\mathscr{L}_Y\omega = Y \rfloor d\omega + d(Y \rfloor \omega)$, where $\rfloor$ denotes the interior product, it may be shown that

$$\mathscr{L}_Y(d\phi^c - Z^c_\alpha\, dx^\alpha) = -\eta^c_\alpha\, dx^\alpha + d\xi^c - Z^c_\alpha\, d\tau^\alpha, \tag{4.9}$$

$$\mathscr{L}_Y(dZ^c_\beta - W^c_{\beta\alpha}\, dx^\alpha) = -\lambda^c_{\beta\alpha}\, dx^\alpha + d\eta^c_\beta - W^c_{\beta\sigma}\, d\tau^\sigma, \tag{4.10}$$

$$\mathscr{L}_Y(F_a) = \tau^\alpha \frac{\partial F_a}{\partial x^\alpha} + \xi^c \frac{\partial F_a}{\partial \phi^c} + \eta^c_\alpha \frac{\partial F_a}{\partial Z^c_\alpha} + \lambda^c_{\alpha\beta} \frac{\partial F_a}{\partial W^c_{\alpha\beta}}. \tag{4.11}$$

Let us now impose the supplementary restriction that τ^α and ξ^c depend only on the variables x^α and ϕ^c; then (4.9) and (4.10) can be shown to read

$$\begin{aligned}\mathscr{L}_Y(d\phi^c - Z^c_\alpha\, dx^\alpha) &= (D_\alpha \xi^c - Z^c_\beta D_\alpha \tau^\beta - \eta^c_\alpha) dx^\alpha \\ &\quad + \left(\frac{\partial \xi^c}{\partial \phi^b} - Z^c_\beta \frac{\partial \tau^\beta}{\partial \phi^b}\right)(d\phi^b - Z^b_\alpha\, dx^\alpha),\end{aligned} \tag{4.12}$$

$$\begin{aligned}\mathscr{L}_Y(dZ^c_\beta - W^c_{\beta\alpha}\, dx^\alpha) &= (D_\alpha \eta^c_\beta - W^c_{\beta\sigma} D_\alpha \tau^\sigma - \lambda^c_{\beta\alpha}) dx^\alpha \\ &\quad + \frac{\partial \eta^c_\beta}{\partial Z^b_\lambda}(dZ^b_\lambda - W^b_{\lambda\alpha}\, dx^\alpha) \\ &\quad + \left(\frac{\partial \eta^c_\beta}{\partial \phi^b} - W^c_{\beta\sigma} \frac{\partial \tau^\sigma}{\partial \phi^b}\right)(d\phi^b - Z^b_\alpha\, dx^\alpha),\end{aligned} \tag{4.13}$$

respectively, with $D_\alpha = \partial/\partial x^\alpha + Z^b_\alpha\, \partial/\partial\phi^b + W^b_{\beta\alpha}\, \partial/\partial Z^b_\beta$. On comparing (4.12) and (4.13) with (4.1)–(4.6) it is easily seen that the Lie derivatives of the forms (4.2) and (4.3) belong to I iff their components along dx^α vanish, that is

$$\eta^c_\alpha = D_\alpha \xi^c - Z^c_\beta D_\alpha \tau^\beta, \tag{4.14}$$

$$\lambda^c_{\beta\alpha} = D_\alpha \eta^c_\beta - W^c_{\beta\sigma} D_\alpha \tau^\sigma. \tag{4.15}$$

Equations (4.14) and (4.15) yield the expressions for the components η^c_α and $\lambda^c_{\beta\alpha}$ in terms of ξ^c and τ^α; as a consequence (4.11) reduces to

$$\begin{aligned}\mathscr{L}_Y(F_a) &= \tau^\alpha \frac{\partial F_a}{\partial x^\alpha} + \xi^c \frac{\partial F_a}{\partial \phi^c} + (D_\alpha \xi^c - Z^c_\beta D_\alpha \tau^\beta) \frac{\partial F_a}{\partial Z^c_\alpha} \\ &\quad + (D_\alpha \eta^c_\beta - W^c_{\beta\sigma} D_\alpha \tau^\sigma) \frac{\partial F_a}{\partial W^c_{\beta\alpha}}.\end{aligned} \tag{4.16}$$

Accordingly, we conclude that the field Y is an isovector if the scalar $\mathscr{L}_Y(F_a)$ given by (4.16) belongs to I. This shows that any symmetry transformation may be associated with an isovector through the correspondence, which is clearly indicated by the notation. In practice, (2.6) implies $\mathscr{L}_Y(F_a) = 0$ by means of the identifications $Z^c_\alpha = \phi^c_\alpha$ and $W^c_{\alpha\beta} = \phi^c_{\alpha\beta}$.

In principle, it appears that the class of isovectors may be larger than the set of symmetry transformations,[18] whence it follows that the role of isovectors as generators of conservation laws is to be discussed again.

To this aim, consider the restriction of any isovector Y of the form (4.8) to an admissible integral manifold of the ideal I. Then it follows that $\phi^c = \phi^c(x^\alpha)$, $Z^c_\alpha = \phi^c_\alpha$, and $W^c_{\alpha\beta} = \phi^c_{\alpha\beta}$; in addition, it is found from (4.9) and (4.10) that η^c_α and $\lambda^c_{\alpha\beta}$ are given by (4.14) and (4.15), respectively, where D_α is now to be interpreted as the total derivative with respect to x^α; finally, the restriction to the given integral manifold of the scalar $\mathscr{L}_Y(F_a)$ must vanish, that is

$$\begin{aligned}&\tau^\alpha \frac{\partial F_a}{\partial x^\alpha} + \xi^c \frac{\partial F_a}{\partial \phi^c} + (D_\alpha \xi^c - \phi^c_\beta D_\alpha \tau^\beta) \frac{\partial F_a}{\partial Z^c_\alpha} \\ &\quad + (D_\alpha \eta^c_\beta - \phi^c_{\beta\sigma} D_\alpha \tau^\sigma) \frac{\partial F_a}{\partial W^c_{\beta\alpha}} = 0.\end{aligned} \tag{4.17}$$

Equation (4.17) looks exactly like (2.6), but τ^α and ξ^c are now allowed to depend on ϕ^c_α and $\phi^c_{\alpha\beta}$. Again, (3.1) can be used to cast (4.17) into the form (3.4) with η^c given by

(3.5). Thus Theorems 3.1 and 3.2 can be extended to the case when τ^α and ξ^c are, respectively, $\partial/\partial x^\alpha$ and $\partial/\partial\phi^c$ components of an isovector. In addition, let us also observe that the self-adjointness of the equation, which now characterizes isovectors, i.e., of (3.4), is a necessary and sufficient condition for the existence of a variational formulation.

We shall now proceed to establish conservation laws by deformation of a given one along the direction of an isovector, thus extending a property that is known to hold for dynamical symmetries of Lagrangian systems in a finite number of degrees of freedom.[11] The discussion of this point requires a preliminary reformulation of conservation laws in terms of differential forms, that may be reviewed as follows.[17]

Consider the volume form $\mu = dx^1 \wedge \cdots \wedge dx^n$ and define the $(n-1)$-forms μ_α by the condition

$$\mu_\alpha = \frac{\partial}{\partial x^\alpha} \lrcorner\, \mu\,, \tag{4.18}$$

so that the following identities hold:

$$d\mu_\alpha = 0\,, \quad dx^\alpha \wedge \mu_\beta = \delta^\alpha_\beta\, \mu\,. \tag{4.19}$$

The most common approach to conservation laws is described in terms of a field $J^\alpha(x^\beta,\phi^c,\phi^c_\beta)$ satisfying $D_\alpha J^\alpha = 0$ on every solution to (2.1). However one may reformulate the condition of vanishing divergence by noting that the J^α's, regarded as functions defined over M, identify the $(n-1)$-form $J^\alpha\mu_\alpha$, which satisfies the identity

$$\begin{aligned} d(J^\alpha\mu_\alpha) = {} & \left(\frac{\partial J^\alpha}{\partial x^\alpha} + Z^c_\alpha \frac{\partial J_\alpha}{\partial \phi^c} + W^c_{\sigma\alpha}\frac{\partial J^\alpha}{\partial Z^c_\sigma}\right)\mu \\ & + \left[\frac{\partial J^\alpha}{\partial\phi^c}(d\phi^c - Z^c_\beta\, dx^\beta)\right. \\ & \left. + \frac{\partial J^\alpha}{\partial Z^c_\sigma}(dZ^c_\sigma - W^c_{\sigma\beta}\, dx^\beta)\right]\wedge\mu_\alpha \end{aligned} \tag{4.20}$$

in view of (4.19). It follows from (4.20) that $D_\alpha J^\alpha = 0$ on every solution to (2.1) iff $d(J^\alpha\mu_\alpha) = 0$ on every integral manifold of I, i.e., iff $\Phi^*(d(J^\alpha\mu_\alpha)) = 0$. On the basis of this remark we obtain a slight extension of the definition of conservation law proposed in Refs. 16, that is any $(n-1)$-form σ identifies a conservation law if

$$\Phi^*(d\sigma) = 0\,. \tag{4.21}$$

On using (4.21), we can prove the following theorem.

Theorem 4.1: Consider an isovector Y. To any conservation law associated with the $(n-1)$-form σ, there corresponds one additional conservation law associated with $\mathscr{L}_Y\sigma$.

Proof: It is to be shown that $\Phi^*(d\mathscr{L}_Y\sigma) = \Phi^*(\mathscr{L}_Y\, d\sigma)$ vanishes identically. To this aim, consider the definition of Lie derivative, that is[15]

$$\mathscr{L}_Y\, d\sigma = \lim \frac{1}{t}\,[\psi_t^*(d\sigma(\psi_t(P))) - d\sigma(P)]\,,$$

where P is a point of M and ψ_t is the local one-parameter group of diffeomorphisms determined by Y. Then observe that in correspondence with every solution Φ for the given exterior system there exists a whole family of solutions that is given by $\Phi_t = \psi_t\circ\Phi$ (see Refs. 7 and 17). It follows that $\mathscr{L}_Y\, d\sigma$ does vanish when restricted to a solution Φ, because both $d\sigma$ and $d\sigma\circ\psi_t$ vanish on the given solution. Q. E. D.

In particular, suppose that σ coincides with $J^\alpha\mu_\alpha$, where J^α yields a conserved current. Then it turns out that $\mathscr{L}_Y(J^\alpha\mu_\alpha)$ may be written as

$$\begin{aligned} \mathscr{L}_Y(J^\alpha\mu_\alpha) = {} & [Y(J^\alpha) + J^\alpha D_\beta\tau^\beta - J^\beta D_\beta\tau^\alpha]\mu_\alpha \\ & + \left[\frac{\partial\tau^\beta}{\partial\phi^c}(d\phi^c - Z^c_\sigma\, dx^\sigma)\right. \\ & \left. + \frac{\partial\tau^\beta}{\partial\phi^c_\sigma}(dZ^c_\sigma - W^c_{\sigma\lambda}dx^\lambda)\right]\wedge(\partial_\beta \lrcorner\, \mu_\alpha)\,, \end{aligned} \tag{4.22}$$

provided τ^α does not depend on $W^c_{\alpha\beta}$. It follows from (4.22) that the field

$$K^\alpha = Y(J^\alpha) + J^\alpha D_\beta\tau^\beta - J^\beta D_\beta\tau^\alpha \tag{4.23}$$

gives rise to a conserved current. Specific examples will be exhibited in the following section.

As a special application of the previous procedure, we can associate a conserved current with any isovector. To this aim it suffices to observe that the $(n-1)$-form $b^\alpha\mu_\alpha$, where the b^α's are constant coefficients, always gives rise to a conservation law, because $d(b^\alpha\mu_\alpha) = 0$ in view of (4.19). Of course, such a conservation law is trivial, nevertheless, it identifies an additional conservation law in correspondence with every isovector Y, as shown in Theorem 4.1. More precisely, on comparison with (4.23) it turns out that the components of the conserved current read

$$K^\alpha = b^\alpha D_\beta\tau^\beta - b^\beta D_\beta\tau^\alpha\,. \tag{4.24}$$

V. APPLICATIONS TO THE NAVIER–STOKES EQUATIONS

Consider the evolution equations for an incompressible viscous fluid in Cartesian coordinates, that is

$$\mathbf{v}_t + (\mathbf{v}\cdot\nabla)\mathbf{v} = -\nabla p + \nu\nabla^2\mathbf{v}\,, \tag{5.1a}$$

$$\nabla\cdot\mathbf{v} = 0\,, \tag{5.1b}$$

where $\mathbf{v}$ is the velocity field of the fluid, p is the pressure, ν is the constant kinematic viscosity, and ∇ denotes the gradient operator. Through the identifications $(x^1, x^2, x^3, x^4) = (x, y, z, t)$ and $(\phi^1, \phi^2, \phi^3, \phi^4) = (\mathbf{v}, p)$ the system (5.1) may be cast into the form (2.1). However, let us point out explicitly that the previous comments on the solvability of the inverse problem of the calculus of variations no longer apply, because (5.1b) is a first-order partial differential equation. Nevertheless, the procedure aiming at the construction of conservation laws continues to hold, since it depends only on the properties of solutions to the equations of variation and of their adjoint system, which make sense also under the given conditions.

The complete class of symmetry transformations for the Navier–Stokes equations (5.1) has already been explicitly computed in Refs. 5–7. It reads

$$\begin{aligned} (\tau^\alpha) = {} & (a_2x - a_3y - a_4z + f,\ a_2y + a_3x - a_5z + g\,, \\ & a_2z + a_4x + a_5y + h,\ a_1 + 2a_2t)\,; \end{aligned} \tag{5.2}$$

$$\begin{aligned} (\xi^c) = (& -a_2u - a_3v - a_4w + f_t, \\ & -a_2v + a_3u - a_5w + g_t\,, \end{aligned}$$

$$-a_2 w + a_4 u + a_5 v + h_t ,$$

$$-2a_2 p + j - x f_{tt} - y g_{tt} - z h_{tt}) , \qquad (5.3)$$

where a_1, a_2, a_3, a_4, and a_5 are five arbitrary parameters; f, g, h, and j are arbitrary functions of t; and u, v, and w are the components of $\mathbf{v}$. The definitions (5.2) and (5.3) can be used to generate a family of isovectors of the form (4.8), provided η^c_α and $\lambda^c_{\alpha\beta}$ are defined as in (4.14) and (4.15). In addition, substitution of τ^α and ξ^c into (3.5) gives rise to a class of solutions to the equations of variation of the Navier–Stokes equations. To exemplify, we give the expression of η^c in three particular cases, when all arbitrary parameters and arbitrary functions are equated to zero, with the only exception of one of them. We get case (1) ($a_1 = 1$),

$$(\eta^c) = (\boldsymbol{\eta}, \eta^4) = (-\mathbf{v}_t, p_t) ; \qquad (5.4)$$

case (2) ($a_2 = 1$),

$$(\eta^c) = (\boldsymbol{\eta}, \eta^4)$$
$$= (-\mathbf{v} - (\mathbf{x}\cdot\nabla)\mathbf{v} - 2t\mathbf{v}_t, -2p - (\mathbf{x}\cdot\nabla)p - 2tp_t) ; \qquad (5.5)$$

and case (3) (f arbitrary),

$$(\eta^c) = (\boldsymbol{\eta}, \eta^4) = (f_t \mathbf{e} - f\mathbf{v}_x, -x f_{tt} - f p_x) ; \qquad (5.6)$$

where spatial vectors are introduced in order to shorten the corresponding formulas, and $\mathbf{e}$ denotes the unit vector of the x axis.

To the aim of finding conserved quantities we also need solutions to the adjoint system of the equations of variation for (5.1), which reads

$$\tilde{\boldsymbol{\eta}}_t + (\mathbf{v}\cdot\nabla)\tilde{\boldsymbol{\eta}} + \nu\nabla^2\tilde{\boldsymbol{\eta}} - (\nabla\mathbf{v})\cdot\tilde{\boldsymbol{\eta}} + \nabla\tilde{\eta}^4 = 0 , \qquad (5.7a)$$

$$\nabla\cdot\tilde{\boldsymbol{\eta}} = 0 , \qquad (5.7b)$$

where $\tilde{\boldsymbol{\eta}}$ denotes a vector of components $(\tilde{\eta}^1, \tilde{\eta}^2, \tilde{\eta}^3)$, and the ith component of $(\nabla\mathbf{v})\cdot\tilde{\boldsymbol{\eta}}$ is given by $(\partial\mathbf{v}/\partial x^i)\cdot\tilde{\boldsymbol{\eta}}$. A three-parameter family of solutions to (5.7) is given by

$$(\tilde{\eta}^c) = (\tilde{\boldsymbol{\eta}}, \tilde{\eta}^4) = (c_1, c_2, c_3, c_1 u + c_2 v + c_3 w) . \qquad (5.8)$$

Henceforth it is assumed $c_1 = 1$, $c_2 = c_3 = 0$, that is

$$(\tilde{\eta}^c) = (1, 0, 0, u) . \qquad (5.9)$$

On comparing (3.14) with (3.5), it follows that substitution into (3.14) of (5.9) and (5.4)–(5.6) gives rise to three conserved currents that may be written as

$$(J^\alpha) = (\mathbf{J}, J^4) = (\eta^1\mathbf{v} + u\boldsymbol{\eta} - \nu\nabla\eta^1 + \eta^4\mathbf{e}, \eta^1) , \qquad (5.10)$$

where the vector $\boldsymbol{\eta}$ and η^4 are determined by comparison with (5.4)–(5.6).

Finally, let us consider the isovector Y determined by the condition $a_2 = 1$ [case (2)]. On using the conserved currents (5.10) we may construct additional conservation laws identified by (4.23). In particular, a lengthy calculation shows that if the expression of J^α is obtained by substitution of (5.4) into (5.10) then (4.23) yields the null current. On the contrary, the conserved vector corresponding to a J^α determined by (5.6) is non-null and reads

$$(K^\alpha) = (\mathbf{K}, K^4)$$
$$= (2t\,\mathbf{J}_t + 3u\boldsymbol{\eta} + 3\eta^1\mathbf{v} + 2f(u\mathbf{v})_x + \nu f\nabla u_x$$
$$- (p_x f + 5x f_{tt})\mathbf{e},\ 2t(f_{tt} - f_t u_x) - f u_x + 3f_t) . \qquad (5.11)$$

Needless to say, the fact that $D_\alpha J^\alpha = D_\alpha K^\alpha = 0$ on solutions to (5.1) can also be checked directly under the obvious assumption that J^α and K^α are given by (5.10) and (5.11), respectively.

VI. COMMENTS AND CONCLUSIONS

The first part of the paper has been devoted to an analysis of some interesting properties of the generators of infinitesimal symmetry transformations for a given system of partial differential equations, say S. A geometric reformulation of the equations of S has brought into evidence the fact that symmetry transformations may be simply interpreted as isovectors. On the basis of this relationship, some previous results concerning the role of symmetry transformations as generators of conserved quantities have been extended to isovectors.

Specifically, it has been shown that any isovector identifies a solution to the equations of variation for S, whence it follows that every pair of isovectors gives rise to a conserved current, provided S is self-adjoint. If this condition is not satisfied, the construction of a conserved current still can be pursued, but now one needs the preliminary knowledge of an isovector and of a solution to the adjoint system of the equations of variation.

A further analysis of the structure of the conservation laws obtained by the previous procedure has led to the introduction of a slight extension of the usual definition of conservation law,[16] which, in turn, has been used to find conservation laws by Lie derivation of a given one along the direction of an isovector. In particular, it has been shown that there exist conserved currents canonically associated with every isovector.

Isovectors and symmetry transformations had been primarily used to imbed known solutions of S into a Lie group of solutions, to find similarity solutions,[6,7,17,19] and to study constitutive equations.[20] The present analysis yields a contribution towards the investigation of their role as generators of conservation laws. Let us point out that this feature of isovectors seems to be the most natural, in a sense, because isovectors give a description of the invariance groups of S, and one is led to conjecture the existence of a strict association between such groups and conservation laws; in addition, the concept of isovector reduces to the well-known definition of dynamical symmetry whenever we consider a dynamical system in a finite number of degrees of freedom, and dynamical symmetries are the most natural generators of first integrals of motion.[9–12] This analogy can be pushed also further by recalling that the self-adjointness of the system that characterizes isovectors—when restricted to a solution of S—is a sufficient condition for the admissibility of a variational formulation for S, as it happens for dynamical symmetries.[10]

ACKNOWLEDGMENTS

This work has been performed under the auspices of GNFM-CNR and partially supported through the MPI 40% research project "Problemi di evoluzione nei fluidi e nei solidi."

The author is also grateful to an anonymous referee for several comments that helped to improve the exposition.

[1]I. M. Gelfand and S. V. Fomin, *Calculus of Variations* (Prentice–Hall, Englewood Cliffs, NJ, 1963); D. Lovelock and H. Rund, *Tensors, Differential Forms and Variational Principles* (Wiley, New York, 1975).
[2]F. Bampi and A. Morro, J. Math. Phys. **23**, 2312 (1982).
[3]R. M. Santilli, Ann. Phys. NY **103**, 354 (1977); **103**, 409 (1977); **105**, 227 (1977).
[4]M. Lutzky, J. Phys. A: Math. Gen. **11**, 249 (1978); Phys. Lett. A **72**, 86 (1979); P. J. Olver, in *Systems of Nonlinear Partial Differential Equations*, edited by G. M. Ball (Reidel, Dordrecht, 1983).
[5]S. P. Lloyd, Acta Mech. **38**, 85 (1981).
[6]R. E. Boisvert, W. F. Ames, and U. N. Srivastava, J. Eng. Math. **17**, 203 (1983).
[7]D. G. B. Edelen, Int. J. Eng. Sci. **20**, 803 (1982).
[8]G. W. Bluman and J. D. Cole, J. Math. Mech. **18**, 1025 (1969).
[9]W. Sarlet and F. Cantrijn, SIAM Rev. **23**, 467 (1981).
[10]G. Caviglia, Inverse Problems **1**, L13 (1985).
[11]G. Caviglia, Int. J. Theor. Phys. **22**, 1051 (1983); J. Math. Phys. **24**, 2065 (1983).
[12]G. E. Prince and C. J. Eliezer, J. Phys. A: Math. Gen. **14**, 587 (1981); P. G. L. Leach, J. Math. Phys. **22**, 679 (1981); J. L. Reid and J. R. Ray, J. Phys. A: Math. Gen. **15**, 2751 (1982); G. E. Prince and M. Crampin, Gen. Relativ. Gravit. **16**, 921 (1984).
[13]E. Cartan, *Les systèmes différentiels exterieurs et leurs applications géométriques* (Hermann, Paris, 1945).
[14]B. K. Harrison and F. B. Estabrook, J. Math. Phys. **12**, 653 (1971); F. B. Estabrook, in *Geometrical Approaches to Differential Equations*, edited by A. Dold and B. Eckmann (Springer, Berlin, 1980).
[15]Y. Choquet-Bruhat, C. De Witt-Morette, and M. Dillard-Bleick, *Analysis, Manifolds and Physics* (North-Holland, Amsterdam, 1977).
[16]H. D. Wahlquist and F. B. Estabrook, J. Math. Phys. **16**, 1 (1975); Y. Akyildiz, J. Math. Phys. **23**, 1723 (1982); E. S. Suhubi, Int. J. Eng. Sci. **22**, 119 (1984).
[17]D. G. B. Edelen, *Isovectors Methods for Equations of Balance* (Sijthoff and Nordhoff, Alphen aan den Rijn, 1980).
[18]It has been shown in Refs. 7 and 17 that isovectors do necessarily satisfy the restrictions $\tau^\alpha = \tau^\alpha(x^\beta,\phi^b)$ and $\xi^c = \xi^c(x^\beta,\phi^b)$, provided Eqs. (2.1) are written in the form of balance equations. However, this condition is not fulfilled in the present case, so that dependence on Z^c_α and $W^c_{\alpha\beta}$ is allowed, at least in principle.
[19]D. Fusco, A. Donato, and W. F. Ames, Wave Motion **6**, 517 (1984).
[20]D. Fusco, Int. J. Non-Linear Mech. **19**, 565 (1984).

Killing spinors on spheres and hyperbolic manifolds

Yasunori Fujii and Kengo Yamagishi[a)]
Institute of Physics, University of Tokyo-Komaba, Meguro-ku, Tokyo 153, Japan

(Received 22 August 1985; accepted for publication 20 November 1985)

Properties of Killing spinors on spheres and hyperbolic manifolds are investigated with an emphasis on the relations to Killing vectors, conformal Killing vectors, and solutions of Maxwell's equations.

I. INTRODUCTION

The concept of Killing spinors[1] proves to be a powerful tool in the study of supersymmetry in Kaluza–Klein theories.[2] Many of its aspects, however, are yet to be investigated. In this paper we present some of the properties of Killing spinors on manifolds of constant curvature; spheres S^n for positive curvature and hyperbolic manifolds for negative curvature, respectively. The simplest example of the latter is provided by the Poincaré manifolds H^n with $ds^2 = x_n^{-2}(dx_1^2 + \cdots + dx_n^2)$, $(x_n > 0)$ (see Ref. 3). With a discrete subgroup Γ one obtains H^n/Γ, which, compact or noncompact, can be made complete with finite volume.[4] This might entitle one of these hyperbolic manifolds to be a candidate of internal space of Kaluza–Klein theories, although global Killing vectors no longer exist after dividing by Γ.

With applications to Kaluza–Klein theories in mind, we put a particular emphasis on the relations to (local) Killing vectors, conformal Killing vectors, and solutions of "Maxwell's equations" for S^n and H^n. We also hope that our studies offer some insights into the applications to Ricci flat manifolds.

II. DEFINITIONS

A Killing spinor is defined by

$$\nabla_m \zeta = (\tfrac{1}{2})\kappa\gamma_m\zeta , \tag{1}$$

on an n-dimensional manifold, where $m = 1,\ldots,n$ label the coordinates, γ_m is the Hermitian Dirac matrix obeying

$$\{\gamma_m,\gamma_n\} = 2g_{mn} ,$$

and the covariant derivative is given by

$$\nabla_m\zeta = (\partial_m + i\tfrac{1}{4}\,\omega^{\underline{mn}}{}_m\sigma_{\underline{mn}})\zeta ,$$

in terms of the spin connection $\omega^{\underline{mn}}{}_m$. The underlined indices are for tangent space equipped with the flat metric $\delta_{\underline{mn}}$. The spin matrix $\sigma_{\underline{mn}}$ is Hermitian as defined by

$$\sigma_{\underline{mn}} = [\gamma_{\underline{m}},\gamma_{\underline{n}}]/2i . \tag{2}$$

The coefficient κ is 0 for Ricci flat manifolds while it is $\pm i$ and ± 1 for manifolds of positive and negative unit curvature, respectively. This is verified by computing

$$[\nabla_m,\nabla_n]\zeta = -\tfrac{1}{4}\kappa^2[\gamma_m,\gamma_n]\zeta , \tag{3}$$

from (1). On the left-hand side we have

$$i\tfrac{1}{4}R^{\underline{mn}}{}_{mn}\sigma_{\underline{mn}}\zeta = i\tfrac{1}{2}\Lambda\sigma_{mn}\zeta , \tag{4}$$

where the curvature tensor is given by

$$R^{\underline{mn}}{}_{mn} = \Lambda(e_m{}^{\underline{m}}e_n{}^{\underline{n}} - e_m{}^{\underline{n}}e_n{}^{\underline{m}}) , \tag{5}$$

with $\Lambda = +1$ and -1 for S^n and H^n, respectively, and

$$\sigma_{mn} = e_m{}^{\underline{m}}e_n{}^{\underline{n}}\sigma_{\underline{mn}} , \tag{6}$$

where $e_m{}^{\underline{m}}$ stands for the vielbein. The right-hand side of (3) is

$$-\kappa^2(i/2)\sigma_{mn}\zeta, \tag{7}$$

due to (2) and (6). Equating (4) and (7) yields $\kappa^2 = -\Lambda$, as stated above.

With any ζ, one can always associate another spinor ζ_- defined by (1) with the sign reversed on the right-hand side:

$$\nabla_m\zeta_- = -\tfrac{1}{2}\kappa\gamma_m\zeta_- . \tag{8}$$

For even n, we find the relation

$$\zeta_- = \gamma_\#\zeta ,$$

where

$$\gamma_\# = \epsilon_\#\gamma_1\gamma_2\cdots\gamma_n = \gamma_\#{}^\dagger ,$$

with $\epsilon_\# = 1$ for $n = 4m$, while $\epsilon_\# = i$ for $n = 4m + 2$.

For odd n, there is no matrix which brings ζ to ζ_-, as long as we limit ourselves to gamma matrices of $2^{[n/2]}\times 2^{[n/2]}$.

For even n, one might also consider[2]

$$\nabla_m\zeta = \tfrac{1}{2}i\kappa\gamma_m\gamma_\#\zeta , \tag{9}$$

which, however, is equivalent to (1), because $i\gamma_m\gamma_\#$ satisfy the same commutation and anticommutation relations as those for γ_m; solution of (9) $\sim\sqrt{\gamma_\#}\times$ solution of (1). We do not consider (9) any more.

III. BILINEAR PRODUCTS

We derive general formulas involving bilinear products;

$$u \equiv \zeta^\dagger\zeta, \quad u_- \equiv \zeta^\dagger_-\zeta , \tag{10}$$

$$v_m \equiv \zeta^\dagger\gamma_m\zeta , \quad v_{-m} \equiv \zeta^\dagger_-\gamma_m\zeta , \tag{11}$$

$$f_{mn} \equiv \zeta^\dagger\sigma_{mn}\zeta , \quad f_{-mn} \equiv \zeta^\dagger_-\sigma_{mn}\zeta . \tag{12}$$

A typical calculation is illustrated by

$$\begin{aligned}\nabla_m v_n &= (\nabla_m\zeta)^\dagger\gamma_n\zeta + \zeta^\dagger\gamma_n(\nabla_m\zeta)\\ &= \tfrac{1}{2}\zeta^\dagger(\kappa^*\gamma_m\gamma_n + \kappa\gamma_n\gamma_m)\zeta ,\end{aligned} \tag{13}$$

where use has been made of (1) and its Hermitian conjugate. Choosing $\kappa = i$ and 1 for S^n and H^n, respectively, we obtain

[a)] Postdoctoral Fellow, Japan Society for Promotion of Sciences. Address after 1 September 1985: Department of Physics, Brown University, Providence, Rhode Island 02912.

0022-2488/86/040979-03$02.50

$$\nabla_m v_n = \begin{cases} (-i/2)\zeta^\dagger[\gamma_m,\gamma_n]\zeta, & \text{for } S^n, \\ \frac{1}{2}\zeta^\dagger\{\gamma_m,\gamma_n\}\zeta, & \text{for } H^n, \end{cases} \tag{14}$$

which yield

$$\nabla_m v_n + \nabla_n v_m = \begin{cases} 0, & \text{for } S^n, \\ 2g_{mn}(\zeta^\dagger\zeta), & \text{for } H^n. \end{cases} \tag{15}$$

It thus follows that

$$v_m = \begin{cases} K_m, & \text{for } S^n, \\ L_m, & \text{for } H^n, \end{cases} \tag{16}$$

where K_m and L_m are a Killing vector and a conformal Killing vector obeying

$$\nabla_m K_n + \nabla_n K_m = 0\,, \tag{17a}$$

$$\nabla_m L_n + \nabla_n L_m = (2/n) g_{mn}(\nabla_k L^k)\,, \tag{17b}$$

respectively. We also find

$$\nabla_k v^k = n(\zeta^\dagger\zeta), \quad \text{for } H^n. \tag{18}$$

The same calculation for v_{-m} results in

$$v_{-m} = \begin{cases} L_m, & \text{for } S^n, \\ K_m, & \text{for } H^n, \end{cases} \tag{19}$$

with

$$\nabla_k v^k_- = n(\zeta^\dagger_-\zeta), \quad \text{for } S^n. \tag{20}$$

We turn to f_{mn}. We obtain

$$\nabla_m f^{mn} = \tfrac{1}{2}\zeta^\dagger(\kappa^*\gamma_m\sigma^{mn} + \kappa\sigma^{mn}\gamma_m)\zeta \tag{21}$$

$$= \begin{cases} (-i/2)\zeta^\dagger[\gamma_m,\sigma^{mn}]\zeta, & \text{for } S^n, \\ \frac{1}{2}\zeta^\dagger\{\gamma_m,\sigma^{mn}\}\zeta, & \text{for } H^n, \end{cases} \tag{22}$$

which replace (13) and (14). Using the relations

$$[\gamma_m,\sigma^{mn}] = -2i(n-1)\gamma^n\,,$$

$$\{\gamma_m,\sigma^{mn}\} = 0\,,$$

we arrive at

$$\nabla_m f^{mn} = \begin{cases} -(n-1)v^n, & \text{for } S^n, \\ 0, & \text{for } H^n. \end{cases} \tag{23}$$

We notice that f^{mn} defined by (12) solves "Maxwell's equations" for H^n, while it solves the "modified Maxwell's equations" for S^n.

It is also straightforward to obtain

$$\nabla_m f_-^{\ mn} = \begin{cases} 0, & \text{for } S^n, \\ i(n-1)v_-^{\ n}, & \text{for } H^n. \end{cases} \tag{24}$$

It is even simpler to calculate

$$\partial_m u = \begin{cases} 0, & \text{for } S^n, \\ \kappa v_m, & \text{for } H^n, \end{cases} \tag{25a}$$

$$\partial_m u_- = \begin{cases} \kappa v_{-m}, & \text{for } S^n, \\ 0, & \text{for } H^n. \end{cases} \tag{25b}$$

Equations (16), (19), and (23)–(25) are summarized in Table I.

TABLE I. The columns give (I) Killing vectors, (II) conformal Killing vectors, (III) solutions of Maxwell's equations, (IV) solutions of modified Maxwell's equations, and (V) constant norm, respectively. Upper and lower rows are for S^n and H^n, respectively.

	I	II	III	IV	V
S^n	v_m	v_{-m}	f_-^{mn}	f^{mn}	u
H^n	v_{-m}	v_m	f^{mn}	f_-^{mn}	u_-

IV. EXPLICIT EXAMPLES

So far we have simply assumed that the Killing spinors exist. We now demonstrate that they do for the examples of $n = 2$ and 3. The line elements are given by

$$S^2\colon\ ds^2 = d\theta^2 + \sin^2\theta\, d\phi^2\,,$$

$$S^3\colon\ ds^2 = d\chi^2 + \sin^2\chi(d\theta^2 + \sin^2\theta\, d\phi^2)\,,$$

$$H^2\colon\ ds^2 = y^{-2}(dx^2 + dy^2) \quad (y>0)\,,$$

$$H^3\colon\ ds^2 = z^{-2}(dx^2 + dy^2 + dz^2) \quad (z>0)\,.$$

The computation will be simplest if we choose $e_m^{\ m} \propto \delta_m^{\ m}$, and the standard Pauli matrices for γ_m. We obtain (C_1, C_2 complex constants)

$$S^2\colon\ \zeta = \begin{pmatrix} C_1 e^{-i\phi/2}\sin(\theta/2) + C_2 e^{i\phi/2}\cos(\theta/2) \\ i[-C_1 e^{-i\phi/2}\cos(\theta/2) + C_2 e^{i\phi/2}\sin(\theta/2)] \end{pmatrix}, \tag{26a}$$

$$S^3\colon\ \zeta = \begin{pmatrix} e^{-i\chi/2}[(C_1 e^{i\phi/2} + C_2 e^{-i\phi/2})e^{i\theta/2} \\ \quad + (C_1 e^{i\phi/2} - C_2 e^{-i\phi/2})e^{-i\theta/2}] \\ ie^{i\chi/2}[(C_1 e^{i\phi/2} + C_2 e^{-i\phi/2})e^{i\theta/2} \\ \quad - (C_1 e^{i\phi/2} - C_2 e^{-i\phi/2})e^{-i\theta/2}] \end{pmatrix}, \tag{26b}$$

with ζ_- obtained by $\chi \to -\chi$,

$$H^2\colon\quad \zeta = \begin{pmatrix} (C_1 + iC_2 x)y^{-1/2} + C_2 y^{1/2} \\ -i[(C_1 + iC_2 x)y^{-1/2} - C_2 y^{1/2}] \end{pmatrix}, \tag{26c}$$

$$H^3\colon\quad \zeta = \begin{pmatrix} C_1 z^{1/2} \\ [C_1(x+iy) + C_2]z^{-1/2} \end{pmatrix}, \qquad \zeta_- = \begin{pmatrix} [-C_1(x-iy) + C_2]z^{-1/2} \\ C_1 z^{1/2} \end{pmatrix}. \tag{26d}$$

V. POSITION DEPENDENCE OF SOME "NORMS"

As we noticed in (25a), $\zeta^\dagger\zeta$ is not constant for H^n. Examples for $n = 2$ and 3 are given by

$$\zeta^\dagger\zeta = 2[|C_1|^2 + i(C_1^*C_2 - C_2^*C_1)x + |C_2|^2x^2]y^{-1} + |C_2|^2 y\,, \quad \text{for } H^2\,.$$

$$\zeta^\dagger\zeta = z^{-2}[|C_1|^2(x^2+y^2+z^2)^2 + |C_2|^2 z + (C_1^*C_2 + C_2^*C_1)zx - i(C_1^*C_2 - C_2^*C_1)zy]\,, \quad \text{for } H^3\,.$$

These are not normalizable even if one integrates over noncompact but finite-volume regions bounded by geodesics. This may not be serious, however, because one can apply the procedure of completion to have compact and complete manifolds.[4]

The nonconstancy of the norms, on the other hand, may also infiltrate S^n, when one calculates "norms of the Maxwell fields" f_{-mn}. Although it is trivially constant for $n = 2$, we encounter some complications for $n = 3$. We form the real parts

$$\hat{f}_{-}^{mn} = f_{-}^{mn} + f_{-}^{*mn} = \epsilon^{mnk}\hat{v}_{-k} ,$$

where

$$\hat{v}_{-k} = v_{-k} + v_{-k}^{*}.$$

We thus find

$$\hat{f}_{-mn}\hat{f}_{-}^{mn} = 2\hat{v}_{-m}\hat{v}_{-}^{m} .$$

By using the Fierz identities like

$$v_{-m}v_{-}^{m} = (\zeta_{-}^{\dagger}\gamma_m\zeta)(\zeta_{-}^{\dagger}\gamma^m\zeta) = (\zeta_{-}^{\dagger}\zeta)^2 ,$$

we finally obtain

$$\begin{aligned}\hat{f}_{-mn}\hat{f}_{-}^{mn} &= 2[2(\zeta^{\dagger}\zeta)(\zeta_{-}^{\dagger}\zeta_{-}) + (\zeta_{-}^{\dagger}\zeta)^2 + (\zeta^{\dagger}\zeta_{-})^2] \\ &= 16[(|C_1|^2 + |C_2|^2)^2(1+\cos^2\chi) \\ &\quad + (|C_1|^2 - |C_2|^2)^2 \sin^2\chi \sin^2\theta] .\end{aligned}$$

This type of position dependence may have some consequences in the process of spontaneous compactification in Kaluza–Klein theories, as will be discussed in a separate article.[5]

VI. CONFORMAL GROUP

We close this paper by adding intuitive explanations of (conformal) Killing vectors on S^n and H^n. For S^n, Killing vectors correspond to rotations M_{IJ} $(I,J = 1,\dots,n+1)$, while conformal Killing vectors correspond to rotations M_{0I} with a hypothetical 0-axis with negative signature. These vectors constitute altogether Lie algebra of conformal group $SO(n+1,1)$: $M_{0,n+1}$ is identified as dilatation D in n dimensions, while $M_{m,n+1}$ and M_{0m} $(m = 1,\dots,n)$ are given by $k_m + P_m$ and $k_m - P_m$, respectively, in terms of special conformal transformations k_m and translations P_m.

The same conformal group occurs also for H^n, but with different identifications: K_m as M_{ij}, k_i and P_i $(i = 1,\dots,n-1)$, and D; L_m as M_{in}, k_n, and P_n. An explicit example of H^2 is

$$P_x = (1,0), \quad k_x = (x^2 - y^2, 2xy), \quad D = (x,y) ,$$

for K^m, while

$$P_y = (0,1), \quad k_y = (2xy, y^2 - x^2), \quad M_{xy} = (-y,x) ,$$

for L^m.

ACKNOWLEDGMENT

One of the authors (Y.F.) thanks Professor S. Kojima for many illuminating discussions on hyperbolic manifolds.

[1] L. F. Abbott and S. Deser, Nucl. Phys. B **195**, 76 (1982); P. Breitenlohner and D. Z. Freedman, Ann. Phys. NY **144**, 249 (1982).

[2] See, e.g., M. J. Duff and C. N. Pope, in *Supersymmetry and Supergravity '82*, edited by S. Ferrara, J. G. Taylor, and P. van Nieuwenhuizen (World Scientific, Singapore, 1983); P. van Nieuwenhuizen, in *Group-Theoretical Methods in Physics* (Nauka, Moscow, 1982).

[3] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry I* (Interscience, New York, 1964).

[4] See, e.g., J. Milnor, Bull. Math. Soc. **6**, 9 (1982).

[5] Y. Fujii and K. Yamagishi, Phys. Lett. B **162**, 317 (1985).

The problem of "global color" in gauge theories

P. A. Horváthy[a)]

Centre de Physique Theorique, Laboratoire propre du Centre National de la Recherche Scientifique, Luminy Case 907, 13288 Marseille Cedex 9 France and U. E. R. de Mathématique, Université de Provence, Marseille, France

J. H. Rawnsley

Centre de Physique Theorique, Laboratoire propre du Centre National de la Recherche Scientifique, Luminy Case 907, 13288 Marseille Cedex 9, France and Mathematics Institute, University of Warwick, Coventry CV4 7AL, England

(Received 12 August 1985; accepted for publication 31 October 1985)

The problem of "global color" (which arose recently in monopole theory) is generalized to arbitrary gauge theories: a subgroup K of the "unbroken" gauge group G is implementable iff the gauge bundle reduces to the centralizer of K in G. Equivalent implementations correspond to equivalent reductions. Such an action is an internal symmetry for a given configuration iff the Yang–Mills field reduces also. The case of monopoles is worked out in detail.

I. INTRODUCTION

One of the most exciting problems that arose recently in monopole theory is that of *global color*.[1–8] We formulate it in two steps[7]: First, we would like to define the action of a fixed element g of the (unbroken) gauge group. Under usual conditions this presents no problem. In topologically nontrivial situations, however, this may not be possible or ambiguous. This is the problem of implementability. Second, if we are able to define such an action, when do we get a *symmetry* (in the sense of Schwarz[9] and Forgács and Manton[10]) for a given field configuration? The importance of these notions is seen, for example, from the role they play in deriving conserved charges in gauge theories.[7,11] In this paper we give the mathematical solution to these problems. Our theory (formulated in fiber-bundle terms[9,12–14]) is valid for any classical gauge theory. Notice that the problem studied here is a special case of dimensional reduction.[10,15]

Our starting point is Proposition 2.2, which states that a "rigid" internal action of a subgroup K of G on P exists if and only if P reduces to an $H = Z_G(K)$ bundle Q. Furthermore, there is a 1-1 correspondence between such equivalent actions and isomorphic reductions (Propositions 2.4 and 2.5).

An action of K on the principal bundle P induces an action of K also on the Yang–Mills (YM) field. Similarly, we can study the action of K on matter fields—sections of bundles associated to P. The condition for such an action to exist is expressed again in terms of bundle reduction (Theorem 3.1).

When is an action a *symmetry* for a given field configuration? Proposition 4.3 tells us that the action of K on (P,G) defined by (Q,H) is an internal symmetry for a Yang–Mills connection A if and only if A reduces to a connection on Q. This happens if and only if H contains the holonomy group of A (see Ref. 15). The implementation of an internal symmetry subgroup is necessarily unique. There is an analogous statement (Proposition 4.4) for matter fields.

These theorems provide us with a complete solution of the color problem—when we are able to construct the corresponding reductions. A first illustration is given by the *non-Abelian Bohm–Aharonov experiment* of Wu and Yang,[11,17,18] where $G = \mathrm{SU}(2)$ admits two inequivalent implementations.

The principal application of our theory is to non-Abelian monopoles.[1–7] Their basic properties[19–25] are geometrically reformulated in Sec. V. The reduction of monopole bundles is worked out in Sec. VI.

The results are summarized as follows: Denote by G the residual symmetry group of a monopole having $[P]\in\pi_1(G)$ as a fundamental topological invariant. A subgroup K of G is *implementable* iff $[P]$ belongs to the image of i_* $\pi_1(Z_G(K)) \to \pi_1(G)$ induced by the inclusion $i: Z_G(K) \to G$. Furthermore, the *inequivalent implementations* are labeled by the elements of $\pi_2(G/Z_G(K))$. In particular, the implementation of the full G is unique (when it does exist).

These results are conveniently expressed in terms of the "non-Abelian charge" Π of Goddard, Nuyts, and Olive[19]: Let us decompose Π as $\Pi = z(\Pi) + \Pi'$, where $z(\Pi)\in Z(\mathscr{G})$ and $\Pi'\in[\mathscr{G},\mathscr{G}]$. We prove that G is implementable iff either (i) $[\exp 4\pi t\Pi]\in\pi_1(G)_{\text{free}}$, and $z(\Pi)$, the projection of the non-Abelian charge onto the center, is quantized: $\exp 4\pi z(\Pi) = 1$; or equivalently, (ii) $\exp 4\pi\Pi' t$, $0\leqslant t\leqslant 1$, is a contractible loop. Here G is a symmetry for a monopole given by Π iff $\Pi\in Z(\mathscr{G})$.

The general results are illustrated on SO(3) monopoles.[19,20]

II. INTERNAL ACTIONS ON PRINCIPAL BUNDLES

Let P be a right principal G-bundle over a connected manifold M. A subgroup K of G *acts internally on* P, if we are given a left action $p \to k\cdot p$ of K on P, which preserves each fiber and commutes with the (right) action of G, $k\cdot(pg) = (k\cdot p)g$, $\forall k\in K$, $g\in G$, $p\in P$, cf. Refs. 1–7. If so, define the map $\tau_p: K \to G$ by $k\cdot p = p(\tau_p(k))$. Here τ_p is well defined, since $k\cdot p$ belongs to the same fiber as p, and G acts on each fiber transitively and freely; $k \to \tau_p(k)$ is a homomorphism of K into G, which satisfies $\tau_{pg} = \mathrm{Ad}\, g^{-1}\cdot\tau_p$, $g\in G$. In what follows we consider only the case when K acts

[a)] Present address: Dublin Institute for Advanced Studies, 10, Burlington Road, Dublin 4, Ireland.

 0022-2488/86/040982-09$02.50

on P freely, i.e., the homomorphism $\tau_p: K \to G$ is injective for each p. This can always be assumed without loss of generality for symmetries (see Sec. IV).

Choosing a local section $s_\alpha: V_\alpha \to P$, τ is given by $\tau^\alpha: V_\alpha \to \mathrm{Hom}(K,G)$, where $\tau^\alpha_x = \tau_{s_\alpha(x)}$. If $h_{\alpha\beta}: V_\alpha \cap V_\beta \to G$ denotes the transition function of P, then $\tau^\alpha_x = \mathrm{Ad}\, h_{\alpha\beta}\, \tau^\beta_x$, $x \in V_\alpha \cap V_\beta$.

An internal action of K on P is called *rigid*, if there exists a local trivialization $\{V_\alpha, s_\alpha\}$ of P such that each τ^α is constant.[1-7] If so, there is no loss of generality in assuming that $\tau^\alpha_x(k) = k$ for each x. In such a gauge,

$$k \cdot s_\alpha(x) = s_\alpha(x)\, k, \quad \forall x \in V_\alpha, \quad k \in K. \tag{2.1}$$

Proposition 2.1: An internal action of K on P is rigid if and only if the image of the associated map $\tau: P \to \mathrm{Hom}(K,G)$ is the orbit of the inclusion map $i: K \hookrightarrow G$ under the adjoint action of G on $\mathrm{Hom}(K,G)$.

Proof: Suppose the action of K on P is rigid in a gauge $\{V_\alpha, s_\alpha\}$. If $p \in P$ is such that $\pi(p) \in V_\alpha$, where π is the projection $\pi: P \to M$, then $p = s_\alpha(\pi(p))g$ for some $g \in G$. By (2.1),

$$p\tau_p(k) = k \cdot p = k \cdot (s_\alpha(\pi(p))g) = (k \cdot s_\alpha(\pi(p)))g$$

$$= (s_\alpha(\pi(p))k)g = p(g^{-1}kg).$$

Hence $\tau_p(k) = \mathrm{Ad}\, g^{-1}k$ and so $\tau_p = \mathrm{Ad}\, g^{-1}i$, and thus the image is the orbit of i. Conversely, if τ has a single orbit as its image, we can always choose local gauges s_α so that τ^α is constant, equal to a base point, which in this case is the inclusion map $i: K \hookrightarrow G$. In this gauge the action of K is rigid.

Requiring rigidity is easily seen to be the same as requiring that, for each p, τ_p is the restriction to K of an automorphism of G (see Refs. 1, 2, 6, and 7).

Let us now consider a rigid internal action of $K \subset G$ and let H denote the stabilizer of $i: K \hookrightarrow G$ under the adjoint action of G,

$$H = Z_G(K) = \{g \in G \mid \mathrm{Ad}\, gk = k, \ \forall k \in K\}. \tag{2.2}$$

The orbit of i is identified with G/H, and τ can be viewed as a section of the associated bundle with fiber G/H. Any such section defines a reduction of P to an H bundle. The reductions to H-bundles are known to be in 1–1 correspondence with sections of the associated bundle $Px_G(G/H) \simeq P/H$, and so with rigid actions of K on P. Hence we have proved the following proposition.

Proposition 2.2: A rigid internal action of a subgroup K of G on P exists if and only if P reduces to $H = Z_G(K)$.

It is easy to see directly that the existence of a rigid action forces the bundle to reduce by using the special gauge (2.1), for, if $s_\beta = s_\alpha h_{\alpha\beta}$, then

$$s_\alpha(kh_{\alpha\beta}) = k \cdot s_\alpha h_{\alpha\beta} = k \cdot s_\beta = s_\beta k = s_\alpha h_{\alpha\beta} k,$$

showing that $h_{\alpha\beta}$ commutes with K, and hence (P,G) reduces to an H-bundle Q. Note that the latter is given by $Q = \{p \in P \mid \tau_p k) = k, \ \forall k \in K\}$.

Corollary 2.3: G itself acts on P internally and rigidly if and only if (P,G) reduces to $Z(G)$, the center of G. In this case the transition functions take their values in $Z(G)$ in a suitable gauge.

An interesting insight is gained by proceeding backwards. Let $i: H \hookrightarrow G$ be a closed Lie subgroup and assume we are given a principal H-bundle (Q,H). Then $(q,g) \simeq (qh, h^{-1}g)$ is an equivalence relation on $Q \times G$, and the set of equivalence classes (denoted here by $\{q,g\}$) yields the associated bundle $P^Q = Q \times_H G$; P^Q is a principal G-bundle with G-action $\{q,g\} \to \{q,gg'\}$, $g' \in G$. Furthermore, (Q,H) is a reduction of (P^Q,G). In fact, $Q \simeq (\{q,e\} \in P^Q \mid q \in Q)$. Conversely, if (Q,H) is a reduction of (P,G), the extended bundle (P^Q,G) is isomorphic to (P,G). Indeed, any $p \in P$ is written as $p = qg$, since (Q,H) is a reduction. Now

$$\sigma(p) = \{q,g\} \tag{2.3}$$

is a well-defined map $\sigma: P \to P^Q$, since if $p' = q' \cdot g'$ with $q' \in Q$, and $g' \in G$, then $q' = qh$, and $g' = h^{-1}g$ for some $h \in H$. Thus

$$(q', g') = (qh, h^{-1}g) \simeq (q, g).$$

Plainly, σ has inverse $\{q, g\} \to qg \in P$, and also commutes with the action of G on P and P^Q, respectively, so it is an isomorphism extending the identity map on Q.

Assume now that we have a reduction of (P,G) to an $H = Z_G(K)$ bundle Q. A rigid internal action of K on (P,G) is defined now as follows:

$$k \cdot \{q, g\} = \{q, k^{-1}g\} \tag{2.4}$$

is a natural action of K on P^Q [since

$$(qh, kh^{-1}g) \simeq (q, hkhg^{-1}) \simeq (q, Kg),$$

for $h \in Z_G(K)$] and thus on $(P,G) \simeq (P^Q,G)$. Observe that this action is given by

$$\tau_p = \mathrm{Ad}\, g^{-1}\, i, \tag{2.5}$$

and is thus rigid. It also satisfies $\tau|Q = i$ as required.

Now we turn to the question of uniqueness of the action of K. To be able to discuss several actions at once we use the following notation: We denote by $\mu: K \times P \to P$ the map $\mu(k,p) = k \cdot p$. Two actions given by μ_1 and μ_2 should be regarded as equivalent if they differ only by a gauge transformation, i.e., if there exists a bundle automorphism $\sigma: P \to P$ preserving fibers, commuting with the G-action, and such that $\mu_2(k,\sigma(p)) = \sigma(\mu_1(k,p))$, $\forall k \in K$, and $p \in P$. Equivalence preserves rigidity, since if $s_\alpha: V_\alpha \to P$ is a gauge in which μ_1 is constant, then σs_α is a gauge in which μ_2 is constant. An equivalence σ determines a map $\gamma: P \to G$ by

$$\sigma(p) = p\gamma(p), \tag{2.6}$$

which satisfies $\gamma(pg) = \mathrm{Ad}\, g^{-1}\gamma(p)$, so γ is a section of the bundle associated to P with G acting on itself by internal automorphisms. If we have two actions μ_1 and μ_2 of K with corresponding maps τ^1 and τ^2: $\mathrm{Hom}(K,G)$ then

$$\sigma(p)\tau^2_{\sigma(p)}(k) = \mu_2(k,\sigma(p)) = \sigma(\mu_1(k,p)) = \sigma(p)\tau^1_p(k),$$

so

$$\tau^1_p(k) = \tau^2_{\sigma(p)}(k) = \mathrm{Ad}\,\gamma(p)^{-1}\tau^2_p(k). \tag{2.7}$$

Conversely, given a section γ of $P \times_G G$ satisfying (2.7), then (2.6) defines an equivalence of the two actions.

If we have two rigid actions μ_1 and μ_2 with corresponding reductions Q_1 and Q_2 then an equivalence σ of μ_1 and μ_2 implies $\tau^1_p = \tau^2_{\sigma(p)}$ by (2.7), so $Q_2 = \sigma(Q_1)$. This suggests a notion of equivalence of reductions: two reductions Q_1 and Q_2 of a principal G-bundle P are equivalent if there is an automorphism σ of P preserving fibers with $Q_2 = \sigma(Q_1)$ (see Ref. 26). Then we have the following proposition.

Proposition 2.4: Two rigid actions μ_1 and μ_2 are equivalent if and only if their corresponding reductions Q_1 and Q_2 are equivalent.

Obviously, if two reductions Q_1 and Q_2 of P are equivalent as reductions, then they are isomorphic as H-bundles. This is in fact the only condition to be satisfied as the next result shows.

Proposition 2.5: There is a 1–1 correspondence between isomorphisms of Q_1 and Q_2 as H-bundles and equivalences of Q_1 and Q_2 as reductions of P.

Proof: It remains only to show how to extend an isomorphism σ_0: $Q_1 \to Q_2$ to an isomorphism of P. Set

$$\sigma(q,g) = (\sigma_0(q),g), \quad q\in Q_1, \quad g\in G; \tag{2.8}$$

$\sigma_0(qh)\, h^{-1}g = \sigma_0(q)g$, since σ_0 is an H-map, and thus (2.8) is a well-defined map from P^{Q_1} to P^{Q_2}. It is clearly 1–1 and commutes with the action of G, so it is a bundle-isomorphism. But (P^{Q_1},G) and (P^{Q_2},G) are both isomorphic to (P,G).

III. INTERNAL ACTIONS ON ASSOCIATED BUNDLES

A matter field Φ is specified by giving a unitary representation U: $G \to U(E)$ (the set of unitary transformation of a linear space E), and by selecting, in each V_α, a local representative Φ^α: $V_\alpha \to E$ such that

$$\Phi^\alpha(x) = U(h_{\alpha\beta}(x))\, \Phi^\beta(x), \quad x\in V_\alpha \cap V_\beta.$$

Suppose a subgroup K of G acts on Φ pointwise and linearly. This means that $k\in K$ sends Φ to an object we denote by $(k\cdot\Phi)$, expressed locally as

$$(k\cdot\Phi)^\alpha(x) = U_x^\alpha(k)\, \Phi^\alpha(x), \tag{3.1}$$

where each U_x^α is a representation of K on E. For (3.1) to be well defined we need the consistency condition

$$U_x^\alpha(k) = U[h_{\alpha\beta}(x)]\, U_x^\beta(k)\, U[h_{\alpha\beta}^{-1}(x)]. \tag{3.2}$$

Now Φ is a section of the associated bundle $\mathscr{A} = P\times_G E$. Equation (3.2) requires, therefore, that the $U_x^\alpha(k)$'s piece together to give sections of the bundle associated to P with fiber $U(E)$, where G acts by conjugation by $U(g)$. If the action of K on the fiber at x is denoted by $U_x(k)$, the $U_x^\alpha(k)$'s are local representatives of this action; $U(k)$ is a section of $U(\mathscr{A})$.

If the representation U is not faithful, denote by N its kernel, where N is a normal subgroup of G, and $G^* = G/N$ is a group to which U descends to give a faithful representation U^*. Now $P\times_G E$ is naturally isomorphic to $P^*\times_{G^*}E$, where $P^* = P/N = P\times_G(G/N)$ is the principal G^*-bundle associated to the homomorphism $G\to G^*$. In this way we may reduce to the case where U^* is faithful, but note that now K need not be a subgroup of G^*. This defect can be avoided if we assume that the action of K is induced locally by gauge transformations, i.e., if we assume that, for each k, there exist functions k_α: $V_\alpha \to G$ such that $U_x^\alpha(k) = U(k_\alpha(x))$ These k_α must satisfy $U[k_\alpha(x)] = U[h_{\alpha\beta}(x)\, k_\beta(x)\, h_{\alpha\beta}^{-1}(x)]$, so

$$k_\alpha(x)^{-1} h_{\alpha\beta}(x)\, k_\beta(x)\, h_{\alpha\beta}^{-1}(x)\in N. \tag{3.3}$$

If we denote by g^* the projection of $g\in G$ into G^*, then

$$k_\alpha^*(x) = h_{\alpha\beta}^*(x)\, k_\beta^*(x)\, h_{\alpha\beta}^*(x)^{-1}. \tag{3.4}$$

It follows that k^* defines a section of $P^*\times_{G^*}[G^*]$ with G^* acting by the adjoint representation on itself. Since each U_x^α is a homomorphism, then, although $k\to k_\alpha(x)$ need not be a homomorphism, $k\to k_\alpha^*(x)$ *is* a homomorphism. Thus, by (3.4), we obtain a section $\tau_x^*(k) = k_\alpha^*(x)$ of $P^*\times_{G^*}[\mathrm{Hom}\,(K,G^*)]$, and hence an action of K on P^*, which commutes with the G^*-action. If further there are local gauges where the k_α are K-valued, then $K\cap N$ acts trivially so we get an action of $K^* = K/(K\cap N)$, which is a subgroup of G^*. Hence we get the situation studied in Sec. II, with K^* acting on P^*.

Finally, if we restrict $k_\alpha(x)$ to be constant (and equal to k) in some gauge, then K^* acts rigidly on P^*. This can happen, as we have seen, if and only if P^* reduces to $Z_{G^*}(K^*)$, the centralizer of K^* in G^*. If

$$H' = \{g\in G \,|\, U(g)\, U(k) = U(k)\, U(g),\ \forall k\in K\}, \tag{3.5}$$

then this centralizer is $H^* = H'/N$. Since $P^* = P/N$, $P^*/H^* = P/H'$ and thus P^* reduces to H^* if and only if P reduces to H'. We summarize.

Theorem 3.1: If K acts pointwise on a generic matter field Φ transforming under a unitary representation U of G on E so that there are local gauges (V_α, s_α) where this action is rigid,

$$(k\cdot\Phi)^\alpha(x) = U(k)\, \Phi^\alpha(x), \tag{3.6}$$

the P reduces to the subgroup (3.5). Conversely, any reduction of P to H' induces an action of K on Φ.

This is the case in particular when K acts internally on P. Indeed, H in (2.2) is a subgroup of H'. Alternatively, observe that if the action of K on P is associated to τ: $P\to \mathrm{Hom}\,(K,G)$, then

$$\tau_{p^*}^*(k) = (\tau_{p^*}(k))^*$$

defines an action of K (and thus of K^*) on P^*. Alternatively, Φ can be viewed as a section of $P^*\times_{G^*}E$, or as an equivariant function $P^*\to E$,

$$\Phi\,(p^*g^*) = U(g^{*-1})\, \Phi\,(p^*), \quad g^*\in G^*, \quad p^*\in P^*.$$

Observe that the action of K on Φ is deduced from that of K^* on P^*,

$$(k\cdot\Phi)\,(p) = U(\tau_p^*(k))\, \Phi\,(p) \tag{3.7}$$

[since $\Phi\,(p) = \Phi\,(p^*)$], whose local form is

$$(k\cdot\Phi)^\alpha(x) = U(\tau_x^{*\alpha}(k))\Phi^\alpha(x). \tag{3.8}$$

The results of this section apply, besides monopoles (Sec. VII), to classical particles in external Yang–Mills fields.[17]

IV. INTERNAL SYMMETRIES

Let us now assume that our principal G-bundle P carries a connection form A, and let K be a subgroup of G acting on P internally. Let this action be given by τ. This allows us to define the action of a $k\in K$ on the Yang–Mills connection A, $(k\cdot A) = (k^{-1})^*A$, where $*$ denotes the pullback of a differential form. We shall call K an *internal symmetry group* for the Yang–Mills field A if this action preserves the connection

$$(k\cdot A)=A, \tag{4.1}$$

cf. Refs. 7, 9, 10, 12–16, 27, and 28. If K is a compact, connected Lie group with Lie algebra $\mathscr{k}$, any $k\in K$ can be written as $k=\exp\kappa$. Equation (4.1) implies that the vector field

$$\tilde{\kappa}(p)=\frac{d}{dt}\Big|_{t=0}[\exp(-t\kappa)]\cdot p \tag{4.2}$$

is invariant under the right action of G on P, and

$$L_{\tilde\kappa}A=0. \tag{4.3}$$

On the other hand, $\tilde\kappa(p)=\hat\xi_p$, the fundamental vectorfield at p associated to the infinitesimal right action of G at $p\in P$ for some $\xi\in\mathscr{G}$. Denote $(\omega_\kappa(p)\hat{\ })_p=\tilde\kappa(p)$. Alternatively, consider

$$\omega_\kappa(p)=\frac{d}{dt}\Big|_{t=0}\tau_p\,[\exp(-t\kappa)]. \tag{4.4}$$

For each $p\in P$, the map $\mathscr{k}\ni\kappa\to\omega_\kappa(p)\in\mathscr{G}$ satisfies

$$\omega_{[\kappa_1,\kappa_2]}(p)=[\omega_{\kappa_1}(p),\omega_{\kappa_2}(p)]\quad\text{and}\quad\omega_\kappa(p_0)=\kappa, \tag{4.5}$$

where $\omega_\kappa(pg)=\mathrm{Ad}\,g^{-1}\omega_\kappa(p)$ for each κ. Therefore, ω_κ is an adjoint "Higgs-type" field.

To express (4.3) another way, observe that $L_{\tilde\kappa}A=dA\,(\tilde\kappa,\cdot)+d\,(A\,(\tilde\kappa))$ by the Cartan lemma. But

$$dA\,(\tilde\kappa,\cdot)=DA(\tilde\kappa,\cdot)-[A\,(\tilde\kappa),A(\cdot)]=[A\,(\cdot),A(\tilde\kappa)]$$

because $DA=F$ is horizontal. Finally,

$$A_p(\tilde\kappa_p)=A_p\,(\omega_\kappa(p)\hat{\ })=\omega_\kappa(p),$$

since $A(\hat\xi)=\xi$ for any connection A on P. So $L_{\tilde\kappa}A=D\omega_{\tilde\kappa}$. This yields the following proposition.

Proposition 4.1: A connected subgroup K of G acting internally on P is an internal symmetry for the connection A if and only if the adjoint "Higgs" field ω_κ associated to its infinitesimal action is covariantly constant for each κ,

$$D\omega_\kappa=0, \tag{4.6}$$

cf. Refs. 5, 7, and 16. When expressed in a local gauge $\{V_\alpha,s_\alpha\}$, Eqs. (4.1) and (4.6) become[7,9,10,12–16,27,28]

$$A^\alpha(x)=\mathrm{Ad}\,\tau_x^\alpha(k)\,A^\alpha(x)-d\tau_x^\alpha(k)\,[\tau_x^\alpha]^{-1} \tag{4.7}$$

and

$$D\omega_\kappa^\alpha=d\omega_\kappa^\alpha+[A^\alpha,\omega_\kappa^\alpha]=0,$$

respectively, where $A^\alpha=s_\alpha^*A$, and

$$\omega_\kappa^\alpha(x)=\frac{d}{dt}\Big|_{t=0}\tau_x^\alpha\,[\exp(-t\kappa)], \tag{4.8}$$

$\omega_\kappa^\alpha(x)$ is just a local representative for ω_κ, and $\omega_\kappa^\alpha(x)=\omega_\kappa(s_\alpha(x))$, as anticipated by the notation.

Conversely, if we can find a bracket-preserving linear map $\kappa\to\omega_\kappa$ satisfying (4.5), that associates convariantly constant adjoint Higgs-type fields ω_κ to each $\kappa\in\mathscr{k}$, $[\exp(-\kappa)]\cdot p=\exp(\omega_\kappa(p)\hat{\ })$ (exponential of a vector field) provides us with an internal action of $k=\exp(-\kappa)\in K$ on P. In fact, $\tau_p\,[\exp(-\kappa)]=\exp(\omega_\kappa(p))$ (exponential in the group).

All solutions of (4.6) are found by parallel transport.[7,20] Therefore, $\omega_\kappa(p)$ belongs, for all $p\in P$, to a single adjoint orbit of G. Hence we get the following proposition.

Proposition 4.2: The action of an internal symmetry group K on P is rigid.

As we have seen in Sec. II, to have a rigid internal action of K is equivalent to requiring that the bundle (P,G) reduce to $H=Z_G(H)$. In terms of ω this reduction is obtained as $Q=\{p\in P\,|\,\omega_\kappa(p)=\kappa\}$ (see Ref. 12). This implies the following proposition.

Proposition 4.3: The action of K on (P,A) defined by the (Q,H) [where $H=Z_G(K)$] is an internal symmetry if and only if the connection form A reduces to Q. This happens if and only if $H=Z_G(K)$ contains the holonomy group of A. In this case Q contains the holonomy bundle and so is unique.

In particular, the *full gauge group G* is a group of internal symmetries if and only if the connection reduces to the $Z(G)$-bundle that characterizes the left action of G on P. This is equivalent to requiring that the generators of the holonomy group lie in the center[25] of $\mathscr{G}$.

Similarly, let us consider a matter field Φ, and assume there is an internal action of K on Φ determined by a section τ^* of $P^*\times_{G^*}[\mathrm{Hom}\,(K^*,G^*)]$. We shall say that this action of K is an *internal symmetry* for the *matter field* Φ, if

$$(k\cdot\Phi)\,(p)=U\,(\tau_p^*(k))\,\Phi\,(p)=\Phi\,(p). \tag{4.9}$$

In a local (in particular in a rigid) gauge, this reads

$$\begin{aligned}&U\,(\tau_x^{*\alpha}(k))\,\Phi^\alpha(x)=\Phi^\alpha(x)\\ &\text{and}\\ &U\,(k)\cdot\Phi^\alpha(x)=\Phi^\alpha(x),\end{aligned} \tag{4.10}$$

respectively. We can work also infinitesimally:

$$\begin{aligned}&\tilde\kappa^*(p^*)=\frac{d}{dt}\Big|_{t=0}[\exp(-t\kappa)]^*\cdot p^*\\ &\text{and}\\ &\omega_{\kappa^*}^*\,(p^*)=\frac{d}{dt}\Big|_{t=0}\tau_{p^*}^*\,[\exp(-t\kappa)]^*\end{aligned} \tag{4.11}$$

provide us with a vertical vectorfield κ^* and an adjoint "Higgs" field ω_κ^* on $P^*=P/N$. The infinitesimal action of $\mathscr{k}^*$ (the Lie algebra of K^*) on Φ reads

$$(\tilde\kappa^*\cdot\Phi)\,(p)=L_{\tilde\kappa^*}\Phi\,(p)=\omega_{\kappa^*}^*(p)\cdot\Phi\,(p), \tag{4.12}$$

where the dot $\cdot$ denotes the action of the Lie algebra induced by U. The definition of a symmetric matter field reads, hence, infinitesimally

$$\omega_{\kappa^*}^*\cdot\Phi\,(p)=0, \tag{4.13}$$

or in a local gauge

$$\omega_{\kappa^*}^{*\alpha}(x)\cdot\Phi^\alpha(x)=0. \tag{4.14}$$

Let us now assume that we have a Yang–Mills potential A and a covariantly constant Higgs field Φ and that we are interested in their simultaneous symmetries. First, K implementable on P implies that K is implementable also on $P\times_G E$. Furthermore, $\omega_\kappa^*(x)=(\omega_\kappa(p))^*$. In particular if N is discrete, $\mathscr{k}^*=\mathscr{k}$ and $\mathscr{G}^*=\mathscr{G}$, so the star can be dropped. Both ω_κ and Φ are now found by parallel transport, $\Phi^\alpha(x)=\mathrm{Ad}\,g^\alpha(x)\,\Phi_0$ and $\omega_\kappa^\alpha(x)=\mathrm{Ad}\,g^\alpha(x)\,\kappa$, where g^α

is the nonintegrable phase factor. Equation (4.13) reduces, hence, to

$$\kappa \cdot \Phi_0 = 0. \tag{4.15}$$

Proposition 4.4: K is a symmetry group for a covariantly constant matter field Φ if and only if K belongs to the little group of a basepoint Φ_0 from the orbit where Φ takes its values.

As a first illustration, consider the *non-Abelian Bohm–Aharonov experiment* [7,11,17,18] proposed by Wu and Yang to test the existence of gauge fields. Here we consider a principal G-bundle P over the punctured plane $M = R^2 \setminus \{0\}$ endowed with a flat connection A. Such bundles are classified by classes $[P]$ in $\pi_0(G)$. The homotopy exact sequence

$$\cdots \to \pi_1(G) \xrightarrow{\mathrm{Ad}_*} \pi_1(G/Z(G)) \xrightarrow{\delta} \underset{[Q]}{\pi_0(Z(G))} \xrightarrow{i_*} \underset{[P]}{\pi_0(G)} \to \cdots, \tag{4.16}$$

analogous to (6.3) and induced by $Z(G) \to G \to G/Z(G)$, shows that G is implementable iff $[P]$ belongs to Im i_*. If this condition is satisfied, there may be still an ambiguity: the different implementations—the inequivalent reductions to $Z(G)$—are classified by Ker I_*. For $G = \mathrm{SU}(2)$ for example, P is trivial because SU(2) is connected:

$$\pi_1(\mathrm{SU}(2)) = 0, \ Z(G) = Z_2, \ G/Z(G) = \mathrm{SO}(3)$$

so there exist *two* gauge-inequivalent implementations corresponding to the reductions of P to a trivial or to a twisted bundle, respectively.[7]

When do we get an internal symmetry? Let us consider more generally a principal G-bundle P over a connected manifold M carrying a flat connection A. The horizontal distribution of A is integrable by the Frobenius theorem. Let us choose a reference point p_0 in P, and denote by Q the leaf of the horizontal distribution through p_0. Here Q is a covering of M, which is a reduction of P with a discrete subgroup Γ of G as structure group. As a matter of fact, Γ is just the holonomy group of A at p_0. According to Proposition 4.3, G acts as a symmetry for (P,A) iff Γ is in $Z(G)$.

In the non-Abelian Bohm–Aharonov experiment, Γ consists of powers of Φ, the nonintegrable phase factor calculated along a loop that winds once around the origin.[11,17,18] Consequently, SU(2) acts as a group of internal symmetries only for $\Phi = 1$ (for the first implementation) or for $\Phi = -1$ (for the second implementation).[7] The physical consequences are explained in Ref. 11.

V. MONOPOLE BUNDLES

The asymptotic properties of monopoles are determined by a principal G-bundle P over the S^2 at infinity, where G, the residual symmetry group, is compact and connected. In grand unified theories (GUT) one starts in general with a trivial "unifying" bundle $\tilde{P} = R^3 \times \tilde{G}$, where $\tilde{G}$—a compact and connected Lie group—is the "unifying group." At large distances the $\tilde{G}$-symmetry is spontaneously broken to a subgroup G of $\tilde{G}$. Geometrically, this means that over S^2, $(\tilde{P},\tilde{G})$ is reduced to a principal G-bundle P. Any such reduction is produced by an equivariant "reducing map."[12] Choosing a global trivialization of $\tilde{P}$, the reducing map can be identified with a map Φ: $S^2 \to \tilde{G}/G$—the physical *Higgs field*. Now Φ defines a homotopy class $[\Phi] \in \pi_2(\tilde{G}/G)$, and the homotopy class $[P]$ is $\delta[\Phi]$, where δ: $\pi_2(\tilde{G}/G) \to \pi_1(G)$ is the connecting homomorphism; δ is an isomorphism if $\tilde{G}$ is simply connected. Both $[P]$ and $[\Phi]$ will be referred to as the Higgs charge in the sequel:

$$\pi_1(G) \simeq \pi_1(G)_{\text{free}} + \pi_1(G_{\text{ss}}),$$

where $\pi_1(G)_{\text{free}} \simeq Z^p$, p is the dimension of $Z(\mathscr{G})$, G_{ss} is the semisimple subgroup of G generated by $[\mathscr{G},\mathscr{G}]$, and $\pi_1(G_{\text{ss}})$ is a finite Abelian group. The free part—which plays a particularly important role—is described as follows[25]: Denote by $\Gamma = \{\xi \in \mathscr{G} \,|\, \exp 2\pi\xi = 1\}$ and let z: $\mathscr{G} \to Z(\mathscr{G})$ be the projection of the Lie algebra of G onto its center. The image of Γ under z, $z(\Gamma)$, is a lattice whose dimension is the same as that of $Z(\mathscr{G})$. In Ref. 25 we proved the following theorem: Define, for any loop γ in G,

$$\rho(\gamma) = \frac{1}{2\pi} \int_\gamma z(\theta) \in Z(\mathscr{G}), \tag{5.1}$$

where $\theta = g^{-1}\,dg$ is the canonical (Maurer–Cartan) one-form of G, and ρ defines an isomorphism of $\pi_1(G)_{\text{free}}$ with $z(\Gamma)$.

Any loop in G is known to be homotopic to one of the form $\gamma(t) = \exp 2\pi\xi t$. For this γ, $\rho(\gamma) = z(\xi)$. If $\zeta_1, \ldots, \zeta_p$ is a Z-basis for $z(\Gamma)$, then $\rho(\gamma) = \Sigma m_i \zeta_i$ provides us with p "quantum" numbers $m_1,\ldots,m_p$.

In Ref. 25 we gave also a second characterization of ρ, namely that $\rho(\Phi) = \rho(\delta[\Phi])$ can also be calculated as the *integral of a two-form* over the two-sphere at infinity,

$$\rho(\Phi) = \frac{1}{2\pi} \int_{S^2} \Phi^*\Omega, \tag{5.2}$$

where Ω is the projection to $\tilde{G}/G$ of the $Z(\mathscr{G})$-valued two-form $z(d\theta)$ on G. In a Chern–Weil framework, (5.2) is also expressed as follows[29]: Choose an arbitrary connection A on P, and denote by F its curvature form $F = DA$. The $Z(\mathscr{G})$-valued two-form $Z(F)$ projects then to S^2, and its cohomology class $[z\,(F)] \in H^2_{dR}(S^2) \times Z(\mathscr{G})$ is independent of the choice of the connection. In fact, the map ρ introduced above is just

$$\rho(p) = \frac{1}{2\pi} \int_{S^2} z(F). \tag{5.3}$$

Monopole fields must also satisfy the *Yang–Mills–Higgs equations*. Assuming a sufficiently rapid falloff at infinity, the Yang–Mills–Higgs equation on S^2 reduces to $D\,{*F} = 0$. The solution has been found by Goddard, Nuyts, and Olive[19–22,29,30]: Let us assume that P is a nontrivial G-bundle over S^2, carrying a connection form A that satisfies the YM equation $D\,{*F} = 0$. Then there is a vector Π in $\mathscr{G}$ generating a homomorphism $U(1) \to G$ such that P is associated to the Hopf bundle over S^2 and the field is $F = \mathscr{F}\Pi$ with $\mathscr{F}$ the area form on the two-sphere; Π is quantized, $\exp 4\pi\Pi = 1$. The vector Π can be chosen without loss of generality in any given Cartan subalgebra of $\mathscr{G}$. The transition function h of a monopole is thus homotopic to $h(t) = \exp 4\pi\Pi t$, $0 \leqslant t \leqslant 1$, so $\rho(P)$ is simply

$$\rho(P) = z(2\Pi). \tag{5.4}$$

This theorem can also be reformulated by saying that the holonomy group of asymptotic monopole bundles is a U(1), generated by the "non-Abelian charge" vector Π (see Refs. 29 and 30). Conversely, given Π we are able to construct an asymptotic monopole configuration, see Sec. VI.

VI. REDUCTION OF MONOPOLE BUNDLES

Theorem 6.1: The monopole bundle (P,G) is reducible to an H-bundle Q (where H is a closed subgroup of G) iff

$$[\Phi]\in \operatorname{Im}\sigma_*, \tag{6.1}$$

where $\sigma_*: \pi_2(\tilde G,H) \to \pi_2(\tilde G/G)$ is induced by the natural projection σ: $\tilde G/H \to \tilde G/G$. Equivalently, iff

$$[P] = \delta\,[\Phi]\in \operatorname{Im} i_*, \tag{6.2}$$

where $i_*: \pi_1(H) \to \pi_1(G)$ is induced by the inclusion i: $H \hookrightarrow G$. The inequivalent reductions from G to H are parametrized by the elements of $\pi_2(G/H)$.

Proof: Suppose first that there exists a reduction (Q,H) of (P,G). Then (Q,H) is a reduction of $(\tilde P,\tilde G)$ also, and is thus determined by a reducing map ("Higgs field") Ψ: $S^2 \to \tilde G/H$. This reduces P if and only if each H-coset is contained in a corresponding G-coset. That is, if and only if $\Phi = \sigma(\Psi)$. But this implies $[\Phi] = \sigma_*\,[\Psi]$.

Conversely, if $[\Phi]\in \operatorname{Im}\sigma_*$, then $[\Phi] = \sigma_*\,[\Psi_1]$ for some Ψ_1: $S^2 \to \tilde G/H$. Hence, Φ and $\sigma(\Psi_1)$ are homotopic and thus gauge-equivalent,[9,10,12] so there exists a map $g(x)$ such that $\Phi(x) = g(x)\cdot \sigma(\Psi_1(x))$. But σ is a G-map, so putting

$$\Psi(x) = g(x)\cdot \Psi_1(x), \quad \Phi(x) = \sigma(\Psi(x)),$$

we get a reduction Q of P defined by Ψ. Consider

$$\begin{tikzcd}
 & 0 \arrow[d] & & \\
 & \pi_2(G/H) \arrow[d, "\delta"] & & \\
0 \arrow[r] & \pi_2(\tilde G/H) \arrow[r, "\delta"] \arrow[d, "\sigma_*"'] & \pi_1(H) \arrow[r, "(j\cdot i)_*"] \arrow[d, "i_*"] & \pi_1(\tilde G) \arrow[r] \arrow[d, equal] & \cdots \\
0 \arrow[r] & \pi_2(\tilde G/G) \arrow[r, "\delta"'] & \pi_1(G) \arrow[r, "j_*"'] \arrow[d] & \pi_1(\tilde G) \arrow[r] & \cdots \\
 & & \vdots & &
\end{tikzcd} \tag{6.3}$$

It follows from this diagram, that $[\Phi]\in \operatorname{Im}\sigma_*$ if and only if (6.2) holds.

Finally, by Theorem 2.5 two reductions (Q_1,H) and (Q_2,H) of (P,G) are equivalent iff $[Q_1] = [Q_2]\in\pi_1(H)$. The inequivalent reductions are hence labeled by $\operatorname{Ker} i_*$, which is, according to the diagram, just $\pi_2(G/H)$.

It is instructive to proceed backwards. Let i: $H\subset G$ be a subgroup to any class in $\pi_1(H)$ that is associated on (up to isomorphism unique) H-bundle (Q,H). The extended bundle (P^Q,G) has class $[P^Q] = i_*[Q]\in\pi_1(G)$. Thus Q is a reduction of (P,G) exactly when $(P^Q,G)\simeq(P,G)$, which happens iff

$$i_*[Q] = [P], \tag{6.4}$$

the transition function of Q and P are homotopic in G. Equation (6.4) shows also that the various reductions of (P,G) to an H-bundle are parametrized by $\operatorname{Ker} i_* \simeq \pi_2(G/H)$.

For example, if $\operatorname{Im} i_* = 0$, (in particular when H is simply connected), then (P,G) reduces to a subbundle Q with structure group H if and only if the Higgs charge of Φ is zero and so the G-bundle P is trivial.[31]

Proposition 6.2: The structure group G of P can be reduced to H if and only if h_P is homotopic to a loop in H. In this case

$$\rho_G([P]) = z_G(\rho_H([Q])). \tag{6.5}$$

If $\pi_1(G)$ is free and H is Abelian, then (6.4) is also sufficient.

Indeed, the reductions (P,G) and (Q,H) of $(\tilde P,\tilde G)$ are compatible if and only if the transition function are homotopic in G, $[\exp 4\pi\Pi_P t] = [\exp 4\pi\Pi_Q t]\in\pi_1(G)$. Next, the projection maps z_G: $\mathscr{G} \to Z(\mathscr{G})$ and z_H: $\mathscr{h} \to Z(\mathscr{h})$ satisfy

$$z_G(z_H(\eta)) = z_G(\eta), \quad \eta\in\mathscr{h} \hookrightarrow \mathscr{G}.$$

This follows from $z_G([\mathscr{G},\mathscr{G}]) = 0$, observing that $\eta = z_H(\eta) + \eta'$, where $\eta'\in[\mathscr{h},\mathscr{h}]\subset[\mathscr{G},\mathscr{G}]$. Notice that any H-connection on Q extends naturally to a G-connection on P. Let F denote the curvature. On Q $z_G F = z_G(z_H F)$, since F is H-valued. Equation (6.4) follows then from the definition.

Finally, let $h_P(t) = \exp 2\pi\xi t$ and $h_Q(t) = \exp 2\pi\eta t$ be the transition functions of the bundles P and Q. Then $\rho_G([P]) = z_G(\xi)$ and $\rho_H([Q]) = \eta$. Thus $\rho_G(h_Q) = z_G(\eta) = z_G(\xi)$. Therefore, if $\pi_1(G)$ is free and H is Abelian, h_P and h_Q are homotopic loops in G since ρ_G is now an isomorphism.

Assume now we have a connection on (P,G). Any bundle reduces to its holonomy. For a monopole this means that its connection reduces asymptotically to a connection on a U(1)-bundle. This remark allows us also to construct asymptotic monopole configurations.[32] The non-Abelian charge vector Π is written as $\Pi = (n/2)\xi$, n an integer and ξ a minimal U(1) generator, because Π is quantized. Denote by Y^n the Hopf-bundle S^3/Z_n; $H = \{\exp 2\pi t\xi\,|0\leqslant t\leqslant 1\}$ is a U(1) subgroup of G; Y^n can be viewed also as a principal H-bundle with H-action $y \to yh = h\,(e^{2\pi t})$, for $h = \exp 2\pi t\xi$. The associated bundle $P^\Pi = Y^n\times_H G$ is a principal G-bundle having transition function $h(\theta) = \exp 2\theta\Pi$, $0\leqslant\theta\leqslant 2\pi$ being the angle parametrizing the equatorial circle of S^2. The natural connection $A^n = n\,\bar y\,d\,y/i$ of Y^n extends to a connection A^Π on P^Π; as a matter of fact, $Y^\Pi = (\{y,e\}|y\in Y^n)$ is the holonomy bundle of (P^Π,A^Π). This latter is an asymptotic monopole bundle iff $[P^\Pi]\in\operatorname{Ker} j_*$ for j: $G \hookrightarrow \tilde G$ (see Refs. 22 and 24). Under suitable conditions such asymptotic solutions can be extended to the interior region.[33,34]

The connection on (P,G) determined by the non-Abelian charge vector reduces to a subbundle (Q,H) iff the latter contains the holonomy bundle. This implies the following theorem.

Theorem 6.3: The Yang–Mills connection A of a principle G-bundle P over S^2 defined by the non-Abelian charge Π reduces to a subbundle (Q,H) if and only if $\Pi\in\mathscr{h}$.

For example, (P,G) reduces to a U(1) subgroup $H = \{\exp 2\pi\eta t\,|0\leqslant t\leqslant 1\}$, where η is a minimal generator, iff

$[\exp 2\pi t\eta] = [\exp 4\pi t\Pi] \in \pi_1(G)$. A necessary condition for this is $z(\eta) = 2z(\Pi)$. This is also sufficient if $\pi_1(G)$ has no finite part. The YM connection A reduces also iff the reduced H-bundle Q contains Y^{Π}, which happens iff $2\Pi = n\eta$ for a suitable integer n.

These results yield topological information concerning the "fate" of monopoles under successive symmetry breaking $G \to H$ (see Refs. 30, 34–36). The topological condition found in Refs. 30 and 34 for its survival means exactly that the G-bundle P reduces to an H-bundle Q. On the other hand, the second condition given in Refs. 30, 35, and 36 requires that the Yang–Mills connection reduces also.

VII. THE COLOR PROBLEM IN MONOPOLE THEORY

Let us consider a grand unified monopole (A_j, Φ) with "residual" symmetry group G, where G is the little group of a basepoint Φ_0 in the orbit where the Higgs field takes asymptotically its values. Our previous results imply the following theorem.

Theorem 7.1: A subgroup $K \subset G$ is (rigidly) implementable if and only if, over S^2, the monopole bundle (P,G) reduces to a $Z_G(K)$-bundle Q_1. The rigid actions of K are in 1–1 correspondence with reductions to $Z_G(K)$. The necessary and sufficient condition of implementability of K is

$$\delta\,[\Phi] \in \operatorname{Im} i_*, \tag{7.1}$$

where i_* is the homomorphism between homotopy groups induced by the inclusion map $i: Z_G(K) \hookrightarrow G$. The inequivalent reductions are in 1–1 correspondence with the elements of $\pi_2(G/Z_G(K))$.

An alternative proof is obtained using the inverse technique. Denote in fact $Z_G(K)$ by H and consider an H-bundle Q over S^2. The extended bundle (P^Q, G) carries, as shown in Sec. II, a natural action of K yielding an action on (P,G) iff $(P,G) \simeq (P^Q,G)$. The condition for this is just (6.4) with $H = Z_G(K)$, with ambiguity parametrized by $\pi_2(G/Z_G(K))$. For $K = G$, we have some more results.[1,2]

Proposition 7.2: G is implementable if and only if the transition function $h(t) = \exp 4\pi t\Pi$, $0 \leqslant t \leqslant 1$ of the bundle P is homotopic to a loop in the center of G. This happens iff

$$\delta\,[\Phi] \in \pi_1(G)_{\text{free}} \tag{7.2}$$

and

$$\exp 2\pi\rho_G(P) = \exp 4\pi z(\Pi) = 1. \tag{7.3}$$

The implementation of G is unique.

Indeed, for $i: Z(G) \to G$ $\operatorname{Im} i_*$ belongs to $\pi_1(G)_{\text{free}}$. On the other hand, a $Z(G)$ bundle Q is represented by a loop $\exp 2\pi\zeta t$, $0 \leqslant t \leqslant 1$, where ζ is in $Z(\mathscr{G})$. Then $\rho_H(Q) = z_H(\zeta) = \zeta$. But $\rho_G(P) = \zeta$ by (6.5). However, $\rho_G(P) = z(2\Pi)$.

Conversely, (7.2) and (7.3) are also sufficient: $\gamma(t) = \exp 2\pi\rho_G(P)t$ is now a loop in Im i whose homotopy class is $\delta[\Phi]$, because $\rho_G(\gamma) = \rho_G(P)$ and ρ_G, when restricted to the free part, is an isomorphism. Now G admits at most one implementation, since $G/Z(G)$ is a Lie group, and has thus trivial second homotopy. So i_* is now injective.

To express this result another way, decompose the non-Abelian charge vector as $\Pi = z(\Pi) + \Pi'$, where Π' belongs to $[\mathscr{G},\mathscr{G}]$. Denote by G_{ss} the semisimple subgroup of G whose Lie algebra is $[\mathscr{G},\mathscr{G}]$, and let G^*_{ss} be the simply connected covering group of G_{ss}.

Proposition 7.3: G is implementable if and only if $\exp 4\pi t\Pi'$, $0 \leqslant t \leqslant 1$ is a contractible loop in G_{ss}. This happens iff

$$\exp^* 4\pi\Pi' = 1, \tag{7.4}$$

where $\exp^*$ is the exponential map in G^*_{ss}.

Proof: $z(\Pi)$ commutes with everything, and thus

$$\exp 4\pi\Pi' = (\exp 4\pi\Pi)\cdot(\exp 4\pi z(\Pi))^{-1} = (\exp 4\pi z(\Pi))^{-1}.$$

Equations (7.3) and (7.4) are thus equivalent; in particular, $\exp 4\pi t\Pi'$, $0 \leqslant t \leqslant 1$ is a loop in G_{ss}. Now $[P]$ is decomposed as

$$[P] = [\exp 4\pi t z(\Pi)] + [\exp 4\pi t\Pi'] \in \pi_1(G)_{\text{free}} + \pi_1(G_{ss}), \tag{7.5}$$

and hence (7.2) is equivalent to $\exp 4\pi t\Pi'$ contractible in G_{ss}. But $\pi_1(G_{ss})$ is known to be Γ/Γ^*, where Γ (resp. Γ^*) are the unit lattices of G_{ss} (resp. of G^*_{ss}.) Thus $\exp 4\pi t\Pi'$ contractible means exactly (7.3).

This is seen alternatively by noting[1,7] that, according to the diagram (6.3), $[P] \in \operatorname{Im} i_*$ exactly when $\operatorname{Ad}_*[P] = 0$, i.e., the transition function $\operatorname{Ad} h$ is contractible in $(\operatorname{Aut} G)_0 \simeq \operatorname{Int} G \simeq G/Z(G)$. But the condition for this is just (7.4), since $G/Z(G)$ has $[\mathscr{G},\mathscr{G}]$ for a Lie algebra.

Equation (7.3) can be translated into numbers: let $\zeta_1,\ldots,\zeta_p$ be a Z-basis for $z(\Gamma)$ (assumed nonempty), then $\rho(P) = \Sigma m_j\zeta_j$. On the other hand, there exist least positive integers such that $M_j\zeta_j \in \Gamma$ (see Ref. 24). Thus (7.3) can hold only if, for each j, m_j/M_j is an integer, say n_j. Consequently, we have the following proposition.

Proposition 7.4: G is implementable if and only if (7.4) is valid and

$$m_j = n_j M_j \tag{7.6}$$

for suitable integers n_j. The case $K \neq G$ is similar but more complicated, cf. Ref. 7.

Next, Proposition 4.4 and Theorem 6.3 imply the following theorem.

Theorem 7.5: $K \subset G$ is an internal symmetry group if and only if the loop $h_p(t) = \exp 4\pi t\Pi$ lies in $Z_G(K)$. This happens iff

$$\operatorname{Ad} k\Pi = \Pi, \quad \forall\, k \in K. \tag{7.7}$$

In particular, G is an internal symmetry iff Π lies in the center. The action is then unique.

Indeed, the holonomy group of a monopole-bundle is generated by the non-Abelian charge Π, (7.7) means exactly that $K \subset Z_G(\Pi)$, $[\mathscr{k},\Phi_0] = 0$ is automatically satisfied, since stabilizes Φ_0. Alternatively, the implementation defined by a reduction (Q,H) is a symmetry iff (Q,H) contains the holonomy bundle Y^{Π}.

As an illustration, consider a GUT with residual group $G = \mathrm{SO}(3)$. Such a situation arises, e.g., when $G = \mathrm{SU}(3)$ is broken by a Higgs $\underline{6}$ (see Refs. 20 and 37). Choose in SO(3) the Cartan algebra

$$\mathscr{T} = aL_3 = \begin{bmatrix} 0 & -a & 0 \\ a & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad a \in R. \tag{7.8}$$

The non-Abelian charge vector can be gauge-rotated into $\mathscr{T}$. Then $\Pi = (m/2)\, L_3$, where m is an integer:

$$\pi_1(\mathrm{SO}(3)) = Z_2, \quad [\exp 2\pi m L_3] = m \pmod 2.$$

Topologically nontrivial solutions arise therefore if *m is odd*. Denote by P the corresponding SO(3)-bundle.

Now, $G = \mathrm{SO}(3)$ is not implementable on P: Equation (7.4) would, in fact, require m to be even.

Consider now a U(1) subgroup K with minimal generator ξ, $K = \{\exp 2\pi\xi t\}$, where ξ is conjugate to L_3, and hence

$$\pi_1(\mathrm{SO}(3)) \ni [\exp 2\pi\xi n t\,] = [\exp 2\pi n L_3] = n \pmod 2,$$

$$Z_{\mathrm{SO}(3)}(K) = K,$$

so the interesting part of diagram (6.3) becomes

$$\begin{array}{ccc} \to \pi_1(K) & \xrightarrow{i_*} & \pi_1(\mathrm{SO}(3)) \to \\ \| & & \| \\ Z & & Z_2 \\ n & \to & n \pmod 2 \end{array} \quad . \tag{7.9}$$

Theorems 7.1 and 7.5 tell us, therefore, that (i) K is implementable on P iff n is odd; (ii) for $n = 2k + 1$, Ker $i_* \simeq Z$, there is a different implementation for each k, corresponding to the different reductions to K; and (iii) K is a symmetry iff Π and ξ are parallel, $\Pi = (n/2)\, \xi$ for some integer n.

We are able to construct the bundles explicitly: Choose an integer n and consider the Hopf bundle $Y^n = S^3/Z_n$ where S^3 is viewed as sitting in C^2. Then Y^n is a two-sided U(1) bundle with actions

$$z: y \to z \cdot y = (zy_1, zy_2)$$

and

$$z': y \to y \cdot z' = (y_1 z', y_2 z'),$$
$$y \simeq (y_1, y_2) \in C^2,$$
$$z, z' \in \mathrm{U}(1).$$

Now Y^n can be viewed alternatively as a two-sided principal K-bundle with $k = \exp 2\pi\xi a$ acting as $k \cdot y = (e^{2\pi i a})\, y$ and $y \cdot k = y\,(e^{2\pi i a})$, respectively.

The associated bundle $P^n = Y^n \times_K \mathrm{SO}(3)$, is a right principal SO(3) bundle; Y^n is identified with $Y^\xi = (\{y,e\} | y \in Y^n)$ and so is a reduction of P^n.

The right action of K on Y^n was used to construct $P^{(n)}$. However, we still have a left action of K on Y^n, which extends to a left action of K on P^n according to

$$k\{y,g\} = \{k \cdot y, g\} = \{y, kg\} = \{y,g\} \cdot \mathrm{Ad}\, g^{-1} k, \tag{7.10}$$

where $k = \exp 2\pi\xi a$. [This is the same as (2.4) since U(1) is commutative.] Hence, for $p = \{y,g\}$, $\tau_p(k) = \mathrm{Ad}\, g^{-1} k$, as it should.

The transition function of the principal K-bundle Y^ξ is $h(\theta) = \exp\ \theta\xi n$. Hence, $\pi_1(\mathrm{SO}(3)) \ni [P^n] = n \pmod 2$: P^n is the trivial bundle for n even and is isomorphic to P for n odd. Our construction provides us, hence, with a rigid action of K on P for each odd integer n, as expected. These actions are obviously inequivalent.

The action of K as constructed above is a symmetry for the monopole field A given by the non-Abelian charge vector Π iff Y^ξ contains the holonomy bundle, which happens iff $\Pi = (n/2)\xi$.

ACKNOWLEDGMENTS

We are indebted to J-M. Souriau for hospitality in Marseille. Discussions with L. O'Raifeartaigh and correspondance with P. Nelson and C. Nash are gratefully acknowledged.

P. A. H. would also like to thank the SERC of Great Britain for financial support during part of this work.

[1] P. Nelson and A. Manohar, "Global color is not always defined," Phys. Rev. Lett. **50**, 943 (1983).

[2] A. P. Balachandran, G. Marmo, N. Mukunda, J. S. Nilsson, E. C. G. Sudarshan, and F. Zaccaria, "Monopole topology and the problem of color," Phys. Rev. Lett. **50**, 1553 (1983).

[3] A. Abouelsaood, "Are there chromodyons?" Nucl. Phys. B **226**, 309 (1984).

[4] A. Abouelsaood, "Chromodyons and equivariant gauge transformations," Phys. Lett. B **125**, 467 (1983).

[5] P. Nelson and S. Coleman, "What becomes of global color?," Nucl. Phys. B **237**, 1 (1984).

[6] A. P. Balachandran, G. Marmo, N. Mukunda, J. S. Nilsson, E. C. G. Sudarshan, and F. Zaccaria, "Non-Abelian monopoles break color I. Classical mechanics; II. Field theory and quantum mechanics," Phys. Rev. D **29**, 2919, 2936 (1984).

[7] P. A. Horváthy and J. H. Rawnsley, "Internal symmetries of non-Abelian gauge field configurations," Phys. Rev. D **32**, 968 (1985).

[8] A. P. Balachandran, F. Lizzi, and V. G. Rodgers, "Topological symmetry breakdown in colesterics, nematics, and ^{3}He," Phys. Rev. Lett. **52**, 1818 (1984).

[9] A. S. Schwarz, "On symmetric gauge fields," Commun. Math. Phys. **56**, 79 (1977).

[10] P. Forgács and N. Manton, "Space-time symmetries in gauge theory," Commun. Math. Phys. **72**, 15 (1980).

[11] P. A. Horváthy "The non-Abelian Bohm–Aharonov effect," Marseille preprint CPT-85/P.1769; submitted to Phys. Rev. D.

[12] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry*, Vols. I and II, (Interscience, New York, 1963 and 1969).

[13] J. Harnad, S. Shnider, and L. Vinet, "Group actions on principal bundles and invariance conditions for gauge fields," J. Math. Phys. **21**, 2719 (1980).

[14] C. Duval and P. A. Horváthy, "Particles with internal structure: The geometry of classical motions and conservation laws," Ann. Phys. (NY) **142**, 10 (1982).

[15] F. A. Bais, K. J. Barnes, P. Forgács, and G. Zoupanos, "Dimensional reduction of gauge theories yielding unified models spontaneously broken to SU(3)×U(1)," Geneva preprint CERN-TH. 4045/85, 1985.

[16] A. E. Fischer, "The internal symmetry group of a connection on a principal fiber bundle with applications to gauge field theories," Santa Cruz preprint 1985.

[17] P. A. Horváthy and J. Kollár, "The non-Abelian Bohm–Aharonov effect in geometric quantization," Class. Quant. Gravit. **1**, L61 (1984).

[18] T. T. Wu and C. N. Yang, "Concept of nonintegrable phase factors and global formulation of gauge fields," Phys. Rev. D **12**, 3845 (1975).

[19] P. Goddard, J. Nuyts, and D. Olive, "Gauge theories and magnetic charge," Nucl. Phys. B **125**, 1 (1977).

[20] P. Goddard and D. Olive, "Magnetic monopoles in gauge field theories," Rep. Prog. Phys. **41**, 1357 (1978).

[21] A. Jaffe and C. H. Taubes, *Vortices and Monopoles*, (Birkhäuser, Boston, 1980).

[22] S. Coleman, "The magnetic monopole fifty years later," in *Proceedings of E. Majorana Summer School*, 1981, edited by A. Zichichi (Plenum, New York, 1983).

[23]C. H. Taubes, "Surface integrals and monopole charges in non-Abelian gauge theories," Commun. Math. Phys. **81**, 299 (1981).
[24]P. A. Horváthy and J. H. Rawnsley, "Topological charges in monopole theories," Commun. Math. Phys. **96**, 497 (1984).
[25]P. A. Horváthy and J. H. Rawnsley, "Monopole charges for arbitrary compact gauge groups and Higgs fields in any representation," Commun. Math. Phys. **99**, 517 (1985).
[26]L. Nikolova and V. A. Rizov, "Geometrical approach to the reduction of gauge theories with spontaneously broken symmetry," Rep. Math. Phys. **20**, 287 (1984).
[27]R. Jackiw and N. Manton, "Symmetries and conservation laws in gauge theory," Ann. Phys. (NY) **127**, 255 (1980).
[28]A. Fischer, "Unified approach to conservation laws in general relativity, gauge theories and elementary particle physics," Gen. Relativ. Gravit. **14**, 683 (1982).
[29]P. A. Horváthy and J. H. Rawnsley, "Monopole invariants," J. Phys. A (to appear).
[30]Th. Friedrich and L. Habermann, "The Yang–Mills equations over the two-sphere," Humbolt University Preprint No. 78, Berlin, 1984.
[31]P. A. Horváthy and J. H. Rawnsley, "Survival of grand unified monopoles," Lett. Math. Phys. **9**, 195 (9185).
[32]P. A. Horváthy and J. H. Rawnsley, "On the existence of monopoles in grand unified theories," J. Phys. A **18**, 2647 (1985).
[33]Z. Horváth and L. Palla, "On the structure of generalized monopole solutions in gauge theories," Nucl. Phys. B **116**, 500 (1976).
[34]Z. Horváth and L. Palla, "Monopoles in grand unification theories," Phys. Lett. B **69**, 197 (1977).
[35]F. A. Bais, "The topology of monopoles crossing a phase boundary," Phys. Lett. B **98**, 437 (1981).
[36]G. Lazarides, M. Magg, and Q. Shafi, "Phase transitions and magnetic monopoles in SO(10)," Phys. Lett. B **97**, 87 (1980).
[37]E. J. Weinberg, D. London, and J. L. Rosner, "Z_N monopoles," Nucl. Phys. B **236**, 90 (1984).

Stochastic averaging by functional differentiation

W. Becker
Center for Advanced Studies, Department of Physics and Astronomy, University of New Mexico, Albuquerque, New Mexico 87131

M. S. Zubairy
Department of Physics, Quaid-i-Azam University, Islamabad, Pakistan and Center for Advanced Studies, Department of Physics and Astronomy, University of New Mexico, Albuquerque, New Mexico 87131

(Received 30 July 1985; accepted for publication 13 November 1985)

The stochastic average of exponential expressions that depend at most quadratically upon a Gaussian random function is evaluated using a functional method. The average can be obtained formally for arbitrary forms of the correlation function. The result, however, depends upon a functional determinant and the inverse of a linear operator. Cases are discussed, in particular the Ornstein–Uhlenbeck process, where the latter quantities can be evaluated explicitly.

I. INTRODUCTION

In many areas of physics one encounters problems where a physical quantity depends on a random function that is specified by a real or complex Gaussian stochastic process.[1] Often the result can be obtained in the form of an explicit functional of the random function. As a next step, then, the ensemble average over all possible realizations of the random function has to be carried out. It is this step that we deal with in the present paper, for the case that the functional is essentially an exponential that depends at most quadratically on the random function. This form arises, e.g., as a solution of the differential equation

$$\dot{\rho}(t) = \Big(A_0(t) + f(t)A_1(t) + \int_0^t dt' f(t)A_2(t,t')f(t')\Big)\rho(t), \tag{1.1}$$

where $f(t)$ is the random function. It also occurs in certain situations as a solution of

$$\dot{\rho}(t) = (A(t) + f(t)B(t) + f^*(t)B^\dagger(t))\rho(t), \tag{1.2}$$

when $A(t)$, $B(t)$, and $B^\dagger(t)$ do not commute. Our interest in this problem is mainly motivated by applications in quantum optics. One may think of the interaction of charged particles with chaotic laser fields, in particular multiphoton ionization induced by these fields, the laser bandwidth problem, the effects of pump field fluctuations on parametric processes, or the interaction of a classical fluctuating current with a quantized radiation field, to name just a few examples.

This paper contains few new results. Rather, we given a unified approach based on using functional methods[2] that is able to generate all known results[3–11] and some new ones. For a Gaussian process and functionals of the above-mentioned type, formally carrying out the ensemble average is almost trivial. This is done in Sec. II, and the result is valid for arbitrary correlation functions of the stochastic function, including, of course, white noise and the Ornstein–Uhlenbeck process. However, the general result depends upon a functional determinant as well as the inverse of an operator that is the solution of an integral equation. The actual determination of these two quantities constitutes the core of the problem. Except for the case where the function $A_2(t,t')$ in Eq. (1.1) factorizes, i.e., $A_2(t,t') = A_2(t)A_2(t')$, this cannot be done in general anymore. In Sec. III we then restrict ourselves to the Ornstein–Uhlenbeck process. In this case, for a fairly wide class of functions $A_2(t,t')$, the just-mentioned integral equation can be converted into an inhomogeneous differential equation with constant coefficients, whose solution is, in principle, straightforward, although imposing the appropriate boundary conditions may be tedious. Following the spirit of most applications of the functional formalism, we do not attempt any mathematical rigor. However, in contrast to applications to relativistic quantum field theory, mathematical problems are much less severe here, since all integrals extend over a finite range. The connection between our approach and path integrals is close, and some results have been alternatively derived by following the latter route.[9,11]

II. FORMULATION OF THE PROBLEM AND FUNCTIONAL SOLUTION

We will consider both real and complex Gaussian processes. The real process is characterized by a stochastic function $f(x)$ whose average satisfies

$$\langle f(x)f(x')\rangle = \Delta(x,x') = \Delta(x',x), \tag{2.1}$$

$$\langle f(x_1)\cdots f(x_n)\rangle = \begin{cases} 0, & \text{for } n \text{ odd}, \\ \sum \prod_{(i,j)} \Delta(x_i,x_j), & \text{for } n \text{ even}, \end{cases} \tag{2.2}$$

where the sum is over all different partitions of the n variables $x_1,\dots,x_n$ into $n/2$ disjoint pairs. For the complex process,

$$\langle f(x)f^*(x')\rangle = \Delta(x,x') = \Delta^*(x',x), \tag{2.3}$$

$$\langle f(x_1)\cdots f(x_n)f^*(x_1')\cdots f^*(x_m')\rangle = \delta_{mn}\sum_P \prod_{i=1}^n \Delta(x_i,x'_{P_i}), \tag{2.4}$$

and the sum is over all permutations P of the variables x_i'. The problem is then to evaluate the stochastic average $\langle F\rangle$ of a functional $F[f]$ or $F[f,f^*]$ in the case of a real or complex process, respectively, as defined by Eqs. (2.1)–(2.4).

 0022-2488/86/040991-05$02.50

Let us first consider the real process. Using the functional shift operator we may write

$$F[f]=\exp\left(\int dx\, f(x)\,\frac{\delta}{\delta g(x)}\right)F[g]\Big|_{g=0}. \quad (2.5)$$

The stochastic function $f(x)$ is now exclusively and linearly contained in the exponent. The average is readily carried out with the help of

$$\left\langle \exp\int dx\, f(x)h(x)\right\rangle = \exp\left(\frac{1}{2}\int dx\,dx'\,h(x)\Delta(x,x')h(x')\right), \quad (2.6)$$

which is valid for any (nonstochastic) function $h(x)$. From Eqs. (2.5) and (2.6) we have

$$\langle F[f]\rangle = \exp\left(\frac{1}{2}\int dx\,dx'\,\frac{\delta}{\delta g(x)} \times\Delta(x,x')\,\frac{\delta}{\delta g(x')}\right)F[g]\Big|_{g=0}. \quad (2.7)$$

The analogous formulas for the complex Gaussian process are

$$F[f,f^*]=\exp\int dx\left(f(x)\,\frac{\delta}{\delta g(x)}+f^*(x)\frac{\delta}{\delta g^*(x)}\right) \times F[g,g^*]|_{g=g^*=0}, \quad (2.8)$$

$$\left\langle \exp\int dx(f(x)h(x)+f^*(x)h^*(x))\right\rangle = \exp\int dx\,dx'\,h(x)\Delta(x,x')h^*(x'), \quad (2.9)$$

and

$$\langle F[f,f^*]\rangle = \exp\left(\int dx\,dx'\,\frac{\delta}{\delta g(x)}\Delta(x,x')\,\frac{\delta}{\delta g^*(x')}\right) \times F[g,g^*]|_{g=g^*=0}. \quad (2.10)$$

Whether or not Eqs. (2.7) and (2.10) are useful depends upon whether or not the functional derivatives can be evaluated in compact form.

In many applications, the functional F has the form

$$F[f]=\exp\left[\int_0^t d\tau\,\alpha(\tau)f(\tau) + \frac{1}{2}\int_0^t d\tau\,d\tau'\,f(\tau)B(\tau,\tau')f(\tau')\right] \quad (2.11)$$

or

$$F[f,f^*]=\exp\left[\sigma\int_0^t d\tau(\alpha(\tau)f(\tau)+\alpha^*(\tau)f^*(\tau)) + \int_0^t d\tau\,d\tau'\,f^*(\tau)B(\tau,\tau')f(\tau')\right]. \quad (2.12)$$

In Eq. (2.11), B may be taken as symmetrical, i.e., $B(\tau,\tau')=B(\tau',\tau)$. In Eq. (2.12), σ denotes an arbitrary complex number. Besides being of interest in themselves, the functionals (2.11) and (2.12) can also serve as generating functions via functional differentiation with respect to $\alpha(\tau)$ and $\alpha(\tau)^*$.

We may now use the well-known formulas[2]

$$\exp\left(\frac{1}{2}\frac{\delta}{\delta g}A\frac{\delta}{\delta g}\right)\exp\left(\frac{1}{2}gBg+hg\right) = \exp\tfrac{1}{2}\{gB(1-AB)^{-1}g+h(1-AB)^{-1}g + g(1-BA)^{-1}h+h(1-AB)^{-1}Ah\} \times\exp(-\tfrac{1}{2}\operatorname{Tr}\log(1-AB)) \quad (2.13)$$

and (σ is an arbitrary complex number)

$$\exp\left(\frac{\delta}{\delta g}A\frac{\delta}{\delta g^*}\right)\exp(g^*Bg+\sigma(hg+h^*g^*)) = \exp\{g^*B(1-AB)^{-1}g+\sigma^2h(1-AB)^{-1}Ah^* + \sigma h(1-AB)^{-1}g+\sigma g^*(1-BA)^{-1}h^*\} \times\exp(-\operatorname{Tr}\log(1-AB)) \quad (2.14)$$

to evaluate the functional derivatives in Eqs. (2.8) and (2.10). In the preceding equation we introduced a shorthand matrix notation that deserves some explanation. Lowercase symbols, such as g, h, or $\delta/\delta g$, correspond to vectors and uppercase symbols to matrices. All products are matrix products, so that, e.g.,

$$\frac{\delta}{\delta g}A\frac{\delta}{\delta g^*} = \int_0^t d\tau\,d\tau'\,\frac{\delta}{\delta g(\tau)}A(\tau,\tau')\,\frac{\delta}{\delta g^*(\tau')},$$

$$g^*Bg=\int_0^t d\tau\,d\tau'\,g^*(\tau)B(\tau,\tau')g(\tau),$$

$$gh=\int_0^t d\tau\,g(\tau)h(\tau).$$

All integrations are extended over the same interval, which is the one introduced in Eq. (2.11) or Eq. (2.12). Inverses have the obvious meaning and 1 in the expression $1-AB$ indicates the unit matrix. We will also use the notation $A(\tau,\tau')=\langle\tau|A|\tau'\rangle$ for any matrix A, so that, e.g.,

$$\langle\tau|B(1-AB)^{-1}|\tau'\rangle = \int_0^t d\tau''\,B(\tau,\tau'')\langle\tau''|(1-AB)^{-1}|\tau'\rangle,$$

$$\langle\tau|1|\tau'\rangle=\delta(\tau-\tau').$$

Finally,

$$\operatorname{Tr}\log(1-AB)=\int_0^t d\tau\langle\tau|\log(1-AB)|\tau\rangle = \log\det(1-AB). \quad (2.15)$$

Now, if the functional F is given by Eq. (2.11) or (2.12), we obtain, respectively,

$$\langle F[f]\rangle = \exp(\tfrac{1}{2}\alpha(1-\Delta B)^{-1}\Delta\alpha) \times\exp(-\tfrac{1}{2}\operatorname{Tr}\log(1-\Delta B)) \quad (2.16)$$

or

$$\langle F[f,f^*]\rangle = \exp(\sigma^2\alpha(1-\Delta B)^{-1}\Delta\alpha^*) \times\exp(-\operatorname{Tr}\log(1-\Delta B)). \quad (2.17)$$

These results are valid for arbitrary α, B, and Δ. However, the inverse $(1-\Delta B)^{-1}$ and the $\operatorname{Tr}\log(1-\Delta B)$ cannot, in general, be obtained explicitly. In Sec. III, we will evaluate these quantities for various cases of interest.

III. EXPLICIT SOLUTIONS FOR SPECIAL CASES

In order to make progress beyond Eqs. (2.16) or (2.17) we will have to impose some conditions on the functions $\Delta(\tau,\tau')$ or $B(\tau,\tau')$. The correlation function Δ refers to the stochastic process and B to the functional that is to be averaged. In what follows we will consider the complex Gaussian process, i.e., Eq. (2.17). All results can immediately be rewritten to apply for the real process.

We want to obtain explicit expressions for

$$L = (1-\Delta B)^{-1}\Delta = \Delta(1-B\Delta)^{-1} \tag{3.1}$$

and $\operatorname{Tr}\log(1-\Delta B)$. The operator L is Hermitian if Δ and B are. With the help of the identity

$$(1-\Delta B)^{-1} = 1 + \Delta B(1-\Delta B)^{-1},$$

we can write down an integral equation for L,

$$L = \Delta + \Delta BL, \tag{3.2}$$

or explicitly

$$L(\tau,\tau') = \Delta(\tau,\tau') + \int_0^t d\sigma\langle\tau|\Delta B|\sigma\rangle L(\sigma,\tau'). \tag{3.3}$$

Most often, if Eq. (3.2) for given B can be solved for L, it can be solved as well when B is replaced by λB, where λ is a real number. In this event

$$L(\lambda) = (1-\lambda\Delta B)^{-1}\Delta$$

is known. We then have

$$\operatorname{Tr}\log(1-\Delta B) = -\operatorname{Tr}\int_0^1 d\lambda\, L(\lambda)B. \tag{3.4}$$

In particular cases it is often easier to calculate $\operatorname{Tr}\log(1-\Delta B)$ via its series expansion rather than to use the general equation (3.4).

We will now consider various special cases.

(A) $B(\tau,\tau')=B_0 b(\tau)b^*(\tau')$: Here and in what follows, B_0 denotes an arbitrary complex number, usually either the real or imaginary unit. In this case a simple series expansion succeeds. We may write

$$B = B_0|b\rangle\langle b|, \tag{3.5}$$

with

$$\langle\tau|b\rangle = b(\tau).$$

Now

$$\begin{aligned} L &= \Delta + B_0\Delta|b\rangle\langle b|\Delta \\ &\quad + B_0^2\Delta|b\rangle\langle b|\Delta|b\rangle\langle b|\Delta + \cdots \\ &= \Delta + B_0\Delta|b\rangle\frac{1}{1-B_0\langle b|\Delta|b\rangle}\langle b|\Delta \end{aligned} \tag{3.6}$$

or

$$\begin{aligned} &L(\tau,\tau') \\ &\quad = \Delta(\tau,\tau') \\ &\qquad + B_0\frac{\int_t^0 d\sigma\,\Delta(\tau,\sigma)b(\sigma)\int_t^0 d\sigma'\,b^*(\sigma')\Delta(\sigma',\tau')}{1-B_0\int_t^0 d\sigma\,d\sigma'\,b^*(\sigma')\Delta(\sigma',\sigma)b(\sigma)}, \end{aligned} \tag{3.7}$$

and, analogously,

$$\operatorname{Tr}\log(1-\Delta B) = \log(1-B_0\langle b|\Delta|b\rangle). \tag{3.8}$$

Hence the final answer is

$$\begin{aligned} &\langle F[f,f^*]\rangle \\ &\quad = (1-B_0\langle b|\Delta|b\rangle)^{-1} \\ &\qquad \times\exp\left[\sigma^2\left(\langle\alpha|\Delta|\alpha\rangle + B_0\frac{|\langle\alpha|\Delta|b\rangle|^2}{1-B_0\langle b|\Delta|b\rangle}\right)\right]. \end{aligned} \tag{3.9}$$

The present case includes the situation where $B(\tau,\tau')$ is just a complex number independent of τ and τ', viz.,

$$B(\tau,\tau') = B_0,\quad b(\tau) = 1,$$

in which case

$$\langle b|\Delta|b\rangle = \int_0^t d\tau\,d\tau'\,\Delta(\tau,\tau'),$$

$$|\langle\alpha|\Delta|b\rangle|^2 = \left|\int d\tau\,d\tau'\,\alpha^*(\tau')\Delta(\tau,\tau')\right|^2.$$

These results or special cases thereof have been obtained with the use of various methods by various authors.[3–5] Notice that we did not have to specify the correlation function Δ yet. Obviously, analogous formulas apply when $\Delta(\tau,\tau') = D_0 d(\tau)d^*(\tau')$ while $B(\tau,\tau')$ is arbitrary.

If both $\alpha(\tau)$ and $b(\tau)$ are constant, i.e., $\alpha(\tau) = a$ and $b(\tau) = b$, we can rewrite Eq. (3.9) using the abbreviations

$$X = \int_0^t d\tau f(\tau),\quad F = \int_0^t d\tau\,d\tau'\,\Delta(\tau,\tau') \tag{3.10}$$

as

$$\begin{aligned} &\exp[\sigma(aX + a^*X^*) + B_0 bb^*XX^*] \\ &\quad = (1-B_0 bb^*F)^{-1}\exp\frac{\sigma^2 aa^*F}{1-B_0 bb^*F} \\ &\quad = \int_{-\infty}^{\infty} dX\,dX^*\frac{e^{-XX^*/F}}{2\pi F} \\ &\qquad \times\exp[\sigma(aX + a^*X^*) + B_0 bb^*XX^*]. \end{aligned} \tag{3.11}$$

The last form of Eq. (3.11), in view of

$$\langle XX^*\rangle = F, \tag{3.12}$$

particularly emphasizes the Gaussian character of the process. Equation (3.11) also constitutes a generating function for the moments $\langle(XX^*)^m\rangle$.

(B) $B(\tau,\tau')=B_0\delta(\tau-\tau')$: This case is already much more complicated. In order to make progress we now have to specify the correlation function. We shall use the Ornstein–Uhlenbeck process,

$$\Delta(\tau',\tau) = (2\Gamma/\tau_c)e^{-|\tau-\tau'|/\tau_c} = \Delta(\tau',\tau), \tag{3.13}$$

which should cover all situations of interest if the Gaussian process is stationary. The quantity τ_c is a correlation time and Γ is proportional to the variance of the fluctuations. If $\tau_c\to\infty$ and $\Gamma\to\infty$ while Γ/τ_c remains finite we recover the case (A) we were just considering. On the other hand, for $\tau_c\to 0$, we have

$$\Delta(\tau,\tau') \xrightarrow{\tau_c\to 0} 4\Gamma\delta(\tau-\tau'), \tag{3.14}$$

which describes white noise.

The integral equation (3.3) now reads

$$L(\tau,\tau') = \Delta(\tau,\tau') + B_0 \int_0^t d\sigma\, \Delta(\tau,\sigma) L(\sigma,\tau') \tag{3.15}$$

and L is symmetric with respect to its arguments. Noting that

$$\left(\frac{\partial^2}{\partial \tau^2} - \frac{1}{\tau_c^2}\right)\Delta(\tau,\tau') = -\frac{4\Gamma}{\tau_c^2}\delta(\tau-\tau'), \tag{3.16}$$

we convert Eq. (3.15) into the differential equation

$$\left(\frac{\partial^2}{\partial \tau^2} - \frac{1}{\tau_c^2}\right)L(\tau,\tau') = -\frac{4\Gamma}{\tau_c^2}\delta(\tau-\tau') - \frac{4\Gamma B_0}{\tau_c^2}L(\tau,\tau'). \tag{3.17}$$

The most general solution that is compatible with the symmetry $L(\tau,\tau') = L(\tau',\tau)$ is

$$\begin{aligned} L(\tau,\tau') &= (2\Gamma\tilde{\tau}/\tau_c^2)e^{-|\tau-\tau'|/\tau_c} \\ &\quad + \alpha \sinh(\tau/\tilde{\tau})\sinh(\tau'/\tilde{\tau}) \\ &\quad + \beta \cosh(\tau/\tilde{\tau})\cosh(\tau/\tilde{\tau}) + \gamma \sinh((\tau+\tau')/\tilde{\tau}). \end{aligned} \tag{3.18}$$

Here

$$\tilde{\tau} = (1-4\Gamma B_0)^{-1/2}\tau_c, \tag{3.19}$$

and the contants α, β, and γ will depend on t. The awkward part comes with the determination of these constants. To this end, in Ref. 6 boundary conditions were derived for $L(\tau,\tau')$ by specifying $L(t,t)$, $L(0,0)$, and $L(t,0)$. A more straightforward approach consists in just inserting the general solution (3.18) of the inhomogeneous differential equation (3.17) into the integral equation (3.15). We will not go into any further details since the solution is given in Ref. 6. More or less special cases differing by the form of the term proportional to $f(\tau)$ can be found in the literature.[7–10] Actually, Ref. 10 investigates the case where

$$B(\tau,\tau') = \delta(\tau-\tau')(a\theta(\tau-t_0) + b\theta(t_0-\tau)) \tag{3.20}$$

and $\alpha(\tau) = \text{const}$. Here $\theta(x)$ denotes the Heaviside step function, viz., $\theta(x) = 1$ if $x>0$ and $\theta(x) = 0$ if $x<0$.

(C) $B(\tau,\tau')=B_0\theta(\tau-\tau')$: We can apply essentially the same method as in the previous case. It is worth mentioning that the case $B(\tau,\tau') = B_0\epsilon(\tau-\tau')$, where $\epsilon(x) = 1$ if $x>0$ and $\epsilon(x) = -1$ if $x<0$, is almost identical with the one we are presently considering. We again assume the process to be governed by the correlation function (3.13). As in case (B) we convert the integral equation (3.3) into a differential equation by applying the differential operator $\partial^2/\partial\tau^2 - \tau_c^{-2}$. With the help of Eq. (3.16) we obtain

$$\left(\frac{\partial^2}{\partial \tau^2} - \frac{1}{\tau_c^2}\right)L(\tau,\tau') = -\frac{4\Gamma}{\tau_c^2}\delta(\tau-\tau') - \frac{4\Gamma B_0}{\tau_c^2}\int_0^\tau d\sigma\, L(\sigma,\tau'). \tag{3.21}$$

A further differentiation with respect to τ yields the third-order differential equation

$$\left(\frac{\partial^3}{\partial \tau^3} - \frac{1}{\tau_c^2}\frac{\partial}{\partial\tau} + \frac{4\Gamma B_0}{\tau_c^2}\right)L(\tau,\tau') = -\frac{4\Gamma}{\tau_c^2}\delta'(\tau-\tau'). \tag{3.22}$$

The general solution can be obtained by standard methods and depends on the three roots of the cubic equation

$$\omega^3 + \omega/\tau_c^2 - 4i\Gamma B_0/\tau_c^2 = 0. \tag{3.23}$$

In the present case, $L(\tau,\tau')$ is lacking any obvious symmetry properties. Hence the general solution of Eq. (3.22) as a function of τ and τ' involves nine arbitrary constants that may be determined via the integral equation (3.3). We cut the discussion off at this point, since this case has been solved by a different approach.[11]

Mainly for curiosity, we will mention a few more cases that allow for an analytic solution. Since we are not aware of any physical application we will not work out the solutions in any detail.

(D) $B(\tau,\tau')=B_0/\tau-\tau/$: Again, by applying $\partial^2/\partial\tau^2 - \tau_c^{-2}$ to Eq. (3.3) and differentiating two more times we obtain the fourth-order equation

$$\left(\frac{\partial^4}{\partial t^4} - \frac{1}{\tau_c^2}\frac{\partial^2}{\partial\tau^2} + \frac{8\Gamma B_0}{\tau_c^2}\right)L(\tau,\tau') = -\frac{4\Gamma}{\tau_c^2}\delta''(\tau-\tau'), \tag{3.24}$$

which leads to the characteristic frequencies

$$\omega^2 = -(1/2\tau_c^2)(1 \pm \sqrt{1-32\Gamma B_0\tau_c^2}). \tag{3.25}$$

Obtaining a particular integral of Eq. (3.24) is straightforward. We will not write down a solution, though, since its explicit shape depends crucially upon whether $1 - 32\Gamma B_0\tau_c^2$ is positive, negative, or complex. The general solution of Eq. (3.24) as a function of τ and τ' contains 16 constants. Because of the symmetry of $B(\tau,\tau')$, $L(\tau,\tau')$ will be symmetric so that ten independent constants need to be determined.

(E) $B(\tau,\tau')=B_0\delta'(\tau-\tau')$: This leads essentially back to case (B) since it yields the second-order equation

$$\left(\frac{\partial^2}{\partial \tau^2} + \frac{4\Gamma B_0}{\tau_c^2}\frac{\partial}{\partial\tau} - \frac{1}{\tau_c^2}\right)L(\tau,\tau') = -\frac{4\Gamma}{\tau_c^2}\delta(\tau-\tau'), \tag{3.26}$$

with

$$\omega = (1/\tau_c^2)(-2i\Gamma B_0 \pm \sqrt{4\Gamma^2 B_0^2 - \tau_c^2}). \tag{3.27}$$

Due to lack of symmetry, four constants have to be specified.

Obviously, whenever $B(\tau,\tau')$ is proportional to a derivative of a delta function we can proceed similarly.

(F) $B(\tau,\tau')=B_0\delta''(\tau-\tau')$: This is equivalent to case (B) if in the latter the replacement

$$\tau_c^2 \to \tau_c^2 + 4\Gamma B_0 \tag{3.28}$$

is made.

Finally, we mention another class of possible solutions. After applying $\partial^2/\partial\tau^2 - \tau_c^2$ to Eq. (3.3) we obtain in general

$$\left(\frac{\partial^2}{\partial \tau^2} - \frac{1}{\tau_c^2}\right)L(\tau,\tau') = -\frac{4\Gamma}{\tau_c^2}\delta(\tau-\tau') - \frac{4\Gamma}{\tau_c^2}\int d\sigma\, B(\tau,\sigma)L(\sigma,\tau'). \tag{3.29}$$

Whenever $B(\tau,\tau')$ satisfies a linear homogeneous differential equation with constant coefficients,

$$D(\tau)B(\tau,\tau') = 0, \tag{3.30}$$

say, Eq. (3.29) can be converted into

$$\left(\frac{\partial^2}{\partial\tau^2} - \frac{1}{\tau_c^2}\right) D(\tau)L(\tau,\tau') = -\frac{4\Gamma D(-\tau')}{\tau_c^2}\,\delta(\tau-\tau'),$$

which may be considered either as a second-order equation for $D(\tau)L(\tau,\tau')$, which then would have to be further integrated in order to obtain $L(\tau,\tau')$, or as a higher-order equation directly for L.

ACKNOWLEDGMENT

This work was supported in part by the U.S. Air Force Office of Scientific Research.

[1]N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981).

[2]For a review, see H. M. Fried, *Functional Methods and Models in Quantum Field Theory* (MIT, Cambridge, MA, 1972), see, in particular, p. 42 and Appendix F.

[3]P. Zoller, J. Phys. B **13**, 1249 (1980).

[4]J. L. van Hemmen and K. Rzążewski, J. Phys. B **15**, 1903 (1982).

[5]R. Daniele, F. H. M. Faisal, and G. Ferrante, J. Phys B **16**, 3831 (1981).

[6]W. Becker, M. O. Scully, K. Wódkiewicz, and M. S. Zubairy, Phys. Rev. A **30**, 2245 (1984).

[7]P. Zoller and P. Lambropoulos, J. Phys. B **13**, 69 (1980).

[8]K. Wódkiewicz and M. S. Zubairy, J. Math Phys. **24**, 1401 (1983).

[9]L. S. Brown, Phys. Rev. Lett. **52**, 2013 (1984); Ann. Phys. (NY) **59**, 62 (1985)

[10]K. E. Süsse, W. Vogel, D. G. Welsch, and D. Kühlke, Phys. Rev. A **31**, 2435 (1985).

[11]M. Kuś, K. Rzążewski, and J. L. van Hemmen, J. Phys. A **17**, 1019 (1984).

Statistical filter theory and the O'Doherty–Anstey effect: Dependence on offset

I. Lerche
Department of Geology, University of South Carolina, Columbia, South Carolina 29208

(Received 5 March 1985; accepted for publication 5 December 1985)

This paper considers the effects of multiples, generated at plane bed interfaces, on the characteristics of a mean seismic wave propagating at an angle to the bedding plane. It is found that (a) the multiples produce frequency- and angle-dependent phase shifts to the coherent wave as well as providing an effective attenuation, which is also frequency and angle dependent; (b) a slim angular pencil of monochromatic waves rapidly loses information about its original angular width due to the multiples as the pencil propagates; (c) a seismic pulse, initially traveling at a fixed angle, has both its envelope amplitude and its phase distorted by multiples, and after a short distance of transmission into the medium, the wave shape is nearly completely determined by the generated multiples and is only slightly beholden to the initial pulse shape; (d) the phase and group directions of the mean seismic wave are different than the incident wave's direction and are canted closer to the horizontal with the group direction being the most highly canted; and (e) lateral spreading information of the mean seismic wave is contained in the cross-correlated response of separated geophones and, in principle, can be extracted from cross-correlated measurements. How the generic response depends on the power spectrum of the reflectivity sequence is illustrated by comparing and contrasting results for a transitional sedimentation pattern with those from a cyclic sedimentation pattern; the former produces both a frequency-dependent time delay and attenuation while the latter produces a pure time delay except in the local vicinity of isolated, but periodic frequencies. Numerical estimates, using parameters believed representative of typical seismic conditions, indicate that all of the effects uncovered are large—in the sense that they fall squarely in the regime where they can be expected to have a significant impact both on the subsurface evolution of seismic waves and on interpretations of subsurface conditions made using surface-received signals.

I. INTRODUCTION

In modeling seismic waveforms and spectra it is usually assumed that the dominant loss mechanism is linear anelastic absorption. There are numerous phenomena, however, that indicate that scattering both contributes to attenuation and affects the estimation of seismic attenuation. Short period seismic P waves from relatively simple sources of short time duration, such as nuclear explosions, have prolonged, complex waveforms and show rapid variations of waveform and amplitude with respect to small changes in the locations of sensors.[1] These apparently random variations in the observed P wave characteristics have long been regarded as the results of propagation in random media, with scale lengths of the heterogeneities on the order of 10 to 50 km.[2–4] It was found[2,3] that the assumptions made by Chernov[5] to make the problem of waves in a random medium mathematically tractable are not valid at frequencies above 1 Hz. The amount of scattering is such that perturbation techniques based on weak scattering approximations, such as the first-order Born and Rytov approximations, are invalid at short periods. This can be readily verified by examining the "beam loss" across any large short period array. Recently Powell and Meltzer[6] reached similar conclusions for the SCARLET array in Southern California where they concluded that multiple scattering must be taking place near 1 Hz. In contrast to the theory of Chernov[5] modern multiple scattering techniques such as mean field theory and the augmented parabolic approximation are not limited to the weak-scattering limit and may be applied to the modeling of multiple scattering of waves above 1 Hz.

The problem of attenuation of the seismic wave by heterogeneity has recently been addressed by many authors.[7–16] They have each addressed several aspects of the problem and derived scalar or elastic solutions for plane or spherical waves in randomly perturbed whole spaces. For the most part, these efforts can be separated into $ka \ll 1$, $ka \gg 1$, where k is the wave number and a is the characteristic dimension of the scatterers, or these efforts make the assumption of a heterogeneity scale-length spectrum to calculate a scattering $Q(f)$.

It has been recognized for some time that the Earth contains a spectrum of heterogeneity scale lengths and that waves are affected quite differently by scattering from the two extremes $ka \gg 1$ and $ka \ll 1$. It is also recognized that the most important scattering scale lengths are those for which ka is approximately unity. Following the analytic tradition of Chernov[5] and Tatarski,[17] the spatial autocorrelation of the refractive index fluctuations often has been assumed to be of an analytic form such as a Gaussian, an exponential, or a von Karman distribution. In the scalar wave case, the heterogeneity is often described by the autocorrelation $N(r)$ of the velocity fluctuation $V(r)$, $N(r) = \langle V(r_1)V(r_2)\rangle$. The Earth obviously possesses a more complicated form of the

 0022-2488/86/040996-19$02.50

heterogeneity spectrum and the use of finite difference calculations allows the investigation of these more complicated distributions of heterogeneity scale length.

The results of scalar theory, where the refractive index is the only variable, generally predict that for $ka \gg 1$, the effects are essentially "forward scattering," where the amplitude fluctuations increase with propagation distance and the results may be considered a form of focusing–defocusing with diffraction. In the range of $ka \gg 1$, the averaged, or mean, seismic field attenuates and the mean field attenuation is seen as "beam loss." The spatial attenuation parameter for the mean field is usually found to be proportional to $v^2 k^2 a$, for $ka \gg 1$, where v is the rms variation of the refractive index.[18] The mean field then attenuates like $\exp(-v^2 kakz)$, where z is the distance propagated. The seismic field as recorded on an individual seismometer attenuates less drastically. References 12 and 16 each have predicted the attenuation of the seismic field as measured by a single seismometer as $\exp[-Cv^2(ka)^\gamma kz]$, where C and γ vary according to the heterogeneity autocorrelation function, for example, $C = \pi^{1/2}$, $\gamma = -1$ for a Gaussian. Menke[19] points out that the high-frequency asymptotic form of the attenuation parameter Q due to the scattering from a distribution of scatterers simply reflects the second moments and derivatives of the autocorrelation function $N(r)$ at $r = 0$. Consequently in the real Earth the high-frequency limit may vary significantly depending on the irregularity of the medium.

In the range of $ka \ll 1$, "backscattering" occurs for the scalar wave. It has been shown[20–22] that for an elastic vector wave in a random solid, the scattered field is not confined to a narrow backward-directed angle, but that the radiation pattern is much more complicated and more isotropic. In the region of $ka \ll 1$, often referred to as Rayleigh scattering, the direct wave attenuates with a spatial attenuation proportional to $\exp[-Cv^2(ka)^3 kz]$. Wu[15] predicts $C = \pi^{1/2}/4$ for a Gaussian autocorrelation function and scalar waves.

Most solutions obtained are strictly valid only for the scalar problem. Hudson and Heritage[20] and Hudson[23] have argued that even the simplest scattering due to velocity and density variations in a solid medium will produce scattering of vector waves significantly different from that of scalar wave scattering. These differences include the radiation pattern of a scattered vector wave compared to a scalar wave as well as the conversion between vector waves ($P \rightleftharpoons S$). Furthermore, the influences of even the simplest deterministic structures, such as layers or gradients, make the solutions quite difficult. Even the most recent works[18,24–26] do not address the problems of non-normally incident elastic waves in a vertically inhomogeneous structure with a horizontal random component. Numerical finite difference methods[27] permit the investigation of the exact problem with fairly arbitrary wavelengths and contrasts in elastic media where the two seismic velocities and the density may vary independently but the general trend of an effect is often difficult to discern through the details of the numerical calculation.[28–30] Richards and Menke[31] and Menke[32] have also investigated, in a one-dimensional simulation, the effects of structural randomness of the media on P waveforms and the associated apparent attenuation. They point out that the scattering dispersion associated with random heterogeneities in high Q random media is not likely to have much effect on the spectral attenuation estimates since the slightly delayed high frequencies are still within the data window. It may, however, have a profound effect on time domain attenuation estimates from the waveforms of the first cycles in P (but see Ref. 33 for a quantitative analytic approach to this problem). The results of Menke[32] are limited to $ka < 1$ because of the Haskell–Thompson layer matrix methods used to compute the response of a random stack. For the randomly layered model, Menke[34] predicts an attenuation parameter $Q^{-1} = \frac{1}{2} v^2 ka$, for $ka < 1$. This work was limited to one-dimensional inhomogeneity. Generalization to a two-dimensional elastic medium must allow for the extra degrees of freedom and, presumably, mode coupling between P and S waves must also be included eventually. Consequently, the scattering attenuation is likely to be most important around $10^{-1} \lesssim ka \lesssim 10$.

Elsewhere[33,35,36] we have examined the effect of multiples from many thin beds on the frequency-dependent transmission properties of the coherent component of a seismic signal. We showed that as well as the effective attenuation reported by O'Doherty and Anstey[37] and confirmed by Schoenburger and Levin,[28–30] a frequency-dependent phase factor also exists as reported,[38] and we tied both of the effects to the power spectrum of the reflectivity sequence through a statistical filter theory based on mean field renormalization. The results of the basic theoretical picture developed were in accord, both in magnitude of the frequency-dependent phase and attenuation and in detailed shape, with the transmission results of synthetic seismogram calculations (but see Refs. 35 and 36 for a detailed appreciation of this point).

However, both the theoretical development and the numerical results from the synthetic seismogram calculations treated with a rigorously one-dimensional situation in which a plane wave was normally incident on a series of plane parallel beds. But, in reality, seismic information (travel time, amplitude, and phase) is collected as a function of offset from the source position as well as a function of frequency. Thus it is important to find out what effect multiples have on an otherwise coherent seismic signal when the wave is non-normally incident upon a series of beds. In particular the multiple effects on the dependence of amplitude with offset would seem to be of considerable interest.

To explore simply some of the more basic effects that occur, in this paper we first consider the statistical filtering produced by multiples on a plane wave non-normally incident on a series of plane parallel beds. The theoretical development in Sec. II of the paper obtains the appropriate dispersion relation and relates it to the power spectrum of the reflectivity sequence. In Sec. III we investigate solutions to the dispersion relation for particular functional forms of the reflectivity power spectrum that model sedimentary deposition patterns, and we also develop the corresponding results for a point source of seismic power and demonstrate how the coherent signal is modified as a function of offset by multiple reflections from the many thin beds. Section V presents our conclusions and suggestions for further research into the basic problem of multiple shaping of waveforms.

To set the stage for the remainder of the paper, we present some rough numerical values for relevant parameters. Sound speeds in different lithologic formations vary from around 6000–7000 ft/sec near the sedimentary depositional surface to around 12 000–20 000 ft as sediment depth increases. A typical change in sound speed across the boundary between sedimentary beds of equal density might correspond to a reflection coefficient (for a normally incident plane wave) of about 10% [i.e., $(V_2 - V_1)/(V_2 + V_1) \approx 0.1$]. The sedimentary density, by and large, varies less rapidly with increasing depth than does the sedimentary sound speed, being about 1.6–1.9 gcm^{-3} at deposition and increasing to about 2.6–2.7 gcm^{-3} for fully compacted sediments.

The length L of correlated bedding plane information has been measured[28–30,37,38] in terms of travel time of waves across beds. The corresponding decorrelation frequency f is around 25–50 Hz, which, for an average sound speed of say, 10 000 ft/sec, corresponds to a spatial length, $L (\equiv \frac{1}{2} V/(2\pi f)) \approx 40$–80 ft.

II. CONSTRUCTION OF THE DISPERSION RELATION

Consider a medium in which both the density ρ and propagation velocity V vary only with depth Z. The particle displacement ξ varies with time according to Newton's Law

$$\rho \frac{\partial^2 \xi}{\partial t^2} = -\nabla p, \tag{1}$$

where p is the driving pressure. For an elastic equation of state in which the pressure and the displacement are connected by a Hooke's Law of the form

$$p = -\rho V^2 \nabla \cdot \xi, \tag{2}$$

we can eliminate ξ between Eqs. (1) and (2) to obtain a wave equation for the pressure:

$$\rho \nabla \cdot (\rho^{-1} \nabla p) - V^{-2} \frac{\partial^2 p}{\partial t^2} = 0. \tag{3}$$

Fourier transforming the pressure in Eq. (3) with respect to time [with dependence $\exp(-i\omega t)$] yields

$$\nabla^2 p + \frac{\omega^2}{V(Z)^2} p - \frac{d \ln \rho}{dZ} \frac{\partial p}{\partial Z} = 0, \tag{4}$$

where use has been made explicitly of the assumption that the density is a function only of depth Z.

In order to illustrate simply the basic influence of multiples from many thin beds on an otherwise coherent seismic wave, we treat here the situation in which both the velocity V and density ρ are, *on the average*, constants independent of depth with small amplitude fluctuations around the constant values (see Fig. 1).

Thus we write

$$V = V_0 + \delta V(Z) \quad \text{and} \quad \rho = \rho_0 + \delta\rho(Z), \tag{5}$$

where V_0 and ρ_0 are constants. To quadratic order in the fluctuations of $\delta\rho$ and δV, we can write Eq. (4) in the form

$$\nabla^2 p + \frac{\omega^2}{V_0^2}\left(1 - 2\frac{\delta V}{V_0} + \frac{(\delta V)^2}{V_0^2}\right) p - \frac{\partial p}{\partial Z} \frac{1}{\rho_0} \frac{d\delta\rho}{dZ}\left(1 - \frac{\delta\rho}{\rho_0}\right) = 0. \tag{6}$$

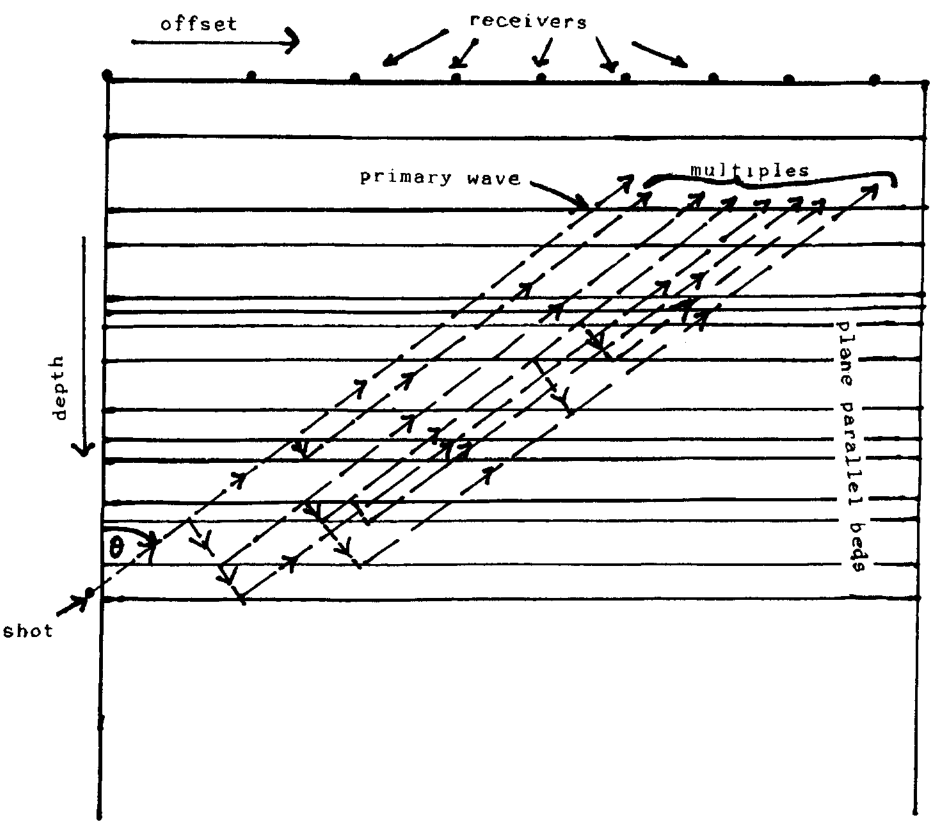

FIG. 1. Schematic illustration of the geometrical situation being envisioned in the paper. The ray associated with the primary wave is that which would occur without bedding plane multiples. A few rays associated with multiple reflection waves are also shown to illustrate their effect in corrupting the otherwise pristine nature of the primary.

Now the Z-dependent fluctuations in velocity and density generate multiples that vary with depth and that impact on what would otherwise have been a completely coherent seismic wave propagating through an homogeneous medium. If the fluctuations in velocity and density occur on a spatial scale small compared to a seismic wavelength, the standard techniques[39,40] of mean field renormalization are particularly appropriate tools to use in unraveling the effect of the multiples on the coherent component of the seismic wave.

The way to reduce the wave equation (6) to a form suitable for investigation of a dispersion relation associated with propagation of the coherent component of the seismic wave is as follows.

First split the pressure field p into a coherent plus an incoherent part

$$p = \langle p\rangle + \delta p\,, \quad \text{with } \langle \delta p\rangle = 0\,, \tag{7}$$

where angular brackets $\langle \quad \rangle$ denote a statistical averaging procedure, i.e., $\langle p\rangle$ is statistically sharp and so represents the coherent component of p.

Then to quadratic order in fluctuating quantities, the coherent component of Eq. (6) is

$$\nabla^2\langle p\rangle + \frac{\omega^2}{V_0^2}\left[1 + \frac{\langle(\delta V)^2\rangle}{V_0^2}\right]\langle p\rangle - \frac{2\omega^2}{V_0^3}\langle \delta V \delta p\rangle + \frac{\partial\langle p\rangle}{\partial Z}\frac{1}{\rho_0^2}\left\langle \delta\rho\frac{d\delta\rho}{dZ}\right\rangle - \left\langle\frac{\partial\delta p}{\partial Z}\frac{1}{\rho_0}\frac{d\delta\rho}{dZ}\right\rangle = 0\,. \tag{8}$$

Subtraction of the coherent equation (8) from Eq. (16) leads to the equation describing the evolution of the fluctuating pressure component correct to quadratic order in fluctuating quantities:

$$\begin{aligned}\nabla^2\delta p + \frac{\omega^2}{V_0^2}\delta p - \frac{2\omega^2}{V_0^3}\delta V\langle p\rangle - \frac{1}{\rho_0}\frac{d\delta\rho}{dZ}\frac{d\langle p\rangle}{dZ} &= \frac{-\omega^2}{V_0^2}\left[(\delta V^2 - \langle\delta V^2\rangle)\langle p\rangle\right] \\ &\quad + \frac{2\omega^2}{V_0^3}\left[\delta V\delta p - \langle\delta V\delta p\rangle\right] \\ &\quad + \left[\frac{\partial\delta p}{\partial Z}\frac{1}{\rho_0}\frac{d\delta\rho}{dZ} - \left\langle\frac{\partial\delta p}{\partial Z}\frac{1}{\rho_0}\frac{d\delta\rho}{dZ}\right\rangle\right].\end{aligned} \tag{9}$$

Now the right-hand side of Eq. (9) contains quadratic products of fluctuating quantities while the left-hand side contains only linear fluctuating quantities. The essence of the mean field renormalization technique is then to declare that the right-hand side of Eq. (9) is ignorable on two counts: first, since the random quadratic products on the right-hand side of Eq. (9) will rapidly become phase incoherent with respect to the linear random terms on the left-hand side of Eq. (9) and, second, because in the coherent mean field equation (8) we see by inspection that the fluctuating pressure component is always multiplied by a second randomly fluctuating quantity prior to statistical averaging. If, therefore, terms on the right-hand side of Eq. (9) were retained, they would give rise to terms in Eq. (8) corresponding to triple products of randomly fluctuating factors. But our initial premise was that we could treat with the full wave equation to quadratic order in fluctuating quantities. Hence as far as the mean field equation (8) is concerned, neglecting the right-hand side of Eq. (9) is a fully self-consistent procedure.[39,40] With this neglection procedure, we can then write Eq. (9) in the form

$$\nabla^2\delta p + \frac{\omega^2}{V_0^2}\delta p = \frac{2\omega^2}{V_0^3}\delta V\langle p\rangle + \frac{1}{\rho_0}\frac{d\delta\rho}{dZ}\frac{d\langle p\rangle}{dZ}\,. \tag{10}$$

Standard methods[41] enable us to write the solution to Eq. (10) in the form

$$\begin{aligned}\delta p = \frac{1}{(2\pi)^3}\int_{-\infty}^{\infty} d^3\mathbf{k}\int_{-\infty}^{\infty}\frac{\exp[i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')]}{k_0^2 - k^2} &\times\left\{2k_0^2\frac{\delta V(Z')}{V_0}\langle p(\mathbf{x}')\rangle\right. \\ &\left. + \frac{1}{\rho_0}\frac{d\delta\rho(Z')}{dZ'}\frac{\partial\langle p(\mathbf{x}')\rangle}{\partial Z'}\right\}d^3\mathbf{x}'\,,\end{aligned} \tag{11}$$

where $k_0 = \omega/V_0$.

Use of Eq. (11) enables us to write the mean field equation (8) in the form

$$\begin{aligned}&\nabla^2\langle p\rangle + k_0^2\left[1 + \frac{\langle\delta V^2\rangle}{V_0^2}\right]\langle p\rangle + \rho_0^{-2}\frac{\partial\langle p\rangle}{\partial Z}\left\langle\delta\rho(Z)\frac{d\delta\rho}{dZ}\right\rangle \\ &\quad - \frac{1}{(2\pi)^3}\int_{-\infty}^{\infty} d^3\mathbf{k}\int_{-\infty}^{\infty} d^3\mathbf{x}'\frac{\exp[i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')]}{k_0^2-k^2} \\ &\quad \times\left\{4k_0^4\langle p(\mathbf{x}')\rangle\left\langle\frac{\delta V(Z)\delta V(Z')}{V_0^2}\right\rangle\right. \\ &\quad \left. + ik_Z\,\rho_0^{-2}\frac{\partial\langle p(\mathbf{x}')\rangle}{\partial Z'}\left\langle\frac{d\delta\rho(Z)}{dZ}\frac{d\delta\rho(Z')}{dZ'}\right\rangle\right\} = 0\,,\end{aligned} \tag{12}$$

where, in the interests of mathematical simplicity and ease of pedagogical exposition, we have assumed that the fluctuations in seismic velocity are completely uncorrelated from the fluctuations in density, and where $k_Z = (\mathbf{k}\cdot\hat{\mathbf{Z}})$.

We note by inspection that Eq. (12) is a linear homogeneous equation in the coherent component of the pressure field. Hence the normal modes of propagation of the mean pressure are controlled by a dispersion relation derivable from Eq. (12). Our immediate task is to obtain the dispersion relation.

To this end we follow conventional wisdom and take the fluctuations in velocity and density to be homogeneously correlated spatially so that we can write

$$\left\langle\frac{\delta V(Z)\delta V(Z')}{V_0^2}\right\rangle = R_{vv}(Z-Z')\,, \tag{13a}$$

$$\left\langle\frac{\delta\rho(Z)\delta\rho(Z')}{\rho_0^2}\right\rangle = R_{\rho\rho}(Z-Z')\,, \tag{13b}$$

with $R_{vv}(-Z) = R_{vv}(Z)$ and $R_{\rho\rho}(-Z) = R_{\rho\rho}(Z)$.

Equation (12) then can be cast in the form

$$\begin{aligned}&\nabla^2\langle p\rangle + k_0^2[1 + R_{vv}(0)]\langle p\rangle \\ &\quad - \frac{1}{(2\pi)^3}\int_{-\infty}^{\infty} d^3\mathbf{k}\int d^3\mathbf{x}'\frac{\exp[i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')]}{k_0^2-k^2} \\ &\quad \times\left\{4k_0^4\langle p(\mathbf{x}')\rangle R_{vv}(Z-Z')\right. \\ &\quad \left. - ik_Z\frac{\partial\langle p(\mathbf{x}')\rangle}{\partial Z'}\frac{\partial^2}{\partial Z'^2}R_{\rho\rho}(Z-Z')\right\} = 0\,,\end{aligned} \tag{14}$$

where use has been made of the symmetry with respect to $\pm Z$ of $R_{\rho\rho}(Z)$ in order to eliminate the term $\langle \delta\rho(Z) d\delta\rho(Z)/dZ \rangle$.

In terms of the Fourier representations

$$\langle p(\mathbf{x}) \rangle = \int_{-\infty}^{\infty} d^3K \, p(\mathbf{K}) \exp[i\mathbf{K}\cdot\mathbf{x}] , \tag{15a}$$

$$R_{vv}(Z) = \int_{-\infty}^{\infty} dK \, R_{vv}(K) \exp(iKZ) , \tag{15b}$$

$$R_{\rho\rho}(Z) = \int_{-\infty}^{\infty} dK \, R_{\rho\rho}(K) \exp(iKZ) , \tag{15c}$$

where $R_{vv}(-K) = R_{vv}(K)$ and $R_{\rho\rho}(-K) = R_{\rho\rho}(K)$ are intrinsically positive definite in order to satisfy Cramer's theorem[42] that the spectral density of an autocorrelation function must be positive definite in order that it represents a real physical correlation,[43] we can write Eq. (14) in the form

$$\int_{-\infty}^{\infty} d^3\mathbf{K} \, p(\mathbf{K}) e^{i\mathbf{K}\cdot\mathbf{x}} \Big\{ k_0^2 [1 + R_{vv}(0)] - K^2 - \int_{-\infty}^{\infty} \frac{d\kappa [4k_0^4 R_{vv}(\kappa) - K_Z \kappa^2 (K_Z + \kappa) R_{\rho\rho}(\kappa)]}{[k_0^2 - K^2 - \kappa^2 - 2\kappa K_Z]} \Big\} = 0 , \tag{16}$$

where $K_Z = (\mathbf{K}\cdot\hat{\mathbf{Z}})$ and $R_{vv}(0) = 2\int_0^\infty R_{vv}(K)\, dK$.

But Eq. (16) must be valid for all values of $\mathbf{x}$. The only way this is possible is if the integrand is identically zero. Since by definition $p(\mathbf{K}) \neq 0$ for all $\mathbf{K}$, it follows that $\mathbf{K}$ must satisfy a dispersion relation given by

$$K^2 = k_0^2 [1 + R_{vv}(0)] - \int_{-\infty}^{\infty} \frac{d\kappa [4k_0^4 R_{vv}(\kappa) - K_Z \kappa^2 (K_Z + \kappa) R_{\rho\rho}(\kappa)]}{[k_0^2 - K^2 - \kappa^2 - 2\kappa K_z]} , \tag{17}$$

which is the relation sought for modes of the mean field. Using the fact that $R_{vv}(\kappa)$ and $R_{\rho\rho}(\kappa)$ are even under the interchange $\kappa \to -\kappa$, in place of Eq. (17) we can write

$$K^2 = k_0^2 [1 + R_{vv}(0)] - 2 \int_0^{\infty} \frac{d\kappa [4k_0^4 R_{vv}(\kappa)(k_0^2 - K^2 - \kappa^2) - \kappa^2 R_{\rho\rho}(\kappa) K_Z^2 (k_0^2 - K^2 + \kappa^2)]}{[(k_0^2 - K^2 - \kappa^2)^2 - 4\kappa^2 K_Z^2]} . \tag{18}$$

Bearing in mind that (a) the modes of the mean pressure field vary as $\exp(i\mathbf{K}\cdot\mathbf{x})$ and (b) the fluctuations in velocity and density vary only with the depth Z, it follows that we can write $\exp(i\mathbf{K}\cdot\mathbf{x}) = \exp(iK_Z Z) \times \exp[i\mathbf{K}_\perp \cdot \mathbf{x}_\perp]$, where $\mathbf{x}_\perp = (x, y, 0)$. Regarding the dispersion relation (18) as an equation determining the complex values of K_Z in terms of real values for $\mathbf{K}_\perp$ and ω, we then can write

$$K_Z^2 = k_0^2 [1 + R_{vv}(0)] - K_\perp^2 - 2 \int_0^{\infty} \frac{d\kappa [4k_0^4 R_{vv}(\kappa)(k_0^2 - \kappa^2 - K_\perp^2 - K_Z^2) - \kappa^2 R_{\rho\rho}(\kappa) K_Z^2 (k_0^2 + \kappa^2 - K_\perp^2 - K_Z^2)]}{[[K_Z^2 - (k_0^2 + \kappa^2 - K_\perp^2)]^2 - 4\kappa^2 (k_0^2 - K_\perp^2)]} . \tag{19}$$

To quadratic order in fluctuating quantities (first order in the power spectra), we can write Eq. (19) in the form

$$K_Z^2 = k_0^2 [1 + R_{vv}(0)] - K_\perp^2 + 2 \int_0^{\infty} \frac{d\kappa [4k_0^4 R_{vv}(\kappa) + \kappa^2 (k_0^2 - K_\perp^2) R_{\rho\rho}(\kappa)]}{[\kappa^2 - 4(k_0^2 - K_\perp^2)]} . \tag{20}$$

The behavior of K_Z with frequency (k_0) clearly depends on the magnitude of the transverse wave number $K_\perp$ relative to k_0, for depending on whether $K_\perp$ is less than or greater than k_0 determines whether the poles in the denominator of the integrand of Eq. (20) lie on the real or imaginary axes, respectively.

Consider each case in turn.

(a) Subcritical waves ($K_\perp < k_0$): In this case we set $K_\perp = k_0 \sin\theta$ so that as $\theta \to 0$ we are dealing with a wave that is incident normally on the layered medium.

With $K_Z = k_0 \cos\theta (1 + F)$, so that the propagation mode can be written $\exp\{i(\omega/V_0)[Z\cos\theta(1+F) + y \times \sin\theta - V_0 t]\}$, we can cast Eq. (20) in the form

$$(1+F)^2 = 1 + R_{vv}(0)\sec^2\theta + 2k_0 \sec^3\theta \int_0^\infty \frac{dx}{(x^2-4)} [4R_{vv}(k_0\cos\theta x) + x^2 \cos^4\theta R_{\rho\rho}(k_0 \cos\theta x)] . \tag{21}$$

For $|F| \ll 1$ Eq. (21) has the approximate solution

$$F = \frac{1}{2} R_{vv}(0)\sec^2\theta + k_0 \sec^3\theta P \int_0^\infty \frac{dx}{(x^2-4)} \times [4R_{vv}(k_0\cos\theta x) + x^2\cos^4\theta R_{\rho\rho}(k_0\cos\theta x)] + i\pi k_0 \sec^3\theta [4R_{vv}(2k_0\cos\theta) + \cos^4\theta R_{\rho\rho}(2k_0\cos\theta)] . \tag{22}$$

For later reference we note that to $O(\theta^2)$ from Eq. (22) we can write $F\cos\theta$ in the form

$$\cos\theta F = F_0(k_0) + \theta^2 F_1(k_0) , \tag{23}$$

with

$$F_0 = \frac{1}{2} R_{vv}(0) + k_0 P \int_0^\infty \frac{dx}{(x^2-4)} \times [4R_{vv}(xk_0) + x^2 R_{\rho\rho}(xk_0)] + i\pi k_0 [4R_{vv}(2k_0) + 4R_{\rho\rho}(2k_0)], \tag{24a}$$

$$F_1 = \frac{1}{4} R_{vv}(0) + k_0 P \int_0^\infty \frac{dx}{(x^2-4)} \times \Big[4R_{vv}(xk_0) - x^2 R_{\rho\rho}(xk_0) - 2x \frac{\partial R_{vv}(xk_0)}{\partial x} - \frac{x^3}{2} \frac{\partial R_{\rho\rho}(xk_0)}{\partial x} \Big] + \frac{1}{2} i\pi k_0 \Big[4R_{vv}(2k_0) - 4R_{\rho\rho}(2k_0) - 2k_0 \frac{\partial}{\partial k_0} R_{vv}(2k_0) - 2k_0 \frac{\partial R_{\rho\rho}(2k_0)}{\partial k_0} \Big], \tag{24b}$$

so that to $O(\theta^2)$ the normal modes of the mean field take the form

$$\exp\Big\{ -i\omega t + \frac{i\omega}{V_0} \times \Big[Z(1+F_0) + y\theta + \theta^2 Z\Big(-\frac{1}{2} + F_1 \Big) \Big] \Big\}. \tag{25}$$

Let us leave aside these mathematical results for the moment and consider the second case.

(b) Supercritical waves ($K_\perp > k_0$): Here we set $K_\perp = k_0 \cosh\theta$ so that as $\theta \to \infty$ we are dealing with a wave that is propagating parallel to the beds of the layered medium. With $K_Z = ik_0 \sinh\theta(1+F)$, so that the normal modes can be written $\exp\{(i\omega/V_0)[iZ\sinh\theta(1+F) + y\cosh\theta - V_0 t]\}$, we can cast Eq. (20) in the form

$$(1+F)^2 = 1 - R_{vv}(0)\operatorname{csch}^2\theta - 2k_0 \operatorname{csch}^3\theta \int_0^\infty \frac{dx}{(x^2+4)} \times [4R_{vv}(k_0 x \sinh\theta) - x^2 \sinh^4\theta R_{\rho\rho}(k_0 x\sinh\theta)]. \tag{26}$$

For $|F| \ll 1$, Eq. (26) has the approximate solution

$$F = -\tfrac{1}{2} R_{vv}(0)\operatorname{csch}^2\theta - k_0 \operatorname{csch}^3\theta \int_0^\infty \frac{dx}{(x^2+4)} \times [4R_{vv}(k_0 x\sinh\theta) - x^2 \sinh^4\theta\, R_{\rho\rho}(k_0 x \sinh\theta)], \tag{27}$$

which is manifestly real so that normal modes that are beyond critical (defined as $K_\perp > k_0$) decay spatially with increasing depth Z.

(c) Critical waves ($K_\perp = k_0$): Some care has to be exercised for waves that are just critical ($K_\perp = k_0$) for then examination of the more rigorous expression (19) yields

$$K_Z^2 = k_0^2 R_{vv}(0) + 2\int_0^\infty \frac{d\kappa[\,+4k_0^4 R_{vv}(\kappa)(\kappa^2 + K_Z^2) + \kappa^2 K_Z^2 R_{\rho\rho}(\kappa)(\kappa^2 - K_Z^2)]}{[K_Z^2 - \kappa^2]^2}, \tag{28}$$

which has a double pole at $\kappa = K_Z$. Evaluation of Eq. (28) is quite tricky since it requires specific, and detailed, knowledge of the functional forms of the power spectra $R_{vv}(\kappa)$ and $R_{\rho\rho}(\kappa)$ due to the highly singular nature of expression (28) coupled with the fact that in the absence of fluctuations K_Z would be precisely zero in this case.

Since only the normal modes that are subcritical ($K_\perp < k_0$) have a propagating behavior with increasing depth in the absence of fluctuations [i.e., they have a factor of the form $\exp(iK_Z Z)$ with K_Z real and positive] and since further the supercritical waves decay exponentially with increasing depth, we concentrate our attention in this paper on the evolution of subcritical waves and modifications to their propagation characteristics brought about by the fluctuations in velocity and density.

Our derivation pertains to the mean field and not necessarily to any single measurement. But what is the mean field? Statistical theory would say it is the average of data recorded through different stratigraphic sections from the same ensemble. Such an average could be achieved by summing along a horizontal profile such that the section changed in detail but kept the same impedance spectrum. More feasible might be to sum in a vertical profile. In either case, the quantity summed is the recorded composite waveform referenced to the primary arrival time.

We can also give an operational interpretation of the statistical fluctuations about the mean field. Equation (A4) of Ref. 35 gives the pressure fluctuations δp in one dimension generated as the mean field encounters reflectors. For a *downgoing* wave the spatial dependence is

$$\langle p(T')\rangle = \langle p(T)\rangle \exp[i\omega(1+F)(T-T')]$$

in terms of travel time T. Substituting this expression into Eq. (A4) gives

$$\frac{\delta p(T)}{\langle p(T)\rangle} = \frac{1}{2}\int_{-\infty}^{T} dT' \frac{\partial \delta\mu(T')}{\partial T'} + \frac{1}{2}\int_T^\infty dT' \frac{\partial\delta\mu(T')}{\partial T'} \exp[i2\omega(T-T')],$$

where we have neglected F compared to unity. The first term vanishes, since it is just the average fluctuation in impedance. This shows that to first order δp represents reflections generated deeper than the detector. (If an *upcoming* mean wave had been chosen instead of a downgoing wave then the fluctuations δp would represent reflections from above.)

Intuitively, the neglect of δp is appropriate for measurements made outside the multiple-generating medium. However, an *in situ* detector such as a VSP (vertical seismic profile) station measures the total field $p = \langle p \rangle + \delta p$, not just the mean field $\langle p \rangle$. In Ref. 47 simulated measurements of attenuation in a VSP profile are presented. They show that estimates of attenuation consist of three components: dissipation, stratigraphic attenuation, and local interference. Our analysis predicts that the stratigraphic component is quantitatively described by the O'Doherty–Anstey formula, while the local interference could be eliminated by using Eq. (10) to subtract δp.

III. SUBCRITICAL WAVES AND SEDIMENTATION PATTERNS

Three aspects of the normal modes of the mean field are of major importance in assessing the coherent response of a seismic wave to the multiples produced by a layered medium. First, we must bear in mind that real seismic signals are not single plane waves propagating at a fixed angle to the bedding plane; rather they are more closely modeled as an angular pencil of waves emanating from an explosive source. We need to investigate how the generic transmission response of a coherent wave is modified by angular superposition of many such waves.

Second, we note that an input seismic signal is made up of many frequencies, not just one, and that the response of a geophone detector is over a band of frequencies. We need to investigate in general how the pristine response of a single normal mode is modified by such bandwidth effects. As we shall see, a consequence of these two aspects entails a discussion of the correlated behavior of the coherent component of the seismic wave in directions parallel and perpendicular to the bedding plane. Third, armed with the results of these two aspects of the normal mode behavior of the mean field, we then can see how various types of sedimentation pattern change the specific and detailed structure of the response. Perhaps two good contrasting illustrations of the type of behaviors possible are derived from sedimentation patterns which O'Doherty and Anstey[37] have labeled transitional and cyclic, respectively. We shall have more to say on these behaviors later in this section of the paper.

We consider each aspect in turn.

A. Angular effects

To illustrate the effects of an angular pencil of waves on the transmission properties of the coherent component of a seismic signal, we consider a slim Gaussian pencil of pressure waves, of angular half width $\Delta\theta$, centered on an initial angle θ_0, emanating from the point $y = 0 = Z$. The frequency-dependent coherent pressure response at position (y, Z) is then given by

$$p(\omega, y, Z) = \frac{p_0}{\pi^{1/2}\Delta\theta}\int_{-\infty}^{\infty} d\theta \exp\bigg[-\frac{(\theta-\theta_0)^2}{(\Delta\theta)^2} + i\frac{\omega}{V_0}(1+F(\theta,\omega))\cos\theta Z + i\frac{\omega}{V_0}\sin\theta y\bigg], \tag{29}$$

with F given by the solution to the dispersion equation (22). Here p_0 is the magnitude of the pressure pulse at angular frequency ω, $(1/\pi^{1/2}\Delta\theta)\exp[-(\theta-\theta_0)^2/(\Delta\theta)^2]$ measures the amplitude of each component of the pressure wave and so defines the Gaussian pencil, while the remaining terms in the exponential represent the propagation effects of the mean wave. The integration range of θ can be accurately represented as $-\infty \leqslant \theta \leqslant \infty$ instead of the true range $-\pi/2 \leqslant \theta \leqslant \pi/2$ as long as the angular width $\Delta\theta$ of the pencil is small compared to $\pi/2$, i.e., the pencil is slim.

To exemplify simply some of the character of the response represented by Eq. (29), we examine here the case of a slim pencil symmetrically located around $\theta_0 = 0$ so that the pencil is organized perpendicularly to the bedding planes. To order θ^2 we can use expression (25) to write the pressure response in the approximate form:

$$p(\omega, y, Z) = \frac{p_0}{\Delta\theta\pi^{1/2}}\int_{-\infty}^{\infty}\exp\bigg\{\frac{i\omega}{V_0}Z(1+F_0) + \frac{i\omega y}{V_0}\theta - \theta^2\bigg[\frac{\omega Z i}{V_0}\Big(+\frac{1}{2}-F_1\Big)+\frac{1}{\Delta\theta^2}\bigg]\bigg\} \times d\theta. \tag{30}$$

Performing the integral over θ in Eq. (30) yields a closed form result for the frequency-dependent pressure response:

$$p(\omega, y, Z) = p_0\frac{\exp\{(i\omega Z/V_0)(1+F_0)\}}{[1+(i\omega Z/V_0)(\Delta\theta)^2(\frac{1}{2}-F_1)]^{1/2}} \times \exp\bigg\{\frac{-\omega^2y^2(\Delta\theta)^2}{4V_0^2[1+(\Delta\theta)^2(i\omega Z/V_0)(\frac{1}{2}-F_1)]}\bigg\}. \tag{31}$$

Inspection of Eq. (31) reveals several interesting effects of multiples on the coherent component of the seismic pencil of waves. First note that as depth Z increases the first exponential factor $\exp[(i\omega Z/V_0)(1+F_0)]$ records both a frequency-dependent fractional phase shift (equivalent to a travel-time delay) of $\mathrm{Re}\,F_0$ relative to the conventional propagation phase variation of $\omega Z/V_0$. In addition the factor $\exp[-(\omega/V_0)Z\,\mathrm{Im}\,F_0]$ records a frequency-dependent effective attenuation of the wave with increasing depth. Since the medium is energy conservative, this effective attenuation of the coherent component of the seismic wave is a consequence of a steady loss of energy from coherent waves being diverted to multiples that are incoherent. These points have been addressed in more detail elsewhere.[35,36]

Of novelty here is the offset-dependent exponential factor in Eq. (31). For distances small compared to the Fresnel length ($Z \lesssim Z_c \equiv V_0/\omega(\Delta\theta)^2$) from the pencil's origin, the offset-dependent exponential factor is $\exp\{-\omega^2y^2(\Delta\theta)^2/4V_0^4\}$ implying that the wave front is laterally coherent in amplitude over a scale

$$y_i \approx (2V_0/\omega\Delta\theta) \simeq Z_c\,\Delta\theta \tag{32}$$

representing the original pencil's structure.

As depth increases to the point where $Z \gtrsim Z_c$, the lateral behavior changes to the form $\exp\{(i\omega y^2/2V_0Z)(1+2F_1)\}$ for $|F_1| \ll 1$, which is insensitive to the original pencil angle. The phase-dependent part represents Fresnel diffraction over a spatial scale

$$y_f \approx [2V_0 Z/\omega(1+2\,\mathrm{Re}\,F_1)]^{1/2}, \qquad (33)$$

which we recognize as the Fresnel length for lateral phase coherence.[43] For $\mathrm{Im}\,F_1 > 0$ the amplitude-dependent part, $\exp\{-(\omega y^2/V_0 Z)\mathrm{Im}\,F_1\}$, represents a frequency-dependent loss of lateral coherence of the mean seismic wave produced by energy diversion to the generated multiples. The offset-scale length y_A that records this effect is

$$y_A \approx \left(\frac{V_0 Z}{\omega\,\mathrm{Im}\,F_1}\right)^{1/2} \approx y_f\left(\frac{1}{2\,\mathrm{Im}\,F_1}\right)^{1/2} \gg y_f. \qquad (34)$$

Thus the lateral Fresnel phase dependence varies on a much shorter spatial scale than the energy diversion scale. For instance, with a seismic wavelength, of about 100 ft ($\lambda = V_0 2\pi/\omega$) at a depth of 10 000 ft the Fresnel phase diffraction length y_f is about 700 ft. As we shall see later, a typical value of F_1 is about 5% so that the amplitude decay offset scale y_A is about 2100 ft. The vertical Fresnel length Z_c is about 2500 ft for a pencil with $\Delta\theta \approx 10^0$.

An alternative way to view the effect is obtained by using cylindrical coordinates $y = R\sin\phi$, $Z = R\cos\phi$ and then a small angle argument ($y \simeq R\phi$, $Z \simeq R$), yielding a scaling behavior for the amplitude decay of $\exp\{-(\omega R/V_0)\phi^2\,\mathrm{Im}\,F_1\}$, i.e., a scattering angle

$$\phi_{sc} \approx (V_0/\omega R\,\mathrm{Im}\,F_1)^{1/2}. \qquad (35)$$

As depth increases ($R \to \infty$), ϕ_{sc} becomes smaller and smaller implying that more and more multiples are removing energy from the coherent beam and that only waves with $\phi \lesssim \phi_{sc}$ remain correlated as the pencil forges deeper into the medium. Again with a wavelength of 100 ft, a depth of 10 000 ft, and the estimate $\mathrm{Im}\,F_1 \approx 5\%$, we obtain the numerical estimate $\phi_{sc} \approx 10^0 \times ((10\,000\ \text{ft})/R)^{1/2}$.

The point to be made here is not the precise numerical values (which can be changed readily by inserting different parameters) but rather that the effects of multiples on the lateral coherence of a pencil of waves sit squarely in the middle of the regime where such effects can have a profound influence on seismic signals. They cannot be ignored with impunity.

As we shall see a little later, the correlated behavior of seismic waves carries a considerable amount of this scattering information.

B. Frequency effects

In the previous section we have seen how generated multiples influence both the infinite angular extent of a pencil of waves and the vertical and lateral dependences of amplitude and phase of what would otherwise be a coherent monochromatic pencil subject only to classical Fresnel wave distortion. Real seismic signals are not monochromatic but rather are made up of many frequencies. The question arises as to the phase and amplitude distortions brought about by the "mixing" of different-frequency multiples into the coherent component of the seismic wave or of the effect of finite bandwidth measurements.

To illustrate this effect we consider the simplified situation of a plane disturbance propagating at a fixed angle θ to the bedding planes and for which the individual Fourier amplitudes of the pressure field on $Z = 0 = y$ are described by $p(\omega) = p_0\exp\{-(\omega-\omega_0)^2/(\Delta\omega)^2\}$, which represents an initial time-dependent disturbance of the form $p(t) \propto \exp(i\omega_0 t - t^2(\Delta\omega)^2/4)$—an oscillatory but decaying pulse bounded by the envelope $\exp(-t^2(\Delta\omega)^2/4)$. The reconstructed mean seismic wave at position (y, Z) at time t is then given by

$$p(\mathbf{x},t) \propto \int_{-\infty}^{\infty} p(\omega)\exp\left\{i\cos\theta\frac{\omega Z}{V_0}(1+F(\omega,\theta)) + iy\sin\theta\frac{\omega}{V_0} - i\omega t\right\}d\omega. \qquad (36)$$

Expanding the exponent in Eq. (36) to quadratic order in $(\omega-\omega_0)$ around $\omega = \omega_0$ yields the form

$$\begin{aligned} p(\mathbf{x},t) \propto{}& \exp\left\{i\frac{\omega_0}{V_0}Z(1+F(\omega_0,\theta)) + iy\frac{\omega_0\sin\theta}{V_0} - i\omega_0 t\right\} \\ &\times\int_{-\infty}^{\infty} dx\,\exp\left\{+ix\left(\frac{y\sin\theta}{V_0}+\frac{Z\cos\theta}{V_0} - t + \cos\theta\frac{Z}{V_0}\omega_0\frac{\partial F(\omega_0,\theta)}{\partial\omega_0}\right)\right. \\ &\left. - x^2\left[\frac{1}{(\Delta\omega)^2} - i\frac{Z\cos\theta}{V_0}\left(\frac{\partial F(\omega_0,\theta)}{\partial\omega_0}+\frac{\omega_0}{2}\frac{\partial^2 F(\omega_0,\theta)}{\partial\omega_0^2}\right)\right]\right\}. \end{aligned} \qquad (37)$$

The integral in Eq. (37) can be performed exactly leading to the expression

$$\begin{aligned} p(\mathbf{x},t) \propto{}& \exp\left\{i\frac{\omega_0 Z}{V_0}(1+F(\omega_0,\theta)) + iy\frac{\omega_0\sin\theta}{V_0} - i\omega t\right\}\left\{1 - i(\Delta\omega)^2\frac{Z\cos\theta}{V_0}\left(\frac{\partial F}{\partial\omega_0}+\frac{\omega_0}{2}\frac{\partial^2 F}{\partial\omega_0^2}\right)\right\}^{1/2} \\ &\times\exp\left\{-\frac{(\Delta\omega)^2}{4}\frac{[y\sin\theta/V_0 + Z\cos\theta/V_0 - t + \omega_0(Z\cos\theta/V_0)(\partial F(\omega_0,\theta)/\partial\omega_0)]^2}{[1 - iZ(\Delta\omega)^2(\cos\theta/V_0)(\partial F/\partial\omega_0 + (\omega_0/2)(\partial^2 F/\partial\omega_0^2))]}\right\}. \end{aligned} \qquad (38)$$

Inspection of Eq. (38) shows that in the absence of multiples ($F \equiv 0$) the pressure pulse is a wave of unchanging shape traveling at the constant speed V_0 and at angle θ_0 through the medium. The influence of multiples is threefold. First we see the presence of multiples reflected in the exponential factor $\exp\{i\omega_0(Z/V_0)(1+F)\}$ representing a phase shift and an effective attenuation of the coherent component of the seismic wave. Second, we see the multiples influencing the position of the peak of the envelope through the factor $\omega_0 Z(\cos\theta/V_0)(\partial F/\partial\omega_0)$ in the second exponential factor. Third, we see a distortion of the pulse shape produced by the factor

$$1 - iZ(\Delta\omega)\,\frac{2\cos\theta}{V_0}\left[\frac{\partial F}{\partial\omega_0} + \frac{\omega_0}{2}\frac{\partial^2 F}{\partial\omega_0^2}\right]$$

in the second exponential.

To estimate the size of the effects we replace the derivatives $\partial F/\partial\omega_0$ and $\partial^2 F/\partial\omega_0^2$ by $|F|/\omega_0$ and $|F|/\omega_0^2$ since this replacement provides a rough idea of the derivatives to within factors of order unity. A typical seismic signal may have a central frequency around 50 Hz and a bandwidth of comparable amount so that when a numerical estimate is required we set $\Delta\omega \approx \omega_0$. Under these conditions the pulse shape distortion factor given above is estimated by $1 - iZ|F|(\Delta\omega)^2\cos\theta/V_0\omega$ so that the distorting influence of the propagation becomes dominant for

$$Z \gtrsim V_0\omega_0/|F|(\Delta\omega)^2 \approx \lambda/|F| = Z_p.$$

Remembering that a typical seismic wavelength is about 100 ft and that $|F| \approx 5\%$, we see that pulse shaping is a major influence for $Z \gtrsim 2000$ ft. Notice also that when $Z \gtrsim Z_p$ the envelope shape becomes

$$\exp\left\{\frac{-i[y\sin\theta + Z\cos\theta - V_0 t + \omega_0 Z\cos\theta\,\partial F/\partial\omega_0]^2}{V_0 4Z\cos\theta(\partial F/\partial\omega_0 + (\omega_0/2)\,\partial^2 F/\partial\omega_0^2)}\right\},$$

which is independent of the bandwidth of the initial pulse representing the fact that once the pulse distortion factor overrides the initial bandwidth effect, the beam shaping is dominated by the multiple production to the extent that the pulse "forgets" its initial spread of frequencies and uses the multiples' travel time delay relative to the primary pulse as its assessment of the "natural" width a pulse should have.

Again we see that the rough order of magnitude estimates of the effects sit squarely in the middle of the range where they can have a significant impact on seismic signals.

Note also that, depending upon the real and imaginary parts of F, the envelope pulse shape changes from real to complex, illustrating the fact that the multiples impress a phase shift as well as an effective attenuation on the coherent seismic wave.

The fractional shift in the peak position of the envelope pulse is about $\omega_0(\partial F/\partial\omega_0) \approx 5\%$ so that, depending on the sign of the derivative $\partial F/\partial\omega_0$ as a function of frequency, the pulse peak may be early or late relative to constant velocity propagation up to $\pm 5\%$ for a total dynamic spread of 10%.

C. Correlated wave behavior and averaging effects

A large majority of the individual effects of multiple reflections can be encompassed by looking at the correlated behavior of the mean seismic wave. Two effects dictate that this is the most likely procedure that will maximize information in a single statement. First, as we have seen in the previous two sections, the initial angular spreading and bandwidth effects are rapidly "forgotten" by the incident wave and replaced by the "natural" spreading angle and "natural" bandwidth produced by the generated multiples. As a consequence, most seismic signals will become independent of initial conditions after penetrating only a few thousand feet into the subsurface. Second, geophones respond over a finite bandwidth and, further, accept incoming signals over a range of angles. The correlation of geophone responses at two different offsets and at different travel times therefore contains information not only of the mean coherent signal at each location but also of the spatial and temporal coherence of that signal.

Perhaps the simplest way to illustrate this behavior is through use of the angular-averaged wave formula (31) in the limit $Z \gtrsim \lambda(\Delta\theta)^2$ so that the wave pressure then has the dominant exponential dependence

$$p(\omega,\mathbf{x}) \propto \frac{p_0}{Z^{1/2}}\exp\left\{\frac{i\omega Z}{V_0}(1 + F_0(\omega)) + \frac{iy^2\omega}{2V_0 Z}(1 + 2F_1(\omega))\right\}. \tag{39}$$

Consider the correlated behavior of the pressure wave as observed at different geophones laterally separated by a distance ξ and with the nearer geophone a distance y from the shot position.

Let the nearer geophone record at a frequency ω_1 and the farther geophone at frequency ω_2.

Define

$$\Gamma_{12}(Z,\xi,y) = p(\omega_1, Z, y)\,p^*(\omega_2, Z, y+\xi). \tag{40}$$

Then from Eq. (39) we obtain

$$\begin{aligned}\Gamma_{12}(Z,\xi,y) = {}& \frac{|p_0|^2}{Z}\exp\left\{-\frac{Z}{V_0}\operatorname{Im}(\omega_1 F_0(\omega_1) + \omega_2 F_0(\omega_2)) - \frac{[y^2\omega_1 F_1(\omega_1) + (y+\xi)^2\omega_2 F_1(\omega_2)]}{V_0 Z}\right\} \\ & \times \exp\left\{\frac{iZ}{V_0}[\omega_1(1 + \operatorname{Re}F_0(\omega_1)) - \omega_2(1 + \operatorname{Re}F_0(\omega_2))]\right. \\ & \left. + \frac{i[y^2\omega_1(1 + 2\operatorname{Re}F_1(\omega_1)) - (y+\xi)^2\omega_2(1 + 2\operatorname{Re}F_1(\omega_2))]}{2V_0 Z}\right\}. \end{aligned} \tag{41}$$

If all pairs of geophones at a fixed lateral separation ξ are considered, then it is possible to construct a statistical average of Eq.

(41) from

$$\gamma_{12}(Z,\xi) \propto \int_{-\infty}^{\infty} dy\, \Gamma_{12}(Z,\xi,y)\,. \tag{42}$$

Performing the integral in Eq. (42) yields

$$\gamma_{12}(Z,\xi) \propto \frac{|p_0|^2}{Z} \exp\left\{\frac{-Z}{V_0} \operatorname{Im}[\omega_1 F_0(\omega_1) + \omega_2 F_0(\omega_2)] + \frac{iZ}{V_0}[\omega_1(1+\operatorname{Re}F_0(\omega_1)) - \omega_2(1+\operatorname{Re}F_0(\omega_2))]\right.$$
$$\left. - \frac{\xi^2}{V_0 Z} \frac{[\omega_2 \operatorname{Im} F_1(\omega_2) + (i/2)\omega_2(1+2\operatorname{Re}F_1(\omega_2))]\,[\omega_1 \operatorname{Im}F_1(\omega_1) - (i\omega_1/2)(1+2\operatorname{Re}F_1(\omega_1))]}{(\omega_1 \operatorname{Im}F_1(\omega_1) + \omega_2 \operatorname{Im}F_1(\omega_2) - (i/2)[\omega_1(1+2\operatorname{Re}F_1(\omega_1)) - \omega_2(1+2\operatorname{Re}F_1(\omega_2))])}\right\}. \tag{43}$$

The conventional definition[43] of a visibility function refers laterally correlated behaviors to their zero separation values as a normalizing factor. Thus with

$$\gamma_{12}(Z,\xi=0) \propto \frac{|p_0|^2}{Z} \exp\left\{-\frac{Z}{V_0} \operatorname{Im}[\omega_1 F_0(\omega_1) + \omega_2 F_0(\omega_2)] + \frac{iZ}{V_0}[\omega_1(1+\operatorname{Re}F_0(\omega_1)) - \omega_2(1+\operatorname{Re}F_0(\omega_2))]\right\}, \tag{44}$$

we can write a laterally dependent normalized visibility function through

$$V_{12}(\xi) = \frac{\gamma_{12}(Z,\xi)}{\gamma_{12}(Z,\xi=0)} = \exp\left\{-\frac{\xi^2[a_1 - ib_1]\,[a_2+ib_2]}{(V_0 Z)[a_1 - ib_1 + a_2 + ib_2]}\right\}, \tag{45}$$

where

$$a_j = \omega_j \operatorname{Im} F_1(\omega_j)\,, \quad b_j = \tfrac{1}{2}\omega_j[1+2\operatorname{Re}F_1(\omega_j)]\,.$$

For $\omega_1 = \omega_2$, the visibility function reduces to what is usually called the intensity visibility function

$$V_{11}(\xi) = \exp\{-\xi^2(a_1^2+b_1^2)/V_0 Z 2a_1\}\,, \tag{46}$$

with

$$\gamma_{11}(Z,\xi=0) \propto (|p_0|^2/Z)\exp\{-(2Z/V_0)\omega_1 \operatorname{Im}F_0(\omega_1)\}\,, \tag{47}$$

so that the correlated lateral behavior declines monotonically with increasing lateral separation ξ with a scale length

$$\xi_c \approx \{8 \operatorname{Im}F_1(\omega) V_0 Z/\omega\}^{1/2}\,, \tag{48}$$

which increases with increasing depth.

While the intensity visibility remains correlated over a larger and larger lateral scale as depth increases, the magnitude of the intensity declines exponentially rapidly with increasing depth as seen in Eq. (47). Note that both the lateral correlation and the vertical decline are both real functions without phase-dependent terms.

Since most geophones can receive over a broad band of frequencies, and the received signals can then be filtered to a narrow band, it is of interest to investigate the correlated behavior given through Eq. (43) when $\omega_2 = \omega_1 + \Delta\omega$, so that phase factors are introduced into the lateral and vertical correlated behaviors.

Then

$$\gamma_{12}(Z,\xi=0) \propto \frac{|p_0|^2}{Z} \exp\left\{-\frac{Z2}{V_0}\omega_1 \operatorname{Im}F_0(\omega_1) - \frac{Z\Delta\omega}{V_0}\right.$$
$$\left.\times\left(\operatorname{Im}F_0(\omega_1) + \omega_1 \frac{\partial \operatorname{Im}F_0}{\partial\omega_1} + i\left[1+\operatorname{Re}F_0(\omega_1) + \omega_1\frac{\partial \operatorname{Re}F_0(\omega_1)}{\partial\omega_1}\right]\right)\right\} \tag{49}$$

and

$$V_{12}(\xi) = \exp\left\{-\frac{\xi^2}{V_0 Z}\right.$$
$$\left.\times \frac{[(\omega I)^2 + (\omega^2/4)(1+2R)^2 + \Delta\omega(I + \omega(\partial I/\partial\omega) + (i/2)(1+2R) + i\omega(\partial R/\partial\omega))(\omega I - (i/2)\omega(1+2R))]}{(2\omega I + \Delta\omega[(i/2)(1+2R) + I + \omega(\partial I/\partial\omega) + i\omega(\partial R/\partial\omega)])}\right\}, \tag{50}$$

where, for brevity, we have written $I \equiv \operatorname{Im}F_0(\omega)$, $R = \operatorname{Re}F_0(\omega)$.

For $\Delta\omega/\omega \lesssim 4I$ we see from Eq. (50) that $V_{12}(\xi)$ reduces to the intensity visibility function (46), while for broader-band signals $\Delta\omega/\omega \gtrsim 4I$, we have

$$V_{12}(\xi) \simeq \exp\{+i\xi^2\omega^2/2V_0 Z\Delta\omega\}\,, \tag{51}$$

which we recognize as a Fresnel zone effect independent of the reflectivity power spectrum, destroying phase coherence for $\xi \gtrsim (2V_0 Z\Delta\omega/\omega^2)^{1/2}$.

In short, the lateral decorrelation of phase information is dominated by geometrical Fresnel zone effects that are independent of the reflectivity power spectrum while the lateral decorrelation of amplitude information is dominated by the generation and propagation of multiples.

And these results are in accord with those reported earlier dealing with angular and bandwidth effects on the propagation characteristics of the mean seismic wave confirming the advertised point that lateral correlation of the wave field contains a considerable fraction of the information embodied in the mean field results of the previous two subsections.

D. Sedimentation pattern effects

In their classic paper, O'Doherty and Anstey[37] identified two major types of sedimentation pattern that have diverse reflectivity power spectra and so produce different effective attenuation effects in the transmission response of a medium to an impressed seismic signal. O'Doherty and Anstey[37] label these patterns as corresponding to transitional and cyclic sedimentation, respectively.

1. Transitional sedimentation

In the case of a transitional sedimentation pattern we[36] have argued that an adequate representation of the power spectrum for illustrative purposes is provided by the generic, Uhlenbeck and Ornstein[44] form

$$R(k) = r^2 K_0/(k^2 + K_0^2)\,, \tag{52}$$

where K_0 is the wave-number scale associated with the decorrelation length [see Refs. 45–47 for explicit examples of well logs satisfying the form given by Eq. (52)]. Note that $\int_{-\infty}^{\infty} R(k)\, dk = \pi r^2$ so that the rms fluctuations are measured by $r\sqrt{\pi}$.

For the purpose of illustrating the expected behavior we choose both $R_{vv}(k)$ and $R_{\rho\rho}(k)$ to be given by the form (52) with different reflectivities r_v^2 and r_ρ^2.

Then, inserting these forms into Eq. (21) enables us to write

$$\begin{aligned}(1+F)^2 - 1 &= \pi r_v^2 \sec^2\theta + 2k_0 K_0 \sec^3\theta \\ &\quad\times \int_0^\infty \frac{dx\,[4r_v^2 + x^2\cos^4\theta\, r_\rho^2]}{(x^2-4)(K_0^2 + x^2 k_0^2\cos^2\theta)} \\ &= F(2+F) \simeq 2F\,. \end{aligned} \tag{53}$$

The integral in Eq. (53) can be done exactly yielding

$$\begin{aligned} F \simeq\, & \frac{\pi}{2(K_0^2 + 4k_0^2\cos^2\theta)} \\ & \times [\, r_\rho^2 K_0^2 + \sec^2\theta\, r_v^2 (K_0^2 - 4k_0^2 \sin^2\theta)\,] \\ & + i\pi k_0 K_0 \sec^3\theta \frac{(r_v^2 + r_\rho^2\cos^4\theta)}{(K_0^2 + 4k_0^2\cos^2\theta)}\,. \end{aligned} \tag{54}$$

Now note that as the frequency is driven to infinity ($k_0 \to \infty$), Eq. (54) yields the asymptotic constant real value

$$F_\infty \simeq (-\pi r_v^2/2)\sin^2\theta\sec^4\theta\,. \tag{55}$$

Following the procedure laid down elsewhere,[1,2] we interpret this value of F_∞ as representing measurements made relative to a travel time associated with infinite frequency. Thus the phase factor $\exp\{ik_0 Z\cos\theta(1+F)\}$ can be written

$$\exp\{ik_0 Z\cos\theta(1+F_\infty)[1 + (F - F_\infty)/(1+F_\infty)]\}\,,$$

illustrating the point that the mean wave phase travels at an effective speed $V_0/(1+F_\infty)$ in the Z direction. Since $|F|$ is considered small compared to unity, to this approximation the wave shaping takes the form $\exp\{ik_0 Z\cos\theta(1+F_\infty)(1+F-F_\infty)\}$ so that $f \equiv F - F_\infty$ measures the frequency-dependent phase and amplitude distortion.

Concern may be expressed that we really cannot run the frequency to infinity since the underlying assumption was that, on the average, the wavelength sampled many beds at once—a condition surely violated as $k_0 \to \infty$. However, our concern here is to find an expression for the phase factor $\exp[ik_0 Z\cos\theta(1+F)]$ from which we can most easily extract the wave—shaping with frequency. To do this effectively it is convenient to write the phase in the form

$$ik_0 Z(1+\Phi)\cos\theta[1 + (F-\Phi)/(1+\Phi)]\,,$$

where Φ is an arbitrary constant of order F. Then to order F we have $ik_0 Z(1+\Phi)[1 + F - \Phi]$. As long as $F(k_0 \to \infty)$ remains finite we may use it as a reference point even if the precise value of F_∞ violates the conditions of derivation of $F(k_0)$. The point is $\Phi(\equiv F_\infty)$ is an arbitrary constant and $F(k_0 \to \infty)$ is a convenient reference.

From Eqs. (54) and (55) we obtain

$$\begin{aligned} f = & \,[\pi K_0^2/2(K_0^2 + 4k_0^2\cos^2\theta)] \\ & \times [r_\rho^2 + \sec^4\theta r_v^2](1 + 2i\cos\theta(k_0/K_0))\,. \end{aligned} \tag{56}$$

At low frequencies such that $2k_0\cos\theta \ll K_0$, the phase shift is nearly independent of frequency at the value $\mathrm{Re} f \approx (\pi/2)[r_\rho^2 + \sec^4\theta\, r_v^2]$ while the attenuative part is linear in frequency with $\mathrm{Im}\, f = (2ik_0/K_0)\mathrm{Re}\, f$.

In terms of an expansion around $\theta = 0$, from Eq. (56) we have

$$f = f_0 + \theta^2 f_1\,, \tag{57}$$

where

$$f_0 = [\pi K_0^2/2(K_0^2 + 4k_0^2)](r_\rho^2 + r_v^2)(1 + 2i\,(k_0/K_0)) \tag{58a}$$

and

$$f_1 = f_0\left[\frac{2k_0^2(1 - ik_0/k_0)}{(K_0^2 + 4k_0^2)} + \frac{2r_v^2}{r_\rho^2 + r_v^2}\right]. \tag{58b}$$

The imaginary part of f_1 is

$$\mathrm{Im}\, f_1 = \frac{k_0\pi K_0(r_\rho^2 + r_v^2)}{(K_0^2 + 4k_0^2)}\left[\frac{2r_v^2}{(r_\rho^2 + r_v^2)} + \frac{(2k_0^2 - K_0^2)}{(K_0^2 + 4k_0^2)}\right], \tag{58c}$$

which is intrinsically positive at all frequencies provided $r_v^2 \gg r_\rho^2$, i.e., provided the rms fluctuations in velocity are greater than those in density.

To provide a numerical estimate of the scattering angle consider the simplified situation when $r_\rho = 0$. Since $\mathrm{Im}\, f_1$ is zero at low and high frequencies, an estimate can be provided using $2k_0 \simeq K_0$ when

$$\mathrm{Im} f_1 \simeq \tfrac{13}{32}\pi r_v^2\,. \tag{59}$$

A typical estimate of a reflection coefficient might be about 0.1 so that $\pi r_v^2 \approx 3\%$, which implies $\mathrm{Im} f_1 \approx 1\%$, which is of the same order as the 5% estimate used previously. Since both r_ρ and the precise ratio of wavelength to scale constant K_0 are variable, the 1% estimate is not at all discordant with the previous estimate that was used for illustrative purposes. With $\mathrm{Im} f_1 = 1\%$ the scattering angle ϕ_{sc} is about 23 (10 000 ft/R)$^{1/2}$ degrees.

Note that the transitional sedimentation pattern, as illustrated with the Ornstein–Uhlenbeck power spectrum, provides for both a frequency-dependent phase shift and effective attenuation, which are angle dependent leading to angular beam spreading and frequency dispersive effects with bandwidth as noted in the previous section of the paper. The attenuative part has a peak around $k_0/K_0 = O(1)$ and declines to zero at both low and high frequencies while the phase shift is constant at low frequencies and declines to zero as the frequency tends to infinity.

2. Cyclic sedimentation

In the case of a cyclic sedimentation pattern, we[36] have represented the power spectrum by the illustrative form

$$R(k) = r^2 \sum_{n=1}^{\infty} \delta(k - nK_0) f(n) , \qquad (60)$$

so that

$$\int_{-\infty}^{\infty} R(k)\, dk = r^2 \sum_{n=1}^{\infty} f(n) ,$$

where the cyclic nature of the sedimentation pattern is mirrored through the δ-function periodicity in wave-number space of Eq. (60), and where each positive $f(n)$ measures the fractional strength of the wave-number periodicity. For purposes of illustrating the expected behavior, we choose both $R_{vv}(k)$ and $R_{\rho\rho}(k)$ to be given by the functional form (60) with scaling values r_v and r_ρ, respectively.

Inserting these functional forms into Eq. (21) then enables us to write, for $|F| \ll 1$, that

$$(F+1)^2 - 1 \simeq 2F \simeq r_v^2 \sec^2\theta \sum_{n=1}^{\infty} f(n) + 2\sec^2\theta \sum_{n=1}^{\infty} f(n) \frac{[4r_v^2 k_0^2 + n^2K_0^2 \cos^2\theta\, r_\rho^2]}{(n^2K_0^2 - 4k_0^2\cos^2\theta)} . \qquad (61)$$

Equation (61) yields an intrinsically real value for F. The only dichotomous points occur when $2k_0\cos\theta = nK_0$, since then the denominator in the second sum in Eq. (61) is zero. Tracing back the origin for this anomalous behavior, we see that it arises because the denominator in Eq. (19) was approximated to quadratic order in fluctuating quantities, which is justifiable only so long as the resulting expression for F is small compared to unity and nowhere infinite. In the particular illustrative case of the cyclic sedimentation pattern represented by Eq. (60) as long as attention is focused away from the critical points $2k_0\cos\theta = nK_0$, F is real and smaller than unity, representing a frequency-dependent phase shift with zero attenuation in accord with the numerical synthetic seismogram results[38] and done for $\theta = 0$, and in line with the one-dimensional wave propagation theoretical results.[35,36]

Before considering the singular points about $2k_0\cos\theta = nK_0$, we first look at Eq. (61) in a little more detail. First note that as $k_0 \to \infty$, Eq. (61) yields

$$F_\infty = \frac{1}{2} r_v^2 \sum_{n=1}^{\infty} f(n) \sec^2\theta (1 - 2\sec^2\theta) , \qquad (62)$$

so that the relative phase delay, $f \equiv F - F_\infty$, is given by

$$f = (r_\rho^2 + r_v^2 \sec^4\theta) \sum_{n=1}^{\infty} \frac{f(n) n^2 K_0^2}{(n^2K_0^2 - 4k_0^2\cos^2\theta)} . \qquad (63)$$

Note that for $2k_0\cos\theta < K_0$ the relative phase f is of one sign and is approximately given by

$$f \simeq (r_\rho^2 + r_v^2\sec^4\theta) \left(\sum_{n=1}^{\infty} f(n) \left[1 + \frac{4k_0^2\cos^2\theta}{n^2K_0^2} \right] \right) > 0 , \qquad (64)$$

which is very nearly independent of frequency so that an overall constant fractional time delay of $(r_\rho^2 + r_v^2\sec^4\theta)\Sigma_{n=1}^{\infty} f(n)$ can be associated with low frequencies in line with the constant time delay results (of 20%) recorded[38] in synthetic seismogram numerical calculations.

For the singular frequencies when $2k_0\cos\theta = nK_0$ ($n \geqslant 1$) more care has to be exercised. Inserting the functional form (60) into the more precise dispersion relation (19) and setting $K_Z = k_0\cos\theta(1+F)$, $K_\perp = k_0\sin\theta$ we obtain

$$F(F+2)\cos^2\theta = r_v^2 \sum_{m=1}^{\infty} f(m) + 2\sec^2\theta \sum_{m=1}^{\infty} f(m) \times \frac{[4r_v^2(4m^2/n^2 + F(F+2)) + (4m^2/n^2)\cos^4\theta\, r_\rho^2 (1+F)^2 (4m^2/n^2 - F(F+2))]}{16(m^2/n^2)(m^2/n^2 - 1) + F(F+2)[F(F+2) - 8m^2/n^2]} . \qquad (65)$$

Inspection of Eq. (65) shows that discarding the terms factored by F in the denominator of the sum on the right-hand side leads immediately to a singular behavior for $m = n$. Since F is considered small, this neglect can readily be justified for $m \neq n$, but the dominant term in the series is then provided by the singular term with $m = n$.

Ignoring the terms with $m \neq n$ in the second sum enables us to cast Eq. (65) in the approximate form

$$F(2+F)\cos^2\theta = r_v^2 \sum_{m=1}^{\infty} f(m) - 2\sec^2\theta f(n) \frac{[4r_v^2(4 + F(2+F)) + 4\cos^4\theta\, r_\rho^2 (1+F)^2 (4 - F(F+2))]}{F(F+2)[8 - F(F+2)]} . \qquad (66)$$

For $|F| \ll 1$, we can write Eq. (66) in the form

$$(2F\cos^2\theta)^2 - r_v^2 \sum_{m=1}^{\infty} f(m)(2F\cos^2\theta) + 4f(n)[r_v^2 + \cos^4\theta\, r_\rho^2] = 0, \tag{67}$$

with solution

$$4F\cos^2\theta = r_v^2 \sum_{m=1}^{\infty} f(m) \pm \left\{\left(\sum_{m=1}^{\infty} f(m)\right)^2 r_v^4 - 16f(n)(r_v^2 + \cos^4\theta\, r_\rho^2)\right\}^{1/2}. \tag{68}$$

For r_v^2 and r_ρ^2 small compared to unity, Eq. (68) provides the approximate attenuative behavior

$$F\cos^2\theta \approx +i\,[f(n)(r_v^2 + \cos^4\theta\, r_\rho^2)]^{1/2}. \tag{69}$$

Thus in the vicinity of the singular points where $2k_0\cos\theta = nK_0$, the dominant response is heavy attenuation; the width around the singular points over which attenuation is the major effect, as opposed to the nearly constant low-frequency time delay, can be estimated by balancing the magnitudes of the two terms in the denominator of Eq. (65), i.e., a dynamic range

$$\Delta k_0 = (K_0/4)\sec\theta\,|F| \tag{70}$$

around $k_0 = nK_0 \sec\theta/2$, where F is given by Eq. (69).

Thus the basic response of the mean seismic wave to the cyclic pattern of sedimentation as represented by Eq. (60) is to produce a pure phase shift that is nearly independent of frequency at low frequencies and so mimics a frequency-independent time delay; exceptions occur in narrow bands centered on isolated wave numbers at multiples of $\frac{1}{2}K_0\sec\theta$, where heavy attenuation is dominant.

Since the imaginary part of F is almost zero at almost all frequencies, in those domains there is very little scattering of the original beam—it preserves its original angular width and its pulse structure very well compared to a sedimentary pattern more dominated by a transitional character. The major effect is to produce an angle-dependent time delay for $2k_0\cos\theta < K_0$, but both the frequency dependence and the angle dependence of the relative phase as given by Eq. (63) are slight.

E. Phase and group velocities and their directions of propagation

The mean coherent wave at angular frequency ω can be written in the form $\exp((i\omega/V_0)\Phi)$, where the complex wave phase Φ is given by

$$\Phi = Z\cos\theta(1 + F_\infty(\cos\theta))\left(1 + \frac{F - F_\infty}{1 + F_\infty}\right) + y\sin\theta - V_0 t. \tag{71}$$

At very high frequencies we have $F \to F_\infty$, so that the complex phase takes on the asymptotic form

$$\Phi_\infty = Z\cos\theta\,[1 + F_\infty(\cos\theta)] + y\sin\theta - V_0 t, \tag{72}$$

which is real.

It is of some considerable interest to determine the directions and magnitudes of propagation of the phase and group velocities for the coherent component of the seismic wave for they define the sort of signature that one could identify as a signal—both in terms of timing and of position. Perhaps the easiest way to place this behavior within a familiar framework is to use the formal method of development set out by Ginzburg.[48]

Basically one notes that any disturbance $\psi(\mathbf{x},t)$ can be written

$$\psi(\mathbf{x},t) = \int \psi(\mathbf{k})\exp[i\mathbf{k}\cdot\mathbf{x} - i\omega(\mathbf{k})t]\, d^3\mathbf{k}, \tag{73}$$

where a dispersion relation [in our case Eq. (19)] connects ω to $\mathbf{k}$ with

$$\omega = \omega(\mathbf{k}).$$

In Ginzburg's[48] development it is $\mathbf{k}$ that is taken to be real and ω that is allowed to be complex. However, in our case we treat with real frequencies and determine the real and imaginary parts of k_z from a dispersion relation. The physical difference is that a complex frequency and real wave vector corresponds to a *truly* attenuative system in which the disturbance $\psi(\mathbf{x},t)$ tends to zero at *all* spatial locations including those close to the source as time becomes large, whereas a complex wave vector and real frequency corresponds to an *apparently* attenuative system in which the disturbance $\psi(\mathbf{x},t)$ tends to zero exponentially only at spatial locations infinitely far removed from the source, i.e., the wave undergoes frustrated transmission but is not absorbed.[43]

To modify the formal mathematical development to our purposes we write

$$\psi(\mathbf{x},t) = \int \psi(k_y,\omega)\exp[ik_z(\omega,k_y)Z + ik_y y - i\omega t]\, dk_y\, d\omega, \tag{74}$$

with ω and k_y real and where a dispersion relation connects k_z to ω and k_y through

$$k_z^2 = k_z(\omega,k_y)^2, \tag{75}$$

and where, in general, k_z has both real and imaginary parts. The standard argument for determining phase and group velocities considers $\psi(k_y,\omega)$ to be a highly peaked function around the particular real values $k_y = K_y$ and $\omega = \Omega$, and it is conventional to write

$$\psi(k_y,\omega) \propto \exp\left[-\frac{(k_y - K_y)^2}{(\Delta k)^2} - \frac{(\omega - \Omega)^2}{(\Delta\omega)^2}\right], \tag{76}$$

where Δk and $\Delta\omega$ are extremely small—in the sense that convergence of the integral in Eq. (74) is controlled by $\psi(k_y,\omega)$. Insertion of (76) into (74) and changing variables through $k_y = K_y + \Delta k u$, $\omega = \Omega + \Delta\omega V$ yields

$$\psi(\mathbf{x},t) \propto \exp[iK_y y - i\Omega t]\int_{-\infty}^{\infty} du \int_{-\infty}^{\infty} dv \times \exp\{-(u^2 + v^2) - i\Delta\omega vt + i\Delta k uy + iZk_z(\Omega + \Delta\omega v, K_y + \Delta k u)\}. \tag{77}$$

The assumption that $\Delta\omega$ and Δk are small then enables an expansion of k_z to be performed to order $\Delta\omega$ and Δk:

$$k_z(\Omega+\Delta\omega v, K_y+\Delta k u)$$
$$\simeq k_z u(\Omega_y, K)+\Delta\omega v\frac{\partial k_z(\Omega,K_y)}{\partial\Omega}$$
$$+\Delta k u\frac{\partial k_z(\Omega,K_y)}{\partial K_y}+\cdots. \tag{78}$$

When this expansion is used in Eq. (77) we obtain the form

$$\psi(\mathbf{x},t)\propto\exp\{ik_z(\Omega,K_y)Z+iK_y y-i\Omega t\}$$
$$\times\int_{-\infty}^{\infty}du\exp\left\{-u^2+iu\Delta k\left(y+Z\frac{\partial k_z(\Omega,K_y)}{\partial K_y}\right)\right\}$$
$$\times\int_{-\infty}^{\infty}dv\exp\left\{-v^2+iv\Delta\omega\left(Z\frac{\partial k_z(\Omega,K_y)}{\partial\Omega}-t\right)\right\}. \tag{79}$$

Performing the u and v integrals in Eq. (79) yields

$$\psi(\mathbf{x},t)\propto\exp[ik_z(\Omega,K_y)Z+iK_y y-i\Omega t]$$
$$\times\exp\left\{-\frac{(\Delta k)^2}{4}\left[y+Z\frac{\partial k_z(\Omega,K_y)}{\partial K_y}\right]^2\right.$$
$$\left.-\frac{(\Delta\omega)^2}{4}\left[t-\frac{Z\,\partial k_z(\Omega,K_y)}{\partial\Omega}\right]^2\right\}. \tag{80}$$

Provided k_z is real, we see that the behavior of ψ is the product of a phase-dependent part and an amplitude-dependent part.

Inspection of the amplitude dependence in Eq. (80) shows that $|\psi(\mathbf{x},t)|$ declines temporally *and* spatially except along a line given by

$$t=Z\frac{\partial k_z(\Omega,K_y)}{\partial\Omega} \tag{81a}$$

and

$$y=-Z\frac{\partial k_z(\Omega,K_y)}{\partial K_y}. \tag{81b}$$

Defining

$$V_{gz}=\left(\frac{\partial k_z(\Omega,K_y)}{\partial\Omega}\right)^{-1} \tag{82a}$$

and

$$V_{gy}=-V_{gz}\frac{\partial k_z(\Omega,K_y)}{\partial K_y}, \tag{82b}$$

the line in Eq. (80) can be written $Z=V_{gz}t$, $y=V_{gy}t$ and represents the line on which the highly peaked (in k_y,ω space) pulse retains its amplitude to order (Δk) and $\Delta\omega$ *provided k_z is real.*

The tight group of waves centered on K_y,Ω then propagates with unchanging amplitude at the group speed

$$V_g=(V_{gz}^2+V_{gy}^2)^{1/2}, \tag{83}$$

and the group direction of propagation is given by

$$\tan\Phi_g=V_{gy}/V_{gz}, \tag{84}$$

where the group angle Φ_g is measured from the vertical to the bedding planes.

Inspection of the phase factor in Eq. (80) shows that the phase of the packet of waves is conserved along a line defined by

$$k_z(\Omega,K_y)Z+K_y y-\Omega t=\text{const}, \tag{85}$$

provided k_z is real.

Differentiating Eq. (85) with respect to time and setting $dZ/dt=V_{pz}$, $dy/dt=V_{py}$ we obtain

$$k_z V_{pz}+k_y V_{py}-\Omega=0. \tag{86}$$

Setting the phase speed

$$V_p=\Omega/(k_z^2+k_y^2)^{1/2}, \tag{87}$$

we can then write the Z and y components of phase velocity as

$$V_{pz}=V_p\cos\psi,\quad V_{py}=V_p\sin\psi, \tag{88}$$

where the angle of propagation ψ of the constant phase line is given by

$$\tan\psi=\mathrm{k}_y/\mathrm{k}_z. \tag{89}$$

When k_z is *complex*, the uses of phase and group velocity are compromised since the separation of the integral of a highly peaked pulse (in k_y,ω space) into a phase and amplitude factor is no longer a simple matter. Indeed we have seen in previous sections of the paper that the effect of a complex k_z is to distort an original pulse both in phase and amplitude to the point that the wave shape "forgets" its initial conditions, and we estimated these phase diffraction and effective attenuation effects for simple situations as functions of frequency and wave propagation angle.

Here we want to concentrate our attention on the pulse propagation properties. In order to do so effectively, it is convenient to consider propagation at high frequencies when, as we have shown earlier, the wave behavior is characterized by

$$k_z=(\omega/V_0)(1+F_\infty(\cos\theta))\cos\theta \tag{90a}$$

and

$$k_y=(\omega/V_0)\sin\theta, \tag{90b}$$

where, for the moment, we regard $\cos\theta$ as a parametric variable that in principle could be eliminated between (90a) and (90b) to provide us with the relationship $k_z^2=k_z^2(\omega,k_y)$, which is required in order to compute group and phase velocities and their directions of propagation.

From Eq. (90) we obtain the phase speed

$$V_p=V_0\{1-F_\infty(\cos\theta)\cos^2\theta+O(F_\infty^2)\} \tag{91a}$$

and phase direction

$$\tan\psi=\tan\theta\,[1-F_\infty(\cos\theta)+O(F_\infty^2)]. \tag{91b}$$

Likewise the group speed components are given by

$$V_{gz}=V_0\cos\theta\left\{1-F_\infty+\sin\theta\cos\theta\frac{\partial F_\infty}{\partial\theta}+O(F_\infty^2)\right\}, \tag{92a}$$

$$V_{gy}=V_0\sin\theta\left\{1-\cot\theta\cos^2\theta\frac{\partial F_\infty}{\partial\theta}+O(F_\infty^2)\right\}, \tag{92b}$$

with

$$V_g=V_0\,[1-F_\infty\cos^2\theta+O(F_\infty^2)] \tag{93a}$$

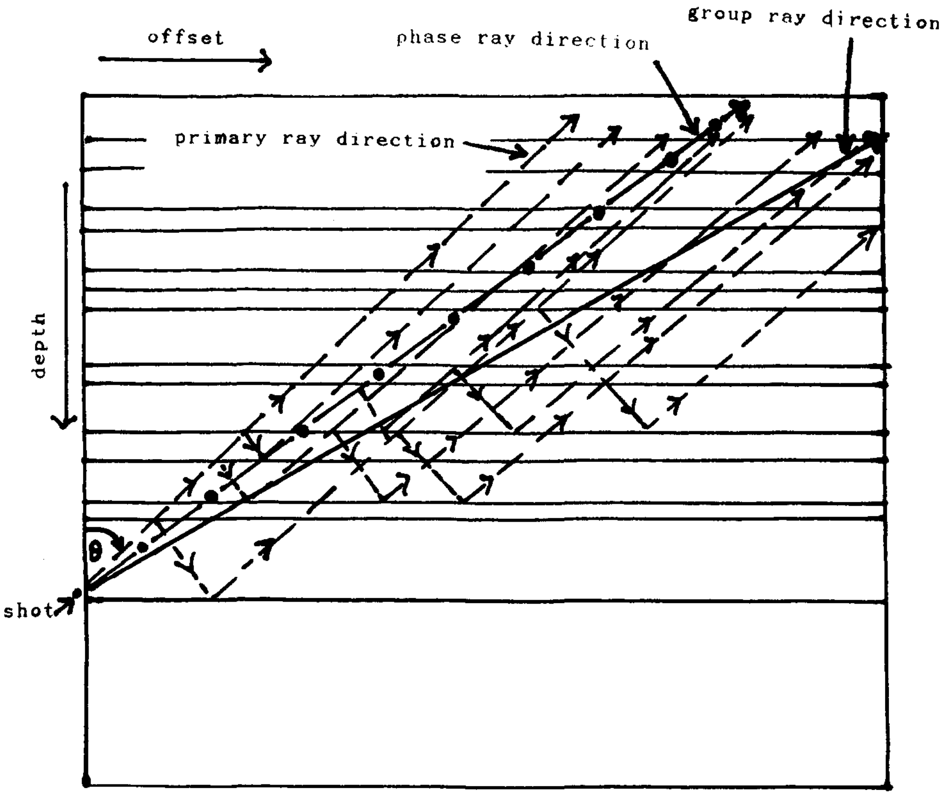

FIG. 2. Sketch of the effect of multiples on the phase and group directions of the mean coherent signal relative to the primary ray direction.

and group direction Φ_g given by

$$\tan \Phi_g = \tan \theta \left[1 + F_\infty - \cot \theta \frac{\partial F_\infty}{\partial \theta} + O(F_\infty^2) \right]. \tag{93b}$$

Note from Eqs. (91a) and (93a) that the group and phase velocities are identical. However, the direction of propagation of the lines of constant phase differs from that of the line of constant amplitude, and further, both of these directions differ in turn from θ, the direction of initial incidence of the plane wave on the layered medium.

For example, in the transitional deposition case, for which, according to Eq. (55), $F_\infty = -(\pi/2)\, r_v^2 \sin^2\theta \sec^4\theta$, we obtain

$$\tan \psi = \tan \theta \left[1 + (\pi/2)\, r_v^2 \sin^2\theta \sec^4\theta + O(F_\infty^2) \right] > \tan\theta \tag{94a}$$

and

$$\tan \Phi_g = \tan \theta \left[1 + (\pi/2)\, r_v^2 \sec^4\theta (2 + \sin^2\theta) + O(F_\infty^2) \right] > \tan\psi > \tan\theta . \tag{94b}$$

In this case we see that for the mean coherent field the direction of the constant phase line is canted more towards the horizontal than the original propagation direction and that the group direction is canted even more towards the horizontal than is the phase direction (see Fig. 2).

The underlying physical reason for this behavior is quite simple. The incident wave makes an angle θ with the bedding planes, the downgoing generated multiples at each bed interface travel either in the θ direction or the $-\theta$ direction (ignoring refraction effects). Those multiples traveling in the $-\theta$ direction rapidly become phase incoherent with respect to the primary wave and so can be ignored safely. Those multiples traveling in the θ direction stay phase coherent with respect to the primary for a longer period of time. Since each generated multiple is laterally offset from the primary, and since each multiple carries some of the energy of the original incident wave, the direction of the coherent component of the seismic wave is basically made up of a superposition of the amplitude and phase of the primary wave plus the amplitudes and phases of the downgoing multiples in the θ direction. Hence the coherent component has its phase and group directions moved slightly laterally (relative to the primary) as it passes each bedding interface by the addition of the multiples' energy and phase. Thus the phase direction and group directions of the coherent component of the seismic signal are canted more towards the horizontal than the original propagation direction as illustrated in the sketch.

In the cyclic depositional case, for which, according to Eq. (62),

$$F_\infty = -r_v^2 \left[\sum_{n=1}^{\infty} f(n) \right] \sec^2\theta \left(\sec^2\theta - \frac{1}{2} \right),$$

we obtain

$$\tan\psi = \tan\theta \left\{ 1 + \frac{1}{2} r_v^2 \sec^2\theta \left(2\sec^2\theta - \frac{1}{2} \right) \times \left(\sum_{n=1}^{\infty} f(n) \right) + O(F_\infty^2) \right\} > \tan\theta \tag{95a}$$

and

$$\tan \Phi_g = \tan \theta \left\{ 1 + \frac{1}{2} r_v^2 \sec^2 \theta (6 \sec^2 \theta - 1) \times \left(\sum_{n=1}^{\infty} f(n) \right) + O(F_\infty^2) \right\} > \tan \psi > \tan \theta, \quad (95b)$$

so that, just as in the transitional deposition case, the direction of constant phase is canted more towards the horizontal than the original propagation direction, and the group direction is canted even more towards the horizontal than the phase direction.

To make a simple numerical estimate of the size of this canting effect, consider a wave initially incident at $\theta = 45°$ to the vertical and with $\Sigma_{n=1}^{\infty} f(n) = 1$, since $f(n)$ is the fractional scattering coefficient in the cyclic case. Then (a) for the transitional deposition pattern, we have

$$\tan \psi = 1 + \pi r_v^2 + O(F_\infty^2),$$
$$\tan \Phi_g = 1 + 10\pi r_v^2 + O(F_\infty^2),$$

so that $\Phi_g = \psi + \frac{9}{2} \pi r_v^2 + O(F_\infty^2)$; and (b) for the cyclic depositional pattern, we have

$$\tan \psi = 1 + 3 r_v^2 + O(F_\infty^2),$$
$$\tan \Phi_g = 1 + 11 r_v^2 + O(F_\infty^2),$$

so that $\Phi_g = \psi + 4 r_v^2 + O(F_\infty^2)$.

For a typical rms reflection coefficient of 0.1 (i.e., $\pi^{1/2} r_v = 0.1$ in the transitional case and $r_v = 0.1$ in the cyclic case) we have $\psi - 45° \approx 0.3°$ and $\Phi_g - 45° \approx 3°$ for the transitional case, and $\psi - 45° \simeq 0.9°$ and $\Phi_g - 45° \approx 3.3°$ for the cyclic case.

Thus for typical seismic parameters the lines of constant phase remain less than about a degree away from the direction of the primary wave, while the group direction is canted of order 3° more to the horizontal than the primary direction. Translated into a lateral displacement equivalent to vertical travel through 10 000 ft of sediment, the ray associated with a 45° primary wave would be at a lateral position of 10 000 ft from the shot whereas the phase line would be a further 100–300 ft laterally from the shot location, and the lateral position of the group of waves would be about 1000 ft further out than the primary ray's 10 000 ft. Upon reflection from a strong reflector at the 10 000 ft level, these values at the surface would double to 200–600 ft for the phase and 2000 ft for the group.

The estimated size of these effects is therefore of importance since velocity and reflector depth analyses on offset seismic signals are often done on either a first-break basis and/or a rms amplitude basis[38]—both of which are influenced by peg-leg multiple effects. While it is not necessarily the case that the group and phase path effects reported here are of paramount importance in all situations, nevertheless the numerical estimates made here suggest that such effects are at least of the order of magnitude where their *a priori* omission from consideration would be unwarranted.

IV. DISCUSSION AND CONCLUSION

In this paper we have presented several major points arising from multiple generation and interference with a primary seismic wave that are outside the scope of the original one-dimensional statistical theory[35–37] devised to account for the effects of multiples on a seismic signal as originally reported by O'Doherty and Anstey[37] extended and corroborated by Schoenburger and Levin[28–30] and revitalized by Bamberger *et al.*,[38] all using numerical codes for generating one-dimensional synthetic seismograms.

A. Summary of main results

The major results obtained can be grouped in four basic classes.

(1) The effects of velocity fluctuations and of density fluctuations enter slightly differently into the dispersion relation describing the propagation characteristics of the coherent component of the seismic wave. The major impact on the role of fluctuations in modifying the transmission response is provided by the angle of propagation of the wave to the bedding planes. Waves that are subcritical propagate with a frequency- and angle-dependent phase and attenuation; as the wave passes through critical and then to supercritical (when the wave does not propagate vertically at all), the attenuation caused by fluctuations increases in magnitude becoming extremely large. Effectively, then, only subcritical waves propagate in the medium, and it is their propagation characteristics that have been investigated extensively in this paper. The remaining major results pertain only to the subcritical waves.

(2) Using the results obtained in Refs. 35 and 36 as a guide, for a single, monochromatic, plane wave propagating initially at an angle to the bedding plane, we found that the transmission response of the mean seismic wave could be written

$$\exp\{(i\omega/V_0)[Z \cos \theta(1+F) + y \sin \theta - V_0 t]\},$$

where the complex factor F, dependent upon the fluctuations in velocity and density, is a function of both frequency and angle.

The magnitude of F is typically a few percent, the real part provides a measure of the frequency-dependent time delay, while the imaginary part measures the effective attenuation of the coherent component of the propagating seismic wave caused by multiple generation and phase incoherence. With the response of a single wave analyzed and understood, our attention then turned to four aspects of the offset-dependent propagation characteristics related to the fact that real seismic waves are not monochromatic nor uniangular. To examine simply the propagation effects caused by angular and frequency spreads of a more representative model of a real seismic signal, we analyzed each factor in isolation.

(A) Effects due to angular spreading: For a pencil of monochromatic waves of wavelength λ, initially spread in a Gaussian manner in amplitude over a half angle $\Delta\theta$ centered on the normal to the bedding plane, we investigated the behavior of the mean seismic pencil as it propagated through the medium. We found the following.

(a) For $Z \lesssim O(\lambda/(\Delta\theta)^2)$, the original pencil was basically unaltered from its original pattern.

(b) For $Z \gtrsim O(\lambda/(\Delta\theta)^2) \equiv Z_f$, the pencil is "splattered" more and more as it moves deeper into the medium with the lateral (y) variation being Gaussian in character for both the phase and the amplitude and dominated by Fresnel diffraction effects for the phase over a scale $y_f \approx (2\lambda Z)^{1/2}$, and dominated by effective attenuation for the amplitude over a scale

$$y_A \approx y_f \left(\frac{\partial^2 \operatorname{Im} F}{\partial \theta^2} \bigg|_{\theta=0} \right)^{-1/2} \gtrsim y_f .$$

(c) For $Z \gtrsim Z_f$, coherence of the mean seismic wave is preserved over a smaller and smaller range of angles around the normal to the bedding plane as propagation into the medium continues. The frequency-dependent scattering angle

$$\phi_{sc} \approx \left(\frac{\lambda}{Z}\right)^{1/2} \left(\frac{\partial^2 \operatorname{Im} F}{\partial \theta^2} \right)^{-1/2}_{\theta=0} \approx \frac{y_A}{Z}$$

decreases to zero as $Z \to \infty$ so that only those waves which make smaller angles than ϕ_{sc} with the normal to the bedding plane contribute to the coherent seismic signal.

(B) Effects due to bandwidth spreading: For a uniangular temporal pulse of waves, initially centered on an angular frequency ω_0 and Gaussianly spread in amplitude over an angular frequency band $\Delta\omega$, we investigated the behavior of the mean seismic pulse as it propagated through the medium. We found the following.

(a) For

$$Z \lesssim O(2\pi(V_0/\omega_0)/(\Delta\omega/\omega_0)^2 |F|) = Z_p ,$$

the original pulse was basically unaltered from its original shape.

(b) for $Z \gtrsim Z_p$, the pulse is distorted from its original shape with the envelope being Gaussianly shaped in the direction of propagation over a scale length of order $(2\lambda_0 Z/|F|)^{1/2}$ and with the phase being Fresnel-diffracted over a scale length of order $(2\lambda_0 Z)^{1/2}$, where $\lambda_0 = 2\pi V_0/\omega_0$.

(C) Effects contained in lateral correlations: Since measurements of seismic signals are made at many geophones which are laterally offset on the surface at different distances from a shot, we investigated the behavior of lateral correlated information from a mean seismic wave for all pairs of receivers at fixed lateral separation, ξ. The results of that investigation depend upon the size of the fractional bandwidth $\Delta\omega/\omega$ relative to $\operatorname{Im} F$.

For narrow bandwidths, $\Delta\omega/\omega \ll \operatorname{Im} F$, we found that the lateral autocorrelation intensity declines in a Gaussian manner for depths in excess of $Z_c = (\lambda/\operatorname{Im} F)$; the normalized visibility function then declining Gaussianly over a lateral scale $\xi_c \approx \lambda(8Z/Z_c)^{1/2}$, and the angular distribution over which waves contribute to the coherent component of the seismic wave also declines in a Gaussian manner over a scale angle of about $\phi_c \approx (Z_c/Z)^{1/2}$.

For broad bandwidths, $\Delta\omega/\omega \gtrsim \operatorname{Im} F$, we found that the character of the lateral correlation shifted so that Fresnel diffraction effects play a more dominant role with a correlated loss of phase over a lateral scale length $\xi_{ph} \approx (Z\Delta\omega/\lambda\omega)^{1/2}$ but with a very small attenuative loss.

(3) The third class comprises the sedimentation pattern effects. In order to illustrate the ranges of both the real and imaginary parts of the transmission response in respect of their frequency and angle dependences, we looked at two extreme situations of possible reflectivity spectra which O'Doherty and Anstey[37] have labeled as characterizing transitional sedimentation patterns and cyclic depositional patterns, respectively. We modeled the transitional pattern quantitatively using an Ornstein–Uhlenbeck power spectrum,[44] as has been done elsewhere,[35,36,45–47] while we modeled the power spectrum of the cyclic sedimentation pattern[37] as a series of equally spaced δ-function spikes.[38]

For the transitional pattern we found that for plane waves propagating at an angle θ to the bedding planes the real part of F, which corresponds to a frequency-dependent time delay, was nearly constant for low angular frequencies $\omega \lesssim \omega_*/(2\cos\theta) \equiv \Omega$, where $\omega_* (\equiv V_0 K_0)$ measure the critical scaling frequency below which the reflectivity power spectrum is effectively quadratically dependent on the frequency and above which the spectrum is nearly constant. In the same frequency regime the fractional attenuation is linearly dependent on the frequency with a peak around $\omega = \Omega$ before declining at high frequencies as also does the time delay.

For the cyclic pattern we found that at almost all frequencies the attenuative effect was negligible—the dominant behavior being a time delay that was nearly constant. The exceptions occurred at multiples of $2\omega\cos\theta = n\,\omega_*$ ($\omega_* = V_0 K_0$), where high attenuation was found in very tight bandwidths with the bandwidth over which attenuation was dominant diminishing with increasing frequency. In short, except for periodic, narrow bands of frequency, the general behavior was that of a pure time delay, in line with the numerical results of Bamberger *et al.*[38]

(4) The fourth class is phase and group directions. The behavior of a pencil of waves in respect of the speeds of propagation and of the overall direction of propagation of the energy in the pencil and of the phase is important in seeing how multiples influence the direction of propagation of a primary beam of waves. To examine this aspect of the problem, we looked at phase and group velocity behavior of a pencil of waves at high frequency where the geometrical optics approximation is expected to be the most accurate. We found that the group and phase speeds were identical to quadratic order in F. However, the directions of propagation differed between the group, the phase and the incident primary wave directions. For both the transitional and the cyclic patterns of sedimentation, the group direction was canted more towards the horizontal than the phase direction, which, in turn, was more highly canted towards the horizontal than the original wave's direction, and both of these effects are due to the generated multiples interfering with the primary wave. Whether the ordering, group angle > phase angle > incident angle, holds for all reflectivity power spectra is unknown, but the fact that this ordering obtains for two such widely disparate power spectra as the representations of transitional and cyclic patterns strongly suggests that it

would indeed be a peculiar reflectivity spectrum which upset the ordering. The magnitude of the effects, particularly for the group angle, is about 5%–10% for an incident wave making a 45° initial angle, which translates into a lateral distance of 500–1000 ft at a depth of 10 000 ft. Since the group and phase velocities differ by only about a percent or so from the mean velocity V_0, the dominant effect is caused by the change in direction of the group relative to the incident wave. Thus, interpretationally, the group takes a longer time to reach a given depth than would the primary wave, with the relative time scales being in the ratio $\sec \Phi_g / \sec \theta$, which is about 5%–10% for an initial wave incident at 45°. Inverting the argument: from a given travel time and a known velocity, an event, determined from a group wave measurement, will be of order 5%–10% *deeper* vertically than it really is, and it will also be located at the wrong lateral position by about 5%–10%.

As the frequency is lowered, both frequency-dependent attenuation and phase dispersion take on larger and larger roles with their precise behaviors with frequency being dependent upon the details of the reflectivity power spectrum.

B. Discussion

The detailed results presented in this paper and the more general arguments and concepts advanced suggest several major points for further consideration.

We have seen that the finite bandwidth and finite angular extent of an incident seismic wave train are rapidly and significantly distorted by the phase and attenuative effects of generated multiples as the wave propagates into the bedded medium. Indeed, beyond a critical depth of penetration (roughly estimated at a few thousand feet), the initial half angle and bandwidth are overshadowed by the effect of multiples. The incident beam's lateral spreading and pulse shaping (amplitude and phase) are then controlled completely by subsurface propagation and are no longer beholden to the initial beam angle spread or bandwidth. We have also seen that the coherent component of the seismic wave is progressively attenuated and time delayed as it propagates and that the direction of propagation of the energy is not the same as the initial propagation direction.

Rough numerical estimates of all of these effects, using typical seismic parameters, show that they sit squarely in the range of values where they can be expected to have a profound influence on seismic signals. They may *not* be declared unimportant with impunity.

The effects discussed raise serious questions that need to be addressed.

(1) In efforts to obtain seismic attenuation of subsurface rocks that can be related to oil and/or gas accumulations, the effective attenuation effect caused by multiples reported here can play a dominant masking role because the magnitude of the imaginary part of F is typically 1%–5% corresponding to a quality factor $Q[\equiv \frac{1}{2}(\text{Im } F)^{-1}]$ of order 10–100—right in the middle of the range normally associated with subsurface formations.[28–30,45–47] How does one separate an intrinsic Q from one caused by multiples?

(2) The low-frequency time delay caused by multiples (due to the real part of F) is estimated to be typically of order 1%–10% so that the travel time one would ascribe to a seismic signal would place an event 1%–10% deeper than it is in actuality. How does one accommodate for this effect?

(3) The 5%–10% change in direction of a wave group means that both the lateral and depth locations of an event will be in error by 5%–10%. How does one separate out these effects so as to avoid misrepresentation?

(4) The progressive phase and amplitude distortion caused by the multiples eventually distorts any original pulsed pencil of seismic waves into a form which causes both lateral spreading and time dispersion. This means that both spatial and temporal resolution at depth degrade as the seismic signal progresses. When do these effects become so dominant that it is pointless to even attempt to construct a depth-migrated section?

(5) The analysis carried through here has been performed for the simple situation of plane parallel beds with a constant, average velocity. But real subsurface formations are not laterally constant, and further, the subsurface velocity normally shows a systematic increase with depth. How significant are the nonlinear distortions produced by such effects when added to the already distorting influence of the basic effects addressed here?

(6) The work reported here treats only with scalar waves, but the conversion of P waves to S waves and vice versa at bedding interfaces is a well-documented fact. How are the results presented here modified by such mode conversion? When does mode conversion dominate over multiples as a primary cause of information loss of an incident P wave? Can the two effects be separated?

We provide no answers to these questions for we do not have any. What we do suggest is that (a) in view of the relevance to seismic signals of the previous[35,36] work on the O'Doherty–Anstey effect,[37] and (b) in light of the theoretical and analytic work reported in this paper, which appears potentially equal, if not greater, in importance than (a), some effort be expended to quantify using a full three-dimensional synthetic seismogram routine, particular regimes of dominance, and the general level of global pervasiveness, of the effects uncovered here. It is our contention that the analytic results, and their rough numerical estimates, are sufficient to indicate that these multiple effects should be a major area of concern for some considerable time to come.

[1] A. Chang and D. H. von Seggern, J. Geophys. Res. **85**, 4811 (1980).
[2] K. Aki, J. Geophys. Res. **78**, 1334 (1973).
[3] J. Capon, Bull. Seismol. Soc. Am. **64**, 235 (1974).
[4] K. A. Berteussen, A. Christofferson, E. S. Husebye, and A. Dahle, Geophys. J. Roy. Astron. Soc. **42**, 403 (1975).
[5] L. A. Chernov, *Wave Propagation in a Random Medium* (McGraw–Hill, New York, 1960).
[6] C. A. Powell and A. S. Meltzer, Geophys. Res. Lett. **11**, 481 (1984).
[7] A. M. Dainty, Geophys. Res. Lett. **8**, 1126 (1981).
[8] A. M. Dainty, J. Geophys. Res. **89**, 3172 (1981).
[9] M. Kikuchi, Phys. Earth Planet. Inter. **25**, 159 (1981).
[10] M. Kikuchi, Phys. Earth Planet. Inter. **27**, 100 (1981).
[11] H. Sato, J. Geophys. Res. **87**, 7779 (1982).
[12] H. Sato, J. Acoust. Soc. Am. **71**, 559 (1982).
[13] K. Aki, A. Christofferson, and E. S. Husebye, Bull. Seismol. Soc. Am. **66**, 501 (1976).
[14] K. Aki, Phys. Earth Planet. Inter. **26**, 241 (1981).
[15] R-S. Wu, Geophys. Res. Lett. **9**, 9 (1982).
[16] R-S. Wu, Wave Motion **4**, 305 (1982).

[17]V. I. Tatarski, *Wave Propagation in a Turbulent Medium* (Dover, New York, 1961).
[18]B. J. Uscinski, *The Elements of Wave Propagation in a Random Medium* (McGraw-Hill, New York, 1977).
[19]W. Menke, Bull. Seismol. Soc. Am. **74**, 1079 (1984).
[20]J. A. Hudson and J. R. Heritage, Geophys. J. R. Astron. Soc. **66**, 221 (1981).
[21]K. L. McLaughlin, Ph.D. Thesis, Univ. Calif., Berkeley, 1983.
[22]K. Aki, Bull. Seismol. Soc. Am. **78**, S319 (1982).
[23]J. A. Hudson, Geophys. J. R. Astron. Soc. **69**, 649 (1982).
[24]A. Ishimaru, *Wave Propagation and Scattering in Random Media* (Academic, New York, 1978), Vols. VI and II.
[25]R. Dashen, J. Math. Phys. **20**, 894 (1979).
[26]J. J. McCoy, Phys. Earth Planet. Inter. **21**, 126 (1980).
[27]K. R. Kelly, R. W. Ward, S. Treitel, and R. M. Alford, Geophys. **41**, 2 (1976).
[28]M. Schoenberger and F. K. Levin, Geophys. **39**, 270 (1974).
[29]M. Schoenberger and F. K. Levin, Geophys. **43**, 730 (1978).
[30]M. Schoenberger and F. K. Levin, Geophys. **44**, 1813 (1979).
[31]P. G. Richards and W. Menke, Bull. Seismol. Soc. Am. **73**, 1005 (1983).
[32]W. Menke, Geophys. Res. Lett. **10**, 1145 (1983).
[33]J. R. Resnick, I. Lerche, and R. T. Shuey, Geophys. J. R. Astron. Soc. (accepted for publication, 1986).
[34]W. Menke, Geophys. J. R. Astron. Soc. **75**, 541 (1983).
[35]N. C. Banik, I. Lerche, and R. T. Shuey, Geophys. **50**, 2768 (1985).
[36]N. C. Banik, I. Lerche, J. R. Resnick, and R. T. Shuey, Geophys. **50**, 2775 (1985).
[37]R. F. O'Doherty and N. A. Anstey, Geophys. Prospect. **19**, 430 (1971).
[38]A. Bamberger, G. Chavent, and P. Lailly, Geophys. **47**, 757 (1982).
[39]F. C. Karal and J. B. Keller, J. Math. Phys. **5**, 537 (1964).
[40]U. Frisch, "Wave propagation in random media," in *Probabilistic Methods in Applied Mathematics*, edited by A. T. Bharucha-Reid (Prentice-Hall, New York, 1968), Vol. I.
[41]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).
[42]H. Cramer, Ann. Math. **41**, 215 (1940).
[43]M. Born and E. Wolf, *Principles of Optics* (Pergamon, New York, 1964).
[44]G. E. Uhlenbeck and L. S. Ornstein, Phys. Rev. **36**, 823 (1930).
[45]R. Godfrey, F. Muir, and F. Rocca, Geophys. **45**, 1351 (1980).
[46]T. W. Spencer, C. M. Edwards, and J. R. Sonnad, Geophys. **47**, 939 (1977).
[47]T. W. Spencer, J. R. Sonnad, and T. M. Butler, Geophys. **47**, 16 (1982).
[48]V. I. Ginzburg, *Propagating of Electromagnetic Waves in Plasma* (Pergamon, New York, 1964).

Is there action-at-a-distance linear confinement?

J. Weiss
Department of Physics, Slovak Technical University, 812 31 Bratislava, Czechoslovakia

(Received 20 March 1985; accepted for publication 28 October 1985)

The possilibity of constructing an action-at-a-distance form of linear confinement is demonstrated. Using the Fokker–Wheeler–Feynman action principle, known from classical action-at-a-distance electrodynamics, with an action containing the relativistically invariant two-particle Heaviside step function, equations of motion and appropriate potentials exhibiting the linearity of their behavior are derived. The plausibility of the generators of motion describing dynamics with the linear potentials is verified on the simple circular-orbit model of a two-component system, and the expected energy spectrum in terms of semiclassical quantization is obtained.

I. INTRODUCTION

Since the appearance of Wheeler–Feynman action-at-a-distance (AAD) electrodynamics[1] many attempts have been made to construct relativistic particle theories in terms of the prototype of this quite consistent theory ignoring field–mediated interaction (a great deal of discussions on this subject may be found in Refs. 2 and 3). Some of them appeared to be equivalent to the corresponding field theories,[1,4] others do not have a field-theoretic analog.[5] However, common features of these theories, important from the point of view of the structure of space-time theories, are that they are derivable from *a Fokker type of action*[6] and are mostly invariant under *the full conformal group* of transformations.[7]

As regards specifically physical ideas, one meets also with relevant properties observed on direct interparticle theories of the Wheeler–Feynman type. Their actions *are not instantaneous* and they alone seem to violate *causality* grasped in the classical sense.[8] By noninstantaneous action is meant that action between particles should propagate with the speed of light and thus it is not of the Newton–Coulomb character. This property is usually expressed by Green's functions, which give nonvanishing contributions on and inside the light cones. Noncausality of an AAD theory demonstrates the idea that motion of a particle is determined by the past and future behavior of the other particles. The fact that both the future-directed and past-directed signals exist on equal footing is an obvious violation of causality in the conventional sense of classical theories. Conventional causality is, however, secured by accepting the condition on the existence of a perfect absorber.[1,8] In cosmology it was found that this condition could be satisfied in the steady-state universe.[9]

The scale of physical applicability of relativistic AAD theories is relatively wide today. It goes from electrodynamics,[1–3] mesodynamics,[4–10] through gravity,[8,11] to Dirac's monopoles.[12] The extension of this set to quarks, even if in a restricted form, could be encouraging, of course.

In this paper we shall try to show that there occurs the real possibility to produce a class of linear potentials on the basis of an AAD dynamics. We have here in mind dynamics *à la* Fokker–Wheeler–Feynman electrodynamics,[1] based on the direct interparticle interaction via time-symmetric fields (half retarded plus half advanced), which exhibits properties of AAD theories mentioned above. Standard Wheeler–Feynman electrodynamics[1] starts with the principle of stationary action and postulates a Fokker-type action with the interaction term determined by the Dirac delta function, the argument of which is the square of the distance between a source and an observation point. The integral of this function taken along the source trajectory generates the Liénard–Wiechert potentials, fundamental potentials of classical relativistic electrodynamics. It is an interesting coincidence that in the dynamics that we are going to propose, it is the Heaviside step function θ of the above argument that plays the same role in constructing linear potentials as the Dirac δ function in the reproduction of electrodynamic potentials. In other words, *the generalized relativistically invariant and genetically associated δ and θ functions are*, in fact, *sources of the classical relativistic electromagnetic and the linear potentials*, respectively.

One of the main goals of this work is to demonstrate how to construct the relativistic linear potentials. The second aim is to give a convenient example of possible applicability of the new type of potentials concretely in the study of simple composite systems. In the following paper[13] we shall derive the set of conformal conservation laws by accepting the requirement of invariance of the AAD linear potentials under the full conformal group of transformations.

We propose first (Sec. II) the derivation of equations of motion by the use of the Fokker–Wheeler–Feynman action principle[1,6] with an action involving the θ function in its interaction term. Due to this θ interaction choice, potentials generated in the variation procedure become functions only of the difference between retarded and advanced positions of pointlike particles considered. As a consequence, in a system of pairwise interacting particles there appears, quite naturally, *linear confinement* motivated just by their mutual retarded and advanced adjunct fields.

The faithfulness of our potentials and the equations of motion are verified in Sec. III on a familiar two-component composite system of the category of circular-orbit models (Schild's problem[14]), consisting of a particle and an antiparticle revolving about a common origin. We give generators of Poincaré's group representing generally our dynamics, and, in application on circular orbits prescribed to the both con-

 0022-2488/86/041015-08$02.50

stituents, we find that they have all expected properties of this type of the bound system: energy levels evaluated in terms of the Bohr–Sommerfeld quantization rules are *very similar* to those of a harmonic-oscillator potential for lowest values n at least. This model may remind us, in a most strongly reduced form, of a classical quarkonium in which a quark and an antiquark act on each other exclusively via the linear potentials.

In this paper our considerations will be concerned only with two-particle dynamics. For this reason we shall accept throughout the practical notation of the famous Wheeler–Feynman work,[1] where the four-vectors of the coordinates of particles a and b are termed a^μ and b^μ, the difference $a^\mu - b^\mu \equiv ab^\mu$.

II. ACTION PRINCIPLE AND THE LINEAR POTENTIALS

We begin with a natural inquiry as to what description of interactions and motions for pointlike particles is possible that demands no direct use of the notion of field, which is at the same time well defined, economical in postulates, and reproduces a constant force acting among particles. The description preserving the mentioned properties is possible within the framework of the Fokker–Wheeler–Feynman action principle.[1,6]

A. Equations of motion

Let us consider a system consisting of two particles a and b, which are subjected to a direct at-a-distance interaction. Equations of motion of the system in which a given pointlike particle interacts only with the other particle (but not with itself) via a constant force may be obtained from the variational principle

$$\delta S = 0, \tag{2.1}$$

where the action S is defined by the expression

$$\begin{aligned} S = & - m_a \int_{\sigma_1^a}^{\sigma_2^a} (\eta_{\mu\nu} \dot a^\mu \dot a^\nu)^{1/2}\, d\lambda_a \\ & - m_b \int_{\sigma_1^b}^{\sigma_2^b} (\eta_{\mu\nu} \dot b^\mu \dot b^\nu)^{1/2}\, d\lambda_b \\ & - g_a g_b \int \int_{(\sigma_1)}^{(\sigma_2)} \theta(ab^\rho ab_\rho) \eta_{\mu\nu}\, da^\mu\, db^\nu. \end{aligned} \tag{2.2}$$

Here m_a and m_b are the masses, g_a and g_b the coupling constants of the objects a and b, respectively, λ_a and λ_b are the path parameters of their world-lines, and θ is the step function that satisfies the condition

$$\theta(z) = \begin{cases} 1, & \text{if } z \geqslant 0, \\ 0, & \text{if } z < 0. \end{cases} \tag{2.3}$$

We use the metric $g_{00} = -g_{11} = -g_{22} = -g_{33} = 1$ and let $\hbar = c = 1$. The dot on the quantities of Eq. (2.2) denotes differentiation with respect to the corresponding path parameter. After the variation one can let this parameter equal the proper time of a particle and calculate with $\dot a^\mu \dot a_\mu = \dot b^\mu \dot b_\mu = 1$. The integrals in the first two terms of Eq. (2.2), representing the action of the free particles, have to be evaluated over those parts of the particle trajectories that lie between two nonintersecting spacelike surfaces σ_1 and σ_2. The limits of the double integrals (σ_1) and (σ_2) in the third, interaction term of Eq. (2.2) must be taken far enough beyond σ_1 and σ_2 to include all effects of any one particle on the parts of the trajectories of the other particle lying between σ_1 and σ_2. The sign of the interaction term is chosen such that the force between particles and antiparticles is of an attractive character.

From Eqs. (2.1) and (2.2) we obtain by variation of the world-line of the particle a with end points held fixed on it the following equations of motion (see Appendix A):

$$m_a \ddot a^\mu = g_a F^{(b)\mu}{}_\nu(a) \dot a^\nu, \tag{2.4}$$

where

$$F^{(b)}_{\mu\nu}(x) = A^{(b)}_{\nu,\mu}(x) - A^{(b)}_{\mu,\nu}(x) \tag{2.5}$$

is the tensor of the adjunct field produced by the particle b at the point x and

$$A^{(b)\mu}(x) = g_b \int_{-\infty}^{+\infty} \theta(xb^\rho xb_\rho)\, db^\mu(\beta) \tag{2.6}$$

is *the linear four-potential* corresponding to this field. The analogs of equations to Eqs. (2.4)–(2.6) can be derived also for the b particle by variation of its world-line, accepting the condition $\delta b(+\infty) = \delta b(-\infty) = 0$. According to Eq. (2.6) the expression $A^{(b)}_\mu(a)$ represents a contribution to the potential of the considered system at the point a of the a particle from the all trajectory of the b particle.

B. AAD linear potential

From the form of Eq. (2.6) it cannot be seen, of course, that one is really dealing with a linear potential. Let us arrange, therefore, the integrand in Eq. (2.6) into a more convenient form. Since $db^\mu(\beta) = \dot b^\mu(\beta) d\beta$, by performing an integration by parts, we obtain from Eq. (2.6) (after subtracting a constant)

$$A^{(b)\mu}(a) = 2g_b \int_{-\infty}^{+\infty} \theta'(ab^\rho ab_\rho)\, ab^\nu(\beta) \dot b_\nu(\beta) b^\mu(\beta)\, d\beta, \tag{2.7}$$

where θ' is the derivative θ with respect to the argument given in the square brackets, i.e., Dirac's δ function of this argument. We see thus that the integral in Eq. (2.7) leads to two integrals with the δ function

$$\int_{-\infty}^{+\infty} (2r_{ab})^{-1} [\delta(r_{ab} - ab^0) + \delta(r_{ab} + ab^0)] b^\mu\, d\beta, \tag{2.8}$$

where we have simply denoted $|\mathbf{a} - \mathbf{b}| = r_{ab}$ and $a^0 - b^0 = ab^0$. Introduce now the new variable

$$s_\pm = r_{ab} \mp ab^0. \tag{2.9}$$

For ds we have

$$ds_\pm = \pm (r_{ab})^{-1} ab^\nu \dot b_\nu\, d\beta. \tag{2.10}$$

Inserting Eqs. (2.9) and (2.10) into Eqs. (2.7) and (2.8) and performing the integration, one finds

$$A^{(b)\mu}(x) = g_b (b^\mu(\beta)_{s_+ = 0} - b^\mu(\beta)_{s_- = 0}). \tag{2.11}$$

This simple formula for the potential, expressing action of the b particle (more correctly, a contribution from its trajectory) on the a particle at the point x, is clearly the sum of the retarded (+) and advanced (−) terms

$$A^{(b)\mu\pm}(x) = \pm g_b b^\mu(\beta)_\pm , \tag{2.12}$$

the dependence of x being implicitly hidden in $s_\pm$. The indices $\pm$ on the brackets in (2.12) mean that b^μ is to be evaluated for values of the parameter β for which $s_\pm = 0$. Accordingly the contribution of the b particle to the potential at the point x is directly proportionate to the difference of its retarded and advanced positions with respect to this point. The mentioned difference will thus increase *linearly* with the increasing distance between the particles. The potential given by Eq. (2.11) defines thus *the classical relativistic AAD linear potential.* The formula (2.11) represents then *a natural manner* of the formulation of *the classical AAD linear confinement.*

In this point, it can be interesting to remember the genesis of the both related potentials: the Liénard–Wiechert and the just derived ones. The result (2.12), evaluated from Eq. (2.6), is very similar to the formula of the four-potential in classical electrodynamics[1]

$$A^\mu(x_i) = e_k \int_{-\infty}^{+\infty} \delta(x_i x_k^\nu x_i x_{k\nu}) \dot{x}_k^\mu \, d\tau_k, \tag{2.13}$$

determining accurately the Liénard-Wiechert potentials. The difference consists "only" in that the "*electromagnetic*" generalized $\delta(ab^\nu ab_\nu)$ *function* is now replaced by the "*linear*" generalized $\theta(ab^\nu ab_\nu)$ *function.* The electrodynamic and linear potentials behave inversely towards each other in asymptotic regions: the potentials of the type (2.13) decrease to 0 (increase to ∞) for $r \to \infty$ ($r \to 0$), while those of the type (2.6) increase to ∞ (decrease to 0) for $r \to \infty$ ($r \to 0$). A remarkable feature of our potentials is also that they are the velocity-independent quantities that depend upon x_i solely implicitly through the dependence of x_k on the retarded and advanced values of the source path parameter, which depend on x_i, while the Liénard–Wiechart potentials depend on the position x_i both explicitly and implicitly (now through x_k and $\dot{x}_k$).[1]

C. Constraint expression for the AAD linear potential

The potentials given by (2.6) or (2.11) will be also discussed in another of our works,[13,15] however in the context with their constraints expressed by the form $s_\pm = 0$ it is needed to add a useful note.

The spatial component of the potential (2.6), written briefly as

$$\mathbf{A}^{(b)}(\mathbf{a}) = g_b(\mathbf{b}^+ - \mathbf{b}^-), \tag{2.14}$$

reads that the contribution of a source (the b particle) to the potential at the point a, the position of the particle a, is proportional to the difference of its retarded b^+ and advanced b^- positions relatively to $\mathbf{a}$. It is convenient to introduce two vectors $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ (see Fig. 1),

$$\boldsymbol{\xi}_a = \mathbf{b}^+ - \mathbf{a}, \quad \boldsymbol{\eta}_a = \mathbf{b}^- - \mathbf{a}, \tag{2.15}$$

by the help of which we are able to express $\mathbf{A}$:

$$\mathbf{A}^{(b)}(\mathbf{a}) = g_b(\boldsymbol{\xi}_a - \boldsymbol{\eta}_a). \tag{2.16}$$

According to Eq. (2.11) the zero component of A^μ is

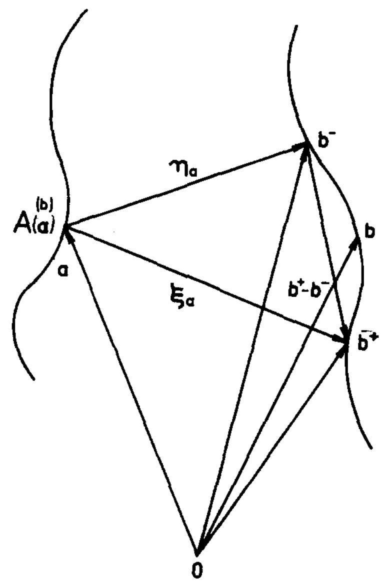

FIG. 1. Illustration of quantities defining the AAD linear potential of a two-body system at the point a on the world line of the particle a.

$$A^{(b)0}(a) = g_b(b^{0+} - b^{0-}). \tag{2.17}$$

The quantities $b^{0\pm}$ are constrained by the light cone condition

$$(b^{0\pm} - a^0)^2 - (\mathbf{b}^\pm - \mathbf{a})^2 = 0. \tag{2.18}$$

Let $\boldsymbol{\xi}_a$ and $\boldsymbol{\eta}_a$ be supplemented on the four-vectors ξ_a^μ and η_a^μ as

$$\xi_a^\mu = b^{\mu+} - a^\mu, \quad \eta_a^\mu = b^{\mu-} - a^\mu, \tag{2.19}$$

for the temporal components of which it holds:

$$\xi_a^0 = b^{0+} - a^0 = |\boldsymbol{\xi}_a|, \quad \eta_a^0 = b^{0-} - a^0 = -|\boldsymbol{\eta}_a|. \tag{2.20}$$

We see thus that the AAD linear potentials can be expressed explicitly in terms of ξ_a^μ and η_a^μ in the following way:

$$A^\mu = g(\xi^\mu - \eta^\mu). \tag{2.21}$$

The new variables must, however, consistently with the requirement (2.18), *obey the constraints*

$$\xi^2 = \eta^2 = 0. \tag{2.22}$$

It is interesting to examine relations between dynamics built up on the potentials (2.21) with the constraints (2.22) and Dirac's constraint dynamics of a pointlike object in which internal dynamic variables are constrained by the condition of the form (2.22).[15] They reveal intrinsic continuity of our dynamics with the theory of rotator models.[16]

D. Tensor $F_{\mu\nu}$ of the adjunct field

Finally, let us determine explicitly the tensor $F_{\mu\nu}$ given by (2.5). To do this, we need know the derivatives of β with respect to x. They can be easily computed from the light cone condition

$$\eta_{\mu\nu} x b^\mu(\beta) x b^\nu(\beta) = 0. \tag{2.23}$$

From here one has

$$\beta_{,\mu} = r_\mu \rho^{-1}. \tag{2.24}$$

where $r_\mu = \eta_{\mu\nu} xb^\nu$ and $\rho = r_\mu \dot{b}^\mu$. Taking into account Eqs. (2.5), (2.12) and (2.24) and accepting the denotation

$$r_{[\mu}\dot{b}_{\nu]} = \tfrac{1}{2}(r_\mu \dot{b}_\nu - r_\nu \dot{b}_\mu), \tag{2.25}$$

we come to the result

$$F^{\pm}_{\mu\nu}(x) = \pm g_b(\rho^{-1} r_{[\mu}\dot{b}_{\nu]})_{\pm}. \tag{2.26}$$

Naturally, the tensor $F_{\mu\nu}$ is considerably simpler than the same tensor electrodynamics[8] and similarly this is case with the equations of motion. In the next section $F_{\mu\nu}$ will be used to determine the equations of motion of a two-particle system carrying out a finite motion with prescribed trajectories of its constituents. Before performing these calculations we notice that the general two-constituent equations of motion have a specific property, as a preliminary analysis shows: a system of *delay* differential equations appears in the three-dimensional case, while in electrodynamics it appears already in two dimensions.[17] Moreover, this analysis indicates[15] that even the two-dimensional equations result in the presence of a certain *oscillatory motion* in a bound system, which is precisely of the same category that characterizes a relativistic rotator.[16]

III. ENERGY SPECTRUM OF THE CIRCULAR-ORBIT MODEL

Now we are trying to illuminate the physical content of the potential defined by Eq. (2.6) or (2.11) and the equations of motion given by Eqs. (2.4) and (2.26) on a simple model of the bound state of a particle–antiparticle pair. The model could be considered to be one of the simplest possible classical relativistic models of a quarkonium with no more than one of the basic "strong charges" of its constituents in a situation when the electromagnetic interaction is negligible and only the linear potentials are present. Within the framework of this most trivial approach to quarks let us put to the test bound states of such a system and compute its energy spectrum. Since one deals here with the classical theory, energy states can be evaluated by way of the quantum rules of early quantum mechanics. For this purpose we give—first in a general form—generators of the Poincaré group that describe our dynamics.

A. Generators of Poincaré's group

We could proceed like Wheeler and Feynman in their classical work[1] and derive these generators straightforwardly from the equations of motion (2.4), computing infinitesimal changes in the corresponding quantities. This method appears to be very effective and in Ref. 13 (see the following paper) it will be used to construct conformal conservation laws motivated by the AAD potentials (2.6).

However, there is another way, too. If we explore the postulated expression (2.2) for the action functional in detail, we find that our ansatz on S fulfills requirements posed on the relativistic action. The S in the formula (2.2) fulfills the requirement of the invariance under *translations*, since the first terms in (2.2) are not dependent on a^μ and b^μ and the third term depends only on the difference ab^μ. It fulfills also the requirement of *the Lorentz invariance*, because θ is a function of the invariant $(ab^\mu)^2$. Our action, therefore, should be invariant under the group of *Poincaré mappings*. From the requirement of the invariance of S under the Poincaré group, taking infinitesimal mappings to be $x^\mu \to x'^\mu = x^\mu + \epsilon^\mu_{\ \nu} x^\nu + \epsilon^\mu$ $(x = a, b)$, we obtain the generators of our system—the four-momentum P^μ and the tensor of the angular momentum $L^{\mu\nu}$. Here the requirement that $\delta\overline{S} = S' - S = 0$ leads to expressions involving variations $\delta\overline{x}^\mu = \epsilon^\mu_{\ \nu} x^\nu + \epsilon^\mu$ and differentiations with respect to the path parameters, which can be taken to be the proper times of the particles. Performing in these expressions offered integrations by parts and using the equations of motion (2.4)–(2.6), we derive the conserved quantities corresponding to the translations and "rotations," respectively [compare with Eqs. (1.11) and (1.14) of Ref. 13]:

$$\begin{aligned} P^\mu(\alpha_0,\beta_0) &= m_a \dot{a}^\mu(\alpha_0) + g_a A^{(b)\mu}(\alpha_0) \\ &\quad + m_b \dot{b}^\mu(\beta_0) + g_b A^{(a)\mu}(\beta_0) \\ &\quad + k\left(\int_{\alpha_0}^{+\infty}\int_{-\infty}^{\beta_0} - \int_{-\infty}^{\alpha_0}\int_{\beta_0}^{+\infty}\right) \\ &\quad \times \delta(ab^\rho ab_\rho) ab^\mu \, da^\nu \, db_\nu, \end{aligned} \tag{3.1}$$

$$\begin{aligned} L^{\mu\nu}&(\alpha_0,\beta_0) \\ &= a^\mu(\alpha_0)(m_a \dot{a}^\nu(\alpha_0) + g_a A^{(b)\nu}(\alpha_0)) \\ &\quad - a^\nu(\alpha_0)(m_a \dot{a}^\mu(\alpha_0) + g_a A^{(b)\mu}(\alpha_0)) \\ &\quad + b^\mu(\beta_0)(m_b \dot{b}^\nu(\beta_0) + g_b A^{(a)\nu}(\beta_0)) \\ &\quad - b^\nu(\beta_0)(m_b \dot{b}^\mu(\beta_0) + g_b A^{(a)\mu}(\beta_0)) \\ &\quad + k\left(\int_{\alpha_0}^{+\infty} d\alpha \int_{-\infty}^{\beta_0} d\beta - \int_{-\infty}^{\alpha_0} d\alpha \int_{\beta_0}^{+\infty} d\beta\right) \\ &\quad \times \{\delta(ab^\rho ab_\rho)(a^\nu b^\mu - a^\mu b^\nu)\dot{a}^\sigma \dot{b}_\sigma \\ &\quad - \tfrac{1}{2}\theta(ab^\rho ab_\rho)(\dot{a}^\nu \dot{b}^\mu - \dot{a}^\mu \dot{b}^\nu)\}, \end{aligned} \tag{3.2}$$

where $k = 2g_a g_b$ and the Wheeler–Feynman denotation $a^\mu - b^\mu = ab^\mu$ was accepted. It is not difficult to show that $[P^\mu(\alpha,\beta)]_{\alpha=\alpha_0,\,\beta=\beta_0}$ is a constant with respect to a given parameter. It is simply sufficient to differentiate Eq. (3.1) with respect to α or β and use Eqs. (2.4)–(2.6). One can easily show that the double integral in Eq. (3.1) has for the component P^0 the meaning of the interaction energy, if one takes the simple case of the stationary particle interaction. In this case we arrive at the familiar result $P^0 = m_a + m_b + kr_{ab}$, where r_{ab} is the distance between two static particles. Thus, in *the static limit*, P^0, given by (3.1), leads directly to the expression defining *the static interaction energy* of two quarks. The appropriate calculation confirms two important facts: the double integrals in (3.1) produce terms compensating the overcount of the potential energies in P^0 and, in addition, one always has to do with the two world-line segments (due to δ) that contribute to the determination of $P^0(\alpha_0,\beta_0)$.

B. Equations of motion of the circular orbits

Now we are going to use, in a manner equivalent to Schild[14] in electrodynamics, the equations of motion (2.4) and the generators (3.1) and (3.2) to compute energy states for the circular-orbit model of a two-particle bound system. We note, in advance, that from the mentioned equations it is evident that even if confinement of our objects a and b is of the linear character, in general the appropriate equations of motion are of the integrodifferential type. Only the two-dimensional case leads, as emphasized in Sec. I, to the differential equations that can be solved at the point initial value conditions.[15] In the present three-dimensional case we are able to find the rigorous solutions, as will be shown in a little while, only because we prescribe the concentric circular orbits, and thus total symmetry of particle trajectories.

Let us assume thus that two particles a and b of a two-component bound system move at first on two concentric circular orbits, the coordinates of which are

$$a^\mu \equiv (a^0, -r_a \cos \omega a^0, -r_a \sin \omega a^0), \qquad b^\mu \equiv (b^0, r_b \cos \omega b^0, \;\; r_b \sin \omega b^0), \tag{3.3}$$

where ω is the angular velocity of the particles about a common origin, which is related to the radii of the orbits and the speeds of the particles by $v_a = \omega r_a$ and $v_b = \omega r_b$. It is convenient to introduce the retardation angle ϑ that means the angle through which one particle travels in the time it takes light to reach it from the other particle. This angle is given by the positive root of the retardation relation

$$\vartheta^2 - v_a^2 - v_b^2 - 2 v_a v_b \cos \vartheta = 0 \tag{3.4}$$

and is pictured by Fig. 2.

We now confront our postulate that the orbits are to be circular with the general form of the equations of motion, inserting the coordinates (3.3) into Eqs. (2.4) and (2.26) for the a particle and into the analogical equations for the b particle. So we obtain six equations for the components of the equations of motion that, together with Eq. (3.4), give certain conditions on ω, v_a, and v_b. The straightforward calculation shows that the temporal and tangential components of these equations are satisfied identically, namely due to applying the time-symmetric θ interaction (in the presence of the purely retarded interaction this would not be the case). The radial components of the equations of motion are

$$m_a \gamma_a \omega v_a = k(\vartheta + v_a v_b \sin \vartheta)^{-1} \times (v_a \gamma_b^{-2} + v_b \gamma_a^{-2} \cos \vartheta + v_b \vartheta \sin \vartheta), \tag{3.5}$$

for the a particle and

$$m_b \gamma_b \omega v_b = k(\vartheta + v_a v_b \sin \vartheta)^{-1} \times (v_b \gamma_a^{-2} + v_a \gamma_b^{-2} \cos \vartheta + v_a \vartheta \sin \vartheta), \tag{3.6}$$

for the particle b [here $\gamma_i = (1 - v_i^2)^{-1/2}$; $i = a,b$]. These manifestly symmetric equations, expressing the certain constraints on the particle speeds for the circular orbits, can serve together with Eq. (3.4) not only to determine ω, v_a, and v_b of the considered objects, but also as a suitable tool for the examination of various dynamic relations (the slow-motion and instantaneous-interaction approximations and so on) of this class of motions. In Appendix B we summarize necessary equations and their solutions for the limit case of the model, when one of the two particles is very heavy, such that it remains at a fixed position, while the other particle orbits around it (a H_2-atom-like model).

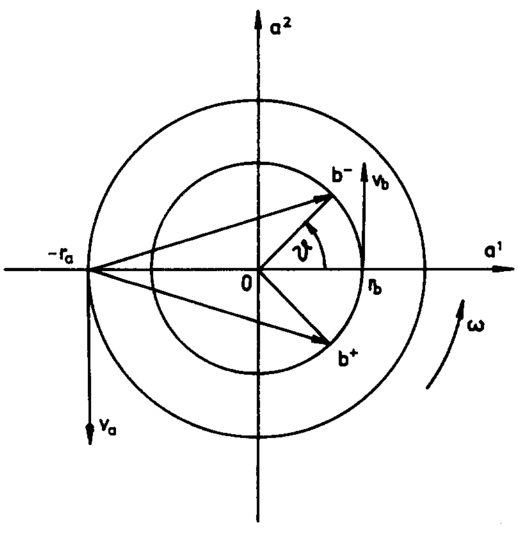

FIG. 2. Circular orbits of the two-particle bound pair. Points b^+ and b^- correspond to the retarded and advanced points, respectively, on the orbit of b, if a is at the point $a^1 = -r_a$, $a^2 = 0$.

C. Generators of motion of circular-orbit dynamics

In terms of Eqs. (3.5) and (3.6) as well as due to the above zero temporal and tangential components, we find, from Eq. (3.1) by inserting the coordinates (3.3), the following result for P^μ:

$$P^1 = P^2 = 0, \qquad P^0(\alpha_0 = 0,\ \beta_0 = 0) = m_a \gamma_a + m_b \gamma_b - k\omega^{-1}\vartheta. \tag{3.7}$$

The formula (3.7) shows that in (3.1) one of two potential energies was completely canceled by the result achieved in the process of integrating the interaction term. The last term in Eq. (3.7) can be replaced, if needed, by the expression that follows from Eq. (3.5) [alternatively from (3.6)], arranged by (3.4). So we have

$$\begin{aligned} P^0(0,0) = {} & m_a \gamma_a^{-1} + m_b \gamma_b^{-1} \\ & - m_a \gamma_a v_a (1 + v_a^2)(v_a \gamma_b^{-2} + v_b \gamma_a^{-2} \cos \vartheta \\ & + v_b \vartheta \sin \vartheta)^{-1} (v_b^2 + v_a v_b \cos \vartheta) \\ & - m_b \gamma_b v_b (1 + v_b^2)(v_b \gamma_a^{-2} + v_a \gamma_b^{-2} \cos \vartheta \\ & + v_a \vartheta \sin \vartheta)^{-1} (v_a^2 + v_a v_b \cos \vartheta) \end{aligned} \tag{3.8}$$

The energy of the two-body system bounded via the linear potential is thus given in the AAD theory by a more ample relation than the energy of two charges bounded to circular orbits in electrodynamics,[14] or in the scalar theory,[11] where, in both the cases, P^0 is represented only by the first two terms of the type indicated in (3.8). According to (3.8) the free motion of the constituents (bounded at first) emerges in the region of the "asymptotical freedom," in the limit $r_a \to 0$, $r_b \to 0$, when $v_a \to 0$ and $v_b \to 0$. For relativistic velocities that are sufficiently high ($v_a \sim c$, $v_b \sim c$), on the contrary, the energy tends to infinity, as expected. One readily sees that the last term in Eq. (3.7), representing the potential energy, contributes, naturally, to both the nonrelativistic and relativistic limits.

In the nonrelativistic limit ($v_a \ll c$, $v_b \ll c$, $\vartheta \to 0$) we have, from (3.8), $E = P^0 = E_k - m_a v_a^2 - m_b v_b^2 - (m_a + m_b) v_a v_b$, where E_k is the kinetic energy of the system. On the other hand, in the relativistic instantaneous-interaction limit ($\vartheta \to 0$, v_a and v_b being any relativistic velocities), Eq. (3.8) provides $P^0 = m_a \gamma_a^{-1} + m_b \gamma_b^{-1} - (m_a \gamma_a + m_b \gamma_b) v_a v_b$. Notice that because of (3.4) the behavior of ϑ in the limit $\vartheta \to 0$ is characterized by the formula $\vartheta \sim (v_a + v_b)(1 + v_a v_b)^{-1/2}$. It is evident that in the first of the mentioned cases the approximation formula for P^0 proves the validity of *the virial theorem*, because the radial equations (3.5) and (3.6) imply here $m_a v_a = m_b v_b$. In the second case, however, from (3.5) and (3.6) for orbits with finite r, one has $m_a \gamma_a v_a = m_b \gamma_b v_b = k\omega^{-1}(1 + v_a v_b)^{-1/2}$ and since the last term in P^0 is zero in the time averaging, we obtain the virial theorem for our relativistic case in the form $E = \langle P^0 \rangle = m_a \langle \gamma_a^{-1} \rangle + m_b \langle \gamma_b^{-1} \rangle$, where the symbol $\langle\ \rangle$ denotes time average. It turns out that exactly the same result may be reproduced for a two-body system with circular orbits that *are bounded* in the AAD theory with the potentials (2.6), using, for the derivation of the virial theorem the conserved *quantity D* (see Ref. 13), the dilation scalar that is deduced as a consequence of scale invariance.[18,19]

The evaluation of the components of the total angular momentum according to (3.2) for the circular orbits of the particles with the coordinates defined by (3.3) yields

$$L^{\mu\nu} = (\delta_1^\mu \delta_2^\nu - \delta_2^\mu \delta_1^\nu) L, \tag{3.9}$$

where

$$L = 2\tilde{g}\omega^{-2}(\vartheta + v_a v_b \sin \vartheta), \quad \tilde{g} = |g_a g_b|. \tag{3.10}$$

Here we have used the above-mentioned conditions for the components of the equations of motion. This simple result for $L^{\mu\nu}$ is obviously one of the consequences of our excessively simple model of the composite system with the regular behavior of its ingredients at small distances. It shows that in the conditions of the "asymptotic freedom" ($r_a \to 0$, $r_b \to 0$, $\vartheta \to 0$), also L disappears, as it might be expected.

D. Energy spectrum

Now we are able to derive the spectrum of the considered system. In the applied theory, clearly, one can use the Bohr–Sommerfeld conditions for quantizing L and write $L = n$. Then in the limit $m_a = m_b = m$, $v_a = v_b = v$, and, for instance, $g_a = -g_b = g$ from Eqs. (3.4), (3.8), and (3.10), the energy levels of this crude type of the "quarkonium" up to the second order in v/c become

$$E_n = 2m - \tfrac{3}{8} k^{-2} m \omega^4 n^2, \tag{3.11}$$

where $k = 2g^2$. Next, one needs to express the angular velocity in (Eq. (3.11) in terms of fundamental constants of the system. Due to the relativistic approximation, accepted in E_n, to this purpose it is allowed to apply the nonrelativistic modifications of the corresponding equations (see Appendix B). Hence, taking Eqs. (3.4), (3.5), or (3.6) and (3.10) with L quantized in the nonrelativistic approximation, one acquires immediately the sought relation

$$\omega = (k^2 m^{-1} n^{-1})^{1/3}.$$

Finally we have

$$E_n = 2m - \tfrac{3}{8}(k^2 m^{-1} n^2)^{1/3}. \tag{3.12}$$

We see that the energy spectrum tends to look like *the spectrum of the linear harmonic oscillator*. The energy levels of the system of the circular-orbit type with the AAD linear potentials appear to be similar to these of the harmonic-oscillator potential, but only for small n. For higher n the levels are shifted down relatively to the harmonic-oscillator spectrum and so in this case the energy spacings are progressively smaller. As a consequence of neglecting the fine and hyperfine structure in this model, the energy levels exhibit the structure of *the H_2-atom-like bound system*. Such a system can be immediately demonstrated by way of use Eqs. (3.4)–(3.6), (3.8), and (3.10) going, for instance, to the limit $m_a \to \infty$, $v_a \to 0$. This may be seen in Appendix B, where we give summarily both the relativistic and nonrelativistic versions of equations and their solutions for this model. We add that one meets with the similar situation also in electrodynamics.[14]

IV. CONCLUDING REMARKS

The question posed in the title of our paper obviously may be answered in the positive sense. Calculations of the previous sections have shown that the construction of AAD linear confinement *can be carried out* and, moreover, in a way *analogical* to the theory of the classical electrodynamic potentials. The fact that the AAD linear potentials *are generated* by the Lorentz invariant *θ function* is not trivial and should be considered *by itself* one of *remarkable and unexpected* outcomes for the AAD theory. Disregarding questions connected with the possible applicability of the new potentials, this finding is supposed to give other motivations *to extend* the theory of direct interparticle action. In the context with contemporary unification trends it could be perhaps interesting to examine whether present dynamics satisfy requirements of *gauge covariant theories*,[7,8] what are conditions for passing to its *Hamiltonian formulation*,[4] and to resolving problems associated with *the quantization*.

Section III was devoted to the current Schild problem[14] of an AAD dynamics, which offers a basis for *qualitative mimicking* the properties of the long range part of the color force and formally allows us to give a *crude* qualitative quarklike interpretation to the potentials (2.6). This most trivial model of the circular-orbit-type quarkonium, reckoning *only* with the linear potentials as the products of the θ function action (1.1), could be, of course, "improved" by adding the "electromagnetic" function part to this action. In such a quarkoniumlike model the short-range region of the color force will be determined by potentials of the Liénard–Wiechert type, while the long-range region can be determined inversely by our potentials.

It is obvious that the question of whether quarks can be confined in general by the linear potential of the form (2.6) remains *open*. We must realize that in the approach considered the equations of motion (2.4) were not derived from the first principles *by eliminating the field variables* in favor of the coordinates of the other particle in *the complete classical action* of chromodynamics. In addition, it is questionable whether the quantum counterpart of Eqs. (2.4) is a Dirac-form equation. This modification of equations of motion, as it has been shown recently,[20] does not admit the quarks to be confined by vector gluon potentials of the linear character.

Nevertheless, AAD dynamics, proposed in this paper, appears to be interesting from the both mathematical and physical point of views. Moreover, here one sees again why dynamics *à la* Wheeler–Feynman is considered one of *most convincing* and *aesthetical* classical *axiomatic* relativistic theories of pointlike objects. The content of the present and following[13] articles offers, we hope, sufficient evidence that there exists a *full justification* for the class of the potentials (2.6) in the AAD theory. It seems that there exists also reason to believe that they can also mean a good starting point for a future, more profound exploration of the quark-sort interaction within the framework of this theory.

ACKNOWLEDGMENT

The author is indebted to M. Petras for numerous discussions.

APPENDIX A: ACTION PRINCIPLE AND EQUATIONS OF MOTION

We derive the equations of motion from the variational principle (2.1) with the action given by (2.2). We consider one of two cases: the application of $\delta S = 0$ on the a particle. The above variation yields

$$m_a \int_{-\infty}^{+\infty} \frac{\partial}{\partial \dot a^\mu} (\eta_{\mu\nu}\dot a^\mu \dot a^\nu)^{1/2} \delta \dot a^\mu \, d\lambda_a$$
$$+ g_a g_b \iint_{(\sigma_1)}^{(\sigma_2)} \left\{ \frac{\partial}{\partial a^\mu} [\theta(ab^\rho ab_\rho)\eta_{\mu\nu}\dot a^\mu \dot b^\nu]\delta a^\mu \right.$$
$$\left. + \frac{\partial}{\partial \dot a^\mu} (\eta_{\mu\nu}\dot a^\mu \dot b^\nu)\delta \dot a^\mu \right\} d\lambda_a \, d\lambda_b = 0,$$

or

$$m_a \int_{-\infty}^{+\infty} \eta_{\mu\nu}\dot a^\nu (\eta_{\rho\sigma}\dot a^\rho \dot a^\sigma)^{-1/2} \frac{d}{d\lambda_a} (\delta a^\mu) d\lambda_a$$
$$+ 2 g_a g_b \iint_{(\sigma_1)}^{(\sigma_2)} \theta'(ab^\rho ab_\rho) ab_\sigma \eta_{\mu\nu} \, da^\mu \, db^\nu \, \delta a^\sigma$$
$$+ g_a g_b \iint_{(\sigma_1)}^{(\sigma_2)} \theta(ab^\rho ab_\rho)\eta_{\mu\nu}\dot b^\nu$$
$$\times \frac{d}{d\lambda_a} (\delta a^\mu) d\lambda_a \, d\lambda_b = 0, \qquad \text{(A1)}$$

where θ' denotes the derivative θ with respect to the argument and one accepts the notation quoted in Sec. I. Next, we perform the integration by parts in the first and third members of Eq. (A1). In addition, since it holds that $\delta a^\mu(\lambda_a = \infty) = \delta a^\mu(\lambda_a = -\infty) = 0$, one has

$$\left\{ m_a \int_{-\infty}^{+\infty} \frac{d}{d\lambda_a} [\eta_{\mu\nu}\dot a^\nu (\eta_{\rho\sigma}\dot a^\rho \dot a^\sigma)^{-1/2}] d\lambda_a \right.$$
$$- 2 g_a g_b \iint_{(\sigma_1)}^{(\sigma_2)} \delta(ab^\rho ab_\rho)(ab_\mu \, da^\nu \, db_\nu$$
$$\left. - ab_\nu \, da^\nu \, db_\mu) \right\} \delta a^\mu = 0, \qquad \text{(A2)}$$

replacing at the same time θ' by δ. We pass now to the proper time in which $\dot a^\nu \dot a_\nu = 1$. Then Eq. (A2) implies

$$dp_{a\mu} = 2 g_a g_b \, da^\nu \int_{-\infty}^{+\infty} \delta(ab^\rho ab_\rho)$$
$$\times (ab_\mu \, db_\nu - ab_\nu \, db_\mu), \qquad \text{(A3)}$$

where $p_a^\mu = m_a \eta^{\mu\nu} \dot a_\nu$ is the momentum of the a particle. Equations (A3) correspond precisely to the equations of motion (2.4) of Sec. II with $F_{\mu\nu}^{(b)}$ and $A_\mu^{(b)}$ given by (2.5) and (2.6), respectively, which can be uniquely verified from here by calculating $dp_{a\mu}$:

$$dp_{a\mu} = m_a \ddot a_\mu \, d\alpha$$
$$= g_a g_b \left(\partial_\mu \int_{-\infty}^{+\infty} \dot b_\nu - \partial_\nu \int_{-\infty}^{+\infty} \dot b_\mu \right) \theta(ab^\rho ab_\rho) d\beta$$
$$= 2 g_a g_b da^\nu \int_{-\infty}^{+\infty} \delta(ab^\rho ab_\rho)$$
$$\times (ab_\mu \, db_\nu - ab_\nu \, db_\mu).$$

APPENDIX B: LIMIT CASES OF THE CIRCULAR-ORBIT MODEL

In this appendix we collect the main equations referred to as the circular-orbit model, specifically to its H_2-atom-like reduced form.

We take the particle a to be at the rest with $m_a \to \infty$, $v_a = 0$, $r_a = 0$, and b to revolve around it. The radial equation (3.6) is now reduced to the equation

$$m_b \gamma_b v_b = k\omega^{-1}, \quad k = 2 g_a g_b, \qquad \text{(B1)}$$

since Eq. (3.4) gives the condition $\vartheta = v_b$. The energy E and the quantized angular momentum ($L = n$) are yielded by the relations (3.7) and (3.10), respectively, and read

$$E = m_a + m_b \gamma_b - k r_b, \qquad \text{(B2)}$$
$$L = k\omega^{-2} v_b = n. \qquad \text{(B3)}$$

The nonrelativistic versions of Eqs. (B1)–(B3) are the familiar forms

$$m_b r_b \omega^2 = k, \qquad \text{(B4)}$$
$$E = \tfrac{1}{2} m_b v_b^2 - k r_b, \qquad \text{(B5)}$$
$$m_b r_b v_b = n. \qquad \text{(B6)}$$

From Eqs. (B4)–(B6) and the formula $v_b = r_b \omega$, one immediately obtains the characteristic parameters of the nonrelativistic model:

$$r_b = (k^{-1} m^{-1} n^2)^{1/3}, \quad \omega = (k^2 m^{-1} n^{-1})^{1/3},$$
$$v_b = (k m^{-2} n)^{1/3}, \quad E_n = -\tfrac{1}{2} (k^2 m^{-1} n^2)^{1/3}. \qquad \text{(B7)}$$

Also the relativistic equations (B1)–(B3) are soluble. They lead, however, to the cubic equation for v_b:

$$m_b^2 v_b^3 + nkv_b^2 - nk = 0. \quad \text{(B8)}$$

Supposing b to be a sufficiently massive particle (and accepting then only the real solution), one finds

$$E = m_a + m_b (1 - v_b^2)^{1/2},$$
$$\omega = (kn^{-1})^{1/2} v_b^{1/2}, \quad r_b = (k^{-1}n)^{1/2} v_b^{-1/2}, \quad \text{(B9)}$$
$$v_b = s^{1/3}(d + D^{1/2})^{1/3} + (d - D^{1/2})^{1/3} - s/3,$$

where $d = 2^{-1} - 3^{-3}s^2$, $D = 2^{-2} - 3^{-3}s^2$, and $s = km^{-2}n$.

It is evident that in the approximation up to $(v/c)^2$ the relativistic formula for E is reduced to completely simple expression

$$E_n = m_a + m_b - \tfrac{1}{2}(k^2 m^{-1} n^2)^{1/3},$$

close by its structure to the energy spectrum of the more general model investigated in Sec. III [compare with the formula (3.12)].

[1]J. A. Wheeler and R. P. Feynman, Rev. Mod. Phys. **17**, 157 (1945); **21**, 425 (1949).

[2]R. Marnelius, Phys. Rev. D **10**, 2535 (1974).

[3]P. Stephas, Am. J. Phys. **46**, 360 (1978).

[4]P. Havas, Phys. Rev. **87**, 309 (1952); **91**, 997 (1953); P. Ramond, Phys. Rev. D **7**, 449 (1973); S. Raha, Phys. Rev. D **19**, 3641 (1979).

[5]J. W. Dettman and A. Schild, Phys. Rev. **95**, 1057 (1954); H. van Dam and E. P. Wigner, Phys. Rev. **138**, B1576 (1965); **142**, 838 (1966).

[6]A. D. Fokker, Z. Phys. **58**, 386 (1929); Physica **9**, 33 (1929).

[7]R. Jackiw, Rev. Mod. Phys. **49**, 681 (1977).

[8]J. L. Anderson, in *Principles of Relativity Physics* (Academic, New York, 1967).

[9]J. V. Narlikar, Found. Phys. **13**, 311 (1983).

[10]C. M. Andersen and H. C. von Baeyer, Ann. Phys. (N.Y.) **60**, 67 (1970). Alternative forms of dynamics for classical relativistic particles can be found in Ref. 5 and also A. Katz, J. Math. Phys. **10**, 1929 (1969).

[11]F. Hoyle and J. V. Narlikar, Proc. Soc. London Ser. A **282**, 191 (1964).

[12]K. Hirata, Nucl. Phys. B **169**, 165 (1980).

[13]J. Weiss, J. Math. Phys. **27**, 1023 (1986).

[14]A. Schild, Phys. Rev. **131**, 2762 (1963).

[15]J. Weiss, preprint STU-K14-P3, 1985.

[16]N. Mukunda, H. van Dam, and L. C. Biedenharn, Phys. Rev. D **22**, 1938 (1980); R. R. Aldinger, A. Bohm, P. Kielanowski, M. Loewe, P. Magnollay, N. Mukunda, W. Drechsler, and S. R. Komy, Phys. Rev. D **28**, 3020 (1983); A. Bohm, M. Loewe, L. C. Biedenharn, and H. van Dam, Phys. Rev. D **28**, 3032 (1983).

[17]R. D. Driver, Phys. Rev. D **19**, 1098 (1979).

[18]J. Weiss, preprint STU-K14-P4, 1985.

[19]C. M. Andersen and H. C. von Baeyer, Am. J. Phys. **39**, 914 (1971).

[20]H. B. Ai and J. P. Hsu, Found. Phys. **15**, 155 (1983).

Action-at-a-distance linear potentials and conformal conservation laws

J. Weiss
Department of Physics, Slovak Technical University, 812 31 Bratislava, Czechoslovakia

(Received 1 August 1985; accepted for publication 28 October 1985)

As a consequence of invariance under the full conformal group of transformations, 15 proper conserved quantities are derived for the system of two massive particles interacting via classical action-at-a-distance linear potentials that have been found in the preceding paper [J. Weiss, J. Math. Phys. **27**, 1015 (1986)] in terms of the step θ function Lagrangian.

I. INTRODUCTION

In the previous paper[1] we have derived on the basis of the action-at-a-distance (AAD) theory a new type of potential that reminds us by its genesis of the origin of the known electrodynamic Liénard–Wiechert potentials. Applying *the Fokker–Wheeler–Feynman approach*,[2] known from classical AAD electrodynamics, to an action functional, defined in terms of the relativistically invariant *Heaviside step function*, we have found the AAD class of potentials exhibiting *linear confinement*. The AAD linear potentials A are defined as follows[1]:

$$A^{(b)\mu}(x) = g_b \int_{-\infty}^{+\infty} \theta(xb^{\nu}xb_{\nu})\,db^{\mu}(\beta)\,. \tag{1}$$

In (1) θ is the step function and it is used here with the Wheeler and Feynman notation[2] of $a^\mu(\alpha)$ and $b^\mu(\beta)$ for the four-vectors of the coordinates of particles (a and b) of a two-component system and $a^\mu - b^\mu = ab^\mu$ for their difference. In the preceding work[1] it was shown that the formula (1) yields actually the linear potentials, expressed by way of their *retarded* A^+ and *advanced* A^- components,

$$A^{(b)\mu}(x) = \tfrac{1}{2}(A^{(b)\mu+}(x) + A^{(b)\mu-}(x))\,, \tag{2}$$

with

$$\begin{aligned} A^{(b)\mu+}(x) &= g_b b^\mu(\beta)_{s_+=0}\,,\\ A^{(b)\mu-}(x) &= -g_b b^\mu(\beta)_{s_-=0}\,, \end{aligned} \tag{3}$$

where α and β are the particle proper times and the dependence of A on x is given *implicitly* in terms of the quantities $s_\pm$:

$$s_\pm = |\mathbf{x} - \mathbf{b}| \mp (x^0 - b^0)\,. \tag{4}$$

In this sense, the potentials that are defined by Eq. (1) or Eqs. (2)–(4) represent in the relativistic region, where the retarded and advanced effects of interactions are not negligible, the linear-potential analogies of the well-known Liénard–Wiechert potentials. They are as created in terms of the Heaviside θ function as the Liénard–Wiechert potentials in terms of the Dirac δ function. Some of properties of these potentials were analyzed in Ref. 1, others will be discussed in this article.

In Ref. 1 we have proved that the potentials (1) can be successfully examined on familiar two-constituent systems such as circular-orbit models of the bound particle pair (Schild's problem[3]). To this aim it has been necessary to find the generators of the Poincaré group that characterize dynamics with our potentials. The generators of Poincaré's group—*the momentum* P^μ and *the angular momentum* $L^{\mu\nu}$ —together with *the dilation scalar D* and *the conformal four-vector K* (see below) form the 15-parameter *conformal group*, which is an O(5,1) group of invariances of the Yang–Mills theory[4] and hence a relevant group of coordinate transformations required by contemporary field theories.

Our AAD dynamics is obliged to satisfy main requirements imposed on a consistent relativistic theory, exactly like AAD electrodynamics, i.e., it *must be invariant* under the full conformal group. Consequently, the potentials (1) must produce all 15 of above conserved quantities. In this paper we link to results of the previous work[1] and try to derive the complete set of these quantities corresponding to the AAD linear potentials.

Starting with the classical works of Wheeler and Feynman[2] up to recent outcomes achieved by Stephas and von Baeyer,[5] the development of AAD theories has shown successively (for more details on the problem see Refs. 5 and 6) that the conformal conservation laws can be constructed for the Liénard–Wiechert as well as for scalar potentials. For both classes of potentials the input tool underlying calculations is a Lagrangian with the Lorentz invariant Dirac δ function, the argument of which is the square of the distance between a source and an observer. In the case of the AAD linear potentials the Lagrangian is formed, as emphasized, by the Heaviside step function θ of the same argument.[1]

To avoid tedious evaluations, accompanied by methods that escape metrics, we use a reliable and flexible method, elaborated by Wheeler and Feynman,[2] convenient also for our purposes. It consists in computing infinitesimal changes in the investigated quantities and in applying appropriate equations of motion.

II. MOMENTUM

We begin from the definition of the canonical momentum P^μ_a associated with the particle a. It reads

$$P^\mu_a(\alpha) = m_a \dot{a}^\mu(\alpha) + g_a A^{(b)\mu}(\alpha)\,, \tag{5}$$

m_a being the mass of the a particle and $A^{(b)\mu}(\alpha)$ is the potential produced by the particle b and experienced by a at the proper time α. Now, the change of P^μ_a in the interval of proper time $d\alpha$ due to action of the b particle is straightforwardly

$$dP^\mu_a(\alpha) = (m_a \ddot{a}^\mu(\alpha) + g_a \dot{A}^{(b)\mu}(\alpha))\,d\alpha\,, \tag{6}$$

 0022-2488/86/041023-04$02.50

where the dot on the quantities denotes differentiation with respect to the proper time. Taking into account the equations of motion

$$m_a \ddot{a}^\mu = g_a F^{(b)\mu}_{\ \ \nu}(a)\dot{a}^\nu, \tag{7}$$

where

$$F^{(b)}_{\mu\nu}(x) = A^{(b)}_{\nu,\mu}(x) - A^{(b)}_{\mu,\nu}(x) \tag{8}$$

and the potentials $A^{(b)}_\mu$ are given by (1) and arranging the term $\dot{A}^{(b)\mu}$ by making use of the obvious relation $da^\nu(\alpha) = \dot{a}^\nu(\alpha)d\alpha$, one obtains

$$dP^\mu_a = 2g_a g_b \int_{-\infty}^{+\infty} \delta(ab^\rho ab_\rho) ab^\mu \dot{a}^\nu \dot{b}_\nu \, d\alpha\, d\beta . \tag{9}$$

In Eq. (9) the Dirac δ function replaces the Heaviside θ function differentiated with respect to the indicated argument ($\delta = \theta'$). Precisely the same result with the negative sign is found, if one uses the formula for P^μ_b and carries out the analogous procedure as above. Thus, an interchange of the roles of both particles in the integrand of (9) results only in the change of its sign. This could be expected, of course, following the form of (9), too. The expression (9) defines the infinitesimal value of *the momentum transfer*. Due to the presence of the delta function in this relation the b particle acts on a (and vice versa) exclusively through both *retarded and advanced forces*. Moreover, as it is seen, the relativistic *law of action and reaction* holds, which ensures that the momentum transferred from a to b via retarded forces is equal in magnitude and opposite in sign to the momentum transferred from b to a via advanced forces over the same infinitesimal world-line intervals (and inversely). Owing to these properties embedded in (9), if we then choose two values α_0 and β_0 of the proper times on the appropriate world-lines and compute finite contributions to the total P^μ, we are always led to two world-line segments (from α^+ to α^- and from β^+ to β^-). Consequently, the *interaction* momentum P^μ_I can be written in the form

$$P^\mu_I = k\left(\int_{\alpha_0}^{+\infty} d\alpha \int_{-\infty}^{\beta_0} d\beta - \int_{-\infty}^{\alpha_0} d\alpha \int_{\beta_0}^{+\infty} d\beta\right) \times \delta(ab^\rho ab_\rho) ab^\mu \dot{a}^\nu \dot{b}_\nu ,\ k = 2g_a g_b , \tag{10}$$

which demands the integrals in (10) to be evaluated over finite segments of the world-lines and demonstrates the law of action and reaction.

So we arrive at the first of four generators of Poincaré's group—the conserved total momentum

$$\begin{aligned} P^\mu(\alpha_0,\beta_0) &= P^\mu_a + P^\mu_b + P^\mu_I \\ &= m_a \dot{a}^\mu(\alpha_0) + g_a A^{(b)\mu}(\alpha_0) \\ &\quad + m_b \dot{b}^\mu(\beta_0) + g_b A^{(a)\mu}(\beta_0) \\ &\quad + k\left(\int_{\alpha_0}^{+\infty} d\alpha \int_{-\infty}^{\beta_0} d\beta - \int_{-\infty}^{\alpha_0} d\alpha \int_{\beta_0}^{+\infty} d\beta\right) \\ &\quad \times \delta(ab^\rho ab_\rho) ab^\mu \dot{a}^\nu \dot{b}_\nu . \end{aligned} \tag{11}$$

It is easy to show that $[P^\mu(\alpha,\beta)]_{\alpha=\alpha_0,\beta=\beta_0}$ is a constant with respect to a given proper time. One needs only differentiate Eq. (11) with respect to α or β and use Eqs. (1) and (7) and (8).

III. ANGULAR MOMENTUM

We proceed as in the previous case. It is sufficient to calculate the change of $L^{\mu\nu}_a$ in the interval of the proper time in order to obtain $L^{\mu\nu}_I$. If one writes $L^{\mu\nu}_a$ to be

$$L^{\mu\nu}_a = a^\mu P^\nu_a - a^\nu P^\mu_a, \tag{12}$$

with P^μ_a given by (1.5), then evidently

$$\begin{aligned} dL^{\mu\nu}_a = g_a (&\dot{a}^\mu A^{(b)\nu} - \dot{a}^\nu A^{(b)\mu} + a^\mu \dot{a}_\sigma A^{(b)\sigma,\nu} \\ &- a^\nu \dot{a}_\sigma A^{(b)\sigma,\mu})\, d\alpha . \end{aligned} \tag{13}$$

In (13) we have used again the equations of motion (7) and (8) and the same arrangement concerning the terms with $\dot{A}^{(b)\mu}$ as above in computing P^μ_I. Next, substitute (1) into (13) and make the analysis of integrals entered. One finds properties of the integrals to be similar to those discussed in the case of the momentum. This implies immediately the expression for the *interaction* term $L^{\mu\nu}_I$. It is involved in the second set of the Poincaré group generators consisting of six antisymmetric tensor components of the conserved total angular momentum:

$$\begin{aligned} L^{\mu\nu}(\alpha_0,\beta_0) &= L^{\mu\nu}_a + L^{\mu\nu}_b + L^{\mu\nu}_I \\ &= (a^\mu P^\nu_a - a^\nu P^\mu_a)_{\alpha_0} + (b^\mu P^\nu_b - b^\nu P^\mu_b)_{\beta_0} \\ &\quad + k\left(\int_{\alpha_0}^{+\infty} d\alpha \int_{-\infty}^{\beta_0} d\beta - \int_{-\infty}^{\alpha_0} d\alpha \int_{\beta_0}^{+\infty} d\beta\right) \\ &\quad \times \{\delta(ab^\rho ab_\rho)(a^\nu b^\mu - a^\mu b^\nu)\dot{a}^\sigma \dot{b}_\sigma \\ &\quad + \tfrac{1}{2}\theta(ab^\rho ab_\rho)(\dot{a}^\mu \dot{b}^\nu - \dot{a}^\nu \dot{b}^\mu)\} . \end{aligned} \tag{14}$$

In the preceding paper the generators (11) and (14) have been derived from the requirement of the invariance of the action with the step function Lagrangian under the Poincaré group of infinitesimal mappings [compare with Eqs. (3.1) and (3.2) of Ref. 1]. The fact that $L^{\mu\nu}$ of (14) is an integral of motion can be verified straightforwardly: one carries out the differentiations with respect to α_0 and β_0 in (14) and takes into account Eqs. (1), (7), and (8).

IV. DILATION SCALAR

So far, we have considered the conserved quantities that form the ten-parameter Poincaré group as a subgroup of the full conformal group. One of special conformal transformations belonging to the full conformal group is the one-parameter dilation transformation. This transformation changes the scale of the space-time interval, but does not change the geometry of particle orbits. The corresponding conserved quantity is the *dilation* scalar D. It is defined in the following way:

$$D = z^\mu P_\mu , \tag{15}$$

where z^μ is the coordinate of a particle and P^μ its canonical momentum. In AAD electrodynamics of the Wheeler and Feynman type, D has been found by Andersen and von Baeyer.[7]

For potentials of the class (1), D can be readily deduced by the method under consideration. We omit details and give only the result achieved from the infinitesimal *change* of $D_a = a^\mu P_{a\mu}$:

$$dD_a = m_a + k\int_{-\infty}^{+\infty} \delta(ab^\rho ab_\rho) ab^\mu (a_\mu + b_\mu) \times \dot a^\nu \dot b_\nu \, d\alpha\, d\beta\,. \qquad (16)$$

It is evident that the conservation of D requires, conformably to electrodynamics,[7,8] the inclusion of a *mass* term. In Eq. (16) m_a is created by the term $P^\mu_a \dot a_\mu$ provided (5) and the condition $\dot a^\mu \dot a_\mu = 1$ hold.

The *interaction* dilation D_I then has the form

$$D_I(\alpha_0,\beta_0) = k\left(\int_{\alpha_0}^{+\infty} d\alpha \int_{-\infty}^{\beta_0} d\beta - \int_{-\infty}^{\alpha_0} d\alpha \int_{\beta_0}^{+\infty} d\beta\right) \times \delta(ab^\rho ab_\rho) ab^\mu (a_\mu + b_\mu)\dot a^\nu \dot b_\nu - m_a\alpha_0 - m_b\beta_0\,, \qquad (17)$$

consistent with the requirement for the total scalar D,

$$D(\alpha,\beta) = P^\mu_a(\alpha) a_\mu(\alpha) + P^\mu_b(\beta) b_\mu(\beta) + D_I(\alpha,\beta)\,, \qquad (18)$$

to be conserved. One can show that the dilation (18) with D_I by (17), used in the deduction of *the virial theorem*, leads for finite and closed orbits to a result[9] that is identical with that yielded by electrodynamics.[7]

V. CONFORMAL FOUR-VECTOR

Finally, it is necessary for us to deal with the four-parameter special conformal transformation that is represented by the *conformal* vector K^μ, reducing in the case of a two-component system to the expression

$$K^\mu = 2a^\mu P^\nu_a a_\nu - P^\mu_a a^\nu a_\nu + 2b^\mu P^\nu_b b_\nu - P^\mu_b b^\nu b_\nu + K^\mu_I\,, \qquad (19)$$

where the *interaction* term is termed K^μ_I. The K^μ vector does not seem to have a fixed interpretation yet.[5] It is not ruled out that it means a generalized coordinate-dependent dilation. For the massless scalar and vector (electromagnetic) potentials, when both time-symmetric and time-asymmetric interactions are present, the K have been evaluated by Stephas and von Baeyer.[5]

The path, sketched by Wheeler and Feynman[2] and applied now to the linear potentials (1), proposes first of all the construction of the infinitesimal change in K^μ_a. We have

$$dK^\mu_a(\alpha) = 2m_a a^\mu + k\int_{-\infty}^{+\infty} \{\theta(ab^\rho ab_\rho)(\dot a^\mu \dot b_\nu - \dot b^\mu \dot a_\nu) a^\nu + \delta(ab^\rho ab_\rho)[(a^\mu + b^\mu) a^\nu a_\nu - 2a^\mu b^\nu b_\nu]\dot a^\sigma \dot b_\sigma\} d\alpha\, d\beta\,, \qquad (20)$$

where the relation (1) is used together with its equivalent version[1]

$$A^{(b)\mu}(a) = 2g_b \int_{-\infty}^{+\infty} \delta(ab^\rho ab_\rho) ab^\nu \dot b_\nu b^\mu \, d\beta\,,$$

obtained by performing an integration by parts. To determine K^μ_I we need to symmetrize the integrand in (20) and recast available integration. One finds

$$\begin{aligned} K^\mu_I = {} & g_a g_b\left(\int_{\alpha_0}^{+\infty} d\alpha \int_{-\infty}^{\beta_0} d\beta - \int_{-\infty}^{\alpha_0} d\alpha \int_{\beta_0}^{+\infty} d\beta\right) \\ & \times [\theta(ab^\rho ab_\rho)(\dot a^\mu \dot b^\nu - \dot b^\mu \dot a^\nu)(a_\nu + b_\nu) \\ & + \delta(ab^\rho ab_\rho) ab^\mu (a^\nu a_\nu - b^\nu b_\nu)\dot a^\sigma \dot b_\sigma] \\ & - g_a g_b\left(\int_{-\infty}^{+\infty} d\alpha \int_{-\infty}^{\beta_0} d\beta + \int_{-\infty}^{\alpha_0} d\alpha \int_{-\infty}^{+\infty} d\beta\right) \\ & \times \{\theta(ab^\rho ab_\rho) ab_\nu (\dot a^\mu \dot b^\nu - \dot b^\mu \dot a^\nu) \\ & + \delta(ab^\rho ab_\rho)[(a^\mu + b^\mu)(a^\nu a_\nu - b^\nu b_\nu) \\ & + b^\mu a^\nu a_\nu - a^\mu b^\nu b_\nu]\dot a^\sigma \dot b_\sigma\} \\ & - 2m_a \int_{-\infty}^{\alpha_0} a^\mu\, d\alpha - 2m_b \int_{-\infty}^{\beta_0} b^\mu\, d\beta\,. \end{aligned} \qquad (21)$$

Additionally, the validity of conservation of the total vector K^μ, defined by (19) and (21), can be verified by analogy with the previous constants of motion.

VI. CONCLUSIONS

Thereby we have saturated the set of the conformal conservation laws and proved that AAD dynamics with *the linear potentials* is *invariant* under the full conformal group of transformations. Consequently, we can say that within the framework of the AAD theory *the existence* of linear potentials (1) is *generally secured.*

The meaning of the derived conserved quantities will be perhaps more apparent in the further development of AAD theories. One sees today that gauge covariant theories here could have a real chance to give appropriate motivations. As noted above, the possibility of applying the generators P^μ and $L^{\mu\nu}$ has been already transferred on the certain quark-like, but very crude, footing, and the typical Schild problem[3] of circular-orbit motion has been explored in the previous paper[1] in continuity with the new potentials. Now we extend the set of conservation laws for this excessively simple model of a "quarkonium" to the conserved dilation. The D rescales the size and is calculated in this model to be *zero*, as expected. The verification of this result is straightforward and demands to use the obvious identity

$$m_a \int_{\alpha_0}^{+\infty} d\alpha = [m_a \gamma_a a^0 (1 - v_a^2)]_{\alpha_0}$$

(and similar for b) having an impact on the elimination of the mass term in (17).

ACKNOWLEDGMENT

I would like to thank Professor H. C. von Baeyer for valuable suggestions, correspondence, and literature.

[1] J. Weiss, J. Math. Phys. **27**, 1015 (1986).
[2] A. D. Fokker, Z. Phys. **58**, 386 (1929); Physica **9**, 33 (1929); **12**, 145 (1932); J. A. Wheeler and R. P. Feynman, Rev. Mod. Phys. **17**, 157 (1945); **21**, 425 (1949).
[3] A. Schild, Phys. Rev. **131**, 2762 (1963).

[4]R. Jackiw, Rev. Mod. Phys. **49**, 681 (1977).
[5]P. Stephas and H. C. von Baeyer, Phys. Rev. D **20**, 3155 (1979).
[6]P. Stephas, Am. J. Phys. **46**, 360 (1978). See also H. van Dam and E. P. Wigner, Phys. Rev. B **138**, 1576 (1965); **142**, 838 (1966); A. Katz, J. Math. Phys. **10**, 1929 (1969).
[7]C. M. Andersen and H. C. von Baeyer, Am. J. Phys. **39**, 914 (1971).
[8]C. Garrod, Phys. Rev. D **1**, 1524 (1970).
[9]J. Weiss, preprint STU-K14-P4, 1985.

Potential envelopes and the large-N approximation

Richard L. Hall
Department of Mathematics, Concordia University, 1455 de Maisonneuve Blvd. W., Montreal, Quebec, Canada H3G 1M8

(Received 28 May 1985; accepted for publication 3 December 1985)

If E is an eigenvalue of the quantum-mechanical Hamiltonian $H = \frac{1}{2}\Delta + V(r)$ in N spatial dimensions, then large-N theory, the potential-envelope method, and scale-optimized variational energies all lead to quasiclassical approximations having the same form given by $E(\alpha) = \min_{r>0}[\frac{1}{2}\alpha r^{-2} + V(r)]$, where α depends on the quantum numbers and on N. Energy bounds provided by the envelope method allow us to prove that in many cases the large-N results are lower energy estimates. For pure power-law potentials all these energies approach the exact eigenvalue in either of the limits $l \to \infty$ or $N \to \infty$.

I. INTRODUCTION

The purpose of this note is to point out some connections between certain aspects of the large-N approximation[1-3] and the potential-envelope method[4,5] in quantum mechanics. The common problem domain for these two theories is the approximation of the discrete spectrum of Schrödinger Hamiltonians given by

$$H = -\tfrac{1}{2}\Delta + V(r), \quad r = |\mathbf{r}|, \quad \mathbf{r} \in \mathbb{R}^N. \tag{1.1}$$

These approaches both lead to approximations for the eigenvalues having the general form

$$E(\alpha) = \min_{r>0}[\tfrac{1}{2}\alpha r^{-2} + V(r)], \quad \alpha > 0, \tag{1.2}$$

where α is kept constant during the optimization process but depends on the quantum numbers of the eigenvalue and on the number N of spatial dimensions.

It is important to realize at the outset that the observations we are able to make concern only a part of large-N theory. This theory originated in the area of critical phenomena,[1] it has been revived and driven by the needs of contemporary field theory,[2] and has recently grown an offshoot[3] devoted to the approximation of Schrödinger eigenvalues. Our remarks pertain overtly only to a portion of this last area. On common ground the two approaches can be compared easily because their results are expressed in terms of the same quasiclassical form (1.2). Since potential envelopes yield energy *bounds* it is possible to use these results to analyze the relationship between the large-N approximation and the corresponding exact eigenvalues. For the class of quark–quark potentials that are convex transformations of the Coulomb potential, we are able to prove that the large-N approximation generates *lower* estimates to the corresponding exact energies.

II. POTENTIAL ENVELOPES

The method of potential envelopes derives from a very simple idea. We suppose that Schrödinger's equation can be solved for the eigenvalues $\mathscr{E}_{nl}(v)$ of the Hamiltonian $-\frac{1}{2}\Delta + vh(r)$. We then consider the eigenvalues E of a new Hamiltonian given by

$$H = -\tfrac{1}{2}\Delta + V(r), \quad V(r) = g(h(r)), \tag{2.1}$$

where the transformation function $g(h)$ is monotone increasing. Tangent lines to $g(h)$ are therefore shifted h-potentials of the form $g(k) - kg'(k) + g'(k)h(r)$ and these potentials lead to exact Schrödinger eigenvalues given by

$$E = g(k) - kg'(k) + \mathscr{E}_{nl}(g'(k)), \tag{2.2}$$

where $k = h(t)$ is the point of contact between $g(h)$ and its tangent. If $g(h)$ is convex, then it lies above its tangents and the min–max characterization of eigenvalues implies that $E < E_{nl}$; similarly, if $g(h)$ is concave, then $E > E_{nl}$. The final step in the derivation is to optimize these lower or upper bounds with respect to k. This theory has been generalized and refined and is described in detail in Refs. 4 and 5. What follows is a very brief derivation tailored to the present application: for simplicity, the quantum-number labels are omitted.

We consider the critical points of (2.2) given by $dE/dk = 0$. Since $g''(k) \neq 0$, we can cancel this factor and obtain the critical equation given by

$$k = \mathscr{E}'(g'(k)). \tag{2.3}$$

This suggests that we set $v = g'(k)$ and look at a new energy function given by

$$E = \mathscr{E}(v) - v\mathscr{E}'(v) + g(\mathscr{E}'(v)). \tag{2.4}$$

For the potential generators $h(r)$ that we shall use (and more generally, too[4]), the trajectory functions $\mathscr{E}(v)$ are concave so that the factor $\mathscr{E}''(v)$ can be canceled from $dE/dv = 0$ to yield the critical equation

$$v = g'(\mathscr{E}'(v)). \tag{2.5}$$

It follows that the critical values of E obtained from (2.2) and (2.4) are the same. In a final Legendre transformation we set

$$K(r) = \mathscr{E}(v) - v\mathscr{E}'(v), \quad h(r) = \mathscr{E}'(v), \tag{2.6}$$

so that $g(\mathscr{E}'(v)) = g(h(r)) = V(r)$ and the energy formula now becomes

$$E = \min_{r>0}[K(r) + V(r)]. \tag{2.7}$$

The point of these transformations is to bring $V(r)$ directly into the equation so that one can know approximately how the eigenvalues depend on the various parameters of the potential. If the potential generator $h(r)$ is a simple power law of the form

 0022-2488/86/041027-04$02.50

$$h(r) = \mathrm{sgn}(q) r^q, \quad q > -1, \quad q \neq 0, \tag{2.8}$$

then by scaling arguments we know that the trajectory function $\mathscr{E}(v)$ has the corresponding form

$$\mathscr{E}(v) = \mathrm{sgn}(q) Q v^{2/(2+q)}, \quad Q > 0. \tag{2.9}$$

It now follows from (2.6) that in these cases the kinetic-energy function $K(r)$ is given explicitly by

$$K(r) = \tfrac{1}{2}\alpha r^{-2}, \quad \alpha = |q| [2Q/(2+q)]^{(2+q)/2}. \tag{2.10}$$

Thus for problems in which the potential *generator* $h(r)$ is a pure power-law potential [the potential $V(r)$ itself, however, is *not* required to be a power law] the prescription for the energy bound always may be expressed in the form

$$E(\alpha) = \min_{r>0} [\tfrac{1}{2}\alpha r^{-2} + V(r)], \quad \alpha > 0. \tag{2.11}$$

Equation (2.11) implies that $dE/d\alpha = \frac{1}{2} r(\alpha)^{-2} > 0$, where $r(\alpha)$ is the critical value of r. Consequently, approximations generated by various different values of α are ordered in the same way as are the α's. Since the elementary ordering result just established will be so important for the various comparisons we shall make, we elevate it to the status of a lemma and write it as follows.

Ordering Lemma:

$$\frac{dE}{d\alpha} = \frac{1}{2} r^{-2} > 0.$$

The form (2.11) is firmly established as the key energy formula of large-N theory. We are able to comment on this theory precisely because our geometrical results also can be represented this way. In our earlier work the optimization variable was chosen to be the mean kinetic energy $s = K(r) = \frac{1}{2}\alpha r^{-2}$. However, in (2.11) and in what follows we express the results we need in a notation that is in harmony with the conventions adopted in large-N theory. The use of r as an optimization parameter and the retention of the vestigial factor of $\frac{1}{2}$ in the kinetic-energy term are both consequencies of this policy.

The potential-envelope method requires exact solutions on which to build. In N spatial dimensions exact solutions are not as abundant as they are for $N = 1$ or for $N = 3$. Two soluble problems that we do have at our disposal are the hydrogenic atom and the harmonic oscillator. The exact eigenvalues for these problems are given by

$$V(r) = -v/r \rightarrow \mathscr{E}_{nl}(v)$$
$$= -\tfrac{1}{2} v^2 [l + n + \tfrac{1}{2}(N-1)]^{-2}, \quad N \geqslant 2, \tag{2.12}$$

$$V(r) = v r^2 \rightarrow \mathscr{E}_{nl}(v) = (2v)^{1/2} [l + 2n + \tfrac{1}{2}N]. \tag{2.13}$$

The square of the total orbital angular momentum is given by $L^2 = l(l + N - 2)$, where $l = 0, 1, 2, 3, \ldots$, and the radial quantum number has values given by $n = 0, 1, 2, 3, \ldots$. The eigenvalues are monotone nondecreasing in n and, so labeled, they each have an "azimuthal" degeneracy, which, in general, depends in a complicated way on l and N, but for the familiar case $N = 3$ is equal to $2l + 1$. We except the Coulomb potential for $N = 1$ because this curious problem has features that are not in harmony with our main purpose.

By comparing these two formulas with (2.10) we can find the values of Q and hence the values of α to be used in (2.11). We have therefore

$$V(r) = g(-1/r) \rightarrow \alpha_1 = [l + n + \tfrac{1}{2}N - \tfrac{1}{2}]^2, \tag{2.14}$$

$$V(r) = g(r^2) \rightarrow \alpha_2 = [l + 2n + \tfrac{1}{2}N]^2. \tag{2.15}$$

Since the energy minima $E(\alpha)$ given by (2.11) are monotone in α we have from (2.14) and (2.15) that $E(\alpha_1) < E(\alpha_2)$.

Various eigenvalue bounds now immediately follow. For example, suppose the potential $V(r)$ has a dual Coulombic-harmonic envelope representation of the form

$$V(r) = g^{(1)}(-1/r) = g^{(2)}(r^2), \tag{2.16}$$

where $g^{(1)}$ is convex on $(-\infty, 0)$ and $g^{(2)}$ is concave on $(0, \infty)$. Then, by substituting, respectively, (2.14) and (2.15) into the minimization formula (2.11) we immediately obtain lower and upper bounds on each of the eigenvalues. In either of the limits $l \rightarrow \infty$ or $N \rightarrow \infty$ the ratio of these two bounds approaches unity. This dual class of potentials includes the quark–quark type of potentials given by

$$V(r) = -a/r + b \log(r) - c/(d + r) + er + fr^2, \tag{2.17}$$

where the non-negative constants $\{a\, b\, c\, d\, e\, f\}$ are not all zero. It is very convenient to be able to treat whole families of problems like this with a single formula and it is a welcome bonus to obtain energy bounds at the same time. Numerical results for spatial dimensions $N = 1, 3$ (and for the many-body problem) may be found in Refs. 4 and 5.

III. THE LARGE-N APPROXIMATION

As we mentioned in the Introduction we by no means look at all aspects of the large-N approximation. We focus attention on those immediate results for simple central potentials for which we are able to provide independent and comparable data by the use of potential envelopes. The radial part of the N-dimensional Schrödinger equation generated by the Hamiltonian $H = -\frac{1}{2}\Delta + V(r)$ may be written

$$\left[-\frac{1}{2} \left\{ \frac{d^2}{dr^2} + (N-1) r^{-1} \frac{d}{dr} \right\} + \frac{1}{2} l(l + N - 2) r^{-2} + V(r) \right] \psi(r) = E\psi(r). \tag{3.1}$$

If the radial wave function $\psi(r)$ is now factored in the form $\psi(r) = r^{-(1/2)(N-1)} R(r)$, then (3.1) becomes

$$-\tfrac{1}{2} R''(r) + [\tfrac{1}{2}\beta r^{-2} + V(r)] R(r) = E R(r), \tag{3.2}$$

where the parameter β is given by

$$\beta = [l + \tfrac{1}{2}(N-1)][l + \tfrac{1}{2}(N-3)]. \tag{3.3}$$

For large values of N it can be argued that the kinetic-energy term $-\frac{1}{2}R''(r)$ is dominated in magnitude by the term in β so that, in these cases, an approximation to the energy at the bottom of the angular-momentum subspace (labeled by l) is provided by

$$E_{0l} \simeq E(\beta) = \min_{r>0} [\tfrac{1}{2}\beta r^{-2} + V(r)]. \tag{3.4}$$

This is the large-N approximation that we consider in this paper. The literature[3] indicates a variety of modes of development after this initial step. The general goal is to obtain an expansion in terms of the variable $1/N$. To this end $V(r)$ is sometimes replaced by $V(9r/N^2)$ and analytical methods are used to estimate the coefficients in the resulting energy expansion. However, since we are no longer able to follow these various paths with our envelope bounds, we stop at the first step (3.4).

Since the envelope results and the large-N results follow from identical prescriptions, (2.11) and (3.4), and since, as we pointed out in Sec. II, the energies are monotonic in the kinetic-energy parameter, comparisons hinge on the relative values of α and β. The strongest statement we can make concerns the class of potentials that are convex transformations of the hydrogenic potential. We shall call these *convex-Coulombic* potentials and they are defined by

$$V(r) = g(-1/r), \tag{3.5}$$

where the function g is monotone increasing and convex on $(-\infty, 0)$. This family is quite broad for it includes the quark–quark potentials (2.17) along with any potential that is an increasing convex function of r. If we set $n = 0$ and compare α_1 in (2.12) with β in (3.3) we see that $\beta < \alpha_1$. Consequently we have the inequalities

$$E(\beta) < E(\alpha_1) \leqslant E_{0l}. \tag{3.6}$$

For the entire class of convex-Coulombic potentials, therefore, we know that the large-N approximation leads to a *lower bound* to the exact energy; this bound is weaker than the lower bound provided by Coulomb potential envelopes. Meanwhile both of the approximations are given by the same quasiclassical energy formula.

If the potential is *convex-harmonic* given by

$$V(r) = g(r^2), \tag{3.7}$$

where g is monotone increasing and convex, then by similar reasoning we find

$$E(\beta) < E(\alpha_1) < E(\alpha_2) \leqslant E_{0l}. \tag{3.8}$$

Thus, for more rapidly increasing potentials like quartic anharmonic oscillators $V(r) = Ar^2 + Br^4$ with A and B positive, the large-N approximation is pushed still further below the exact energy.

IV. POWER-LAW POTENTIALS

So far in this paper we have used power-law potential generators $h(r)$ but the potentials $V(r)$ themselves have been restricted only to the rather broad classes of transformations of the generators. Lieb[6] sought to find in classical terms both upper and lower bounds on the partition function of certain quantum spin systems. Although we are only dealing with the one-body problem in the present paper, we now try to bound the exact quantum mechanical energies for pure power-law potentials by the quasiclassical energy form that is common to three different approaches to the problem.

Throughout this section we suppose that the potential is given by the pure power law:

$$V(r) = \mathrm{sgn}(q) v r^q, \quad q \geqslant -1, \quad q \neq 0. \tag{4.1}$$

The generic energy form we are using in this paper, that is to say,

$$E(\alpha) = \min_{r>0} [\tfrac{1}{2}\alpha r^{-2} + V(r)], \tag{4.2}$$

now can be written explicitly as the function $E(\alpha)$, where

$$E(\alpha) = \alpha^{q/(q+2)} |qv|^{2/(q+2)} \{\tfrac{1}{2} + 1/q\}. \tag{4.3}$$

We have already obtained two general results that apply to this particular problem. First, since $V(r)$ is a convex function of the Coulomb potential $h(r) = -1/r$, we can use the envelope lower bound given by (2.11), that is,

$$E(\alpha) \leqslant E_{nl}, \quad \alpha = [l + n + \tfrac{1}{2}N - \tfrac{1}{2}]^2. \tag{4.4}$$

Second, we have the large-N approximation (3.4) for the bottom of the spectrum in each angular-momentum subspace ($n = 0$), that is to say,

$$E(\beta) \simeq E_{0l}, \quad \beta = [l + \tfrac{1}{2}(N-1)][l + \tfrac{1}{2}(N-3)]. \tag{4.5}$$

Thus, for the bottom of each angular momentum subspace, we have, by the ordering lemma, that $(\beta < \alpha) \Rightarrow E(\beta) < E(\alpha)$; also the ratio of $E(\alpha)$ to $E(\beta)$ approaches 1 for *either* of the limits $l \to \infty$ or $N \to \infty$. All we need to do now is to find an upper energy bound.

An upper bound for the bottom of the spectrum in each angular momentum subspace is conveniently found with the aid of a variational trial function of the form

$$\psi(r) = r^l \exp(-r^2/\sigma^2), \tag{4.6}$$

where σ is a variational parameter. If we apply $\psi(r)$ in the Rayleigh quotient $(\psi, H\psi)/(\psi, \psi)$ and then perform the optimization with respect to σ, we find that the result can eventually be reworked into the standard large-N form (4.2), (4.3), with

$$\begin{aligned} &E_{0l} \leqslant E(\gamma), \\ &\gamma = (r/\sigma)^2 \{l + \tfrac{1}{2}N\} \\ &\quad = [\Gamma(l + \tfrac{1}{2}N + \tfrac{1}{2}q)/\Gamma(l + \tfrac{1}{2}N)]^{2/q} \{l + \tfrac{1}{2}N\}, \end{aligned} \tag{4.7}$$

and where the r in (4.7) is the critical value arising from (4.2).

Now we can collect all our results for $n = 0$ into the following inequalities:

$$E(\text{large-}N) = E(\beta) < E(\alpha) \leqslant E_{0l} \leqslant E(\gamma), \tag{4.8}$$

where E_{0l} is the (unknown) exact energy. It is clear from the expressions (4.4), (4.5), and (4.7) for (α, β, γ) that in *either* limit $l \to \infty$ or $N \to \infty$, all three energy estimates approach the exact energy E_{0l}. This is the best we can do in general with these methods for the pure power-law problem in N spatial dimensions [the lower bound (4.4) applies to all the eigenvalues].

V. CONCLUSION

The main point of this article is to say something definite about the quality of the large-N approximation. That the method often leads to lower energy estimates already may have been observed numerically but it is useful to have definite information about this, even in a restricted environment. The neglect of the positive-definite term $-\tfrac{1}{2}R''(r)$ certainly suggests a lower estimate but this observation does not account for possible side effects introduced by the quasiclassical approximation; this reservation is particularly germane for finite N.

Our upper and lower bounds for the energies of the problem with pure power-law potentials determine the exact eigenvalues asymptotically for $n = 0$ and in the limits $l \to \infty$ or $N \to \infty$. These bounds and the "exact solutions" that they determine may provide useful test examples for large-N theory. The extension of some of these results to the problem of $\nu \geqslant 3$ identical particles interacting by pair potentials is straightforward if one applies the methods of Ref. 4.

ACKNOWLEDGMENTS

The author would like to thank the Department of Physics of the National Taiwan University in Taipei for its hospitality during a visit when this work was carried out. He would also like to thank the referee for suggesting the inclusion of a variational upper bound such as the one we have used for power-law potentials in Sec. IV.

Partial financial support of this work by the Natural Sciences and Engineering Research Council Canada (Grant No. A3438) is gratefully acknowledged.

[1]T. H. Berlin and M. Kac, Phys. Rev. **86**, 821 (1952); H. E. Stanley, *ibid.* **176**, 718 (1968).

[2]G. t'Hooft, Nucl. Phys. B **72**, 461 (1974); G. Rossi and G. Veneziano, *ibid.* **123**, 507 (1977); L. G. Yaffe, Rev. Mod. Phys. **54**, 407 (1982).

[3]L. D. Mlodinow and N. Papanicolaou, Ann. Phys. **128**, 314 (1980); L. D. Mlodinow, Prog. Part. Nucl. Phys. **8**, 387 (1982); G. Moreno and A. Zepeda, J. Phys. B **17**, 21 (1984).

[4]R. L. Hall, J. Math. Phys. **24**, 324 (1983).

[5]R. L. Hall, J. Math. Phys. **25**, 2708 (1984).

[6]E. H. Lieb, Commun. Math. Phys. **31**, 327 (1973).

Energy eigenvalues of *d*-dimensional quartic anharmonic oscillator

M. Seetharaman and S. S. Vasan
Department of Theoretical Physics, University of Madras, Guindy Campus, Madras 600 025, India

(Received 3 September 1985; accepted for publication 20 November 1985)

On the basis of a radial generalization of the JWKB quantization rule, which incorporates higher orders of the approximation, an explicit analytical formula is derived for the energy levels of the three-dimensional quartic anharmonic oscillator. The formula exhibits the scaling property of the exact eigenvalues, and is readily generalized to any dimension. Together with the Hellmann–Feynman theorem, it yields the values of the diagonal moments of r^{2k}. The predicted energies and moments are in excellent agreement with known numerical results.

I. INTRODUCTION

The JWKB approximation is a well-known nonperturbative technique for computing energy eigenvalues in quantum mechanical bound-state problems. In the lowest order of the approximation, one obtains the Bohr–Sommerfeld energy quantization rule (with n replaced by $n + \frac{1}{2}$). This standard textbook prescription is generally satisfactory insofar as one needs only a first estimate. In many cases of interest, the accuracy of the computed energy values can be significantly increased by including corrections to the Bohr–Sommerfeld rule that arise when the higher orders of the approximation are considered. In the context of one-dimensional problems, Dunham[1] showed long ago how this could be done. His elegant analysis resulted in a quantization rule that may be termed exact in the sense that it takes into account all orders of the approximation. Several authors[2–4] have applied Dunham's method to calculate energy eigenvalues for different potentials in one dimension. It has been found that, when sufficiently high orders are included, very high accuracies (1 in 10^{15}) are attainable. A nice feature of the Dunham formula is that it is expressed in terms of complex contour integrals over a closed curve. Consequently, all the higher-order JWKB integrals are well defined, and one has the freedom to simplify them considerably before actual evaluation.

The extension of the JWKB approximation to three-dimensional problems with spherical symmetry is not entirely straightforward, even in the lowest order. This fact has been known for a long time. The problem is compounded if one seeks to incorporate higher orders. The difficulties involved are traceable to the presence of the centrifugal barrier term in the Schrödinger equation, and the related requirement that the reduced radial wave function should vanish as r^{l+1} at $r = 0$. Any generalization of the JWKB method to radial problems must yield this correct behavior of the wave function near the origin. At least two approaches are known for extending Dunham's JWKB formalism to three dimensions that meet the above criterion. In the method of Beckel and Nakhleh,[5] which has been further studied by Fröman and Fröman,[6] and by ourselves,[7] one modifies the effective potential in the radial problem, treating the strength of the centrifugal barrier as an adjustable parameter. In the other method, developed by Krieger and Rosenzweig,[8] one performs a Langer transformation[9] by which the radial problem is turned into a truly one-dimensional eigenvalue problem, to which the original Dunham method is applied. The equivalence of the two methods has been recently discussed by us.[10] As in one-dimensional cases, when higher-order corrections are included in the JWKB treatment of radial eigenvalue problems, the accuracy of the computed energy values increases appreciably. We have demonstrated this in the case of the potential $V(r) = r^4$ by including corrections up to the fourth order.[10]

In the present work we apply the JWKB formalism of Krieger and Rosenzweig to the anharmonic oscillator (AHO) characterized by the potential $V(r) = \frac{1}{2}\alpha_1 r^2 + \alpha_2 r^4$, $\alpha_1,\alpha_2 > 0$. Our principal objective is to derive an explicit analytical formula for the energy eigenvalues of the AHO. We take into account the lowest, the second, and the fourth orders of the approximation. By a suitable expansion of the higher-order terms in the quantization condition, we show that it can be put in the form

$$n + \frac{3}{2} = W^{3/4} \sum_{k=0}^{8} C_k W^{-k/2} + O(W^{-15/4}),$$

where n is the principal quantum number and W is the energy. The coefficients C_k are functions of l, α_1, and α_2, which can be determined exactly. By inverting this series, we obtain an explicit formula for the energy levels. A by-product of the energy formula is that we can derive expressions for the diagonal moments $\langle r^{2k} \rangle$, with the use of the Hellmann–Feynman theorem.[11] The computed energy values and the moments are found to be excellent when compared with accurate numerical values for these computed by Bhargava.[12] As shown in the text, our formula could be trivially extended to the case of the AHO in d dimensions.

II. THE JWKB QUANTIZATION RULE

Consider a particle of unit mass moving in a central potential $V(r)$ that has a single minimum at a positive value of r and satisfies the condition $r^2 V(r) \to 0$ as $r \to 0$. The bound state energies of the particle may be determined from the JWKB quantization rule, which in the fourth order of the approximation can be written as

$$(2n_r + 1)\pi = J_0 + J_2 + J_4, \tag{1}$$

where

$$J_0 = \sqrt{2} \oint dr\, r^{-1} F^{1/2}, \tag{2}$$

 0022-2488/86/041031-06$02.50

with

$$F(r) = r^2[E - V(r)] - (l+\tfrac{1}{2})^2/2, \tag{3}$$

and

$$J_2 = -\frac{\sqrt{2}}{64}\oint dr\, r F^{-5/2}(F')^2, \tag{4}$$

$$J_4 = -\frac{\sqrt{2}}{8192}\oint dr\Big[49r^3F^{-11/2}(F')^4 - 16F^{-7/2}F'\Big(r\frac{d}{dr}\Big)^3F\Big]. \tag{5}$$

In the above F' stands for dF/dr, and J_0, J_2, and J_4 are the contributions coming from the lowest, the second, and the fourth orders of the JWKB approximation.[8] These JWKB integrals are contour integrals in the r plane over a closed curve, which encloses a cut along the real axis between $r = a$ and $r = b$, a and b being the classical turning points defined to be the positive real roots of $F(r) = 0$. The contour does not enclose any other singularity of the integrand. We take the contour to be traversed clockwise; it is then necessary to choose the branch of $F^{1/2}$ that is positive real on the upper lip of the cut.

By an obvious deformation of the contour, it is easy to check that J_0 reduces to

$$2\int_a^b dr\left[2(E-V) - \frac{(l+\frac{1}{2})^2}{r^2}\right]^{1/2},$$

which gives the lowest-order expression for the right-hand-side of (1). The expressions for the higher-order integrals in (4) and (5) are due to Krieger and Rosenzweig.[8] They applied the Langer transformation to the radial Schrödinger equation so that the one-dimensional analysis of Dunham could be used, without any modification, to derive higher-order corrections in radial problems. As is well known, the appearance of $(l+\frac{1}{2})^2$ in the place of $l(l+1)$ is a consequence of the Langer transformation.

We shall employ the above quantization rule to derive an expression for the energy levels of the quartic anharmonic oscillator, taking it to be defined by the potential

$$V(r) = \tfrac{1}{2}\alpha_1 r^2 + \alpha_2 r^4, \quad \alpha_1,\alpha_2 > 0. \tag{6}$$

For this potential, it is known that the exact Schrödinger eigenvalues satisfy the following scaling relation:

$$E(\alpha_1,\alpha_2) = \alpha_2^{1/3}E(\alpha_1\alpha_2^{-2/3},1). \tag{7}$$

This exact scaling relation is also obeyed by the JWKB energy values determined by the condition (1). To see this, consider the lowest-order approximation to (1). We have

$$\begin{aligned}(2n_r+1)\pi &= \sqrt{2}\oint dr\, r^{-1}\Big[r^2\Big(E - \frac{1}{2}\alpha_1 r^2 - \alpha_2 r^4\Big) - \frac{1}{2}\Big(l+\frac{1}{2}\Big)^2\Big]^{1/2}\\ &= \sqrt{2}\oint dy\, y^{-1}\Big[y^2\Big(E\alpha_2^{-1/3} - \frac{1}{2}\alpha_1\alpha_2^{-2/3}y^2 - y^4\Big) - \frac{1}{2}\Big(l+\frac{1}{2}\Big)^2\Big]^{1/2},\end{aligned}$$

the second step resulting from a change of scale of r. From this it follows that $E\alpha_2^{-1/3}$ is the energy corresponding to the potential $V(r) = \frac{1}{2}\alpha_1\alpha_2^{-2/3}r^2 + r^4$, which is precisely what is implied by the exact scaling relation (7). The relation (7) is found to hold even when J_2 and J_4 are included in (1). In fact, for polynomial potentials, it can be shown that the exact scaling relation is obeyed by the JWKB energy values in every order of the approximation.

For later convenience we introduce at this point the following new variable and parameters:

$$\begin{aligned}&W = E\alpha_2^{-1/3}, \quad \sigma = (l+\tfrac{1}{2})^2W^{-3/2}/2,\\ &\alpha = \tfrac{1}{8}\alpha_1\alpha_2^{-2/3}W^{-1/2}, \quad z = r^2W^{-1/3}\alpha_2^{1/3},\\ &H(z) = -z^3 + z - 4\alpha z^2 - \sigma.\end{aligned} \tag{8}$$

In terms of these, we get

$$\begin{aligned}&J_0 = 2^{-1/2}W^{3/4}\oint dz\, z^{-1}H^{1/2},\\ &J_2 = -\frac{W^{-3/4}}{48\sqrt{2}}\oint dz(1 - 24\alpha z - 15z^2)H^{-3/2},\\ &J_4 = J_{41} + J_{42},\end{aligned} \tag{9}$$

with

$$\begin{aligned}J_{41} &= W^{-9/4}\oint dz\Big[\frac{16}{21}(9z^5 + 24\alpha z^4 + 16\alpha^2z^3)H^{-7/2} - \frac{64}{35}(5z^2 + 4\alpha z)H^{-5/2}\Big],\\ J_{42} &= W^{-9/4}\oint dz(27z^3 + 32\alpha z^2 - z)\times(-3z^2 - 8\alpha z + 1)H^{-7/2}.\end{aligned} \tag{10}$$

In obtaining the above forms for J_2 and J_4 from (4) and (5), some integrations by parts have been carried out, the integrated terms vanishing because the contour is a closed curve. We note that the W dependence of the J_k is only through α and σ, apart from the overall explicit powers of W. We also note that, to first order in α and σ, the relevant branch points are located at $z = \sigma$ and $z = 1 - \alpha/2$.

III. EXPANSION OF J_k IN POWERS OF α AND σ

In the case of the quartic AHO all the JWKB integrals J_k can be evaluated in terms of complete elliptic integrals of the three kinds. (One such evaluation of J_0 can be found in Ref. 13.) One may then numerically solve for the energies, as was done by Kesarvani and Varshni[3] for the one-dimensional quartic AHO. Since our present aim is to derive an analytical formula for the energy levels, we proceed in a different way. Instead of expressing the J_k in terms of complete elliptic integrals whose arguments are complicated functions of W, we obtain directly a series representation of the rhs of (1) in powers of W. The rationale of this procedure is the following. Generally, the JWKB method is expected to work well for states characterized by large values of the relevant quantum numbers. In particular, the accuracy of the energy values determined by (1) increases with n_r. In the present case, large values of n_r mean large values of W. When W is large, the parameters α and σ are small, for fixed l, α_1, and α_2. It is then a natural step to expand J_k in powers of α and σ, and obtain a series representation of the rhs of (1). This series may then be inverted to get a relation for W in terms of the quantum numbers n_r, l, and the parameter α. An additional motivation for adopting the method described is that a similar treatment in the case of the potential $V(r) = r^{2m}$ has yielded excellent results.[14]

It turns out that the leading term in J_k is $\sim W^{-3(k-1)/4}$. Since in the quantization condition (1) integrals J_k with $k > 6$ have been neglected, consistency requires that terms of order $W^{-15/4}$ be dropped from the series expansions of J_0, J_2, and J_4.

A. Series for J_0

Consider first J_0. A careful inspection shows that the integral in it can be expanded as a series in α and $\sqrt{\sigma}$. It turns out, however, that all the terms involving odd powers of $\sqrt{\sigma}$ higher than the first vanish. Hence we can write

$$J_0 = \frac{W^{3/4}}{\sqrt{2}} \left[-2\pi\sqrt{\sigma} + \sum_{k=0}^{8} a_k \alpha^k + \sigma \sum_{k=0}^{5} b_k \alpha^k + \sigma^2 \sum_{k=0}^{2} c_k \alpha^k \right] + O(W^{-15/4}). \tag{11}$$

The coefficients are determined by differentiating J_0 suitably, and then setting $\alpha = \sigma = 0$. In the process one needs to make use of the following result: if m, n are odd integers with $m > -1$, then

$$\oint dz\, z^{m/2}(1 - z^2)^{n/2} = B\left(\frac{1}{2} + \frac{m}{4}, 1 + \frac{n}{2}\right), \tag{12}$$

where the contour surrounds a cut from $z = 0$ to $z = 1$. In (12), B is the beta function defined by

$$B(x,y) = \Gamma(x)\Gamma(y)/\Gamma(x + y). \tag{13}$$

The result (12) is derived in the Appendix. Using the properties of the gamma function, one can express the rhs of (12) as a function of m and n multiplied by either $B(\frac{1}{4},\frac{1}{2})$ or $B(\frac{3}{4},\frac{1}{2})$. As an illustration of the entire process, we show how to evaluate a_4:

$$\begin{aligned} a_4 &\equiv \frac{1}{4!} \left(\frac{\partial^4}{\partial \alpha^4} J_0 \sqrt{2} W^{-3/4} \right)_{\alpha = \sigma = 0} \\ &= \frac{256}{24} \left(-\frac{15}{16} \right) \oint dz\, z^7 (-z^3 + z)^{-7/2} \\ &= -10 \oint dz\, z^{7/2} (1 - z^2)^{-7/2} \\ &= -10\, B(\tfrac{9}{4}, -\tfrac{5}{2}) = -\tfrac{5}{12} B(\tfrac{1}{4},\tfrac{1}{2}). \end{aligned}$$

The values of the coefficients in (11) calculated as above are given below:

$$\begin{aligned} &a_0 = \tfrac{2}{3} B(\tfrac{1}{4}, \tfrac{1}{2}) \equiv \tfrac{2}{3} B_1, \\ &a_1 = -2B(\tfrac{3}{4}, \tfrac{1}{2}) \equiv -2B_3, \\ &a_2 = B_1, \quad a_3 = -B_3, \quad a_4 = -\tfrac{5}{12} B_1, \quad a_5 = \tfrac{21}{20} B_3, \\ &a_6 = \tfrac{5}{8} B_1, \quad a_7 = -\tfrac{77}{40} B_3, \quad a_8 = -\tfrac{585}{448} B_1; \\ &b_0 = \tfrac{1}{2} B_3, \quad b_1 = -\tfrac{1}{2} B_1, \quad b_2 = \tfrac{3}{4} B_3, \quad b_3 = \tfrac{5}{12} B_1, \\ &b_4 = -\tfrac{21}{16} B_3, \quad b_5 = -\tfrac{15}{16} B_1, \\ &c_0 = -\tfrac{5}{48} B_1, \quad c_1 = \tfrac{21}{16} B_3, \quad c_2 = -\tfrac{45}{32} B_1. \end{aligned} \tag{14}$$

B. Series for J_2

The expansion for J_2 is somewhat simpler, since we need to keep terms only up to $\sigma\alpha^2$, owing to the fact that J_2 has $W^{-3/4}$ as an overall factor. An inspection of J_2 shows that the expansion is free of all odd powers of $\sqrt{\sigma}$. Explicitly we have

$$J_2 = -\frac{1}{48\sqrt{2}} W^{-3/4} \left[\sum_{k=0}^{5} d_k \alpha^k + \sigma \sum_{k=0}^{2} e_k \alpha^k \right] + O(W^{-15/4}). \tag{15}$$

The coefficients are evaluated as in the previous case, and their values are given below:

$$\begin{aligned} &d_0 = 6 B_3, \quad d_1 = -2 B_1, \quad d_2 = -3 B_3, \quad d_3 = -5 B_1, \\ &d_4 = \tfrac{105}{4} B_3, \quad d_5 = \tfrac{105}{4} B_1, \\ &e_0 = -\tfrac{15}{2} B_1, \quad e_1 = \tfrac{147}{2} B_3, \quad e_2 = -225 B_1. \end{aligned} \tag{16}$$

C. Series for J_4

The expansion for J_4 is similar to that of J_2. Due to the overall factor $W^{-9/4}$ in J_4, we need to keep terms only up to α^2. Referring to (9) and (10), we can expand J_{41} and J_{42} as

$$J_{4i} = W^{-9/4} \sum_{k=0}^{2} f_{ki}\, \alpha^k + O(W^{-15/4}), \quad i = 1,2. \tag{17}$$

On evaluation, the values of f_{ki} are

$$\begin{aligned} &f_{01} = -\tfrac{86}{21} B_1, \quad f_{11} = \tfrac{146}{5} B_3, \quad f_{21} = -\tfrac{43}{7} B_1, \\ &f_{02} = 13 B_1, \quad f_{12} = -\tfrac{469}{5} B_3, \quad f_{22} = \tfrac{51}{2} B_1. \end{aligned} \tag{18}$$

Combining J_{41} and J_{42}, we have

$$J_4 = -\frac{\sqrt{2}}{1024} W^{-9/4} \sum_{k=0}^{2} f_k \alpha^k + O(W^{-15/4}), \tag{19}$$

with

$$f_0 = \tfrac{22}{3} B_1, \quad f_1 = -70 B_3, \quad f_2 = 107 B_1. \tag{20}$$

IV. ANALYTICAL FORMULA FOR THE ENERGY LEVELS

The series expansions derived above for the J_k's enable us to write the quantization condition (1) in the form

$$\begin{aligned} &\pi\sqrt{2}(2n_r + 1 + l + \tfrac{1}{2}) \\ &\quad \equiv \pi\sqrt{2}(n + \tfrac{3}{2}) \\ &\quad = W^{3/4} \sum_{k=0}^{8} g_k W^{-k/2} + O(W^{-15/4}), \end{aligned} \tag{21}$$

where $n = 2n_r + l$ is the principal quantum number. The coefficients g_k are functions of l, α_1, and α_2, and are obtained by substituting for α and σ in terms of W, l, α_1, and α_2, and regrouping terms according to powers of W. With the definitions

$$\beta = \alpha_1 \alpha_2^{-2/3}, \quad L^2 = (l + \tfrac{1}{2})^2, \tag{22}$$

the explicit expressions for the g_k's are as follows:

$$
\begin{aligned}
g_0 &= \tfrac{2}{3} B_1, \quad g_1 = -B_3\beta, \quad g_2 = \tfrac{1}{4} B_1 \beta^2, \\
g_3 &= \tfrac{1}{8}(-1-\beta^3+2L^2)B_3, \\
g_4 &= \tfrac{1}{48} B_1 \beta(1-\tfrac{5}{4}\beta^3-6L^2), \\
g_5 &= \tfrac{1}{64} B_3 \beta^2(1+\tfrac{21}{10}\beta^3+6L^2), \\
g_6 &= \tfrac{1}{192} B_1 [-\tfrac{11}{4}+\tfrac{5}{2}\beta^3+\tfrac{15}{8}\beta^6 \\
&\quad + (15+5\beta^3)L^2-5L^4], \qquad (23) \\
g_7 &= \tfrac{1}{512} B_3\beta [140-\tfrac{35}{2}\beta^3-\tfrac{77}{10}\beta^6 \\
&\quad - (196+21\beta^3)L^2+84L^4], \\
g_8 &= \tfrac{1}{2048} B_1 \beta^2[-107-35\beta^3-\tfrac{585}{56}\beta^6 \\
&\quad + (300-30\beta^3)L^2-180L^4].
\end{aligned}
$$

It may be noted that inclusion of J_6 and higher-order integrals in (1) will not affect the values of the g_k given above, and therefore the expressions in (23) are exact to all orders.

The relation (21), with the g_k given by (22) and (23), is the basis for obtaining an analytical formula for the energy values. If the leading term alone on the rhs of (21) is retained, the expression for W is immediately seen to be

$$W = [(3\pi/B_1\sqrt{2})(n+\tfrac{3}{2})]^{4/3}.$$

The inclusion of other terms results in corrections to this leading approximation. An inspection shows that in the present case we may write

$$W(\beta) = N^{4/3} \sum_{k=0}^{8} G_k(\beta) N^{-2k/3} + O(N^{-14/3}), \qquad (24)$$

where

$$N = [9\pi\gamma/8]^{1/2}(n+\tfrac{3}{2}) \qquad (25)$$

with

$$\gamma = [2\Gamma(\tfrac{3}{4})/\Gamma(\tfrac{1}{4})]^2 = 0.456\,9466. \qquad (26)$$

Substitution of (24) into (21) yields the G_k in terms of β and γ. A long and tedious calculation leads to the following explicit expressions:

$$
\begin{aligned}
G_0 &= 1, \quad G_1 = 2\beta\gamma, \quad G_2 = -\tfrac{1}{2}\beta^2(1-3\gamma^2), \\
G_3 &= \tfrac{1}{24}\gamma[3+8\beta^3\gamma^2-12l(l+1)], \\
G_4 &= \tfrac{1}{48}\beta[1-3\gamma^2+\beta^3(1-5\gamma^4)+12(1+\gamma^2)l(l+1)], \\
G_5 &= -\tfrac{1}{80}\beta^2\gamma[5+4\beta^3+60l(l+1)], \\
G_6 &= \tfrac{1}{192}[-\tfrac{11}{8}-\tfrac{15}{8}\gamma^2+5\beta^3(-1+4\gamma^2+\gamma^4) \\
&\quad + \tfrac{14}{3}\beta^6\gamma^2(3+\gamma^4) + [-25+15\gamma^2 \\
&\quad + \beta^3(20+240\gamma^2-20\gamma^4)]l(l+1) \\
&\quad + (10-30\gamma^2)l^2(l+1)^2], \qquad (27) \\
G_7 &= \tfrac{7}{192}\beta\gamma[-\tfrac{39}{4}+\tfrac{3}{4}\gamma^2+\beta^3(\tfrac{7}{2}-\tfrac{10}{3}\gamma^2-\tfrac{1}{2}\gamma^4) \\
&\quad - \beta^6\gamma^2(\tfrac{32}{15}+\tfrac{8}{21}\gamma^4) \\
&\quad + [28-6\gamma^2+\beta^3(-14-40\gamma^2+2\gamma^4)]l(l+1) \\
&\quad + (-8+12\gamma^2)l^2(l+1)^2], \\
G_8 &= \tfrac{9}{256}\beta^2[\tfrac{23}{24}+\tfrac{95}{2}\gamma^2-\tfrac{9}{8}\gamma^4+\beta^3(\tfrac{1}{9}-\tfrac{45}{5}\gamma^2+3\gamma^4) \\
&\quad + \beta^6(\tfrac{1}{63}+\tfrac{9}{5}\gamma^4) + [-\tfrac{31}{3}-96\gamma^2+9\gamma^4 \\
&\quad + \beta^3(\tfrac{4}{3}+\tfrac{192}{5}\gamma^2+36\gamma^4)]l(l+1) \\
&\quad + (\tfrac{14}{3}+16\gamma^2-18\gamma^4)l^2(l+1)^2].
\end{aligned}
$$

The relation (24), together with (22) and (25)–(27), provides an explicit formula for the energy eigenvalues (labeled by n and l) of the Hamiltonian $H = \frac{1}{2}\mathbf{p}^2 + \frac{1}{2}\alpha_1 r^2 + \alpha_2 r^4$ through the relation

$$E(\alpha_1,\alpha_2) = \alpha_2^{1/3} W(\beta). \qquad (28)$$

If we set $\alpha_1 = 0$, $\alpha_2 = 1$, we obtain the energies of the pure quartic oscillator, and these are in agreement with our earlier results.[10]

A formula for the one-dimensional AHO can be obtained in the following way. Noting first that the $l = 0$ levels of the three-dimensional problem correspond to the odd-parity level of the one-dimensional case, we can get the latter by setting $l = 0$ and replacing $n + \frac{3}{2}$ by $n' + \frac{1}{2}$, where n' is an odd integer. Since the coefficient G_k are independent of n, the same formula holds also for even values of n', thus yielding all the levels of the one-dimensional AHO.

For a d-dimensional AHO ($d > 1$), the energy levels are again easily obtained by making the replacements

$$l \to l + \tfrac{1}{2}(d-3), \quad n + \tfrac{3}{2} \to n + d/2,$$

the second being a consequence of the first. The above replacement of l follows from the fact that the radial Schrödinger equation in d dimensions has the form[15]

$$\psi'' + 2\left[E - V(r) - \frac{1}{2r^2}\left(l + \frac{d-3}{2}\right)\left(l + \frac{d-1}{2}\right)\right]\psi = 0.$$

V. DIAGONAL MOMENTS OF r^{2k}

A fallout of our analytical formula for the energy levels is that the moments $\langle r^{2k} \rangle$ in states of the AHO can be calculated, using the Hellmann–Feynman (HF) theorem and the hypervirial relations.[16] Consider the Hamiltonian $H = \frac{1}{2}\mathbf{p}^2 + \frac{1}{2}\beta r^2 + r^4$ with $H\psi = W(\beta)\psi$. The values of W (in the JWKB approximation) are given by (24). It follows from the HF theorem that

TABLE I. Energy values of the three-dimensional AHO with $H = \frac{1}{2}(\mathbf{p}^2 + r^2 + r^4)$.

n	l	W_0	W_2	W_4	W_{exact}
0	0	2.295 160	2.325 816	2.322 300	2.324 406
1	1	4.305 746	4.192 130	4.189 304	4.190 172
2	0	6.560 951	6.578 486	6.578 126	6.578 400
2	2	6.560 951	6.241 279	6.242 469	6.242 778
5	1	14.406 097	14.339 079	14.338 825	14.338 860
5	3	14.406 097	13.946 033	13.949 236	13.949 200
5	5	14.406 097	13.238 551	13.266 575	13.264 450
10	0	30.056 858	30.064 766	30.064 755	30.064 761
10	4	30.056 858	29.507 297	29.510 048	29.510 012
10	10	30.056 858	26.998 687	27.103 115	27.092 490
50	0	213.988 009	213.990 907	213.990 907	213.990 906
50	20	213.988 009	209.448 561	209.483 137	209.482 265
50	50	213.988 009	186.412 380	187.698 966	187.529 7

TABLE II. Energy values of the two-dimensional AHO with $H=\frac{1}{2}(\mathbf{p}^2+r^2+r^4)$.

n	l	W_0	W_2	W_4	$W_{\rm exact}$
0	0	1.406 913	1.474 576	1.473 891	1.476 025
1	1	3.265 776	3.233 766	3.230 118	3.231 453
2	0	5.406 277	5.441 025	5.440 763	5.441 218
2	2	5.406 277	5.195 942	5.194 749	5.195 314
5	1	13.003 409	12.984 736	12.984 530	12.984 581
5	3	13.003 409	12.655 105	12.656 652	12.656 690
5	5	13.003 409	11.995 842	12.018 144	12.016 583
10	0	28.374 614	28.389 922	28.389 922	28.389 930
10	4	28.374 614	27.931 652	27.933 395	27.933 383
10	10	28.374 614	25.525 732	25.621 211	25.611 650
50	0	211.252 237	211.257 875	211.257 875	211.257 880
50	20	211.252 237	206.904 498	206.936 604	206.935 820
50	50	211.252 237	184.049 270	185.317 269	185.150 586

$$\langle r^2\rangle = 2\frac{\partial W}{\partial\beta}. \tag{29}$$

This relation, together with the scaling law (28), yields

$$\langle r^4\rangle = \tfrac{1}{3}(W-\beta\langle r^2\rangle). \tag{30}$$

Still higher diagonal moments can be calculated using the hypervirial relations.[16] In the limit $\beta\to 0$, (29) and (30) give the moments $\langle r^2\rangle$ and $\langle r^4\rangle$ with respect to the pure quartic oscillator states.

VI. RESULTS AND DISCUSSION

We present in Tables I–III the predictions of our formula for the AHO energy levels. In these tables, W_0 is the lowest-order result obtained by neglecting all the terms beyond G_2 in (24); W_2 is the value incorporating the second-order corrections G_3 to G_5, while W_4 corresponds to keeping all the coefficients up to G_8 in (24); and $W_{\rm exact}$ denotes the set of highly accurate numerical eigenvalues computed by Bhargava.[12] It is evident from our results that the inclusion of higher orders has the effect of increasing significantly the accuracy of our formula. One would expect that, for a given n, the accuracy would decrease with increasing l, since σ, one of the expansion parameters, increases with l. On the contrary, the results are found to be good even in the case of levels with $n=l$. The maximum error occurs in the ground state values. Even here the error decreases as the dimension increases. In Table IV we give the values of $\langle r^2\rangle$ for the three-dimensional AHO. These also are seen to be in very good agreement with the numerical results of Bhargava.[12] To our knowledge, this is the first time explicit analytical formulas have been given for the energies and the diagonal moments of the quartic AHO in any dimension.

TABLE III. Energy values of the one-dimensional AHO with $H=\frac{1}{2}(p^2+x^2+x^4)$.

n	W_0	W_2	W_4	$W_{\rm exact}$
0	0.622 978	0.682 242	0.722 183	0.696 176
1	2.295 160	2.325 816	2.322 300	2.324 406
2	4.305 746	4.327 675	4.326 749	4.327 525
5	11.635 785	11.648 749	11.648 652	11.648 721
10	26.716 151	26.724 557	26.724 542	26.724 551
50	208.525 195	208.528 132	208.528 132	208.528 132

TABLE IV. Values of $\langle r^2\rangle$ for the three-dimensional AHO.

n	l	$\langle r^2\rangle$	$\langle r^2\rangle_{\rm exact}$
0	0	0.800 161	0.801 251
1	1	1.212 315	1.214 551
4	0	2.013 992	2.014 468
5	1	2.280 292	2.280 362
8	0	2.938 350	2.938 362
9	1	3.157 616	3.151 628
49	1	9.137 001	9.137 001
50	0	9.257 019	9.257 019

We conclude this work with some general observations. The JWKB series is known to be an asymptotic series. Hence, the accuracy of the JWKB energies cannot be increased arbitrarily by including corrections from higher and higher orders. There appears to be no general criterion that will determine how far in the series one should go, in order to get the best results for a given set of quantum numbers and parameters. Our quartic AHO results show that, for n values that are not too small, the accuracy keeps increasing with order. For low values of n, the second order seems to yield a better result than the fourth, which is an indication that the series is not a convergent series. Similar results have been reported earlier, in the context of the one-dimensional AHO.[2] In general, as n increases, the accuracy will increase with the inclusion of higher and higher orders, and one can even get in this way an accuracy of 1 in 10^{15}. However, given the algebraic complexity of the sixth- and higher-order integrals, it does not appear to be worthwhile to continue along the lines of the present work beyond the fourth order. If one wishes to obtain greater accuracy by including still higher orders, it is best to resort to numerical computation.

Finally, it may be noted that our formula does not yield the correct harmonic oscillator energy levels. This does not occasion any surprise, for the expansion of the JWKB integrals is about the pure quartic oscillator levels (which correspond to $\alpha=0$). The harmonic oscillator results can of course be recovered from the integrals directly (in the limit $\alpha\to\infty$), before they are expanded in powers of α.

APPENDIX: PROOF OF FORMULA (12)

We give below a proof of the formula

$$\oint dz\, z^{m/2}(1-z^2)^{n/2} = B\left(\frac{1}{2}+\frac{m}{4},1+\frac{n}{2}\right) \tag{A1}$$

quoted in the text. In this m, n are odd integers with $m \geqslant -1$. The contour of integration encloses a branch cut along the real axis between the points $z=0$ and $z=1$.

Consider first the case when $n \geqslant -1$. In this case the integrand in (A1) has only integrable singularities at $z=0$ and $z=1$. Therefore the contour can be deformed until it consists of two straight line segments lying just above and below the cut, and two small circular arcs of radius ϵ around $z=0$ and $z=1$. In the limit $\epsilon\to 0$ the contributions from the circular arcs vanish, and the above integral reduces to

$$2\int_0^1 dx\, x^{m/2}(1-x^2)^{n/2}.$$

By the change of variable $y = x^2$, this integral can be seen to be equal to the beta function given in (A1).

It remains to prove (A1) for the case $n < -1$. Let $n = -(2k+1)$, $k = 1,2,3,\dots$. To evaluate the integral, which is well defined as a contour integral, we consider

$$I(m,k,a) = \oint dz\, z^{m/2}(a-z^2)^{-k-1/2}$$

and take the required integral to be $I(m,k,1)$. We can then write

$$I(m,k,a) = C_k \frac{\partial^k}{\partial a^k} \oint dz\, z^{m/2}(a-z^2)^{-1/2},$$

$$C_k = \frac{(-2)^k}{(2k-1)!!}.$$

The a dependence of the integrand can be factored out by the change of variable $y = a^{-1/2}z$. This yields

$$I(m,k,a) = \frac{1}{2} C_k \left[\frac{\partial^k}{\partial a^k} a^{m/4}\right] \oint dy\, y^{m/2}(1-y^2)^{-1/2}.$$

We observe that this integral is one with $n = -1$. Hence it can be evaluated as before. Carrying out the differentiations and then setting $a = 1$, we get

$$\begin{aligned} I(m,k,1) &= C_k \frac{m}{4}\left(\frac{m}{4}-2\right)\cdots\left(\frac{m}{4}-k+1\right) \\ &\quad \times B\left(\frac{1}{2}+\frac{m}{4},\frac{1}{2}\right) \\ &= B\left(\frac{1}{2}+\frac{m}{4},\frac{1}{2}-k\right) \\ &= B\left(\frac{1}{2}+\frac{m}{4},1+\frac{n}{2}\right), \end{aligned}$$

the last step resulting from the use of the identity

$$B(x,y) = [(x+y)/y]B(x,y+1)$$

k times.

[1]J. L. Dunham, Phys. Rev. **41**, 713 (1932).
[2]C. M. Bender, K. Olaussen, and P. S. Wang, Phys. Rev. D **16**, 1740 (1977).
[3]R. N. Kesarwani and Y. P. Varshni, J. Math. Phys. **22**, 1983 (1981); **23**, 803 (1982).
[4]M. Seetharaman, S. S. Vasan, and S. Ravichandran, Pramāna **24**, 695 (1985).
[5]C. Beckel and J. Nakhleh, J. Chem. Phys. **39**, 94 (1963).
[6]N. Fröman and P. O. Fröman, Nuovo Cimento B **20**, 121 (1974).
[7]M. Seetharaman and S. S. Vasan, J. Phys. A **17**, 2485 (1984).
[8]J. B. Krieger and C. Rosenzweig, Phys. Rev. **164**, 171 (1967).
[9]R. E. Langer, Phys. Rev. **51**, 669 (1937).
[10]S. S. Vasan and M. Seetharaman, J. Phys. A **17**, 2493 (1984).
[11]See, for instance, C. Quigg and J. L. Rosner, Phys. Rep. C **56**, 167 (1979).
[12]V. T. A. Bhargava, Ph. D. thesis, University of Madras, 1982 (unpublished).
[13]M. Seetharaman, S. Raghavan, and S. S. Vasan, J. Phys. A **15**, 1537 (1982).
[14]M. Seetharaman and S. S. Vasan, J. Phys. A **18**, 1041 (1985).
[15]F. T. Hioe, J. Chem. Phys. **69**, 204 (1978).
[16]D. E. Hughes, J. Phys. B **10**, 3167 (1977).

Massless fields in curved space-time: The conformal formalism

Mario A. Castagnino
Instituto de Astronomía y Física del Espacio, Casilla de Cerreo 67, Suc. 28, Ciudad Universitaria, 1428 Buenos Aires, Argentina and Instituto de Física de Rosario, CONICET–U.N.R., Av. Pellegrini 250, 2000 Rosario, Argentina

Jorge B. Sztrajman
Instituto de Astronomía y Física del Espacio, Casilla de Correo 67, Suc. 28, Ciudad Universitaria, 1428 Buenos Aires, Argentina

(Received 22 November 1983; accepted for publication 31 October 1985)

A conformally invariant theory for massless quantum fields in curved space-time is formulated. We analyze the cases of spin-0, -$\frac{1}{2}$, and -1. The theory is developed in the important case of an "expanding universe," generalizing the particle model of "conformal transplantation" known for spin-0 to spins-$\frac{1}{2}$ and -1. For the spin-1 case two methods introducing new conformally invariant gauge conditions are stated, and a problem of inconsistency that was stated for spin-1 is overcome.

I. INTRODUCTION

Quantum field theories in curved space-time have been studied intensely in the last years.[1] In the Robertson–Walker universe, and for spin-0 massless particles, the solutions, those we call "conformal transplantation," have been widely used, and all authors consider their vacuum as very satisfactory. However, this vacuum has not been extended to higher spins (although other features of those spins are well known.[2–5]

In this work, we shall see that the extension of this vacuum to spin-$\frac{1}{2}$ is done without difficulty, but the extension to spin-1 is more involved and it requires at least the introduction of new kinds of gauge.

In the frame of quantum field theory in curved space-time is a generally accepted postulate that classical formalism of the massless fields must be invariant when we perform a "conformal transformation." Such a transformation is a local change of the metric scale of space-time simultaneous with an appropriate change of the fields scale. The main arguments for these invariances are that (i) the classic trajectories of massless particles are not disturbed by conformal transformations; and (ii) the inexistence of a mass in the field equation deprives it of any scale (in the system of natural units where $c = \hbar = 1$).

In this work we try to build a consistent quantum field theory for massless fields in curved space-time, preserving the conformal invariance in order to overcome the inconsistencies pointed out in Ref. 3. The study is stated in the frame of general relativity, introducing the gravitation as an unquantized field by means of the space-time geometry.

In Sec. II we introduce the conformal transformation and we obtain the transformation laws of the characteristic geometric objects of a Riemannian manifold. Moreover, we obtain the transformation laws for (tensor and spinor) fields in order to use them later.

In Sec. III, and in the frame of a Lagrangian formulation, we prove the conformal covariance of Euler–Lagrange equations for Weyl fields.

In Sec. IV, and as an example, we analyze the well-known case of the massless scalar field, pointing out the main characteristics of a conformal theory.

Section V is devoted to the neutrino field; it is considered as a massless Dirac field. Principal objects are shown (Dirac's matrices, spinors, and spinor connection), analyzing in particular their behavior by conformal transformations and proving the conformal covariance of the theory.

In Sec. VI A we point out the problems for the canonical quantization of the electromagnetic field in Minkowski's space and we show two ways to solve these problems: (A) the quantization in the temporal gauge and (B) the Gupta–Bleuler quantization method. In Sec. VI B and in order to obtain a conformal generalization of both methods to curved space-time and to overcome the inconsistency just mentioned, we introduce the timelike vector field U^μ, which we believe could be a field of observers covering all space. In Sec. VI C we show the field equations for the cases (A) and (B). Section VI D is devoted to the construction of a "conformal derivative" for the four-potential A_μ in order to use it in (B). In Sec. VI E we introduce suitable inner products in order to orthonormalize the solutions of field equations. In Sec. VI F we introduce conformally invariant gauge conditions in order to eliminate the unphysical photons.

In Sec. VII we point out the advantages of using a conformal theory to treat massless fields ("conformal transplantation").

In Sec. VIII we develop the previous theories in the important case of an expanding universe, obtaining the orthonormal basis of solutions of the field equations for spins-0, -$\frac{1}{2}$, and -1. For spin-1, we show that both theories (A) and (B) yield the same particle model overcoming the inconsistency mentioned before. The explanation is that we use a conformal theory, i.e., the theory where the inconsistency was found, not conformal.

II. CONFORMAL TRANSFORMATIONS

Let V_4 be an arbitrary Riemannian space-time and let $g_{\mu\nu}$ be its metric tensor, with principal diagonal $(1, -1, -1, -1)$ in its diagonal form. We suppose V_4 to be of C^∞ class. Greek indices run from 0 to 3 whereas Latin

 0022-2488/86/041037-11$02.50

indices run from 1 to 3. A conformal transformation in V_4 is a local change of scale defined by

$$g_{\mu\nu}\to\tilde{g}_{\mu\nu}=\lambda(x)g_{\mu\nu}, \quad (2.1)$$

where $\lambda(x)$ is positive defined arbitrary function that characterizes the transformation. In order to satisfy the property

$$\tilde{g}_{\mu\rho}\,\tilde{g}^{\rho\nu}=g^{\nu}_{\mu}, \quad (2.2)$$

we take the contravariant components of a metric tensor:

$$g^{\mu\nu}\to\tilde{g}^{\mu\nu}=\lambda^{-1}(x)g^{\mu\nu}. \quad (2.3)$$

From transformation laws (2.1) and (2.3) we can obtain the transformation laws for all geometric objects of V_4. So, for the Riemannian connection $\Gamma^{\mu}_{\nu\rho}$, defined by

$$\Gamma^{\mu}_{\nu\rho}=\tfrac{1}{2}g^{\mu\sigma}(\partial_\nu g_{\sigma\rho}+\partial_\rho g_{\sigma\nu}-\partial_\sigma g_{\nu\rho}), \quad (2.4)$$

we obtain

$$\tilde{\Gamma}^{\mu}_{\nu\rho}=\Gamma^{\mu}_{\nu\rho}+\tfrac{1}{2}(\delta^{\mu}_{\nu}\,\partial\rho\ln\lambda+\delta^{\mu}_{\rho}\,\partial\nu\ln\lambda-g_{\nu\rho}\,\partial^{\mu}\ln\lambda), \quad (2.5)$$

$$\tilde{\Gamma}^{\mu}_{\mu\rho}=\Gamma^{\mu}_{\mu\rho}+2\,\partial\rho\ln\lambda. \quad (2.6)$$

For the Ricci's tensor $R_{\mu\nu}$ defined by

$$R_{\mu\nu}=\partial\rho\Gamma^{\rho}_{\mu\nu}-\partial\nu\Gamma^{\rho}_{\rho\mu}+\Gamma^{\rho}_{\rho\sigma}\Gamma^{\sigma}_{\mu\nu}-\Gamma^{\rho}_{\sigma\nu}\Gamma^{\sigma}_{\mu\rho}, \quad (2.7)$$

we obtain

$$\tilde{R}_{\mu\nu}=R_{\mu\nu}+\tfrac{1}{2}\nabla_\mu\nabla_\nu\ln\lambda+\tfrac{1}{2}(g_{\mu\rho}\,g_{\sigma\nu}-g_{\mu\nu}\,g_{\sigma\rho})\times(\nabla^{\rho}\ln\lambda\nabla^{\sigma}\ln\lambda-\nabla^{\rho}\nabla^{\sigma}\ln\lambda). \quad (2.8)$$

From (2.8), by contraction, we obtain for the curvature scalar $R=R^{\mu}\mu$,

$$\tilde{R}=\lambda^{-1}(R-3\nabla^{\mu}\nabla_\mu\ln\lambda-\tfrac{3}{2}\nabla^{\mu}\ln\lambda\nabla_\mu\ln\lambda). \quad (2.9)$$

For the differentials of the coordinates, we have

$$\widetilde{dx}^{\mu}=dx^{\mu}, \quad (2.10)$$

$$\widetilde{dx}_\mu=\lambda\,dx_\mu. \quad (2.11)$$

The volume element $d\eta$ and the surface element $d\sigma_\mu$ of V_4 are defined by

$$d\eta=\sqrt{-g}d^4x=\sqrt{-g}(1/4!)\varepsilon_{\mu\nu\rho\sigma}\,dx^{\mu}\,dx^{\nu}\,dx^{\rho}\,dx^{\sigma}, \quad (2.12)$$

$$d\sigma_\mu=\sqrt{-g}dS_\mu=\sqrt{-g}(1/3!)\varepsilon_{\mu\nu\rho\sigma}\,dx^{\nu}\,dx^{\rho}\,dx^{\sigma}, \quad (2.13)$$

where $g=\det(g_{\mu\nu})$ and $\varepsilon_{\mu\nu\rho\sigma}$ is the Levi-Civita symbol.

Then, we shall have

$$\widetilde{d\eta}=\lambda^2\,d\eta, \quad (2.14)$$

$$\widetilde{d\sigma}_\mu=\lambda^2\,d\sigma_\mu. \quad (2.15)$$

Since we shall consider tensor and spinor fields in V_4, we must study their behavior under conformal transformations. A tensor (spinor) ψ will be called a Weyl tensor (spinor) of weight n if its covariant components transform under (2.1) as

$$\tilde{\psi}=\lambda^{n}\psi. \quad (2.16)$$

Examples are the interval $ds=(dx^{\mu}\,dx_\mu)^{1/2}$, which is a Weyl scalar of weight $\frac{1}{2}$, $d\sigma_\mu$ (Weyl vector of weight 2), and $g_{\mu\nu}$ (Weyl tensor of weight 1).

If ϕ, V_μ, and $T_{\mu\nu}$ are a Weyl scalar, Weyl vector, and Weyl tensor of weights n, respectively, we obtain the following transformation laws:

$$\widetilde{\nabla_\mu\phi}=\lambda^{n}(\nabla_\mu\,\phi+n\phi\nabla_\mu\,\ln\lambda), \quad (2.17)$$

$$\widetilde{\nabla_\mu V_\nu}=\lambda^{n}(\nabla_\mu V_\nu-\tfrac{1}{2}V_\mu\nabla_\nu\ln\lambda+(n-\tfrac{1}{2})V_\nu\nabla_\mu\ln\lambda+\tfrac{1}{2}g_{\mu\nu}V_\rho\nabla^{\rho}\ln\lambda), \quad (2.18)$$

$$\widetilde{\nabla_\mu V^{\mu}}=\lambda^{n-1}[\nabla_\mu V^{\mu}+(n+1)V^{\mu}\nabla_\mu\ln\lambda],$$

$$\widetilde{\nabla_\mu T^{\mu\nu}}=\lambda^{n-2}[\nabla_\mu T^{\mu\nu}+(n+\tfrac{1}{2})T^{\mu\nu}\nabla_\mu\ln\lambda+\tfrac{1}{2}T^{\nu\mu}\nabla_\mu\ln\lambda-\tfrac{1}{2}T^{\mu}{}_{\mu}\nabla^{\nu}\ln\lambda]. \quad (2.19)$$

III. CONFORMAL COVARIANCE OF EULER–LAGRANGE EQUATIONS

The action S for a tensor or spinor field ψ is the Riemann's scalar

$$S=\int\mathscr{L}(\psi,\nabla_\mu\psi)d\eta, \quad (3.1)$$

where $\mathscr{L}(\psi,\nabla_\mu\psi)$ is the Lagrangian density.

Since we want (3.1) to result in a conformal invariant (Weyl scalar of weight zero), the Lagrangian density must be a Weyl scalar of weight -2, i.e.,

$$\tilde{\mathscr{L}}=\lambda^{-2}\mathscr{L}. \quad (3.2)$$

The field equations are obtained by means of the Euler–Lagrange equations

$$\nabla_\mu\,\frac{\partial\mathscr{L}}{\partial(\nabla_\mu\psi)}-\frac{\partial\mathscr{L}}{\partial\psi}=0. \quad (3.3)$$

Therefore, it is important in order to assure the conformal covariance of the field equations, to study the conformal behavior of (3.3).

Let $\psi_{\nu_1\cdots\nu_p}$ be a Weyl tensor of rank p and weight n. Then, we have for ψ and its covariant derivatives, the following transformation laws under (2.1):

$$\tilde{\psi}_{\nu_1\cdots\nu_p}=\lambda^{n}\psi_{\nu_1\cdots\nu_p}, \quad (3.4)$$

$$\widetilde{\nabla_\mu\psi_{\nu_1\cdots\nu_p}}=\lambda^{n}\Big[\nabla_\mu\psi_{\nu_1\cdots\nu_p}+n\psi_{\nu_1\cdots\nu_p}\nabla_\mu\ln\lambda-\sum_{i=1}^{p}(\tilde{\Gamma}^{\beta}_{\mu\nu_i}-\Gamma^{\beta}_{\mu\nu_i})\psi_{\nu_1\cdots\nu_{i-1}\beta\nu_{i+1}\cdots\nu_p}\Big]. \quad (3.5)$$

Free variables of the Lagrangian density are $\psi_{\nu_1\cdots\nu_p}$ and $\nabla_\mu\psi_{\nu_1\cdots\nu_p}$, whch can be expressed in terms of $\tilde{\psi}_{\nu_1\cdots\nu_p}$ and $\tilde{\nabla}_\mu\tilde{\psi}_{\nu_1\cdots\nu_p}$ by means of (3.4) and (3.5):

$$\psi_{\nu_1\cdots\nu_p}=\lambda^{-n}\tilde{\psi}_{\nu_1\cdots\nu_p}, \quad (3.6)$$

$$\nabla_\mu\psi_{\nu_1\cdots\nu_p}=\lambda^{-n}\Big[\widetilde{\nabla_\mu\psi_{\nu_1\cdots\nu_p}}-n\tilde{\psi}_{\nu_1\cdots\nu_p}\nabla_\mu\ln\lambda+\sum_{i=1}^{p}(\tilde{\Gamma}^{\beta}_{\mu\nu_i}-\Gamma^{\beta}_{\mu\nu_i})\tilde{\psi}_{\nu_1\cdots\nu_{i-1}\beta\nu_{i+1}\cdots\nu_p}\Big]. \quad (3.7)$$

From (3.2), (3.6), and (3.7), we obtain

$$\frac{\partial \tilde{\mathscr{L}}}{\partial \tilde{\psi}_{\nu_1 \cdots \nu_p}} = \lambda^{-(n+2)} \Bigg[\frac{\partial \mathscr{L}}{\partial \psi_{\nu_1 \cdots \nu_p}} - n \nabla_\mu \ln \lambda \frac{\partial \mathscr{L}}{\partial \nabla_\mu \psi_{\nu_1 \cdots \nu_p}} + \sum_{i=1}^{p} (\tilde{\Gamma}^{\nu_i}_{\mu\beta} - \Gamma^{\nu_i}_{\mu\beta}) \frac{\partial \mathscr{L}}{\partial \nabla_\mu \psi_{\nu_1 \cdots \nu_{i-1}\beta\nu_{i+1} \cdots \nu_p}} \Bigg], \tag{3.8}$$

$$\frac{\partial \mathscr{L}}{\partial \widetilde{\nabla_\mu \psi_{\nu_1 \cdots \nu_p}}} = \lambda^{-(n+2)} \frac{\partial \mathscr{L}}{\partial \nabla_\mu \psi_{\nu_1 \cdots \nu_p}}. \tag{3.9}$$

Now, using (3.9), and on account of (2.6), we deduce

$$\tilde{\nabla}_\mu \frac{\partial \tilde{\mathscr{L}}}{\partial \tilde{\nabla}_\mu \tilde{\psi}_{\nu_1 \cdots \nu_p}} = \lambda^{-(n+2)} \Bigg[\nabla_\mu \frac{\partial \mathscr{L}}{\partial \nabla_\mu \psi_{\nu_1 \cdots \nu_p}} - n \nabla_\mu \ln \lambda \frac{\partial \mathscr{L}}{\partial \nabla_\mu \psi_{\nu_1 \cdots \nu_p}} + \sum_{i=1}^{p} (\tilde{\Gamma}^{\nu_i}_{\mu\beta} - \Gamma^{\nu_i}_{\mu\beta}) \frac{\partial \mathscr{L}}{\partial \nabla_\mu \psi_{\nu_1 \cdots \nu_{i-1}\beta\nu_{i+1} \cdots \nu_p}} \Bigg]. \tag{3.10}$$

In consequence, (3.9) and (3.10) yield

$$\left(\nabla_\mu \frac{\partial \mathscr{L}}{\partial \nabla_\mu \psi_{\nu_1 \cdots \nu_p}} - \frac{\partial \mathscr{L}}{\partial \psi_{\nu_1 \cdots \nu_p}} \right)^{\sim} = \lambda^{-(n+2)} \left(\nabla_\mu \frac{\partial \mathscr{L}}{\partial \nabla_\mu \psi_{\nu_1 \cdots \nu_p}} - \frac{\partial \mathscr{L}}{\partial \psi_{\nu_1 \cdots \nu_p}} \right). \tag{3.11}$$

The relationship (3.11) shows that $\nabla_\mu \partial \mathscr{L} / \partial \nabla_\mu \psi_{\nu_1 \cdots \nu_p} - \partial \mathscr{L} / \partial \psi_{\nu_1 \cdots \nu_p}$ is a Weyl tensor, and in this way we assure the conformal covariance of the field equations for all massless tensor fields.

An analogous proof can be realized, "mutatis mutandi," for spinor fields.

IV. CONFORMAL MASSLESS SCALAR FIELD

From previous discussions, we know that the problem of finding a field equation invariant under conformal transformations is equivalent to finding a Lagrangian density that transforms like (3.2) under conformal transformations. More exactly, the Lagrangian density $\mathscr{L}$ must transform like

$$\tilde{\mathscr{L}} = \lambda^{-2} (\mathscr{L} + \nabla_\mu f^\mu), \tag{4.1}$$

where f_μ is an arbitrary vector function that vanishes at infinity quickly enough. In fact, the term $\nabla_\mu f^\mu = (-g)^{-1/2} \partial_\mu (\sqrt{-g} f^\mu)$ in (4.1) disappears from the action integral using Gauss's theorem and the action turns out, clearly, conformally invariant.

We consider as an introduction, the simple and well-known case of a spin-0 massless field ϕ. We postulate as the Lagrangian density for ϕ,

$$\mathscr{L} = \tfrac{1}{2} \nabla_\mu \phi \nabla^\mu \phi + \tfrac{1}{2} \xi R \phi^2, \tag{4.2}$$

where R is the curvature scalar and ξ is a parameter to be determined (coupling parameter). Using the expressions (1.9), (1.16), and (1.17),

$$\tilde{\mathscr{L}} = \lambda^{2n-1} [\mathscr{L} + \tfrac{1}{2} n \nabla_\mu (\phi^2 \nabla^\mu \ln \lambda) + \tfrac{1}{2} (n^2 - \tfrac{3}{2}\xi) \phi^2 \nabla_\mu \ln \lambda \nabla^\mu \ln \lambda - \tfrac{1}{2} (n + 3\xi) \phi_2 \nabla_\mu \nabla^\mu \ln \lambda]. \tag{4.3}$$

Therefore, comparing (4.3) with (4.1), it follows that $n = -\frac{1}{2}$ and $\xi = \frac{1}{6}$ (conformal coupling). Then we find

$$\mathscr{L} = \lambda^{-2} [\mathscr{L} - \nabla_\mu (\tfrac{1}{4} \phi^2 \nabla^\mu \ln \lambda)], \tag{4.4}$$

which has exactly the form required. Consequently, we take, as Lagrangian density for ϕ,

$$\mathscr{L} = \tfrac{1}{2} (\nabla_\mu \phi \nabla^\mu \phi + \tfrac{1}{6} R \phi^2). \tag{4.5}$$

The field transforms like

$$\tilde{\phi} = \lambda^{-1/2} \phi. \tag{4.6}$$

Applying the Euler–Lagrange equations to (4.5) we establish the field equation for

$$\nabla^\mu \nabla_\mu \phi - \tfrac{1}{6} R \phi = 0. \tag{4.7}$$

Orthonormal complex solutions of Eq. (4.7) can be obtained with respect to a suitable inner product $\langle \, , \rangle$ in the vector space of solutions of (4.7). So, if $U(x)$ and $V(x)$ are complex solutions of (4.7), we define the inner product between U and V as

$$\langle U, V \rangle = i \int_\Sigma (V \nabla_\mu U^* - U^* \nabla_\mu V) d\sigma^\mu, \tag{4.8}$$

where Σ is a Cauchy surface of V_4. On account of the fact that $U(x)$ and $V(x)$ satisfy (4.7), (4.8) is independent of the Cauchy surface used to perform the integration. Moreover, this inner product is clearly a Riemann's scalar and it results a Weyl scalar of weight 0 also; i.e., (4.8) is invariant under conformal transformations.

V. CONFORMAL NEUTRINO FIELD

We consider, now, the case of a spin-$\frac{1}{2}$ massless field (neutrino field), which is represented by a Weyl spinor ψ of weight n.

We introduce the 4×4 Dirac's matrices that satisfy the following anticommutation relations at every point $x \in V_4$:

$$\gamma_\mu \gamma_\nu + \gamma_\nu \gamma_\mu = -2 g_{\mu\nu} I, \tag{5.1}$$

where I is the 4×4 unit matrix. From (5.1) it follows that under conformal transformations

$$\tilde{\gamma}_\mu = \lambda^{1/2} \gamma_\mu, \quad \tilde{\gamma}^\mu = \lambda^{-1/2} \gamma^\mu. \tag{5.2}$$

Clearly, from γ_μ we can obtain another set of Dirac's matrices, say $\check{\gamma}_\mu$, using a similarity transformation:

$$\check{\gamma}_\mu = S \gamma_\mu S^{-1}. \tag{5.3}$$

In fact, evidently the matrices $\check{\gamma}_\mu$ satisfy the same anticommutation relations (5.1) as γ_μ. The matrix S that defines the transformation (5.3) is subject to a unique condition, to be nonsingular. The transformation (5.3) can be considered as a change of basis in the spinor space (change of spinor basis).

Under change of spinor basis, the contravariant spinor ψ (column spinor) transforms like

$$\check{\psi} = S \psi. \tag{5.4}$$

The Dirac adjoint of ψ, which will be written $\bar{\psi}$, is defined by

$$\bar{\psi} = \psi^\dagger \beta, \tag{5.5}$$

where $\dagger$ is a conjugation and a transportation, i.e., the ad-junction, and β is the matrix that performs the following change of spinor basis:

$$-\gamma_\mu^\dagger = \beta\gamma_\mu\beta^{-1}. \tag{5.6}$$

The covariant spinor $\bar\psi$ (row spinor) transforms under change of spinor basis as

$$\check{\bar\psi} = \bar\psi S. \tag{5.7}$$

We postulate, as the Lagrangian density for the neutrino field,

$$\mathscr{L} = \tfrac{1}{2}[(\nabla_\mu\bar\psi)\gamma^\mu\psi - \bar\psi\gamma^\mu\nabla_\mu\psi], \tag{5.8}$$

where the covariant derivatives are defined by means of the spinor connection σ_μ as[2]

$$\nabla_\mu\psi = \partial_\mu\psi + \sigma_\mu\psi, \tag{5.9}$$

$$\nabla_\mu\bar\psi = \partial_\mu\bar\psi - \bar\psi\sigma_\mu. \tag{5.10}$$

The field equations for ψ and $\bar\psi$ will be

$$\nabla_\mu\frac{\partial\mathscr{L}}{\partial\nabla_\mu\psi} - \frac{\partial\mathscr{L}}{\partial\psi} = 0, \tag{5.11}$$

$$\nabla_\mu\frac{\partial\mathscr{L}}{\partial\nabla_\mu\bar\psi} - \frac{\partial\mathscr{L}}{\partial\bar\psi} = 0. \tag{5.12}$$

We must take into consideration the symbolic character of the derivatives in these relations since ψ is a column spinor and $\bar\psi$ is a row spinor. Therefore, $\partial\mathscr{L}/\partial\psi$ is a row spinor, etc.

The free variables ψ, $\bar\psi$, $\nabla_\mu\psi$, and $\nabla_\mu\bar\psi$ in terms of $\tilde\psi$, $\tilde{\bar\psi}$, $\widetilde{\nabla_\mu\psi}$, and $\widetilde{\nabla_\mu\bar\psi}$ turn out to be

$$\begin{aligned}
\psi &= \lambda^{-n}\tilde\psi,\\
\bar\psi &= \lambda^{-n}\tilde{\bar\psi},\\
\nabla_\mu\psi &= \lambda^{-n}[\widetilde{\nabla_\mu\psi} - (n\nabla_\mu\ln\lambda + \Lambda_\mu)\tilde\psi],\\
\nabla_\mu\bar\psi &= \lambda^{-n}[\widetilde{\nabla_\mu\bar\psi} - \tilde{\bar\psi}(n\nabla_\mu\ln\lambda - \Lambda_\mu)],
\end{aligned} \tag{5.13}$$

where $\Lambda_\mu \equiv \tilde\sigma_\mu - \sigma_\mu$.

In order to preserve the property

$$\check{\nabla}_\mu\check\psi = S\nabla_\mu\psi, \tag{5.14}$$

the spinor connection σ_μ must transform under the change of spinor basis like

$$\check\sigma_\mu = S\sigma_\mu S^{-1} + S\partial_\mu S^{-1}. \tag{5.15}$$

Clearly, we note that (4.8) is a Riemann invariant and besides it is invariant under change of the spinor basis. Turning to conformal behavior of (5.8), using (5.13), (5.8) yields

$$\tilde{\mathscr{L}} = \lambda^{2n+3/2}[\mathscr{L} - \tfrac{1}{2}\bar\psi(\Lambda_\mu\gamma^\mu + \gamma^\mu\Lambda_\mu)\psi]. \tag{5.16}$$

In order to satisfy (4.1) we take

$$n = -\tfrac{3}{4}, \tag{5.17}$$

$$\Lambda_\mu\gamma^\mu + \gamma^\mu\Lambda_\mu = 0, \tag{5.18}$$

where Λ_μ is the conformal variation of σ_μ:

$$\Lambda_\mu = \tilde\sigma_\mu - \sigma_\mu. \tag{5.19}$$

Applying Eq. (5.12) to the Lagrangian density (5.8) we obtain the field equation for ψ:

$$\gamma^\mu\nabla_\mu\psi + \tfrac{1}{2}(\nabla_\mu\gamma^\mu)\psi = 0. \tag{5.20}$$

However, because the condition (5.18) does not determine σ_μ univocally, we can impose another condition on σ_μ. The very useful and usually introduced condition consists of demanding that the covariant derivative vanishes for the Dirac matrices. In this way, we take

$$\nabla_\mu\gamma_\nu = 0. \tag{5.21}$$

In order to assure the conformal consistency of the condition (5.21), it is necessary to study the behavior of $\nabla_\mu\gamma_\nu$ under conformal transformations. Since

$$\nabla_\mu\gamma_\nu = \gamma_{\nu;\mu} + \sigma_\mu\gamma_\nu - \gamma_\nu\sigma_\mu, \tag{5.22}$$

where the symbol ; is the covariant derivative of γ_ν "as if it were only a vector," we have

$$\widetilde{\nabla_\mu\gamma_\nu} = \lambda^{1/2}[\nabla_\mu\gamma_\nu + \Lambda_\mu\gamma_\nu - \gamma_\nu\Lambda_\mu - \tfrac{1}{2}(g_{\nu\rho}\gamma_\mu - g_{\mu\nu}\gamma_\rho)\nabla^\rho\ln\lambda]. \tag{5.23}$$

Therefore, the conformal consistency of (5.21) implies

$$\Lambda_\mu\gamma_\nu - \gamma_\nu\Lambda_\mu = \tfrac{1}{2}(g_{\nu\rho}\gamma_\mu - g_{\mu\nu}\gamma_\rho)\nabla^\rho\ln\lambda. \tag{5.24}$$

In the Appendix we prove that the only expression that verifies (5.18) and (5.24) is

$$\Lambda_\mu = \tfrac{1}{8}(\gamma_\nu\gamma_\mu - \gamma_\mu\gamma_\nu)\nabla^\nu\ln\lambda. \tag{5.25}$$

With (5.25), it is established as the transformation law of the spinor connection that verifies (5.21) and it has conformal consistency.

Of course, on account of (5.21), the field equation (5.20) reads

$$\gamma^\mu\nabla_\mu\psi = 0. \tag{5.26}$$

From the paper of Loos[4] we know that the most general form of the spinor connection σ_μ that verifies (5.21) is

$$\sigma_\mu = \tfrac{1}{4}(I - \tfrac{7}{6}\mathscr{A} + \tfrac{4}{3}\mathscr{A}^2 - \tfrac{2}{3}\mathscr{A}^3)(\gamma_{\nu;\mu}\gamma^\nu) + v_\mu I, \tag{5.27}$$

where v_μ is an arbitrary vector and $\mathscr{A}$ is the following linear operator on 4×4 matrices $\mathscr{M}$:

$$\mathscr{A}\mathscr{M} = \tfrac{1}{4}\gamma_\mu\mathscr{M}\gamma^\mu. \tag{5.28}$$

Since

$$\tilde\gamma_{\nu;\mu}\tilde\gamma^\nu = \gamma_{\nu;\mu}\gamma^\nu + \tfrac{1}{2}(\gamma_\nu\gamma_\mu - \gamma_\mu\gamma_\nu)\nabla^\nu\ln\lambda, \tag{5.29}$$

and also

$$\mathscr{A}(\gamma_\nu\gamma_\mu - \gamma_\mu\gamma_\nu) = 0, \tag{5.30}$$

from (5.27) it follows that

$$\tilde\sigma_\mu = \sigma_\mu + \tfrac{1}{8}(\gamma_\nu\gamma_\mu - \gamma_\mu\gamma_\nu)\nabla^\nu\ln\lambda + (\tilde v_\mu - v_\mu)I. \tag{5.31}$$

Then, we see that (5.31) is consistent with (5.25) if we take

$$\tilde v_\mu = v_\mu. \tag{5.32}$$

In a way similar to the scalar case, we can define an inner product in the vector space of the complex solutions of Eq. (5.26) and its adjoint one. In fact, if $U(x)$ and $V(x)$ are two Dirac spinors that verify (5.26), we define

$$\langle U,V\rangle = -i\int_\Sigma \bar U\gamma^\mu V\,d\sigma_\mu, \tag{5.33}$$

where Σ is a Cauchy surface of V_4.

Since the adjoint field $\bar\psi$ satisfies the equation

$$(\nabla_\mu\bar\psi)\gamma^\mu = 0, \tag{5.34}$$

(5.33) is independent of the Cauchy surface used to perform the integration. In fact,

$$\nabla_\mu (\bar{U}\gamma^\mu V) = (\nabla_\mu \bar{U})\gamma^\mu V + \bar{U}\gamma^\mu \nabla_\mu V = 0, \tag{5.35}$$

which proves our statement.

The inner product (5.33) is clearly invariant under coordinate transformations in V_4 and invariant under the change of the spinor basis. Moreover, since

$$\tilde{U} = \lambda^{-3/4} U, \quad \tilde{\bar{U}} = \lambda^{-3/4} \bar{U}, \tag{5.36}$$

(5.33) is invariant under conformal transformations, too.

We can define a helicity operator γ^5 in a covariant way as

$$\gamma^5 = -(i/4!)\eta^{\mu\nu\lambda\rho}\gamma_\mu \gamma_\nu \gamma_\lambda \gamma_\rho, \tag{5.37}$$

where η is the "element of volume" pseudotensor, i.e.,

$$\eta^{\mu\nu\lambda\rho} = (1/\sqrt{-g})\varepsilon^{\mu\nu\lambda\rho}. \tag{5.38}$$

From (5.2) and on account of $\sqrt{-g} = \lambda^2\sqrt{-g}$,

$$\tilde{\gamma}^5 = \gamma^5. \tag{5.39}$$

Using at every point of V_4 an orthonormal base, where $g_{\mu\nu}$ = diagonal matrix and the relations (5.1) we can deduce that

$$(\gamma^5)^2 = 1. \tag{5.40}$$

Therefore, the eigenvalues of γ^5 are ± 1.

Analogously to the flat case, we can define the projection operators

$$\begin{aligned} h_+ &= \tfrac{1}{2}(1+\gamma^5), \\ h_- &= \tfrac{1}{2}(1-\gamma^5). \end{aligned} \tag{5.41}$$

The operators h_+ and h_- determine the subspaces of positive and negative helicity, respectively.

Note that

$$\tilde{h}_\pm = h_\pm, \tag{5.42}$$

so the notion of helicity turns out to be invariant under conformal transformations.

VI. CONFORMAL ELECTROMAGNETIC FIELD

A. The problem of quantization in Minkowski space

From classic electrodynamics, the Lagrangian density for an electromagnetic field is (in the flat space-time)

$$\mathcal{L} = -\tfrac{1}{4} F_{\mu\nu} F^{\mu\nu}, \tag{6.1}$$

where $F_{\mu\nu}$ is the electromagnetic field tensor

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{6.2}$$

Nevertheless, as we know from quantum electrodynamics in Minkowski space, it becomes a problem when we intend to perform the canonical quantization. In fact, from (6.1), the conjugate momentum of A_0 vanishes,

$$\pi^0 = \frac{\partial \mathcal{L}}{\partial(\partial_0 A_0)} = F^{00} = 0, \tag{6.3}$$

and, in consequence, the canonical commutation rules,

$$[A_\mu(t,\mathbf{x}); \pi^\nu(t,\mathbf{y})] = i\delta_\mu^\nu \delta(\mathbf{x}-\mathbf{y}), \tag{6.4}$$

are inconsistent.

There are several ways to solve this problem; we shall study two of them: (A) the quantization in the temporal gauge, and (B) the Gupta–Bleuler quantization method.

(A) The quantization in the temporal gauge introduces[6] the following gauge condition:

$$A_0 = 0. \tag{6.5}$$

Therefore, A_0 disappears as a degree of freedom and then does not appear in the commutation rules.

(B) The Gupta–Bleuler quantization method in Minkowski space consists of the change of the Lagrangian density (6.1) by[7]

$$\mathcal{L} = -\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} - (\alpha/2)(\partial_\mu A^\mu)^2, \tag{6.6}$$

where α is a dimensionless constant $\neq 0$. From (6.6),

$$\pi^0 = -\alpha\, \partial_\mu A^\mu \neq 0, \tag{6.7}$$

and in this way, the inconsistency disappears. The elimination of unphysical photons is obtained by means of the following gauge condition:

$$(\partial_\mu A^\mu)^- |\text{Phys}\rangle = 0, \tag{6.8}$$

where $(\partial_\mu A^\mu)^-$ is the negative-frequency part of the operator $\partial_\mu A^\mu$.

B. The problem of quantization in curved space-time. The field $U^\mu(x)$

In order to obtain a conformal generalization of both methods to curved space-time, we are forced to introduce new entities. These could be certain ghosts,[5] but we prefer to search the solution introducing systematically a timelike vector field $U^\mu(x)$, which we shall take to be unitary[8]:

$$g_{\mu\nu} U^\mu U^\nu = U^\mu U_\mu = 1. \tag{6.9}$$

Clearly, from (6.9), U_μ transforms under conformal transformations as

$$\tilde{U}^\mu = \lambda^{-1/2} U^\mu, \quad \tilde{U}_\mu = \lambda^{1/2} U_\mu. \tag{6.10}$$

In this way, U_μ results in a Weyl vector of weight $\tfrac{1}{2}$.

C. Field equations

Now we will try to generalize to the curved case the Lagrangian densities (6.1) and (6.6).

(A) The generalization in this case is straightforward. In fact, we take

$$\mathcal{L} = -\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{6.11}$$

with

$$F_{\mu\nu} = \nabla_\mu A_\nu - \nabla_\nu A_\mu = \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{6.12}$$

In order to satisfy the condition (3.2), we must assume that A_μ is a Weyl vector of weight 0, i.e.,

$$\tilde{A}_\mu = A_\mu, \quad \tilde{A}^\mu = \lambda^{-1} A^\mu. \tag{6.13}$$

As we shall see in Sec. VI F, we take as the gauge condition $A_0 = 0$ (in any coordinate system), and then, from (6.11), by means of Eq. (3.3), we obtain the field equations

$$\nabla_\mu F^{\mu i} = 0 \tag{6.14}$$

(see Ref. 9).

(B) The most natural generalization of (6.6) to curved space-time would be, in principle,

$$\mathcal{L} = -\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} - (\alpha/2)(\nabla_\mu A^\mu)^2. \tag{6.15}$$

However, as we shall see, the Lagrangian density (6.15) is unsuitable for a conformal theory. In fact, if we assume (6.13), then from (2.19) we have

$$\widetilde{\nabla_\mu A^\mu} = \lambda^{-1}(\nabla_\mu A^\mu + A^\mu \nabla_\mu \ln \lambda). \tag{6.16}$$

Therefore, $\nabla_\mu A^\mu$ is not a Weyl scalar, and in consequence (6.15) is unsuitable as we said.

Then, as we see, the conformal generalization of (6.6) to the curved case is not trivial. Since the change $\partial_\mu A^\mu \to \nabla_\mu A^\mu$ is not adequate, we propose the change $\partial_\mu A^\mu \to D_\mu A^\mu$, where $D_\mu A^\mu$ is an entity that transforms under (2.1) as a Weyl scalar of weight -1:

$$\widetilde{D_\mu A^\mu} = \lambda^{-1} D_\mu A^\mu. \tag{6.17}$$

In the next section we shall face the construction of $D_\mu A^\mu$.

D. Conformal derivative

Now, we propose the introduction of an appropriate conformal derivative for A_μ. Therefore, let A_μ be an Weyl vector of weight 0, i.e., A_μ satisfies (6.13). The ordinary covariant derivative for A_μ transforms under conformal transformations as

$$\widetilde{\nabla_\nu A_\mu} = \nabla_\nu A_\mu - \tfrac{1}{2}(\delta^\rho_\mu \nabla_\nu \ln \lambda + \delta^\rho_\nu \nabla_\nu \ln \lambda - g_{\mu\nu}\nabla^\rho \ln \lambda)A_\rho. \tag{6.18}$$

Clearly, $\nabla_\nu A_\mu$ is not a Weyl tensor, that is to say, ∇_ν is not a Weyl vector. Of course, the nonconformal behavior of ∇_ν is produced by the nonconformal behavior of the Riemannian connection $\Gamma^\rho_{\mu\nu}$ [cf. Eq. (2.5)]:

$$\widetilde{\Gamma}^\rho_{\mu\nu} = \Gamma^\rho_{\mu\nu} + \tfrac{1}{2}(\delta^\rho_\mu \nabla_\nu \ln \lambda + \delta^\rho_\nu \nabla_\mu \ln \lambda - g_{\mu\nu}\nabla^\rho \ln \lambda). \tag{6.19}$$

Therefore, for a conformal theory, we would need a "conformal connection," $C^\rho_{\mu\nu}$, that is to say

$$\widetilde{C}^\rho_{\mu\nu} = C^\rho_{\mu\nu}. \tag{6.20}$$

Then, we would define the conformal derivative of A_μ as

$$D_\nu A_\mu = \partial_\nu A_\mu - C^\rho_{\mu\nu} A_\rho. \tag{6.21}$$

So, our problem is to find the conformal connection $C^\rho_{\mu\nu}$.

The transformation law (6.19) suggests the following expression for $C^\rho_{\mu\nu}$:

$$C^\rho_{\mu\nu} = \Gamma^\rho_{\mu\nu} + \tfrac{1}{2}(\delta^\rho_\mu Q_\nu + \delta^\rho_\nu Q_\mu - g_{\mu\nu}Q^\rho), \tag{6.22}$$

where, in order to satisfy (6.20), Q_μ must transform as

$$\widetilde{Q}_\mu = Q_\mu - \nabla_\mu \ln \lambda. \tag{6.23}$$

With the geometric objects of the manifold it is not possible to obtain a vector, thus in order to define Q_μ we are forced to use the timelike vector field U_μ introduced in Sec. VI B.

The most general expression with appropriate dimension is

$$Q_\mu = bU^\nu \nabla_\nu U_\mu + aU_\mu \nabla_\nu U^\nu, \tag{6.24}$$

where a and b are dimensionless constants to be determined [we do not include $U_\nu \nabla_\mu U^\nu$ because from (6.9) it vanishes identically].

For the covariant derivatives of U_μ we have

$$\widetilde{\nabla_\nu U_\mu} = \lambda^{+1/2}(\nabla_\nu U_\mu - \tfrac{1}{2}U_\nu \nabla_\mu \ln \lambda + \tfrac{1}{2}g_{\mu\nu}U_\rho \nabla^\rho \ln \lambda), \tag{6.25}$$

$$\widetilde{\nabla_\nu U^\nu} = \lambda^{-1/2}(\nabla_\nu U^\nu + \tfrac{3}{2}U^\nu \nabla_\nu \ln \lambda). \tag{6.26}$$

Then, by means of (6.10), (6.24), (6.25), and (6.26), we obtain, for Q_μ,

$$\widetilde{Q}_\mu = Q_\mu - (b/2)\nabla_\mu \ln \lambda + \tfrac{1}{2}(b+3a)U_\mu U_\nu \nabla^\nu \ln \lambda, \tag{6.27}$$

and comparing (6.27) with (6.23), we deduce

$$b = 2, \quad a = -\tfrac{2}{3}. \tag{6.28}$$

Thus, Q_μ is univocally determined in terms of U_μ:

$$Q_\mu = 2U^\nu \nabla_\nu U_\mu - \tfrac{2}{3}U_\mu \nabla_\nu U^\nu. \tag{6.29}$$

Now, we can obtain the expression of the conformal connection in terms of U_μ. From (6.22), on account of (6.29),

$$C^\rho_{\mu\nu} = \Gamma^\rho_{\mu\nu} + U^\sigma \nabla_\sigma(\delta^\rho_\mu U_\nu + \delta^\rho_\nu U_\mu - g_{\mu\nu}U^\rho) - \tfrac{1}{3}(\delta^\rho_\mu U_\nu + \delta^\rho_\nu U_\mu - g_{\mu\nu}U^\rho)\nabla_\sigma U^\sigma, \tag{6.30}$$

and consequently, the conformal derivative of A_μ will be

$$D_\nu A_\mu = \nabla_\nu A_\mu - \tfrac{1}{2}(\delta^\rho_\mu Q_\nu + \delta^\rho_\nu Q_\nu - g_{\mu\nu}Q^\mu)A_\rho, \tag{6.31}$$

or, in terms of U_μ,

$$D_\nu A_\mu = \nabla_\nu A_\mu - U^\sigma \nabla_\sigma(\delta^\rho_\mu U_\nu + \delta^\rho_\nu U_\mu - g_{\mu\nu}U^\rho)A_\rho + \tfrac{1}{3}(\delta^\rho_\mu U_\nu + \delta^\rho_\nu U_\mu - g_{\mu\nu}U^\rho)A_\rho. \tag{6.32}$$

Clearly, $D_\nu A_\mu$ is a Weyl tensor weight 0:

$$\widetilde{D_\nu A_\mu} = D_\nu A_\mu. \tag{6.33}$$

Now, we can calculate the "conformal divergence" of A_μ, defined as $D_\mu A^\mu = g^{\mu\nu}D_\nu A_\mu$; it results from (6.31):

$$D_\mu A^\mu = \nabla_\mu A^\mu + Q_\mu A^\mu, \tag{6.34}$$

or, in terms of U_μ,

$$D_\mu A^\mu = \nabla_\mu A^\mu + (2U^\nu \nabla_\nu U_\mu - \tfrac{2}{3}U_\mu \nabla_\nu U^\nu)A^\mu. \tag{6.35}$$

Thus, $D_\mu A^\mu$ results a Weyl scalar of weight -1:

$$\widetilde{D_\mu A^\mu} = \lambda^{-1}D_\mu A^\mu. \tag{6.36}$$

Then, we shall take, as the Lagrangian density for A_μ,

$$\mathscr{L} = -\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} - (\alpha/2)(D_\mu A^\mu)^2. \tag{6.37}$$

In this way, $\mathscr{L}$ turns out to be a Weyl scalar of weight -2 as we wanted.

From (6.37), by means of Lagrange equations (3.3), we obtain the following equations for A_μ:

$$\nabla_\mu F^{\mu\nu} + \alpha \nabla^\nu(D_\mu A^\mu) - \alpha Q^\nu D_\mu A^\mu = 0. \tag{6.38}$$

E. Conformal inner product

In order to define an orthonormal basis in the space of complex solutions of the field equations, and on account of the conformal covariance of the theory, it is necessary to introduce an inner product $\langle W^\mu; V_\mu \rangle_\Sigma$, invariant under conformal transformations, where W_μ and V_μ are complex solutions of the field equations and Σ is a Cauchy surface of V_4.

Moreover, we demand the invariance of $\langle\,;\,\rangle_\Sigma$ under changes of the Cauchy surface Σ (Σ invariance).

The expression (4.8) suggests for $\langle\,;\,\rangle$ the form

$$\langle W^\mu;V_\mu\rangle = i\int_\Sigma (V^\mu\nabla_\nu W^*_\mu - \dot{W}^\mu\nabla_\nu V_\mu)d\sigma^\nu. \tag{6.39}$$

However, this inner product is unsuitable because under conformal transformations we have

$$(V^\mu\nabla_\nu W^*_\mu - \dot{W}^\mu\nabla_\nu V_\mu)^\sim = \Big[V^\mu\nabla_\nu W^*_\mu - \dot{W}^\mu\nabla_\nu V_\mu + (V_\nu W^*_\mu - W^*_\nu V_\mu)\nabla^\mu \ln\lambda\,\Big]\lambda^{-1}, \tag{6.40}$$

and then

$$\langle\widetilde{W^\mu;V_\mu}\rangle \neq \langle W^\mu;V_\mu\rangle. \tag{6.41}$$

In consequence, if we want to preserve the conformal covariance of the theory, we must modify (6.39).

(A) We propose, for Eq. (6.14), the following expression:

$$\langle W^\mu;V_\mu\rangle = i\int_\Sigma \Big[V^\mu\nabla_\nu \dot{W}_\mu - \dot{W}^\mu\nabla_\nu V_\mu + a(V^\mu\nabla_\mu W^*_\nu - \dot{W}^\mu\nabla_\mu V_\nu) + b(V_\nu\nabla_\mu \dot{W}^\mu - \dot{W}_\nu\nabla_\mu V^\mu)\Big]d\sigma^\nu, \tag{6.42}$$

where a and b are dimensionless constants to be determined.

Under (2.1) we have, from (6.42),

$$\langle\widetilde{W^\mu;V_\mu}\rangle = \langle W^\mu;V_\mu\rangle + i(1+a+b)\times\int_\Sigma (V_\nu W^*_\mu - \dot{W}_\nu V_\mu)\nabla^\mu\ln\lambda\, d\sigma^\nu, \tag{6.43}$$

and then, in order to preserve the conformal invariance, we must take

$$1+a+b=0. \tag{6.44}$$

On account of Gauss's theorem, the Σ invariance is equivalent to

$$\nabla^\nu\Big[V^\mu\nabla_\nu W^*_\mu - \dot{W}^\mu\nabla_\nu V_\mu + a(V^\mu\nabla_\mu \dot{W}_\nu - \dot{W}^\mu\nabla_\mu V_\nu) + b(V_\nu\nabla_\mu\dot{W}^\mu - \dot{W}_\nu\nabla_\mu V^\mu)\Big]=0. \tag{6.45}$$

Developing (5.45) and using the field equations (6.14), we obtain

$$(1+a+b)(V^\mu\nabla^\nu\nabla_\nu W^*_\mu - \dot{W}^\mu\nabla^\nu\nabla_\nu V_\mu)=0. \tag{6.46}$$

In consequence, the Σ-invariance condition is

$$1+a+b=0. \tag{6.47}$$

Therefore the conformal invariance condition and the Σ-invariance condition are the same.

Thus, we take for $\langle\,;\,\rangle$ the following expression:

$$\langle W^\mu;V_\mu\rangle = i\int_\Sigma\Big[V^\mu\nabla_\nu\dot{W}_\mu - \dot{W}^\mu\nabla_\mu V_\mu - (1+b)(V^\mu\nabla_\mu\dot{W}_\nu - \dot{W}^\mu\nabla_\mu V_\nu) + b(V_\nu\nabla_\mu\dot{W}^\mu - \dot{W}_\nu\nabla_\mu V^\mu)\Big]d\sigma^\nu, \tag{6.48}$$

where the constant b remains arbitrary.

(B) In this case we propose, for $\langle\,;\,\rangle$,

$$\langle W^\mu;V_\mu\rangle = i\int_\Sigma (V^\mu D_\nu W^*_\mu - \dot{W}^\mu D_\nu V_\mu)d\sigma^\nu, \tag{6.49}$$

where D_ν is the conformal derivative defined in (6.31).

Clearly, on account of (6.33), (6.49) results conformally invariant.

From (6.31) we obtain the following expression for (6.49) in terms of the Riemannian derivatives of W^*_μ and V_μ:

$$\langle W^\mu;V_\mu\rangle = i\int_\Sigma\Big[V^\mu\nabla_\nu W^*_\mu - \dot{W}^\mu\nabla_\nu V_\mu + Q^\mu(\dot{W}_\mu V_\nu - V_\mu\dot{W}_\nu)\Big]d\sigma^\nu. \tag{6.50}$$

In order to obtain the Σ invariance of (5.49) we impose the condition[10]

$$\nabla^\nu(V^\mu D_\nu W^*_\mu - \dot{W}^\mu D_\nu V_\mu)=0. \tag{6.51}$$

Developing (6.51) and using the field equations (6.38),

$$(1-\alpha)\nabla^\nu(V_\nu D_\mu\dot{W}^\mu - W_\nu D_\mu V^\mu)=0. \tag{6.52}$$

So, we assure the Σ invariance of (6.49) by taking in (6.38), $\alpha=1$. Then, the field equations are

$$\nabla_\mu F^{\mu\nu} + \nabla^\nu(D_\mu A^\mu) - Q^\nu D_\mu A^\mu = 0. \tag{6.53}$$

We note that the expression of $\langle\,;\,\rangle$ is different for the cases (A) and (B) because the field equations are different.

F. Gauge conditions

Now, we must face the construction of appropriate gauge conditions in order to eliminate the unphysical photons. Naturally, we shall require conformally invariant conditions.

(A) For this case we take, as a generalization of (6.5),

$$U^\mu A_\mu = 0, \tag{6.54}$$

where U^μ is the timelike vector field introduced in Sec. VI B.

In the coordinate system where $U^\mu=(U^0,0,0,0)$, (6.54) turns out to be $A_0=0$.

Since U^μ and A^μ are Weyl vectors, (6.54) turns out to be a conformally invariant condition.

(B) In this case, the gauge condition will be

$$(D_\mu A^\mu)^-|\mathrm{Phys}\rangle = 0, \tag{6.55}$$

where $(D_\mu A^\mu)^-$ is the negative-frequency part of the operator $D_\mu A^\mu$. Now, the conformal invariance of (6.55) becomes evident.

VII. CONFORMAL TRANSPLANTATION

Conformal theories present a very convenient formulation for the study of fields in space-times with conformal metrics to the Minkowski one; i.e., those metrics can be written in any coordinate system $(\xi^0,\xi^1,\xi^2,\xi^3)$ as

$$(ds)^2 = \lambda(\xi^\mu)[(d\xi^0)^2 - (d\xi^1)^2 - (d\xi^2)^2 - (d\xi^3)^2]. \tag{7.1}$$

If $\lambda(\xi^\mu) = \lambda(\xi^0)$, we find the very important case of the Robertson–Walker metric, corresponding to an expanding universe,

$$(ds)^2 = (dt)^2 - a^2(t)[(dx)^2 + (dy)^2 + (dz)^2], \qquad (7.2)$$

which we can write as

$$(ds)^2 = a^2(\eta)[(d\eta)^2 - (dx)^2 - (dy)^2 - (dz)^2], \qquad (7.3)$$

where η is the "conformal time," defined by

$$d\eta = dt/a(t). \qquad (7.4)$$

So, we have a metric that is conformal to the flat one. In this case we have

$$\lambda = a^2(t). \qquad (7.5)$$

We shall call, in the general case (7.1), the coordinates ξ^μ "conformal coordinates." The advantage of working with conformal coordinates is that in this case we can solve the equations in such coordinates as in the flat case (if we have a conformal theory).

In the spin-1 case, we can do it, now, because we have a fully conformal theory. Naturally, in conformal coordinates, the solutions will be the ones corresponding to Minkowski space. Then, we perform the conformal transformation in order to work in the metric we want to study. We shall call this method the "conformal transplantation." A very important feature of the conformal transplantation is that it allows us to define univocally (by analogy with the flat case) the particle model, i.e., the decomposition of the solutions into the positive- and negative-frequency parts (in the Robertson–Walker universe, at least).

VIII. MASSLESS FIELDS IN THE EXPANDING UNIVERSE

Now, we shall limit ourselves to the particular case of an expanding universe, i.e., the space-time where the metric is (7.2). We shall call V_4 an expanding universe even in the case where $a(t)$ is not monotonically increasing with time t.

A. Scalar field

We shall start the study of the massless fields in the expanding universe with the well-known case of the scalar field. Using Eq. (4.7) and the inner product (4.8), we obtain the basis of solutions in flat space-time:

$$^{\mathrm{FL}}\phi_{\underline{k}}(x) = (2\pi)^{-3/2}(2\kappa)^{-1/2}\exp[\mp i(\kappa t + k\mathbf{x})]. \qquad (8.1)$$

Since ϕ is a Weyl scalar of weight $-\frac{1}{2}$ and taking into account Eq. (7.5), the basis of solutions for the metric (7.2) will be

$$\begin{aligned} ^{\mathrm{RW}}\phi_{\underline{k}}(x) &= a^{-1}(t)\,^{\mathrm{FL}}\phi_{\underline{k}}(\eta,\mathbf{x}) \\ &= a^{-1}(t)(2\pi)^{-3/2}(2\kappa)^{-1/2}\exp[\mp i(\kappa\eta + k\mathbf{x})]. \end{aligned} \qquad (8.2)$$

This is the well-known "conformal transplantation" for spin-0, which now we extend to another spins.

B. Neutrino field

We shall consider, as another example, the spin-$\frac{1}{2}$ field. An orthonormal basis of solutions of Eq. (5.26) in flat space-time, on account of the inner product (5.33) is

$$\begin{aligned} ^{\mathrm{FL}}\psi_{\underline{k}}^{(-)}(x) &= (2\pi)^{-3/2}e^{(-)}(\mathbf{n})\exp[-i(\kappa t + k\mathbf{x})], \\ ^{\mathrm{FL}}\psi_{\underline{k}}^{(+)}(x) &= (2\pi)^{-3/2}e^{(+)}(\mathbf{n})\exp[i(\kappa t + k\mathbf{x})], \end{aligned} \qquad (8.3)$$

where $e^{(-)}(\mathbf{n})$ and $e^{(+)}(\mathbf{n})$ are four-component spinors corresponding to negative and positive helicity, respectively.

Since ψ is a Weyl spinor of weight $-\frac{3}{4}$, we deduce, for the metric (7.2),

$$^{\mathrm{RW}}\psi(t,\mathbf{x}) = a^{-3/2}(t)\,^{\mathrm{FL}}\psi(\eta,\mathbf{x}), \qquad (8.4)$$

i.e., that the "conformal transplanted" basis is

$$\begin{aligned} ^{\mathrm{RW}}\psi_{\underline{k}}^{(-)}(x) &= (2\pi a)^{-3/2}e^{(-)}(\mathbf{n})\exp[-i(\kappa\eta + k\mathbf{x})], \\ ^{\mathrm{RW}}\psi_{\underline{k}}^{(+)}(x) &= (2\pi a)^{-3/2}e^{(+)}(\mathbf{n})\exp[i(\kappa\eta + k\mathbf{x})]. \end{aligned} \qquad (8.5)$$

C. Electromagnetic field

Now we shall apply the conformal transplantation method to analyze the electromagnetic field in the Robertson–Walker universe. We shall solve this problem in the two gauges, the temporal gauge (A) and the Gupta–Bleuler gauge (B), and we shall show that in both cases the particle model obtained is the same.

In order to apply the conformal transplantation method, we need to solve the problem in the flat case.

(A) We begin our study in Minkowski space. We take the timelike field $U^\mu = (1,0,0,0)$ in order to reproduce the temporal gauge condition (6.5). In fact, with such a choice of U^μ, (6.54) turns out to be

$$U^\mu A_\mu = A_0 = 0. \qquad (8.6)$$

Then A_0 disappears as a dynamic coordinate and the field equations are

$$\partial^\mu \partial_\mu A_j - \partial^i \partial_j A_i = 0, \qquad (8.7)$$

or

$$\ddot{A}_j - \partial_i\,\partial_i\,A_j + \partial_j\,\partial_i\,A_i = 0. \qquad (8.8)$$

In order to solve (8.8) we postulate the solution

$$\phi_j(k,x) = f_j(k,t)e^{-ikx}. \qquad (8.9)$$

By replacing (8.9) and (8.8) we obtain, for f_j,

$$\ddot{f}_j + \kappa^2(\delta_{ij} - n_i n_j)f_i = 0, \qquad (8.10)$$

where $\kappa = |\mathbf{k}|$, $n_i = k_i/k$, and $\dot{\ } = d/dt$.

If we take for convenience $k = (0,0,-k)$, Eqs. (8.10) read

$$\begin{aligned} \ddot{f}_1 + \kappa^2 f_1 &= 0, \\ \ddot{f}_2 + \kappa^2 f_2 &= 0, \\ \ddot{f}_3 &= 0, \end{aligned} \qquad (8.11)$$

Clearly, from (8.11): $f_{1,2} \sim e^{\pm ikt}$ and $f_3 \sim a + b\cdot t$ with a,b constant.

In consequence, we obtain transverse modes (T) and longitudinal ones (L):

$$^{(\mathrm{T})}\phi_i^{(m)}(x,k) = (2\pi)^{-3/2}(2\kappa)^{-1/2} \times \epsilon_i^{(m)}(k)\exp[-i(\kappa t + k\mathbf{x})], \tag{8.12}$$

$$^{(\mathrm{L})}\phi_i(x,k) = (2\pi^{-3/2}(2\kappa)^{-1/2}n_i(f+g\kappa t)\exp(-ik\mathbf{x}).$$

The index m takes the values 1,2 corresponding to the two transverse modes; f and g are dimensionless constants to be determined with the $\epsilon_i^{(m)}(k)$, by demanding on ϕ_i, (a) orthonormality in the inner product, and (b) canonical commutation relations.

Precisely, (a) we demand

$$\begin{aligned}
&\langle {}^{(\mathrm{T})}\phi^i_{(j)}(x,k);\, {}^{(\mathrm{T})}\phi_{i(m)}(x,k')\rangle = -\delta_{jm}\delta(k'-k),\\
&\langle {}^{(\mathrm{L})}\phi^i(x,k);\, {}^{(\mathrm{L})}\phi_i(x,k')\rangle = -\delta(k'-k),\\
&\langle {}^{(\mathrm{T})}\phi^i_{(j)}(x,k);\, {}^{(\mathrm{T})}\phi^*_{i(m)}(x,k')\rangle = 0,\\
&\langle {}^{(\mathrm{L})}\phi^i(x,k);\, {}^{(\mathrm{L})}\phi^*_i(x,k')\rangle = 0,
\end{aligned} \tag{8.13}$$

where the inner product $\langle\, ;\, \rangle$ is (6.48).

Introducing (8.12) in (8.13), we obtain, after calculation,

$$\begin{aligned}
&\epsilon^{*(j)}\cdot\epsilon^{(m)} = \delta^{jm},\\
&\mathrm{Im}(g^*f) = 1.
\end{aligned} \tag{8.14}$$

(b) We demand the following commutation relations at $t=$ const:

$$[E_i(t,\mathbf{x}); A_j(t,\mathbf{y})] = i\delta_{ij}\delta(\mathbf{x}-\mathbf{y}), \tag{8.15}$$

where $E_i(t,x) = -\dot{A}_i(t,x)$, and all others $=0$.

Then we write the field A_i as

$$A_i(t,\mathbf{x}) = \int d^3\kappa \left\{\sum_{(j)} [{}^{(\mathrm{T})}\phi_{i(j)}(x,k)a_{(j)}(k) + {}^{(\mathrm{T})}\phi^*_{i(j)}(x,k)a^\dagger_{(j)}(k)] + {}^{(\mathrm{L})}\phi_i(x,k)^{\mathrm{L}}a(k) + {}^{(\mathrm{L})}\phi^*_i(x,k)^{\mathrm{L}}a^\dagger(k)\right\}, \tag{8.16}$$

and we require the usual commutation relations for the creation and annihilation operators:

$$\begin{aligned}
&[a_{(j)}(k), a^\dagger_{(m)}(k')] = \delta_{jm}\delta(k'-k),\\
&[{}^{\mathrm{L}}a(k), {}^{\mathrm{L}}a^\dagger(k')] = \delta(k'-k),
\end{aligned} \tag{8.17}$$

and all others $=0$.

Replacing (8.16) in (8.15), in account of (8.17), we have

$$\begin{aligned}
&\sum_m \left[\epsilon_i^{(m)}(k)\overset{*}{\epsilon}{}_j^{(m)}(k) - \overset{*}{\epsilon}{}_i^{(m)}(-k)\epsilon_d^{(m)}(-k)\right] = 0,\\
&\frac{1}{2}\sum_m \left[\epsilon_i^{(m)}(k)\overset{*}{\epsilon}{}_j^{(m)}(k) + \overset{*}{\epsilon}{}_i^{(m)}(-k)\epsilon_j^{(m)}(-k)\right] = \delta_{ij} - n_i n_j.
\end{aligned} \tag{8.18}$$

Then, the requirements (a) and (b) imply

$$\begin{aligned}
&\overset{*}{\epsilon}{}^{(j)}(k)\cdot\epsilon^{(m)}(k) = \delta^{jm},\\
&\sum_m \left[\epsilon_i^{(m)}(k)\epsilon_j^{*(m)}(k) - \overset{*}{\epsilon}{}_i^{(m)}(-k)\epsilon_j^{(m)}(-k)\right] = 0,\\
&\frac{1}{2}\sum_m \left[\epsilon_i^{(m)}(k)\epsilon_j^{*(m)}(k) + \overset{*}{\epsilon}{}_i^{(m)}(-k)\epsilon_j^{(m)}(-k)\right] = \delta_{ij} - n_i n_j,\\
&I_m(g^*f) = 1.
\end{aligned} \tag{8.19}$$

Note that

$$\begin{aligned}
&{}^{(\mathrm{T})}\dot{\phi}_i^{(m)}(x,k) = -i(2\pi^{-3/2})(k/2)^{1/2}\epsilon_i^{(m)}(\kappa) \times\exp[-i(\kappa t + k\mathbf{x})],\\
&{}^{(\mathrm{L})}\dot{\phi}_i(x,k) = g(2\pi)^{-3/2}(k/2)^{1/2}n_i\exp(-\mathrm{i}k\mathbf{x}).
\end{aligned} \tag{8.20}$$

In consequence, the unphysical modes (L) can be eliminated *a posteriori* by taking

$$g = -i\beta, \tag{8.21}$$

where we shall take $\beta\to 0$ after calculations.[11]

Moreover, in order to satisfy (7.19), we must take

$$f = 1/\beta. \tag{8.22}$$

Finally, we can take, as the basis of solutions of (8.8),

$$\begin{aligned}
&{}^{(\mathrm{T})}\phi_\mu^{(1)}(x,k) = (2\pi)^{-3/2}(2\kappa)^{-1/2} \times(0,-1,0,0)\exp[-i(\kappa t + k\mathbf{x})],\\
&{}^{(\mathrm{T})}\phi_\mu^{(2)}(x,k) = (2\pi)^{-3/2}(2\kappa)^{-1/2} \times(0,0,-1,0)\exp[-i(\kappa t + k\mathbf{x})],\\
&{}^{(\mathrm{L})}\phi_\mu(x,k) = (2\pi)^{-3/2}(2\kappa)^{-1/2}((1/\beta) - i\beta\kappa t) \times(0,0,0,-1)\exp(-ik\mathbf{x}),
\end{aligned} \tag{8.23}$$

and its conjugate complex.

Since A_μ is a Weyl vector of weight 0, the corresponding solutions for the metric (6.2) will be, in coordinates t, x, y, z,

$$\begin{aligned}
&{}^{\mathrm{RW(T)}}\phi_\mu^{(1)}(x,k) = (2\pi)^{-3/2}(2\kappa)^{-1/2} \times(0,-1,0,0)\exp[-i(\kappa\eta + k\mathbf{x})],\\
&{}^{\mathrm{RW(T)}}\phi_\mu^{(2)}(x,k) = (2\pi)^{-3/2}(2\kappa)^{-1/2} \times(0,0,-1,0)\exp[-i(\kappa\eta + k\mathbf{x})],\\
&{}^{\mathrm{RW(L)}}\phi_\mu(x,k) = (2\pi)^{-3/2}(2\kappa)^{-1/2}((1/\beta) - i\beta\kappa\eta) \times(0,0,0,-1)\exp(-ik\mathbf{x}),
\end{aligned} \tag{8.24}$$

and its conjugate complex.

(B) For the Gupta–Bleuler method, in Minkowski space, and taking as in (A), $U^\mu = (1,0,0,0)$, we obtain, from (6.53),

$$\partial^\mu\partial_\mu A_\nu = 0. \tag{8.25}$$

The solutions are orthonormalized by means of

$$\begin{aligned}
&\langle \phi^\mu(k,s); \phi_\mu(k',s')\rangle = -\eta_{ss'}\delta(k-k'),\\
&\langle \phi^\mu(k,s); \phi^*_\mu(k',s')\rangle = 0,\\
&\langle {}^*\phi^\mu(k,s); {}^*\phi_\mu(k',s')\rangle = \eta_{ss'}\delta(k-k'),
\end{aligned} \tag{8.26}$$

where $\eta_{ss'}$ is the Minkowski tensor ($\eta_{00} = -\eta_{11} = -\eta_{22} = -\eta_{33} = 1$ and $\eta_{ss'} = 0, \forall s' \neq s$) and the inner product is (6.49), which in this case takes the form

$$\langle W^\mu; V_\mu \rangle = i \int_\Sigma (V^\mu \partial_\nu \overset{*}{W}{}^\mu - \overset{*}{W}{}^\mu \partial_\nu V_\mu) d\sigma^\nu. \tag{8.27}$$

We postulate for (8.16) the solution, as in (A),

$$\phi_\nu(k,s;x) = \epsilon_\nu(k,s)\exp(-i\kappa_\mu x^\mu). \tag{8.28}$$

By replacing (8.19) in (8.17) and performing the integration on a surface $t = \text{const}$, we obtain the following conditions on ϵ_ν:

$$\epsilon^\mu(k,s)\overset{*}{\epsilon}_\mu(k,s') = \eta_{ss'}/(2\pi)^3 2\kappa. \tag{8.29}$$

Then we can take, as the basis of solutions of (8.16) [we take $\mathbf{k} = (0,0,k)$],

$$\begin{aligned}
{}^{\mathrm{FL}}\phi_\mu(k,0) &= (2\pi)^{-3/2}(2\kappa)^{-1/2}(1,0,0,0) \\
&\quad \times \exp[-i(\kappa t + k\mathbf{x})], \\
{}^{\mathrm{FL}}\phi_\mu(k,1) &= (2\pi)^{-3/2}(2\kappa)^{-1/2}(0,-1,0,0) \\
&\quad \times \exp[-i(\kappa t + k\mathbf{x})], \\
{}^{\mathrm{FL}}\phi_\mu(k,2) &= (2\pi)^{-3/2}(2\kappa)^{-1/2}(0,0,-1,0) \\
&\quad \times \exp[-i(\kappa t + k\mathbf{x})], \\
{}^{\mathrm{FL}}\phi_\mu(k,3) &= (2\pi)^{-3/2}(2\kappa)^{-1/2}(0,0,0,-1) \\
&\quad \times \exp[-i(\kappa t + k\mathbf{x})],
\end{aligned} \tag{8.30}$$

and its conjugate complex.

The unphysical modes $s = 0,3$ are removed from the physical states, by means of the gauge condition (6.55), which in this case is simply (6.8).[7]

Now, we obtain the corresponding solutions for the Robertson–Walker metric by performing the conformal transplantation.

So, in coordinates (t, x, y, z), we have

$$\begin{aligned}
{}^{\mathrm{RW}}\phi_\mu(k,0) &= a^{-1}(t)(2\pi)^{-3/2}(2\kappa)^{-1/2}(1,0,0,0)\exp[-i(\kappa\eta + k\mathbf{x})], \\
{}^{\mathrm{RW}}\phi_\mu(k,1) &= (2\pi)^{-3/2}(2\kappa)^{-1/2}(0,-1,0,0)\exp[-i(\kappa\eta + k\mathbf{x})], \\
{}^{\mathrm{RW}}\phi_\mu(k,2) &= (2\pi)^{-3/2}(2\kappa)^{-1/2}(0,0,-1,0)\exp[-i(\kappa\eta + k\mathbf{x})], \\
{}^{\mathrm{RW}}\phi_\mu(k,3) &= (2\pi)^{-3/2}(2\kappa)^{-1/2}(0,0,0,-1)\exp[-i(\kappa\eta + k\mathbf{x})].
\end{aligned} \tag{8.31}$$

Of course, in this case, the unphysical modes $s = 0,3$ are removed, in the same way as in the flat space, by means of the condition (6.55), which takes the form

$$(\partial_\mu A^\mu + HA_0)^-|\text{Phys}\rangle = 0, \tag{8.32}$$

where $H = a^{-1}\, da/dt$ is the Hubble coefficient.

The demonstration is unnecessary on account of the conformal invariance of condition (6.55). In fact, all the flat space-time formalism can be conformally mapped to the Robertson–Walker universe.

So, we obtain in the cases (A) and (B) the same particle model for the Robertson–Walker universe.

These results have been possible because we dispose of a fully conformal theory (field equations, inner products, and gauge conditions are conformally invariant) and it explains the inconsistency of Ref. 3 where the theory was not conformally invariant. We believe that this is the cause of the inconsistency, not the quantum equivalence principle; in fact, it seems to us quite impossible to formulate a nonconformal quantum theory for massless particles (at least without ghosts).

Another important feature of this theory is the fact that the introduction of the timelike vector field U^μ in the formalism is necessary in order to obtain the elimination of the unphysical photons. Also, on account of condition (6.55), we obtain, for Eq. (6.53),

$$\langle \text{Phys}|\nabla^\mu F_{\mu\nu}|\text{Phys}\rangle = 0. \tag{8.33}$$

Thus, the field U^μ disappears on average, in the Gupta–Bleuler method, from the field equations. It seems that the field U^μ is necessary for us to perform the quantization in curved space-time, while classically it is irrelevant.

IX. CONCLUSIONS

We see that the conformal model, known for a spin-0 massless field, can be extended to higher spins.

For the spin-1 case, however, it is necessary to introduce a timelike vector field U^μ. We can perform this generalization in two ways: the temporal gauge method and the Gupta–Bleuler method. Both methods yield the same particle model when they are applied to the Robertson–Walker universe. Besides, the field U^μ disappears on average.

APPENDIX: UNIQUENESS OF Λ_μ

We want to determine the expression of Λ_μ that verifies

$$\{\Lambda_\mu, \gamma^\mu\} = 0, \tag{A1}$$

$$[\Lambda_\mu, \gamma_\nu] = \tfrac{1}{2}(\gamma_\mu\, \partial_\nu \ln\lambda - g_{\mu\nu}\, \gamma^\rho\, \partial_\rho \ln\lambda), \tag{A2}$$

where [,] and { , } are the commutator and anticommutator, respectively.

In order to obtain Λ_μ we write Λ_μ as a Clifford expansion:

$$\Lambda_\mu = \partial_\mu + b_\mu\, \gamma^5 - ic_{\mu\nu}\, \gamma^\nu + d_{\mu\nu}\, \gamma^\nu\gamma^5 - iE_{\mu\rho\nu}\sigma^{\rho\nu}, \tag{A3}$$

where γ^5 is the matrix defined in (5.37) and $\sigma_{\mu\nu} = \frac{1}{2}[\gamma_\mu, \gamma_\nu]$. The coefficients $a_\mu, b_\mu, c_{\mu\nu}, d_{\mu\nu}$, and $E_{\mu\rho\nu}$ must be determined by means of (A1) and (A2).

For calculations it is convenient to take at every point of V_4 an orthonormal base, where $g_{\mu\nu}$ = diagonal matrix.

After a straightforward, but tedious, calculation using the properties of Dirac matrices, we obtain from (A1), (A2), and (A3) the following values:

$$a_\mu = b_\mu = c_{\mu\nu} = d_{\mu\nu} = 0,$$
$$E_{\mu\nu\rho} = -(i/8)(g_{\mu\rho}\,\partial_\nu \ln\lambda - g_{\mu\nu}\,\partial_\rho \ln\lambda), \tag{A4}$$

and then, introducing (A4) in (A3), it follows that

$$\Lambda_\mu = -\tfrac{1}{4}\sigma_{\mu\nu}\nabla^\nu \ln\lambda = \tfrac{i}{8}(\gamma_\nu\,\gamma_\mu - \gamma_\mu\,\gamma_\nu)\nabla^\nu \ln\lambda. \tag{A5}$$

Therefore, the uniqueness of (5.25) is proved.

[1]N. D. Birrel and P. C. W. Davies, *Quantum Fields in Curved Space* (Cambridge, U.P., London, 1982).

[2]M. Castagnino, Ann. Inst. H. Poincaré **35**, 55 (1981).

[3]H. Ceccatto, A. Foussats, H. Giacomini, and O. Zandron, J. Math. Phys. **23**, 1865 (1982).

[4]H. G. Loos, Nuovo Cimento **30**, 901 (1963).

[5]T. Kugo and I. Ojima, Progr. Theor. Phys. Suppl. **66**, 1 (1979).

[6]M. Creutz, Ann. Phys. (NY) **117**, 471 (1979).

[7]C. Itzykson and J. B. Zuber, *Quantum Field Theory* (McGraw–Hill, New York, 1980).

[8]We believe that this vector field could have a physical interpretation; it could be considered as the field of velocities of a set of infinite observers covering all space-time. In fact, our final objective will be that the formalism removes the scalar and longitudinal photons (without physical sense) which cannot be defined in a covariant way in curved space-time if we do not define a time direction, which we may think of as the speed of the observer at each point of the manifold. We assimilate the field of observers to an ideal referential fluid that fills the manifold, [G. Cattaneo, Nuovo Cimento **10**, 318 (1958)], which is defined by the set of space-time trajectories of the particles. Then, U will be the unitary vector tangent at each point of the fluid, which will be a timelike vector. We shall try to clarify this interpretation with other examples in forcoming papers.

[9]The relativistic covariance of Eq. (6.14) will be assured by imposing, as we shall see, the Gauss law $\nabla_\mu F^{\mu 0} = 0$. Then, we shall have the equation $\nabla_\mu F^{\mu\nu} = 0$, which is obviously covariant.

[10]Let us note that (6.51) has conformal consistency. In fact, because $V^\mu D_\nu W^*_\mu - \mathring{W}^\mu D_\nu V_\mu$ is a Weyl vector of weight -1, we deduce, from (2.19),

$$\left[\nabla^\nu(V^\mu D_\nu W^*_\mu - \mathring{W}^\mu D_\nu V_\mu)\right]^\sim = \lambda^{-2}\nabla^\nu(V^\mu D_\nu \mathring{W}_\mu - \mathring{W}^\mu D_\nu V_\mu),$$

and therefore the condition (6.51) is conformally invariant. Clearly, the same property holds for (6.45).

[11]When we take $\beta \to 0$ we obtain Gauss's law, div $E = 0$. In fact, from Eqs. (8.16), (8.20), and (8.21) it follows that

$$\operatorname{div} \mathbf{E}(t,\mathbf{x}) = -\beta \int d^3\mathbf{k}(2\pi)^{-3/2}\kappa(\kappa/2)^{1/2} \times\{{}^{\mathrm{L}}a(k)e^{-ikx} + {}^{\mathrm{L}}a^\dagger(k)e^{ikx}\}.$$

Then, to impose Gauss's law we must take the limit $\beta \to 0$.

Compatibility of weak rigidity with some types of elastic schemes

V. del Olmo[a)] and J. Olivert
Departament de Mecànica i Astronomia, Facultat de Matemàtiques, Burjassot, València, Spain

(Received 25 June 1985; accepted for publication 13 November 1985)

The behavior of the hypoelastic-Synge, hypoelastic-Maugin, and hypoelastic-Carter and Quintana almost-thermodynamic material schemes, under weak rigidity hypotheses, is studied. In every case, the absence of principal transverse shock waves (or the vanishing of the corresponding speeds) is obtained. The same result follows for the longitudinal shock waves when the Lamé coefficient μ does not vanish. A definition of an elastic almost-thermodynamic material scheme based on the Fermi–Walker transport is proposed and compared with the above elastic schemes. The speeds of the principal shock waves associated are attained and its compatibility with the Ferrando–Olivert incompressibility condition is proved. In the presence of weak rigidity the elastic schemes here defined lead (assuming $\mu \neq 0$) to the Born-rigidity condition.

I. INTRODUCTION

Controversy exists between different authors about the way in which the theory of elasticity in general relativity must be formulated. Thus, for instance, Synge[1] refuses to establish a linear relation between a strained state (with respect to an initial one) and the stress that produces the deformation; consequently, he postulates a linear relation between rate of strain and rate of change of stress by using the covariant derivative. Rayner,[2] in order to define the strain tensor, introduces two metrics: the one corresponding to the strained body and another that describes its "natural state," i.e., the state of the elastic body previous to the deformation; the latter treatment follows directly the classical one.

Nevertheless, in Rayner's work, the dimensions of the elastic body do not take part, unlike the classical Hooke's law. Choquet-Bruhat and Lamoureux-Brousse,[3] Carter and Quintana,[4] and Maugin[5] partially solve this deficiency by inserting the term

$$\frac{\text{stress}}{\text{mass-energy density}}$$

in their respective formulations.

Reference 3 deals with a formalization of Cattaneo's work[6] about the "state of reference" and introduces a Hooke's law which in Ref. 4 is obtained as a consistency condition for the equations of motion. In this way, Rayner's work is generalized, given that the Carter and Quintana material structures contain, in an implicit manner, the unstrained state. Also, the use of the formalism of the convected derivative[7] is implicit, as Oldroyd proves, in Ref. 2.

A Hooke's law relates the stress with the increase of length without presence of any effect due to the internal rotations. This fact is not considered in the generalized laws quoted above; to take it into account is one of the purposes of this paper.

If we wish to formulate a Hooke's law in the manner explained, we must remove the rotational effects due to the different observers that describe the worldlines of our scheme. With this intention, the use of the Fermi–Walker transport seems suitable to us.

In Sec. II we establish a Hooke's law without deformant torsions in which the rate of change of the stress is given by means of the Fermi–Walker transport [Eq. (2.8)]. Then we express this law in the form in which the elastic schemes are usually presented in the literature [Eqs. (2.11) and (2.15)] and compare it with the ones presented by Maugin and by Carter and Quintana. What is deduced from our study is that the formulation here suggested appears as an intermediate stage between both of them and does not contain terms in the rotations.

In Sec. III we study the compatibility of the elastic schemes defined by Synge, Maugin, and Carter and Quintana with the weak rigidity presented in Ref. 8. This compatibility must be considered from the kinematical point of view since the concept of a material scheme evolves from a given continuous medium. In all the cases analyzed we obtain absence of principal transverse shock waves (or vanishing of their speeds) but we do not get information with respect to the longitudinal ones except for the case in which the Lamé coefficient μ does not vanish.

Section IV studies the behavior of the elastic schemes introduced in Sec. II and their compatibility with the Ferrando–Olivert incompressibility condition[9] and with the weak rigidity.[8] Previously to this work, we obtain (Proposition 4.1) the expressions for the speeds of the principal longitudinal and transverse shock waves in such schemes. The results attained lead us (Proposition 4.2) to positive outcomes with respect to the compatibility with the incompressibility. Finally, Proposition 4.3 collects the results obtained. If we join the weak rigidity hypotheses to Eq. (2.15) we get Born rigidity when the Lamé coefficient μ does not vanish.

Everywhere in this work we consider that the space-time manifold M will be pseudo-Riemannian, connected, and of the Hausdorff type, endowed with a metric hyperbolic tensor field g of signature (3,1) and with a linear connection ∇, compatible with g, and without torsion. (Second-order Christoffel symbols will be noted as $\Gamma^{\cdot}_{\cdot\,\cdot}$.)

According to Ref. 10, an almost-thermodynamic material scheme defined in M is a domain D of the space-time manifold in which a second-order energy-momentum tensor T is defined that is normal and such that if u is its four-velocity vector and $-\rho$ ($\rho > 0$) the associated eigenvalue, we have the decomposition

$$\rho = r(1+\epsilon), \tag{1.1}$$

[a)] Present address: Catedra de Matematicas, E. U. de Informática, Universidad Politecnica de València, València, Spain.

 0022-2488/86/041048-08$02.50

ρ being the proper mass-energy density of the scheme D, r its matter density, and ϵ its specific internal energy.

By defining the spatial projector γ by means of the equation

$$\gamma = g + u \otimes u, \tag{1.2}$$

the energy-momentum tensor is expressed as

$$T = \rho(u \otimes u) + t, \tag{1.3}$$

where t is the relativistic stress tensor (projection of T by means of γ). Spatial tensors will be the orthogonal ones to the four-velocity u.

Concerning the tensorial expressions, Latin indices will be meant to range from 1 to 4 and Greek indices will range from 1 to 3.

The tangent fiber bundle of the space-time manifold M will be noted as TM and, for every $p \in M$, T_pM will symbolize the tangent space at p.

The rate strain tensor d and the rotation tensor Ω are defined, respectively, by means of the tensor symmetrization and skew symmetrization

$$e_{ij} = \gamma^k{}_i \gamma^l{}_j \nabla_k u_l, \tag{1.4}$$

d being also expressed as

$$d = \tfrac{1}{2} L_u \gamma, \tag{1.5}$$

where L_u symbolizes the Lie derivative with respect to u. Also, the symbols F_u and C_u will be used, respectively, for the Fermi derivative and for the convected derivative.

Since we apply in Secs. III and IV the method of the Hadamard discontinuities, we will say that the notations are the same as in Ref. 10. Notwithstanding, let us say that U symbolizes the speed of the corresponding shock waves, λ^i the components of the shock waves spatial propagation direction vector, and δ the infinitesimal discontinuity of the tensor on which it acts. The Maugin projector is defined as

$$S = \gamma - \lambda \otimes \lambda, \tag{1.6}$$

and we get the decomposition

$$\delta_u{}^i = \delta u^i_\perp + \lambda^i \delta u_\parallel, \tag{1.7}$$

where

$$\delta u^i_\perp = S^i{}_j \delta u^j \tag{1.8}$$

and

$$\delta u_\parallel = \lambda_i \delta u^i. \tag{1.9}$$

Moreover, when we use principal shock waves, the corresponding directions will be given by $\lambda^i, d^i{}_{(\alpha)}$ (α:2,3) and, in this case,

$$\delta u^i_\perp = a_2 d^i{}_{(2)} + a_3 d^i{}_{(3)}. \tag{1.10}$$

Finally, let us say that everywhere throughout this paper, the units have been chosen so that the light velocity in the vacuum has the constant value 1.

II. HOOKE'S LAW AND FERMI DERIVATIVE

With the purpose of contrasting our definition of weak rigidity in elastic almost-thermodynamic material schemes now we intend to give a definition of elastic scheme according to the evolution of observable magnitudes.

First we reformulate the Fermi derivative from the Fermi connection[11,12] in the usual form in which the derivation operators are presented.

From the system of differential equations determined by

$$F_u X = 0, \tag{2.1}$$

X being a vector field defined on a congruence of streamlines c of the scheme D, we obtain, in a coordinate neighborhood of a local chart in which $X = (X^1, X^2, X^3, X^4)$, $c = (c^1, c^2, c^3, c^4)$, and $u^i = dc^i/d\tau$,

$$\frac{dX^i}{d\tau} + A^i{}_k X^k(c) = 0, \tag{2.2}$$

where

$$A^i{}_k = \Gamma^i_{kj} u^j - g_{kl}(u^i \nabla_u u^l - u^l \nabla_u u^i). \tag{2.3}$$

Let us define now the vector spaces

$$E = \{X \cdot c \text{ that verify Eq. (2.1)}\}$$

and

$$E_\tau = \{X(c(\tau)) \in T_{c(\tau)} D / X \cdot c \in E\}.$$

The linear map

$$\mu_\tau : E \to E_\tau,$$

defined by $\mu_\tau(X) = X(c(\tau))$ is, by virtue of the existence and uniqueness theorem for the solutions of the system given by Eq. (2.1), a bijection.

The Fermi–Walker transport is defined as

$$\mathcal{T}_{\tau_1,\tau_2} = \mu_{\tau_2} \circ \mu_{\tau_1}^{-1} : E_{\tau_1} \to E_{\tau_2},$$

an isomorphism that verifies the properties

(1) $\mathcal{T}_{\tau,\tau} = I_{T_{c(\tau)}} M$,

(2) $\mathcal{T}_{\tau,\tau'} = \mathcal{T}^{-1}_{\tau',\tau}$,

(3) $\mathcal{T}_{\tau,\tau'} \circ \mathcal{T}_{\tau'\tau''} = \mathcal{T}_{\tau,\tau''}$.

Moreover, if we define $\mathcal{T}_\tau = \mathcal{T}_{0,\tau}$, we get

(4) $F_u X = 0$ implies $\mathcal{T}_\tau X(c(0)) = X(c(\tau))$.

For more details on this process see Refs. 13–15.

Now, it is possible to define the Fermi derivative of the vector field X with respect to the four-velocity u of the scheme in $c(0)$, as

$$\lim_{\tau \to 0} \left\{ \frac{\mathcal{T}_\tau^{-1}(X(c(\tau)) - X(c(0)))}{\tau} \right\}, \tag{2.4}$$

if this limit exists. Besides, we obtain that the expression given by (2.4) coincides with $(F_u X)_{c(0)}$. Thus, according to property (4), a necessary and sufficient condition for a vector field X to be Fermi–Walker transported along the streamlines of the scheme is the vanishing of $F_u X$.

The process followed above is generalized, in a natural way, for the tensor fields and we could obtain a transport that extends the one given previously and allows us to propose an expression like (2.4) for the Fermi derivative, with respect to the four-velocity u of the scheme, of the tensor fields. This form will be used to establish our relativistic Hooke's law.

Every classical Hooke's law expresses the change of length proportionally to the stress, which, for a stick of length l and section s, is given by

$$l\cdot F/s = k\cdot \Delta l, \tag{2.5}$$

where F/s is the normal stress acting per unit of surface and k is the elastic constant. Equation (2.5) can be expressed, equivalently, as

$$V\cdot F/s = K\cdot \Delta l, \tag{2.6}$$

V being the volume of the stick and K the product $k\cdot s$.

In relativity we will proceed in the following way. Let us consider, at first, the congruence of timelike worldlines in which the deformation takes place. If φ_τ is the flow that u generates,[16] the change of length in each direction could be determined by means of

$$\varphi^*_\tau\gamma - \gamma \tag{2.7}$$

and, according to the Introduction, the Fermi–Walker transport will be used to formulate the term corresponding to the stress.

As is known, the mass is conserved in the almost-thermodynamic material schemes; so, we will take volumes with unit mass, i.e., equaling $1/r$.

Taking into account Eq. (2.6), we postulate the relativistic Hooke's law

$$\mathscr{T}_\tau^{-1}(t/r) - (t/r) = C'(\varphi^*_\tau\gamma - \gamma), \tag{2.8}$$

in which C' is an elastic tensor field of the usual type.[4,5] In particular, it will be a fourth-order spatial tensor field whose components are C'_{ijkl} in some coordinate basis.

Equation (2.8) is susceptible to transformation by dividing both sides of it by τ and taking limits. Thus,

$$\lim_{\tau\to 0}\left\{\frac{\mathscr{T}_\tau^{-1}(t/r) - (t/r)}{\tau}\right\} = C'\cdot\lim_{\tau\to 0}\left\{\frac{\varphi^*_\tau\gamma - \gamma}{\tau}\right\}, \tag{2.9}$$

which, having in mind the generalization of (2.4) and the concept of the Lie derivative, leads us to

$$F_u(t/r) = C'(L_u\gamma), \tag{2.10}$$

or, by virtue of the definition of the strain rate tensor, to

$$F_u(t/r) = C(d), \tag{2.11}$$

where $C = 2C'$.

Equation (2.11) expresses the change of the relativistic stress tensor per unit of volume by means of the change of the spatial metric. The elastic tensor C could have the form of the Maugin zero-order hypoelastic tensor[5] and so it will be considered henceforth.

At this moment our purpose is to establish local expressions equivalent to the one given by Eq. (2.11) and compare them with another used in the literature.

If we develop the left side of Eq. (2.11), we will get

$$\begin{aligned} F_u(t/r) &= (1/r)F_u t + t\cdot F_u(1/r)\\ &= (1/r)F_u t + t\cdot u(1/r)\\ &= (1/r)F_u t - (t/r^2)\nabla_u r. \end{aligned} \tag{2.12}$$

After considering the continuity equation

$$\nabla_u r = -rd^i_{\ i} \tag{2.13}$$

and replacing it in Eq. (2.12), the result is

$$F_u(t/r) = (1/r)F_u t + (t/r)d^i_{\ i} \tag{2.14}$$

and we deduce that every elastic almost-thermodynamic material scheme defined by Eq. (2.11) verifies

$$F_u t_{ij} + t_{ij}d^c_{\ c} = rC_{ijkl}d^{kl} = J_{ijkl}d^{kl}. \tag{2.15}$$

At first sight, Eq. (2.15) contains on one hand a change of the stress of Synge type[1] while the other terms remind us of the Maugin and Carter and Quintana elastic schemes.

Previously to a more detailed comparison between the different types of elastic schemes, we study the general relation between the Fermi and the convected derivatives, the last one introduced by Oldroyd[7] to establish rheological invariance for a set of state equations in general relativity and used by Carter and Quintana to formulate their elasticity theory for perfect solids.

Lemma 2.1: Let s be a tensor field of type (a,b) defined in an almost-thermodynamic material scheme D of the space-time manifold M. The following general relation between the Fermi and convected derivatives of s with respect to the four-velocity u of the scheme, holds:

$$\begin{aligned} &(F_u s)^{i_1\cdots i_a}_{j_1\cdots j_b} - (C_u s)^{i_1\cdots i_a}_{j_1\cdots j_b}\\ &\quad = \sum_{k=1}^{a} s^{i_1\cdots i_{k-1}ci_{k+1}\cdots i_a}_{j_1\cdots\cdots\cdots\cdots j_b}(\Omega^{i_k}_{\ c} + d^{i_k}_{\ c})\\ &\qquad - \sum_{k=1}^{b} s^{i_1\cdots\cdots\cdots\cdots i_a}_{j_1\cdots j_{k-1}cj_{k+1}\cdots j_b}(\Omega^{c}_{\ j_k} + d^{c}_{\ j_k}). \end{aligned} \tag{2.16}$$

Proof: For a given local chart, the local expression of $C_u s$ is given by[4]

$$\begin{aligned} &(C_u s)^{i_1\cdots i_a}_{j_1\cdots j_b}\\ &\quad = (\nabla_u s)^{i_1\cdots i_a}_{j_1\cdots j_b}\\ &\qquad - \sum_{k=1}^{a} s^{i_1\cdots i_{k-1}ci_{k+1}\cdots i_a}_{j_1\cdots\cdots\cdots\cdots j_b}(\nabla_c u^{i_k} + u^{i_k}\nabla_u u_c)\\ &\qquad + \sum_{k=1}^{b} s^{i_1\cdots\cdots\cdots\cdots i_a}_{j_1\cdots j_{k-1}cj_{k+1}\cdots j_b}(\nabla_{j_k}u^c + u^c\nabla_u u_{j_k}). \end{aligned} \tag{2.17}$$

On the other hand, from the Fermi connection F_uX, given by Sachs and Wu,[11] and after expressing the Fermi derivative for tensor fields in a local form (a way of making this would be to apply the process pointed out by the Willmore theorem[17]), we have

$$\begin{aligned} &(F_u s)^{i_1\cdots i_a}_{j_1\cdots j_b}\\ &\quad = (\nabla_u s)^{i_1\cdots i_a}_{j_1\cdots j_b}\\ &\qquad + \sum_{k=1}^{a} s^{i_1\cdots i_{k-1}ci_{k+1}\cdots i_a}_{j_1\cdots\cdots\cdots\cdots j_b}(u_c\nabla_u u^{i_k} - u^{i_k}\nabla_u u_c)\\ &\qquad + \sum_{k=1}^{b} s^{i_1\cdots\cdots\cdots\cdots i_a}_{j_1\cdots j_{k-1}cj_{k+1}\cdots j_b}(u^c\nabla_u u_{j_k} - u_{j_k}\nabla_u u^c). \end{aligned} \tag{2.18}$$

From Eqs. (2.17) and (2.18) we obtain Eq. (2.16) by subtracting them and after using the definition of d_{ij} and Ω_{ij}; i.e., taking into account the relation

$$\nabla_j u_i = d_{ij} + \Omega_{ij} - u_j\nabla_u u_i.$$

The general relation (2.16) shows us that the convected and the Fermi derivatives are not independent operators. The first of them is simple when it is calculated in properly comoving reference systems.[7] Moreover, the Fermi derivative of the metric tensor field vanishes so the process of raising and lowering indices is easily performed.

Let us consider now the different types of elastic schemes.

The Maugin Hooke's law[5] has the form

$$\gamma^k{}_i\gamma^l{}_j L_u t_{kl} + t_{ij} d^c{}_c = J_{ijkl} d^{kl}, \tag{2.19}$$

which differs from the one expressed by Eq. (2.15) in the term containing the Lie derivative of the relativistic stress tensor.

Let us symbolize by M_{ij} the components of the second-order tensor field given by the left hand of Eq. (2.19):

$$M_{ij} = \gamma^k{}_i\gamma^l{}_j L_u t_{kl} + t_{ij} d^c{}_c. \tag{2.20}$$

Besides, the covariant form of the Hooke's law defined by Carter and Quintana in Ref. 4 is

$$\gamma_{ik}\gamma_{jl} L_u t^{kl} + t_{ij} d^c{}_c = J_{ijkl} d^{kl}, \tag{2.21}$$

which allows us to define the tensor field with components

$$(\mathrm{CQ})_{ij} = \gamma_{ik}\gamma_{jl} L_u t^{kl} + t_{ij} d^c{}_c. \tag{2.22}$$

Finally, let us define, according to Eq. (2.15), the tensor field whose components are

$$E_{ij} = F_u t_{ij} + t_{ij} d^c{}_c \tag{2.23}$$

and let up keep in mind that the term $F_u t_{ij}$ can be expressed in an equivalent way, by virtue of the properties of the Fermi derivative, as

$$\gamma_{ik}\gamma_{jl} F_u t^{kl}.$$

Our purpose is to compare, at first, the Maugin Hooke's law with the one established in this paper. The term

$$\gamma^k{}_i\gamma^l{}_j (L_u t_{kl} - F_u t_{kl})$$

is written, taking into account the spatial covariant character of the relativistic stress tensor t,as[4]

$$\gamma^k{}_i\gamma^l{}_j (C_u t_{kl} - F_u t_{kl}),$$

which, using the general relation (2.16) between the Fermi and the convected derivative, leads to

$$\gamma^k{}_i\gamma^l{}_j (t_{kc}(\Omega^c{}_l + d^c{}_l) + t_{cl}(\Omega^c{}_k + d^c{}_k))$$

or, given that t and Ω are spatial tensor fields, to

$$t_{ic}(\Omega^c{}_j + d^c{}_j) + t_{cj}(\Omega^c{}_i + d^c{}_i)$$
$$= t_i{}^c\Omega_{cj} + t^c{}_j\Omega_{ci} + t_i{}^c d_{cj} + t^c{}_j d_{ci}.$$

Consequently, after defining the second-order tensor fields with components

$$F_{ij} = t_i{}^c\Omega_{cj} + t^c{}_j\Omega_{ci} \tag{2.24}$$

and

$$D_{ij} = t_i{}^c d_{cj} + t^c{}_j d_{ci} \tag{2.25}$$

and using Eqs. (2.20) and (2.23), we can write

$$M_{ij} = E_{ij} + F_{ij} + D_{ij}. \tag{2.26}$$

Let us see the corresponding result for the Carter and Quintana elastic schemes. The term

$$\gamma_{ik}\gamma_{jl} (L_u t^{kl} - F_u t^{kl})$$

is expressed as

$$\gamma_{ik}\gamma_{jl} (C_u t^{kl} + t^{cl} u^k \nabla_u u_c + t^{kc} u^l \nabla_u u_c - F_u t^{kl})$$

from the relation between the convected derivative and the Lie derivative.[4]

The latter expression is transformed, using again Eq. (2.16), into

$$-\gamma_{ik}\gamma_{jl}(t^{kc}(\Omega^l{}_c + d^l{}_c) + t^{cl}(\Omega^k{}_c + d^k{}_c));$$

i.e., into

$$F_{ij} - D_{ij},$$

by virtue of the skew-symmetric character of Ω.

At last, from Eqs. (2.22) and (2.23), we get

$$(\mathrm{CQ})_{ij} = E_{ij} + F_{ij} - D_{ij}. \tag{2.27}$$

Let us remind the reader that the tensor fields F_{ij} and D_{ij} would represent, in each direction, the work produced for the stress and the rotations (power due to the internal rotation)[10] and the work due to the stress and deformations. From Eqs. (2.26) and (2.27), we deduce

$$M_{ij} - (\mathrm{CQ})_{ij} = 2D_{ij}; \tag{2.28}$$

therefore, the Maugin and the Carter and Quintana elastic schemes only differ in the terms due to the deformation and coincide under the Born-rigidity hypothesis ($d = 0$). This result was found in a certain way in Ref. 10 where we obtained identical results in both cases when the strain rate tensor vanished.

On the other hand, from Eqs. (2.26) and (2.27), one deduces

$$E_{ij} = \tfrac{1}{2}(M_{ij} + (\mathrm{CQ})_{ij}) - F_{ij}; \tag{2.29}$$

thus, the elastic almost-thermodynamic material schemes here defined present intermediate characteristics between the Maugin and the Carter and Quintana elastic schemes with the particularity that, according to our initial aim of giving an observational meaning to the relativistic elasticity, our definition does not contain the terms due to the rotations.

III. WEAK RIGIDITY IN ELASTIC SCHEMES

In a former work,[10] the Born-rigidity condition was contrasted with the elastic schemes such as the ones proposed by Synge, Carter and Quintana, and Maugin with the purpose of checking the incompressibility condition and the compatibility with respect to the inexistence of perturbations. The negative results obtained for the hypoelastic-Maugin and for the hypoelastic-Carter and Quintana schemes led us to modify the Born rigidity by adding to the nullity of the strain rate tensor, the vanishing of the spatial change of the relativistic stress tensor. Nevertheless, the restrictions that suppose in relativity the Born-rigidity conditions induced us[8] to propose the definition of weakly rigid almost-thermodynamic material scheme from which some results of the geometric type were obtained.

Note that the weak-rigidity condition contains the vanishing of the spatial change of the relativistic stress tensor ($F_u t_{ij} = 0$). This fact, besides the conservation equations, led us[10] to the absence of shock waves corresponding to the infinitesimal discontinuities of the four-velocity of the scheme.

Our present aim is to check the weak rigidity concept with the elastic schemes proposed by Synge, Maugin, and Carter and Quintana as we announced in Ref. 8. The incompressibility condition is not affected given that the elastic schemes quoted above are compatible with the Ferrando and Olivert condition[9] and, as it is proved in Ref. 8, the weak rigidity leads to the incompressibility.

Consequently, we study the weak rigidity conditions in the elastic almost-thermodynamic material schemes with the purpose of verifying the absence of shock waves due to the discontinuities of the scheme four-velocity or, at least, the vanishing of their speeds. In any case, we will use the method of the Hadamard discontinuities.

With respect to the elastic-Synge almost-thermodynamic material schemes, we take into account the equations

$$\nabla_u t_{ij} = C_{ijkl} d^{kl} \tag{3.1}$$

and

$$\nabla_u r = \nabla_u \epsilon = 0 = F_u t_{ij}, \tag{3.2}$$

i.e., a Synge Hooke's law[10] and the equations defining the weak rigidity.[8] Here, the elastic tensor components C_{ijkl} have the form (isotropic case)

$$C_{ijkl} = -\lambda g_{ij} g_{kl} - \mu (g_{ik} g_{jl} + g_{il} g_{jk}), \tag{3.3}$$

λ and μ being the Lamé coefficients.

By means of the contraction of Eq. (3.1) with $\gamma^i{}_a \gamma^j{}_b$ and after using Eqs. (3.2) and (3.3), we get

$$-\mu(\gamma_a{}^k \gamma_b{}^l + \gamma_a{}^l \gamma_b{}^k) d_{kl} = 0. \tag{3.4}$$

If we now use the definition of the strain rate tensor d_{kl} and take Hadamard discontinuities in it, we obtain the following system in the infinitesimal discontinuities of the four-velocity of the scheme D:

$$\mu(\gamma_{bl} \lambda_a + \gamma_{al} \lambda_b) \delta u^l = 0. \tag{3.5}$$

The symmetric spatial character of the tensor

$$P_{bl} = \gamma_{bl} + \lambda_b \lambda_l \tag{3.6}$$

allows us to apply the methods used in Ref. 10 and to get the system

$$\mu(\gamma_{bl} - \lambda_b \lambda_l) \delta u_\perp^l = 0, \tag{3.7}$$

$$2\mu \delta u_\parallel = 0. \tag{3.8}$$

If $\mu = 0$, the results obtained in Ref. 10 allow us to affirm that the speeds of the principal transverse shock waves (if these shock waves exist) vanish.

If $\mu \neq 0$, from Eq. (3.8) we deduce the absence of longitudinal shock waves. On the other hand, by contracting Eq. (3.7) with the principal directions $d^l{}_{(\alpha)}$ $(\alpha\colon 2,3)$, we obtain straightaway

$$\delta u_\perp = 0$$

and, consequently, the absence of transverse shock waves. Now let us consider the hypotheses of the hypoelastic-Maugin schemes and weak rigidity:

$$\gamma^k{}_i \gamma^l{}_j L_u t_{kl} + t_{ij} k^k{}_k = J_{ijkl} d^{kl}, \tag{3.9}$$

$$\nabla_u r = \nabla_u \epsilon = F_u t_{ij} = 0. \tag{3.10}$$

Proceeding as in the former case (taking into account that weak rigidity implies the vanishing of the expansion velocity scalar $d^k{}_k$) and developing the term containing the Lie derivative, we get

$$\lambda_i t_{aj} \delta u^a + \lambda_j t_{bi} \delta u^b = -r\mu(\lambda_i \delta u_j + \lambda_j \delta u^i); \tag{3.11}$$

and after contracting with $\gamma^i{}_k \lambda^i$ and again for principal shock waves, we obtain the system

$$(t_{ki} + r\mu \gamma_{ki} + (t_\parallel + r\mu)\lambda_k \lambda_i) \delta u^i = 0. \tag{3.12}$$

The spatial and symmetric character of the tensor, which is in brackets, allows us the use of the method of Hadamard, which leads to

$$(t_{ab} + r\mu\gamma_{ab} - (t_\parallel + r\mu)\lambda_a \lambda_b) \delta u_\perp^b = 0, \tag{3.13}$$

$$2(t_\parallel + r\mu)\delta u_\parallel = 0. \tag{3.14}$$

By contracting Eq. (3.13) with the principal directions $d^a{}_{(\alpha)}$ $(\alpha\colon 2,3)$, the former system can be expressed as

$$a_\alpha (t_\alpha + r\mu) = 0 \quad (\alpha\colon 2,3), \tag{3.15}$$

$$(t_\parallel + r\mu)\delta u_\parallel = 0. \tag{3.16}$$

The vanishing of the terms $t_\alpha + r\mu$ $(\alpha\colon 2,3)$ leads us, according to the results of Ref. 10, to the nullity of the speeds of the transverse shock waves (if they exist). If, by means of some kind of physical arguments, we could ensure that the terms $t_\alpha + r\mu$ $(\alpha\colon 2,3)$ would not vanish—let us remind the reader that the principal stresses are divided by the light velocity and the corresponding terms would vanish in the classical limit, so their nullity seems to be reduced to the vanishing of μ—Eq. (3.15) implies the absence of transverse shock waves. Furthermore the condition $t_\parallel + r\mu \neq 0$ together with Eq. (3.16) precludes the existence of longitudinal shock waves.

Finally, a hypoelastic-Carter and Quintana weakly rigid almost-thermodynamic material scheme verifies

$$\gamma^k{}_i \gamma^l{}_j L_u t^{ij} + t_{kl} d^i{}_i = J^{kl}{}_{ab} d^{ab}, \tag{3.17}$$

$$\nabla_u r = \nabla_u \epsilon = F_u t_{ij} = 0. \tag{3.18}$$

If we work in a similar way to the Maugin case and again for principal shock waves, we get the system

$$(t_\parallel - \mu)(\gamma_{ij} - \lambda_i \lambda_j)\delta u_\perp^i = 0, \tag{3.19}$$

$$(t_\parallel - \mu)\delta u_\parallel = 0, \tag{3.20}$$

which, after contracting Eq. (3.19) with $d^j{}_{(\alpha)}$ $(\alpha\colon 2,3)$, gives

$$a_\alpha (t_\parallel - \mu) = 0, \tag{3.21}$$

$$(t_\parallel - \mu)\delta u_\parallel = 0. \tag{3.22}$$

Similarly, if we take into account the results obtained for the speeds of principal shock waves in the hypoelastic-Carter and Quintana almost-thermodynamic material schemes,[10] we get, for $t_\parallel - \mu = 0$, the vanishing of the speeds of the principal transverse shock waves (if they exist); while if $t_\parallel - \mu \neq 0$, from Eqs. (3.21) and (3.22) we deduce the absence of both transverse and longitudinal principal shock waves.

In the three cases studied, we can extract as a general conclusion the absence of the principal transverse shock waves (of the vanishing of their velocities if they exist, which has the same physical meaning). With respect to the principal longitudinal shock waves, they are undetermined if $\mu = 0$. Relatively, we could have in mind some classical considerations as the ones which Landau and Lifshitz make in Ref. 18 and we could assume, in conditions of thermodynamical equilibrium, that the Lamé coefficient μ is positive defined. This fact leads us to the absence of principal shock waves, for every scheme considered, under weak rigidity hypotheses.

IV. BEHAVIOR OF THE ELASTIC ALMOST-THERMODYNAMIC MATERIAL SCHEMES GIVEN BY EQ. (2.15)

Our present purpose is to consider the elastic almost-thermodynamic material schemes as they were defined in Sec. I. According to the general method followed in Ref. 10, we will work with the Hadamard infinitesimal discontinuities to obtain the speeds of the principal shock waves with the purpose of studying the compatibility of such a schemes with the Ferrando–Olivert incompressibility condition. At last, we will study the behavior of these schemes under weak rigidity hypotheses.

Let us consider the Hooke's law given by Eq. (2.15), written in the form

$$\gamma_{ri}\gamma_{lj}\nabla_u t^{ij} + t_{rl} d^c{}_c = rC_{rlij} d^{ij}, \tag{4.1}$$

where the C_{rlij} are the components of the zero-order hypo-elastic-Maugin tensor:

$$C_{rlij} = -\lambda\gamma_{rl}\gamma_{ij} - \mu(\gamma_{ri}\gamma_{lj} + \gamma_{rj}\gamma_{li}). \tag{4.2}$$

By applying Hadamard discontinuities to Eq. (4.1), we get

$$-U\gamma_{ri}\gamma_{lj}\delta t^{ij} + t_{rl}\lambda_i\delta u^i = rC_{rlij}\lambda^j\delta u^i, \tag{4.3}$$

which, contracted with λ^l, gives

$$(-rU^2 f_{ri} + \lambda^l t_{rl}\lambda_i - rC_{rlij}\lambda^l\lambda^j)\delta u^i = 0, \tag{4.4}$$

f_{ri} being the components of the tensor index f of the scheme D:

$$f = (1+\epsilon)\gamma + t/r.$$

If we define

$$Q_{ri} = -rU^2 f_{ri} + \lambda^l t_{rl}\lambda_i - rC_{rlij}\lambda^l\lambda^j, \tag{4.5}$$

taking into account the expression given by Eq. (4.2), we obtain, for principal shock waves,

$$Q_{ri} = -rU^2 f_{ri} + r\mu\gamma_{ri} + (r(\lambda+\mu) + t)\lambda_r\lambda_i. \tag{4.6}$$

We can now state the following proposition.

Proposition 4.1: In every elastic almost-thermodynamic material scheme given by Eq. (2.15), with the former particularizations with respect to the elastic tensor, the speeds of the principal shock waves in the longitudinal and transverse cases are, respectively,

$$U_L^2 = \frac{r(\lambda+2\mu) + t_\parallel}{r(1+\epsilon) + t_\parallel}, \tag{4.7}$$

$$U_{T_\alpha}^2 = \frac{r\mu}{r(1+\epsilon) + t_\alpha} \quad (\alpha{:}\ 2,3). \tag{4.8}$$

Proof: If we consider the expression Q_{ri} given by Eq. (4.6) for the principal shock waves, given its spatial and symmetric character, the system (4.4) is equivalent to

$$Q_{\perp ri}\delta u^i_\perp = 0, \tag{4.9}$$

$$Q_\parallel \delta u_\parallel = 0, \tag{4.10}$$

with

$$Q_{\perp ri} = S^s{}_r S^j{}_i(-rU^2 f_{sj} + r\mu\gamma_{sj}), \tag{4.11}$$

$$Q_\parallel = -(r(1+\epsilon) + t_\parallel)U^2 + r(\lambda+2\mu) + t_\parallel. \tag{4.12}$$

In order to obtain the speeds of the longitudinal shock waves, we assume $\delta u_\parallel \neq 0$, $\delta u_\perp = 0$. Thus, from Eqs. (4.10) and (4.12), we obtain Eq. (4.7).

For the transverse shock waves, if $\delta u_\parallel = 0$ and $\delta u_\perp \neq 0$, Eqs. (4.9) and (4.11) lead us to

$$rS^s{}_r S^j{}_i(U^2 f_{sj} - \mu\gamma_{sj})\delta u^i_\perp = 0, \tag{4.13}$$

which, after contracting with $d^r{}_{(\alpha)}$ $(\alpha{:}2,3)$, gives

$$U^2_{T_\alpha}(r(1+\epsilon)a_\alpha + a_\alpha t_\alpha) - r\mu a_\alpha = 0, \tag{4.14}$$

from which we deduce Eq. (4.8).

The results presented in Eqs. (4.7) and (4.8) generalize the classic ones. If we use an elastic tensor of the Carter and Quintana type,[4] these equations would not contain the matter density r in the numerators.

The following result can be stated with respect to the contrasting of the Ferrando–Olivert incompressibility condition in elastic scheme, which are given by Eq. (2.15).

Proposition 4.2: Every elastic almost-thermodynamic material scheme according to Eq. (2.15), for the principal shock waves, is compatible with the incompressibility condition.

Proof: Let us consider Eq. (4.3) and the expression obtained by taking Hadamard discontinuities in the incompressibility condition[9,10]

$$\gamma_{ri}\gamma_{lj}U\delta t^{ij} = -\gamma_{rl}t^k{}_i\lambda_k\delta u^i + (2t^{rl} + r(1+\epsilon)\gamma^{rl})\lambda_i\delta u^i. \tag{4.15}$$

If we add both of them and contract with λ^l, we get

$$rC_{rlij}\lambda^l\lambda^j\delta u^i - t_{rl}\lambda^l\lambda_i\delta u^i - \lambda_r\lambda_k t^k{}_i\delta u^i + (2t_{rl}\lambda^l + r(1+\epsilon)\lambda_r)\lambda_i\delta u^i = 0, \tag{4.16}$$

which results in

$$R_{ri}\delta u^i = 0, \tag{4.17}$$

where

$$R_{ri} = rC_{rlij}\lambda^l\lambda^j - t_{rl}\lambda^l\lambda_i - t^k{}_i\lambda_r\lambda_k + (r(1+\epsilon)\lambda_r + 2t_{rl}\lambda^l)\lambda_i. \tag{4.18}$$

If we work in the principal shock waves, taking into account the expression for C_{rlij}, which is given by Eq. (4.2), we obtain

$$R_{ri} = r(1+\epsilon-\lambda-\mu)\lambda_r\lambda_i - r\mu\gamma_{ri}, \tag{4.19}$$

which are the components of a spatial and symmetric tensor field. So, the system defined by Eq. (4.17) is equivalent to

$$R_{\perp ri}\delta u^i_\perp = 0, \tag{4.20}$$

$$R_\parallel \delta u_\parallel = 0, \tag{4.21}$$

where

$$R_{\perp ri} = -r\mu S^s{}_r S^j{}_i\gamma_{sj} \tag{4.22}$$

and

$$R_\parallel = r(1+\epsilon) - r(\lambda+2\mu). \tag{4.23}$$

For the longitudinal shock waves ($\delta u_\parallel \neq 0$, $\delta u_\perp = 0$), we obtain

$$R_{\parallel} = 0, \tag{4.24}$$

i.e.,

$$r(1+\epsilon) = r(\lambda + 2\mu), \tag{4.25}$$

that leads us, having in mind Eq. (4.7), to

$$U_L = 1. \tag{4.26}$$

By working in transverse shock waves, we have $\delta u_{\parallel} = 0$, $\delta u_{\perp} \neq 0$, if they exist. From Eqs. (4.20) and (4.22), we deduce

$$r\mu S^s{}_r S^j{}_i \gamma_{sj} \delta u^i_{\perp} = 0 \tag{4.27}$$

or, equivalently,

$$r\mu S^s{}_r \gamma_{sj} \delta u^j_{\perp} = 0, \tag{4.28}$$

by virtue of the spatial character of $\delta u_{\perp}$. From the fact that $S^s{}_r \gamma_{sj} = S_{rj}$, we get

$$r\mu S_{rj} \delta u^j_{\perp} = 0, \tag{4.29}$$

thus,

$$r\mu \delta u_{\perp r} = 0. \tag{4.30}$$

Equation (4.8) allows us to state that the speeds of the principal transverse shock waves $U_{T\alpha}$ (α: 2,3) vanish when the Lamé coefficient $\mu = 0$, while, according to Eq. (4.30), such shock waves do not exist if $\mu \neq 0$. This result leads us, in every case, to the stated compatibility. Positive results would be also obtained by using hypoelastic tensors of the Carter and Quintana type.

To close this section, let us consider the hypotheses of elastic almost-thermodynamic material scheme given by Eq. (2.15):

$$F_u t_{ij} + t_{ij} d^c{}_c = J_{ijkl} d^{kl} \tag{4.31}$$

and the rigidity conditions

$$\nabla_u r = \nabla_u \epsilon = F_u t = 0. \tag{4.32}$$

The use of Eq. (4.32) allows us to write Eq. (4.31) as

$$J_{ijkl} d^{kl} = 0, \tag{4.33}$$

i.e.,

$$r\lambda \gamma_{ij} \gamma_{kl} d^{kl} + r\mu(\gamma_{ik}\gamma_{jl} + \gamma_{il}\gamma_{jk}) d^{kl} = 0, \tag{4.34}$$

which gives, using the spatial character of d and the vanishing of the expansion velocity scalar,[8]

$$r\mu d_{ij} = 0. \tag{4.35}$$

From it, the vanishing of μ allows us to obtain null speeds for the transverse principal shock waves as in the Synge, Maugin, and Carter and Quintana cases. Nevertheless, if we assume μ to be positive defined, our elastic schemes lead far away; thus, we state the following proposition.

Proposition 4.3: Under the hypothesis $\mu > 0$, every elastic almost-thermodynamic material scheme given by Eq. (2.15) defined by means of a zero-order hypoelastic-Maugin tensor, weakly rigid, is Born rigid and, consequently, does not present shock waves due to the infinitesimal discontinuities of its four-velocity u.

Proof: The result follows from Eq. (4.35) and Theorem 4.1 or Ref. 10.

V. DISCUSSION

By using the Fermi–Walker transport, we have proposed in this paper the definition of an elastic almost-thermodynamic material scheme according to the observational requirements in general relativity.

The general relation (given by Lemma 2.1) between the convected and the Fermi derivatives facilitated us in the comparison of our Hooke's law with the ones defined by Carter and Quintana[4] and Maugin.[5]

With the purpose of contrasting the weak rigidity concept given in Ref. 8 in the relativistic elasticity, we worked in Sec. III in the elastic schemes of the Synge, Carter and Quintana, and Maugin types. The results obtained with respect to the behavior of the principal shock waves are undetermined for the longitudinal shock waves if the Lamé coefficient μ vanishes.

Concerning the elastic schemes here defined (Sec. IV), we obtained in Proposition 4.1 the speeds of the principal longitudinal and transverse shock waves and Proposition 4.2 proved the compatibility of such a scheme with the Ferrando and Olivert incompressibility condition. The behavior of these schemes under weak rigidity hypotheses seems better than the one of the elastic schemes above studied: if the Lamé coefficient μ does not vanish, we obtain Born rigidity and, consequently, absence of shock waves of any type. So, in this case, we do not take into account the restriction that supposes the study of principal shock waves.

The latter fact seems to us logical since the elastic schemes defined here are free of rotations and in a former paper[8] we obtained the same conclusion under some additional conditions (irrotationality) on the congruence of the scheme streamlines.

Since the Lamé coefficient μ is positive defined in the classical elasticity theory under thermodynamical equilibrium conditions,[18] taking into account the results obtained in this work, it would be desirable to study in a more detailed manner its character from the relativistic point of view.

[1] J. L. Synge, "A theory of elasticity in general relativity," Math. Z. **72**, 82 (1959).
[2] C. B. Rayner, "Elasticity in general relativity," Proc. R. Soc. London Ser. A **272**, 44 (1963).
[3] Y. Choquet-Bruhat and L. Lamoureux-Brousse, "Sur les équations de l'élasticité relativiste," C. R. Acad. Sci. Paris A **276**, 1317 (1973).
[4] B. Carter and H. Quintana, "Foundations of general relativistic high-pressure elasticity theory," Proc. R. Soc. London Ser. A **331**, 57 (1972).
[5] G. A. Maugin, "Ondes infinitésimales principales dan les milieux élastiques précontraints en relativité," C. R. Acad. Sci. Paris A **284**, 393 (1977).
[6] C. Cattaneo, "Essai d'une théorie relativiste de l'élasticité," C. R. Acad. Sci. Paris A **272**, 1421 (1971).
[7] J. G. Oldroyd, "Equations of state of continuous matter in general relativity," Proc. R. Soc. London Ser. A **316**, 1 (1970).
[8] V. del Olmo and J. Olivert, "Weak rigidity in almost-thermodynamic material schemes," J. Math. Phys. **26**, 1311 (1985).
[9] J. J. Ferrando and J. Olivert, "Incompressibility in relativistic continuous media," J. Math. Phys. **22**, 2223 (1981).
[10] V. del Olmo and J. Olivert, "Compatibility of rigid motions with the relativistic incompressibility condition," J. Math. Phys. **24**, 1172 (1983).
[11] R. K. Sachs and H. Wu, *General Relativity for Mathematicians* (Springer, New York, 1977).
[12] S. W. Hawking and G. F. R. Ellis, *The Large-scale Structure of Space-time*

(Cambridge U. P., Cambridge, 1973).
[13]W. Greub, S. Halperin, and R. Vanstone, *Connections, Curvature and Cohomology* (Academic, New York, 1972).
[14]S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962).
[15]R. L. Bishop and S. I. Goldberg, *Tensor Analysis on Manifolds* (Macmillan, New York, 1968).
[16]F. Brickell and R. S. Clark, *Differentiable Manifolds* (Van Nostrand, Princeton, NJ, 1970).
[17]R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin/Cummings, Reading, MA, 1978).
[18]L. D. Landau and E. M. Lifshitz, *Theory of Elasticity* (Pergamon, London, 1959).

A Kerr object embedded in a gravitational field[a)]

K. D. Krori and R. Bhattacharjee
Mathematical Physics Forum, Cotton College, Gaunati-781001, India

(Received 10 June 1985; accepted for publication 6 November 1985)

The exact expression for the Ernst potential for a Kerr object embedded in a gravitational field is derived using a formalism developed by Kramer and Neugebauer.

I. INTRODUCTION

Kerns and Wild[1] recently constructed an exact solution of the vacuum Einstein field equations for a Schwarzschild black hole embedded in an external gravitational field using a formalism developed by Ernst.[2] (The same solution may, however, be directly derived from the Ernst equation.[3]) But, as pointed out by Kerns and Wild in their paper (and also by the latter in a private communication), "it would be very interesting to develop a technique to imbed the Kerr metric in an external vacuum field as this may have application to black holes that exist under very unusual physical environments." Having addressed ourselves to the search for such a technique we have achieved a certain measure of success. Using a formalism developed by Kramer and Neugebauer[4,5] and Kramer,[6] we derive in this paper the Ernst potential for a Kerr object embedded in a gravitational field for $a^2 \geqslant m^2$. The problem, however, remains open for $a^2 < m^2$, which is now under our consideration.

II. KRAMER-NEUGEBAUER FORMALISM

The Einstein–Maxwell equations for a stationary axisymmetric field[7] in terms of the Ernst potential $\mathscr{E}$ and electromagnetic potential Φ are given by

$$\begin{aligned} &f\Delta\mathscr{E} = (\nabla\mathscr{E} + 2\overline{\Phi}\nabla\Phi)\nabla\mathscr{E},\\ &f\Delta\Phi = (\nabla\mathscr{E} + 2\overline{\Phi}\nabla\Phi)\nabla\Phi,\\ &f = \mathrm{Re}\,\mathscr{E} + \overline{\Phi}\Phi, \end{aligned} \tag{1}$$

where a bar denotes complex congugation.

These nonlinear equations for $\mathscr{E}$ and Φ represent the integrability condition of the following linear eigenvalue equations:

$$\begin{aligned} \Omega_{,z} &= \left[\begin{pmatrix} B_1 & 0 & E_1 \\ 0 & A_1 & 0 \\ -F_1 & 0 & \frac{1}{2}(A_1+B_1) \end{pmatrix} + \lambda \begin{pmatrix} 0 & B_1 & 0 \\ A_1 & 0 & -E_1 \\ 0 & -F_1 & 0 \end{pmatrix}\right]\Omega,\\ \Omega_{,\bar z} &= \left[\begin{pmatrix} B_2 & 0 & E_2 \\ 0 & A_2 & 0 \\ -F_2 & 0 & \frac{1}{2}(A_2+B_2) \end{pmatrix} + \frac{1}{\lambda} \begin{pmatrix} 0 & B_2 & 0 \\ A_2 & 0 & -E_2 \\ 0 & -F_2 & 0 \end{pmatrix}\right]\Omega, \end{aligned} \tag{2}$$

where the 3×3 pseudopotential matrix $\Omega(\lambda) = \Omega(\lambda, z, \bar z)$ is normalized according to

$$\Omega(1, z, \bar z) = \begin{pmatrix} \overline{\mathscr{E}} + 2\Phi\overline{\Phi} & 1 & \sqrt{2}i\Phi \\ \mathscr{E} & -1 & -\sqrt{2}i\Phi \\ -2i\overline{\Phi}f^{1/2} & 0 & \sqrt{2}f^{1/2} \end{pmatrix}, \tag{3}$$

at $\lambda = 1$. Here λ is given by

$$\lambda = (K - i\bar z)^{1/2}(K + iz)^{-1/2},$$

K being the spectral parameter. Further,

$$\begin{aligned} &A_1 = \tfrac{1}{2}f^{-1}(\mathscr{E}_{,z} + 2\overline{\Phi}\Phi_{,z}),\\ &B_1 = \tfrac{1}{2}f^{-1}(\overline{\mathscr{E}}_{,z} + 2\Phi\overline{\Phi}_{,z}),\\ &E_1 = if^{-1/2}\Phi_{,z},\\ &F_1 = if^{-1/2}\overline{\Phi}_{,z},\\ &f = \mathrm{Re}\,\mathscr{E} + \overline{\Phi}\Phi, \end{aligned} \tag{4}$$

and $A_2 \cdots F_2$ are the corresponding expressions for $\bar z$ in place of z. For a given solution $(\mathscr{E}_0, \Phi_0)$ there is an associated matrix Ω_0 that satisfies the conditions (2) and (3). By means of the ansatz

$$\begin{aligned} &\Omega = T(\lambda)\Omega_0,\\ &T(\lambda) = T(\lambda, z, \bar z),\\ &T(-\lambda) = \epsilon T(\lambda)\epsilon,\\ &\epsilon := \mathrm{diag}(1, -1, 1),\\ &T(\lambda) = \alpha(K)(K + iz)^{n/2}\sum_{s=0}^{n} X_s\lambda^s, \quad n = 2N,\\ &T(1) = \sum_{s=0}^{n} X_s, \end{aligned} \tag{5}$$

with λ-independent 3×3 matrices X_s and a suitably chosen constant $\alpha(K)$, we can build from Ω_0 a new matrix Ω obeying (2) and (3).

The 3×3 λ-independent matrices X_s can be determined completely from the equations

$$\begin{aligned} &\sum_{s=0}^{n} X_s\lambda_K^s\Omega_0(\lambda_K)C_K = 0,\\ &T_{11}(1) - T_{12}(1) = 1,\\ &T_{13}(1) + T_{23}(1) = 0,\\ &T_{22}(1) - T_{21}(1) = 1,\\ &T_{31}(1) - T_{32}(1) = 0,\\ &T_{33}(1) = \overline{T_{33}(1)} = (1 + T_{12}(1) + T_{21}(1))^{1/2}. \end{aligned} \tag{6}$$

For a suitably chosen constant $\alpha(K)$ using (6) in (5) one

[a)] This work is humbly dedicated in honor of Professor P. C. Vaidya who derived the "Vaidya metric."

 0022-2488/86/041056-03$02.50

can obtain the new matrix Ω that also satisfies (2) and (8).

The zeros λ_K of the det T and the constant vector C_K satisfy the conditions

$$\lambda_{3m} = \lambda_{3m-2}, \quad \lambda_{3m-1} = 1/\lambda_{3m}, \quad m = 1,...,N; \quad (7)$$

$$C^+_{3m-1}\sigma C_{3m} = 0 = C^+_{3m-1}\sigma C_{3m-2},$$

$$\sigma := \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \quad (8)$$

Thus for a given Ω_0 and the values of λ_K and C_K determined from (7) and (8) one can obtain the transformation matrix T from (6) and consequently the pseudopotential matrix Ω. From Ω one obtains

$$\mathscr{E} = \Omega_{21}(1), \quad \sqrt{2}\Phi = i\Omega_{23}(1). \quad (9)$$

III. ERNST POTENTIAL FOR A KERR OBJECT IN THE BACKGROUND OF A GRAVITATIONAL FIELD

As the background we choose the gravitational field as derived by Ernst[2]

$$\mathscr{E}_0 = e^U = f_0, \quad \Phi = 0. \quad (10)$$

For any u there is a function $V(\lambda) = V(\lambda, z, \bar{z})$ defined by

$$V_{,z} = \lambda U_{,z}, \quad V_{,\bar{z}} = (1/\lambda) U_{,\bar{z}}. \quad (11)$$

The constant of integration in the case of (11) is chosen to satisfy

$$V(1) = U \quad (12)$$

The definition (11) implies the properties

$$V(-\lambda) = -V(\lambda), \quad V^*(\lambda) = \overline{V(1/\lambda)} = V(\lambda). \quad (13)$$

The pseudopotential matrix Ω_0 for the solution (10) is

$$\Omega_0(\lambda) = \begin{pmatrix} e^{(U+V)/2} & e^{(U-V)/2} & 0 \\ e^{(U+V)/2} & -e^{(U-V)/2} & 0 \\ 0 & 0 & \sqrt{2}e^{U/2} \end{pmatrix}. \quad (14)$$

For the derivation of the Kerr ($N = 1$) solution ($a^2 \geqslant m^2$) from flat space time ($U = V = 0$), the $\lambda_K s$ and $C_K s$ are given by (7) and (8).

The spectral parameters K_s are taken as $K_1 = K_3 = -K_2 = ix$, x being real and

$$\lambda_1 = \lambda_3 = [(x - \bar{z})/(x + \bar{z})]^{1/2},$$
$$\lambda_2 = \overline{\lambda_1^{-1}} = [(x + \bar{z})/(x - \bar{z})]^{1/2}. \quad (15)$$

The components p_K, q_K, and r_K of the vectors $\Omega_0(\lambda_K)C_K$,

$$\begin{pmatrix} p_K \\ q_K \\ r_K \end{pmatrix} := \Omega_0(\lambda_K)C_K, \quad K = 1,...,3, \quad (16)$$

are given by

$$p_K = e^{U/2}\left[\frac{p_K^0 + q_K^0}{2} e^{V_K/2} + \frac{p_K^0 - q_K^0}{2} e^{-V_K/2}\right],$$
$$q_K = e^{U/2}\left[\frac{p_K^0 + q_K^0}{2} e^{V_K/2} - \frac{p_K^0 - q_K^0}{2} e^{-V_K/2}\right], \quad (17)$$
$$r_K = r_K^0 e^{U/2},$$

where $V_K = V(\lambda_K, z, \bar{z})$, and the superscript 0 refers to the case $U = V = 0$.

The quantities p_K, q_K, r_K are restricted by

$$\bar{p}_2 p_3 - \bar{q}_2 q_3 - \bar{r}_2 r_3 = 0 = \bar{p}_2 p_1 - \bar{q}_2 q_1 - \bar{r}_2 r_1, \quad (18)$$

if the corresponding relations held for p_K^0, q_K^0, and r_K^0. The ratios

$$\alpha_K^0 = p_K^0/q_K^0, \quad \alpha_K = p_K/q_K,$$
$$\beta_K^0 = r_K^0/q_K^0, \quad \beta_K = r_K/q_K, \quad (19)$$

will occur in the potential functions of the new solutions. The α_K^0's and β_K^0's are prescribed according to

$$\alpha_1^0 = \alpha_3^0 = i(a - x)/m,$$
$$\alpha_2^0 = (\bar{\alpha}_1^0)^{-1} = i(a + x)/m, \quad (20)$$
$$\beta_1^0 = \beta_2^0 = 0, \quad \beta_3^0 = 1, \quad x^2 = a^2 - m^2 \geqslant 0,$$

a and m being the real Kerr parameters.

For α_K and β_K one gets the expressions

$$\alpha_1 = \frac{\alpha_1^0 + \tanh(V_1/2)}{\alpha_1^0 \tanh(V_1/2) + 1} = \alpha_3,$$
$$\alpha_2 = \frac{\alpha_2^0 + \tanh(V_2/2)}{\alpha_2^0 \tanh(V_2/2) + 1},$$
$$\beta_1 = 0 = \beta_2 \quad (21)$$
$$\beta_3 = \frac{2}{(\alpha_1^0 + 1)e^{V_1/2} - (\alpha_1^0 - 1)e^{-V_1/2}}.$$

Using (15)–(17), (20), and (21) in (16) the components of $\Omega_{\alpha\beta}$ of the new pseudopotential matrix can be determined.

Equation (9) leads to the Ernst potential $\mathscr{E}$ in the following form:

$$\mathscr{E} = e^u\left[1 - 2\left(\frac{i(a-x) + m\tanh(V_1/2)}{i(a-x)\tanh(V_1/2) + m} \cdot \frac{x\cos\theta + i(r-m)}{2x} + \frac{i(a+x) + m\tanh(V_2/2)}{i(a+x)\tanh(V_2/2) + m} \cdot \frac{x\cos\theta - i(r-m)}{2x} + 1\right)^{-1}\right]. \quad (22)$$

IV. DISCUSSION

Equation (22) gives the exact expression for the Ernst potential for a Kerr object embedded in a gravitational field for $a^2 \geqslant m^2$. For $U = V = 0$ (absence of the gravitational field), (22) reduces to the Kerr solution for $a^2 \geqslant m^2$. On the other hand, for $a = m = 0$, from (22) we have only the background gravitational field.[8]

ACKNOWLEDGMENT

The authors express their profound gratitude to the Government of Assam, Dispur, Gauhati (India) for the financial assistance to conduct the investigations reported in this paper.

[1]R. M. Kerns and W. J. Wild, Gen. Relativ. Gravit. **14**, 1 (1982).
[2]F. J. Ernst, J. Math. Phys. **19**, 1986 (1978).
[3]F. J. Ernst, Phys. Rev. **167**, 1175 (1968).
[4]D. Kramer and G. Neugebauer, J. Phys. A **14**, 1333 (1981).
[5]G. Neugebauer and D. Kramer, J. Phys. A **16**, 1927 (1983).
[6]D. Kramer, J. Phys. A **15**, 2201 (1982).
[7]F. J. Ernst, Phys. Rev. **168**, 1415 (1968).
[8]This is done by making first $x \to 0$ (i.e., $a^2 \to m^2$) and then $a \to 0$ and $m \to 0$ in (22). When $x \to 0$, $V_1 = V_2 = V_0$ (say).

A generalization of Penrose's helicity theorem for space-times with nonzero dual mass

Anne Magnon[a)]

Physics Department, University of Syracuse, Syracuse, New York 13244-1130 and Département de Mathématiques, Université de Clermont-Fd, 63170, Aubière, France

(Received 25 July 1985; accepted for publication 31 October 1985)

An algebraic definition of the helicity operator $\mathscr{H}$ is proposed for vacuum stationary and asymptotically flat wormholes (i.e., space-times where the manifold of orbits of the stationary Killing field has $S^2 \times R$ topology). The definition avoids the use of momentum space or Fourier decomposition of the gravitational degrees of freedom into positive and negative frequency parts, and is essentially geared to emphasize the role of nontrivial topology. It is obtained via the introduction of a total spin vector S^α derived from the dual Bondi four-momentum $*P^\alpha$, both vectors originating in the presence of nontrivial homotopy groups. (Space-times with nonzero dual mass can be characterized by a conformal null boundary $\mathscr{I}$ having the topology of an S^1 fiber bundle over S^2 with possible identifications along the fiber—lens space—or equivalently vanishing Bondi–News.) It is shown that S^α is a constant multiple of P^α, the total Bondi four-momentum, and if in addition the space-time admits a point at spacelike infinity, there is strong support for the past limit of S^α to be a null vector. This can be viewed as a generalization of Penrose's result on the Pauli–Lubanski vector for classical zero rest-mass particles. The helicity operator at null infinity is rooted in the topology and turns out to be essentially the Hodge duality operator (*). The notion of duality appears as a global concept. Under such conditions, self- and anti-self-dual modes of the Weyl curvature could be viewed as states originating in the nontrivial topology. These results depend crucially on the presence of topological charges; it is tempting to speculate that such wormholes might be basic building blocks.

I. INTRODUCTION

It is often stated that, since the strength of the gravitational coupling constant is so weak, the inclusion of gravitational effects in quantum theory would induce corrections that would be irrelevent at the laboratory scale. However, one must acknowledge the fact that quantum field theory finds its success in its ability to describe particles and predict effects that are essentially qualitative and independent of the strength of the coupling constant: for instance, the prediction that to every particle is associated an antiparticle arises in the transition from Newtonian to special relativistic framework with the availability of the Poincaré group. Consequently it is of interest to search for features that are specific of general relativity and that could generate new qualitative effects. One such feature is the availability of nontrivial topologies. Of course, there is still no observational evidence that the space-time topology could be anything other than that of R^4. Nevertheless, assuming an absolute topology would be as contrary to the spirit of physics as assuming the existence of an absolute time and consequently rejecting the basic principle of relativity.

In this paper, we shall focus on a class of vacuum space-times, usually quoted as "gravitational dyons" and derive the notion of right and left helicity (eigen) states of the gravitational degrees of freedom via an algebraic method. Although this notion has been extensively studied for massless free fields,[1] our investigation is of interest since it aims at underlining the role of nontrivial topologies.

Recall that for classical zero rest-mass particles, with four-momentum P^α, and angular momentum $M_{\beta\gamma}$, the spin vector S^α is defined by $S^\alpha = \epsilon^{\alpha\beta\gamma\delta} P_\beta M_{\gamma\delta}$, where $\epsilon_{\alpha\beta\gamma\delta}$ is the alternating tensor of Minkowski space-time. Here S^α and P^α have been shown[2] to be proportional, the helicity being the proportionality coefficient $\mathfrak{s}$ between these two null vectors: $S^\alpha = \mathfrak{s} P^\alpha$.

For a self-gravitating isolated system, the definition of total four-momentum and total angular momentum has been introduced[3] within the framework of general relativity using the behavior of the Weyl curvature of the underlying space-time at large spacelike separations from sources. The construction may be summarized as follows. The notion of asymptotic flatness at spatial infinity is introduced via a conformal completion in the sense of Ref. 4, which attaches to the physical space-time a point analogous to the Minkowskian i°. The resulting conformally rescaled metric is smooth in "angular directions" and has finite radial discontinuities in its first derivatives, which in fact measure the total energy-momentum of the space-time. The group of asymptotic symmetries is then introduced, and displays most of the features of the Bondi–Mezner–Sachs (BMS) group. Next, physical fields that incorporate the asymptotic behavior of the gravitational field, are isolated. These are represented by two second rank, symmetric, trace-free "electric" and "magnetic" tensor fields $E_{\alpha\beta}$ and $B_{\alpha\beta}$ on the hyperboloid $\mathscr{D}$ of unit spacelike vectors in the tangent space at i°—the point at spa-

[a)] Détachée du Ministère des Relations Extérieures, Paris, France.

 0022-2488/86/041059-07$02.50

tial infinity. The $1/r^3$ behavior of the physical Weyl curvature is contained in this pair of fields. Finally, using the asymptotic symmetries and asymptotic fields, conserved quantities can be defined. The four-momentum P_α arises as a mapping from the preferred four-dimensional normal subgroup of the asymptotic group (the subgroup of translations at spacelike infinity) into the reals, and involves the asymptotic field $E_{\alpha\beta}$. The corresponding quantity, involving $B_{\alpha\beta}$, and representing a "magnetic gravitational charge," vanishes. The angular momentum emerges from the $1/r^4$ contribution to the magnetic part of the Weyl curvature and is obtained after introduction of an additional fall-off condition. This condition causes the reduction of the asymptotic symmetry group to the Poincaré group and angular momentum emerges as a skew tensor $M_{\alpha\beta}$ generating a linear mapping from the Lorentz Lie albegra into the reals. The total spin vector S^α can be introduced: $S^\alpha = \epsilon^{\alpha\beta\gamma\delta} P_\beta M_{\gamma\delta}$, where $\epsilon^{\alpha\beta\gamma\delta}$ is the alternating tensor of the Minkowski space tangent at i°. This fixed vector at i° could be viewed as an element of the Poincaré Lie algebra at spacelike infinity. If the four-momentum is null-like, $S^\alpha = \mathfrak{s} P^\alpha$, where $\mathfrak{s}$ denotes the total helicity of the space-time. This definition of helicity is a natural generalization, in the case of the gravitational field, of that introduced for zero rest-mass particles.

Ambiguities[5] arising in the definition of angular momentum at null infinity (presence of local fluxes even if the Bondi News function vanishes, presence of various incompatible definitions) have prevented the introduction of a total spin vector for radiative degrees of freedom, which would mimic that available at spacelike infinity. As a result the notion of helicity has been introduced via different procedures. For instance the scheme provided by the quantum description of Maxwell fields in Minkowski space, can be summarized as follows. First, "true" degrees of freedom of the field (radiative modes) are isolated via a Fourier transform: a solution $F_{\alpha\beta}(x)$ to the source free field equations can be completely characterized by the equivalence class $\{A_\alpha(k)\}$ of vector fields on the light cone in the momentum space, where two $A_\alpha(k)$ are equivalent if they differ by a multiple of k^α. These equivalence classes represent the radiative degrees of freedom. Each $\{A_\alpha(k)\}$ leads to a creation and annihilation operator (a photon state). Each photon state can finally be decomposed into right and left helicity components, the positive and negative eigenvalues of the duality operator. The corresponding treatment for gravity can then be obtained as follows.[6] The radiative degrees of freedom are equivalence classes $\{D\}$ of connections at null infinity $\mathscr{I}$. Each class is characterized by two tensor fields on $\mathscr{I}$: $N_{\alpha\beta}$, the News tensor, and ${}^*K_{\alpha\beta}$, a symmetric traceless tensor field associated to the pullback of $\Omega^{-1}\,{}^*C_{\alpha\mu\beta\nu} n^\mu n^\nu$ to $\mathscr{I}$ (where n^ν denotes the null normal to $\mathscr{I}$). A natural decomposition into positive and negative frequency parts of the field operators $N(f)$ induced by the action of operator valued distributions $N_{\alpha\beta}(x)$ on test fields f's is provided by the affine parameter u along the integral curves of n^α. Right-handed (resp. left-handed) graviton states emerge as eigenstates with eigenvalue i (resp. $-i$) of the natural alternating tensor available on the space of generators of $\mathscr{I}$.

In this paper, we shall present a new approach to the notion of helicity states of the gravitational degrees of freedom, which will not require the introduction of Fourier decomposition into positive and negative frequency parts, will be effective when the News function is zero, and will be based on the nonvanishing of a charge, called the "dual mass," obtained from ${}^*K_{\alpha\beta}$ in vacuum stationary space-times, due to the presence of a nontrivial topology: the manifold of orbits of the stationary Killing vector field has topology $S^2 \times R$, and $\mathscr{I}$ must be an S^1 fiber bundle over S^2. The method is algebraic and is based on the introduction of a helicity operator $\mathscr{H}$ derived from a spin vector S^α, which incorporates the total angular momentum of the space-time and is shown to be a multiple of the total Bondi four-momentum P^α. Here $\mathscr{H}$ is essentially the Hodge duality operator (*). Its eigenstates will provide the self- and anti-self-dual modes of the gravitational degrees of freedom. The presence of these modes is essentially rooted in the space-time topology. Such features are qualitative and specific of general relativity. We would like to consider them as an example of these effects, which are expected to play a role in quantum gravity.

II. PRELIMINARIES

The class of space-times that is going to be investigated is that of vacuum stationary solutions to Einstein's equation encompassing most of the features of NUT solutions[7] (wormholes): the manifold of orbits of the stationary Killing vector field has topology $S^2 \times R$, and $\mathscr{I}$ is an S^1 bundle over S^2. (The structure at spacelike infinity i° will not play a leading role: it is not expected to be always reasonable for such space-times.) Under these conditions the total Bondi dual mass (magnetic angular monopole moment) does not vanish (reflecting the presence of nontrivial homotopy groups) and is generated by the field ${}^*K_{\alpha\beta}$ on $\mathscr{I}$. Since such space-times are characterized by $N_{\alpha\beta} = 0$, together with their topological structure, we shall briefly outline the technique for analyzing the asymptotic structure at null infinity.

Definition: A space-time $(\hat{M}, \hat{g}_{\alpha\beta})$ will be said to be asymptotically empty and flat at null infinity if there exists a manifold M with boundary $\mathscr{I}$ such that $M = \hat{M} \cup \mathscr{I}$, together with hyperbolic metric $g_{\alpha\beta}$ and a smooth scalar field Ω on M such that (i) $g_{\alpha\beta} = \Omega^2 \hat{g}_{\alpha\beta}$ on $\hat{M}$, (ii) on $\mathscr{I}$, $\Omega = 0$, $n_\alpha = \nabla_\alpha \Omega \neq 0$, (iii) there exists a neighborhood N of $\mathscr{I}$ (in the space-times under consideration $N = M$) such that in $N \cap \hat{M}$, the vacuum Einstein's equations $\hat{R}_{\alpha\beta} = 0$ are satisfied, and (iv) $n^\alpha = g^{\alpha\beta} \nabla_\beta \Omega$ is complete and $\mathscr{S}$, the space of its orbits, has the topology of a two-sphere.

The above conditions imply that the asymptotic structure resembles sufficiently that of Minkowski space to enable the introduction of familiar notions such as the radiation field, peeling properties, energy momentum four-vectors, "balance laws," etc. They also imply that $\mathscr{I}$ is a null three-surface with n_α as a normal. The topology of $\mathscr{I}$ will be relaxed to that of a principal S^1 fiber bundle over S^2 in order to enable the introduction of the conserved quantity known as the dual mass: if the topology of $\mathscr{I}$ is $S^2 \times R$ and $N_{\alpha\beta} = 0$, the dual mass is essentially the integral over $\mathscr{S}$ of a curl[8] $(\mathrm{Im}\, \eth^2 \sigma)$ preventing the existence of a two-sphere cross section on $\mathscr{I}$. The metric $g_{\alpha\beta}$ induces, via pullback, a degenerate metric $q_{\alpha\beta}$, with signature $(0, +, +)$, such that $q_{\alpha\beta}v^\beta = 0$ iff v^β is proportional to n^α. Using the gauge freedom associated with the rescaling of $g_{\alpha\beta}$ ($\tilde{g}_{\alpha\beta} = \omega^2 g_{\alpha\beta}$), $\mathscr{I}$ can be assumed to be divergence free ($\nabla_\alpha \nabla_\beta \omega\Omega$ vanishes on $\mathscr{I}$). We are thus left with the gauge freedom allowed by the equation $\mathscr{L}_n \omega = 0$. Alternating tensors on $\mathscr{I}$, $\epsilon^{\alpha\beta\gamma}$, and $\epsilon_{\alpha\beta\gamma}$ are fixed by $\epsilon^{\alpha\beta\gamma}\epsilon^{\lambda\mu\nu}q_{\alpha\lambda}q_{\beta\mu} = 2n^\gamma n^\nu$, $\epsilon^{\alpha\beta\gamma}\epsilon_{\alpha\beta\gamma} = 6$. The second-order structure is coded in the derivative operator D, induced on $\mathscr{I}$ by ∇, the derivative operator on $(M, g_{\alpha\beta})$. This D satisfies

$$D_\alpha q_{\beta\gamma} = 0, \quad D_\alpha n^\beta = 0,$$

and (1)

$$D_\alpha V_\beta = D_{[\alpha} V_{\beta]} + \tfrac{1}{2}\mathscr{L}_V q_{\alpha\beta}, \quad \text{if } V_\alpha n^\alpha = 0,$$

where V^α is any vector field on $\mathscr{I}$ satisfying $V^\alpha q_{\alpha\beta} = V_\beta$. Using the fact that the space of generators of $\mathscr{I}$ is diffeomorphic to S^2, we can show the existence, on $\mathscr{I}$, of a unique symmetric tensor field $\rho_{\alpha\beta}$ satisfying

$$\rho_{\alpha\beta}n^\beta = 0, \quad \rho_{\alpha\beta}q^{\alpha\beta} = \mathscr{R}, \quad D_{[\alpha}\rho_{\beta]\gamma} = 0, \tag{2}$$

where $q^{\alpha\beta}$ is any inverse of $q_{\alpha\beta}$ and $\mathscr{R}$ is the lift to $\mathscr{I}$ of the scalar curvature $\overline{\mathscr{R}}$ of the metric $\bar{q}_{\alpha\beta}$ induced on the space of generators of $\mathscr{I}$ by $q_{\alpha\beta}$. Finally the Riemann tensor $R_{\alpha\beta\gamma\delta}$ and Ricci tensor $R_{\alpha\beta}$ of $g_{\alpha\beta}$ induce a useful tensor field on $\mathscr{I}$: $S_\alpha{}^\beta$ the pullback to $\mathscr{I}$ of the combination $R_\alpha{}^\beta - \frac{1}{6}Rg_\alpha{}^\beta$, which satisfies

$$S_\alpha{}^\beta n^\alpha = \sigma n^\beta; \quad S_{\alpha\beta} \equiv S_\alpha{}^\gamma q_{\gamma\beta} = S_{(\alpha\beta)}; \quad S_{\alpha\beta}q^{\alpha\beta} = \mathscr{R}, \tag{3}$$

for some scalar field σ on $\mathscr{I}$, while the Weyl tensor induces two symmetric fields $K^{\alpha\beta}$ and ${}^*K^{\alpha\beta}$ via $K_{\alpha\beta\gamma\delta}$ the pullback to $\mathscr{I}$ of $\Omega^{-1}C_{\alpha\beta\gamma\delta}$:

$$K^{\alpha\beta} = \epsilon^{\alpha\lambda\mu}\epsilon^{\beta\nu\tau}K_{\lambda\mu\nu\tau}, \quad {}^*K^{\alpha\beta} = \epsilon^{\alpha\lambda\mu}\epsilon^{\beta\nu\tau}\, {}^*K_{\lambda\mu\nu\tau}, \tag{4}$$

which satisfy

$$K^{\alpha\beta}q_{\alpha\beta} = 0; \quad {}^*K^{\alpha\beta}q_{\alpha\beta} = 0. \tag{5}$$

The useful properties of these fields are listed below:

$$D_\alpha(S^\alpha{}_\beta - S\delta^\alpha{}_\beta) = 0, \quad S \equiv S^\mu{}_\mu; \tag{6a}$$

$$D_{[\alpha}S_{\beta]}{}^\gamma = \tfrac{1}{4}\epsilon_{\alpha\beta\mu}\, {}^*K^{\mu\gamma}; \tag{6b}$$

$$D_\alpha\, {}^*K^{\alpha\beta} = 0, \quad D_\alpha K^{\alpha\beta} = 0; \tag{6c}$$

$$q_{\alpha\mu}K^{\mu\beta} = -\epsilon_{\alpha\mu\nu}n^\nu\, {}^*K^{\mu\beta}, \quad q_{\alpha\mu}\, {}^*K^{\mu\beta} = \epsilon_{\alpha\mu\nu}n^\nu K^{\mu\beta}. \tag{6d}$$

Under conformal rescalings $\tilde{g}_{\alpha\beta} = \omega^2 g_{\alpha\beta}$, which leave $\mathscr{I}$ divergence-free, we have the following transformations:

$$\tilde{n}^\alpha = \omega^{-1}n^\alpha, \quad \tilde{\epsilon}_{\alpha\beta\gamma} = \omega^3\epsilon_{\alpha\beta\gamma}, \tag{7a}$$

$$\tilde{K}^{\alpha\beta} = \omega^{-5}K^{\alpha\beta}, \quad {}^*\tilde{K}^{\alpha\beta} = \omega^{-5}\, {}^*K^{\alpha\beta}, \tag{7b}$$

$$\tilde{S}_\alpha{}^\beta = \omega^{-2}S_\alpha{}^\beta - 2\omega^{-3}D_\alpha\omega^\beta + 4\omega^{-4}\omega^\beta D_\alpha\omega - \omega^{-4}\delta_\alpha{}^\beta\omega^\mu D_\mu\omega, \tag{7c}$$

where ω_α is the pullback to $\mathscr{I}$ of $\nabla_\alpha\omega$.

An important gauge invariant field is the Bondi–News function $N_{\alpha\beta}$, the gauge invariant part of $S_\alpha{}^\beta$ defined by

$$N_{\alpha\beta} = S_{\alpha\beta} - \rho_{\alpha\beta},$$

and satisfying (8)

$$D_{[\alpha}N_{\beta]\gamma} = \tfrac{1}{4}\epsilon_{\alpha\beta\lambda}\, {}^*K^{\lambda\nu}q_{\nu\gamma}.$$

From its algebraic properties, it can be deduced that $N_{\alpha\beta}$ has only two independent components, which yield, in a Bondi frame, the complex scalar denoted by $\dot{C}$ by Bondi *et al.*,[9] and Sachs,[10] denoted by N by Penrose,[11] and by $\dot{\sigma}^\circ$ by Newman and Unti.[12] If $N_{\alpha\beta} \neq 0$, the physical space-time admits gravitational waves. In the Newman–Penrose[13] formalism, the five independent components of ${}^*K^{\alpha\beta}$ correspond to Ψ_4^0, Ψ_3^0, Im Ψ_2^0, while those of $K^{\alpha\beta}$ correspond to Ψ_4^0, Ψ_3^0, Re Ψ_2^0. It is usually stated[14] that ${}^*K^{\alpha\beta}$ carries only radiative information and vanishes in stationary space-times. We shall see that it is not so in the stationary solutions under consideration. More specifically, we shall see that, while $K^{\alpha\beta}$ and $N_{\alpha\beta}$ carry information concerning the energy momentum radiated away to infinity, ${}^*K^{\alpha\beta}$ carries that concerning the space-time angular momentum.

Finally let us mention that an asymptotic symmetry group is available at $\mathscr{I}$, which admits the BMS group as its covering space, since each generator of $\mathscr{I}$ has the topology of an S^1. The supertranslations are of the form $t^\lambda = \alpha n^\lambda$, where α, a scalar function on $\mathscr{I}$, satisfies $L_n\alpha = 0$, and is restricted by $D_\lambda D_\mu\alpha + \frac{1}{2}\alpha\rho_{\lambda\mu} = \kappa q_{\lambda\mu}$ (for some function κ on $\mathscr{I}$) in the case of an infinitesimal translation.

III. TOTAL SPIN VECTOR, DUALITY AND HELICITY OPERATORS

In this section we would like to introduce the notion of spin operator for zero rest-mass fields, and more specifically for gravitational fields. Recall that for a classical zero rest-mass particle with four-momentum P_β and angular momentum $M_{\gamma\delta}$, the spin vector is defined by

$$S^\alpha = \epsilon^{\alpha\beta\gamma\delta}P_\beta M_{\gamma\delta}, \tag{9}$$

which can be interpreted as the action of ${}^*M^{\alpha\beta}$ on P_β. Furthermore it has been shown[2] that $S^\alpha = \mathfrak{s}P^\alpha$, the proportionality coefficient $\mathfrak{s}$ being identified with the helicity of the particle.

We shall now consider the four-momentum P_α and angular momentum form $M_{\gamma\delta}$ as elements of the Poincaré Lie algebra, and evaluate the corresponding structure constants to obtain an expression of the associated spin vector. We shall introduce Latin indices for the labeling of components in adapted charts, keeping Greek indices as abstract indices[4] for the labeling of (vector) spaces and associated geometrical objects. In an adapted chart, P_α can be represented by the derivation $P^a_\alpha\partial/\partial x^a = \delta^a_\alpha\partial/\partial x^a$. It is clear that the corresponding commutators $[P_\alpha, P_\beta]$ vanish. Similarly the derivation generated by the angular momentum two-form, via its dual, is represented by

$$M_{\gamma\delta} = (\delta_\gamma{}^c\delta_\delta{}^d - \delta_\delta{}^c\delta_\gamma{}^d)x_c\frac{\partial}{\partial x^d} = \Delta^{cd}_{\gamma\delta}\eta_{ck}x^\kappa\frac{\partial}{\partial x^d},$$

where n_{ck}'s denote the components of the Minkowskian metric in the adapted chart, and

$$\Delta_{\gamma\delta}{}^{cd} = \delta_\gamma{}^c\delta_\delta{}^d - \delta_\delta{}^c\delta_\gamma{}^d.$$

A straightforward calculation leads to

$$[M_{\alpha\beta}, M_{\gamma\delta}] = M_{\alpha\delta}\eta_{\beta\gamma} - M_{\beta\delta}\eta_{\alpha\gamma} + M_{\beta\gamma}\eta_{\alpha\delta} - M_{\alpha\gamma}\eta_{\beta\delta},$$

(10)

$$[P_\alpha, M_{\beta\gamma}] = \eta_{\alpha\beta}P_\gamma - \eta_{\alpha\gamma}P_\beta,$$

where the second Lie bracket gives the action of the angular momentum two-form on the four-momentum vector. Following the definition available for particles, it is natural to define the action of the spin vector on fields via that of the form

$$[{}^*M_{\alpha\beta}, P_\gamma] = \tfrac{1}{2}\epsilon_{\alpha\beta}{}^{\lambda\mu}(P_\lambda\eta_{\mu\gamma} - P_\mu\eta_{\lambda\gamma}) = \epsilon_{\alpha\beta}{}^{\lambda}{}_{\gamma}P_\lambda .$$

We shall now display its action on zero rest-mass fields. For a spin-1 field $F_{\alpha\beta}$, we obtain

$$F^{\alpha\beta}\epsilon_{\alpha\beta}{}^{\lambda}{}_{\gamma}P_\lambda \equiv P_\lambda\, {}^*F^{\lambda}{}_{\gamma}. \qquad (11)$$

In an adapted chart, this last expression can be written $\delta_\lambda{}^a\nabla_a{}^*F^\lambda{}_\gamma$, or equivalently $\nabla_\lambda\ {}^*F^\lambda{}_\gamma$. In the case of a spin-2 field $K_{\alpha\beta\tau\kappa}$ having the symmetries of the Weyl tensor, we obtain[15]:

$$K^{\alpha\beta}{}_{\tau\kappa}\epsilon_{\alpha\beta}{}^{\lambda}{}_{\gamma}P_\lambda + K_{\tau\kappa}{}^{\alpha\beta}\epsilon_{\alpha\beta}{}^{\lambda}{}_{\gamma}P_\lambda, \qquad (12)$$

which, in an adapted chart, can be written $2\delta^a_\lambda\nabla_a\ {}^*K^{\lambda}{}_{\gamma\tau\kappa}$. (In the generic case of spin-s fields, the corresponding expression will display the spin-s as a coefficient.) Following the notation suggested by the expression of the spin vector for particles, we shall introduce the helicity operator $\mathscr{H}$ defined by

$$\mathscr{H} = s(*), \qquad (13)$$

where $(*)$ denotes the Hodge duality operator. The action of the spin operator on the field $K_{\alpha\beta\lambda\tau}$ will thus be given by

$$sP_\lambda\ {}^*K^{\lambda}{}_{\gamma\tau\kappa} = s(*)P_\lambda K^{\lambda}{}_{\gamma\tau\kappa} = \mathscr{H}P_\lambda K^{\lambda}{}_{\gamma\tau\kappa}.$$

In the case of complex fields, P_λ should be replaced by a Hermitian vector, with components $i\delta_\lambda{}^a$ (in an adapted chart) the corresponding helicity operator being

$$\mathscr{H} = is\ (*). \qquad (14)$$

Hence the helicity operator appears as a purely algebraic operator such that $\mathscr{H}^2 = -s^2\mathbf{I}$, where $\mathbf{I}$ denotes the identity operator. It is clear that $\mathscr{H}$ induces a complex structure.

Definition: We shall call self-dual and anti-self-dual states of a zero rest-mass field the eigenvectors of the helicity operator with negative and positive eigenvalue, respectively.

In the forthcoming sections, we shall be concerned with the introduction of a helicity operator at null infinity $\mathscr{I}$, for the wormhole solutions that are under investigation. The key point will be the introduction of the Poincaré generators P_α and $M_{\gamma\delta}$. The Bondi four-momentum, and dual Bondi four-momentum will play a crucial role. Furthermore, as we shall see, both of them find their origin in the nontrivial topology of the space-time, implying a notion of helicity operator, together with self- and antiself-dual states of the gravitational field, which are rooted in the topology.

IV. BONDI AND DUAL BONDI FOUR-MOMENTA, SPIN VECTOR AT NULL INFINITY

Using the various results available at null infinity (which have been summarized in Sec. II) we shall be concerned here with the introduction of a spin vector at $\mathscr{I}$. This vector is to be viewed as a generalization of the null vector $S_\alpha = \epsilon_{\alpha\beta\gamma\delta}P^\beta M^{\gamma\delta}$, which has been investigated by Penrose[1,2] in the case of zero rest-mass particles. For the space-time (with wormholes) under investigation, this vector will describe the total angular momentum of the space-time, and will be a constant multiple of the Bondi four-momentum: these two properties suffice to enable us to view it as a generalization of the spin vector (the Pauli–Lubanski vector) for a gravitational field. The presence of nontrivial homotopy groups related to the bundle structure of $\mathscr{I}$ (or lens space structure if identifications are made along the fibers) and an $S^2\times R$ topology for the manifold of orbits of the stationary Killing field, will be crucial.

Let us first make various remarks concerning the notion of dual mass in vacuum stationary solutions with nontrivial topology: t^α, T, and π will denote the stationary Killing vector field, its manifold of orbits and the projection from $\hat{M}$ into T, respectively. The twist and norm of t^α are defined by $\omega_\alpha = \epsilon_{\alpha\beta\gamma\delta}t^\beta\nabla^\gamma t^\delta$, $\lambda = -g_{\alpha\beta}t^\alpha t^\beta$. It is straightforward to check that the curl-free two-form $F_{\alpha\beta} = \nabla_{[\alpha}\lambda^{-1}t_{\beta]} = \frac{1}{2}\lambda^{-3/2}\epsilon_{\alpha\beta}{}^{\gamma}{}_{\delta}\zeta^\delta\omega_\gamma$ (with $\zeta^\alpha = \lambda^{-1/2}t^\alpha$) is the lift to $\hat{M}$ of a curl free two-form on T defined by $\tilde{F}_{\alpha\beta} = -\lambda^{-3/2}\epsilon^\mu{}_{\alpha\beta}\tilde{\omega}_\mu$, where the alternating tensor is that compatible with the induced metric on T, and $\tilde{\omega}_\alpha$ the gradient of the twist potential $\tilde{\omega}$. The quantity

$$M^* = \frac{1}{8\pi}\int_{S^2}\tilde{F}_{\alpha\beta}\,dS^{\alpha\beta} \qquad (15)$$

is independent of the two-sphere S_2 surrounding the point at infinity on T and does not vanish when the second homotopy group of T is nontrivial. This conserved charge is called the dual mass of the space-time. For all vacuum, stationary solutions for which T has $S^2\times R$ topology, M^* is to be compared to the NUT parameter.[7]

If T, the manifold of orbits of t^α, is asymptotically flat in the sense of Ref. 16, we have available the Hansen dipole angular momentum moment $\underset{\sim}{S}_\alpha$, which is defined by $\underset{\sim}{S}_\alpha = \frac{1}{2}\lim_{\to\Lambda}\mathrm{grad}_\alpha(\tilde{\Omega}^{-1/2}\tilde{\lambda}\tilde{\omega})$, where $\tilde{\Omega}$ is the conformal factor that attaches Λ, the point at infinity, to the manifold T. A simple calculation yields

$$\underset{\sim}{S}\cdot V = \frac{1}{2}\oint_{S^2}(\underset{\sim}{\omega}^\alpha\eta_\alpha)(V\cdot\eta)\,dS,$$

where S^2 is the two-sphere of unit directions η at Λ, V is an arbitrary fixed vector at Λ, and $\underset{\sim}{\omega}_\alpha$ the limiting value of $\tilde{\omega}_\alpha\tilde{\Omega}^{-1/2}$. On the other hand, if the space-time is asymptotically flat at spacelike infinity,[3] with total four-momentum P_α and angular momentum $M_{\gamma\delta}$, the spin vector $S_\alpha = \epsilon_{\alpha\beta\gamma\delta}P^\beta M^{\gamma\delta}$ satisfies[17]

$$S\cdot V = \frac{m}{2}\oint_S(\overline{\omega}^\alpha\eta_\alpha(V\cdot\eta)\,dS,$$

where m is the space-time mass, V any vector at i° (the point at spacelike infinity) chosen such that $V\cdot P = 0$, $\overline{\omega}_\alpha = \lim\Omega^{-1/2}\omega_\alpha$, and S is the two-sphere cross section of the hyperboloid of spacelike directions at infinity defined by $P\cdot\eta = 0$. Thus the spin vector at i° can be viewed as the "pullback" of the Hansen dipole angular momentum moment. Furthermore, if a rotational Killing vector field R^α is available on the space-time, it has been shown[17] that $S_\alpha = jA_\alpha$, where j denotes the value of the Komar integral involving R^α, and A^α is the axis vector induced at i° by R^α.

Theorem: Hansen's dipole angular momentum moment and the spin vector at i° define the same linear mapping on the space of translations at spacelike infinity.

Proof: Recall[18] that translations at spacelike infinity are characterized by functions $v(\eta) = V\cdot\eta$ on the hyperboloid of spacelike directions η at i°, where V is a fixed vector at i°. The result follows provided one identifies any cross section of the hyperboloid with the "pullback" of the two-sphere of directions at Λ (the point at infinity of the manifold of orbits of the stationary Killing field).

Finally we consider the expression of the dual mass as a two-sphere integral on the manifold of orbits T. Using the expression of M^* as introduced in (15) together with the expression of $\widetilde{F}_{\alpha\beta}$ we obtain immediately

$$M^* = \lim_{\to\Lambda}\left(\frac{-1}{8\pi}\right)\oint_{S^2}\lambda^{-3/2}\epsilon^{\mu}{}_{\alpha\beta}\widetilde{\omega}_\mu\, dS^{\alpha\beta}.$$

Using the fact that $\Omega^{-1/2}\widetilde{\omega}_\alpha$ admits a direction dependent limit $\underline{\omega}_\alpha$ at Λ, we obtain $M^* = \lim_{\to\Lambda}\oint_{S^2}\underline{\omega}_\mu\eta^\mu\, dS$. This result is to be compared with the expressions of the total spin vector at i° and Hansen's dipole angular momentum. It suggests that the dual mass can describe the angular momentum of (vacuum) space-times with nontrivial topologies and leads us to the definition of the spin vector at null infinity.

We shall now present this definition. Recall that the total Bondi four-momentum P^α is already available at $\mathscr{I}$. We want to introduce another vector S^α, which should be a multiple of P^α, and display the space-time angular momentum information. The "dual Bondi four-momentum," appropriately defined to encompass the dual mass will be crucial. We first prove useful properties at $\mathscr{I}$.

Lemma 1: If the News tensor $N_{\alpha\beta}$ vanishes, ${}^*K^{\alpha\beta} = f n^\alpha n^\beta$ and $K^{\alpha\beta} = g n^\alpha n^\beta$, where f and g are scalar fields on the two-sphere of orbits of n^α.

Proof: From (6b) and (8) we deduce

$${}^*K^{\alpha\beta} = -2\epsilon^{\alpha\mu\nu}D_\mu S_\nu{}^\beta = -2\epsilon^{\alpha\mu\nu}D_\mu(g^{\beta\tau}\rho_{\nu\tau}),$$

since $S_{\alpha\beta} = \rho_{\alpha\beta}$ and $N_{\alpha\beta} = 0$. Furthermore, $\rho_{\alpha\beta}n^\beta = 0$ and $D_\alpha n^\beta = 0$ imply that ${}^*K^{\alpha\beta} = f n^\alpha V^\beta$; next ${}^*K^{\alpha\beta} = f n^\beta n^\alpha$ follows from the symmetry of ${}^*K^{\alpha\beta}$. Finally $D_\alpha\, {}^*K^{\alpha\beta} = 0$ implies immediately that $n^\alpha D_\alpha f = 0$; f is a function on the manifold of orbits of n^α. Similarly, using (6d) together with ${}^*K^{\alpha\beta} = f n^\alpha n^\beta$, we obtain $K^{\alpha\beta} = g\, n^\alpha n^\beta$, where g is a function on the manifold of orbits of n^α.

Lemma 2: If $\hat{f} n^\alpha$ is the generator of a translation $T_{\hat{f}}$ at $\mathscr{I}$, there is a choice of gauge such that $\hat{f}$ is constant. If $\hat{f}$ is positive, the translation will be called "timelike."

Proof: We know from [Ref. 14, Eq. (37)] that for an infinitesimal translation $\hat{f} n^\alpha$ the equation

$$g^{\alpha\beta}D_\alpha\hat{f}D_\beta\hat{f} - g^{\alpha\beta}D_\alpha D_\beta\hat{f} - \tfrac{1}{2}\mathscr{R}\hat{f}^2 = \text{const} \qquad (16)$$

must be satisfied. Using the available gauge freedom at $\mathscr{I}$, which rescales $\hat{f}$, one can choose $\hat{f} = $ const, from which it follows that $\mathscr{R}$ [as introduced in (2)], the Ricci scalar on the manifold of orbits of n^α, is constant.

Lemma 3: Let l_α denote a vector field at $\mathscr{I}$ such that $l\cdot n = -1$. The vectors $p^\alpha = \hat{f}K^{\alpha\mu}l_\mu$ and ${}^*p^\alpha = \hat{g}\,{}^*K^{\alpha\mu}l_\mu$ are conserved, where $\hat{f}$ and $\hat{g}$, functions on the space of generators of $\mathscr{I}$, generate supertranslations $\hat{f}n^\alpha$ and $\hat{g}n^\alpha$.

Proof: The result follows immediately from Lemma 1 and formula (1).

We shall now introduce several definitions. (From now onwards, $\mathscr{S}$ will denote the space of null generators of $\mathscr{I}$.)

Definition 1: The total Bondi four-momentum $\underline{P}_\alpha$ is defined as an element of the dual of the vector space of translations at null infinity

$$\underline{P}\cdot T_{\hat{f}} = \int_{\mathscr{S}} p^\alpha(\hat{f})\epsilon_{\alpha\mu\nu}\, dS^{\mu\nu}, \qquad (17)$$

where $\hat{f}$ is a function on $\mathscr{S}$ generating the infinitesimal translation $\hat{f}n^\alpha$.

Definition 2: The total Bondi four-momentum ${}^*\underline{P}_\alpha$ is defined as an element of the dual of the vector space of translations at null infinity

$${}^*\underline{P}\cdot T_{\hat{g}} = \int_{\mathscr{S}} {}^*p^\alpha(\hat{g})\epsilon_{\alpha\mu\nu}\, dS^{\mu\nu}, \qquad (18)$$

where $\hat{g}$ is a function on $\mathscr{S}$ generating the infinitesimal translation $\hat{g}n^\alpha$.

Remarks: (i) If $N_{\alpha\beta}$ does not vanish, the previous definitions of $\underline{P}^\alpha$ and ${}^*\underline{P}^\alpha$ have to be modified accordingly. The generalizations, possibly not unique, lead to the notion of four-momentum associated with a partial cross section of $\mathscr{I}$.

(ii) Although the relation between the action of the dual four-momentum on a translation at $\mathscr{I}$ and the dual mass as defined via formula (15) has not been displayed so far, a direct verification could be considered, for instance on the NUT solutions, or within the Newman–Penrose formalism where M^* is given by

$$\int_{\mathscr{S}} \mathrm{Im}\, \eth^2\sigma\, dS, \qquad (19)$$

details being available in Ref. (8). However, the following argument can be proposed.

(iii) Recall that the dual mass, expressed as an integral over the two-sphere of directions at the point at infinity of the manifold of orbits of the stationary Killing vector field is given by

$$\int_{S^2}\underline{\omega}_\mu\eta^\mu\, dS. \qquad (20)$$

Recall also that translations are characterized by functions of the type $\underline{V}\cdot\eta$ (for a fixed vector at i°) on the hyperboloid of spacelike directions at i°. Consider a function f on the manifold of generators of $\mathscr{I}$, generating the translation $f n^\alpha$ at $\mathscr{I}$. The action of the dual Bondi four-momentum ${}^*\underline{P}^\alpha$ on this translation is given by

$${}^*\underline{P}\cdot T_{\hat{f}} = \int_{\mathscr{S}} f\cdot\hat{f}\, dS, \qquad (21)$$

where $f\cdot\hat{f}$ is a function on $\mathscr{S}$. Formulas (20) and (21) suggest that the dual mass is the "projection" (on the manifold of trajectories of the stationary Killing field) of the action of the total dual Bondi four-momentum on a translation at $\mathscr{I}$. Following the notations introduced in Sec. II, we can display the action of ${}^*P^\alpha$ as that of an angular momentum two-form $M_{\gamma\delta}$, element of the Poincaré Lie algebra at $\mathscr{I}$ with the following action:

$$M_{\gamma\delta}(f)\, F^{\gamma\delta}(\hat{f}) = M_{\gamma\delta}(f)\,\epsilon^{\gamma\delta\alpha}(T_{\hat{f}})_\alpha = \int_{\mathscr{S}} f\cdot\hat{f}\, dS,$$

and conjecture that the past limit of the dual Bondi four-momentum must be the total angular momentum at i° if the space-time is asymptotically flat at spacelike infinity.

Since we are concerned with the introduction at null infinity of a spin vector that would mimic the total spin vector at spacelike infinity, we shall now prove the following theorem.

Theorem: If $N_{\alpha\beta} = 0$, the total Bondi four-momentum $\underline{P}^\alpha$ and the total dual Bondi four-momentum $*\underline{P}^\alpha$ are proportional.

Proof: Let us choose four linearly independent translations $T_{\hat f}$ spanning the space of translations at $\mathscr{I}$ and generated by functions $\hat f$ such that Eq. (16) is satisfied. We have already mentioned that we can use the available gauge freedom to impose $\hat f = \text{const}$ on $\mathscr{S}$, the space of orbits of n^α. Using the definition of p^α and $*p^\alpha$, as well as the expressions derived for $\underline{P}^\alpha$ and $*\underline{P}^\alpha$, we obtain immediately

$$\underline{P}\cdot T_{\hat f} = \hat f \int_{\mathscr{S}} g\, dS, \tag{22}$$

$$*\underline{P}\cdot T_{\hat f} = \hat f \int_{\mathscr{S}} f\, dS. \tag{23}$$

This implies that for any choice of a translation $T_{\hat f}$, the quotient $s^{-1}(\underline{P}\cdot T_{\hat f})(*\underline{P}\cdot T_{\hat f})^{-1}$ is independent of $\hat f$, implying that the two four-momenta, $\underline{P}^\alpha$ and $*\underline{P}^\alpha$ are colinear.

Consequence: since $*\underline{P}^\alpha$ carries the angular momentum information and is colinear to the four-momentum $\underline{P}^\alpha$, we want to view it as a natural candidate for the spin vector at $\mathscr{I}$, the generalization of the Pauli–Lubanski vector.[19] We shall set

$$S^\alpha = *\underline{P}^\alpha = \mathscr{s}\underline{P}^\alpha, \tag{24}$$

where $\mathscr{s}$, the (constant) coefficient of proportionality, denotes the total helicity of the space-time.

Definition: The total Bondi dual four-momentum at null infinity is identified with the total spin vector of the wormhole.

Remarks: (i) The results displayed in this section enable us to give further support to this definition. If the space-time is asymptotically flat at spacelike infinity and $N_{\alpha\beta} = 0$, we know, from Ref. 20 that the past limit of $\underline{P}_\alpha$, the total Bondi four-momentum is the Arnowitt–Deser–Misner four-momentum $\underline{P}_\alpha^{\text{ADM}}$. Furthermore, if a rotational Killing vector field is present, and if $\underline{P}_\alpha^{\text{ADM}}$ is a null vector, the total spin vector

$$S^\alpha = \epsilon^{\alpha\beta\gamma\delta}\underline{P}_\beta^{\text{ADM}}\underline{M}_{\gamma\delta}^{\text{ADM}} = \mathscr{s}A^\alpha, \tag{25}$$

where $\mathscr{s}$ is the total helicity of the space-time and A^α the axis of rotation induced at i° by the rotational Killing field.

(ii) The spin vector at null infinity could be introduced using the colinearity of the Bondi and dual Bondi four-momenta when $N_{\alpha\beta} = 0$, i.e., when the asymptotic degrees of freedom of the gravitational field are characterized by $*K_{\alpha\beta}$. One could conjecture that the onset of radiative modes described by $N_{\alpha\beta}$ would destroy the colinearity of $\underline{P}^\alpha$ and $*\underline{P}^\alpha$ these vectors being evaluated on a local (cross) section of $\mathscr{I}$. [Recall that the dual mass, as expressed by formula (19) where $\eth^2$ is essentially a curl, prevents that $\mathscr{I}$ could admit a global cross section when $M^* \neq 0$. In fact, the number of twists in the S^1 fiber bundle over $\mathscr{S}$ could be viewed as a measure of M^*, which vanishes when $\mathscr{I}$ has topology $S^2\times R$.]

(iii) The spin vector at $\mathscr{I}$ is also essentially rooted in the presence of nontrivial homotopy groups. It is tempting to identify vacuum stationary wormholes with gravitational dyons for which the spin vector would play a role analogous to that investigated by Penrose in the case of massless fields.[21]

V. HELICITY OPERATOR AT NULL INFINITY

Recall that the helicity operator $\mathscr{H}$ has been introduced in Sec. III via a purely algebraic method essentially based on the existence of the spin vector S^α. From this vector a spin operator was derived, which could be formally written

$$\underline{S}^\alpha = \epsilon^{\alpha\beta\gamma\delta}\underline{P}_\beta\underline{M}_{\gamma\delta},$$

where $\underline{P}$ and $\underline{M}$ were considered as elements of a Poincaré Lie algebra. As a result, $\underline{S}$ was shown to be a multiple of the Hodge duality operator $\underline{(*)}$ the coefficient of proportionality being the spin s of the field under consideration, implying the following expression for the helicity operator:

$$\mathscr{H} = is(*). \tag{26}$$

For the vacuum stationary wormhole solutions under consideration, we have proved that the presence of a topological charge, the dual mass of the space-time, enabled the introduction at null infinity of a vector S^α, which could be considered as the "pullback" of Hansen's dipole angular momentum vector defined at the point at infinity of the manifold of orbits of the stationary Killing vector field. It was shown that S^α is a constant multiple of the total Bondi four-momentum. Furthermore we know that when the News function vanishes, the past limit of the Bondi four-momentum is the ADM four-momentum if the space-time is asymptotically flat at spacelike infinity. Finally we recalled the fact that, in presence of a rotational Killing vector field, ADM four-momentum and total spin vector at i° are proportional if they are null, the coefficient of proportionality being the space-time helicity. This line of argument supported the identification of S^α with the total spin vector of the (wormhole) space-time. The introduction of the helicity operator at $\mathscr{I}$ follows.

Since there is strong support to consider S^α as the canonical generalization of the Pauli–Lubanski vector available for zero rest-mass particles, the notion of helicity induced by $\mathscr{H}$ on the gravitational degrees of freedom is a generalization of the usual one available (i.e.) for Maxwell fields. We must underline that the gravitational degrees of freedom that are relevant here, and that induce all the topological charges from which the definition of $\mathscr{H}$ is obtained, are those attached to the field $*K_{\alpha\beta}$, since the News function $N_{\alpha\beta}$ vanishes (a crucial point in our derivation of the proportionality of the Bondi, and dual Bondi four-momenta).

In the case of classical zero rest-mass particles, the proportionality of the four-momentum P^α and the spin vector $S^\alpha = \epsilon^{\alpha\beta\gamma\delta}P_\beta M_{\gamma\delta}$ has been obtained in Ref. (2). It seems natural to consider the proportionality of the Bondi and dual Bondi four-momenta as a generalization of this result for the class of topologically nontrivial space-times investigated here.

Definition: We shall call self-dual and anti-self-dual modes of (the degrees of freedom of a) complex zero rest-mass field at $\mathscr{I}$ the eigenstates of the helicity operator $\mathscr{H}$.

Consequence: For the class of vacuum, stationary, topologically nontrivial space-times considered here (wormholes), the existence of the helicity operator at $\mathscr{I}$ is to be traced back to the presence of nontrivial homotopy groups. The notion of self- (resp. anti-self-) duality has thus been shown to be rooted in the topology.

VI. DISCUSSION

We already know[22,23] that for vacuum stationary wormholes (asymptotically flat at spacelike infinity), the presence of a topological charge, the total mass of the space-time given by the Komar integral

$$\int_{S^2} (\nabla_\alpha t_\beta)\epsilon^{\alpha\beta}{}_{\gamma\delta}\, dS^{\gamma\delta} = \int_{S^2} {}^*\mathring{F}_{\alpha\beta}\, dS^{\alpha\beta} \tag{27}$$

over a two-sphere at infinity (or equivalently by an integral over S^2 of an appropriately rescaled component of the Weyl tensor) can be used to define superselection rules for the quantized source-free Maxwell field. The idea is to introduce automorphisms of the *-algebra of field operators $\underline{F}(t)$:

$$T_{m,s}\underline{F}(t) = \underline{F}(t) + \left\{ \int_M (s\mathring{\underline{F}}_{\alpha\beta}(x) + m\, {}^*\mathring{\underline{F}}_{\alpha\beta}(x))\, t^{\alpha\beta}\, dV \right\} \mathrm{I}, \tag{28}$$

which fail to be unitarily implemented on the usual Fock sector if the charge does not vanish (The nonvanishing of this quantity is essentially related to the existence of a nontrivial Poincaré homotopy group on the manifold of orbits of the Killing vector field.) Thus the charges that are superselected on new "photonic sectors" labeled by mass and spin, are the electric and magnetic charge of the Maxwell field.

On another hand, if a rotational Killing vector field R^α is present, there is also available another topological charge, the Komar integral

$$\int_{S^2} (\nabla_\alpha R_\beta)\epsilon^{\alpha\beta}{}_{\gamma\delta}\, dS^{\gamma\delta} \tag{29}$$

(or equivalently a two-sphere integral over an appropriately rescaled component of the Weyl tensor), the total angular momentum of the space-time. This charge could also be used to induce new "photonic" sectors.

In absence of a rotational Killing field, one might be interested to follow other tracks for the superselection of spin. An interesting quantity is the charge called the dual mass, which originates in the nontrivial topology of the manifold or orbits of the stationary Killing field

$$M^* = \int_{S^2} \nabla_{[\alpha}\lambda^{-1}t_{\beta]}\, dS^{\alpha\beta}. \tag{30}$$

We have related the dual mass to the angular momentum and we have been concerned by qualitative effects that could be induced by the presence of this topological charge. We have been working in the asymptotically null regime, since it is expected that a reasonable structure at spacelike infinity might not be available in the generic case. We have proved that the existence of the dual mass enables the introduction at null infinity of a total spin vector of the space-time, derived from the dual Bondi four-momentum, which can be viewed as a natural generalization of the Pauli–Lubanski vector investigated by Penrose for zero rest-mass particles. This spin vector has been rooted in the nontrivial topology of the wormhole. Next, a helicity operator has been introduced at $\mathscr{I}$ (using this spin vector), which turns out to be essentially the Hodge duality operator. The concept of duality turns out to be global. The eigenstates of the helicity operator, the self- and anti-self-dual modes have "knowledge" of the global structure of the wormhole.

The notation of dual mass has been a key point in our investigation. Since this new topological charge emerges when the conformal null boundary of a wormhole is exhibiting a nonstandard topology, and if we take into account its implications concerning the global aspects of the notion of self- (anti-self-) duality, we are tempted to speculate that such wormholes might be basic building blocks.

ACKNOWLEDGMENT

I would like to thank Professor Roger Penrose for his suggestion.

[1]R. Penrose and R. S. Ward, in *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980), Vol. 2.
[2]R. Penrose, M. A. H. MacCallum, Phys. Lett. C 6, 242 (1973).
[3]A. Ashtekar and R. O. Hansen, J. Math. Phys. **19**, 1542 (1978).
[4]R. Penrose, *Battelle Rencontre* (Benjamin, New York, 1968).
[5]A. Ashtekar and M. Streubel, Proc. R. Soc. London Ser. A **376**, 585 (1981).
[6]A. Ashtekar, J. Math. Phys. **22**, 12 (1981); Phys. Rev. Lett. **46**, 9 (1981).
[7]E. T. Newman, L. Tamburino, and T. Unti, J. Math. Phys. **4**, 915 (1963).
[8]E. T. Newman and P. T. Tod, in *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980), Vol. 2.
[9]H. Bondi, M. G. J. Van der Burg, and A. W. K. Mezner, Proc. R. Soc. London Ser. A **269**, 21 (1962).
[10]R. Sachs, Proc. R. Soc. London Ser. A **270**, 103 (1962); Phys. Rev. **128**, 2851 (1962).
[11]R. Penrose, Phys. Rev. Lett. **10**, 66 (1963).
[12]E. T. Newman and T. Unti, J. Math. Phys. **3**, 891 (1962).
[13]E. T. Newman and R. Penrose, Proc. R. Soc. London Ser. A **305**, 175 (1967).
[14]R. Geroch, *Asymptotic Structure of Space-Time*, edited by P. Esposito and L. Witten (Plenum, New York, 1977).
[15]Since a Poincaré generator induces an action described on tensor fields by the Lie derivative, a direct evaluation of this result can be obtained.
[16]R. O. Hansen, J. Math. Phys. **15**, 46 (1974).
[17]A. Ashtekar and A. Magnon-Ashtekar, J. Math. Phys. **20**, 5 (1979).
[18]A. Ashtekar, in *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980), Vol. 2.
[19]P. Roman, *Introduction to Quantum Field Theory* (Wiley, New York, 1969).
[20]A. Ashtekar and A. Magnon-Ashtekar, Phys. Rev. Lett. **43**, 3 (1979).
[21]R. Penrose, in *Quantum Gravity*, edited by C. J. Isham, R. Penrose, and D. Sciama (Clarendon, Oxford, 1975).
[22]A. Magnon-Ashtekar, "Décomposition spatio-temporelle du cadre a l'infini spatial, in *Proceedings of the 1983 Journées Relativises* (Inst. Lagrange, Torino, Italy, 1984).
[23]A. Magnon, "Penrose's conformal techniques, Komar integrals, and quantum phenomena in asymptotically flat space-times with wormholes (submitted to Phys. Rev. Lett. 1985).

Global techniques, dual mass, and causality violation

Anne Magnon[a)]

Physics Department, University of Syracuse, Syracuse, New York 13244-1130 and Département de Mathématiques, Université de Clermont-Fd, 63170, Aubière, France

(Received 9 September 1985; accepted for publication 6 November 1985)

A mathematical framework is presented for the description of (magnetic) monopoles or their gravitational analogs. Using Penrose's global techniques, a proof of a theorem to the effect that, for vacuum space-times with wormhole and an everywhere timelike and complete Killing vector field, the topology must be that of a principal S^1 fiber bundle over $S^2 \times \mathbb{R}$ (topology of the manifold of orbits of the stationary Killing vector field) if the dual mass (the gravitational analog of the magnetic monopole) does not vanish, is presented. Hence the presence of this magnetic charge induces a causality violation: it is shown that it measures the number of windings of the space-time bundle around its fiber, or the periodicity of the timelike closed loops. If, in addition, the manifold of orbits of the stationary Killing field is asymptotically flat, or if a rotational Killing field is present, the resulting expressions of the dual mass reinsure the fact that it should be viewed as a monopole source of angular momentum. The NUT solution is presented as an example of space-time exhibiting the above features. The role of dual mass solutions in quantum gravity is considered.

I. INTRODUCTION

Ever since Dirac advocated their possible existence on account of the observed quantized electric charge, magnetic monopoles have raised increasing interest in spite of the fact that they have not been detected in the laboratory. On one hand it has been speculated that they might have played an important role in the early stages of the evolution of the Universe. On another hand, various (semiclassical) analyses indicate that monopole configurations may play an important role in quantum theories. Finally the role of monopoles has been ever increasing in the context of non-Abelian gauge theories.

Since these various questions are of relevence in the context of general relativity, which also shares many similarities with gauge theories, one might be interested to see whether Einstein's equations allow configurations (solutions) that would be gravitational analogs of the magnetic monopoles. Along these lines it has been suggested that the NUT[1] solution to Einstein's vacuum equation might be an example of a gravitational configuration with a monopole source of mass (or electric charge) and a monopole source of magnetic mass (the NUT parameter).

In analogy with Maxwell theory, it is expected that gravitational fields with nonvanishing magnetic mass will exhibit wire singularities. These singularities could be absorbed in topology changes. As a result causality violations might occur[2] (the NUT solution exhibits closed timelike curves). Consequently, such space-times are not likely to be of interest as models for macroscopic objects. Their interest lies rather within the context of quantum gravity. For instance, if one adopts the Euclidean path integral approach to quantization, one is led to allow four-geometries that might be singular, or even complex, the only requirement being that these paths have finite action.[3] Hence there is *a priori* no reason to rule out magnetic mass configurations. Not only should they be considered as admissible paths, but their role should be crucial since such vacuum solutions extremize the gravitational action functional.

Recently various attempts have been made to present suitable definitions of the notion of magnetic mass.[4,5] Recall that, in the case of a vacuum, stationary and topologically nontrivial space-time, such a definition has already been proposed. On the manifold of orbits of the stationary Killing vector field, one has available the Hansen potentials,[4]

$$\Phi_M = (1/4\lambda)\,(\lambda^2 + \omega^2 - 1) \quad \text{and} \quad \Phi_J = \omega/2\lambda\,, \tag{1}$$

where λ and ω are, respectively, the norm and the twist potential of the Killing vector field. If the manifold of orbits is asymptotically flat in the sense of Ref. 4, the monopole moments at infinity (corresponding to these potentials) are, respectively, the mass and the angular momentum monopole moment or dual mass. In the case of the Schwarzschild space-time, $\Phi_M = (1/4\lambda)(\lambda^2 - 1)$, and $\Phi_J = 0$. Hence the mass monopole (in Φ_M) is nonzero (since $\lambda = -1 + 2M/r$), while the dual mass Φ_J vanishes.

A rotation in the (Φ_M, Φ_J) plane leads to

$$\Phi'_M = \Phi_M \cos 2\theta + \Phi_J \sin 2\theta\,,$$
$$\Phi'_J = -\Phi_M \sin 2\theta + \Phi_J \cos 2\theta\,. \tag{2}$$

For $\theta = \pi/4$, one obtains a new solution with potentials $\Phi'_M = 0$ and $\Phi'_J = -\Phi_M$. This solution is asymptotically flat in the sense of Hansen, and is precisely the NUT metric with NUT parameter $l = -M$; this parameter appears clearly as dual to the mass.

The two forthcoming sections will lead us to a definition that uses crucially the presence of an everywhere timelike Killing vector field but does not require asymptotic flatness of its manifold of orbits. The definition involves the flux of the torsion field—a curvature two-form on the space-time bundle—induced by the twist (covector) of the stationary Killing field. A relation between the two definitions will be proposed in Sec. V, when the manifold of orbits is asymptotically flat in the sense of Hansen.

In absence of stationarity, various strategies can be proposed to introduce a notion that could be viewed as canoni-

[a)] Détachée du Ministère des Affaires extérieures, Paris, France.

 0022-2488/86/041066-08$02.50

cally dual to the mass. A natural approach consists in focusing on the asymptotic degrees of freedom of the gravitational field, and to mimicking the introduction of the magnetic charge in Maxwell's theory. One can summarize the situation as follows. The asymptotic regime at large null separation from the sources is described via the introduction of a three-dimensional null manifold,[5] $\mathscr{I}$, with a degenerate metric, the standard topology being $S^2\times\mathbb{R}$, topology of the null cone at infinity. There exists on $\mathscr{I}$ an asymptotic symmetry group, the BMS group, and physical fields exhibiting the appropriate asymptotic behavior are registered as tensor fields on $\mathscr{I}$. These tensor fields, together with the asymptotic symmetries, give rise to integrals similar to the conserved integrals that arise from symmetries on flat space-time. For instance, the Bondi–Sachs momentum, a real-valued linear function on the asymptotic translations, is expressed as an integral over a cross section of $\mathscr{I}$ (representing an instant of time), the integrand being an appropriate component of the rescaled Weyl tensor. A definition of the dual Bondi momentum has already been proposed. The radiative modes of the gravitational field can be represented[6] by certain equivalence classes $\{D\}$ of derivative operators on $\mathscr{I}^+$ (or $\mathscr{I}^-$, $\mathscr{I}^+\cup\mathscr{I}^-$ playing the role of the abstract null cone at infinity). The physical observables, the Bondi–News N_{ab}, and the radiative part of the dual Weyl tensor $^*K_{ab}$ can be constructed from the curvature of $\{D\}$. The idea, then, is to allow the connection $\{D\}$ to develop "wire singularities" in such a way that the fields N_{ab} and $^*K_{ab}$ remain smooth. Then, the magnetic analog of the Bondi–Sachs energy momentum, the dual four-momentum, can be introduced, which vanishes if $\{D\}$ is smooth. This situation is to be compared with that of the magnetic charge in Maxwell's theory. It has been proved[6] that, unlike the Bondi–Sachs four-momentum, which satisfies balance equations, the dual energy momentum cannot be radiated away, even if gravitational waves carry away energy momentum across $\mathscr{I}^+$. Hence the gravitational field presents itself as analogous to non-Abelian Yang–Mills fields in its electric aspects, while its magnetic aspects mimic the Abelian Maxwell field, reflecting the dual role of the gravitational field in the description of the space-time geometry and the gravitational interaction as well.

If one wants to restrict oneself to C^∞ space-times with C^∞ conformal boundary, it is to be expected that wire singularities could be absorbed in topology changes. In absence of radiation, such a framework has been proposed.[7] A four-momentum involving the asymptotic dual Weyl tensor is introduced at $\mathscr{I}$. This dual four-momentum vanishes identically if $\mathscr{I}$ has the usual $S^2\times\mathbb{R}$ Minkowskian topology. The resulting dual mass, which appears as a generalization of the NUT parameter, does not vanish provided $\mathscr{I}$ is an S^3. A simple consequence of the existence of this dual mass is that the space-time is not stably causal (for a definition of stable causality, see, e.g., Ref. 2). Although no theorem has been proved so far, to this effect, one thus expects that these space-times will exhibit causality violations.

In presence of stationarity, it is also an open question whether the definition involving the asymptotic dual Weyl tensor reduces to that which will be used (and presented in Sec. III) here. This will be studied in a forthcoming paper.

We shall rather focus here on the causal structure, and show that the nonvanishing of the dual mass implies that the orbits of the stationary Killing field must be closed loops, thus violating causality.

Section II will be devoted to mathematical preliminaries suitable for the definition of (magnetic-gravitational) monopoles. Although various frameworks have already been proposed for the description of such monopoles (most of them trying to explain their nonexistence, see, e.g., Ref. 8, and references therein) our definition will be geared to handle nontrivial topologies and extract the geometrical features that are suitable for the proof of the existence of causality violations.

This will be performed in Sec. III. We shall focus on stationary space-times, and the derivation will not require any knowledge of the asymptotic structure of the space-time. We shall make crucial use of the presence of an everywhere timelike and complete Killing vector field, and of an $S^2\times\mathbb{R}$ topology on its manifold of orbits. The result will be that these orbits are closed loops if and only if the dual mass does not vanish. Hence the space-time appears as an S^1 fiber bundle over $S^2\times\mathbb{R}$. Furthermore, the dual mass is proportional to the number of times the bundle winds around its fiber, being essentially the flux of the twist of the stationary Killing field through a two-sphere surrounding the nontrivial topological features.

In Sec. IV, the NUT space-time will be presented as an example of vacuum stationary solution exhibiting these features, the dual mass being the NUT parameter. Closed orbits of the stationary Killing field are obvious in an appropriate chart. In addition, an asymptotic structure is available at null infinity, which confirms (the conformal null boundary being an S^3) the presence of causal "pathologies."

In Sec. V we shall deal with situations suggesting that the dual mass should be viewed as a monopole source of angular momentum, and will consider the relation with the Hansen magnetic mass monopole.

Remarks concerning the role which dual-mass solutions to Einstein's equation might be led to play (e.g., in quantum gravity), in spite of pathologies that seem to make them inappropriate for the description of macroscopic systems, will be presented in the Conclusion.

II. PRELIMINARIES: G-FIELDS, TRANSITION MAPPINGS, MONOPOLES

In this section, we want to display a mathematical framework suitable for the description of (magnetic) monopoles.

A. G-fields, the algebraic structure

The basic elements of our construction will be (a) a finite-dimensional smooth manifold M, and (b) an Abelian group G (displaying the features of the cyclic group) with additive law denoted $+$, and zero element denoted e. Here L will denote the Latin alphabet and N the set of integers. Any element $(i,p)\in L\times \mathrm{N}$ will be called an index letter, for which the symbol i_p will be used, as a subscript or superscript, and $\mathscr{J}$ is the collection of all arrangements ι of subscripts or superscripts, where no index letter appears more than once. Such an arrangement is an index structure. Denote by V the vector space of smooth scalar fields on M, and by V^ι the (infinite-dimensional) real vector space of smooth fields on M with a given index structure $\iota \in \mathscr{I}$. Assume that, for any element $g \in G$, there exists a copy V^ι_g of V^ι: $V^\iota_g = \Psi^\iota_g V^\iota$, where the isomorphism Ψ^ι_g is not necessarily unique, with the exception of Ψ^ι_e (V^ι_e is a natural copy of V^ι and is to be identified with it).

Let $\mathscr{F}_g = \cup_{\iota \in \mathscr{I}} V^\iota_g$ and $\mathscr{S} = \cup_{g \in G} \mathscr{F}_g$. Any element $T^\iota_g \in V^\iota_g$ is called a G-field with weight g and index structure ι. The collection $\mathscr{F}_g$ of G-fields with the same weight g is called the g-fiber. The collection $\mathscr{S}$ of G-fields with arbitrary weight and arbitrary index structure is the set of G-fields.

The tensor algebra structure on $\mathscr{S}$ consists of four laws.

(a) There is an additive law that associates, to any T^ι_g and T''^ι_g in V^ι_g, a third element $T^\iota_g + T''^\iota_g$. This law is that of an Abelian group.

(b) There is an outer product that associates, to any T^ι_g in V^ι_g and any $T'^{\iota'}_{g'}$ in $V^{\iota'}_{g'}$, a third element $T^{\iota''}_{g+g'}$ in $V^{\iota''}_{g+g'}$, where ι and ι' do not have any index letter in common, and $\iota'' = \iota\,\iota'$ is obtained by attaching the index structure ι' to the index structure ι. This outer product will be written $T^\iota_g \cdot T'^{\iota'}_{g'} = T^{\iota\iota'}_{g+g'}$; it is associative and distributive with respect to the addition.

(c) There is a contraction law that consists of suppressing a superscript and a subscript when they are identical.

(d) There is an index substitution that consists of replacing, in the index structure of a G-field, a chosen index letter by another one. Contraction and index substitution commute with the other laws.

Remark: Note that any Φ_g in V_g induces, via outer product, a family of mappings, $\vec{\Phi}_g$: $V^\iota_\kappa \to V^\iota_{\kappa+g}$ ($\forall \kappa \in G$, $\forall \iota \in \mathscr{I}$) such that

$$\forall\, T^\iota_\kappa \in V^\iota_\kappa\,, \quad \vec{\Phi}_g\, T^\iota_\kappa = \Phi_g \cdot T^\iota_\kappa\,.$$

Conversely, if $\Phi_{-g} \cdot \Phi_g$ does not vanish on M, $\overleftarrow{\Phi}_g [\vec{\Phi}_g\, T^\iota_\kappa] \equiv \Phi_{-g} \cdot \Phi_g T^\iota_\kappa = T^\iota_\kappa$. Mappings Φ_g will be called charging mappings for reasons that will become clear. We shall see that for physical reasons, such mappings might not be always regular. It is to be noticed that products $\Phi'_{g+g'} \cdot \Phi_{-(g+g')}$ give rise to functions $\underset{e}{K}(g+g') = (\Phi' \cdot \Phi)_{(g+g')-(g+g')}$.

B. Derivative operators

The next structure to be introduced on $\mathscr{S}$ is a collection of derivative operators together with their curvature and torsion. Their definition will be axiomatic and will suggest a classification of G-fields.

1. Definition

Let $\iota(\underset{\cdot}{a}, \underset{\cdot}{b})$ be, in $\mathscr{I}$, the index structure obtained by attaching the two subscripts a and b. We require that, for any a and $b \in L \times \mathrm{N}$ there exists an element of $V_e^{\iota(a,b)}$ denoted by F_{ab}, such that $\forall g, g_1, g_2 \in G$,

$$\text{(i)}\ F_{ab}(g) = -F_{ba}(g)\,, \tag{3a}$$

$$\text{(ii)}\ F_{ab}(e) = 0\,, \tag{3b}$$

$$\text{(iii)}\ F_{ab}(g_1 + g_2) = F_{ab}(g_1) + F_{ab}(g_2)\,. \tag{3c}$$

A derivative operator ∇_a on $\mathscr{S}$ is a mapping from V^ι_g to $V^{\iota'}_g$, ι' being obtained by attaching the subscript a to the index structure ι, with the following properties:

(i) $\forall\, T^\iota_g$ and $T''^\iota_g \in V^\iota_g$,

$$\nabla_a (T^\iota_g + T''^\iota_g) = \nabla_a T^\iota_g + \nabla_a T''^\iota_g\,, \tag{4}$$

(ii) $\forall\, T^\iota_g \in V^\iota_g$ and $S^{\iota'}_{g'} \in V^{\iota'}_{g'}$,

$$\nabla_a (T^\iota_g \cdot S^{\iota'}_{g'}) = T^\iota_g\, \nabla_a S^{\iota'}_{g'} + (\nabla_a T^\iota_g) S^{\iota'}_{g'}\,, \tag{5}$$

$$\text{(iii)}\ \forall\, \Phi_g \in V_g\,,\ \nabla_{[a} \nabla_{b]} \Phi_g = F_{ab}(g) \Phi_g\,. \tag{6}$$

Here $F_{ab}(g)$ will be called the torsion field of ∇_a. All ∇_a's are assumed to coincide on V_e, where they will be denoted by D_a. The existence of derivative operators on $\mathscr{S}$ will be assumed.

2. Uniqueness

Let ∇_a and ∇'_a be two derivative operators on $\mathscr{S}$. Suppose $\Phi_g \cdot \Phi_{-g}$ does not vanish on M. Then $(\nabla'_a - \nabla_a)\Phi_g$ can also be written $\Phi_{-g}[(\nabla'_a - \nabla_a)\Phi_g]\Phi_g$. Furthermore the expression $\Phi_{-g}(\nabla'_a - \nabla_a)\Phi_g$ does not depend on the choice of the weighted scalar field Φ_g, i.e.,

$\forall\, \Phi_g,\ \hat{\Phi}_g \in V_g$,

$$\Phi_{-g}[(\nabla'_a - \nabla_a)\Phi_g] = \hat{\Phi}_{-g}(\nabla'_a - \nabla_a)\hat{\Phi}_g\,. \tag{7}$$

The proof is the following: since ∇'_a and ∇_a coincide on V_e, we have

$$(\nabla'_a - \nabla_a)(\hat{\Phi}_{-g} \cdot \Phi_g) = 0\,,$$

i.e.,

$$\hat{\Phi}_{-g}(\nabla'_a - \nabla_a)\Phi_g = -\Phi_g(\nabla'_a - \nabla_a)\hat{\Phi}_{-g}\,,$$

hence the result after left multiplication of both members by $\Phi_{-g}\hat{\Phi}_g$. This implies the existence of $\underset{e}{\mu_a}(g) \in \mathscr{F}'_e$ such that

$$(\nabla'_a - \nabla_a)\Phi_g = \underset{e}{\mu_a}(g)\Phi_g.$$

$$\left[\text{Notice that } \underset{e}{\mu_a}(e) = 0.\right] \tag{8}$$

Next, consider $(\nabla'_a - \nabla_a)\underset{g}{K_b}$. Since ∇'_a and ∇_a coincide on V_e, we have for any $f_e \in V_e$: $[\nabla'_a - \nabla_a](f_e \underset{g}{K_b}) = f_e\left[(\nabla'_a - \nabla_a)\underset{g}{K_b}\right]$. Hence $\nabla'_a - \nabla_a$ is a linear functional on $V_g^{\iota(b)}$. Choose any $\Psi_{-g} \in V_{-g}$ such that $\Psi_g \cdot \Psi_{-g}$ does not vanish on M, and consider $[\nabla'_a - \nabla_a](\Psi_{-g} \cdot \underset{g}{K_b})$. Since $\Psi_{-g} \cdot \underset{g}{K_b}$ has zero weight and since $\nabla'_a - \nabla_a$ is linear, there exists $\underset{e}{C_{ab}}{}^m(e) \in V_e^{\iota(a,b,\bar{m})}$ such that $[\nabla'_a - \nabla_a](\Psi_{-g} \cdot \underset{g}{K_b}) = \underset{e}{C_{ab}}{}^m(e)\Psi_{-g}\underset{g}{K_m}$. It follows immediately (using the Liebnitz rule) that

$$[\nabla'_a - \nabla_a]\underset{g}{K_b}$$

$$= \underset{g}{K_m}\left[\underset{e}{C_{ab}}{}^m(e) - \delta_b{}^m \Psi_g(\nabla'_a - \nabla_a)\Psi_{-g}\right]. \tag{9}$$

Using the definition of $\underset{e}{\mu}_a(g)$, the bracket can be written

$$\underset{e}{C}_{ab}{}^m(e) + \delta_b{}^m \underset{e}{\mu}_a(g) \equiv \underset{e}{C}_{ab}{}^m(g) , \tag{10}$$

hence the result.

Similarly, for any $\underset{g}{K}{}^b \in V_g^{\iota(\bar b)}$, one shows (after expansion of $[\nabla'_a - \nabla_a](\underset{g}{K}{}^b \cdot \underset{-g}{\tilde K}{}_b) = 0,\ \forall \underset{-g}{\tilde K}{}_b)$ that $[\nabla'_a - \nabla_a]\underset{g}{K}{}^b = -\underset{e}{C}_{am}{}^b(-g)\underset{g}{K}{}^m$ with $\underset{e}{C}_{ab}{}^c(-g) = \underset{e}{C}_{ab}{}^c(e) - \delta_b{}^c \underset{e}{\mu}_a(g)$ since $\underset{e}{\mu}_a(-g) = -\underset{e}{\mu}_a(g)$. More generally, if $\underset{g}{T}_b{}^c$ is a G-field with mixed indices, we obtain

$$[\nabla'_a - \nabla_a]\underset{g}{T}_b{}^c = \underset{e}{C}_{ab}{}^m(e)\underset{g}{T}_m{}^c - \underset{e}{C}_{am}{}^c(e) \times \underset{g}{T}_b{}^m + \underset{e}{\mu}_a(g)\underset{g}{T}_b{}^c . \tag{11}$$

Hence, given any derivative operator ∇_a on $\mathscr{S}$, the operator ∇'_a is completely determined by the fields $\underset{e}{\mu}_a(g)$ and $\underset{e}{C}_{ab}{}^c(e)$ provided the scalar fields Φ_g under consideration give rise to regular mappings $\vec{\Phi}_g$. It is important to notice that, when the charging mappings Φ_g's fail to be regular, these results could not have been derived, the situation being comparable to a loss of differentiability on a manifold; the charging mappings could be viewed as charts in an atlas where each region would be given a particular weight g.

Finally, it is easy to prove that

(i) $\underset{e}{C}_{ab}{}^c(e) = \underset{e}{C}_{(ab)}{}^c(e)$, (12a)

(ii) $D_{[a}\underset{e}{\mu}_{b]}(g) = 0$, (12b)

(iii) $\forall g, g' \in G,\ \underset{e}{\mu}_a(g+g') = \underset{e}{\mu}_a(g) + \underset{e}{\mu}_a(g')$, (12c)

and

(iv) when G is the Abelian group of complex numbers, $\underset{e}{\mu}_a(g) = g\ \underset{e}{\mu}_a(1)$ and therefore

$$\underset{e}{C}_{ab}{}^c(g) = \underset{e}{C}_{ab}{}^c(e) + g\underset{e}{\mu}_a(1)\,\delta_b{}^c . \tag{12d}$$

C. Charging mappings

In this section we want to show that all derivative operators that coincide on $\mathscr{F}_e$ can be generated from one of them via the action of the charging mappings $\vec{\Psi}_g$. As a result, two structures $(\mathscr{S}, \nabla_a, \mathscr{F}_e)$, $(\mathscr{S}, \nabla'_a, \mathscr{F}_e)$ will be distinct if ∇_a and ∇'_a do not coincide on $\mathscr{F}_e$.

Consider any $T^\iota_g \in V^\iota_g$; choose any $\Psi_g \in V_g$ and $\Psi'_{-g} \in V_{-g}$ such that $\Psi_g \Psi'_{-g}$ does not vanish on M. Then, it is easy to show that

$$\Psi_g \Psi'_{-g} \nabla_a(\Psi'_g \Psi_{-g} T^\iota_g) = \nabla'_a T^\iota_g \tag{13}$$

defines a derivative operator ∇'_a on $\mathscr{S}$: ∇'_a is obviously linear, ∇'_a satisfies the Leibnitz rule (this is easily obtained setting $g = g_1 + g_2$ in the definition of ∇'_a); finally ∇'_a satisfies the torsion equality

$$\nabla'_{[a}\nabla'_{b]}\Phi_g = F_{ab}(g)\Phi_g . \tag{14}$$

Consider now two derivative operators ∇_a and ∇'_a that coincide on $\mathscr{F}_e$ [i.e., such that $\underset{e}{C}_{ab}{}^c(e) = 0$]. We claim the existence of a mapping $\underset{e}{\vec{K}}(g) = \vec{\Psi}'_g \overleftarrow{\Psi}_{-g}$ that sends derivative operator to derivative operator, i.e., which satisfies

$$\forall T^\iota_g \in V^\iota_g ,\quad \underset{e}{K}(-g)\nabla_a\Big[\underset{e}{K}(g)T^\iota_g\Big] = \nabla'_a T^\iota_g . \tag{15}$$

This follows from the fact that $(\nabla'_a - \nabla_a)T^\iota_g = \underset{e}{\mu}_a(g)T^\iota_g$ and $\underset{e}{\mu}_a(g) = D_a\alpha(g)$ imply $\nabla'_a T^\iota_g = e^{-\alpha(g)}\nabla_a[e^{\alpha(g)}T^\iota_g]$, where $\alpha(g+g') = \alpha(g) + \alpha(g')$ and $\alpha(e) = 0$ have enabled us to set $\underset{e}{K}(g) = \exp\alpha(g)$. [Note that $\alpha(g)$ being defined up to the addition of a constant $c(g)$, the mapping generated by $\underset{e}{K}(g)$ is defined up to multiplication by $\exp c(g)$.]

D. Relation between two systems $(\mathscr{S}, \nabla_a)$ and $(\tilde{\mathscr{S}}, \tilde\nabla_a)$

Let $(\mathscr{S}, \nabla_a)$ and $(\tilde{\mathscr{S}}, \tilde\nabla_a)$ be two systems of G-fields with respective (fixed) derivative operators ∇_a and $\tilde\nabla_a$. We claim the existence of a unique one to one onto mapping χ: $\mathscr{S} \to \tilde{\mathscr{S}}$, which takes derivative operator to derivative operator, i.e., such that

$$\forall T^\iota_g \in \mathscr{S},\quad \chi^{-1}\tilde\nabla_a(\chi T^\iota_g) = \nabla_a T^\iota_g$$

provided ∇_a and $\tilde\nabla_a$ coincide on $\mathscr{F}_e$. This follows immediately from $\tilde\nabla_a T^\iota_g = (\exp -\alpha(g))\nabla_a[\exp\alpha(g)T^\iota_g]$, the mapping χ being unique up to multiplication by $\exp c(g)$ in $\mathscr{F}_g$ and $\exp\tilde c(g)$ in $\tilde{\mathscr{F}}_g$.

E. Curvature and torsion of derivative operators

Let $\underset{g}{K}_c \in V^{\iota(c)}$ and Ψ_g generator of a regular charging mapping $\overleftrightarrow{\Psi}_g$. From the definition of the Riemann tensor on zero-weighted fields, and the Leibnitz rule, it follows immediately that

$$\nabla_{[a}\nabla_{b]}\underset{g}{K}_c = \{\underset{e}{R}_{abc}{}^m(e) + F_{ab}(g)\delta_c^m\}\underset{g}{K}_m \equiv \underset{e}{R}_{abc}{}^m(g)\underset{g}{K}_m . \tag{16}$$

Here, $\underset{e}{R}_{abc}{}^d(g)$ will be called the curvature of the system $(\mathscr{S}, \nabla_a)$. Hence, on a flat manifold M, the presence of curvature for derivative operators acting on G-fields is identical to that of torsion:

$$R_{abc}{}^d(g) = \delta^d{}_c F_{ab}(g) . \tag{17}$$

Next, since $\underset{e}{R}_{[abc]}{}^d(g) = \delta_{[c}{}^d F_{ab]}(g)$, one has an identity that reduces to the first Bianchi identity when $g = e$.

Finally, if we expand the equality

$$\nabla_{[d}\,(R_{ab]c}{}^{m}(g)\underset{g}{K}{}_{m}) = \nabla_{[d}\,(\nabla_a\nabla_{b]}\underset{g}{K}{}_{c}\,,$$

it is easy to obtain a generalization of the second Bianchi identity

$$\nabla_{[d}\,(\underset{e}{R}{}_{ab]c}{}^{m}(g) = 0 \quad \text{and} \quad \nabla_{[a}F_{bc]}(g) = 0\,. \tag{18}$$

Hence, the torsion $F_{ab}(g)$ is a closed two-form.

Remark: It is interesting to notice that, if we identify g with a charge, $F_{ab}(g)$ can be viewed as a charged Maxwell field.

Let us assume, now, the existence on $\mathscr{F}_e$ of a derivative operator ∇_a and of a family of two-forms $F_{ab}(g)$ satisfying

$$\text{(i)}\ \nabla_{[a}F_{bc]}(g) = 0\,, \tag{19a}$$

$$\text{(ii)}\ F_{ab}(g+g') = F_{ab}(g) + F_{ab}(g')\,, \tag{19b}$$

$$\text{(iii)}\ F_{ab}(-g) = -F_{ab}(g)\,. \tag{19c}$$

Then, if $A_a(g)$ is any potential for $F_{ab}(g)$, one can define derivative operators on $\mathscr{S}$ via

$$\underset{g}{\nabla}{}_b T^{\iota} = \underset{g}{\Psi}\nabla_b(\underset{-g}{\Psi}\underset{g}{T}{}^{\iota}) + A_b(g)\underset{g}{T}{}^{\iota}\,. \tag{20}$$

The torsion of these operators is $F_{ab}(g)$.

Remark: In the above formula, Ψ_g and A_b determine the derivative operator ∇_a on G-fields.

Conversely, Ψ_g and ∇_a determine A_b; we necessarily have

$$A_b(g) = \Psi_{-g}\nabla_b\Psi_g\,. \tag{21}$$

It is straightforward to show that ∇_a and ∇'_a determined by $\{\Psi_g, A_b(g)\}$ and $\{\Psi'_g, A_b(g)\}$, respectively, satisfy

$$\nabla'_b\underset{g}{T} - \nabla_b\underset{g}{T} = \underset{g}{T}\nabla_b\log(\Psi_g\Psi'_{-g})\,. \tag{22}$$

Conversely, the two vectors $A_b(g)$ and $A'_b(g')$ determined by $\{\Psi_g, \nabla_b\}$ and $\{\Psi'_{g'}, \nabla_b\}$, respectively, satisfy

$$A'_b(g') - A_b(g) = \nabla_b\log(\Psi'_{g'}\Psi_{-g})\,. \tag{23}$$

F. *G*-fields and monopoles

In the previous sections, we have introduced derivative operators and studied their properties. As we already mentioned, most of these properties could be derived because the charging mappings were assumed to be regular. In the absence of such a regularity, the torsion $F_{ab}(g)$ might not even be curl-free, i.e., be viewed as a source free Maxwell field. This leads us to the following considerations.

Definition of a monopole: Consider in M, a two-sphere S^2 (surrounding a topologically trivial region), and a point P located in its interior (see Fig. 1). Let S denote a two-surface passing through P, and intersecting S^2 along a one-dimensional loop γ that can be considered as the boundary of a disk D through P. We shall say that $F_{ab}(g)$ admits a monopole at P if $\epsilon^{abcd}\nabla_b F_{cd}(g) = \rho u^a$ at P, where ρ is called the strength of the monopole and u^a its current. Furthermore, any potential $A_b(g)$ at P, must exhibit a singularity on $M-P$. Suppose $A_b(g)$ were singularity-free on $M-P$. Then, the loop integral $\oint_\gamma A_b\, dS^b = 0$, since the loop is shrinkable to a point Q within the singularity-free region. However, the in-

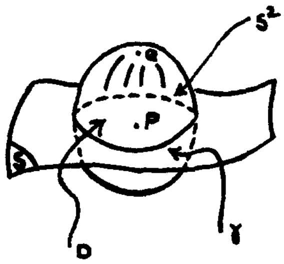

FIG. 1. Monopole inside a two-sphere.

tegral is also equal to $\int\int_\Sigma F_{ab}\, dS^{ab}$, the flux of F_{ab} through a cap of S^2 bordered by γ. Hence the contradiction, and the result. Another possible description of the presence of a monopole inside S^2 consists in removing the point P via creation of topologically nontrivial features inside this two-sphere; for instance, one could introduce a wormhole via removal (as in Fig. 2) of two connected balls surrounding the point P. In this case, $F_{ab}(g)$ could be a source free Maxwell field and yet exhibit a flux through S^2 since this two-sphere is not shrinkable to a point.

Consequence: Suppose now that $M-P$ is covered by charts such that in each chart the torsion field $F_{ab}(g)$ is curl-free with singularity-free potential $A_b(g)$. Let R and R' be two such regions (see Fig. 3). The previous results imply that $A_b(g) = \Psi_{-g}\nabla_b\Psi_g$ in R, and $A'_b(g) = \Psi'_{-g}\nabla_b\Psi'_g$ in R'. Consequently $A_b(g) - A'_b(g) = \nabla_b\log(\Psi'_{-g}\cdot\Psi_g)$ in the intersection of the two regions. Hence the total flux through the two-sphere surrounding the monopole is given by

$$\int\!\!\int_{S^2} F_{ab}\, dS^{ab} = \oint_\gamma (A'_b - A_b)\,dS^b$$

$$= \oint_\gamma \nabla_b\log(\Psi_{-g}\,\Psi'_g)\,dS^b\,, \tag{24}$$

which proves that the flux depends only on the transition mapping between region R and region R'.

The next section will provide an application of the results obtained here.

III. DUAL-MASS MONOPOLES AND CAUSALITY VIOLATION

In this section we want to prove the existence of closed timelike curves in stationary (i.e., with twist) space-times with wormholes. These space-times exhibit a nonvanishing dual mass or magnetic monopole moment. We shall first focus on a definition of the dual mass available for such space-times. The presence of the stationary Killing vector field will be crucial and we shall make no use, whatsoever, of the asymptotic behavior of the space-time.

Let (M, g_{ab}, t^a) be a vacuum stationary space-time, i.e., a four-manifold M with a metric g_{ab} of signature $(-,+,+,+)$ solution of $R_{ab} = 0$, and t^a an *everywhere*

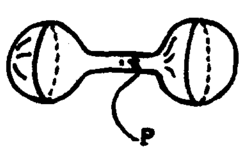

FIG. 2. Creation of a monopole via topological modification.

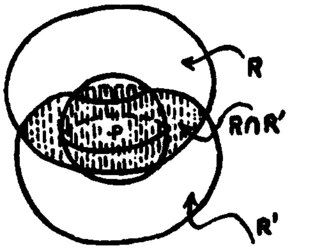

FIG. 3. Chart in the neighborhood of a monopole.

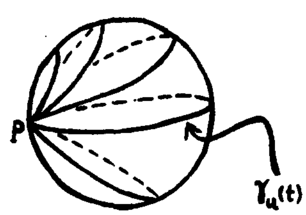

FIG. 4. Spanning of a two-sphere by a one-parameter family of loops.

timelike and complete Killing vector field. Let π: $M \to$T denote the projection map from M into T, the manifold of orbits of t^a. We shall assume that T has topology $S^2 \times \mathbb{R}$, topology of a wormhole. Hence, each spacelike hypersurface Σ_t in M carries a handle. Our purpose is to prove that (M, g_{ab}, t^a) must exhibit a causality violation.

Proof: Since t^a is not hypersurface orthogonal, one has

$$\nabla_a t_b = \lambda^{-1} t_{[b} \nabla_{a]} \lambda + (\lambda)^{1/2} \epsilon_{abcd} t^d \omega^c ,$$

where $t^a t_a = -\lambda$ denotes the norm of t^a and $\omega_a = (\lambda)^{-1/2} \epsilon_{abcd} t^b \nabla^c t^d$ its twist. Consequently there exists on M a curl-free two-form

$$F_{ab} = \nabla_{[a} \lambda^{-1} t_{b]} ,$$

which can be viewed as the pullback $\pi^* (\widetilde{F}_{ab})$ to M, of a curl-free two-form $\widetilde{F}_{ab}$ on T, since $t^a F_{ab} = 0$, and $\mathscr{L}_t F_{ab} = 0$. It is straightforward to check that

$$\widetilde{F}_{ab} = (\lambda)^{-3/2} \tilde{\epsilon}_{mab} \widetilde{\omega}^m = \alpha \epsilon_{ab} , \quad (25)$$

where $\widetilde{\omega}_a$ is a one-form on T inducing, via pullback, the twist of t^a $(\pi^*(\widetilde{\omega}_a) = \omega_a)$, and $\tilde{\epsilon}_{mab}$ is the alternating tensor compatible with $h_{ab} = g_{ab} + \lambda^{-1} t_a t_b$, the metric induced on T by g_{ab} (note that since $R_{ab} = 0$, $\widetilde{\omega}_a$ admits a scalar potential $\widetilde{\omega}_a = \operatorname{grad}_a \widetilde{\omega}$). We shall thereafter denote by $\widetilde{D}_a$ the derivation compatible with h_{ab}. The *dual* (or magnetic) *mass* Q^* associated with the stationary Killing vector field t^a (and the presence of the wormhole) is defined via the flux of $\widetilde{F}_{ab}$ through a two-sphere on T surrounding the topologically nontrivial features

$$Q^* = \int_{S^2_T} \widetilde{F}_{ab}\, dS^{ab} = \int_{S^2_T} (\lambda)^{-3/2} \tilde{\epsilon}_{mab}\, \widetilde{\omega}^m\, dS^{ab}, \quad (26)$$

where the integral is independent of the choice of S^2_T (since $\widetilde{F}_{ab}$ is curl-free) and does not vanish since the second homotopy group of T is nontrivial.

Remark: According to the results obtained in the previous section, it is natural to call $\widetilde{F}_{ab}$ the torsion of the wormhole. It is clearly related to the twist $\widetilde{\omega}^a$.

Consequence: The dual mass is defined as the flux through a two-sphere surrounding the nontrivial topological features, of the torsion field $\widetilde{F}_{ab}$, or equivalently of the twist of the stationary Killing vector field.

We shall now display a geometrical reformulation of Q^*. Let us choose on the manifold of orbits T, a two-sphere S^2_T surrounding the handle. We shall view this two-sphere as a one-parameter family of loops $\gamma_u(t)$, $0<t<1$, $0<u<1$, originating at the same point p $[\gamma_u(0) = \gamma_u(1) = p]$ (see Fig. 4). Denote by Γ the two-dimensional timelike submanifold of M (or cylinder) representing the pullback to M of such a loop, and generated by the integral curves of t^a (see Fig. 5). On this Γ, p is represented by such an integral curve, denoted thereafter by $\mathscr{P}$. Since Γ is timelike, there exists, through each point q of Γ, two null curves C^+ and C^-, which can be parametrized in such a way that their respective tangent vectors k_a and k'_a satisfy $k_a t^a = k'_a t^a = -1$. Let D denote the derivative operator compatible on Γ with the induced metric. From $\mathscr{L}_t k_b = 2 t^m D_{[m} k_{b]}$ and $D_{[m} k_{b]} = \alpha k_{[m} t_{b]}$, one deduces immediately $D_{[m} k_{b]} = 0$. Hence the integral $\int_{\mathscr{C}} k_a\, dS^a$ is independent of the closed curve $\mathscr{C}$ chosen on Γ. Choose first the closed curve $\mathscr{C}_1^+$ obtained moving along C^+ from q to m and along $\mathscr{P}$ from m to q. Since k^a is null the integral $I = \int_{\mathscr{C}_1^+} k_a\, dS^a$ reduces to $\int_m^q \lambda^{-1/2} (k_a t^a) ds$. Similarly one can consider the integral $I' = \int_{\mathscr{C}} k'_a\, dS^a$ and choose for its evaluation the closed loop $\mathscr{C}_2^+$ obtained moving along $\mathscr{P}$ from q to m' and along C^- from m' to q. The integral reduces to $\int_q^{m'} \lambda^{-1/2} k'_a t^a\, ds$. Hence,

$$I + I' = \int_{\mathscr{C}} (k_a + k'_a) dS^a = \lambda_q^{+1/2} \int_{\mathscr{C}} \lambda^{-1} t_a\, dS^a$$

$$= \lambda_q^{-1/2} \int_m^{m'} ds,$$

provided the affine parameter along $\mathscr{P}$ is defined via $\lambda^{-1/2} t^a D_a s = 1$. Since $\mathscr{C}$ can be considered as the lift, to M, of some loop $\gamma_u(t)$ through p, if Σ denotes the area of the cap bordered, on S^2_T, by this loop, one has

$$\lambda_q^{-1/2} \int_m^{m'} ds = \lambda_q^{1/2} \int_{\mathscr{C}} \lambda^{-1} t_a\, dS^a$$

$$= \lambda_q^{1/2} \int_\Sigma \nabla_{[a} \lambda^{-1} t_{b]} dS^{ab}, \quad (27)$$

the flux of the torsion field $\widetilde{F}_{ab}(\omega)$ through the cap.

It is clear that, when u varies from 0 to 1, while $\gamma_u(t)$ sweeps the two-sphere S^2_T, starting from the trivial loop at p (the point) and ending at p, the whole fiber over p will be described. Hence Q^* (the flux of $\widetilde{F}_{ab}$ through S^2_T) will be finite (and nonvanishing) if the fiber is a closed loop. Since the argument does not depend on the point p, M is clearly an S^1 fiber bundle over $S^2 \times \mathbb{R}$, and conversely.

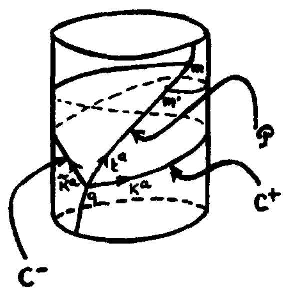

FIG. 5. The cylinder Γ.

Let L denote the period of the fiber (or affine length) and n the number of times the bundle winds around its fiber, then one has the following theorem.

Theorem: The dual mass (or magnetic mass monopole moment) does not vanish iff the space-time is an S^1 fiber bundle over $S^2 \times \mathbb{R}$, and is a measurement of the period of the fiber times the number of windings of the space-time bundle around its fiber.

IV. AN EXAMPLE

We want to consider here the NUT space-time as an example of the considerations displayed in the previous sections. Let[1] (t, r, θ, φ) be the chart in which the metric is given by

$$ds^2 = -U(dt + A\,d\varphi)^2 + U^{-1}\,dr^2 + (r^2 + l^2)(d\theta^2 + \sin^2\theta\,d\varphi^2), \quad (28)$$

where $U = 1 - 2(mr + l^2)/(r^2 + l^2)$ and

$A = 2l(1 - \cos\theta)$.

The constants m and l are, respectively, the mass and the NUT parameter. The coordinate singularity at $\theta = \pi$ can be removed via the introduction of $t' = t + 4l\varphi$. The expression of the metric can be kept if one sets $A' = -2l(1 + \cos\theta)$. It is clear that the orbits of $\partial/\partial t'$ are periodic with period $8\pi l$. This removal of the coordinate singularity induces the topology $S^3 \times \mathbb{R}$ for the space-time manifold, which thus appears as an S^1 fiber bundle over $S^2 \times \mathbb{R}$, the fiber being generated by the orbits of $\partial/\partial t'$. The causality violation induced by the presence of closed timelike loops cannot be removed since the space-time is simply connected. The NUT parameter is a measure of the flux of the torsion field or of the twist of the Killing field $\partial/\partial t'$ through any two-sphere (on T, manifold of orbits of $\partial/\partial t'$) surrounding the topologically nontrivial features. The twist potential ω and norm λ of $\partial/\partial t'$ are smooth functions on T and generate the Hansen potentials defined by $\Phi_M = (1/4\lambda)(\lambda^2 + \omega^2 - 1)$, $\Phi_J = \omega/2\lambda$. Furthermore since T is asymptotically flat,[4] the monopole moments induced at infinity by these potentials give, respectively, the mass and the dual mass (or angular momentum monopole). The angular momentum monopole is dual to the mass in the following sense. Consider, in the Φ_M, Φ_J plane, the rotation defined by

$$\Phi'_M = \Phi_M \cos 2\theta + \Phi_J \sin 2\theta,$$
$$\Phi'_J = -\Phi_M \sin 2\theta + \Phi_J \cos 2\theta.$$

For $\theta = \pi/4$, one obtains a new solution with potentials $\Phi'_M = 0$ and $\Phi'_J = -\Phi_M$. This solution is precisely the NUT metric with NUT parameter $l = -M$. Hence this parameter is in duality with respect to the mass parameter.

Although this solution has no asymptotically reasonable spacelike surfaces, it is possible to study its behavior in the asymptotic null regime. One can define new coordinates

$$x = r^{-1} \quad \text{and} \quad u = t - \int U^{-1}\,dr.$$

Then the metric can be written as

$$\begin{aligned} ds^2 = &-U[du^2 - 2U^{-1}x^{-2}\,du\,dx \\ &+ 2A(du - U^{-1}x^{-2}\,dx)d\varphi] \\ &+ [x^{-2}(1 + l^2x^2)\sin^2\theta - UA^2]d\varphi^2 \\ &+ x^{-2}(1 + l^2x^2)d\theta^2. \end{aligned} \quad (29)$$

Let $\Omega = x$ be a rescaling factor, then $d\tilde{s}^2 = \Omega^2\,ds^2$ is given by

$$\begin{aligned} d\tilde{s}^2 = &-U[x^2\,du^2 - 2U^{-1}\,du\,dx \\ &+ 2A(x^2\,du - U^{-1}\,dx)d\varphi] + (1 + l^2x^2)d\theta^2 \\ &+ [(1 + l^2x^2)\sin^2\theta - UA^2x^2]d\varphi^2. \end{aligned} \quad (30)$$

The conformal boundary at infinity $\mathscr{I}$ is attached via the addition of the points with coordinate $x = 0$. The degenerate metric on $\mathscr{I}$ is given by $d\sigma^2 = d\theta^2 + \sin^2\theta\,d\varphi^2$, hence the manifold of orbits of $\partial/\partial u$ is a metric two-sphere, $\mathscr{I}$ being an S^1 bundle over this two-sphere. Hence the acausality of the space-time is also present in the asymptotic structure.

V. DUAL MASS AS A MONOPOLE SOURCE OF ANGULAR MOMENTUM

We shall consider here two situations that underline the idea that the dual mass could be considered as a monopole source of angular momentum.

Suppose the stationary space-time is now equipped with a rotational Killing vector field R^a such that $\mathscr{L}_t R^a = 0$. Since, at each point of (M, g_{ab}) the two-dimensional subspace (of the tangent space) generated by t^a and R^a is integrable (since t^a and R^a are commuting, $\mathscr{L}_t R^a = 0$), we shall identify the cylinder Γ with their integral manifold. In this case, the previous analysis is still valid. If we choose for $\mathscr{C}$ an integral curve of R^a, we obtain

$$\begin{aligned} &\lambda^{1/2} \int_{\mathscr{C}} \lambda^{-1} t_a\,dS^a \\ &= \lambda^{1/2} \int_{\mathscr{C}} \lambda^{-1} t_a\,R^a\,ds \\ &= \lambda^{-1/2}\,(t \cdot R)(R \cdot R)^{-1/2} \text{ affine length of } \mathscr{C} \end{aligned}$$

(note that $\mathscr{L}_R \lambda = 0$, $\mathscr{L}_t \lambda = 0$ implies that λ is constant on the integral manifold of the two Killing fields).

As a result, the total length of the orbits of t^a is proportional to the scalar product $t \cdot R$ describing the state of rotation of the space-time. The dual mass is therefore related to the total angular momentum of the space-time.

Let us assume now that T, the manifold of orbits of t^a is asymptotically flat in the sense of Ref. 4. One has available in this case the Hansen spin vector defined by

$$\underline{S}_a = \tfrac{1}{2} \lim \operatorname{grad}_a(\tilde{\Omega}^{-1/2}\tilde{\lambda}\tilde{\omega}), \quad (31)$$

where $\tilde{\Omega}$ is the conformal factor which attaches Λ—the point at infinity—to the manifold T. It is then straightforward to show that

$$\underline{S} \cdot V = \frac{1}{2} \int_{\underline{S}^2} (\underline{\omega}^a \eta_a) V \cdot \eta\,dS, \quad (32)$$

where $\underline{S}^2$ is the two-sphere of unit directions η at Λ, V is an arbitrary fixed vector at Λ, and $\underline{\omega}_a$ the limiting value of $\tilde{\omega}_a \tilde{\Omega}^{-1/2}$.

Since the spacelike "translations at infinity" can be characterized by functions of the type $V \cdot \eta$ on the two-sphere of unit directions at Λ (V being a fixed vector at Λ), the Hansen spin vector defines a linear mapping from the vector space of asymptotic spacelike translations into the reals.

On another hand, if the two-sphere S^2_{T}, which has been used in the definition of the dual mass (Sec. III) is "pushed" to infinity on a manifold of orbits T, asymptotically flat in the sense of Ref. 4, the resulting expression of Q^* will be

$$Q^* = \lim_{\to\Lambda} \int_{\underset{\sim}{S}^2} \lambda^{-3/2} \tilde{\epsilon}_{mab} \tilde{\omega}^m \, dS^{ab} = \lim_{\to\Lambda} \int_{\underset{\sim}{S}^2} \omega_a \eta^a \, dS. \quad (33)$$

Since the function associated to t^a on the two-sphere of unit directions at Λ cannot be direction dependent (under reasonable asymptotic conditions), one is led to consider Q^* as the action of a linear mapping on asymptotic "timelike translations." The previous result suggests that Q^* encompasses the angular-momentum content of the space-time. Hence, $(\underset{\sim}{S} \cdot V, Q^*)$ defines the action of a four-momentum vector on the vector space of asymptotic translations, which has the attributes of a spin vector. On the other hand, if a conformal completion is available at null infinity, it has been shown[9] that in the absence of radiation (vanishing Bondi–News) a spin vector is also available, and is a constant multiple of the total Bondi four-momentum. We conjecture that, if the space-time is stationary, with a conformal completion in the null directions, and if T is asymptotically flat in the sense of Ref. 4, $(\underset{\sim}{S} \cdot V, Q^*)$ induces the spin-vector at null infinity.

All these considerations support the fact that the dual mass should be viewed as a monopole source of angular momentum, the state of rotation of the space-time, contained in ω (torsion of the closed orbits), being displayed in a four-spin-vector (at Λ or at $\mathscr{I}$).

VI. CONCLUDING REMARKS

In the main body of this paper, we have proved that (irrespective of their asymptotic behavior) stationary space-times with nonvanishing dual mass must exhibit causality violation: the orbits of the stationary Killing vector field are closed loops. In the absence of stationarity, if a conformal boundary is available at null infinity, the dual Bondi four-momentum involving the asymptotic degrees of freedom of the gravitational field[5] suggests also that dual-mass space-times must be acausal since $\mathscr{I}$ must have the topology of an S^3. Although the agreement of the two definitions of the dual mass is yet to be proved (this will be done in a forthcoming paper) in the presence of an everywhere timelike and complete Killing vector field, the result brings consistency and is of interest in itself since it displays features that could characterize gravitational (magnetic) monopoles.

It has been frequently objected that space-times with nontrivial closed causal curves cannot be realistic for the following reasons: solutions of equations describing the propagation of physical fields might require severe consistency conditions. An observer (with free will) following closed timelike curves should have no difficulty altering past events etc. Consequently it is believed that such space-times might not be suitable for the description of macroscopic systems. Nevertheless there are various features that suggest that their role cannot be neglected in the context of quantum gravity. We shall list some of them.

First, there is available on stationary dual mass space-times two source-free Maxwell fields, $F_{ab} = \nabla_a t_b$ and ${}^*F_{ab}$. If one considers the quantum description, it is known that the automorphism associated to the isometry generated by t^a cannot be unitarily implemented on the usual Fock sector. As a result, one obtains generalized photon states, and a superselection of mass and spin, the new quantum sectors being labeled by parameters emerging from the electric (resp. magnetic) charge of F_{ab} (resp. ${}^*F_{ab}$), which turns out to be the total energy of the space-time. Since in this case, the dual mass $\nabla_{[a}\lambda^{-1} t_{b]}$ is more appropriate for the description of the angular momentum of the space-time, it should be of interest to investigate whether the presence of this charge gives rise to a phenomenon of superselection, and possibly new photon states.

It has also been noticed[9] that dual mass solutions with an asymptotic null regime, and vanishing Bondi–News, enable the introduction of a total spin vector S^a at null infinity, a constant multiple of the total Bondi four-momentum vector, which can be viewed as a generalization of Penrose's result on the Pauli–Lubanski vector for classical zero rest-mass particles. The resulting notions of helicity, helicity operator, self- and anti-self-dual modes are therefore essentially induced by the presence of a nontrivial topology (and nonvanishing dual mass). One could speculate (in view of the possible existence of magnetic (gravitational) monopoles in the early stages of the universe,) whether "dual mass handles" could be viewed as particles "carrying" their own spin vector.

Furthermore, dual mass solutions extremize the gravitational action functional in Euclidean quantum gravity and therefore cannot be discarded in spite of their causal pathologies. They might have a role to play in the S-matrix description of quantum gravity.

ACKNOWLEDGMENT

I would like to thank Professor Roger Penrose for his suggestion.

[1] C. W. Misner, J. Math. Phys. 4, 924 (1963); E. T. Newman, L. Tamburino, and T. Unti, J. Math. Phys. 4, 915 (1963).
[2] R. M. Wald, *General Relativity* (Univ. Chicago P., Chicago, 1984).
[3] A. Magnon-Ashtekar, J. Math. Phys. 22, 2012 (1981); A. Magnon-Ashtekar, "Contribution à l'étude des instantons gravitationnels," in *Proceedings on the 1980 Caen Journées Relativistes*, France (Math. Dept., Caen U., Caen, France, 1980).
[4] R. Hansen, J. Math. Phys. 15, 46 (1974).
[5] A. Ashtekar and A. Sen, J. Math. Phys. 23, 2168 (1982).
[6] R. Penrose, *Structure of Space-Time* (Benjamin, New York, 1967).
[7] S. Ramaswamy and A. Sen, J. Math. Phys. 22, 11 (1981).
[8] R. Sorkin, J. Phys. A: Math. Gen. 10, 5 (1977); R. Sorkin, Phys. Rev. Lett. 51, 2 (1983).
[9] A. Magnon, J. Math. Phys. 27, 1059 (1986).

Birkhoff-type theorem in self-creation cosmology

Tarkeshwar Singh
Department of Mathematics, Harish Chandra Post Graduate College, Varanasi, 221001, India

(Received 23 August 1985; accepted for publication 6 November 1985)

It is shown that an analog of Birkhoff's theorem of general relativity exists in a self-creation theory of gravitation, proposed by Barber, when the scalar field is independent of time.

I. INTRODUCTION

Many theories have been proposed as an alternative to Einstein's theory. The most important among them being the scalar–tensor theory proposed by Brans and Dicke.[1] This theory develops Mach's principle in a relativistic framework by assuming that inertial masses of fundamental particles are not constant, but are dependent upon the particles' interaction with some cosmic scalar field coupled to the large-scale distribution of matter in motion. The Brans–Dicke theory[1] does not allow the scalar field to interact with fundamental particles and photons. By allowing the scalar field to otherwise interact with particle and photon momentum four-vectors, and by thus modifying the Brans–Dicke theory to allow for the continuous creation of matter, Barber[2] has developed a continous creation theory. In this theory the universe is seen to be created out of self-contained gravitational, scalar, and matter fields. The field equations given by Barber[2] are

$$R_{ij}-\frac{1}{2}g_{ij}R=\frac{8\pi}{\phi}T_{(m)ij}-\frac{2}{3\phi\alpha}\phi_{;ij}+\frac{2}{3\phi\alpha}g_{ij}\Box\phi\,, \tag{1}$$

and

$$\Box\phi=4\pi\alpha T_{(m)}\,, \tag{2}$$

where $\Box\phi=\phi_{;k}^{;k}$ is the invariant d'Alembertion and the contracted tensor $T_{(m)}$ is the trace of energy momentum tensor describing all nongravitational and nonscalar field matter and energy. Here α is a coupling constant chosen so that gravitational constant G, which is now a functionof ϕ, can be defined as $G=1/\phi$. Recently Singh and Singh[3] have discussed some general results on spatially homogeneous stationary cosmological models in this theory.

In this paper we have shown that Birkhoff's theorem of general relativity exists in the self-creation theory proposed by Barber[2] when the scalar field is independent of time.

II. BIRKHOFF'S THEOREM IN SELF-CREATION THEORY

It was shown by Birkhoff[4] that every spherically symmetric solution of the Einstein vacuum field equations is static. This fact is known as Birkhoff's theorem. Shücking[5] has shown that this theorem is valid in Jordan's[6] extended theory of gravitation when the gravitational invariant of the theory is independent of time. Reddy[7] has shown that Birkhoff's theorem holds in the Brans–Dicke theory of gravitation when the scalar field introduced in the theory is independent of time. It is also shown, therein, that Birkhoff's theorem is valid in the Sen and Dunn theory,[8] irrespective of the nature of the scalar field. On similar lines we show, here, that Birkhoff's theorem exists in the self-creation theory[2] of gravitation when the scalar field is independent of time.

We consider the spherically symmetric metric in the form

$$ds^2=e^{\nu}\,dt^2-e^{\lambda}\,dr^2-r^2(d\theta^2+\sin^2\theta\,d\Phi^2)\,, \tag{3}$$

where $\lambda=\lambda(r,t)$, $\nu=\nu(r,t)$, with the scalar field $\phi=\phi(r,t)$.

The field equations (1) and (2) for the metric (3) read, in vacuum, as

$$-e^{-\lambda}\left(\frac{\nu'}{r}+\frac{1}{r^2}\right)+\frac{1}{r^2}=-\frac{2}{3\alpha\phi}\left[e^{-\lambda}\phi'\left(\frac{2}{r}+\frac{\nu'}{2}\right)-e^{-\nu}\left(\ddot{\phi}-\frac{1}{2}\dot{\nu}\dot{\phi}\right)\right], \tag{4}$$

$$\begin{aligned}&-e^{-\lambda}\left(\frac{\nu''}{2}-\frac{\lambda'\nu'}{4}+\frac{\nu'^2}{4}+\frac{\nu'-\lambda'}{2r}\right)+e^{-\nu}\left(\frac{\ddot{\lambda}}{2}+\frac{\dot{\lambda}^2}{4}-\frac{\dot{\lambda}\dot{\nu}}{4}\right)\\&=-\frac{2}{3\alpha\phi}\left[e^{-\lambda}\left\{\frac{\phi'}{r}-\frac{1}{2}\lambda'\phi'+\phi''+\frac{1}{2}\phi'\nu'\right\}-e^{-\nu}\left\{\ddot{\phi}-\frac{1}{2}\dot{\phi}\dot{\nu}+\frac{1}{2}\dot{\phi}\dot{\lambda}\right\}\right],\end{aligned} \tag{5}$$

$$e^{-\lambda}\left(\frac{\lambda'}{r}-\frac{1}{r^2}\right)+\frac{1}{r^2}=-\frac{2}{2\alpha\phi}\left[-e^{-\lambda}\left(\frac{1}{2}\phi'\lambda'-\phi''-\frac{2\phi'}{r}\right)-\frac{e^{-\nu}}{2}\dot{\phi}\dot{\lambda}\right], \tag{6}$$

$$-e^{-\lambda}\frac{\dot{\lambda}}{r}=-\frac{2e^{-\lambda}}{3\alpha\phi}\left[\frac{1}{2}\dot{\lambda}\phi'-\dot{\phi}'+\frac{1}{2}\dot{\phi}\nu'\right], \tag{7}$$

$$e^{-\lambda}\left\{\frac{1}{2}\lambda'\phi'-\phi''-\phi'\left(\frac{2}{r}+\frac{\nu'}{2}\right)\right\}+e^{-\nu}\left\{\ddot{\phi}-\frac{1}{2}\dot{\phi}\dot{\nu}+\frac{1}{2}\dot{\phi}\dot{\lambda}\right\}=0\,, \tag{8}$$

where primes denote partial differentiation with respect to r and overdots denote partial differentiation with respect to t.

When the scalar field is a function of r only, that is,

$$\dot{\phi}=0\,, \tag{9}$$

 0022-2488/86/041074-02$02.50

then from Eq. (7) we have

$$\dot{\lambda}\left(\frac{1}{r}-\frac{\phi'}{3\alpha\phi}\right)=0, \tag{10}$$

which implies that either

$$\dot{\lambda}=0 \tag{11}$$

or

$$\frac{1}{r}=\frac{\phi'}{3\alpha\phi}, \quad \text{i.e.,} \quad \phi=\phi_0 r^{3\alpha}, \quad \phi_0=\text{const}, \tag{12}$$

when $\dot{\phi}=0$; from Eqs. (4), (6), and (8) it follows that

$$e^{\lambda}=1-\frac{2r\phi'}{2\alpha\phi}+\frac{r}{2}(\nu'-\lambda'). \tag{13}$$

Using Eq. (12) in Eq. (8) we get

$$\nu'-\lambda'=-(2/r)(3\alpha+1). \tag{14}$$

Substituting this value in Eq. (13) we have

$$e^{\lambda}=-(2+3\alpha), \quad \text{i.e.,} \quad \text{const.} \tag{15}$$

In this case no solution exists. So the only other possibility is valid, i.e., $\dot{\lambda}=0$.

Now differentiation of (14) with respect to t along with the use of (9) and (11) gives

$$\dot{\nu}'=0. \tag{16}$$

From this, we have

$$\nu=f(r)+g(t), \tag{17}$$

where f and g are arbitrary functions of r and t, respectively. We can now use the transformation[9]

$$dt'=e^{g(t)/2}\,dt.$$

This turns ν into a function of r only. This together with (11) reduces the metric (3) to the static case.

Thus we have shown that Birkhoff's theorem of general relativity is true in the self-creation theory proposed by Barber[2] when the scalar field is independent of time.

ACKNOWLEDGMENTS

The author is thankful to Dr. T. Singh for his help and encouragement.

[1]C. Brans and R. H. Dicke, Phys. Rev. **124**, 925 (1961).
[2]G. A. Barber, Gen. Relativ. Gravit. **14**, 117 (1982).
[3]T. Singh and Tarkeshwar Singh, Rep. Math. Phys. (1985) (to appear).
[4]G. D. Birkhoff, *Relativity and Modern Physics* (Harvard U. P., Cambridge, MA, 1927).
[5]E. Shücking, Z. Phys. **148**, 72 (1957).
[6]P. Jordan, *Schwerkraft und Weltall* (Vieweg, Braunschweig, 1952).
[7]D. R. K. Reddy, J. Phys. A: Math. Nucl. Gen. **6**, 1867 (1973).
[8]D. K. Sen and K. A. Dunn, J. Math. Phys. **12**, 578 (1971).
[9]A. Das, Prog. Theor. Phys. **24**, 915 (1960).

Conformally Ricci-flat perfect fluids

N. Van den Bergh
Theoretical Astronomy Unit, School of Mathematical Sciences, Queen Mary College, London, United Kingdom and Physics Department, University of Antwerp, Antwerp, Belgium

(Received 23 September 1985; accepted for publication 20 November 1985)

Classes of inhomogeneous perfect fluid solutions can be obtained by requiring that the associated Weyl tensor corresponds to a nonflat vacuum solution of Einstein's field equations. It is shown how one derives from this assumption useful information on the Newman–Penrose variables. Some particular classes of shear-free perfect fluid solutions are discussed, which all turn out to be locally rotationally symmetric.

I. GENERALITIES AND NOTATION

When searching for inhomogeneous perfect fluid solutions of the Einstein field equations, the most trivial restriction one can put on the Weyl tensor is that it vanishes; this assumption leads to the conformally flat perfect fluid solutions that are all explicitly known.[1] In a next step one can put weaker algebraic restrictions on the Weyl tensor (e.g., concerning Petrov type, the vanishing of its electric or magnetic parts, etc.), which has yielded a rich variety of results.[1–3] Alternatively one could restrict one's attention to perfect fluid solutions for which the space-time metric is conformally related to a vacuum solution, or, more generally, to an Einstein space. Properties of such space-times ("conformally Einstein spaces") were studied a long time ago,[4,5] but their use for generating perfect fluid solutions has been neglected because the restrictions on the geometry are—at first sight—not strong enough.

Suppose in fact that $\hat{g}_{ab}$ is the metric of a vacuum space-time and g_{ab} the metric of a perfect fluid solution with pressure p, matter density w, and fluid velocity u, then

$$\hat{R}_{ab} = 0 \tag{1.1}$$

and

$$R_{ab} - \tfrac{1}{2}Rg_{ab} = (w+p)\, u_a u_b + p g_{ab}. \tag{1.2}$$

When a scalar field ϕ exists such that

$$\hat{g}_{ab} = e^{2\phi} g_{ab}, \tag{1.3}$$

we will call (g_{ab}, p, w) a conformally Ricci-flat perfect fluid. A search for all such fluids (without further restrictions on p and w) would be too ambitious, as these would contain as a subset all the vacuum solutions, by taking $\phi = \text{const}$ in (1.3). Hence we impose the additional—although customary—restriction

$$p + w \neq 0. \tag{1.4}$$

Now the curvature tensors of the two solutions $\hat{g}_{ab}$ and g_{ab} are connected by the relations[1]

$$e^{2\phi}\hat{R}^{ab}{}_{cd} = R^{ab}{}_{cd} + 4Y^{[b}_{[c}\delta^{a]}_{d]}, \tag{1.5}$$

with

$$Y_{ab} = \phi_{,a;b} - \phi_{,a}\phi_{,b} + \tfrac{1}{2}g_{ab}\phi_{,e}\phi^{,e}_{,} \tag{1.6}$$

(; denoting the covariant derivative with respect to g_{ab}). From (1.1) one then obtains

$$2Y_{ab} = R_{ab} - \tfrac{1}{6}Rg_{ab} \tag{1.7}$$

and hence, by (1.2),

$$2Y_{ab} = (w+p)\, u_a u_b + \tfrac{1}{3}wg_{ab}. \tag{1.8}$$

A necessary and sufficient condition for (g_{ab}, p, w) being conformally Ricci-flat, is that a solution ϕ exists for Eq. (1.8). This leads then to a set of integrability conditions on (g_{ab}, p, w),

$$2Y_{a[b;c]} = C^{d}{}_{abc}\phi_{,d}, \tag{1.9}$$

which, with the aid of the Bianchi identities, can be rewritten as

$$C^{a}{}_{bcd;a} = C^{a}{}_{bcd}\phi_{,a}. \tag{1.10}$$

The full set of integrability conditions for (1.7) was given recently by Kozameh *et al.*,[6] without, however, insisting on the fact that the Ricci tensor should be of the perfect fluid type. As these conditions (in particular the vanishing of the Bach tensor) contain second-order derivatives of the curvature, they do not directly loan themselves to extract useful information on the possible perfect fluid solutions. Therefore it is preferable to investigate, e.g., the tetrad components of (1.7) or (1.8), directly. As heavy use has to be made of the Bianchi identities it is preferable to employ the Newman–Penrose formalism, which, by its use of complex quantities, has the advantage of reducing the length of the occurring expressions. It is the latter approach that will be followed in the next paragraphs. In Sec. II, Eqs. (1.8) and (1.9) will be reformulated in the Newman–Penrose (NP) formalism, whereas in Sec. III, as an example, a discussion will be given of shear-free perfect fluid solutions. These solutions all turn out to be conformally flat or to be locally rotationally symmetric (admitting a G_4 on T_3); their explicit form is given in the Appendix. It is hoped that in the general case (for nonvanishing shear) the present method might lead to some new physically relevant inhomogeneous cosmological models, or alternatively, to new exact solutions for perfect fluid matter in local equilibrium (e.g., stationary axisymmetric and differentially rotating fluids).

II. CONFORMALLY RICCI-FLAT PERFECT FLUIDS IN THE NEWMAN–PENROSE FORMALISM

Notation and sign conventions of Kramer *et al.*[1] are followed throughout: the null tetrad is taken as $e_a = (m, \bar{m}, l, k)$ with $k_a l^a = -1$ and $m^a \bar{m}_a = 1$; the direc-

 0022-2488/86/041076-06$02.50

tional derivatives are $D = k^a\partial_a, \Delta = l^a\partial_a$, and $\delta = m^a\partial_a$.

The tetrad can always be chosen such that

$$\delta\phi = u^1 = u^2 = 0 \quad \text{and} \quad u^3 = u^4 = 1/\sqrt{2}. \tag{2.1}$$

It is then fixed up to rotations in the $(m,\bar{m})$ plane, except when $D\phi = \Delta\phi$ (such that u and $\nabla\phi$ are aligned): in the latter case it is preferable to fix the null direction k (choosing it as one of the principal null directions of the Weyl tensor in the algebraically general case, or as the most repeated principal null direction in the algebraically special case) after which a boost and a null rotation about k can be applied to arrive again at the conditions (2.1). The remaining tetrad freedom is then given by the spatial rotations as in the general case.

The components of (1.6) and (1.8), i.e., of

$$\phi|_{a;b} = \phi|_a\phi|_b - \tfrac{1}{2}g_{ab}\phi|_m\phi|^m + \tfrac{1}{2}(w+p)\, u_a u_b + \tfrac{1}{6}w g_{ab} \tag{2.2}$$

(with $|_a$ denoting the directional derivative along e_a) read then, using $\phi|_e\phi|^e = -2D\phi\Delta\phi$,

$$\bar{\lambda}D\phi - \sigma\Delta\phi = 0, \tag{2.3}$$

$$\bar{\nu}D\phi - \tau\Delta\phi = 0, \tag{2.4}$$

$$\bar{\pi}D\phi - \kappa\Delta\phi = 0, \tag{2.5}$$

$$\bar{\mu}D\phi - \rho\Delta\phi + D\phi\Delta\phi + \tfrac{1}{6}w = 0, \tag{2.6}$$

$$\Delta^2\phi - (\Delta\phi)^2 + (\gamma + \bar{\gamma})\Delta\phi = (w+p)/4, \tag{2.7}$$

$$D^2\phi - (D\phi)^2 - (\epsilon + \bar{\epsilon})D\phi = (w+p)/4, \tag{2.8}$$

$$D\Delta\phi + (\epsilon + \bar{\epsilon})\Delta\phi = (w+3p)/12. \tag{2.9}$$

The remaining components can be obtained equally from the commutator relations, which yield [using (2.4) and (2.5)]

$$(\Delta D - D\Delta)\phi = (\gamma + \bar{\gamma})D\phi + (\epsilon + \bar{\epsilon})\Delta\phi, \tag{2.10}$$

$$\delta D\phi = (\bar{\alpha} + \beta)D\phi, \tag{2.11}$$

$$\delta\Delta\phi = -(\bar{\alpha} + \beta)\Delta\phi, \tag{2.12}$$

$$(\bar{\mu} - \mu)D\phi + (\bar{\rho} - \rho)\Delta\phi = 0. \tag{2.13}$$

Notice that

$$D\phi \neq 0 \neq \Delta\phi, \tag{2.14}$$

otherwise (2.7) or (2.6) would yield $w + p = 0$.

Finally one has to substitute in the NP equations the following expressions for the Ricci tetrad coefficients:

$$\Phi_{00} = \Phi_{22} = 2\Phi_{11} = (w+p)/4 \quad \text{and} \quad R = w - 3p$$

(all others being 0). (2.15)

Henceforth we will refer to the NP equations and Bianchi identities by their numeration in Kramer *et al.*,[1] preceded by *; e.g., *29 will stand for Eq. 7.29 in Kramer *et al.* The conservation laws for the fluid can then be expressed as (*69–71):

$$Dw + \Delta w = (\gamma + \bar{\gamma} - \epsilon - \bar{\epsilon} + \rho + \bar{\rho} - \mu - \bar{\mu})\,(w+p), \tag{2.16}$$

$$Dp - \Delta p = -(\gamma + \bar{\gamma} + \epsilon + \bar{\epsilon})\,(w+p), \tag{2.17}$$

$$\delta p = \tfrac{1}{2}(\kappa + \tau - \bar{\pi} - \bar{\nu})\,(w+p). \tag{2.18}$$

A first step is now made by taking derivatives of (2.3)–(2.6), eliminating the second-order terms in ϕ by (2.7)–(2.10), and eliminating the terms containing derivatives of the spin coefficients by taking suitable linear combinations of the NP equations. The resulting algebraic expressions between the Weyl tetrad coefficients ψ_i, the spin coefficients, and $D\phi$, $\Delta\phi$ are listed below:

$$\psi_0\Delta\phi = \tfrac{1}{4}(w+p)\,(\bar{\lambda} - \sigma), \tag{2.19}$$

$$\bar{\psi}_4 D\phi = \tfrac{1}{4}(w+p)\,(\bar{\lambda} - \sigma), \tag{2.20}$$

$$\bar{\psi}_2 D\phi = \tfrac{1}{4}(w+p)\,(\rho - \bar{\mu}) - \tfrac{1}{6}Dw, \tag{2.21}$$

$$\psi_2\Delta\phi = \tfrac{1}{4}(w+p)\,(\rho - \bar{\mu}) - \tfrac{1}{6}\Delta w, \tag{2.22}$$

$$\bar{\psi}_3 D\phi - \psi_1\Delta\phi = \tfrac{1}{4}(w+p)\,(\tau + \bar{\pi} - \kappa - \bar{\nu}), \tag{2.23}$$

$$\bar{\psi}_3 D\phi + \psi_1\Delta\phi = -\tfrac{1}{6}\delta w. \tag{2.24}$$

From (2.21) and (2.22) one obtains $\bar{\psi}_2 D\phi - \psi_2\Delta\phi = \tfrac{1}{6}(\Delta w - Dw)$, and hence, by taking the conjugate,

$$(\bar{\psi}_2 - \psi_2)\,(D\phi + \Delta\phi) = 0. \tag{2.25}$$

I also list below the expressions for the kinematical quantities, as derived from the usual decomposition of the covariant derivative of the velocity $u_{a;b}$:

$$\dot{u}_1 = -\tfrac{1}{2}(\kappa + \tau - \bar{\pi} - \bar{\nu}), \tag{2.26}$$

$$\dot{u}_3 = -\dot{u}_4 = -\tfrac{1}{2}(\gamma + \bar{\gamma} + \epsilon + \bar{\epsilon}),$$

$$\omega^1 = -(i/2)\,(\bar{\tau} - \bar{\kappa} + \pi - \nu - 2\alpha - 2\bar{\beta}), \tag{2.27}$$

$$\omega^3 = -\omega^4 = (i/2)\,(\rho - \bar{\rho} + \mu - \bar{\mu}),$$

$$\sigma_{12} = \sigma_{34} = -\sigma_{33} = -\sigma_{44} = (1/6\sqrt{2})\,(2\gamma + 2\bar{\gamma} - 2\epsilon - 2\bar{\epsilon} - \rho - \bar{\rho} + \mu + \bar{\mu}),$$

$$\sigma_{14} = -\sigma_{13} = (1/4\sqrt{2})\,(\tau + \bar{\pi} - \kappa - \bar{\nu} + 2\bar{\alpha} + 2\beta), \tag{2.28}$$

$$\sigma_{11} = (1/\sqrt{2})\,(\bar{\lambda} - \sigma),$$

$$\theta = -(1/\sqrt{2})\,(\gamma + \bar{\gamma} - \epsilon - \bar{\epsilon} + \rho + \bar{\rho} - \mu - \bar{\mu}). \tag{2.29}$$

In the next paragraph an example will be given of how to obtain explicit perfect fluid solutions by imposing, e.g., the extra condition that the fluid has vanishing shear.

III. SHEAR-FREE CONFORMALLY RICCI-FLAT PERFECT FLUIDS

Let us suppose first that the gradient of ϕ is aligned with the fluid velocity:

$$D\phi = \Delta\phi. \tag{3.1}$$

Then (2.4) and (2.5) imply

$$\bar{\nu} = \tau \quad \text{and} \quad \bar{\pi} = \kappa, \tag{3.2}$$

whereas (2.7), (2.8), (2.10), and (2.13) yield, respectively,

$$\gamma + \bar{\gamma} + \epsilon + \bar{\epsilon} = 0 \tag{3.3}$$

and

$$\bar{\mu} - \mu + \bar{\rho} - \rho = 0. \tag{3.4}$$

Furthermore one obtains from (2.11) and (2.12),

$$\bar{\alpha} + \beta = 0, \tag{3.5}$$

which, when substituted together with (3.2)–(3.4) in (2.26) and (2.27), shows that the fluid is irrotational ($\omega_{ab} = 0$) and geodesic ($\dot{u}_a = 0$).

When it is also shear-free, the evolution equations for the kinematical quantities[7] show immediately that solutions are conformally flat.

As suggested by (2.25) we will now further restrict to solutions for which the gradient of ϕ is *orthogonal* to the fluid velocity:

$$D\phi + \Delta\phi = 0. \tag{3.6}$$

(In fact this is the only case in which shear-free and nonconformally flat solutions can exist.)[8]

Let us first see to which relations this assumption leads, without imposing the additional restriction that $\sigma_{ab} = 0$: From (3.6) and (2.3)–(2.12) one infers

$$\begin{aligned}\bar\lambda + \sigma &= \bar\nu + \tau = \bar\pi + \kappa \\ &= \epsilon + \bar\epsilon - \gamma - \bar\gamma = \bar\alpha + \beta = 0.\end{aligned} \tag{3.7}$$

This implies $(\Delta D - D\Delta)\phi = 0$, and hence $D^2\phi = \Delta^2\phi = -D\Delta\phi$. From (2.7)–(2.9) one obtains then the relations

$$(D\phi)^2 + 2(\epsilon + \bar\epsilon)D\phi + (2w + 3p)/6 = 0 \tag{3.8}$$

and

$$D^2\phi + (\epsilon + \bar\epsilon)D\phi + (w + 3p)/12 = 0. \tag{3.9}$$

Eliminating next ψ_2 from (2.21) and (2.22) yields

$$\tfrac{1}{2}(w + p)\,(\rho + \bar\rho - \mu - \bar\mu) = \tfrac{1}{3}(Dw + \Delta w)$$

or, by (2.16),

$$\rho + \bar\rho = \mu + \bar\mu. \tag{3.10}$$

On the other hand (2.6) shows that $\rho + \bar\mu = \bar\rho + \mu$, such that

$$\rho = \mu, \tag{3.11}$$

and hence, by (2.21) and (2.22),

$$Dw + \Delta w = 0. \tag{3.12}$$

Equation (3.11) allows one now to eliminate $D\phi$ from (3.8) and (2.6), yielding the following algebraic relationships between w, p, $\epsilon + \bar\epsilon$, $\rho + \bar\rho$, and $D\phi$:

$$2w\Omega^2 - 6(w + p)\,(\rho + \bar\rho)\,\Omega - 3(w + p)^2 = 0 \tag{3.13}$$

and

$$\Omega D\phi = -\tfrac{1}{2}(w + p), \tag{3.14}$$

with Ω defined by

$$\Omega = \rho + \bar\rho + 2(\epsilon + \bar\epsilon). \tag{3.15}$$

Considering now (2.23) one has

$$\bar\psi_3 + \psi_1 = (\kappa - \tau)\Omega. \tag{3.16}$$

On the other hand Eqs. (*31,32) and (*42,45) yield the relations

$$\delta(\epsilon + \bar\epsilon) = -2\bar\pi(\epsilon + \bar\epsilon) + \kappa\mu - \bar\pi\bar\rho - \psi_1 \tag{3.17}$$

and

$$\delta(\gamma + \bar\gamma) = 2\tau(\gamma + \bar\gamma) + \tau\mu - \overline{\nu\rho} + \bar\psi_3, \tag{3.18}$$

which, by (3.7), are compatible with (3.16) and which can be used to eliminate the $\delta(\epsilon + \bar\epsilon)$ term arising by taking the δ derivative of (3.8).

This results in [using (2.18) and (2.24) to substitute also for δp and δw]

$$2D\phi(\kappa\Omega - \psi_1) = 2D\phi\,(\bar\psi_3 - \psi_1) - \tfrac{1}{2}(\kappa + \tau)\,(w + p)$$

or, cf. (3.14),

$$\bar\psi_3 = \tfrac{1}{2}(\kappa - \tau)\Omega, \tag{3.19}$$

and hence, by (3.16),

$$\bar\psi_3 = \psi_1 = \tfrac{1}{2}(\kappa - \tau)\Omega. \tag{3.20}$$

Then (3.24) becomes

$$\delta(\epsilon + \bar\epsilon) = \tfrac{1}{2}(\kappa + \tau)\Omega \tag{3.21}$$

and (2.24) yields

$$\delta w = 0. \tag{3.22}$$

Further information can be inferred from the Bianchi identities. This is at present under investigation, but only the shear-free case has been worked out so far, because it gives additional considerable simplifications.

If one assumes $\sigma_{ab} = 0$, then (2.28) gives [using (3.7)]

$$\lambda = \sigma = 0 \quad \text{and} \quad \kappa = \tau \tag{3.23}$$

such that by (2.19), (2.20), and (3.20) the solutions are obviously of Petrov type D:

$$\psi_0 = \psi_4 = \psi_1 = \psi_3 = 0. \tag{3.24}$$

Furthermore Bianchi identity (*61) yields $\kappa\psi_2 = 0$, such that, for nonconformally flat solutions, one necessarily has

$$\nu = \pi = \tau = \kappa = 0. \tag{3.25}$$

As it is clear from (2.26) and (2.27) that the fluid's velocity, acceleration, and rotation are now confined to the (k,l) plane, it follows from (3.23) and (3.25) that the solutions—if they exist— are likely to be locally rotationally symmetric.[1] In order to check this, further information from the NP equations and Bianchi identities is required. Listed below are the remaining nontrivial NP equations (3.26)–(3.31) and Bianchi identities (3.32)–(3.35), which have been simplified as far as possible under the present conditions:

$$\delta\rho = 0, \tag{3.26}$$

$$D\rho = \rho^2 + \rho(\epsilon + \bar\epsilon) + (w + p)/4, \tag{3.27}$$

$$D\alpha - \bar\delta\epsilon = (\rho + \bar\epsilon - \epsilon)\alpha, \tag{3.28}$$

$$D\gamma - \Delta\epsilon = -(\epsilon + \bar\epsilon)\,(\epsilon + \gamma) + (w + 3p)/12 + \psi_2, \tag{3.29}$$

$$\delta\alpha + \overline{\delta\alpha} = \rho^2 + 4\alpha\bar\alpha + (\rho - \bar\rho)\,(\gamma + \epsilon) + w/6 - \psi_2, \tag{3.30}$$

$$\Delta\bar\alpha + \delta\gamma = \bar\alpha(\gamma - \bar\gamma - \rho), \tag{3.31}$$

$$D\psi_2 - \tfrac{1}{6}Dw = 3\rho\psi_2 + (\rho - \bar\rho)\,(w + p)/4, \tag{3.32}$$

$$(D + \Delta)\psi_2 = 0, \tag{3.33}$$

$$\delta\psi_2 = \delta p = \delta(\epsilon + \bar\epsilon) = 0, \tag{3.34}$$

with ψ_2 given by (2.21) and (3.14) as

$$\psi_2 = \tfrac{1}{2}(\rho - \bar\rho)\Omega + \tfrac{1}{3}\Omega(w + p)^{-1}Dw. \tag{3.35}$$

Notice also that the commutator relations (*55–58) simplify to the following:

$$\Delta D - D\Delta = (\epsilon + \bar\epsilon)\,(D + \Delta), \tag{3.36}$$

$$\bar\delta\delta - \delta\bar\delta = (\bar\rho - \rho)\,(D + \Delta) + 2(\alpha\delta - \overline{\alpha\delta}), \tag{3.37}$$

$$\delta D - D\delta = (\bar{\epsilon} - \epsilon - \bar{\rho})\delta, \tag{3.38}$$

$$\delta\Delta - \Delta\delta = (\rho - \gamma + \bar{\gamma})\delta. \tag{3.39}$$

By (3.36) one can see that Eq. (3.29) [together with (3.35)] is precisely the integrability condition for the following system:

$$D\theta = (i/2)\ (\bar{\rho} - \rho - 2\bar{\epsilon} + 2\epsilon),$$
$$\Delta\theta = (i/2)\ (\bar{\rho} - \rho - 2\bar{\gamma} + 2\gamma). \tag{3.40}$$

Taking then any solution θ of (3.40) and applying the spatial rotation $m \to m' = e^{i\theta}m$, one obtains in the new tetrad [using (3.7), (3.21), (3.25), and (3.26)]

$$\epsilon = \gamma, \tag{3.41}$$

$$\bar{\epsilon} - \epsilon = \tfrac{1}{2}(\bar{\rho} - \rho), \tag{3.42}$$

$$\delta\epsilon = \bar{\delta}\epsilon = 0. \tag{3.43}$$

The situations for a twisting or nontwisting null ray k will now be worked out separately.

A. Twisting null rays ($\bar{\rho} - \rho \neq 0$)

From (3.26), (3.34), and (3.37) one infers immediately

$$(D+\Delta)\rho = (D+\Delta)\ (\epsilon + \bar{\epsilon}) = (D+\Delta)\psi_2$$
$$= (D+\Delta)p = 0. \tag{3.44}$$

By (2.17) and (3.29) one has then

$$Dp = -\ (\epsilon + \bar{\epsilon})\ (w + p) \tag{3.45}$$

and

$$D(\epsilon + \bar{\epsilon}) = -\ (\epsilon + \bar{\epsilon})^2 + (w + 3p)/12$$
$$+ \tfrac{1}{3}\Omega(w+p)^{-1}Dw. \tag{3.46}$$

With the aid of (3.27) this allows one to reduce the imaginary part of (3.32) to

$$\tfrac{1}{3}\Omega(w+p)^{-1}Dw = (w + 3p)/6 - 2\rho\bar{\rho} + \tfrac{1}{2}\Omega(\rho + \bar{\rho}), \tag{3.47}$$

which, when substituted in (3.35), gives

$$\psi_2 = (w + 3p)/6 - 2\rho\bar{\rho} + \rho\Omega. \tag{3.48}$$

Herewith the real part of (3.32) becomes identically satisfied under Eqs. (3.27), (3.45), (3.46), and (3.47), which form a system of five (real) equations for $\rho + \bar{\rho}$, $\rho - \bar{\rho}$, $\epsilon + \bar{\epsilon}$, w, and p. Between the latter quantities there is also the algebraic relationship (3.13), which, however, turns out to be an integral of this system.

Introducing the Stewart and Ellis[9] variables,

$$\dot{u} = \sqrt{2}\dot{u}_4 = \sqrt{2}\ (\epsilon + \bar{\epsilon}), \tag{3.49a}$$

$$\omega = -\ (1/\sqrt{2})\omega_4 = (i/\sqrt{2})\ (\rho - \bar{\rho}), \tag{3.49b}$$

$$a = (1/\sqrt{2})\ (\rho + \bar{\rho}), \tag{3.49c}$$

and

$$r = \rho^2 + \bar{\rho}^2 - 2(\epsilon + \bar{\epsilon})\ (\rho + \bar{\rho}) - p$$
$$= a^2 - 2a\dot{u} - \omega^2 - p, \tag{3.49d}$$

the relevant equations go over to the particular case $\tau = 0$ of the field equations of Stewart and Ellis (substitute $\sqrt{2}D$ for their ∂_1 derivative):

$$2\sqrt{2}D\dot{u} = 3p + w - 4\omega^2 - 2\dot{u}^2 + 4a\dot{u}, \tag{3.50}$$

$$2\sqrt{2}Da = p + w - 2\omega^2 + 2a^2 + 2a\dot{u}, \tag{3.51}$$

$$\sqrt{2}Dp = -\ (p + w)\dot{u}, \tag{3.52}$$

$$\sqrt{2}D\omega = \omega(2a + \dot{u}), \tag{3.53}$$

$$\sqrt{2}Dr = 2ar, \tag{3.54}$$

with w given by (3.13), i.e.,

$$4w(a + \dot{u})^2 - 12(w + p)\ a\ (a + \dot{u}) - 3(w + p)^2 = 0, \tag{3.55}$$

and with (3.54) identically satisfied under (3.50)–(3.53).

Once a solution of (3.50)–(3.55) is obtained, the NP coefficient α follows from (3.28), (3.30), and (3.31), i.e.,

$$D\alpha = \tfrac{1}{2}(\rho + \bar{\rho})\alpha, \tag{3.56}$$

$$\Delta\alpha = -\ \tfrac{1}{2}(\rho + \bar{\rho})\alpha, \tag{3.57}$$

$$\delta\alpha + \bar{\delta}\bar{\alpha} = 4\alpha\bar{\alpha} + \tfrac{1}{2}r. \tag{3.58}$$

Introducing a solution R of

$$DR = -\ \Delta R = (\rho + \bar{\rho})R$$

and

$$\delta R = 0 \tag{3.59}$$

(e.g., $R = \frac{1}{8}|r|$ if $r \neq 0$), one has, with $k = \operatorname{sgn} r$,

$$\alpha = R^{1/2}(u + iv), \tag{3.60}$$

$$D(u + iv) = \Delta(u + iv) = 0, \tag{3.61}$$

and

$$\delta(u + iv) + \bar{\delta}(u - iv) = 4R^{1/2}(u^2 + v^2 + k).$$

Note that, when $r = 0$, (3.50)–(3.53) reduces to a system of three equations in $\dot{u}$, a, and ω, as (3.52) becomes then an identity. Also note that no other solutions with $r = \text{const}$ are possible, because then $a = 0$, and substitution of the solutions, given by Stewart and Ellis for this particular case, in (3.55) leads to an inconsistency. The solution of (3.60) and (3.61) can easily be found after the introduction of a coordinate system. Note that α can be made real with the aid of an additional spatial rotation $m \to e^{i\theta}m$, keeping at the same time $D\theta = \Delta\theta = 0$.

Some standard manipulations lead then to

$$\omega^1 = \tfrac{1}{4}(dy + i\Sigma(y,k)dz),$$
$$\omega^3 = (1/4\sqrt{2})\omega R^{-1}(dt + \sigma(y,k)dz) - dx, \tag{3.62}$$
$$\omega^4 = (1/4\sqrt{2})\omega R^{-1}(dt + \sigma(y,k)dz) + dx,$$

with

$$\Sigma(y,1) = \sin y, \quad \sigma(y,1) = -\cos y,$$
$$\Sigma(y,0) = y, \quad \sigma(y,0) = \tfrac{1}{2}y^2, \tag{3.63}$$
$$\Sigma(y,-1) = \sinh y, \quad \sigma(y,-1) = \cos y,$$

which gives us the canonical form of the class I locally rotationally symmetric (LRS) space-times[1]

$$ds^2 = dx^2 - \tfrac{1}{16}R^{-2}\omega^2\,(dt + \sigma(y,k)dz)^2$$
$$+ \tfrac{1}{8}R^{-1}(dy^2 + \Sigma^2(y,k)dz^2). \tag{3.64}$$

As the conformal scalar field ϕ also satisfies $(D+\Delta)\phi = \delta\phi = \bar{\delta}\phi = 0$, one has $\phi = \phi(x)$, such that the vacuum solution $\hat{g}_{ab}$ corresponding to (g_{ab},p,w) [cf. (1.31)] is necessarily LRS of class I, too. The latter's form being explicitly known,[10] this enables one to reduce the system (3.50)–(3.55) to a single first-order ordinary differential equation (cf. the Appendix), the solutions of which can all be explicitly obtained.

B. Nontwisting null rays ($\rho - \bar{\rho} = 0$)

It is clear now from (3.42) that $\epsilon = \gamma$ is real. However it is not immediately obvious that one should have as before $(D+\Delta)p = (D+\Delta)\epsilon = 0$ [which could be derived earlier from the commutator relation (3.37) and $\delta p = \delta\epsilon = 0$]. To demonstrate that the same relations still hold, take first the Δ derivative of (3.9), yielding

$$(D+\Delta)(2\epsilon D\phi + (w+3p)/12) = 0 \tag{3.65}$$

and hence, by (3.12),

$$(D+\Delta)(2\epsilon D\phi + \tfrac{1}{4}p) = 0. \tag{3.66}$$

On the other hand, $D+\Delta$ acting on (3.14) yields

$$(D+\Delta)(2\epsilon D\phi + \tfrac{1}{4}p) + (D+\Delta)\rho = 0, \tag{3.67}$$

such that

$$(D+\Delta)\rho = 0. \tag{3.68}$$

With the aid of (3.27) and (3.36), one obtains then

$$2\rho(D+\Delta)\epsilon + \tfrac{1}{4}(D+\Delta)p = (D+\Delta)D\rho = (D+2\epsilon)(D+\Delta)\rho = 0. \tag{3.69}$$

Together with (3.66) this implies

$$(\rho - D\phi)\cdot(D+\Delta)\epsilon = 0. \tag{3.70}$$

Now $\rho = D\phi$ is impossible, as then (3.14), (3.27), and (3.68) would yield $Dp = \Delta\rho = 0$, such that by (3.26), ρ and $D\phi(=\rho)$ would be constants. Substituting this in (3.10) and (3.27) gives then

$$\rho^2 + 2\rho\epsilon + (w+p)/4 = 0 = 2\rho\epsilon + (w+3p)/12,$$

which would imply $w = \text{const}$ and hence, by (3.35), $\psi_2 = 0$ such that the solution would be conformally flat.

Hence from (3.69) and (3.70) one can conclude

$$(D+\Delta)\epsilon = (D+\Delta)p = 0. \tag{3.71}$$

As before one recovers now Eqs. (3.45) and (3.46), but now with ϵ real.

The one remaining difficulty is that the differential equation (3.47) does not follow in a straightforward way [as (3.32) is now a real equation].

Therefore consider first the following identity, which results from the commutator relations (3.38) and (3.39):

$$\delta\Delta\alpha + \bar{\delta}\Delta\bar{\alpha} - \Delta(\delta\alpha + \overline{\delta\alpha}) \equiv \rho\,(\delta\alpha + \overline{\delta\alpha}). \tag{3.72}$$

Using (3.30) and (3.31) together with $\delta\gamma = 0$, this leads us to

$$\psi_2 = 4\epsilon\rho + (w+3p)/6, \tag{3.73}$$

which is precisely the $(\epsilon = \bar{\epsilon},\ \rho = \bar{\rho})$ particular case of (3.48). Substituting this in (3.32) gives us (3.47) as before,

$$\tfrac{1}{3}\Omega(w+p)^{-1}Dw = (w+3p)/6 + 4\epsilon\rho. \tag{3.74}$$

Hence, all equations governing the twisting case are recovered (subject only to $\epsilon = \bar{\epsilon}, \rho = \bar{\rho}$), the only difference lying in the form of the coordinate system: the solution being now irrotational ($\rho - \bar{\rho} = 0$), one can introduce coordinates $\zeta = (1/\sqrt{2})(X + iY)$ and x,t (real) such that

$$\omega^1 = \xi\, d\zeta \quad \omega^3 = A\, dt, \quad \omega^4 = \omega^3 + \sqrt{2}\, dx. \tag{3.75}$$

Acting with the commutators (3.36)–(3.39) on ζ, x, and t leads then to

$$\xi = \tfrac{1}{2}R^{-1/2}(x)\cdot P^{-1}(\zeta,\bar{\zeta}) \quad (P \text{ real}), \tag{3.76}$$

$$\frac{\partial}{\partial x}(\log A) = 2\sqrt{2}\,\gamma(x) \quad \text{and} \quad \frac{\partial P}{\partial \zeta} = \bar{\alpha} = u - iv.$$

After a spatial rotation $m' = e^{i\theta}m$ with $\theta = \theta(\zeta,\bar{\zeta})$ such that α becomes real, one has in addition from (3.58) and (3.76)

$$u_X = \sqrt{2}P^{-1}(u^2 + k),$$
$$P_X = \sqrt{2}u \tag{3.77}$$

[and with as before $k = \operatorname{sgn} r = \operatorname{sgn}(2\rho^2 - 8\epsilon\rho - p)$].

This shows that the spaces ($x = \text{const}$, $t = \text{const}$) are two-spaces of constant curvature $K \equiv 2P^2(\partial^2/\partial\zeta\,\partial\bar{\zeta})\log P = 2k$, such that the solutions are locally rotationally symmetric of Ellis' class II (in particular IIc, as a and $\dot{u}$ in (3.50) and (3.51) are now necessarily $\neq 0$). Choosing a new t coordinate and (y,z) to be well-determined functions of X and Y, the metric is then

$$ds^2 = dx^2 - A^2(x)\,dt^2 + \tfrac{1}{8}R^{-1}(dy^2 + \Sigma^2(y,k))\,dz^2, \tag{3.78}$$

with A given by (3.76), i.e., $(\partial/\partial x)(\log A) = \dot{u}$.

As in the twisting case, the conformal scalar field ϕ depends on x only, such that the system [(3.50) and (3.51)] can be reduced to a single first-order ordinary differential equation, by making use of the canonical form of the class II LRS vacuum solutions. The explicit solutions are given in the Appendix.

ACKNOWLEDGMENTS

I thank J. Aman, G. Joly, R. Kerr, and M. MacCallum (Queen Mary College) for helpful comments and Professor I. W. Roxburgh for the hospitality provided at The School of Mathematical Sciences, Queen Mary College.

I also thank the Royal Society and the National Fund for Scientific Research (Belgium) for a grant under which this work was carried out.

APPENDIX: A CLOSED FORM EXPRESSION FOR THE SOLUTIONS OF SEC. III

We will obtain here explicit expressions for the solutions discussed in Sec. III from the form of the corresponding vacuum solutions.

First one rewrites both twisting and nontwisting metrics (3.64) and (3.78) as

$$ds^2 = dt^2 - A^2(t)\,(dx + l\sigma(y,k)\,dz)^2 + \tfrac{1}{8}R^{-1}(t)\,(dy^2 + \Sigma^2(y,k)\,dz^2), \tag{A1}$$

with σ and Σ as defined by (3.63) and with $(d/dx)\log A = \dot{u}$, $(d/dx)\log R = 2a$, and $l = +1$ or 0 for (3.64) or (3.78), respectively. In both cases the rotation scalar ω can be written as [(3.49) and (3.59)]

$$\omega = 4lAR. \tag{A2}$$

Defining now

$$f = A^2, \quad Y = (8R)^{-1/2}, \tag{A3}$$

a coordinate transformation

$$y = y(\zeta,\bar{\zeta}), \quad z = z(\zeta,\bar{\zeta}),$$
$$X = \int A(t)\, dt, \quad x = x(T,\zeta,\bar{\zeta}) \tag{A4}$$

exists, such that (A1) becomes[1]

$$ds^2 = -f(X)\left[dT + \frac{il}{2}\frac{\zeta\, d\bar{\zeta} - \bar{\zeta}\, d\zeta}{1 + (k/2)\zeta\bar{\zeta}}\right]^2$$
$$+ f^{-1}(X)\, dX^2 + Y^2(X)\frac{2\, d\zeta\, d\bar{\zeta}}{(1 + (k/2)\zeta\bar{\zeta})^2}. \tag{A5}$$

The general vacuum solution for (A5) is[10] (with X replaced by χ)

$$f = h(\chi) \equiv (\chi^2 + l/4)^{-1}[k(\chi^2 - l/4) - 2m\chi],$$
$$Y = U(\chi) \equiv (\chi^2 + l/4)^{1/2}. \tag{A6}$$

Notice that m is the only integration constant appearing in the solution, as the two other integration constants entering in the general solution of (3.50), (3.51), and (3.53) (with $p = w = 0$) can be eliminated by rescalings and translations of χ. As remarked in Sec. III, one has for the scalar field $\phi = \phi(X)$, such that by (1.3) the coefficients of (A5) are related to those of (A6) by

$$Y = e^{-\phi}U,$$
$$f = e^{-2\phi}h \tag{A7}$$

and

$$X' = e^{-2\phi}$$

(with $'$ denoting henceforth the derivative with respect to χ). From $\dot{u} = (d/dx)\log A$ and $2a = (d/dx)\log R$ one obtains then

$$\dot{u} = e^{\phi}h^{1/2}(\tfrac{1}{2}h^{-1}h' - \phi') \tag{A8}$$

and

$$a = e^{\phi}h^{1/2}(\phi' - U^{-1}U'). \tag{A9}$$

Furthermore (A2) yields

$$\omega = (l/2)\, e^{\phi}h^{1/2}U^{-2}. \tag{A10}$$

The two remaining functions w and p, appearing in (3.50)–(3.53), can be equally expressed in terms of ϕ and the known functions h and U, by using (3.14) and (3.49d), which yield, respectively,

$$p = e^{2\phi}h\,[3\phi'^2 + U^{-2}U'^2 - 4\phi'U^{-1}U' - h^{-1}h'$$
$$\times(\phi' - U^{-1}U') - \tfrac{1}{4}lU^{-4} - kh^{-1}U^{-2}] \tag{A11}$$

and

$$w + p = e^{2\phi}h\phi'(2U^{-1}U' - h^{-1}h'). \tag{A12}$$

Substitution of (A8)–(A12) in (3.50)–(3.55) gives rise now to a single second-order ordinary differential equation in ϕ, which is reducible to a first-order one, as one integration constant can be made redundant by a global scale transformation.

First, it is obvious that (3.53) and (3.54) are identically satisfied. One can show that the same holds for (3.55). This leaves one just with the equations (3.50) and (3.51), as (3.52) will be automatically satisfied under the other equations. Now both equations (3.50) and (3.51) can be reduced, by making use of the explicit forms (A6), to the single condition

$$\phi'' + \phi'^2 + \phi'(\tfrac{1}{2}h^{-1}h' - 2U^{-1}U') = 0, \tag{A13}$$

i.e., to

$$e^{\phi} = 2\int Uh^{-1/2}\, d\chi \tag{A14}$$

(after elimination of an integration constant by rescaling ds^2).

This yields, with χ_0 constant,

$k = 0$ (and, e.g., $m < 0, x \geqslant 0$),

$$e^{\phi} = \sqrt{2}\cdot|m|^{-1/2}(\tfrac{2}{5}\chi^{5/2} + (l/2)\chi^{1/2} + \chi_0); \tag{A15}$$

$k = \pm 1$,

$$e^{\phi} = 3k\epsilon\,(m^2 + l/4)\cos(h)^{-1}(|\chi - km|/(m^2 + l/4)^{1/2})$$
$$+ (3m + k\chi)\,(k\chi^2 - 2m\chi - kl/4)^{1/2} + \chi_0 \tag{A16}$$

[$\epsilon = \mathrm{sgn}(\chi - km)$, cosh when $k = +1$ and cos when $k = -1$].

Substituting (A14) back in (A8)–(A12) yields, with h, U, and ϕ given by (A6), (A15), and (A16),

$$\dot{u} = e^{\phi}h^{-1/2}U^{-4}(m\chi^2 + kl\chi/2 - ml/4) - 2U, \tag{A17}$$

$$a = 2U - \chi e^{\phi}h^{1/2}U^{-2}, \tag{A18}$$

$$p = 24U^2 - 2e^{\phi}h^{-1/2}U^{-3}(4k\chi^3 - 6m\chi^2 - ml/2), \tag{A19}$$

$$w = 24U^2 + 2\chi e^{\phi}h^{1/2}U^{-1}. \tag{A20}$$

Notice that it is clear from the latter two expressions that solutions with $k = 0$ cannot have both $p > 0$ and $w > 0$.

[1]D. Kramer, H. Stephani, M. A. H. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (Cambridge U. P., Cambridge, 1980).
[2]C. B. Collins, J. Math. Phys. **25**, 995 (1984).
[3]A. J. White and C. B. Collins, J. Math. Phys. **25**, 332 (1984).
[4]H. W. Brinkman, Math. Ann. **91**, 269 (1924).
[5]A. S. Petrov, *Einstein-Räume* (Akademie, Berlin, 1964).
[6]C. N. Kozameh, E. T. Newman, K. P. Tod, Gen. Relativ. Gravit. **17**, 343 (1985).
[7]A. Barnes, Gen. Relativ. Gravit. **4**, 105 (1973).
[8]N. Van den Bergh, Lett. Math. Phys. (to be published).
[9]J. M. Stewart and G. F. R. Ellis, J. Math. Phys. **9**, 1072 (1968).
[10]M. Cahen and L. Defrise, Commun. Math. Phys. **11**, 56 (1968).

On the microscopic derivation of the finite-temperature Josephson relation in operator form

A. Rieckers and M. Ullrich
Institut für Theoretische Physik, D-7400 Tübingen, Federal Republic of Germany

(Received 27 September 1984; accepted for publication 22 October 1985)

As a microscopic description of the Josephson junction, two BCS models are studied in the strict pair formulation with quite an arbitrary weak coupling potential. The modular formalism, the separate gauge transformations, and the limiting dynamics are analyzed for the interacting system in terms of the GNS representation of the uncoupled limiting Gibbs state. By means of the Connes theory the condensed Cooper pair and the quasiparticle spectrum is shown to be stable against weak perturbations. The modular formalism is used to construct a local approximation to the renormalized particle number operator and, by this, its time dependence, in spite of this observable not being affiliated with the von Neumann algebra of the temperature representation. The time derivation from this unbounded operator-valued function coincides with the limit of the local currents and splits under a natural assumption into a sum of the Josephson and the quasiparticle current operator extending the two-fluid picture also to the coupled model.

I. INTRODUCTION

One of the most fundamental questions of quantum mechanics, the relationship to macroscopic physics, has obtained additional actuality by the discovery of the so-called macroscopic quantum phenomena. Experiments, especially with weakly coupled superconductors, demonstrated that typical quantum properties may be amplified to the macroscopic level by the process of quantum condensation (cf., e.g., Refs. 1 and 2). A typical observable is the supercurrent, which is manifestly macroscopic but seems to have quantum features also. At first sight it appears almost impossible to combine the attributes "macroscopic" and "quantum mechanical" within the description of one phenomenon; in fact "macroscopic" is often used synonymously with "classical."

We discuss these problems here in terms of a simple model for the Josephson junction consisting of the weak coupling of two (spatially homogeneous) BCS models. By the use of operator algebraic techniques we are able to formulate a self-consistent point of view on the indicated questions that is generalizable to other systems with quantum condensation and spontaneous symmetry breakdown. A detailed analysis of this procedure will be given in the Conclusions after the exposition of the model construction. Here let us merely indicate the central idea, how to include macroscopic aspects into a quantum theory. In the framework of operator algebraic quantum statistics[3] the purely microscopic observables are contained in the so-called quasilocal C^*-algebra. The macroscopic classical observables appear only in representation-dependent extensions (by means of the weak operator topology) of this algebra as central elements. A macroscopic quantum theory—and this is a new aspect of our investigation—is obtained, if the generators of the broken symmetries in a representation are included in the set of observables. Both types of macroobservables are not present in usual many-body physics, which in some sense refers to the quasilocal algebra only. They arise theoretically only by representation-dependent limiting procedures following well-established mathematical techniques, but they constitute an essential extension of the explicative power of microscopic many-body physics. The treatment of the macroscopic observables belongs also to those features, where our investigation differs principally from the treatment in Ref. 4 of the same model.

For the present investigation we make use of earlier results on the single BCS model,[5–8] where especially the latter two references are adapted to the discussion to follow. There (in Ref. 7) one finds the appropriate particle number operator, which also counts, besides normal electrons, condensed Cooper pairs, if the temperature is low enough, and which constitutes, in virute of this together with the macroscopic phase operator, a canonical pair of observables in the sense of quantum mechanics. Since the macroscopic phase operator is in the center of the representation von Neumann algebra its ambivalent nature of being partly quantum mechanical and partly classical has found a theoretical formulation. In this reference it is also worked out that the physical dynamics (in contradistinction to the reduced KMS dynamics) acts nontrivially on some central elements as, e.g., the phase operator. We shall here also refer to the gauge-covariant quasiparticle formalism and the resulting two-fluid picture of Ref. 8.

The description of the Josephson junction to be discussed in this work consists of a composition of two strong coupling BCS models in the strict pair formulation (sometimes called the quasispin formulation), which interact by means of a tunneling potential that is weak in a twofold sense: a finite number of pairs (since single electrons are excluded in this formalism) tunnels with a finite frequency and a global tunneling process involving infinitely many pairs in the thermodynamic limit takes place with a rate that decreases rapidly with the size of the system. The latter type of interaction is the origin for the supercurrent and its dominant part leads exactly to the Josephson expression. The normal part of the current is kept quite arbitrary.

In Sec. II we discuss the limiting Gibbs states of the uncoupled and interacting model in terms of their central

 0022-2488/86/041082-11$02.50

decompositions, where the result for the coupled system is taken over from Ref. 9. For any weak coupling in the sense specified above and for any inverse temperature $\beta\in(0,+\infty)$ the two kinds of limiting Gibbs states lead to unitarily equivalent GNS representations, in sharp contrast to strong interactions. Due to the low-temperature behavior of the condensed Cooper pair exchange term the limiting ground states ($\beta=+\infty$) of the coupled and the uncoupled systems, however, are singular with respect to each other and no Josephson effect is predicted. For finite temperatures one may use the uncoupled GNS representation also for the interacting system, and we elaborate there the modular formalism connected with the noninteracting equilibrium state. That is, we show that the modular quantities are reduced by a family of local sub-von Neumann algebras and there may be written out explicitly.

In Sec. III we discuss the renormalized particle number operators for the subsystems and suggest a local approximation procedure by making use of the modular involution. This is the starting point to extend the theoretical formalism beyond those observables that are locally approximable in the usual sense, which amounts to being affiliated with the representation von Neumann algebra. The limiting dynamics is analyzed in terms of the perturbation expansion with respect to the weak interaction and various local limiting relations are worked out, among which a certain approximation of the finite-time Heisenberg transformations is most important for the later current calculation.

The quasiparticle formalism of Ref. 8 is used to write the multiphase von Neumann algebra of the coupled system as a W^*-tensor product of the center and a factor. Under a natural assumption the limiting Heisenberg dynamics factorizes accordingly into a central part without any coupling and a part in the factor, which, by means of the Connes theory, is shown to contain always the uncoupled quasiparticle spectrum.

In Sec. IV time-dependent renormalized particle number operators are introduced by means of the local approximation constructed earlier. By time differentiation one obtains a time-dependent current that also may be obtained by direct local approximation. Under the mentioned natural splitting assumption the total current operator decomposes into the supercurrent of the Josephson form and the quasiparticle current.

In the Conclusions we discuss to what extent the operator algebraic model construction of weakly coupled superconductors reflects basic features of macroscopic quantum phenomena and describes a state of the composite system, which is "intermediate between complete separation and complete union" as anticipated by Josephson.[10]

II. THE LOCAL MODEL AND ITS THERMODYNAMIC LIMITING REPRESENTATION

As an extremely simplified description of a Josephson junction we consider here a closed system consisting of two BCS models in the strict pair formulation with a weak interaction. As is elaborated in, e.g., Refs. 7 and 11, the relevant wave vectors for the electrons of each superconductor are localized in a shell $\mathscr{K}_x$ around the Fermi surface, where $x\in\{a,b\}$ is the index for the subsystems. To every finite volume corresponds a finite subset $\Lambda_x\subset\mathscr{K}_x$ and we shall assume $\mathscr{K}_x$ to be denumerable. For a given Λ_x we form the electron pair algebra $\mathscr{A}^P_{\Lambda x}$ which is generated $*$-algebraically by the pair annihilation operators

$$b_{kx}:=C_{-k\downarrow x}C_{k\uparrow x},\quad k\in\Lambda_x, \tag{2.1}$$

where the $C_{k\sigma x}$ are the usual electron annihilation operators with $\sigma\in\{\uparrow,\downarrow\}$ the spin eigenvalues, and $\mathscr{A}^P_{\Lambda x}$ has $\otimes_{k\in\Lambda x}[b_{kx},b^*_{kx}]_+$ as unit element. The quasilocal pair algebra for a subsystem is denoted by $\mathscr{A}^P_x$. The (quasi-) local structure of the combined system is based on the finite subsets $\Lambda_a\times\Lambda_b$ of the denumerable set $\mathscr{K}=\mathscr{K}_a\times\mathscr{K}_b$. Denoting by $|\Lambda|$ the cardinality of Λ we write

$$\mathscr{L}:=\{\Lambda\subset\mathscr{K};\quad |\Lambda|<+\infty\}, \tag{2.2}$$

and obtain the inductive system of local algebras $\{\mathscr{A}_\Lambda:=\mathscr{A}^P_{\Lambda a}\otimes\mathscr{A}^P_{\Lambda b};\ \ \Lambda\in\mathscr{L}\}$ by means of the usual injections $i_{\Lambda',\Lambda}:\mathscr{A}_\Lambda\rightarrow\mathscr{A}_{\Lambda'}$, $\Lambda\subset\Lambda'$. The C^*-inductive limit has the form

$$\mathscr{A}=\mathscr{A}^P_a\otimes\mathscr{A}^P_b, \tag{2.3}$$

where $\otimes$ for C^*-algebras always denotes the so-called injective tensor product with minimal cross norm,[12,13] and $\mathscr{A}$ is an abstract, simple C^*-algebra with unit and trivial center. It is antiliminary[14] and thus quite different from an algebra $\mathscr{B}(\mathscr{H})$ consisting of all bounded operators on the Hilbert space $\mathscr{H}$ and characterizing traditional quantum mechanics.

For a single subsystem the local pairing Hamiltonian is

$$H_{\Lambda x}=\sum_{k\in\Lambda_x}2\epsilon_x b^*_{kx}b_{kx}-\frac{g_x}{|\Lambda_x|}\sum_{k,\,k'\in\Lambda_x}b^*_{kx}b_{k'x}, \tag{2.4}$$

where ϵ_x is the averaged kinetic energy and g_x is the positive coupling constant. The chemical potential μ_x is assumed fixed to its zero point value, and in order to have a phase transition we require $2\epsilon'_x/g_x<1$, where $\epsilon'_x:=\epsilon_x-\mu_x$. The local reduced Hamiltonian $H'_{\Lambda x}$ is obtained by replacing ϵ_x by ϵ'_x in $H_{\Lambda x}$.

In a weak link the two superconductors a and b are spacially separated by a normal conductor, an isolator, or the vacuum. Since the essential features of the composite system are independent of the specific barrier we only have to imitate the weakness of the interaction by our model. The usual single electron exchange term[2,15] first is excluded by our strict pair formulation and second gives no hint for the weakness assumption. We start, therefore, from the strong internal pair interaction in (2.4) and impose weakness in two different ways. On the one hand we restrict the number of modes partaking in the exchange to a finite one, also in the thermodynamic limit. This makes the damping factor $1/|\Lambda_x|$ of the coupling constant in (2.4) meaningless and we have a transition of normal pairs. Admitting also higher-order pair transfer we shall work with a general exchange polynomial in the normal pair operators with modes in the fixed set $(X_a,X_b)=X\in\mathscr{L}$. This will eventually give us the normal tunneling current. On the other hand one may increase the damping factor for the exchange of infinitely many modes to $(1/|\Lambda|)^{1+\epsilon}$, $\epsilon>0$. If $\epsilon>1$, then the limit of such an interaction tends to zero in the temperature representation. If $\epsilon \in (0, 1)$ the limit of the interaction is unbounded similarly to the strong coupling case. Thus, the value $\epsilon = 1$ is a distinguished one and we shall adhere to this choice. Since in this part of the interaction infinitely many modes are involved, the collective phenomenon may persist during the exchange, and will give us, in fact, the spectacular phenomenon of a tunneling super current. Introducing

$$m_{\Lambda x} := \sum_{k \in \Lambda_x} \frac{b_{kx}}{|\Lambda_x|}, \tag{2.5}$$

we can restate our assumption in the way that the interaction Hamiltonian h_Λ will be for all $\Lambda \in \mathscr{L}$ a Λ-independent polynomial in the quantities $m_{\Lambda x}^{(*)}$, $b_{kx}^{(*)}$, $k \in X_x$, $x \in \{a, b\}$ (and their Hermitian adjoints). The most important part of the collective tunneling is that of second order, and is uniquely determined by self-adjointness, total particle number conservation, and symmetry under a–b permutation to be of the form

$$h_\Lambda^s = g_s (m_{\Lambda a}^* m_{\Lambda b} + m_{\Lambda a} m_{\Lambda b}^*). \tag{2.6}$$

The local uncoupled and interacting Hamiltonians of the composite system are

$$H_\Lambda := H_{\Lambda a} + H_{\Lambda b}, \quad H'_\Lambda := H_\Lambda + h_\Lambda \tag{2.7}$$

and determine in their reduced form the uncoupled and interacting local grand-canonical equilibrium states. Due to the results on the single BCS model[5,6] we know that the uncoupled equilibrium states converge for $\Lambda_x \to \infty$ to a state on $\mathscr{A}$ with the following central decomposition:

$$\omega^\beta = \omega_a^\beta \otimes \omega_b^\beta = \int \omega^{\beta\vartheta} \, d\vartheta, \tag{2.8}$$

where $\vartheta := (\vartheta_a, \vartheta_b)$, $d\vartheta := d\vartheta_a \, d\vartheta_b/(2\pi)^2$, and the $d\vartheta$ integration tends always over $[0, 2\pi) \times [0, 2\pi)$. Here

$$\omega^{\beta\vartheta} = \omega^{\beta\vartheta_a} \otimes \omega^{\beta\vartheta_b} \tag{2.9}$$

and $\omega^{\beta\vartheta_x}$ is locally given on Λ_x by the density operator

$$\rho_{\Lambda x}^\vartheta = \exp(-\zeta_{\Lambda x} - \beta H_{\Lambda x}^{r\vartheta}) \in \mathscr{A}_{\Lambda x}^P, \tag{2.10}$$

with

$$H_{\Lambda x}^{r\vartheta} = 2\epsilon_x^r \sum_{k \in \Lambda_x} b_{kx}^* b_{kx} - g_x w_x \sum_{k \in \Lambda_x} (e^{-i\vartheta_x} b_{kx}^* + e^{i\vartheta_x} b_{kx}). \tag{2.11}$$

The positive constants $w_x = w_x(\beta)$ vanish for β smaller than the critical temperature β_{cx}. If $w_x(\beta) = 0$, the $d\vartheta_x$ integration may be deleted in (2.8).

The GNS triplet of ω^β is desintegrated into the GNS triplets of the pure phase states $\omega^{\beta\vartheta}$ according to Ref. 3, Chap. 4, as

$$(\pi_\beta, \mathscr{H}_\beta, \Omega_\beta) = (\pi_\beta^a \otimes \pi_\rho^b, \mathscr{H}_\beta^a \otimes \mathscr{H}_\beta^b, \Omega_\beta^a \otimes \Omega_\beta^b) = \int^\oplus (\pi_{\beta\vartheta}, \mathscr{H}_{\beta\vartheta}, \Omega_{\beta\vartheta}) d\vartheta. \tag{2.12}$$

In virtue of (2.9) we have also

$$\pi_{\beta\vartheta} = \pi_{\beta\vartheta_a} \otimes \pi_{\beta\vartheta_b} \tag{2.13}$$

and the analog for $\mathscr{H}_{\beta\vartheta}$ and $\Omega_{\beta\vartheta}$, where $(\pi_{\beta\vartheta_x}, \mathscr{H}_{\beta\vartheta_x}, \Omega_{\beta\vartheta_x})$ is the GNS triplet of $\omega^{\beta\vartheta_x}$ as a state on $\mathscr{A}_x^P$, $x \in \{a, b\}$.

We shall henceforth identify $\mathscr{A}$ with $\pi_\beta(\mathscr{A})$, selecting an arbitrary but fixed temperature $\beta \in (0, +\infty)$.

The weak closure $\mathscr{M}^\beta$ of $\mathscr{A}$ in $\mathscr{H}_\beta$ has the decomposition

$$\mathscr{M}^\beta = \mathscr{M}_a^\beta \bar{\otimes} \mathscr{M}_b^\beta = \int^\oplus \mathscr{M}^{\beta\vartheta} \, d\vartheta, \tag{2.14}$$

where

$$\mathscr{M}_x^\beta = \pi_\beta^x(\mathscr{A}_x^P)'', \quad x \in \{a, b\}, \tag{2.15a}$$

and

$$\mathscr{M}^{\beta\vartheta} := \pi_{\beta\vartheta}(\mathscr{A})'' = \pi_{\beta\vartheta_a}(\mathscr{A}_a^P)'' \bar{\otimes} \pi_{\beta\vartheta_b}(\mathscr{A}_b^P)'' =: \mathscr{M}^{\beta\vartheta_a} \bar{\otimes} \mathscr{M}^{\beta\vartheta_b}. \tag{2.15b}$$

Here the tensor product $\bar{\otimes}$ of von Neumann algebras has been used in (2.14) and (2.15) (cf. Ref. 13, Chap. IV), which arises from the product representations according to Proposition 4.13 of Ref. 13. For all $\Lambda \in \mathscr{L}$ we introduce

$$\mathscr{M}_\Lambda^\beta := \int^\oplus \pi_{\beta\vartheta}(\mathscr{A}_\Lambda) d\vartheta \subset \mathscr{M}^\beta. \tag{2.16}$$

The center $\mathscr{Z}^\beta$ of $\mathscr{M}^\beta$ is nontrivial for $\beta > \min\{\beta_{ca}, \beta_{cb}\}$ and is contained in $\mathscr{M}_\Lambda^\beta$ for all $\Lambda \in \mathscr{L}$. Denoting by tr_Λ the usual trace on the finite-dimensional (full matrix) algebra $\mathscr{A}_\Lambda$ we introduce on $\mathscr{M}_\Lambda^\beta$, for any $\Lambda \in \mathscr{L}$, the functional

$$\mathrm{tr}_\Lambda^\beta\{M\} := \int \mathrm{tr}_\Lambda\{A^\vartheta\} d\vartheta, \quad M \in \mathscr{M}_\Lambda^\beta, \tag{2.17}$$

where here and in the following the typical element M of $\mathscr{M}_\Lambda^\beta$ is associated with an essentially bounded, measurable function $\vartheta \to A^\vartheta \in \mathscr{A}_\Lambda$ by its integral decomposition

$$M = \int^\oplus \pi_{\beta\vartheta}(A^\vartheta) d\vartheta, \tag{2.18}$$

in an a.e. unique manner. Since $\mathrm{tr}_\Lambda^\beta$ is a normal, finite, faithful trace on $\mathscr{M}_\Lambda^\beta$, the latter is a finite W^*-algebra. Observe that the family $\{\mathrm{tr}_\Lambda^\beta; \Lambda \in \mathscr{L}\}$ does not define a trace on

$$\mathscr{M}_0^\beta := \bigcup_{\Lambda \in \mathscr{L}} \mathscr{M}_\Lambda^\beta. \tag{2.19}$$

It is only for small β that $\{\mathscr{M}_\Lambda^\beta; \Lambda \in \mathscr{L}\}$ constitutes a type-I funnel in the sense of Ref. 16. With

$$\rho_\Lambda^\vartheta := \rho_{\Lambda a}^\vartheta \otimes \rho_{\Lambda b}^\vartheta \in \mathscr{A}_\Lambda \tag{2.20}$$

and

$$\rho_\Lambda^\beta := \int^\oplus \pi_{\beta\vartheta}(\rho_\Lambda^\vartheta) d\vartheta \in \mathscr{M}_\Lambda^\beta, \tag{2.21}$$

we have for all $M \in \mathscr{M}_\Lambda^\beta$

$$\langle \omega^\beta; M \rangle := (\Omega_\beta, M\Omega_\beta) = \mathrm{tr}_\Lambda^\beta\{\rho_\Lambda^\beta M\}. \tag{2.22}$$

In the strong operator topology of $\mathscr{H}_\beta$, it holds that

$$\operatorname*{s-lim}_{\Lambda_x} m_{\Lambda x} = s_x = \int^\oplus w_x e^{-i\vartheta_x} \pi_{\beta\vartheta}(1) d\vartheta, \quad x \in \{a, b\}. \tag{2.23}$$

We have already given in Ref. 8 some arguments for the interpretation of s_x as the condensed pair field operator and shall confirm this point of view by the present investigation. Since on norm-bounded sets the operator product is bicontinuous with respect to the strong operator topology we have, according to our model assumptions in $\mathscr{H}_\beta$,

$$\text{s-}\lim_{\Lambda_x} h_\Lambda = h_\beta =: \int^{\oplus} \pi_{\beta\vartheta}(h_{\beta\vartheta})d\vartheta \in \mathscr{M}^\beta_X. \tag{2.24}$$

Proposition 2.1: The net of local grand canonical equilibrium states of the coupled model converges (after extension to states on $\mathscr{A}$ in the w-*-topology) to a state with the central decomposition

$$\omega'^\beta = \int \omega'^{\beta\vartheta}\, d\vartheta, \tag{2.25}$$

where the restriction of $\omega'^{\beta\vartheta}$ to $\mathscr{A}_\Lambda$, $\Lambda \supset X$, is given by the density operator

$$\rho'^\vartheta_\Lambda = \exp(-\zeta'_\Lambda - \beta(H^{r\vartheta}_{\Lambda a} + H^{r\vartheta}_{\Lambda b} + h_{\beta\vartheta})). \tag{2.26}$$

The GNS representation associated with ω'^β is unitary equivalent to that associated with ω^β.

Proof: (i) The existence of the limiting state ω'^β follows from the general investigation in Ref. 9. In Theorem 2.3 of this reference ω'^β is written in the form ($A \in \mathscr{A} \subset \mathscr{M}^\beta$)

$$\langle \omega'^\beta; A\rangle = \langle \omega^\beta; A\Gamma_{i\beta}\rangle / \langle \omega^\beta; \Gamma_{i\beta}\rangle, \tag{2.27}$$

where

$$\Gamma_{i\beta} := \sum_{n=0}^{\infty} (-1)^n \int_0^\beta ds_n \cdots \int_0^{s_2} ds_1 \times h_\beta(s_1) \cdots h_\beta(s_n) \in \mathscr{M}^\beta_X, \tag{2.28}$$

with

$$h_\beta(s) := \exp(-sH^{r\beta}_X) h_\beta \exp(sH^{r\beta}_X) \tag{2.29}$$

and [cf. (2.11)]

$$H^{r\beta}_\Lambda := \int^{\oplus} \pi_{\beta\vartheta}(H^{r\vartheta}_{\Lambda a} + H^{r\vartheta}_{\Lambda b})d\vartheta \in \mathscr{M}^\beta_\Lambda. \tag{2.30}$$

Using (2.22) if $A \in \mathscr{A}_\Lambda \subset \mathscr{M}^\beta_\Lambda$ and summing up the perturbation series gives

$$\langle \omega'^\beta; A\rangle = \mathrm{tr}^\beta_\Lambda\{\rho'^\beta_\Lambda A\}, \tag{2.31}$$

with

$$\rho'^\beta_\Lambda = \int^{\oplus} \pi_{\beta\vartheta}(\rho'^\vartheta_\Lambda)d\vartheta, \tag{2.32}$$

ρ'^ϑ_Λ being given by (2.26). Thus, (2.25) is obtained locally and, hence, on all of $\mathscr{A}$.

(ii) The explicit formula (2.26) shows that the $\omega'^{\beta\vartheta}$ are local perturbations of the $\omega^{\beta\vartheta}$ and thus are factorial and mutually disjoint. Therefore, (2.25) is the central decomposition of ω'^β.

(iii) Denoting $(\pi'_\beta, \mathscr{H}'_\beta, \Omega'_\beta)$ the GNS triple of ω'^β we introduce

$$U: \mathscr{H}'_\beta \to \mathscr{H}_\beta,$$

by

$$U\pi'_\beta(A)\Omega'_\beta := \pi_\beta(A)\Gamma^{1/2}_{i\beta}\Omega_\beta / \langle \omega^\beta; \Gamma_{i\beta}\rangle^{1/2}, \tag{2.33}$$

for all $A \in \mathscr{A}$. This is a well-defined mapping on a dense domain since Ω'_β is cyclic and separating for $\pi'_\beta(\mathscr{A})$ in virtue of the approximating KMS states (cf., e.g., Ref. 17) and π'_β is faithful. From (i) and from Ref. 3 II, p. 153, it follows that U is an isometry. If $\Phi \in \mathscr{H}_\beta$ has the form of the right-hand side (rhs) in (2.33) then $A \in \mathscr{A}$ is unique, since $\Gamma^{1/2}_{i\beta}$ is invertible, Ω_β separating, and π_β faithful. Thus the inverse mapping U^{-1} exists on these vectors. If $\Psi \in \mathscr{H}_\beta$, then there is a sequence $\{A_n;\ n \in \mathbb{N}\} \subset \mathscr{A}$ with uniform norm-bound such that

$$\Psi = \lim_n A_n \Omega_\beta = \lim_n A_n \Gamma^{-1/2}_{i\beta}\Gamma^{1/2}_{i\beta}\Omega_\beta.$$

Approximating $\Gamma^{-1/2}_{i\beta}$ by a strongly converging net $\{B_\alpha;\ \alpha \in I\} \subset \mathscr{A}$, Ψ may be approximated by vectors $C_{\alpha'}\Gamma^{1/2}_{i\beta}\Omega_\beta$, $C_{\alpha'} \in \mathscr{A}$, $\alpha' \in I'$. Thus U^{-1} is densely defined, U is unitary, and $\Omega'_\beta = U^{-1}\Gamma^{1/2}_{i\beta}\Omega_\beta / \langle \omega^\beta; \Gamma_{i\beta}\rangle^{1/2}$. Using this in (2.33) gives the unitary equivalence of π'_β and π_β. □

Observe that for the special case $h_\Lambda = h^s_\Lambda$ [cf. (2.6)]

$$h_\beta = g_s(s^*_a s_b + s_a s^*_b) = 2g_s w_a w_b \int^{\oplus} \pi_{\beta\vartheta}(\cos(\vartheta_a - \vartheta_b))d\vartheta. \tag{2.34}$$

Equation (2.25) specializes then to

$$\omega'^\beta = \int \omega^{\beta\vartheta}\, d\mu^\beta(\vartheta), \tag{2.35}$$

with

$$d\mu^\beta(\vartheta) = C_\beta \exp(-2\beta w_a w_b \cos(\vartheta_a - \vartheta_b))d\vartheta, \quad C_\beta \in \mathbb{R}_+. \tag{2.36}$$

In this case the interacting system has the same pure phase states $\omega^{\beta\vartheta}$ as the uncoupled model, the interaction potential showing up only in the statistical distribution of the macroscopic phase difference. This situation escapes the treatment of usual many-body physics where only pure phase quantities are considered. Also interesting is the nonequipartition of the pure phase. (For a more detailed discussion of these aspects, cf. Ref. 18.) For finite β, Eq. (2.36) shows that $d\mu^\beta$ and $d\vartheta$ are equivalent as measures on $[0, 2\pi) \times [0, 2\pi)$. In the low-temperature limit $\beta \to +\infty$, however, we have in the vague topology

$$d\mu^\beta(\vartheta) \to \delta(\vartheta_a - \vartheta_b)d\vartheta,$$

which gives a measure singular to $d\vartheta$, and the corresponding two GNS-groundstate representations are no more equivalent. Physically this means that also the weak coupling of the considered type is sufficient to equalize the two macroscopic phases at zero (absolute) temperature, where no temperature fluctuations disturb the mutual dynamic influences of the subsystems. Thus the careful evaluation of the limiting equilibrium states shows that at $T = 0$ the total system behaves like a single superconductor and that, in contradistinction to the common point of view and to the conclusion of Ref. 4, the prerequisite for the Josephson effect is lost. The deviation of Ref. 4 from our result is due to an inappropriate *ad hoc* choice of the representation Hilbert space.

Being interested in the finite-temperature physics we use $(\pi_\beta, \mathscr{H}_\beta, \Omega_\beta)$ also for the interacting system and benefit from the better invariance properties of Ω_β. The modular quantities for the standard triple $(\mathscr{M}^\beta, \mathscr{H}_\beta, \Omega_\beta)$ derive from the closure S of the antilinear mapping $M\Omega_\beta \to M^*\Omega_\beta$, $M \in \mathscr{M}^\beta$, with the polar decomposition $S = J\Delta^{1/2}$ (cf., e.g., Ref. 3 and 19). Here J is an antiunitary operator in $\mathscr{H}_\beta$ with $J^2 = 1$ and $\Delta^{1/2}$ is a positive, self-adjoint operator. The modular automorphism group is then given by $\sigma^\beta_t(M) := \Delta^{it}M\Delta^{-it}$, and the canonical antilinear isomorphism

$$j: \mathscr{M}^{\beta} \to \mathscr{M}^{\beta'} \tag{2.37}$$

by

$$j(M) = JMJ, \quad M \in \mathscr{M}^{\beta}. \tag{2.38}$$

Defining

$$\mathscr{H}^{\beta}_{\Lambda} := \overline{\mathscr{M}^{\beta}_{\Lambda}\Omega_{\beta}}, \quad \Lambda \in \mathscr{L}, \tag{2.39}$$

we have the following statement.

Proposition 2.2: For all $\Lambda \in \mathscr{L}$ and all $t \in \mathbb{R}$ it holds that

$$\sigma^{\beta}_{t}(\mathscr{M}^{\beta}_{\Lambda}) = \mathscr{M}^{\beta}_{\Lambda}. \tag{2.40}$$

The operators S, Δ, and J are reduced by $\mathscr{H}^{\beta}_{\Lambda}$. The restrictions S_{Λ}, Δ_{Λ}, and J_{Λ} are bounded and are the modular quantities of the standard triple $(\mathscr{M}^{\beta}_{\Lambda}, \mathscr{H}^{\beta}_{\Lambda}, \Omega_{\beta})$. For $M \in \mathscr{M}^{\beta}_{\Lambda}$ we have

$$\Delta^{1/2}_{\Lambda} M\Omega_{\beta} = \rho^{\beta\,1/2}_{\Lambda} M \rho^{\beta\,-1/2}_{\Lambda}\Omega_{\beta}, \tag{2.41}$$

$$J_{\Lambda} M\Omega_{\beta} = \rho^{\beta\,1/2}_{\Lambda} M^{*} \rho^{\beta\,-1/2}_{\Lambda}\Omega_{\beta}. \tag{2.42}$$

Proof: (i) From the product form of the $\rho^{\vartheta}_{\Lambda}$ one derives for $M \in \mathscr{M}^{\beta}_{K}$, $K \subset \Lambda$,

$$\rho^{\beta}_{\Lambda} M \rho^{\beta\,-1}_{\Lambda} = \rho^{\beta}_{K} M \rho^{\beta\,-1}_{K}. \tag{2.43}$$

Thus,

$$\hat{\sigma}^{\beta}_{t}(M) := \rho^{\beta it}_{\Lambda} M \rho^{\beta\,-it}_{\Lambda}, \quad M \in \mathscr{M}^{\beta}_{\Lambda}, \tag{2.44}$$

constitutes for t varying in $\mathbb{R}$ not only an $*$-automorphism group in $\mathscr{M}^{\beta}_{\Lambda}$ but also in $\mathscr{M}^{\beta}_{0}$, which is σ-weakly continuous in t. By means of (2.22) one finds that

$$\langle \omega^{\beta}, C^{*}_{1}\hat{\sigma}^{\beta}_{t}(M)C_{2}\rangle = \langle \omega^{\beta}; \hat{\sigma}^{\beta}_{-t}(C^{*}_{1})M\hat{\sigma}^{\beta}_{-t}(C_{2})\rangle,$$

for all $C_{1/2} \in \mathscr{M}^{\beta}_{0}$. Thus $\hat{\sigma}^{\beta}_{t}$ has a norm-densely defined preadjoint acting in $\mathscr{M}^{\beta}_{*}$ and is σ-weakly closable by Ref. 3, Lemma 3.1.9. The closures constitute a W^{*}-automorphism group of $\mathscr{M}^{\beta}$, which satisfies the KMS condition with respect to ω^{β} for the natural temperature unity. From the uniqueness of the modular automorphism group we find $\hat{\sigma}^{\beta}_{t} = \sigma^{\beta}_{t}$ for all $t \in \mathbb{R}$, showing the latter to leave $\mathscr{M}^{\beta}_{\Lambda}$ invariant.

(ii) With $\lambda_{+(-)} > 0$ the supremum (infimum) of the spectrum of the positive, invertible operator ρ^{β}_{Λ}, we obtain for $M \in \mathscr{M}^{\beta}_{\Lambda}$

$$\begin{aligned}
&\|SM\Omega_{\beta}\|^{2} \\
&\quad = \operatorname{tr}^{\beta}_{\Lambda}\{\rho^{\beta}_{\Lambda} MM^{*}\} \leqslant \lambda_{+}\operatorname{tr}^{\beta}_{\Lambda}\{MM^{*}\} \\
&\quad = (\lambda_{+}/\lambda_{-})\lambda_{-}\operatorname{tr}^{\beta}_{\Lambda}\{M^{*}M\} \\
&\quad \leqslant (\lambda_{+}/\lambda_{-})\operatorname{tr}^{\beta}_{\Lambda}\{\rho^{\beta}_{\Lambda} M^{*}M\} \leqslant (\lambda_{+}/\lambda_{-})\|M\Omega_{\beta}\|^{2},
\end{aligned}$$

which gives the uniform boundedness of S_{Λ} on $\mathscr{M}^{\beta}_{\Lambda}\Omega_{\beta}$. By the closedness of S one observes that $\mathscr{H}^{\beta}_{\Lambda} \subset D(S)$. The restriction of S to $\mathscr{H}^{\beta}_{\Lambda}$ gives then rise to the modular quantities of the standard triple $(\mathscr{M}^{\beta}_{\Lambda}, \mathscr{H}^{\beta}_{\Lambda}, \Omega_{\beta})$. Since J is bounded, $\Delta^{1/2}_{\Lambda} = J_{\Lambda}S_{\Lambda}$ and $\Delta^{-1/2}_{\Lambda} = S_{\Lambda}J_{\Lambda}$ both are bounded, too.

(iii) For $M \in \mathscr{M}^{\beta}_{\Lambda}$ we calculate

$$\begin{aligned}
\Delta^{1/2}_{\Lambda} M\Omega_{\beta} &= \Delta^{1/2}_{\Lambda} M \Delta^{-1/2}_{\Lambda}\Omega_{\beta} = \sigma^{\beta}_{-i/2}(M)\Omega_{\beta} \\
&\overset{\text{(i)}}{=} \rho^{\beta\,1/2}_{\Lambda} M \rho^{\beta\,-1/2}_{\Lambda}\Omega_{\beta}
\end{aligned}$$

and

$$\begin{aligned}
JM\Omega_{\beta} &= J(J\Delta^{1/2}M^{*})\Omega_{\beta} \\
&= \rho^{\beta\,1/2}_{\Lambda} M^{*} \rho^{\beta\,-1/2}_{\Lambda}\Omega_{\beta}.
\end{aligned}$$

Because of boundedness all expressions are well defined. □

III. INNER LIMITING DYNAMICS AND PARTICLE STRUCTURE

The particle structure of a nonrelativistic (reconstructed) quantum mechanics is determined by the unitary representation of the gauge group, its generator counting the particlelike excitations, and by the (renormalized) limiting Hamiltonian, its diagonalization specifying the quasiparticles.

Beginning with the gauge transformations let us take the gauge angles $\Theta = (\Theta_a, \Theta_b)$ from $\mathbb{R}^2$, which sometimes is understood as $(\mathbb{R}/[0, 2\pi))^2$. Referring to the single electron formalism we define the local particle number operators by

$$N_{\Lambda x} := \sum_{k \in \Lambda_x} 2b^{*}_{kx} b_{kx}, \quad x \in \{a, b\}, \tag{3.1}$$

and the local gauge automorphisms by

$$\alpha_{\Theta}(A) := \exp\Big(i \sum_{x} \Theta_x N_{\Lambda x}\Big) A \exp\Big(-i \sum_{x} \Theta_x N_{\Lambda x}\Big), \tag{3.2}$$

if $A \in \mathscr{A}_{\Lambda}$. Being compatible with the injections of $\mathscr{A}_{\Lambda}$ into $\mathscr{A}_{\Lambda'}$, $\Lambda \subset \Lambda'$, Eq. (3.2) gives rise to a C^{*}-dynamical system in $\mathscr{A}$, if Θ varies in $\mathbb{R}^2$.

In order to have a particle interpretation for the coupled model one has to require that h_{Λ} be invariant under α_{Θ}, if $\Theta_a = \Theta_b$; the possibility of particle transfer demands that h_{Λ} not be gauge invariant, if $\Theta_a \neq \Theta_b$. For the unitary implementation one has to use, therefore, the uncoupled equilibrium state and to define $W^{\beta}_{\Theta} A\Omega_{\beta} := \alpha_{\Theta}(A)\Omega_{\beta}$, $A \in \mathscr{A}$. The extension of this transformation to a unitary operator, denoted by the same symbol, reflects special properties of the representation and implements the extended gauge automorphisms $\alpha^{\beta}_{\Theta} \in \operatorname{Aut}(\mathscr{M}^{\beta})$.

Let us denote by $\mathscr{M}^{\beta}_{P} \subset \mathscr{M}^{\beta}_{0}$ the $*$-algebra of all polynomials in $s^{(*)}_{x}$, $b^{(*)}_{kx}$, k taken from a finite set Λ_x, $x \in \{a, b\}$ (and in the $*$-conjugate quantities). For $M \in \mathscr{M}^{\beta}_{P} \cap \mathscr{M}^{\beta}_{\Lambda}$ the generators N^{β}_{x}, $x \in \{a, b\}$, of $\{W^{\beta}_{\Theta}; \Theta \in \mathbb{R}^2\}$ are shown[7] to act as

$$N^{\beta}_{x} M\Omega_{\beta} = [N_{\Lambda x}, M]\Omega_{\beta} + \delta^{x}_{T}(M)\Omega_{\beta}. \tag{3.3}$$

Here δ^{x}_{T} designates the antisymmetric $*$-derivation

$$\delta^{x}_{T}: \mathscr{M}^{\beta}_{P} \to \mathscr{M}^{\beta}, \tag{3.4a}$$

given by

$$\delta^{x}_{T}(M) := \int^{\oplus} \pi_{\beta\vartheta}\Big(\frac{2}{i}\frac{d}{d\vartheta_x} A^{\vartheta}\Big) d\vartheta. \tag{3.4b}$$

The explicit form of s_x (2.23) shows that δ^{x}_{T} leaves $\mathscr{M}^{\beta}_{P}$ in fact invariant. Restricted to $\mathscr{M}^{\beta}_{P} \cap \mathscr{M}^{\beta}_{\Lambda}$ the operator δ^{x}_{T} is closable with the closure domain consisting of all $M \in \mathscr{M}^{\beta}_{\Lambda}$ with absolutely continuous functions $\vartheta \to A^{\vartheta} \in \mathscr{A}_{\Lambda}$ having essentially bounded derivatives. On $\mathscr{M}^{\beta}_{P}$ the $*$-derivation δ^{x}_{T} is, however, not closable,[7] and the corresponding classical part of N^{β}_{x} is not essentially self-adjoint on $\{M\Omega_{\beta};\ M \in \mathscr{M}^{\beta}_{P}\}$. From (3.3) and (3.4) one derives that the condensed pair states $s_x^n \Omega_\beta$, $n \in \mathbb{Z}$, are counted with the number $2n$.

Since elements in $\mathscr{M}^\beta$ act as left multiplication operators on vectors $M\Omega_\beta$, $M \in \mathscr{M}^\beta$, it is clear that in (3.3) not even the normal part, and, even more, not the classical part, is affiliated with $\mathscr{M}^\beta$. Nevertheless, N_a^β and N_b^β are physically the most important observables since their time dependence determines the current between a and b.

We face here the problem of how to extend the dynamic theory beyond the von Neumann algebra $\mathscr{M}^\beta$, in a way that still admits a kind of local approximation. In our opinion, the latter requirement is indispensable since the quantum mechanics of infinite systems obtain its meaning and justification by the approximating traditional quantum mechanical descriptions. The kind of approximation that seems suitable for external symmetry generators follows from the observation that, for $A \in \mathscr{A}_\Lambda$,

$$N_x^\beta A\Omega_\beta = (N_{\Lambda x} - j(N_{\Lambda x}))A\Omega_\beta. \tag{3.5}$$

In fact, Eq. (3.3) leads to

$$N_x^\beta A\Omega_\beta = N_{\Lambda x} A\Omega_\beta - AN_{\Lambda x}\Omega_\beta.$$

In order to treat the last term, we choose an arbitrary $B \in \cup_\Lambda \mathscr{A}_\Lambda$. Without any restriction in generality we may assume that A and B are both in $\mathscr{A}_\Lambda$. We calculate

$$\begin{aligned}(B\Omega_\beta, \delta_T^x(\rho_\Lambda^{\beta\,-1/2})\Omega_\beta) &\overset{(2.22)}{=} \operatorname{tr}_\Lambda^\beta\{\rho_\Lambda^\beta B^*\delta_T^x(\rho_\Lambda^{\beta\,-1/2})\} \\ &\overset{(2.17)(3.4b)}{=} \frac{2}{i}\int \frac{\partial}{\partial\vartheta_x}\operatorname{tr}_\Lambda\{B^*\rho_\Lambda^{\vartheta\,1/2}\}d\vartheta = 0.\end{aligned}$$

Since $\cup_\Lambda \mathscr{A}_\Lambda \Omega_\beta$ is dense in $\mathscr{H}_\beta$ we conclude

$$\delta_T^x(\rho_\Lambda^{\beta\,-1/2})\Omega_\beta = 0.$$

From $\alpha_{\theta x}^\beta(\rho_\Lambda^\beta) = \rho_\Lambda^\beta$ we get

$$N_x^\beta \rho_\Lambda^{\beta\,-1/2}\Omega_\beta = 0 \overset{(3.3)}{=} [N_{\Lambda x}, \rho_\Lambda^{\beta\,-1/2}]\Omega_\beta.$$

This gives

$$A\rho_\Lambda^{\beta\,1/2} N_{\Lambda x}\rho_\Lambda^{\beta\,-1/2}\Omega_\beta = AN_{\Lambda x}\Omega_\beta.$$

Observing Proposition 2.2 we arrive at

$$AN_{\Lambda x}\Omega_\beta = j(N_{\Lambda x})A\Omega_\beta,$$

which leads to (3.5). Being invariant under the W_θ^β, $\Theta \in \mathbb{R}^2$, $\cup_\Lambda \mathscr{A}_\Lambda \Omega_\beta$ is a core for N_x^β and we may verify condition IV of Ref. 20, Theorem 3.17. This implies

$$N_x^\beta = \text{s-resolvent-lim}\, N_{\Lambda x} - j(N_{\Lambda x}), \tag{3.6}$$

demonstrating that the macroscopic features are not locally anticipated. From the mentioned theorem in 20 we obtain also

$$W_\Theta^\beta = \text{s} - \lim_\Lambda \exp\Big[i\sum_x \Theta_x (N_{\Lambda x} - j(N_{\Lambda x}))\Big], \tag{3.7}$$

uniformly in Θ, where we have used the invariance of Ω_β to identify the gauge-implementing unitaries.

The discussion of the limiting dynamics starts with the convergence relation, following easily from (2.23) and (2.24) for $A \in \mathscr{A}_\Lambda$,

$$\begin{aligned}\text{s} - \lim_{\Lambda'} L'_{\Lambda'}(A) :&= \text{s} - \lim_{\lambda'}[H'_{\Lambda'}, A] \\ &= [H_\Lambda'^\beta, A] =: L_0'^\beta(A),\end{aligned} \tag{3.8}$$

where

$$H_\Lambda'^\beta = H_\Lambda^\beta + h_\beta \tag{3.9}$$

and

$$H_\Lambda^\beta = \sum_x \sum_{k \in \Lambda_x} (2\epsilon_x b_{kx}^* b_{kx} - g_x(s_x b_{kx}^* + s_x^* b_{kx})). \tag{3.10}$$

Observe that the pure phase components of H_Λ^β coincide with (2.11) if one replaces there ϵ_x' by ϵ_x. The analog of (3.7) is valid in the GNS representations of many states defining an antisymmetric *-derivation, which in the present case is

$$L_0'^\beta: \ \cup_\Lambda \mathscr{A}_\Lambda \to \mathscr{M}^\beta. \tag{3.11}$$

But only under very restrictive conditions is such a *-derivation closable to a generator of a W^*-automorphism group. That these conditions are fulfilled in the representations of general equilibrium states is demonstrated in Ref. 21. From this work or from the results on the physical dynamics of the single BCS model[7] it follows for the uncoupled composed system, the GNS representation of which we are using, that the commutator with H_Λ^β is closable to a generator L^β of a W^*-automorphism group $\{\tau_t^\beta; t \in \mathbb{R}\}$. In order to formulate detailed results for the coupled dynamics let us introduce the notion that a net $\{P_\Lambda; \Lambda \in \mathscr{L}\} \subset \cup_\Lambda \mathscr{A}_\Lambda$ is a "density approximation for $P \in \mathscr{M}_P^\beta$," if the P_Λ are the same polynomials as P but the $s_x^{(*)}$ arguments are replaced there, respectively, by the local density observables $m_{\Lambda x}^{(*)}$.

Theorem 3.1: (i) $L_0'^\beta$ of (3.8) is σ-weakly closable to a σ-weak generator L'^β of a W^*-automorphism group $\{\tau_t'^\beta; t \in \mathbb{R}\}$ in $\mathscr{M}^\beta$. If M is in the domain $\mathscr{D}(L'^\beta)$ and if $\{A_\alpha \in \mathscr{A}_{\Lambda\alpha}; \alpha \in I\}$ is an arbitrary net with $\|A_\alpha\| < \|M\|$ and s-$\lim_\alpha A_\alpha = M$, then

$$\begin{aligned}\sigma - \text{w-lim}_\alpha L'_{\Lambda\alpha}(A_\alpha) &= \sigma - \text{w-lim}_\alpha L'^\beta(A_\alpha) \\ &= L'^\beta(M).\end{aligned} \tag{3.12}$$

(ii) $\mathscr{M}_p^\beta$ is an invariant core for L'^β, and for $P \in \mathscr{M}_p^\beta \cap \mathscr{M}_\Lambda^\beta$ one has

$$L'^\beta(P) = [H_\Lambda'^\beta, P] + \mu_a \delta_T^a(P) + \mu_b \delta_T^b(P). \tag{3.13}$$

If $\{P_\Lambda; \Lambda \in \mathscr{L}\}$ is a density approximation of $P \in \mathscr{M}_P^\beta$ then

$$\text{s-}\lim_\Lambda L_\Lambda'^n(P_\Lambda) = \text{s-}\lim_\Lambda L'^{\beta n}(P_\Lambda) = L'^{\beta n}(P), \tag{3.14}$$

for all $n \in \mathbb{N}$. Moreover, it holds

$$\text{s-}\lim_\Lambda \tau_t'^\Lambda(P_\Lambda) = \tau_t'^\beta(P), \tag{3.15}$$

for all $t \in \mathbb{R}$.

Proof: (i) From (3.8) follows that the closure L'^β of $L_0'^\beta$ exists and that

$$\mathscr{D}(L'^\beta) = \mathscr{D}(L^\beta), \quad L'^\beta = L^\beta + [h_\beta, \ldots]. \tag{3.16}$$

By a general semigroup argument (cf., e.g., Ref. 3, Theorem 3.1.33), L'^β is the σ-weak generator of the σ-weakly continuous group of transformations in $\mathscr{M}^\beta$,

$$\tau_t'^{\beta}(M) := \sum_{n=0}^{\infty} i^n \int_0^t dt_n \cdots \int_0^{t_2} dt_1 \times \mathrm{Ad}\, h_\beta(t_1)(\cdots \mathrm{Ad}\, h_\beta(t_n)(\tau_t^\beta(M))\cdots), \tag{3.17}$$

where here $M \in \mathscr{M}^\beta$, $h_\beta(t) := \tau_t^\beta(h_\beta)$, and $\mathrm{Ad}\, h(M) := [h,M]$. The series in (3.17) converges in norm and the time integrals in the σ-weak topology. From this explicit expression follows that the $\tau_t'^\beta$ are W^*-automorphisms of $\mathscr{M}^\beta$. Relation (3.12) is a consequence of the corresponding relation for L^β on $\mathscr{D}(L^\beta) = \mathscr{D}(L'^\beta)$ derived in Ref. 7.

(ii) From (3.16) and formula (3.4) of Ref. 7 one obtains (3.13) if one has verified that the right-hand side is defined. The application of δ_T^x to a $P \in \mathscr{M}_P^\beta$ leads in fact to a polynomial of the same kind. Since also $H_\Lambda'^\beta \in \mathscr{M}_P^\beta$, for all $\Lambda \in \mathscr{L}$, the right-hand side of (3.13) is a well-defined element of $\mathscr{M}_P^\beta$. Explicit evaluation of $[H_\Lambda', P_\Lambda]$ gives again a polynomial in the density observables $m_{\Lambda x}^{(*)}$ and in those $b_{kx}^{(*)}$, which appear in P_Λ. Thus s-$\lim_\Lambda L_\Lambda'(P_\Lambda)$ exists and must then coincide with the σ-weak limit $L'^\beta(P)$ following from (3.12). This leads to (3.14) for $n = 1$ and by iteration for general n.

In order to obtain (3.15) we use Appendix B of Ref. 9 where, for the uncoupled dynamics,

$$\|L_\Lambda^n(P_\Lambda)\| \leqslant MC^n n!, \quad \forall \Lambda \in \mathscr{L},$$

and from this the strong operator convergence $\Pi_1^n \tau_{t_\nu}^\Lambda(P_\Delta) \to \Pi_1^n \tau_{t_\nu}^\beta(P)$ has been derived, for all $n \in \mathbb{N}$ and uniformly for the region $|t_\nu| < 1/C$, $1 \leqslant \nu \leqslant n$. (The uniform convergence in all finite t-intervals will be provided in Ref. 21.)

The perturbation series for $\tau_t'^\Lambda(P_\Lambda)$ looks like (3.17), τ_t^β being replaced by τ_t^Λ and $h_\beta(t)$ by $h_\Lambda(t) := \tau_t^\Lambda(h_\Lambda)$. By the norm convergence of the series, uniform in Λ, the Λ-limit may be interchanged with the summation. The time integrals are taken locally over matrices which have a Λ-independent norm bound. Since for arbitrary $t \in \mathbb{R}$ the iterated time integration is finite, we may interchange it with the Λ limit according to the dominated convergence theorem and obtain (3.15). □

Remark: it seems not possible to modify (3.8) to hold in some norm topology (e.g., in the predual $\mathscr{M}_*^\beta$ or in $\mathscr{H}_\beta$), which would permit the application of general semigroup arguments and lead directly to the convergence of the finite time translations. Relation (3.12), which is crucial for obtaining (3.14) and (3.15), has been generally derived[23] from the boundedness of the vector norm $\|(H_\Lambda' - \Delta_\Lambda H_\Lambda')\Omega_\beta\|$ uniformly in Λ. (In the Appendix of Ref. 23, the Hamiltonians used cover the uncoupled case, but the addition of the bounded perturbation causes no difficulty.) If the mentioned norms would tend to zero for $\Lambda \to \infty$, the preadjoints $L_{\Lambda *}'$ (acting in $\mathscr{M}_*^\beta$) would converge in the strong resolvent sense. If the closely related relation $\lim_\Lambda \|(H_\Lambda - j(H_\Lambda))\Omega_\beta\| = 0$ would be valid [cf. (2.41) and (2.42)], then one would obtain the strong resolvent convergence of $\{H_\Lambda + h_\Lambda - j(H_\Lambda);\ \Lambda \in \mathscr{L}\}$ to the renormalized Hamiltonian K'^β, cf. below, in close parallel to (3.6). The investigations in Ref. 23 indicate that neither the one nor the other improvement will be possible, in sharp contrast to the case of the renormalized Hamiltonians in the ground state representations.

In order to find the eigenexcitations of the system we construct the representation-dependent, effective Hamiltonian. For the uncoupled case, ω^β is invariant under $\tau_t^{\beta*}$ and we may employ the usual procedure,

$$K^\beta A \Omega_\beta := L^\beta(A)\Omega_\beta, \quad A \in \cup_\Lambda \mathscr{A}_\Lambda, \tag{3.18}$$

to introduce by self-adjoint extension the generator (also denoted by K^β) of a group of unitaries, which implement τ_t^β. As for the gauge transformations we obtain in this way just the standard implementation of τ_t^β, which belongs to the triple $(\mathscr{M}^\beta, \mathscr{H}_\beta, \Omega_\beta)$ and which leaves invariant the self-dual, pointed, closed, and convex generating cone $\{Mj(M)\Omega_\beta; M \in \mathscr{M}^\beta\}$ (cf. e.g., Ref. 3, Chap. 2.5.4).

In the coupled case ω'^β is not invariant under $\tau_t'^{\beta*}$. (It is invariant under the reduced, interacting dynamics.) Since in general neither ω^β is invariant under $\tau_t'^{\beta*}$ we have no natural standard implementation of $\tau_t'^\beta$. For the subsequent discussions

$$\tau_t'^\beta(M) = \exp(it(K^\beta + h_\beta))M \exp(-it(K^\beta + h_\beta)), \tag{3.19}$$

which follows from the perturbation series (3.17), seems to be a suitable unitary implementation. Thus we declare

$$K'^\beta := K^\beta + h_\beta \tag{3.20}$$

as the effective Hamiltonian for the interacting dynamics. For $P \in \mathscr{M}_P^\beta \cap \mathscr{M}_\Lambda^\beta$ we obtain

$$K'^\beta P \Omega_\beta = [H_\Lambda^{r\beta}, P]\Omega_\beta + (\mu_a N_a^\beta + \mu_b N_b^\beta + h_\beta)P\Omega_\beta, \tag{3.21}$$

where $H_\Lambda^{r\beta}$ is (3.10) with ϵ_x replaced by ϵ_x^r, $x \in \{a, b\}$, and where (3.13) and (3.3) have been used.

In order to diagonalize K^β we follow the method of Ref. 8 for the single superconductor and introduce for every $x \in \{a, b\}$, and $\vartheta_x \in [0, 2\pi)$ a Bogoliubov–Valatin transformation as a C^*-automorphism of $\mathscr{A}_x^P$, which transforms b_{kx} into the quasiparticle operators $q_{kx}^{\vartheta x} \in \mathscr{A}_x^P$. The decisive step is to introduce also the quasiparticle operators in $\mathscr{M}^\beta$:

$$q_{ka} := \int^\oplus \pi_{\beta\vartheta}(q_{ka}^{\vartheta a} \otimes 1_b)\, d\vartheta \tag{3.22}$$

and q_{kb} (defined analogously). With the constants in the Bogoliubov–Valatin transformation chosen in a suitable manner the various quasi-particle operators diagonalize $H_{\Lambda x}^{r\vartheta}$, $\rho_{\Lambda x}^\vartheta$ as well as $H_\Lambda^{r\beta}$, ρ_Λ^β, respectively, with always the same eigenvalues. This gives also for the composed system

$$(\Omega_{\beta\vartheta}, \pi_{\beta\vartheta}(P(q_{kx}^{\vartheta x}))\Omega_{\beta\vartheta}) = (\Omega_\beta, P(q_{kx})\Omega_\beta), \tag{3.23}$$

where $P(q_{kx}^{\vartheta x})$ is to symbolize a polynomial in $q_{kx}^{\vartheta x}$ and the $*$-conjugate quantities, where k and x vary. Thus, the prescription

$$q_{kx}^{\vartheta x} \to q_{kx}, \quad x \in \{a,b\}, \quad k \in \mathscr{K}_x, \tag{3.24}$$

is extensible to a W^*-isomorphism

$$\mathscr{M}^{\beta\vartheta} \to \mathscr{M}_q^\beta \tag{3.25}$$

by unitary implementation, where $\mathscr{M}_q^\beta$ denotes the von Neumann algebra in $\mathscr{H}_\beta$, which is generated by the q_{kx}, $x \in \{a, b\}$, $k \in \mathscr{K}_x$. Denoting

$$\mathscr{H}^s_\beta := \overline{\mathscr{Z}^\beta}\Omega_\beta, \quad \mathscr{H}^q_\beta := \overline{\mathscr{M}^\beta_q}\Omega_\beta, \tag{3.26}$$

one finds

$$\mathscr{H}_\beta = \mathscr{H}^s_\beta \otimes \mathscr{H}^q_\beta, \tag{3.27}$$

where we have used (3.23) and the integral decomposition (2.12) to derive a natural unitary equivalence between $\mathscr{H}_\beta$ and $\mathscr{H}^s_\beta \otimes \mathscr{H}^q_\beta$, which gives rise to the identification.

By means of an analogous identification we infer from (2.14)

$$\mathscr{M}^\beta = \mathscr{Z}^\beta \bar{\otimes} \mathscr{M}^\beta_q. \tag{3.28}$$

The combination of (3.20) and (3.21) makes explicit the form of K^β in terms of H'^β_Λ and demonstrates that the condensed and quasiparticle excitations $\Pi_x s_x^{n_x} \Pi_k q^*_{kx} \Pi_e q_{ex} \Omega_\beta$ are eigenstates of this operator. By means of the inverse Bogoliubov–Valatin transformation, h_β may be expressed as a polynomial in the $s_x^{(*)}$, $q_{kx}^{(*)}$. If we make the assumption that during the tunneling the status of a pair to be condensed or normal is not altered we are led to the decomposition

$$h_\beta = h^s_\beta + h^q_\beta, \tag{3.29}$$

with $h^s_\beta \in \mathscr{Z}^\beta$ and $h^q_\beta \in \mathscr{M}^\beta_q$. This assumption is strongly supported by the fact that the averaging process (2.23) performed with quasipair operators gives zero: a direct condensation of quasiparticles is not possible.

Theorem 3.2: Under the assumption (3.29) the coupled dynamics decomposes according to

$$\tau'^\beta_t = \tau^{\beta s}_t \otimes \tau'^{\beta q}_t, \tag{3.30}$$

where

$$\tau^{\beta s}_t = \alpha^\beta_{(\mu_a t, \mu_b t)} \mid_{\mathscr{Z}^\beta} \tag{3.31}$$

and $\tau'^{\beta q}_t \in \mathrm{Aut}\, \mathscr{M}^\beta_q$. The spectrum of the reduced form of $\tau'^{\beta q}_t$ contains always the reduced quasiparticle spectrum.

Proof: (i) According to Ref. 8, Theorem 5, the uncoupled Hamiltonian, applied on vectors in $\mathscr{M}^\beta_P \Omega_\beta \subset \mathscr{H}^s_\beta \otimes \mathscr{H}^q_\beta$, has the form

$$K^\beta = K^\beta_s + K^\beta_q$$
$$= \sum_x \mu_x \delta^x_T + \sum_x 2(E_x + \mu_x)\Big[\sum_k q^*_{kx} q_{kx}, \ldots\Big], \tag{3.32}$$

where $E_x = (\epsilon'^2_x + g^2_x \omega^2_x)^{1/2}$. By virtue of (3.20) and (3.29) we have on the same domain

$$K'^\beta = K^\beta_s + h^s_\beta + K^\beta_q + h^q_\beta. \tag{3.33}$$

The problem of forming the self-adjoint extensions separately is only nontrivial for the uncoupled parts. But for the differential operators δ^x_T on $\mathscr{H}^s_\beta$ and the number operators on $\mathscr{H}^q_\beta$ they are known to exist. Since K^β_s and K^β_q are defined on invariant domains in $\mathscr{H}^s_\beta$ and $\mathscr{H}^q_\beta$, respectively, the self-adjoint extensions are unique and generate unique unitary groups, which in turn implement the mentioned automorphism groups. Since $h^s_\beta \in \mathscr{Z}^\beta$ it drops out from the implementing expression for $\tau^{\beta s}_t$.

(ii) Let us denote the restriction of ω^β to $\mathscr{M}^\beta_q$ by ω^β_q. Since $\mathscr{M}^\beta_q$ is W^*-isomorphic to $\mathscr{M}^{\beta\vartheta}$ it contains in some way $\mathscr{A}^P_a \otimes \mathscr{A}^P_b$. By the associativity of the C^*-tensor product (cf. Ref. 12) we may interpret $\mathscr{A}^P_a \otimes \mathscr{A}^P_b$ as a single Glimm algebra in the sense of Ref. 14 and ω^β_q as a permutation-invariant product state thereon, which still is faithful. Now we can take over the reasoning of Ref. 14, 8.15.12, and conclude that the invariant algebra of the modular automorphism group of ω^β_q is a factor. Up to a time-scale transformation the reduced uncoupled dynamics $\{\tau^{\beta q}_t; t \in \mathbb{R}\}$ constitutes this modular automorphism group. Its spectrum is, therefore, the Connes invariant for $\mathscr{M}^\beta_q$. Since the reduced coupled dynamics is—up to the same time-scale factor—the modular dynamics of the faithful normal state ω'^β_q on $\mathscr{M}^\beta_q$ it contains the quasiparticle spectrum as a spectral part. □

IV. EXTERIOR DYNAMICS AND CURRENT

In order to define the particle current $J^a_\beta(t)$ at the temperature β from a to b as usual by the time derivative of $N^\beta_a(t)$, we first have to formulate an appropriate dynamics for the latter quantity. Since there is no direct relationship to τ'^β_t we look for a natural time dependence for the local approximation (3.6). The part $j(N_{\Lambda_a}) \in \mathscr{M}^{\beta'}$ plays the role of an operator-valued renormalization term and should, therefore, be kept time independent, whereas the first is an element of $\mathscr{A}_\Lambda$ and has the time variation $\tau'^\Lambda_t(N_{\Lambda a})$. One may now perform first the thermodynamic limit and then the differentiation to the time, or first calculate

$$J^a_\Lambda(t) := \frac{d\tau'^\Lambda_t(N_{\Lambda a})}{dt} \tag{4.1}$$

and then do the thermodynamic limit.

Theorem 4.1: The following two families of limits exist:

$$\text{s-resolvent-}\lim_\Lambda \tau'^\Lambda_t(N_{\Lambda a}) - j(N_{\Lambda a}) = : N^\beta_a(t), \tag{4.2}$$

where for all $t \in \mathbb{R}$ the unbounded self-adjoint operators $N^\beta_a(t)$ have the same domain as N^β_a and $N^\beta_a(t) - N^\beta_a \in \mathscr{M}^\beta$;

$$\text{s-}\lim_\Lambda J^a_\Lambda(t) = : J^a_\beta(t), \tag{4.3}$$

where $J^a_\beta(t) \in \mathscr{M}^\beta$ for all $t \in \mathbb{R}$. Furthermore it holds for all $t \in \mathbb{R}$

$$J^a_\beta(t) = \tau'^\beta_t(J^a_\beta(0)) \tag{4.4}$$

and

$$J^a_\beta(t) = \frac{dN^\beta_a(t)}{dt} \tag{4.5}$$

on the domain $\mathscr{D}(N^\beta_a)$.

Proof: (i) Since the commutators of b_{kx} and $m_{\Lambda x}$ with $N_{\Lambda a}$ are either zero or minus times the original quantity, $J^a_\Lambda(0) = [ih_\Lambda, N_{\Lambda a}]$ is a density approximation (cf. Theorem 3.1) for a polynomial in $\mathscr{M}^\beta_P \cap \mathscr{M}^\beta_x$. Thus

$$\text{s-}\lim_\Lambda J^a_\Lambda(t) = \text{s-}\lim_\Lambda \tau'^\Lambda_t(J^a_\Lambda(0))$$
$$= \tau'^\beta_t(J^a_\beta(0)) = J^a_\beta(t),$$

by virtue of (3.15). This gives (4.3) and (4.4).

(ii) The local relation

$$\int_0^t J^a_\Lambda(t')dt' = \tau'^\Lambda_t(N_{\Lambda a}) - N_{\Lambda a}$$

is obtained by integrating (4.1) and takes place in the finite-dimensional algebra $\mathscr{A}_\Lambda$. In view of (i) and the dominated convergence theorem the left-hand side converges in the strong operator topology of $\mathscr{H}_\beta$. In the strong resolvent sense one has then

$$\lim_\Lambda \tau_t'^\Lambda(N_{\Lambda a}) - j(N_{\Lambda a}) = \int_0^t J_\beta^a(t')dt' + N_a^\beta,$$

where (3.6) has been observed. Since the right-hand side is defined on $\mathscr{D}(N_a^\beta)$, the left-hand side is also, and gives, by definition, $N_a^\beta(t)$ for all t. On this domain one can then also differentiate. Thus we have demonstrated (4.2) and (4.5). □

Among the classical observables of the separated superconductors the macroscopic phase operators

$$\Theta_x^\beta : (i/2)\log(s_x/w_x), \quad x \in \{a, b\}, \tag{4.6}$$

are especially important. (We assume now $\beta > \beta_{ca}, \beta_{cb}$.) If the splitting of the interaction (3.28) is valid then the macroscopic phases have the time dependence

$$\Theta_x^\beta(t) = \tau_t'^\beta(\Theta_x^\beta) = \Theta_x^\beta + \mu_x t\, 1, \tag{4.7}$$

by virtue of Theorem 3.2. We realize that this simple time behavior, which is usually motivated mostly by heuristic arguments, is exact under quite general circumstances. It is a logical prerequisite for the formulation of the Josephson relation. More precisely one obtains in this situation the following relations for the current.

Theorem 4.2: If the interaction splits according to (3.29), then the current decomposes as

$$J_\beta^a(t) = J_{\beta s}^a(t) + J_{\beta q}^a(t), \tag{4.8}$$

into a condensed part in $\mathscr{Z}^\beta$ and a quasiparticle part in $\mathscr{M}_q^\beta$, where

$$J_{\beta q}^a(t) = \tau_t'^{\beta q}(J_{\beta q}^a(0)). \tag{4.9}$$

Written as a function of the macroscopic phases the condensed current depends on time via the relation (4.7), and its part due to the dominant interaction (2.34) has the exact form

$$J_{\beta s}^a(t) = 4g_s w_a w_b \sin(2\Theta_a^\beta(t) \cdot - 2\Theta_b^\beta(t)) \tag{4.10}$$

and may also be obtained by the extended Heisenberg equation

$$J_{\beta s}^a(t) = [ih_\beta^s(t), N_a^\beta(t)]. \tag{4.11}$$

Proof: (i) Under the splitting assumption (3.29) we have

$$J_\beta^a(0) = \text{s-}\lim_\Lambda L'_\Lambda(N_{\Lambda a})$$
$$= \text{s-}\lim_\Lambda [ih_\Lambda^s + ih_\Lambda^q, N_{\Lambda a}], \tag{*}$$

where h_Λ^s and h_Λ^q are the density approximations of h_β^s and h_β^q and the gauge invariance of H_Λ has been employed. Since the density approximations $m_{\Lambda x}$ and $q_{k,\Lambda x}$ of s_x and q_{kx} satisfy, due to their correct gauge behavior,

$$[m_{\Lambda a}, N_{\Lambda a}] = 2m_{\Lambda a}, \quad [m_{\Lambda b}, N_{\Lambda a}] = 0,$$
$$[q_{k,\Lambda a}, N_{\Lambda a}] = 2q_{k\Lambda a}, \quad [q_{k\Lambda b}, N_{\Lambda a}] = 0,$$

the two parts in the local commutators (*) converge to elements in $\mathscr{Z}^\beta$ and $\mathscr{M}_q^\beta$, respectively. For the dominant condensed part, one obtains from (2.6), (2.23), and (4.6),

$$J_\beta^a(0) = 4g_s w_a w_b \sin(2\Theta_a^\beta - 2\Theta_b^\beta).$$

From (4.4) and (3.30), together with (4.7), follows the asserted time behavior for both parts of the current.

(ii) In evaluating the right-hand side of (4.11) we arrive at

$$[ig_s(s_a^*(t)s_b(t) + s_a(t)s_b^*(t)), \delta_T^a],$$

which leads directly to (4.10), since the differentiation to Θ_a^β contained in δ_T^a is the same as to $\Theta_a^\beta(t)$. □

V. CONCLUSIONS

The model for two weakly coupled superconductors treated above is simple enough to allow for a deductive step-by-step construction by means of operator algebraic techniques, where, besides the thermodynamic limit, no further approximation is required. Since on the other hand the model exhibits a kind of quantum condensation that is related to observable phenomena, it provides a valuable opportunity for a detailed analysis of the theoretical origin of these unique effects. Here we shall draw some first conclusions on the relationship of macroscopic quantum phenomena to a many-body theory and on the peculiarities of weak coupling models, deferring the comparison with more complex models of the literature to forthcoming treatments.

By virtue of previous investigations[7–9] we are able to give here the existence and explicit form of the limiting Gibbs state for a natural class of weak coupling interactions between two quasispin BCS models. This contrasts our investigation from the outset with all conventional Greens' functions formalisms, where spontaneous symmetry breakdown is treated either by self-consistency equations for the anomalous expectations or by a thermodynamic limit in the presence of an external field. Both methods refer thus to the asymmetric pure phase states. Our closed gauge invariant model, however, displays spontaneous symmetry breakdown at low temperatures by the nontriviality of the central decomposition of the limiting Gibbs states. The first important consequence of the coupling to be weak is the identical parametrization of the coupled (2.25) and uncoupled (2.8) central decompositions by means of two macroscopic gauge angles. That means that the same macroscopic observable values identify the pure phase states in spite of their completely different properties as quantum distributions by virtue of the presence or absence of quasiparticle exchange. The onset of the condensed Cooper pair exchange alters the probability distribution for the pure phases itself from the equipartition to a weighted one, where the leading tunneling term produces the Josephson cosine potential in the exponent of this weight (2.35). In contradistinction to the usual reasoning this potential does not show up in the specific free energy (being averaged out) and does not play a role in the reconstructed electron field dynamics in Sec. III (being associated with a central observable).

A further and related consequence of the weakness of the coupling is the unitary equivalence of the GNS representations corresponding to the coupled and uncoupled equilibrium states. The quasilocal electron pair algebra $\mathscr{A}$ of (2.3) may thus be represented in a Hilbert space $\mathscr{H}_\beta$, which contains the uncoupled limiting Gibbs state as a cyclic vector Ω_β. The representation-dependent weak closure $\mathscr{M}^\beta$ of $\pi_\beta(\mathscr{A})$ represents the canonical extension of the electron algebra appropriate for thermodynamic equilibrium. Any observable, which is locally approximable by electron observables in the weak operator topology of $\mathscr{H}_\beta$, is contained in $\mathscr{M}^\beta$. The center $\mathscr{Z}^\beta$ of $\mathscr{M}^\beta$ is spanned by the (measurable, essentially bounded) functions of the condensed pair field operators s_a and s_b [(2.23)] or, equivalently, by the functions of the macroscopic phase operators Θ_a^β and Θ_b^β [(4.6)]. The concept of the GNS representation provides thus a distinguished set of observables ($\mathscr{M}^\beta$) and especially of macroscopic classical observables ($\mathscr{Z}^\beta$). States that can be discriminated by the values of a central observable are macroscopically different. This is the way in which a microscopic theory determines its macroscopic classical superstructure by itself.

In order to obtain also the macroscopic quantum mechanical superstructure one has to consider the generators of the spontaneously broken symmetries. Due to the separately gauge-invariant, uncoupled, cyclic state vector Ω_β the gauge transformations for the separate subsystems may be unitarily implemented and give the unitary operators W_Θ^β, $\Theta\in\mathbb{R}^2$ [before (3.3)]. Their self-adjoint generators $N_{a/b}^\beta$ count the particles in the different subsystems and are, therefore, of fundamental physical importance. They are, however, not affiliated with $\mathscr{M}^\beta$ and may have properties, which are completely foreign to Fock space number operators. To analyze their structure we have developed two mathematical techniques, which use both certain sub-von Neumann algebras $\mathscr{M}_\Lambda^\beta \subset \mathscr{M}^\beta$ (2.16) in which the central observables are combined with the "local" ones of $\pi_\beta(\mathscr{A}_\Lambda)$, Λ a finite set of electron momenta. Applied to local perturbations of Ω_β the $N_{a/b}^\beta$ count only normal electron pairs and pairs of normal holes, whereas on collective perturbations effectuated by elements from $\mathscr{M}_\Lambda^\beta$ the condensed Cooper pairs and condensed holes are taken into account by means of differential operators to the macroscopic phase angles [(3.3) and (3.4)]. As demonstrated in Refs. 7 and 23 these differential expressions arise by a unique mathematical closure construction and are, therefore, also determined by the microscopic theory. Since, as already mentioned above, they are not sensitive to local perturbations [from a subalgebra $\pi_\beta(\mathscr{A}_\Lambda)\subset\mathscr{M}^\beta$] they belong also to the purely macroscopic aspects of the model theory. Neither are they, however, affiliated with commutant $\mathscr{M}^{\beta\prime}$ of $\mathscr{M}^\beta$, but they give rise to the canonical commutation relations between the $N_{a/b}^\beta$ and $\Theta_{a/b}^\beta$, respectively. And this is exactly the point, where macroscopic quantum theory has its theoretical origin. Of direct experimental relevance is, of course, only the canonical pair of difference observables $(N_a - N_b)/2$ and $\Theta_a^\beta - \Theta_b^\beta$. Their mutual incompatibility signifies the principal incompatibility between a sharp difference number of condensed Cooper pairs and a precise value of the macroscopic phase difference, the latter corresponding to a sharp value of the tunneling super current [cf. (4.10)]: the junction is either open or closed for the supercurrent flow. As will be pointed out in forthcoming investigations, this macroscopic incompatibility is related to macroscopic quantum coherence, to the quantum nature of flux tunneling, and to the commutation relations of quantum electrodynamic circuit theory.[24]

In the present investigation we have concentrated on the careful microscopic foundation of the dynamical equations for the closed system of the weakly coupled superconductors. Since for the bulk properties of superconductors only long-range interacting models (in momentum space) are available the construction of the limiting Heisenberg dynamics for the usual (locally approximable) electron observables is already a highly delicate task. In Sec. III the limiting dynamics is realized as a W^*-automorphism group in $\mathscr{M}^\beta$, which gives, besides other things, the rigorous justification for the macroscopic phase angle dynamics without any dressing from the weak interaction (4.7) and for the extension of the two-fluid picture to this composite system (3.30).

The technically completely new step is the construction of the dynamics for the exterior generators $N_{a/b}^\beta$ in Sec. IV. Here we use for the second time the sub-von Neumann algebras $\mathscr{M}_\Lambda^\beta$ and show that they reduce the modular quantities of the Tomita–Takesaki theory (Proposition 2.2). By means of the corresponding modular involutions every local Fock number operator obtains a canonical operator-valued subtraction term in $\mathscr{M}^{\beta\prime}$ such that the combined expressions converge to $N_{a/b}^\beta$ (in the strong resolvent sense of $\mathscr{H}_\beta$) and constitute, so to speak, a pseudolocal approximation of the latter. Applying the local weak coupling dynamics only to the local Fock parts one obtains in the thermodynamic limit a well-behaved time dependence for $N_{a/b}^\beta$.

The inner and exterior limiting dynamics provide us with microscopically derived relations between the time-dependent particle number, current, and phase operators. Observing that quasiparticle pairs do not condense, the current decomposes into a normal and a condensed part. The expectation values of the supercurrent and the phase operators in the pure phase states reproduce the usual Josephson relations. The operator expressions for themselves provide, however, a much more detailed picture. Those operators which commute with all local elements of the electron algebra constitute the purely macroscopic part of the model and are divided into macroscopic classical and macroscopic quantum observables if they are affiliated with the center $\mathscr{Z}^\beta$ or are incompatible with some central elements, respectively. In this terminology the phase and supercurrent operators are macroscopic classical and those parts of the number operators that count condensed Cooper pairs are macroscopic quantum mechanical.

The prerequisites for the Josephson effect are quantum condensation, which makes the macroscopic part nontrivial, and weak coupling, which leaves the macroscopic classical part of the original superconductors unchanged but which is strong enough to give a tunneling supercurrent. The Josephson cosine potential has (beside its statistical role) a dynamic effect only on the macroscopic quantum part [cf. (4.11)]. That the latter effect may be formulated in a single GNS representation is also due to the weakness of the coupling. If the effective, temperature-dependent coupling poential increases, as in the low-temperature limit, also the classical macroscopic structure will become homogeneous in equilibrium, as is indicated by the equality of the macroscopic phase values [cf. the reasoning after (2.36)]. From our discussion it is unlikely, that there is a potential which leads to a Josephson effect both for finite and zero temperatures.

The most important conclusion of the present model investigation is the principal possibility to reduce all macroscopic classical and quantum mechanical notions to canonically described limits of traditional quantum mechanical concepts.

ACKNOWLEDGMENT

This work has been supported by the Deutsche Forschungsgemeinschaft.

[1]J. E. Mercereau, in *Superconductivity*, edited by R. D. Parks (Dekker, New York, 1969).

[2]B. D. Josephson, in Ref. 1.

[3]O. Bratteli and D. W. Robinson, *Operator Algebras and Quantum Statistical Mechanics*, I and II (Springer, New York, 1979 and 1981).

[4]K. Hepp, Ann. Phys. **90**, 285 (1975).

[5]W. Thirring, Commun. Math. Phys. **7**, 181 (1968).

[6]W. Fleig, Acta Phys. Austriaca **55**, 135 (1983).

[7]A. Rieckers and M. Ullrich, Acta Phys. Austriaca **56**, 131 (1985).

[8]A. Rieckers and M. Ullrich, Acta Phys. Austriaca. **56**, 259 (1985).

[9]M. Ullrich, "Calculation of the limiting Gibbs states for weakly coupled macroscopic quantum systems with application to the Josephson oscillator," to appear in Rep. Math. Phys.

[10]B. Josephson, thesis, Cambridge University, 1964.

[11]M. Tinkham, *Introduction to Superconductivity* (McGraw–Hill, Kogalinska, Tokyo, 1975).

[12]S. Sakai, *C*-Algebras and W*-Algebras* (Springer, Berlin, 1971).

[13]M. Takesaki, *Theory of Operator Algebras. I* (Springer, New York, 1979).

[14]G. K. Pedersen, *C*-Algebras and their Automorphism Groups* (Academic, London, 1979).

[15]A. Barone and G. Paterno, *Physics and Applications of the Josephson Effect* (Wiley, New York, 1982).

[16]R. Haag, R. V. Kadison, and D. Kastler, Commun. Math. Phys. **16**, 81 (1970).

[17]M. Winnink, in *Statistical Mechanics and Feild Theory*, edited by R. N. Sen and C. Weil (Keter, Jerusalem, 1972).

[18]A. Rieckers, Acta Phys. Austriaca, Suppl. **XXVI**, 423 (1984).

[19]S. Stratila and L. Zsido, *Lectures on von Neumann Algebras* (Abacus, Tunbridge Wells, Kent, 1979).

[20]E. B. Davies, *One-Parameter Semigroups* (Academic, London, 1980).

[21]E. Duffner, Physica A **133**, 187 (1985).

[22]E. Duffner, "Nonequilibrium pure phase dynamics for the strong coupling Josephson tunnel junction," to appear in Z. Phys. B.

[23]A. Rieckers, J. Math. Phys. **25**, 2593 (1985).

[24]A. Widom, J. Low Temp. Phys. **37**, 449 (1979).

Exact Bogoliubov limits for the Bassichis–Foldy model and continued fractions

D. Masson
Department of Mathematics, University of Toronto, Toronto M5S 1A1, Canada

(Received 10 October 1985; accepted for publication 4 December 1985)

The critical values of coupling for the Bassichis–Foldy model are explained. For a finite number of bosons N they are associated with changes in the bounded above and below properties of the Hamiltonian. Exact $N \to \infty$ spectral properties are obtained using continued fractions and a duality in terms of competing Bogoliubov-type limits is exhibited. The critical coupling limits are then associated with transitions from either a discrete to a continuous spectrum or from one Bogoliubov-type limit to another.

I. INTRODUCTION

The Bassichis–Foldy (BF) model[1] was introduced in 1964 in order to test the validity of the Bogoliubov c-number approximation.[2] The basic BF model of N bosons was shown to yield low-energy eigenvalues that approached the standard Bogoliubov approximation as $N \to \infty$ for a certain range of two-body interactions. However, for a range of interaction of a modified BF model there was lack of agreement and indications of a transition to a different (incorrectly given) Bogoliubov-type limit.

The basic model was later used to rederive the standard Bogoliubov limit using group theory methods.[3] These methods suggested (but did not prove) the existence of a continuous spectrum limit for a certain range of coupling.

More recently[4] the model was used to examine an algebraic approach to a $1/N$ expansion. Extensive numerical calculations indicated certain critical values of coupling and an interesting duality between attraction and repulsion. Correct Bogoliubov-type limits were rederived for a range of couplings but what happens for other ranges of coupling was left open for speculation and interpretation.

From the above works there emerges no clear consensus regarding the precise role of the Bogoliubov-type limits and the critical couplings or the existence of continuous spectrum limits and/or phase transitions.

In the present paper we reexamine the BF model for an entire range of two-body interactions both for finite N and $N \to \infty$. The critical values of coupling are explained. For finite N they are associated with changes in the bounded above and below properties of the Hamiltonian. Exact $N \to \infty$ spectral properties are obtained using continued fractions and a duality in terms of competing Bogoliubov-type limits is exhibited. For $N \to \infty$ the critical values of coupling are then associated with either transitions from a discrete to a continuous spectrum or a transition from one Bogoliubov-type limit to another. It is the latter that may be regarded as a true phase transition.

The basic model is introduced in Sec. II and finite N bounds derived in Sec. III. Exact $N \to \infty$ limits are obtained in Sec. IV for the basic model and in Sec. V for the modified model. The results hinge on the convergence properties of some simple continued fractions for which some relevant formulas are given in Appendices A and B.

II. BASIC BF MODEL[1]

In the absence of interactions, the model consists of three types of bosons having creation and annihilation operators denoted by $a^*, a, a_+^*, a_+, a_-^*, a_-$, with the first pair associated with a zero-energy state and the latter two pairs associated with degenerate excited states of unit energy.

The interacting Hamiltonian of the basic model is given by

$$H = a_+^* a_+ + a_-^* a_- + g[a^*a(a_+^* a_+ + a_-^* a_-) + a^{*2}a_+a_- + a^2 a_+^* a_-^*], \tag{1}$$

where

$$N = a^*a + a_+^* a_+ + a_-^* a_- \tag{2}$$

is a constant of motion. The Hilbert space is given by

$$\mathscr{H} = \text{closed span}\left\{\frac{a^{*n_1}a_+^{*n_2}a_-^{*n_3}|0\rangle}{\sqrt{n_1!n_2!n_3!}}\right\}_{n_i=0}^{\infty}.$$

One also has

$$\Delta = a_+^* a_+ - a_-^* a_-,$$

a constant of motion, and the invariant subspaces

$$\mathscr{H}_{N,\Delta} = \text{span}\{|N,\Delta,n\rangle\}_{n=0}^{(N-\Delta)/2},$$

where

$$|N,\Delta,n\rangle = \frac{a^{*N-\Delta-2n}a_+^{*n+\Delta}a_-^{*n}|0\rangle}{\sqrt{(N-\Delta-2n)!}\sqrt{(n+\Delta)!}\sqrt{n!}}. \tag{3}$$

For simplicity we restrict attention to the subspace $\mathscr{H}_{N,0}$ so that the eigenvalue equation $H\psi = E\psi$ is then unitarily equivalent to a $(N/2+1)\times(N/2+1)$ tridiagonal matrix eigenvalue problem with

$$\psi = \sum_{n=0}^{N/2} c_n |N,0,n\rangle, \tag{4}$$

$$U\psi = (c_0, c_1, \ldots, c_{N/2})^t, \tag{5}$$

$$UHU^* = \begin{pmatrix} a_0 & b_1 & & 0 \\ b_1 & a_1 & \ddots & \\ & \ddots & \ddots & b_{N/2} \\ 0 & & b_{N/2} & a_{N/2} \end{pmatrix}, \tag{6}$$

 0022-2488/86/041093-06$02.50

$$a_n = 2n[1 + g(N-2n)], \tag{7}$$

$$b_n = gn[(N-2n+2)(N-2n+1)]^{1/2}, \tag{8}$$

and U a unitary map from $\mathscr{H}_{N,0}$ to $\mathbb{C}^{N/2+1}$.

For $g \neq 0$ the eigenvalues of H can then be identified with the poles of the finite fraction

$$\langle N,0,0|(E-H)^{-1}|N,0,0\rangle = \cfrac{1}{E-a_0-\cfrac{b_1^2}{E-a_1-\cfrac{b_2^2}{\ddots\, \overline{E-a_{N/2}}}}}. \tag{9}$$

It is this last expression that will be exploited in Sec. IV A to obtain $N\to\infty$ spectral properties since, with $g = c/N$ and $N\to\infty$, (9) becomes a continued fraction with known convergence properties.

III. FINITE-N BOUNDS

For finite N, Bassichis and Foldy[1] and Carr[4] either have made numerical calculations or resorted to large N approximations. Here we exploit the exact form of H to obtain exact Hamiltonian bounds.

Let $A = a^*a_- + aa^*_+$ and $n = a^*_+a_+ = a^*_-a_-$. Then (1) can be expressed as

$$H = 2n - g(N-n) + gA^*A. \tag{10}$$

This yields the inequalities

$$H \geqslant 2n - g(N-n), \quad g \geqslant 0,$$
$$H \leqslant 2n - g(N-n), \quad g \leqslant 0,$$

and hence (noting that $0 \leqslant n \leqslant N/2$) the bounds

$$H \geqslant -gN, \quad g \geqslant 0, \tag{11}$$

$$H \leqslant N(1-g/2), \quad -2 \leqslant g \leqslant 0. \tag{12}$$

With $B = a^*a_- - aa^*_+$ one can reexpress H as

$$H = 2n + g[4n(N-2n) + N - n] - gB^*B. \tag{13}$$

This now yields the inequalities

$$H \geqslant -8gn^2 + (2 + 4gN - g)n + gN, \quad g \leqslant 0,$$
$$H \leqslant -8gn^2 + (2 + 4gN - g)n + gN, \quad g \geqslant 0,$$

and hence (by completing the square of the quadratic in n or $m = N/2 - n$, respectively) the bounds

$$H \geqslant gN, \quad -2/(4N-1) \leqslant g \leqslant 0, \tag{14a}$$

$$H \geqslant gN + (4gN + 2 - g)^2/32g, \quad g \leqslant -2/(4N-1), \tag{14b}$$

$$H \leqslant N(1+g/2), \quad 0 \leqslant g \leqslant 2/(4N+1), \tag{15a}$$

$$H \leqslant N(1+g/2) + (4gN - 2 + g)^2/32g, \quad g \geqslant 2/(4N+1). \tag{15b}$$

From (14a), (14b), (15a), and (15b), one sees that $g = \pm 2/(4N \pm 1)$ are critical values of coupling where the derived bounded above and below properties of H undergo a transition. This transition is supported by numerical calculations with the above bounds surprisingly tight.[4] In terms of $c = gN$ and $N\to\infty$ the critical values correspond to $c = \pm\frac{1}{2}$. These critical value limits are interpreted in terms of Bogoliubov limits in the next section.

IV. BOGOLIUBOV LIMITS

Boboliubov limits are traditionally obtained from an *ad hoc* replacement of certain operators by c-numbers.[1] In Ref. 4 the approach is different, with $1/N$ projections and rotations exploited. Here we note that the approach to the Bogoliubov limit may be expressed rigorously in terms of generalized strong operator (resolvent) convergence. For our present purposes it suffices to consider only two different matrix elements.

A. Standard Bogoliubov

Let $g = c/N$ and consider the weak coupling limit $N\to\infty$ of (9). One obtains the infinite continued fraction representation

$$\lim_{N\to\infty} \langle N,0,0|(E-H)^{-1}|N,0,0\rangle = f(E,c),$$

where

$$f(E,c) = \cfrac{1}{E-\alpha_0-\cfrac{\beta_1^2}{E-\alpha_1-\cfrac{\beta_2^2}{\ddots}}}, \tag{16}$$

$$\alpha_n = \lim_{N\to\infty} a_n = 2n(1+c), \tag{17}$$

$$\beta_n = \lim_{N\to\infty} b_n = nc. \tag{18}$$

This continued fraction can be exactly "summed" to yield (see Appendix A)

$$f(E,c) = -\frac{\Gamma(x - E/2\sqrt{1+2c})}{2\sqrt{1+2c}\,\Gamma(x - E/2\sqrt{1+2c}+1)} \times {}_2F_1(1,1;x - E/2\sqrt{1+2c}+1;x), \tag{19}$$

with $x = (1 - (1+c)/\sqrt{1+2c})/2$, $1 + 2c > 0$. Since ${}_2F_1(a,b;c;z)/\Gamma(c)$ is an entire function of c one sees that the only E singularities are given by $\Gamma(x - E/2\sqrt{1+2c})$ singular. That is, $x - E/2\sqrt{1+2c} = -n$, $n = 0,1,\ldots$, yields poles in the continued fraction giving the $N\to\infty$ eigenvalues

$$E_n = -(1+c) + \sqrt{1+2c}(2n+1), \quad n = 0,1,\ldots, \quad c > -\tfrac{1}{2}, \tag{20}$$

and the expression (see also Ref. 1)

$$f(E,c) = (1-x)^{-1}\sum_{n=0}^{\infty}\frac{(-1)^n(x/(1-x))^n}{E - 2(x+n)\sqrt{1+2c}}, \quad c > -\tfrac{1}{2}.$$

For the special case $c = -\frac{1}{2}$ one has (see Ref. 5, Eq. 92.7)

$$f(E, -\tfrac{1}{2}) = \int_0^\infty \frac{e^{-u}\,du}{E + (1-u)/2}, \tag{21}$$

and hence a continuous spectrum $[-\frac{1}{2},\infty)$.

For $c<-\frac{1}{2}$, $\mathrm{Im}\,E=0$ is a singular boundary and one obtains

$$f(E,c)=\int_{-\infty}^{\infty}\frac{\sigma(u,c)du}{E-u},\quad \mathrm{Im}\,E\neq 0,\quad c<-\frac{1}{2},\tag{22}$$

with $\sigma(u,c)$ real, positive, and continuous. The complex poles (quasiparticles or resonances)

$$E_n=-(1+c)\pm\sqrt{|1+2c|}(2n+1)i,\tag{20'}$$

which correspond to (20), now occur on the second sheet in E as one continues (from above or below, respectively) through the spectral cut on $(-\infty,\infty)$. In particular (see Appendix B), for $c=-1$, one has

$$\begin{aligned}f(E,-1)&=(1/2i)[\psi((3-iE)/4)-\psi((1-iE)/4)]\\&=2\sum_{n=0}^{\infty}\frac{(-1)^n}{E+(2n+1)i},\quad \mathrm{Im}\,E>0,\\&=\frac{1}{2}\int_{-\infty}^{\infty}\frac{\mathrm{sech}(\pi u/2)du}{E-u},\quad \mathrm{Im}\,E\neq 0,\end{aligned}\tag{23}$$

where $\psi(z)=\Gamma'(z)/\Gamma(z)$.

The above $N\to\infty$ limits may be collectively expressed in terms of the standard Bogoliubov Hamiltonian H_{B} obtained from H by replacing a^* and a by $\sqrt{N}$. That is,

$$\lim_{N\to\infty}\langle N,0,0|(E-H)^{-1}|N,0,0\rangle={}_{\mathrm{B}}\langle 0|(E-H_{\mathrm{B}})^{-1}|0\rangle_{\mathrm{B}},$$

where

$$H_{\mathrm{B}}=a_+^*a_++a_-^*a_-+c(a_+^*a_++a_-^*a_-+a_+^*a_-^*+a_+a_-)\tag{24}$$

acts in the Hilbert space

$$\mathscr{H}_{\mathrm{B}}=\text{closed span}\left\{\frac{a_+^{*^n}a_-^{*^n}|0\rangle_{\mathrm{B}}}{n!}\right\}_{n=0}^{\infty}$$

and has spectrum

$$\sigma(H_{\mathrm{B}})=\begin{cases}-(1+c)+\sqrt{1+2c}(2n+1),\\ \qquad n=0,1,\ldots,\quad c>-\frac{1}{2},\\ [-\frac{1}{2},\infty),\quad c=-\frac{1}{2},\\ (-\infty,\infty),\quad c<-\frac{1}{2}.\end{cases}$$

B. Inverted Bogoliubov

In (4) the basis vectors were ordered with respect to n. If instead one orders them with respect to $m=N/2-n$ and renormalizes the Hamiltonian by subtracting N, one obtains an equivalent tridiagonal matrix eigenvalue problem

$$V(H-N)V^*V\psi=E'V\psi,\tag{25}$$

with

$$\psi=\sum_{m=o}^{N/2}c'_m|N,0,(N-2m)/2\rangle,\tag{26}$$

$$V\psi=(c'_0,c'_1,\ldots,c'_{N/2})^t,\tag{27}$$

$$V(H-N)V^*=\begin{pmatrix}a'_0 & b'_1 & & 0\\ b'_1 & & \ddots & \\ & \ddots & & \\ & & \ddots & b'_{N/2}\\ 0 & & b'_{N/2} & a'_{N/2}\end{pmatrix},\tag{28}$$

$$a'_m=2m[(N-2m)g-1],\tag{29}$$

$$b'_m=g(N-2m+2)[(2m-1)(2m)]^{1/2}/2,\tag{30}$$

where, of course, $c'_m=c_{N/2-m}$, $a'_m=a_{N/2-m}-N$, $b'_m=b_{N/2-m+1}$, and $E'=E-N$.

From this inverted viewpoint the eigenvalues of $H-N$ are the E' poles of the finite fraction

$$\langle N,0,N/2|(E'-H+N)^{-1}|N,0,N/2\rangle=\cfrac{1}{E'-a'_0-\cfrac{b'^2_1}{\ddots\;\overline{E'-a'_{N/2}}}}.\tag{31}$$

With $g=c/N$ the $N\to\infty$ limit of (31) becomes the continued fraction

$$F(E',c)=\cfrac{1}{E'-\alpha'_0-\cfrac{\beta'^2_1}{E'-\alpha'_1-\cfrac{\beta'^2_2}{\ddots}}},\tag{32}$$

with

$$\alpha'_m=\lim_{N\to\infty}a'_m=2m(c-1),\tag{33}$$

$$\beta'_m=\lim_{N\to\infty}b'_m=c[(2m-1)(2m)]^{1/2}/2.\tag{34}$$

This may also be "summed" exactly to obtain (see Appendix A)

$$F(E',c)=\frac{\Gamma((x+E'/\sqrt{1-2c})/2)}{2\sqrt{1-2c}\,\Gamma((x+E'/\sqrt{1-2c})/2+1)}\times{}_2F_1(1,1/2;(x+E'/\sqrt{1-2c})/2+1;x),\tag{35}$$

with $x=(1+(c-1)/\sqrt{1-2c})/2$, $1-2c>0$. The limiting eigenvalues are thus given by the condition $x+E'/\sqrt{1-2c}=-2m$, $m=0,1,\ldots$, yielding the $N\to\infty$ eigenvalues (see also Ref. 4)

$$E'_m=(1-c)/2-\sqrt{1-2c}(2m+\tfrac{1}{2}),\quad m=0,1,\ldots,\quad c<\tfrac{1}{2}.\tag{36}$$

For the special case $c=\frac{1}{2}$ one has [Ref. 6, Eq. (92.7)]

$$F\left(E',\frac{1}{2}\right)=\frac{1}{\sqrt{\pi}}\int_0^{\infty}\frac{e^{-u}u^{-1/2}du}{E'-\frac{1}{4}+u/2},\quad E'\notin(-\infty,\tfrac{1}{4}],\tag{37}$$

while for $c>\frac{1}{2}$ one obtains

$$F(E',c)=\int_{-\infty}^{\infty}\frac{\tau(u,c)du}{E'-u},\quad \mathrm{Im}\,E'\neq 0,\quad c>\tfrac{1}{2},$$

with τ real, positive, and continuous.

These inverted Bogoliubov limits may also be summarized according to

$$\lim_{N\to\infty} \langle N,0,N/2|(E'-H+N)^{-1}|N,0,N/2\rangle = {}_{B'}\langle 0|(E'-H_{B'})^{-1}|0\rangle_{B'},$$

where

$$H_{B'} = a^*a(c-1) + c(a^{*2}+a^2)/2 \tag{38}$$

acts in the Hilbert space

$$\mathscr{H}_{B'} = \text{closed span}\left\{ \frac{a^{*2m}|0\rangle_{B'}}{\sqrt{(2m)!}} \right\}_{m=0}^{\infty}$$

and has spectrum

$$\sigma(H_{B'}) = \begin{cases} (1-c)/2 - \sqrt{1-2c}(2m+\frac{1}{2}), \\ \quad m=0,1,\ldots, \quad c<\frac{1}{2}, \\ (-\infty,\frac{1}{4}], \quad c=\frac{1}{2}, \\ (-\infty,\infty), \quad c>\frac{1}{2}. \end{cases}$$

Conclusion: For $N\to\infty$ the basic BF Hamiltonian can be associated, through resolvent convergence, with two Bogoliubov-type limits each acting in a different Hilbert space. For $-\frac{1}{2}<c<\frac{1}{2}$ one has discrete spectrum limits with the low-energy BF spectrum dominated by the standard Bogoliubov limit[1] and the high-energy BF spectrum dominated by the inverted Bogoliubov limit.[4] For the other ranges of coupling ($c\leqslant -\frac{1}{2}$ and $c\geqslant\frac{1}{2}$) one of these Bogoliubov-type limits has a continuous spectrum.

V. MODIFIED MODEL

The modified BF model[1,4,6] takes into account the additional two-body interaction $a_+^* a_+ a_-^* a_-$ with a modified Hamiltonian

$$H_h = H + h a_+^* a_+ a_-^* a_-, \tag{39}$$

where h is a new independent coupling parameter.

From (10) one obtains

$$H_h \geqslant hn^2 + (2+g)n - gN, \quad g\geqslant 0,$$
$$H_h \leqslant hn^2 + (2+g)n - gN, \quad g\leqslant 0,$$

and hence the modified bounds

$$H_h \geqslant -gN, \quad g\geqslant 0, \quad h\geqslant -(4+2g)/N, \tag{40a}$$

$$H_h \geqslant N(1-g/2) + hN^2/4, \quad g\geqslant 0, \quad h\leqslant -(4+2g)/N, \tag{40b}$$

$$H_h \leqslant N(1-g/2) + hN^2/4, \quad -2\leqslant g\leqslant 0, \quad h\geqslant -(2+g)/N, \tag{41a}$$

$$H_h \leqslant -gN - (2+g)^2/4h, \quad -2\leqslant g\leqslant 0, \quad h\leqslant -(2+g)/N. \tag{41b}$$

These bounds thus undergo a transition at the critical values $h=-(4+2g)/N$ (for $g\geqslant 0$) and $h=-(2+g)/N$ (for $-2\leqslant g\leqslant 0$). With $c=gN$ and $d=hN$ and $N\to\infty$ this suggests the existence of critical weak coupling limits given by $d=-4$, $c>0$ and $d=-2$, $c<0$ associated with the low- and high-energy modified BF spectrums, respectively. The existence of these critical value limits is substantiated by the results of Ref. 4 and the inverted Bogoliubov limit for the modified BF model derived below.

With the modified BF model the standard Bogoliubov limit results of Sec. IV A remain unchanged. The inverted Bogoliubov limit is, however, completely altered.

One now has

$$\lim_{N\to\infty} \langle N,0,N/2|(E'-H_h+N(1+hN/4))^{-1}|N,0,N/2\rangle = F(E',c,d), \tag{42}$$

with

$$F(E',c,d) = \cfrac{1}{E' - \cfrac{c^2(1)(\frac{1}{2})}{E'-2(c-1-d/2) - \cfrac{c^2(2)(\frac{3}{2})}{E'-4(c-1-d/2)-\ddots}}}, \tag{43}$$

and $E' = E - N(1+d/4)$.

From Appendix A, one obtains, for $(1-c+d/2)^2>c^2$,

$$F(E',c,d) = \pm \frac{\Gamma(\gamma)}{2\sqrt{(1-c+d/2)^2-c^2}\,\Gamma(1+\gamma)} \times {}_2F_1(1,\tfrac{1}{2};1+\gamma;x),$$

$$\gamma = (x \pm E'/\sqrt{(1-c+d/2)^2-c^2})/2, \tag{44}$$

$$x = (1 \pm (c-1-d/2)/\sqrt{(1-c+d/2)^2-c^2})/2,$$

with the upper sign for $c-1-d/2<0$ and the lower sign for $c-1-d/2>0$. This yields the eigenvalue condition $\gamma=-m$, $m=0,1,\ldots$, and hence the eigenvalues

$$E'_m = -(c-1-d/2)/2 \mp \sqrt{(1-c+d/2)^2-c^2}(2m+\tfrac{1}{2}). \tag{45}$$

[Note that for $d=-4$ one has $E'_0 = E_0/2$, where E_0 is the standard Bogoliubov lowest-energy eigenvalue given by (20).]

Thus one has an inverted spectrum limit which is discrete for $(1-c+d/2)^2>c^2$ and bounded below (above) for $c-1-d/2>0$ (<0) and which is continuous for $(1-c+d/2)^2\leqslant c^2$, $c\neq 0$. The critical values for the transition from a discrete to a continuous spectrum are given by $d=-2$, $c\neq 0$ and $c=(2+d)/4$, $d\neq -2$.

It is only the critical value limit $d_T=-4$, $c>-\frac{1}{2}$, that can be associated with a true phase transition. For this critical value, numerical calculations[1,4,6] indicate that the modified BF model has a ground state eigenvalue, which, as $N\to\infty$, makes a transition from the standard Bogoliubov limit to the inverted Bogoliubov limit.

For finite N the approximate transition value should thus be given by equating $E'_0 + (1+d/4)N$ and E_0. Since, as previously noted, one has $E'_0=E_0/2$ at $d=-4$, this yields the asymptotic expansion

TABLE I. Values of d_T. Column 4 is from Ref. 6 with $d = -Fc$. Column 5 is the corrected Bogoliubov approximation given by Eq. (46).

$c=gN$	g	N	$d_T(BF)$	$-4+2E_0/N$
0.20	0.01	20	−4.001 68	−4.001 68
0.64	0.01	64	−4.004 06	−4.004 06
1.00	0.01	100	−4.005 36	−4.005 36
2.00	0.01	200	−4.007	−4.007 64
3.20	0.05	64	−4.046 0	−4.046 25
6.40	0.10	64	−4.114	−4.115 15

$$d_T(\mathrm{BF}) = -4 + 2E_0/N + O(1/N^2). \tag{46}$$

This corrects the formula given in Refs. 1 and 6, where the inverted Bogoliubov limit is incorrectly given. That is, one should approximate H_h by $N(1+d/4) + H_{\mathrm{B'}}$, where

$$H_{\mathrm{B'}} = a^*a(c-1-d/2) + c(a^{*2}+a^2)/2. \tag{47}$$

For an alternative derivation of the correct inverted Bogoliubov limit for the case $(1-c+d/2)^2 > c^2$, see Ref. 4.

In Table I some (Ref. 6) finite N transition values are listed together with the corrected Bogoliubov approximation given by (46). Agreement is now seen to be excellent.

ACKNOWLEDGMENTS

The author is thankful to R. R. Carr for several interesting discussions.

This work was partially supported by the Natural Sciences and Engineering Research Council of Canada.

APPENDIX A: A GENERAL FORMULA

The continued fractions encountered in previous sections are all of the form

$$\text{C.F.} = d + \cfrac{a+b+c}{d+e+\cfrac{a+2b+4c}{d+2e+\cfrac{a+3b+9c}{d+3e+\ddots}}}, \tag{A1}$$

with $c \neq 0$. Perron[7] (p. 488) proves the following convergence formula for the case $c,e,e^2+4c \neq 0$ and $|\arg(e^2/(e^2+4c))| < \pi$:

$$\text{C.F.} = \sqrt{e^2+4c}\,\frac{\Gamma(\gamma+1)}{\Gamma(\gamma)} \times \frac{{}_2F_1(\alpha,\beta;\gamma;x)}{{}_2F_1(\alpha+1,\beta+1;\gamma+1;x)},$$
$$\gamma = [(b+c)/2c](1-e/\sqrt{e^2+4c}) + d/\sqrt{e^2+4c}, \tag{A2}$$
$$x = (1-e/\sqrt{e^2+4c})/2,$$

where α,β are the roots of $cz^2 - bz + a = 0$ and the branch of $\sqrt{e^2+4c}$ is chosen so that $\mathrm{Re}(e/\sqrt{e^2+4c}) > 0$.

In Ref. 8, Eq. (A2) is shown to be also valid for the case $e=0$, $c,d \neq 0$, $|\arg(c/d^2)| < \pi$ with the square root branch chosen so that $\mathrm{Re}(d/\sqrt{c}) > 0$.

APPENDIX B: PARTICULAR FORMULAS

We collect here some representation formulas associated with the continued fraction (16) with $c = -1$. One has (Rogers[9])

$$f(E,-1) = \cfrac{1}{E - \cfrac{1}{E - \cfrac{4}{E - \cfrac{9}{\ddots}}}}, \tag{B1}$$

having a large-E asymptotic expansion

$$f(E,-1) \sim \frac{1}{E} + \frac{1}{E^3} + \frac{5}{E^5} + \cdots + (-1)^n \frac{E_{2n}}{E^{2n+1}} + \cdots \tag{B2}$$

(E_{2n} are the Euler numbers $E_0 = 1$, $E_2 = -1$, $E_4 = 5$, $E_6 = -61$, etc.) and an integral representation

$$f(E,-1) = -i\int_0^\infty e^{iEt}\,\mathrm{sech}\,t\,dt, \quad \mathrm{Im}\,E > 0. \tag{B3}$$

One also has the convergence formula [see Ref. 5, Eq. (94.11), see also Ref. 8]

$$f(E,-1) = (1/2i)G(\tfrac{1}{2} - iE/2), \quad \mathrm{Im}\,E > 0, \tag{B4}$$

where

$$G(z) = \psi((1+z)/2) - \psi(z/2) \tag{B5}$$

and $\psi(z)$ is the digamma function $\Gamma'(z)/\Gamma(z)$.

From (B4) and the expression (see Ref. 10, p. 20)

$$G(z) = 2\sum_{n=0}^{\infty}(-1)^n(z+n)^{-1},$$

one obtains

$$f(E,-1) = 2\sum_{n=0}^{\infty}\frac{(-1)^n}{E+(2n+1)i}, \quad \mathrm{Im}\,E > 0. \tag{B6}$$

Finally one can obtain the integral representation

$$f(E,-1) = \frac{1}{2}\int_{-\infty}^{\infty}\frac{\mathrm{sech}(\pi u/2)du}{E-u}, \quad \mathrm{Im}\,E \neq 0. \tag{B7}$$

To derive (B7) one may note that (Ref. 10, p. 42)

$$E_{2n} = (-1)^n\int_0^\infty u^{2n}\,\mathrm{sech}(\pi u/2)du \tag{B8}$$

and use the fact that (B1) is a real J-fraction for a determinant moment problem.[5]

An alternative derivation is to use the standard integral representation (see Ref. 10, p. 18)

$$\psi(z) = \ln z - \frac{1}{2z} - 2\int_0^\infty \frac{s\,ds}{(s^2+z^2)(e^{2\pi s}-1)}, \quad \mathrm{Re}\,z > 0, \tag{B9}$$

which one can write as

$$\psi(z) = \ln z - \frac{1}{2z} - \lim_{\epsilon\to 0^+}\int_\epsilon^\infty\left(\frac{1}{t+iz} + \frac{1}{t-iz}\right) \times (e^{2\pi t}-1)^{-1}\,dt, \quad \mathrm{Re}\,z > 0.$$

From this one obtains (with changes of integration variable $t = s \pm i/2$ together with a change of contour integration)

$$\psi\left(\frac{1+z}{2}\right)=\ln\frac{z}{2}+\int_0^\infty \frac{2s(e^{2\pi s}+1)^{-1}\,ds}{s^2+z^2/4},$$
$$\operatorname{Re} z>0. \tag{B10}$$

From (B5), (B9), and (B10) one obtains (with $v=2s$)

$$G(z)=\frac{1}{z}+\int_0^\infty \frac{2v\,dv}{(v^2+z^2)(\sinh \pi v)}, \quad \operatorname{Re} z>0. \tag{B11}$$

From (B11) with $u=2v\pm i$ it follows that [just as one obtains (B10) from (B9)]

$$G\left(\frac{1}{2}+\frac{z}{2}\right)=\int_0^\infty \frac{2z\,du}{(u^2+z^2)\cosh(\pi u/2)},$$
$$\operatorname{Re} z>0, \tag{B12}$$

which, using (B4), yields (B7).

[1]W. H. Bassichis and L. L. Foldy, Phys. Rev. A **133**, 935 (1964). For a bibliography see Ref. 4.
[2]N. N. Bogoliubov, J. Phys. (USSR) **9**, 23 (1947).
[3]A. I. Solomon, J. Math. Phys. **12**, 390 (1971); A. I. Solomon, J. Phys. A **14**, 2177 (1981); A. M. Perelomov, Sov. Phys. Usp. **20**, 703 (1977).
[4]R. R. Carr, thesis, University of Toronto, 1985.
[5]H. S. Wall, *Analytic Theory of Continued Fractions* (van Nostrand, Princeton, NJ, 1948).
[6]M. Razavy and E. S. Krebes, Can. J. Phys. **51**, 292 (1973).
[7]O. Perron, *Die Lehre von den Kettenbruchen* (Druck, Leipzig, 1929).
[8]D. Masson, Can. Math. Bull. **28**, 411 (1985).
[9]L. J. Rogers, Proc. Lond. Math. Soc. (2) **4**, 72 (1907).
[10]Bateman Manuscript Project, *Higher Transcendental Functions*, Vol. 1, edited by A. Erdélyi (McGraw-Hill, New York, 1953).

Thermodynamics and molecular freedom of dimers on plane triangular lattices

Alain J. Phares and Francis J. Wunderlich
Department of Physics, Villanova University, Villanova, Pennsylvania 19085

(Received 26 June 1985; accepted for publication 20 November 1985)

This is an extension of three previous papers dealing with dimers on rectangular lattices (one, two, and three dimensions). The technique presented in the first paper in this series continues to be fruitful for dimers on plane triangular lattices. Entropy, isothermal compressibility, constant pressure heat capacity, and molecular freedom per dimer at close packing are obtained exactly for lattices infinite in one direction and finite in the other. Observations made in the third paper of the series concerning molecular freedom per dimer at close packing on rectangular lattices are used to extrapolate our results to infinite plane triangular lattices. At close packing, the molecular freedom per dimer on an infinite plane triangular lattice is calculated to be 2.356 527... in agreement with the value obtained by Nagle. Based on our earlier findings, the value of 2.356 527... was used to obtain the analytic fit for the thermodynamic quantities in terms of the normalized number density.

I. INTRODUCTION

Three previous articles,[1–3] referred to as papers I, II, and III, respectively, were devoted to the study of dimers on rectangular lattices. Paper I presented general mathematical techniques for dealing with such problems. Following McQuistan and various coauthors,[4] recurrence relations among the numbers of possible arrangements of dimers on lattices of different sizes are obtained. These recurrence relations are linear and can be decoupled.[5] The decoupled equations lead to a polynomial $P(x, z)$, whose largest z-root as functions of x, the absolute activity of dimers, is simply related to the partition function of dimers. The proof of this was presented and preliminary results were obtained for $x = 1$ in paper I. Paper II extended these results for values of x in the range (0, 10) for planar rectangular lattices. Paper III presented a further extension, which enables us to obtain a closed form analytic solution of dimers on infinite two- and three-dimensional lattices.

Our objective in this paper is the study of dimers on plane triangular lattices using the same techniques developed in papers I and III. One major ingredient in this study is the bending by 30° of the rectangular lattice to form the plane trigonal lattice, increasing the coordination number from 4 to 6, as shown in Fig. 1. In this lattice, lattice sites are at the centers of rhombic boxes, the connections being made at 60° angles. A similar, but distinct, study has just been completed by Hock and McQuistan,[6] where lattice sites are at the center of hexagons. Although they use the terminology of "hexagonal lattice," we think it would be more appropriate to call it "triangular," since in such a case one generates a plane triangular lattice with different edges than the one we generate by bending the rectangular lattice. The advantage of our approach is that it can be extrapolated to obtain the thermodynamic quantities associated with dimers on an infinite plane triangular lattice.

Another advantage of bending the rectangular lattice by 30° is that it enables us to generate plane hexagonal or "honeycomb" and Kagomé lattices by a systematic way of avoiding certain lattice sites. A study of dimers on plane-honeycomb and Kagomé lattices will be presented in a separate article.[7]

The organization of the paper is as follows. Section II is a brief summary of how the partition function and all other quantities can be calculated exactly from the knowledge of the largest z-root of polynomial $P(x, z)$. Sections III, IV, and V briefly show how polynomials $P(x, z)$ are derived for infinite plane triangular lattices with a finite number of rows $M = 2$, 3, and 4, respectively. Section VI presents the numerical results and gives a one-parameter fit of the data. Extrapolation of the results for the infinite plane triangular lattice with an infinite number of rows is discussed and comparison is made with dimers on an infinite square lattice.

II. THERMODYNAMIC QUANTITIES IN TERMS OF THE LARGEST z-ROOT OF $P(x, z)$

As mentioned in the Introduction, we bend the rectangular lattice by 30° as shown in Fig. 1. The lattice sites are at the center of the rhombic boxes. "Nearest" neighbor occupation means that if one end of a dimer occupies one site (see Fig. 1) the other end of a dimer may occupy one of the six closest equidistant sites; the two neighboring sites that are

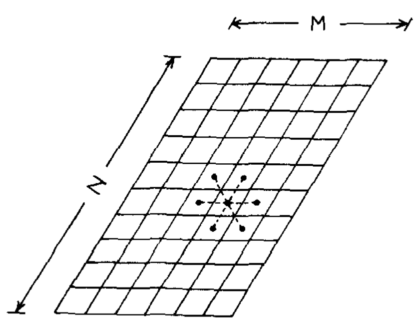

FIG. 1. Diagram of "bent" lattice showing the coordination number $Q = 6$ for a plane trigonal lattice.

 0022-2488/86/041099-11$02.50

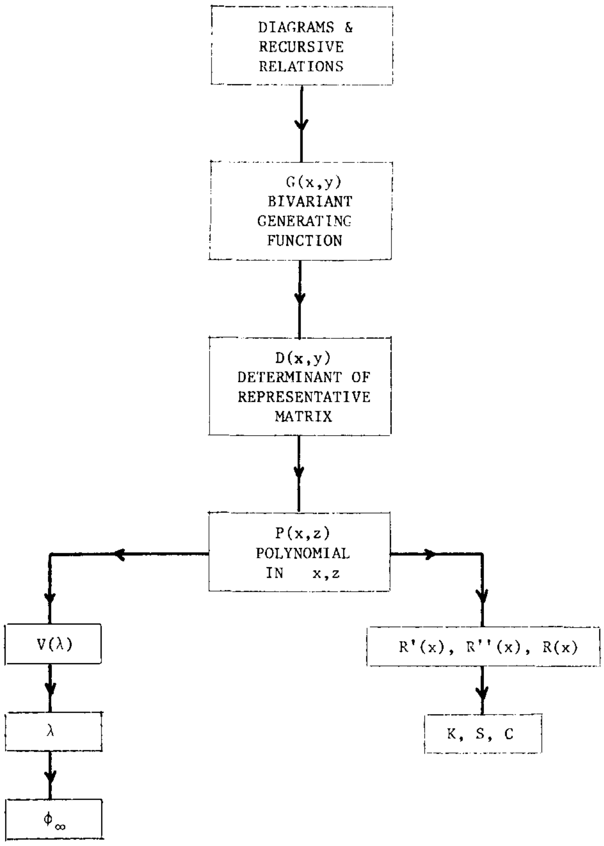

FIG. 2. Flow chart showing the general method of computation of the molecular freedom per dimer at close packing and other thermodynamic quantities.

farther away are automatically excluded. Therefore, the coordination number Q is 6 for the triangular lattice as compared to 4 for the plane rectangular lattice. Size M of the lattice is fixed at a finite value, and size N is allowed to become infinite. The case $M = 1$ is the one-dimensional problem previously solved.[1–4] We will study the triangular lattices for $N = 2$, 3, and 4.

For the reader's convenience, a flow chart summarizing the procedure followed in obtaining the various thermodynamic quantities is given in Fig. 2.

(1) We follow McQuistan and various coauthors[4] (also see paper I and Ref. 5), and identify various types of $M \times N$ lattices: A-type, B-type, C-type, etc. (Figs. 3–5), corresponding to lattices missing no cell in the Nth row, missing one cell, two cells, etc., respectively. Obviously, there is only one A-type lattice, but there may be several B-type, C-type, etc., depending on the size of M.

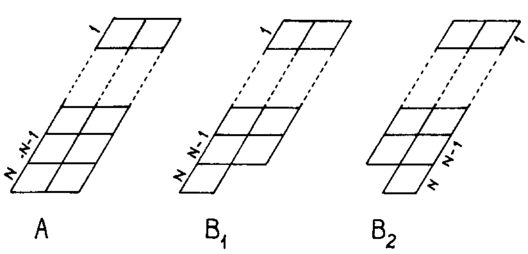

FIG. 3. Types of lattices for $M = 2$.

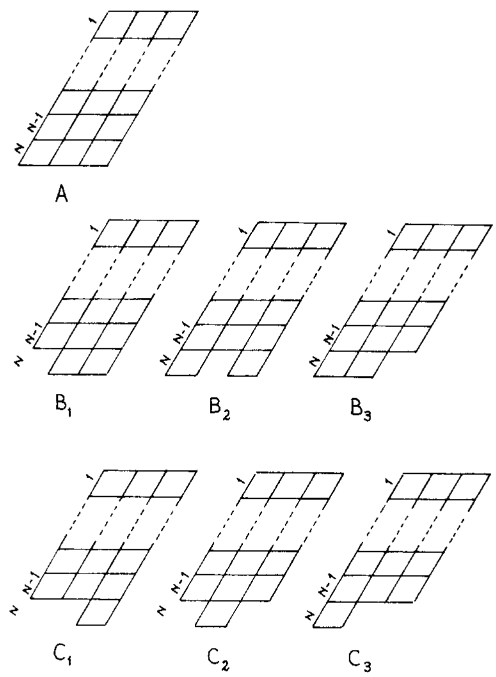

FIG. 4. Types of lattices for $M = 3$.

(2) We call $A(q,N)$, $B_i(q,N)$, $C_i(q,N)$, etc., the number of possible arrangements of q dimers on the A, B_i, C_i lattices, etc. Then coupled recursive relations among these arrangements are generated depending on the distinct number of ways of occupying cells in the Nth row of a $M \times N$ lattice of a given type. The diagrammatic method used is straightforward and offers no difficulties. However, the number of terms involved in this analysis increases very rapidly with the size M.

(3) Bivariant generating functions are introduced in the

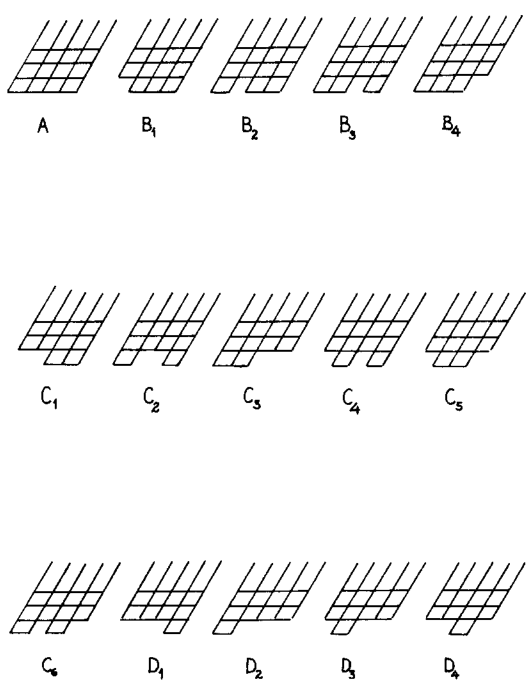

FIG. 5. Types of lattices for $M = 4$.

usual way for every T-type lattice (T stands for A, B, C, etc.)[5]:

$$G_T(x,y) = \sum_{N=0}^{\infty} \sum_{q}^{q_{max}} T(q,N) x^q y^N. \tag{2.1}$$

The variable x is later identified with the absolute activity of dimers. The system of coupled recursive relations is then replaced by a system of linear equations, where the G's play the role of the unknowns. As shown in paper I and Ref. 5, a term of the form $T(q-j, N-k)$ in any given recursive relation is to be identified with

$$T(q-j, N-k) \to x^j y^k G_T. \tag{2.2}$$

(4) It is then obvious that the solution of the system of linear equations with the G's playing the role of unknowns are obtained as the ratio of two polynomials in x and y[5]:

$$G_T(x,y) = N_T(x,y)/D(x,y). \tag{2.3}$$

The polynomial appearing in the denominator, $D(x,y)$, is the determinant of the matrix associated with the system of linear equations. Computing this determinant becomes increasingly difficult as the size M of the lattice increases.

(5) We have also shown in paper I that the partition function $Z(x)$ of the system of dimers with absolute activity x is given by

$$Z(x) = R(x)^{1/M}, \tag{2.4}$$

where $R(x)$ is the largest z-root of polynomial $D(x,y=1/z)$. From the knowledge of the partition function, all the other thermodynamic functions follow, namely, the grand potential $\Gamma(x)$, the number density $\rho(x)$, the entropy per unit volume $S_v(x)$, the isothermal compressibility per unit volume $K_T(x)$, and the constant specific heat capacity per unit volume $C_v(x)$:

$$\Gamma(x) = (2/Q)\ln(Z), \tag{2.5}$$

$$\rho(x) = x\frac{d\Gamma}{dx}, \tag{2.6}$$

$$S(x) = S_v(x)/k_B = -\rho\ln(x) + \Gamma(x), \tag{2.7}$$

$$K(x) = K_T(x) k_B T \rho^2 = x\frac{d\rho}{dx}, \tag{2.8}$$

$$C(x) = C_v(x)/k_B = K(x)(\Gamma/\rho)^2. \tag{2.9}$$

In these equations k_B is Boltzmann's constant and T is the absolute temperature. In paper II, we computed all these quantities numerically and obtained an approximate analytic fit for a system of dimers on an infinite square lattice for values of x in the range 0 to 10. Basically, $R(x)$ was obtained numerically as well as all derivatives involved. In paper III, we proposed an exact analytic solution of the problem of dimers on an infinite square lattice (two-dimensional problem), and on an infinite cubic lattice (three-dimensional problem), based on the fact that the roots $R(x,M)$ approach an exact exponential behavior with the size M of the lattice as M becomes increasingly large. In this article, we propose to limit the uncertainties involved in numerical evaluations of the thermodynamic quantities. The only quantity being evaluated numerically is the root $R(x)$. All the thermodynamic quantities are then computed from closed form analytic expressions involving the root R. This is a major improvement on previous approaches. Starting from the expression of $D(x,y)$, we replace y by $(1/z)$ and multiply the result by z raised to the highest power h of y occurring in $D(x,y)$:

$$P(x,z) = z^h D(x,1/z). \tag{2.10}$$

Clearly, the z-roots of polynomial $D(x,y=1/z)$ are the same as the z-roots of $P(x,z)$. We are only interested in the largest z-root of this polynomial and we will refer to it as $R(x)$. It then follows that the first and second derivatives of any z-root, and in particular the root $R(x)$, may be obtained directly from various partial derivatives of $P(x,z)$, namely,

$$R'(x) = \left(-\frac{(\partial P/\partial x)}{(\partial P/\partial z)}\right)_{z=R}, \tag{2.11}$$

$$R''(x) = \left(-\frac{\partial^2 P/\partial x^2}{(\partial P/\partial z)} + 2\frac{(\partial P/\partial x)(\partial^2 P/\partial x\,\partial z)}{(\partial P/\partial z)^2} - \frac{(\partial P/\partial x)^2(\partial^2 P/\partial z^2)}{(\partial P/\partial z)^3}\right)_{z=R}. \tag{2.12}$$

(6) The normalized number density θ and other thermodynamic quantities follow from the knowledge of these derivatives, namely,

$$\theta = Q\rho = (2x/M)(R'/R), \tag{2.13}$$

$$K = [(R'/R) + x(R''/R) - x(R'/R)^2](2x/QM), \tag{2.14}$$

$$S = -\rho\ln(x) + (2/QM)\ln(R), \tag{2.15}$$

$$C = (4K/Q^2M^2\rho^2)\ln^2(R). \tag{2.16}$$

(7) Finally, the molecular freedom per dimer as a function of the absolute activity x is given as (paper III)

$$\phi(x) = \exp[Q\,S(x)] = R(x)^{2/M}/x^{\theta}. \tag{2.17}$$

The molecular freedom per dimer at close packing[8,9] follows by taking the limit as x becomes infinite and the normalized number density θ approaches unity. This means that the larger root $R(x)$ in the large x-limit should behave like $(x\phi)$ raised to the power $(M/2)$. This is explained in paper III. A polynomial $V(\lambda)$ is obtained:

$$V(\lambda) = \lim_{x\to\infty} [P(x,\lambda x^{M/2})/x^a], \tag{2.18}$$

where a is the highest power of x in the polynomial $P(x,z)$. The largest root of $V(\lambda)$ gives the molecular freedom per dimer at close packing, namely,

$$\phi = \lambda^{2/M}. \tag{2.19}$$

III. PLANE-TRIANGULAR M = 2 LATTICE

Following the technique of paper I, three types of lattices are considered: the nontruncated A-type, and two B-type lattices B_1 and B_2 as shown in Fig. 3. As mentioned in Sec. II, $A(q,N)$, $B_1(q,N)$, and $B_2(q,N)$ are the numbers of arrangements of q dimers on the A, B_1, and B_2 lattices, respectively. Figures 6(a) and 6(b) show how recursive relations are generated among these arrangements. We introduce the associated generating functions $G_A(x,y)$, $G_{B_1}(x,y)$, and $G_{B_2}(x,y)$. According to Eq. (2.2), each diagram in a given set contributes to the linear system of equations involving the G's as un- knowns. We leave it to the reader to write down the system of linear equations; here, we simply give the determinant $D(x,y)$ for this system, namely,

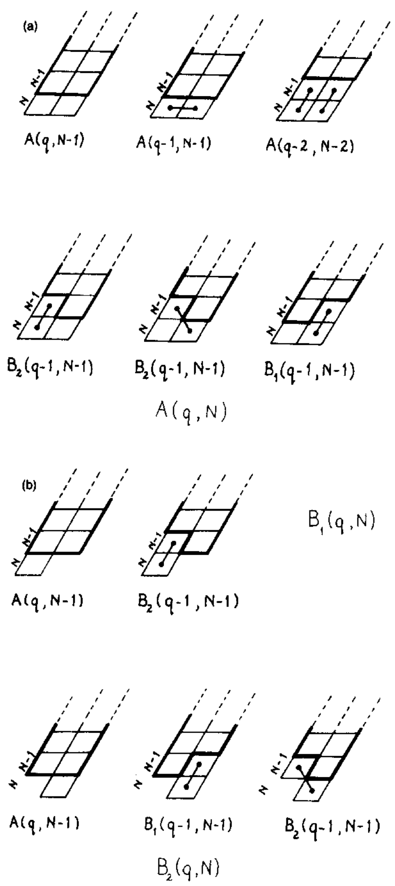

FIG. 6. (a) Diagrams of dimers on an $M = 2$ plane trigonal lattice and recursive relation for $A(q,N)$. (b) Diagrams of dimers on an $M = 2$ plane trigonal lattice and recursive relation for $B_1(q,N)$ and $B_2(q,N)$.

$$D(x,y) = 1 - y(1+2x) - xy^2(2+x) + x^2y^3(2x-1) + x^4y^4. \tag{3.1}$$

The polynomial $P(x,z)$ is obtained according to Eq. (2.10)

$$P(x,z) = z^4 - (1+2x)z^3 - x(2+x)z^2 + x^2(2x-1)z + x^4. \tag{3.2}$$

The molecular freedom per dimer at close packing is computed by generating the polynomial $V(\lambda)$ according to Eq. (2.18), in this case, one has

$$V(\lambda) = \lambda^4 - 2\lambda^3 - \lambda^2 + 2\lambda + 1 = (\lambda^2 - \lambda - 1)^2. \tag{3.3}$$

This polynomial has two double roots and the largest root is the one giving the molecular freedom per dimer at close packing, namely,

$$\phi_\infty = (1+\sqrt{5})/2. \tag{3.4}$$

Surprisingly, this number is the same as the one calculated for dimers on a plane rectangular lattice[8] with $M = 2$. Technically, the roots of $P(x,z)$ can be obtained analytically, since it is a quartic equation in z, and, therefore, one can claim that all thermodynamic quantities can be obtained in closed form for the $M = 2$ lattice.

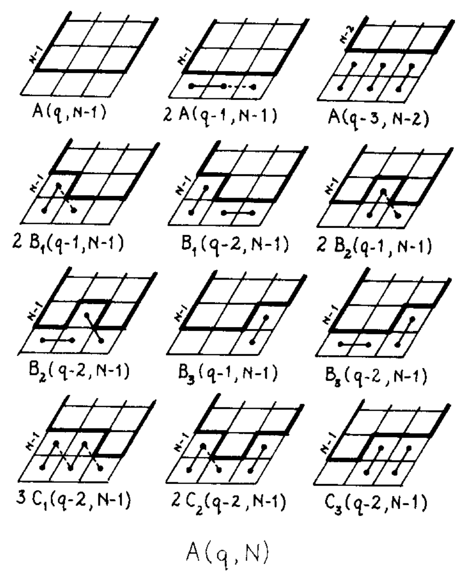

FIG. 7. Diagrams of dimers on an $M = 3$ plane trigonal lattice and recursive relations for $A(q,N)$.

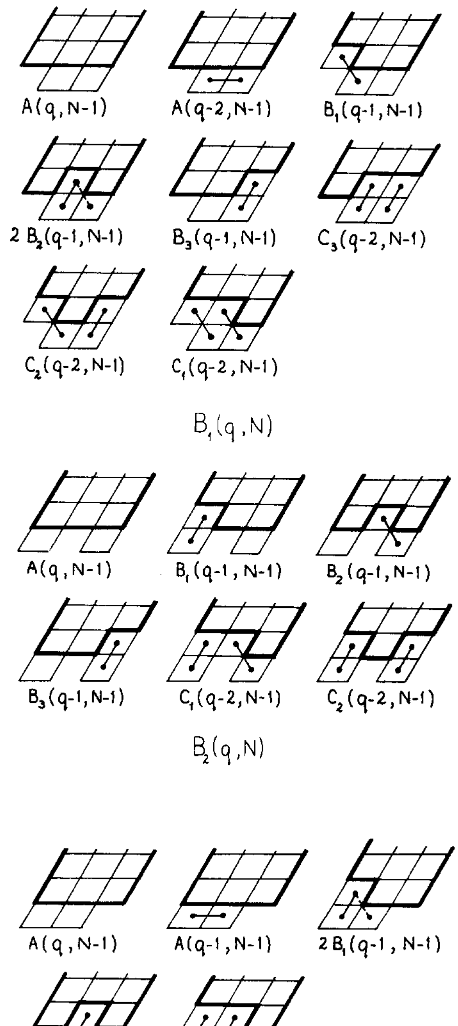

FIG. 8. Diagrams of dimers on an $M = 3$ plane trigonal lattice and recursive relations for $B_i(q,N)$, $i = 1$, 2, and 3.

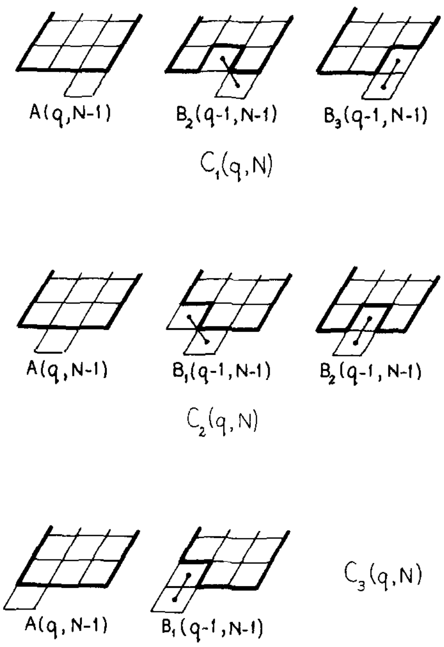

FIG. 9. Diagrams of dimers on an $M=3$ plane trigonal lattice and recursive relations for $C_i(q,N)$, $i=1$, 2, and 3.

IV. PLANE-TRIANGULAR M = 3 LATTICE

In addition to the nontruncated A-type lattice, there are three B-type and three C-type lattices (Fig. 4). Recursive relations are developed in the usual way, from Figs. 7–9, appropriate bivariant generating functions are introduced, and polynomial $P(x,z)$ is found to be

$$\begin{aligned}P(x,z) = {} & z^8 + (-4x-1)z^7 + (-8x^3-12x^2-3x)z^6 \\ & + (-4x^4-7x^3-3x^2)z^5 \\ & + (14x^6+4x^5+3x^4-x^3)z^4 \\ & + (12x^7-x^6+3x^5)z^3 \\ & + (-8x^9+4x^8-3x^7)z^2 \\ & + (x^9-4x^{10})z + x^{12},\end{aligned} \tag{4.1}$$

and the polynomial $V(\lambda)$ follows:

$$\begin{aligned}V(\lambda) &= \lambda^8 - 8\lambda^6 + 14\lambda^4 - 8\lambda^2 + 1 \\ &= (\lambda-1)^2(\lambda+1)^2 \\ &\quad \times(\lambda^2-2\lambda-1)(\lambda^2+2\lambda-1).\end{aligned} \tag{4.2}$$

The largest root is $(1+\sqrt{2})$, from which, according to Eq. (2.19), the molecular freedom per dimer at close packing is calculated to be

$$\phi_\infty = (1+\sqrt{2})^{2/3} = 1.799\,632\,345\,151\,997\ldots. \tag{4.3}$$

This is to be compared with the molecular freedom per dimer at close packing for an $M=3$ plane rectangular lattice $(2+\sqrt{3})^{1/3} = 1.551\,133\ldots$.

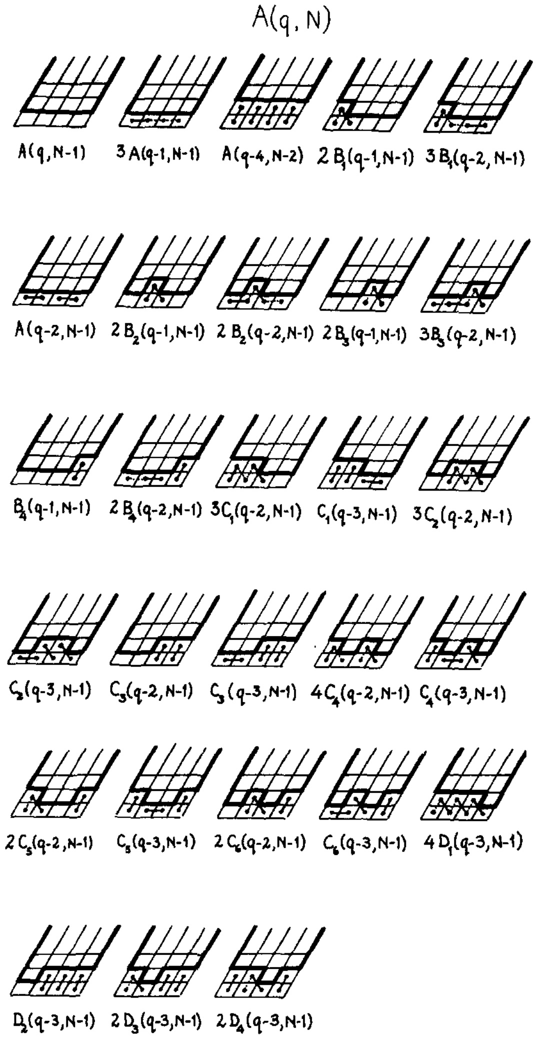

FIG. 10. Diagrams of dimers on an $M=4$ plane trigonal lattice and recursive relation for $A(q,N)$.

V. PLANE-TRIANGULAR M = 4 LATTICE

As usual, one has the nontruncated A-type lattice, and in addition four B-type, six C-type, and four D-type lattices, as shown in Fig. 5. Recursive relations are developed from Figs. 10–13. This leads to a set of 15 bivariant generating functions playing the role of 15 unknowns in a set of 15 linear equations. The polynomial $P(x,z)$ is then found to be

$$\begin{aligned}P(x,z) = {} & z^{16} + (-4x^2-6x-1)z^{15} + (-14x^4-60x^3-33x^2-4x)z^{14} + (44x^6-6x^5-109x^4-50x^3-6x^2)z^{13} \\ & + (107x^8+372x^7+210x^6-12x^5-24x^4-4x^3)z^{12} \\ & + (-148x^{10}+250x^9+471x^8+322x^7+81x^6+6x^5-x^4)z^{11} \\ & + (-396x^{12}-752x^{11}-382x^{10}-152x^9+35x^8+26x^7+6x^6)z^{10}\end{aligned}$$

$$
\begin{aligned}
&+ (128x^{14} - 664x^{13} - 724x^{12} - 434x^{11} - 276x^{10} - 82x^{9} - 15x^{8})z^{9}\\
&+ (645x^{16} + 592x^{15} + 460x^{14} + 280x^{13} + 268x^{12} + 84x^{11} + 20x^{10})z^{8}\\
&+ (128x^{18} + 488x^{17} + 388x^{16} + 262x^{15} - 52x^{14} - 22x^{13} - 15x^{12})z^{7}\\
&+ (- 396x^{20} - 368x^{19} - 286x^{18} - 288x^{17} - 29x^{16} - 22x^{15} + 6x^{14})z^{6}\\
&+ (- 148x^{22} - 122x^{21} - 25x^{20} + 94x^{19} - 15x^{18} + 18x^{17} - x^{16})z^{5}\\
&+ (107x^{24} + 164x^{23} + 82x^{22} - 36x^{21} + 24x^{20} - 4x^{19})z^{4}\\
&+ (44x^{26} - 10x^{25} - 53x^{24} + 22x^{23} - 6x^{22})z^{3}\\
&+ (- 14x^{28} - 28x^{27} + 15x^{26} - 4x^{25})z^{2} + (- 4x^{30} + 6x^{29} - x^{28})z + x^{32}. \qquad (5.1)
\end{aligned}
$$

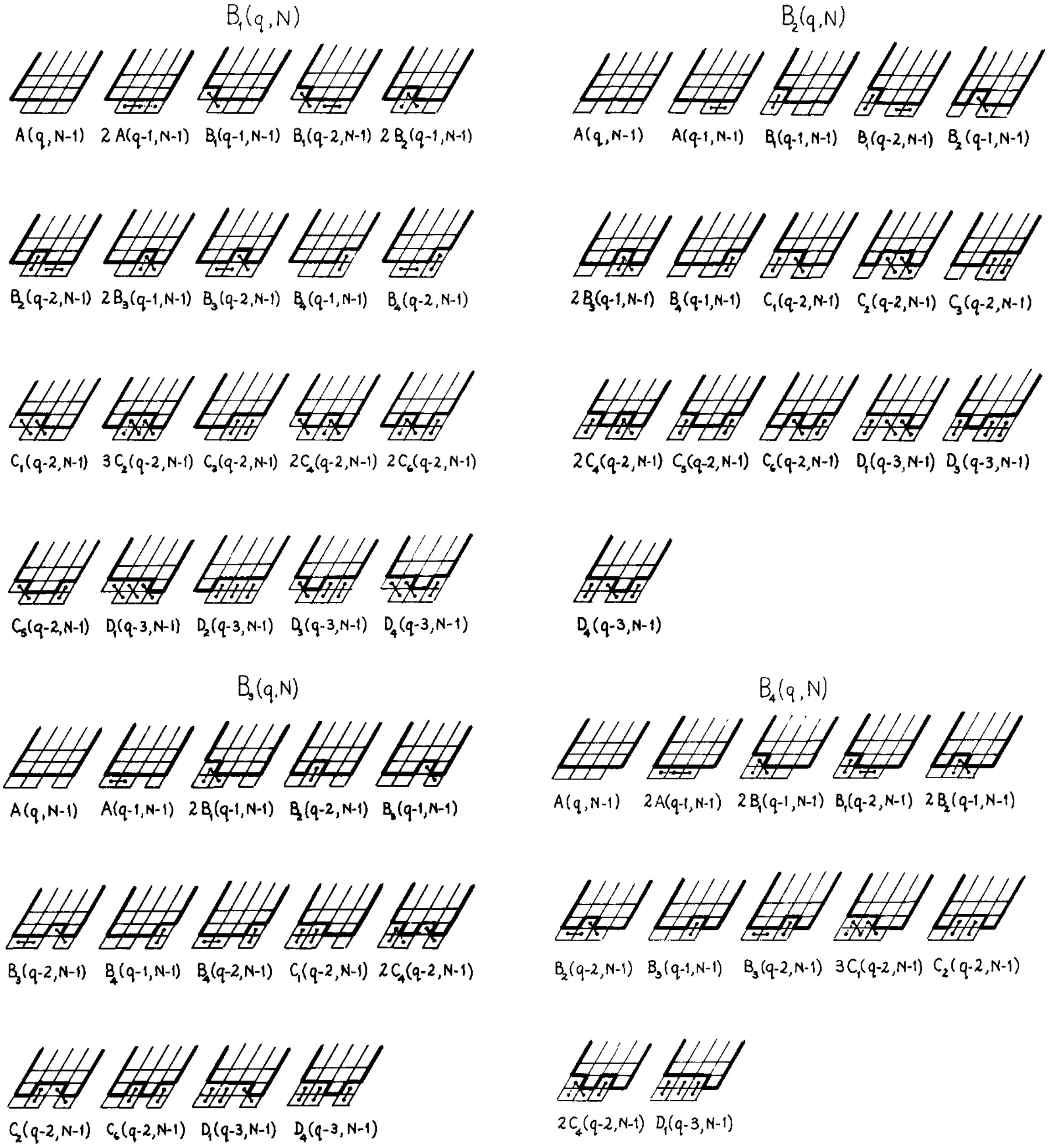

FIG. 11. Diagrams of dimers on an $M = 4$ plane trigonal lattice and recursive relations for $B_i(q,N)$, $i = 1$–4.

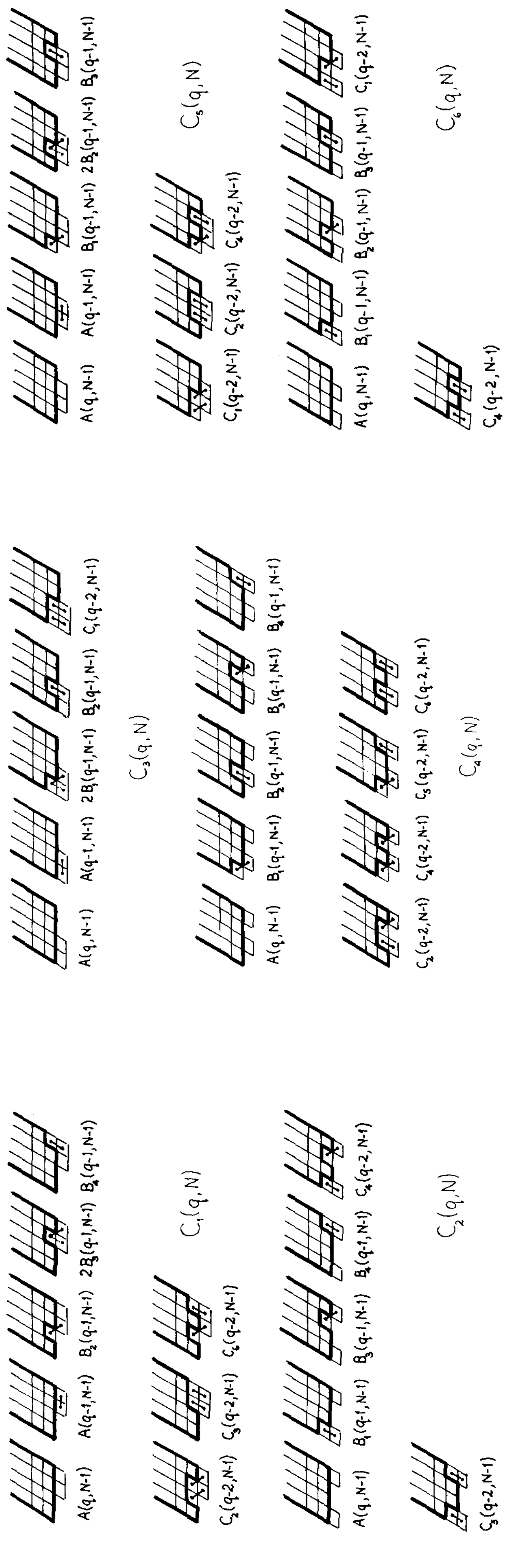

FIG. 12. Diagrams of dimers on an $M = 4$ plane trigonal lattice and recursive relations for $C_i(q,N)$, $i = 1$–6.

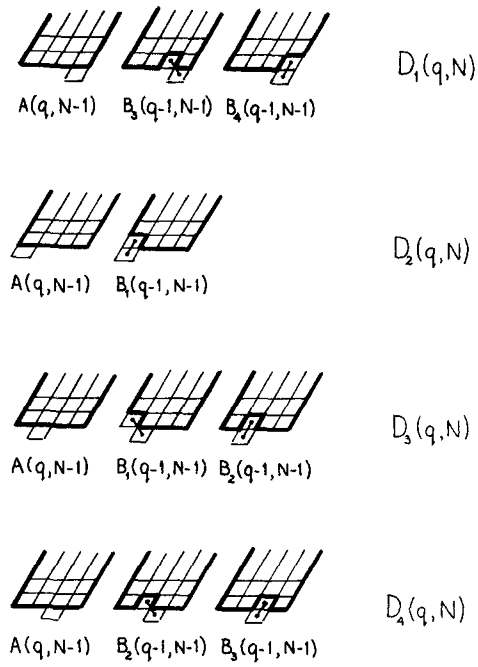

FIG. 13. Diagrams of dimers on an $M = 4$ plane trigonal lattice and recursive relations for $D_i(q,N)$ for $i = 1$–4.

We were able to factor the polynomial $V(\lambda)$ and found

$$V(\lambda) = (\lambda^4 - 3\lambda^3 - 3\lambda^2 + \lambda + 1)^2 \times (\lambda^4 + \lambda^3 - 3\lambda^2 - 3\lambda + 1)^2. \quad (5.2)$$

The root of largest modulus is found, in the first factor of this equation, to be 3.715 495 169 327 638... . According to Eq. (2.19), the actual molecular freedom per dimer at close packing for this lattice is the square root of this number

$$\phi_\infty = (3.715\,495...)^{2/M} = 1.927\,561\,975\,482\,925... . \quad (5.3)$$

Technically, again, this molecular freedom at close packing can be written down analytically in closed form, since it is the square root of the solution of a quartic equation as exhibited in Eq. (5.2). This is to be compared with the molecular freedom per dimer at close packing for a plane rectangular lattice with $M = 4$, which is[3,8] 1.685 389... . As speculated by many authors before, the molecular freedom should normally increase with the coordination number Q for the lattice. This is certainly the case when comparing results for dimers on square and plane triangular lattices. The coordination number has increased from 4 to 6, and so did the molecular freedom from 1.685... to 1.927... . One thing that could be said is that the increase in molecular freedom is not in proportion to the increase in coordination number.

VI. THERMODYNAMIC QUANTITIES AND ANALYTIC FIT OF THE DATA

We first analyze our exact analytic results for the molecular freedom per dimer at close packing obtained for sizes $M = 2$, $M = 3$, and $M = 4$ as given in Eqs. (3.4), (4.3), and (5.3), respectively. In the case of a square lattice filled with dimers, Kasteleyn[8] was able to obtain, analytically and in

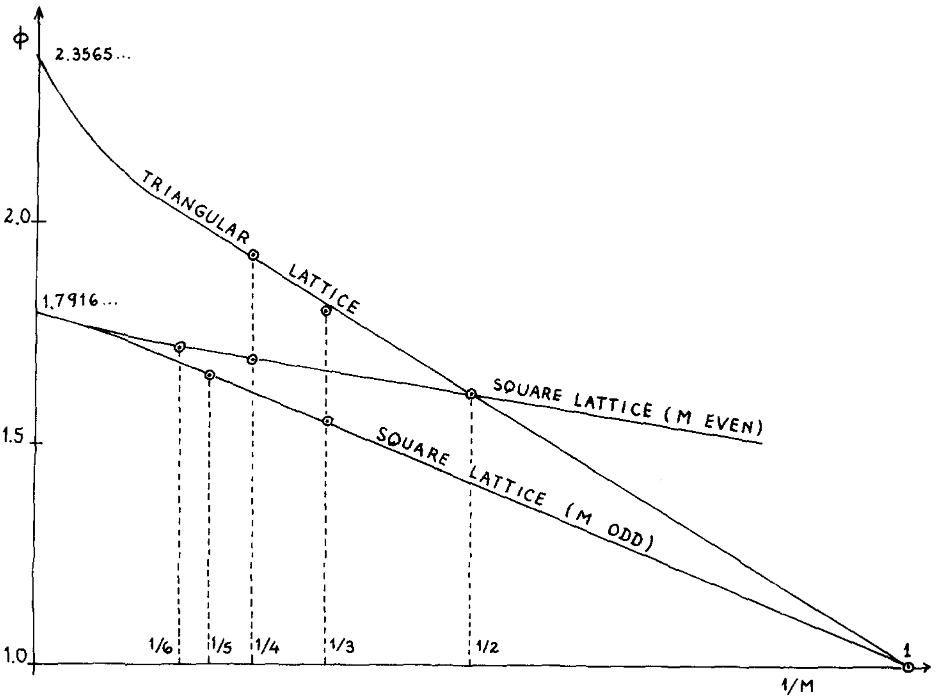

FIG. 14. Plots of the molecular freedom per dimer at close packing for square lattices (even and odd values of M) and triangular lattices ($M = 1, 2, 3, 4$) versus the reciprocal of M.

closed form, the molecular freedom for any sizes M and N of the two-dimensional lattice. In this case, lattices finite in one direction (M) and infinite in the other exhibit molecular freedoms that can be grouped in two sets, those corresponding to even values of M and those corresponding to the odd values of M. In Fig. 14, we plotted the molecular freedom at close packing versus $(1/M)$ for both square and triangular lattices. For square lattices, we observe the odd and even series of data to follow almost straight-line trajectories, then bend without crossing beyond $M = 10$, and intercept the ordinate axis at the value obtained by Kasteleyn,[8] namely, 1.7916..., which is the molecular freedom at close packing for the square lattice infinite in both directions. For triangular lattices, the situation is different. The data shown in Fig. 14 follow an almost straight-line trajectory, which bends upward beyond $M = 10$ and intercepts the ordinate axis at the value given by Nagle[9] with five significant figures to be 2.3565. We believe that this number was obtained by using Kasteleyn's theorem[10] on dimers fully covering planar lattices combined with the technique of Stephenson.[11] For completeness, we give the answer in the form of a double integral[10,11]

TABLE I. Thermodynamic properties of dimers on a plane trigonal lattice for $M = 2$ for selected values of the absolute activity x of dimers.

X	θ	S	K	C
0.1	0.242 670	0.144 233	0.025 436	0.040 611
0.2	0.357 400	0.181 489	0.029 105	0.060 129
0.3	0.428 850	0.198 237	0.029 413	0.072 453
0.4	0.479 240	0.207 143	0.028 875	0.081 254
0.5	0.517 420	0.212 266	0.028 104	0.087 878
0.6	0.547 730	0.215 309	0.027 288	0.093 166
0.7	0.572 606	0.217 109	0.026 494	0.097 492
0.8	0.593 535	0.218 122	0.025 743	0.101 118
0.9	0.611 482	0.218 614	0.025 043	0.104 192
1.0	0.627 109	0.218 752	0.024 392	0.106 850
2.0	0.719 179	0.213 616	0.019 846	0.121 602
3.0	0.764 241	0.206 924	0.017 219	0.127 693
4.0	0.792 426	0.201 100	0.015 455	0.130 780
5.0	0.812 240	0.196 157	0.014 161	0.132 463
6.0	0.827 180	0.191 927	0.013 157	0.133 383
7.0	0.838 974	0.188 250	0.012 349	0.133 849
8.0	0.848 598	0.185 028	0.011 678	0.134 027
9.0	0.856 648	0.182 159	0.011 110	0.134 015
10.0	0.863 520	0.179 580	0.010 620	0.133 875

TABLE II. Thermodynamic properties of dimers on a plane trigonal lattice for $M = 3$ for selected values of the absolute activity x of dimers.

X	θ	S	K	C
0.1	0.261 999	0.157 517	0.025 939	0.044 1537
0.2	0.377 283	0.194 981	0.028 890	0.064 258
0.3	0.447 831	0.211 521	0.028 905	0.076 795
0.4	0.497 218	0.220 252	0.028 235	0.085 632
0.5	0.534 487	0.225 254	0.027 396	0.092 296
0.6	0.564 003	0.228 217	0.026 544	0.097 547
0.7	0.588 183	0.229 967	0.025 730	0.101 811
0.8	0.608 496	0.230 949	0.024 969	0.105 356
0.9	0.625 895	0.231 427	0.024 270	0.108 354
1.0	0.641 029	0.231 560	0.023 614	0.110 926
2.0	0.729 935	0.226 605	0.019 117	0.124 875
3.0	0.773 291	0.220 167	0.016 549	0.130 385
4.0	0.800 361	0.214 573	0.014 834	0.133 045
5.0	0.819 375	0.209 832	0.013 581	0.134 414
6.0	0.833 698	0.205 775	0.012 612	0.135 080
7.0	0.845 001	0.202 256	0.011 834	0.135 355
8.0	0.854 222	0.199 164	0.011 188	0.135 365
9.0	0.861 935	0.196 416	0.010 642	0.135 218
10.0	0.868 513	0.193 949	0.010 173	0.134 969

TABLE III. Thermodynamic properties of dimers on a plane trigonal lattice for $M = 4$ for selected values of the absolute activity x of dimers.

X	θ	S	K	C
0.1	0.272 670	0.164 646	0.026 402	0.046 029
0.2	0.389 196	0.202 529	0.029 013	0.066 401
0.3	0.459 790	0.219 083	0.028 825	0.078 945
0.4	0.508 935	0.227 770	0.028 029	0.987 713
0.5	0.545 874	0.232 729	0.027 112	0.094 291
0.6	0.575 050	0.235 658	0.026 204	0.099 437
0.7	0.598 896	0.237 385	0.025 354	0.103 614
0.8	0.618 896	0.238 352	0.024 567	0.107 065
0.9	0.636 002	0.238 821	0.023 845	0.109 984
1.0	0.650 863	0.238 952	0.023 182	0.112 484
2.0	0.737 844	0.234 111	0.018 714	0.126 207
3.0	0.780 074	0.227 842	0.016 260	0.132 174
4.0	0.806 386	0.222 407	0.014 658	0.135 565
5.0	0.824 847	0.217 806	0.013 508	0.137 783
6.1	0.840 043	0.213 483	0.012 163	0.135 120
6.5	0.844 617	0.212 081	0.011 846	0.135 208
7.0	0.849 804	0.210 430	0.011 485	0.135 253
7.2	0.851 734	0.209 800	0.011 350	0.135 255
9.0	0.866 235	0.204 764	0.010 324	0.134 967

$$\Phi\begin{pmatrix}\text{triangular}\\ \text{lattice}\end{pmatrix} = \exp\left[\frac{1}{32\,\pi^2}\int\!\!\int_{-\pi}^{+\pi} d\theta\, d\phi \ln A(\theta,\phi)\right], \tag{6.1}$$

where the quantity $A(\theta,\phi)$ is given by

$$\begin{aligned} A(\theta,\phi) = 962 &- 68\cos\theta + 2\cos 2\theta - 32\cos 2\phi \\ &+ 192\cos(\theta + 2\phi) - 32\cos(2\theta + 2\phi) \\ &- 2\cos(\theta + 4\phi) \\ &+ 4\cos(2\theta + 4\phi) - 2\cos(3\theta + 4\phi). \end{aligned} \tag{6.2}$$

The double integration done numerically gives the value of the molecular freedom for a triangular lattice infinite in both directions and fully covered with dimers to be 2.356 527... .

Other thermodynamic quantities are calculated for each lattice size M of the triangular lattice using Eqs. (2.13)–(2.16). In these equations, the value of R is computed numerically for a given value of the absolute activity x, using polynomials $P(x,z)$ given in Eqs. (3.2), (4.1), and (5.1), for $M = 2$, 3, and 4, respectively. Also, from the knowledge of R for a given value of x, we compute the derivatives R' and R'' according to Eqs. (2.11) and (2.12). In the latter equations, the partial derivatives of polynomials $P(x,z)$ are needed, namely, $\partial P/\partial x$, $\partial P/\partial z$, $\partial^2 P/\partial x\,\partial z$, $\partial^2 P/\partial x^2$, and $\partial^2 P/\partial z^2$. These derivatives are evaluated at the chosen value of x and for $z = R$ (R computed numerically for the same value of x). The results are listed in Tables I, II, and III, for $M = 2$, 3, and 4, respectively.

We have plotted the quantities $S(\theta)$, $K(\theta)$, and $C(\theta)$ versus the normalized number density θ, for lattice sizes $M = 2$, 3, and 4. Figures 15–17 show these curves. As in the case of plane-rectangular lattices (paper II), we observe that the roots $R(x;M)$ increase exponentially with the value of M considered. It is reasonable to expect that this is the case in general, and under this assumption, we have shown in paper III that one can find closed form analytic expressions for S, C, K, and x in terms of ϕ, the molecular freedom at close packing, θ the normalized number density, and the coordination number Q, namely,

$$x(\theta) = \theta(2-\theta)/4\phi(L,M)(1-\theta)^2, \tag{6.3}$$

$$S(\theta) = -\frac{\theta}{Q}\ln\left(\frac{\theta(2-\theta)}{4\phi(L,M)(1-\theta)^2}\right) + \frac{2}{Q}\ln\left(\frac{2-\theta}{2-2\theta}\right), \tag{6.4}$$

$$K(\theta) = \theta(2-\theta)(1-\theta)/2Q, \tag{6.5}$$

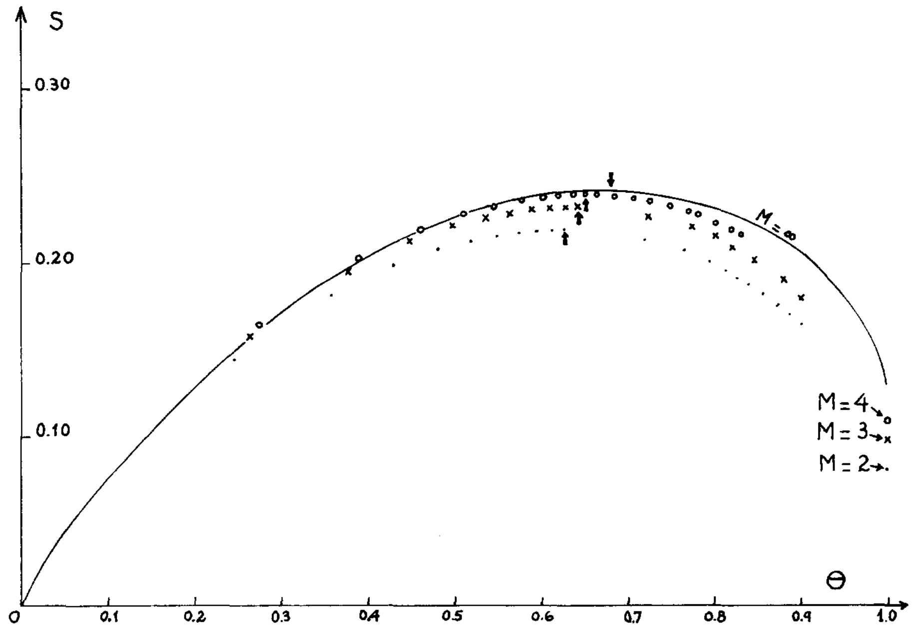

FIG. 15. Plot of S versus the normalized number density θ for $M = 2$, 3, 4, and the analytic curve for $M = \infty$.

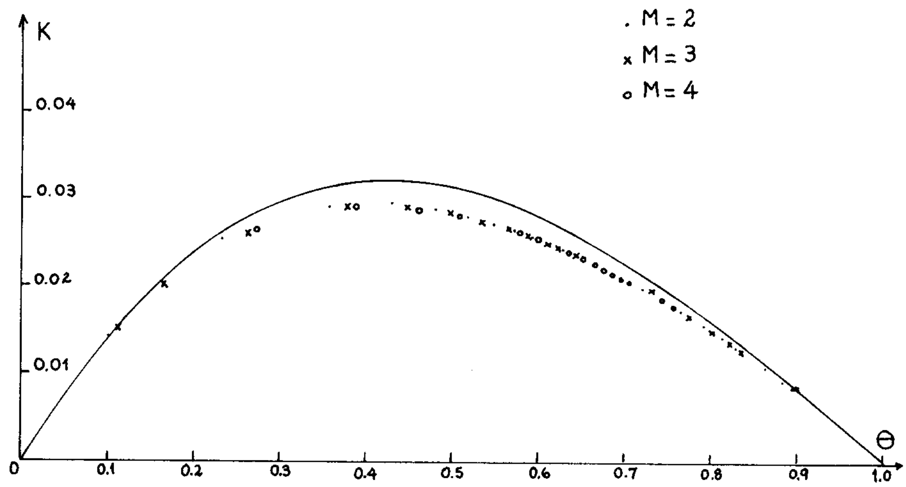

FIG. 16. Plot of K vs θ for $M = 2, 3, 4$, and the analytic curve for $M = \infty$.

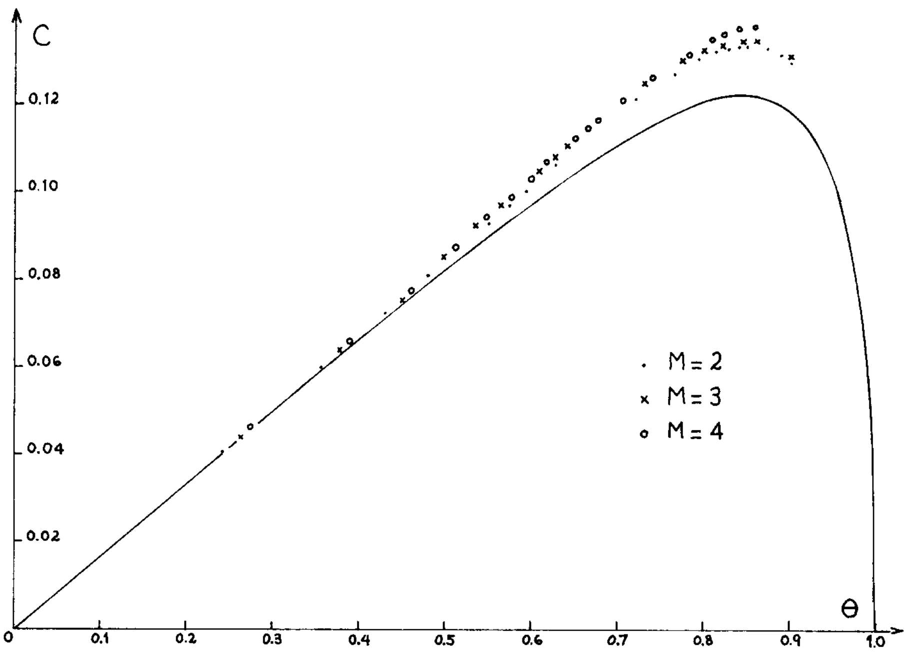

FIG. 17. Plot of C vs θ for $M = 2, 3, 4$, and the analytic curve for $M = \infty$.

$$C(\theta) = \frac{2(1-\theta)(2-\theta)}{\theta Q} \ln^2\!\left(\frac{2-\theta}{2-2\theta}\right). \tag{6.6}$$

Here, $\phi(L,M)$ is the molecular freedom at close packing, where we have used the notation of paper III, $L = 1$ corresponds to two-dimensional lattices, and $L > 1$ corresponds to several two-dimensional layers (L), i.e., three-dimensional lattices. In this article, we only studied two-dimensional lattices ($L = 1$). More layers would lead to three-dimensional trigonal lattices (rhombohedral). Equations (6.3)–(6.6) are expressions that are expected to be exact in the limit of infinite two-dimensional ($L = 1, M = \infty$) and three-dimensional ($L = \infty, M = \infty$) lattice spaces. The dependence on the lattice size is hidden in the expression of $\phi(L,M)$, thus showing that only x and S depend on the lattice size. Figures 15–17 are the plot of S, K, and C vs θ for two-dimensional lattices with $M = 2$, 3, and 4, respectively. The entropy plot $S(\theta)$ in Fig. 15 shows that the curves obtained numerically, following the flow chart of Fig. 2 approach rather rapidly the limiting curve at $M = \infty$. This is not the case for the curves $K(\theta)$ and $C(\theta)$ of Figs. (16) and (17), respectively. There, we observe some oscillations and the data obtained numerically for values of θ beyond 0.8 can only be trusted to within the numerical capabilities of the Commodore VIC-20 used.

[1]A. J. Phares, J. Math. Phys. 25, 1756 (1984).
[2]A. J. Phares, D. E. Shaw, and F. J. Wunderlich, J. Math. Phys. 26, 1762 (1985).
[3]A. J. Phares and F. J. Wunderlich, J. Math. Phys. 26, 2491 (1985).
[4]D. Lichtman and R. B. McQuistan, J. Math. Phys. 8, 2441 (1967); R. B. McQuistan, Nuovo Cimento B 58, 86 (1968); R. B. McQuistan and S. J. Lichtman, J. Math. Phys. 11, 3095 (1970); R. B. McQuistan and J. L. Hock, Fibonacci Quart. 21, 196 (1983).
[5]A. J. Phares, J. Math. Phys. 25, 2169 (1984).
[6]J. L. Hock and R. B. McQuistan, J. Chem. Phys. 83, 3626 (1985).
[7]A. J. Phares and F. J. Wunderlich, "Thermodynamics of Dimers on Plane Honeycomb and Kagome Lattices," to be submitted for publication.
[8]P. W. Kasteleyn, Physica 27, 1209 (1961).
[9]J. F. Nagle, Phys. Rev. 152, 190 (1966).
[10]P. W. Kasteleyn, J. Math. Phys. 4, 287 (1963).
[11]J. Stephenson, J. Math. Phys. 5, 1009 (1964).

Solution of a generalized Chandrasekhar H-equation

B. L. Willis
Center for Transport Theory and Mathematical Physics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

(Received 8 August 1985; accepted for publication 16 October 1985)

The Chandrasekhar H-equations are generalized to problems relevant to multigroup transport equations that have nondiagonal cross-section matrices. These equations are shown to have a unique solution in a ball of a Banach space, which satisfies the necessary analyticity properties.

I. INTRODUCTION

In Ref. 1, Willis and van der Mee have shown that multigroup transport equations with nondiagonal cross-section matrices arise when the method of weighted residuals is applied to problems in energy-dependent transport theory. In Ref. 1, we were able to completely solve the two-group, half space problem if the scattering matrix was noninvertible and the cross-section matrix nondiagonalizable. In this paper, we will study the same problem, but with an invertible scattering matrix. The formulas given in Ref. 1 for the exit distribution and solution still remain valid for this case, but the formulas for the Wiener–Hopf factorization are not. If the scattering matrix C is invertible, then we are unable to explicitly construct the factorization. For this case we derive equations that the factors must satisfy, and then develop a numerical method for their solution, which automatically satisfies the necessary analyticity properties.

II. DERIVATION OF THE H-EQUATIONS

We will study a two-group transport equation with a nondiagonalizable cross-section matrix and half space boundary conditions. In Ref. 1 we showed that this transport equation is equivalent to a Wiener–Hopf integral equation, and that the symbol of the integral equation (the identity matrix minus the Fourier transform of the kernel) is given by

$$W(\lambda) = I - (\lambda^{-1}\tan^{-1}\lambda)C + (1+\lambda^2)^{-1}MC, \tag{1}$$

where I is the 2×2 identity matrix, and M is the nilpotent part of the cross-section matrix. Without loss of generality we may assume that

$$M = \begin{vmatrix} 0 & \alpha \\ 0 & 0 \end{vmatrix}. \tag{2}$$

The crucial step in the solution to any Wiener–Hopf (WH) equation is the construction of the WH factorization of the symbol W. Mullikian[2] has shown that a sufficient condition for the existence of a WH factorization is that the spectral radius of the matrix B, defined by

$$B_{ij} = \int_{-\infty}^{\infty} |K_{ij}(x)|\,dx, \tag{3}$$

is less than 1, where K is the kernel of the integral operator. It is easy to check that the spectral radius of B will be less than 1 if

$$\|C\| + \|MC\| < 1, \tag{4}$$

where $\|\ \|$ is the matrix norm. Furthermore, Mullikian[2] has derived nonlinear integral equations, which the factors of the symbol satisfy. They are, for $\operatorname{Im} z > 0$,

$$H_r^{-1}(z) = I + \frac{1}{2\pi i}\int_{-\infty}^{\infty} \frac{H_l(t)\hat{K}(-t)}{t+z}\,dt, \tag{5a}$$

$$H_l^{-1}(z) = I + \frac{1}{2\pi i}\int_{-\infty}^{\infty} \frac{\hat{K}(t)H_r(t)}{t+z}\,dt, \tag{5b}$$

where $\hat{K}$ is the Fourier transform of the kernel of the Wiener–Hopf equation, and I is the identity matrix. The matrix valued functions H_l and H_r have the following properties: (i) H_r and H_l are analytic in the half plane $\operatorname{Im} z > 0$, and continuous for $\operatorname{Im} z \geqslant 0$; (ii) H_r and H_l are invertible in the half plane $\operatorname{Im} z \geqslant 0$; and (iii) $I - \hat{K}(z) = H_l^{-1}(z)H_r^{-1}(z)$.

We now specialize Eqs. (5a) and (5b) to the two-group problem given in Ref. 1 with $\det C \neq 0$ by introducing the explicit form of $\hat{K}$;

$$\hat{K}(z) = ((1/z)\tan^{-1} z)C - (1+z^2)^{-1}MC, \tag{6}$$

where M is the nilpotent matrix given by Eq. (2). Recall that $\hat{K}$ is analytic on $\mathbb{C}\setminus\{z\in\mathbb{C}: z = it, |t| \geqslant 1\}$. The integrals appearing in the H-equations [Eqs. (5a) and (5b)] can be rewritten into a more familiar form by making use of Cauchy's theorem. The calculation for Eq. (5b) will be shown; the procedure for Eq. (5a) is essentially the same. Substitute Eq. (1) into Eq. (5b); the result is

$$H_l^{-1}(z) = I + \frac{1}{2\pi i}\int_{-\infty}^{\infty}\left[\frac{1}{t}(\tan^{-1} t)C - (1+t^2)^{-1}MC\right] \times H_r(t)/(t+z)\,dt. \tag{7}$$

We analyze each term of the integrand separately. First, consider the term $(1+t^2)^{-1}MC$. Note that $H_r(t)/(t+z)$ is analytic in the upper half plane, and $H_r(t)/(t+z)(1+t^2)$ vanishes at infinity as t^{-2}. Therefore, Cauchy's theorem yields

$$\int_{-\infty}^{\infty} \frac{H_r(t)}{(1+t^2)(t+z)} = \pi\frac{H_r(i)}{i+z}. \tag{8a}$$

Next, consider the term containing $t^{-1}\tan^{-1} t$. The contour can be completed in the upper half plane if the branch cut $[i, i\infty)$ is avoided. The result is

$$\int_{-\infty}^{\infty} \frac{\tan^{-1}(t)H_r(t)}{t(t+z)}\,dt = \int_{i\infty}^{i} \frac{1}{t}\,[\tan^{-1}(it+0) - \tan^{-1}(it-0)] \times [H_r(t)/(t+z)]\,dt, \tag{8b}$$

 0022-2488/86/041110-03$02.50

where $\tan^{-1}(it \pm 0)$ are the boundary values of $\tan^{-1}$ along the imaginary axis. Using

$$\tan^{-1}(it+0) - \tan^{-1}(it-0) = \pi, \quad |t| \geqslant 1,$$

the right-hand side of Eq. (8b) can be rewritten as

$$\int_{i\infty}^{i} \frac{\pi}{it} \frac{H_r(t)}{t+z} dt. \tag{9}$$

A simple change of variables gives

$$\int_{i\infty}^{i} \frac{\tan^{-1}(t)}{t} \frac{H_r(t)}{t+z} = \int_0^1 \frac{iH_r(i/t)}{t(z+i/t)} dt. \tag{10}$$

It is convenient to define functions X and Y by

$$X(z) = H_l(i/z) \tag{11a}$$

and

$$Y(z) = H_r(i/z). \tag{11b}$$

The final form of the H-equations is derived by substituting Eqs. (8) and (10) into Eq. (7). The result is

$$X^{-1}(z) = I + \frac{1}{2} C \int_0^1 d\mu \frac{z}{z+\mu} Y(\mu) + \frac{1}{2} MC \frac{z}{z+1} Y(1). \tag{12a}$$

Similarly, the H-equation corresponding to Eq. (5a) is

$$Y^{-1}(z) = I + \frac{1}{2} \int_0^1 d\mu \frac{z}{z+\mu} X(\mu) C + \frac{1}{2} \frac{z}{z+1} X(1) MC. \tag{12b}$$

Note that if $M=0$, then Eqs. (12a) and (12b) reduce to the Chandrasekhar H-equations. It is straightforward to show that every solution of Eqs. (12a) and (12b) provides a factorization of the symbol. A Banach space analysis of these equations is the next topic.

III. A BANACH SPACE ANALYSIS OF THE *H*-EQUATIONS

In Ref. 3, Bowden and Zweifel presented a Banach space analysis of the Chandreskhar H-equations. They showed that the H-equations could be solved by iteration. Of course the H-equations must be supplemented by constraints, which are the analyticity requirements given by conditions (i) and (ii) of Sec. II. In Ref. 4, Bowden, Menikoff, and Zweifel generalized their results to H-equations relevant to multigroup problems. In this section a similar analysis is given for Eqs. (12a) and (12b). In Refs. 3 and 4, it was shown that the H-equations have solutions in the Banach space $L_1(0,1)$ by making use of a contracting mapping principle. For the H-equations [Eqs. (12a) and (12b)] the terms $X(1)$ and $Y(1)$ prevent one from using a contracting principle in $L_1(0,1)$ because point evaluation is an unbounded operator in $L_1(0,1)$. Thus, instead of $L_1(0,1)$, we must use a Banach space with a supremum norm. Consequently the following analysis parallels the one given by Rall.[5]

First, we define some Banach spaces: Let X_0 be the vector space $X_0 = \{T\colon T \text{ is a } 2\times 2 \text{ matrix valued function, with } T_{ij} \text{ continuous on } [0,1]\}$ and define a norm

$$\|U\|_{X_0} = \| \, \|U\|_m(\cdot) \|, \tag{13a}$$

where $\|U\|_m(\cdot)$ is defined for each $s \in [0,1]$ by

$$\|U\|_m(s) = \sup_{\|x\|=1} \|U(s)x\|_2, \tag{13b}$$

where $x \in \mathbf{R}^2$, and $\| \; \|_2$ is the usual Euclidean norm on $\mathbf{R}^2$. That is, for each fixed s, $\|U\|_m$ is the operator norm of U, when U is viewed as an operator on $\mathbf{R}^2 \to \mathbf{R}^2$. Define another Banach space by

$$X = X_0 \oplus X_0, \tag{14}$$

with norm

$$\|T\|_X = \|T_1\|_{X_0} + \|T_2\|_{X_0}, \tag{15}$$

where T_1 and T_2 are the components of T.

We now rewrite the H-equations [Eqs. (12a) and (12b)] in a form more suitable for analysis: Postmultiply Eq. (12a) by $X(z)$, and premultiply Eq. (12b) by $Y(z)$. The results are

$$X(z) = I - \frac{1}{2} \int_0^1 d\mu \frac{z}{z+\mu} CY(\mu)X(z) - [z/(z+1)]MCY(1)X(z), \tag{16a}$$

$$Y(z) = I - \frac{1}{2} \int_0^1 d\mu \frac{z}{z+\mu} Y(z)X(\mu)C - [z/(z+1)]Y(z)X(1)MC. \tag{16b}$$

Define $\widetilde{X} \in X$ to be the ordered pair (X,Y), then Eqs. (16a) and (16b) can be rewritten in operator form as

$$\widetilde{X} = F\widetilde{X}. \tag{17}$$

We hope to show that Eq. (17) can be solved by iteration. That is, we hope to show that the sequence $\{\widetilde{X}^{(m)}\}_{m=0}^{\infty}$, defined recursively by

$$\widetilde{X}^{(m+1)} = F\widetilde{X}^{(m)}, \tag{18}$$

will converge to the solution of Eq. (17). We shall always choose $\widetilde{X}^{(0)} = I$. First we must determine if F is a contraction. Let $\widetilde{X} = (X_1, X_2)$ and $\widetilde{Y} = (Y_1, Y_2)$. Then

$$\begin{aligned} &F\widetilde{X}(z) - F\widetilde{Y}(z) \\ &= \Big(\frac{1}{2} \int_0^1 d\mu \frac{z}{z+\mu} C[X_2(\mu)X_1(z) - Y_2(\mu)Y_1(z)] + \frac{1}{2} MC \frac{z}{z+1} [X_2(1)X_1(z) - Y_2(1)Y_1(z)], \\ &\quad \frac{1}{2} \int_0^1 d\mu \frac{z}{z+\mu} [X_2(z)X_1(\mu) - Y_2(z)Y_1(\mu)]C + \frac{1}{2} C \frac{z}{z+1} M[X_1(1)X_2(z) - Y_1(1)Y_2(z)] \Big). \end{aligned} \tag{19}$$

The first component can be rewritten as

$$\frac{1}{4}C\int_0^1 d\mu\,\frac{z}{z+\mu}[X_2(\mu)-Y_2(\mu)][X_1(z)+Y_1(z)]+\frac{1}{4}C\int_0^1 d\mu\,\frac{z}{z+\mu}[X_2(\mu)+Y_2(\mu)][X_1(z)-Y_1(z)]$$
$$+\frac{1}{4}MC\frac{z}{z+1}[X_2(1)-Y_1(1)][X_1(z)+Y_1(z)]+\frac{1}{4}MC\frac{z}{z+1}[X_2(1)+Y_2(1)][X_1(z)-Y_1(z)],$$

and the second component can be rewritten in a similar form. Let $B(I;r)$ be the ball $\{T\in X: \|I-T\|\leqslant r\}$. If $\widetilde{X}$ and $\widetilde{Y}\in B$ then

$$\|\widetilde{X}+\widetilde{Y}\|_X\leqslant 2(1+r). \tag{20}$$

Using Eq. (19) and the inequality

$$\left|\int_0^1 d\mu\,\frac{z}{z+\mu}\right|\leqslant \ln 2,\quad \text{for } z\in[0,1], \tag{21}$$

it is possible to estimate the norm of $F\widetilde{X}-F\widetilde{Y}$ by

$$\|F\widetilde{X}-F\widetilde{Y}\|\leqslant(1+r)[\ln 2\|C\|+\|MC\|]\|\widetilde{X}-\widetilde{Y}\|. \tag{22}$$

It is convenient to define

$$a=\ln 2\|C\|+\|MC\|. \tag{23}$$

Therefore, F is a contraction operator provided

$$a(1+r)<1. \tag{24}$$

So F is a contraction operator for sufficiently small a and r. We must now show that r can be chosen large enough so that F: $B\to B$. For F to map B into B it is sufficient to require

$$r\geqslant(\|F\widetilde{X}^{(0)}-\widetilde{X}^{(0)}\|)/[1-a(1+r)]. \tag{25}$$

It is straightforward to estimate $\|F\widetilde{X}^{(0)}-\widetilde{X}^{(0)}\|$ by

$$\|F\widetilde{X}^{(0)}-\widetilde{X}^{(0)}\|\leqslant a. \tag{26}$$

Therefore, inequality (26) can be rewritten as

$$ar^2+(a-1)r+a\leqslant 0. \tag{27}$$

The largest value of a that is consistent with inequality (25) is

$$a=\tfrac{1}{3}. \tag{28}$$

For this value of a, r can be determined to be

$$r=1. \tag{29}$$

All that remains to be checked is inequality (24). With $a=\frac{1}{3}$ and $r=1$, we have

$$a(1+r)=\tfrac{2}{3}<1. \tag{30}$$

Therefore, if

$$\ln 2\|C\|+\|MC\|<\tfrac{1}{3}, \tag{31}$$

then Eq. (17) has a unique solution inside the ball $B(I;1)$. Furthermore, the sequence defined by Eq. (18) converges to this solution if $\widetilde{X}^{(0)}\in B(I;1)$.

Finally, we must check that $\widetilde{X}^{(m)}$ converges to a solution with the required analyticity and invertibility properties. First, we have the following lemma.

Lemma 2: If $\widetilde{X}=(X,Y)$ is the unique solution to Eq. (17) in the ball $B(I;1)$, and $a<\frac{1}{3}$, then

$$\det\left[I+\frac{1}{2}C\int_0^1 d\mu\,\frac{z}{z+\mu}Y(\mu)+\frac{z}{z+1}MCY(1)\right]\neq 0,$$

for $\operatorname{Re} z\geqslant 0$.

Proof: Since $(X,Y)\in B(I;1)$ we have

$$1\geqslant\|I-X\|\geqslant|1-\|X\||. \tag{32}$$

Thus we have

$$\|X\|\leqslant 2. \tag{33}$$

Suppose that for some value of z

$$\det\left[I+\frac{1}{2}C\int_0^1 d\mu\,\frac{z}{z+\mu}Y(\mu)+\frac{z}{z+1}MCY(1)\right]=0. \tag{34}$$

This implies that

$$\left\|\frac{1}{2}C\int_0^1 d\mu\,\frac{z}{z+\mu}Y(\mu)+\frac{z}{z+1}MCY(1)\right\|\geqslant 1, \tag{35}$$

or,

$$\|C\|\int_0^1\left|\frac{z}{z+\mu}\right|d\mu+\|MC\|\geqslant 1. \tag{36}$$

But, if $\operatorname{Re} z\geqslant 0$, then

$$\int_0^1\left|\frac{z}{z+\mu}\right|d\mu\leqslant 1. \tag{37}$$

So inequality (34) implies

$$\|C\|+\|MC\|\geqslant 1. \tag{38}$$

But by assumption

$$a=\ln 2\|C\|+\|MC\|<\tfrac{1}{3},$$

which is a contradiction. Therefore, $\hat{X}(z)$, defined by

$$\hat{X}(z)=I+\frac{1}{2}C\int_0^1 d\mu\,\frac{z}{z+\mu}Y(\mu)+\frac{z}{z+1}MCY(1), \tag{39}$$

is invertible for $\operatorname{Re} z\geqslant 0$. The same proof shows that

$$\hat{Y}(z)=I+\frac{1}{2}\int_0^1 d\mu\,\frac{z}{z+\mu}X(\mu)C+\frac{z}{z+1}X(1)MC$$

is also invertible for $\operatorname{Re} z\geqslant 0$. Furthermore, $\hat{X}$ and $\hat{Y}$ are analytic on $\mathbb{C}/[-1,0]$. Finally observe that $\hat{X}$ and $\hat{Y}$ satisfy the H-equations [Eqs. (12a) and (12b)]. Thus the iteration scheme defined by Eq. (18) converges to the solution of the H-equations, and satisfies the constraints.

ACKNOWLEDGMENTS

This work was supported in part by Department of Energy Grant No. DE-AS05-80ER10711 and National Science Foundation Grant No. DMS-8312451.

[1]B. L. Willis and C. V. van der Mee, "Transport equations with nondiagonal cross-section matrices," to appear in J. Math. Phys.
[2]T. W. Mullikian, "Weiner–Hopf factorizations," Transp. Theory Stat. Phys. 3, 215 (1973).
[3]R. L. Bowden and P. F. Zweifel, "A Banach space analysis of the Chandrasekhar H-equation," Astrophys. J. **210**, 178 (1976).
[4]R. L. Bowden, P. F. Zweifel, and R. Meinikoff, "On the solution of nonlinear matrix integral equations in transport theory," J. Math. Phys. **17**, 1722 (1976).
[5]L. B. Rall, *Computational Solution of Nonlinear Operator Equations* (Wiley, New York, 1969), pp. 74–85.

Unbounded representations of symmetry groups in gauge quantum field theory. II. Integration

A. H. Völkel
Institut für Mathematik III, Freie Universität Berlin, Berlin, West Germany

(Received 26 July 1985; accepted for publication 5 December 1985)

Within the gauge quantum field theory of the Wightman–Gårding type, the integration of representations of Lie algebras is investigated. By means of the covariance condition (substitution rules) for the basic fields, it is shown that a form skew-symmetric representation of a Lie algebra can be integrated to a form isometric and in general unbounded representation of the universal covering group of a corresponding Lie group provided the conditions (Nelson, Sternheimer, etc.), which are well known for the case of Hilbert or Banach representations, hold. If a form isometric representation leaves the subspace from which the physical Hilbert space is obtained via factorization and completion invariant, then the same is proved to be true for its differential. Conversely, a necessary and sufficient condition is derived for the transmission of the invariance of this subspace under a form skew-symmetric representation of a Lie algebra to its integral.

I. INTRODUCTION

In recent years the evidence that non-Abelian gauge quantum field theories are of outstanding importance for elementary particle physics has become overwhelming. For a consistent formulation of such a theory, it seems to be unavoidable[1–4] to introduce an indefinite metric formalism.[5] Such a formalism makes use of three Hilbert spaces $H_0 \subset H \subset \mathscr{H}$ with scalar product $(\cdot,\cdot)$ and a Hermitian sesquilinear form $\langle\cdot,\cdot\rangle := (\cdot,\eta\cdot)$ generated by a bounded symmetric operator η on $\mathscr{H}$. The sesquilinear form is semidefinite on H and induces a definite scalar product on the factor space H/H_0. The completion of this factor space is considered as the physical Hilbert space $\mathscr{H}_{\text{phys}} = \overline{H/H_0}$. The source of all grave structural differences between gauge quantum field theories and "classical" Wightman theories[6–8] is that physically interesting quantities are exclusively described by means of the sesquilinear form $\langle\cdot,\cdot\rangle$, whereas the basic fields are linear operators in $\mathscr{H}$ or more generally the topology is prescribed by the positive definite scalar product $(\cdot,\cdot)$ of $\mathscr{H}$. This "splitting" on the one hand allows the occurrence of physically most important phenomena like confinement[3,9–12] and Higgs[2] mechanisms, however, on the other hand a variety of nice, for the physical interpretation as well as for the calculations, vital (topological) properties of operators that hold in Wightman theories get lost. For instance, the representations of symmetry groups in $\mathscr{H}$ have to be isometric with respect to the form $\langle\cdot,\cdot\rangle$ (in the following called η-isometry), but in general they are not unitary and not even bounded in $\mathscr{H}$. An example is the Lorentz group in the Gupta–Bleuler formulation of the free electromagnetic vector potential (Refs. 5, Sec. III, 11, and 13). Moreover apart from the Schwinger model (with trivial interaction) in two space-time dimensions,[9,10] up to now a confinement mechanism has been proved to exist only on the basis of nonisometric representations of the translation group.[11,12] Furthermore, the operators representing observables (after multiplication by $i = \sqrt{-1}$) have to be skew-symmetric with respect to the form $\langle\cdot,\cdot\rangle$ (in the following called η-skew-symmetry), but in general they need not and cannot be essentially skew-adjoint in $\mathscr{H}$.

In any quantum theory (from nonrelativistic quantum mechanics to relativistic quantum field theories) continuous symmetry groups gain their particular importance from the close connection between representations of Lie groups and those of their Lie algebras, which is mathematically well established in *both* directions up until now only for *strongly continuous* representations in *Banach spaces*.[14–22] In these cases the observables (after multiplication by $\sqrt{-1}$) form just a maximal Abelian subalgebra of an essentially skew-adjoint representation for the Lie algebra of a Lie (symmetry) group of the system.

In contrast to the "classical" Wightman theories this concept obviously breaks down in gauge quantum field theories with an indefinite metric, unless the above connection between representations of Lie groups, respectively of their Lie algebras, can be generalized to the unbounded representations in $\mathscr{H}$ by means of the basic assumptions, which define a gauge quantum field theory.

In a gauge quantum field theory of the Wightman–Gårding type[2,5] the covariance condition for the basic field operators closely relates the (unbounded) representation U of a symmetry group G in $\mathscr{H}$ to *continuous* (and hence bounded) representations of G in the underlying complete, countably normed test function spaces.[23] For the latter the differentiation (Lie group repr.→Lie algebra repr.) is also a well understood process (Ref. 14, Chap. 4). In a recent publication[11] (in the following quoted as paper I) the author was able to prove, by means of the covariance condition and the differentiation in the test function spaces, that in a Wightman–Gårding theory for every η-isometric (unbounded) representation U of a Lie group G on a dense subspace $D_G \subseteq \mathscr{H}$ the differential ∂U exists and is η-skew-symmetric; i.e., the subspace $\mathscr{D}^\infty(U) \subseteq D_G$ of all $\mathscr{C}^\infty$-vectors for U is dense in D_G and on $\mathscr{D}^\infty(U)$ there exists a unique η-skew-symmetric representation ∂U of the Lie algebra $\mathfrak{g}$ of G

 0022-2488/86/041113-15$02.50

obtained from U by the usual strong limit process (Ref. 14, Chap. 4).

The main purpose of the present paper is to fill the remaining gap, namely the integration of a given η-skew-symmetric representation δU of a Lie algebra $\mathfrak{g}$ on a dense domain in $\mathscr{H}$ to an η-isometric (in general unbounded) representation U for a simply connected Lie group G belonging to $\mathfrak{g}$ such that $\partial U \supseteq \delta U$.

A major burden will be the integration of Lie algebra representations in the complete, countably normed test function spaces, which in contrast to the differentiation, unfortunately, cannot be found in the literature. Already in the case of Hilbert or Banach spaces some additional condition on a representation of a Lie algebra is required in order that its integral exists.[14–22] One cannot expect more than this in complete, countably normed spaces. It will turn out that we do not need stronger ones than those familiar from the case of Hilbert spaces, since the test function spaces can be represented as a countable intersection of Hilbert spaces.

The paper is organized as follows. In Sec. II we briefly review the basic concepts of a Wightman–Gårding gauge quantum field theory[2,5] with particular emphasis on the structure of the test function spaces and the covariance condition with respect to symmetry group G. In Sec. III we introduce the concept of an infinitesimal symmetry (i.e., a symmetry on the level of representations for the Lie algebra $\mathfrak{g}$ of G), and formulate and discuss the precise integrability conditions, as well as the main results of this paper. Sections IV and V contain the existence proofs for unique continuous representations of G in the countably normed test function spaces, respectively, of an η-isometric (in general unbounded) representation of G in the physical Hilbert space $\mathscr{H}$.

Of particular physical importance are the symmetries that leave the subspace H (with non-negative η-norm) invariant, because only their representations lift to unitary representations of G, respectively skew-symmetry ones of $\mathfrak{g}$, in the physical Hilbert space $\mathscr{H}_{\text{phys}}$. We call them *strict* global, respectively *strict* infinitesimal, symmetries and show in Sec. VI that the differential of every strict global symmetry is a strict infinitesimal one, and that the integral of a strict infinitesimal is a strict global one if and only if the subspace $\mathscr{D}^\infty(U) \cap H$ contains a dense set of analytic vectors that is invariant under the differential ∂U.

Appendix A contains a simple proof for the extension of Nelson's famous result[15] on the density of analytic vectors for continuous representations in Banach spaces to complete, countably normed spaces. In Appendix B we present a list of the most important integrability conditions in Hilbert and Banach spaces collected from the literature.

II. GQFT, GEL'FAND SPACES, AND (GLOBAL) SYMMETRIES

As indicated in the Introduction, a distinguished role will be played by the test function spaces. In the following, $S(\mathbb{R}_{4L}, \mathbb{C}_N)$ ($L, N \in \mathbb{N}$, natural numbers) denotes a Gel'fand space of type $K\{M_q\}$ satisfying the Gel'fand conditions[23,24] (N) and (P) or the Fourier transform of such a space. In more familiar terms it is a separable nuclear Frechét space of complex $\mathscr{C}^\infty$-functions f: $\mathbb{R}_{4L} \to \mathbb{C}_N$, $x \to f(x) = (f^1(x),\ldots,f^N(x))$ in $4L$ real variables $x = (x_1^0,\ldots, x_1^3, \ldots, x_L^0,\ldots, x_L^3)$. The topology is given by a countable set of pairwise compatible norms $\|f\|_q$, ($q \in \mathbb{N}^0 = \mathbb{N} \cup \{0\}$) with $\|\cdots\|_q \leqslant \|\cdots\|_{q+1}$ (Ref. 23, Chap. I). Four important properties of these spaces, which will be needed below, are the following (See Ref. 23, Chap. I, § 3–6 and Ref. 25, Part III, Proposition 50.2).

(S.0) Every space $S(\mathbb{R}_{4L}, \mathbb{C}_N)$ is perfect; this means every bounded subset is relatively sequentially compact.

(S.I) If $S_n(\mathbb{R}_{4L}, \mathbb{C}_N)$ denotes the completion of $S(\mathbb{R}_{4L}, \mathbb{C}_N)$ with respect to the norm $\|\cdots\|_n$, then

$$S(\mathbb{R}_{4L}, \mathbb{C}_N) = \bigcap_{n \in \mathbb{N}^0} S_n(\mathbb{R}_{4L}, \mathbb{C}_n).$$

(S.II) Every linear continuous functional F on $S(\mathbb{R}_{4L}, \mathbb{C}_N)$ is of finite order; i.e., there exist $C \in \mathbb{R}^+$ (positive real numbers) and a minimal $q_0 \in \mathbb{N}^0$ such that for all $q \geqslant q_0$, $|F(f)| \leqslant C\|f\|_q$.

(S.III) The nuclear theorem holds.

Physically important examples of such spaces are, besides the Schwartz space of strongly decreasing $\mathscr{C}^\infty$-functions, all strictly localizable spaces of Jaffe $S(\mathbb{R}_{4L}, \mathbb{C}_N) = \mathscr{L}(\mathbb{R}_{4L}) \bar{\otimes} \mathbb{C}_N$ ($\bar{\otimes}$ denotes the completed tensor product).[26] The countable set of norms generating their topologies is defined by the following equations:

$$M^L_{m,\rho}(p) := \prod_{t=1}^{L} g\Big(\rho \sum_{s=0}^{3} (p_t^s)^2\Big) \prod_{i=0}^{3} (1 + |p_t^i|)^m, \quad (2.1)$$

$$\|f\|_{\nu(m,\rho)} := \sup\Big\{ M^L_{m,\rho}(p) \Big| \partial_p^{|r|} \int d^{4L}x\, f^\mu(x) \times \exp\Big[-i \sum_{k=1}^{L} (p_k \cdot x_k)\Big]\Big| \;\Big|\; p \in \mathbb{R}_{4L};\ r_j^i \in \mathbb{N}^0, \sum_{t=1}^{L}\sum_{s=0}^{3} r_t^s \leqslant m;\ \mu = 1,\ldots,N \Big\};$$
$$m \in \mathbb{N}^0, \quad \rho \in \mathbb{N}. \quad (2.2)$$

We use the abbrevations

$$(p \cdot x) := p^0 x^0 - \sum_{\alpha=1}^{3} p^\alpha x^\alpha$$

and

$$\partial_p^{|r|} := \frac{\partial^{r_1^0 + \cdots + r_n^3}}{\Pi_{j=1}^{L} \Pi_{i=0}^{3} (\partial p_j^i)^{r_j^i}}.$$

Furthermore, ν denotes a bijection from $\mathbb{N}^0 \times \mathbb{N}$ onto $\mathbb{N}^0$; g: $\mathbb{C} \to \mathbb{R}$ is some entire function, which on $\mathbb{R}^+ \cup \{0\}$ is positive, monotonically growing, and satisfies the condition (strict localizability)

$$\int_1^\infty dt\, t^{-2} \ln g(t^2) < +\infty. \quad (2.3)$$

The Schwartz space is the special case $g \equiv 1$.

According to (S.I), the space $S(\mathbb{R}_{4L}, \mathbb{C}_N)$ is the intersection of a countable set of Banach spaces. However, the great majority of integrability conditions for representations of Lie algebras[14–21] is formulated in terms of scalar products of Hilbert spaces. Hence, one would prefer a topology that is equivalent to the one above, but in addition admits $S(\mathbb{R}_{4L},$

$\mathbb{C}_N$) to be represented as an intersection of a countable number of Hilbert spaces. Indeed for the case of the Gel'fand spaces $K\{M_q\}$ satisfying the Gel'fand condition (N), there exists a countable set of scalar products $[f,h]_q$ ($q\in\mathbb{N}^0$) on $S(\mathbb{R}_{4L}, \mathbb{C}_N)$, such that the topology obtained from the corresponding set of norms $|f|_q := \sqrt{[f,f]_q}$ ($q\in\mathbb{N}^0$) is equivalent to the original one (Ref. 24, Chap. I, §3]. But then property (S.I) reads as follows.

(S.$\tilde{\text{I}}$) If H_n ($\mathbb{R}_{4L}$, $\mathbb{C}_N$) denotes the completion of $S(\mathbb{R}_{4L}, \mathbb{C}_N)$ with respect to the Hilbert norm $|f|_n = \sqrt{[f,f]_n}$, then

$$S(\mathbb{R}_{4L},\mathbb{C}_N) = \bigcap_{n\in\mathbb{N}^0} H_n(\mathbb{R}_{4L},\mathbb{C}_N).$$

Furthermore without less of generality it can be assumed that

$$|\cdots|_q \leqslant |\cdots|_{q+1}, H_{q+1}(\mathbb{R}_{4L},\mathbb{C}_N) \subseteq H_q(\mathbb{R}_{4L},\mathbb{C}_N). \quad (2.4)$$

In the remainder of this paper, we exclusively refer to the topology defined by these Hilbert norms. For the explicit examples of strictly localizable spaces given above, the scalar products read

$$[f,h]_{v(m,\rho)} := \int d^{4L}p\, M^L_{m,\rho}(p)^2 \times \sum_{r\in R^L_m} \sum_{\mu=1}^{N} [\overline{\partial_p^{|r|} f^\mu_F(p)}]\,\partial_p^{|r|} h^\mu_F(p). \quad (2.5)$$

Here $f^\mu_F(p)$ denotes the Fourier transform of $f^\mu(x)$,

$$f^\mu_F(p) = (2\Pi)^{-3/2L}\int d^{4L}x\, f^\mu(x)\exp\left[-i\sum_{j=1}^{L}(p_j\cdot x_j)\right], \quad (2.6)$$

and $\bar{f}$ is the complex conjugate of f and R^L_m is the subset of non-negative integers:

$$R^L_m := \left\{ r^j_t\in\mathbb{N}^0 \,\middle|\, |r_t| = \sum_{j=0}^{3} r^j_t \leqslant m;\ t=1,\dots,L \right\}. \quad (2.7)$$

With these preparations we can proceed to the formulation of the basic assumptions that define a gauge quantum field theory (GQFT) of the Wightman–Gårding type.[2,5]

A.I: Field operators: Let $\mathscr{H}$ denote a Hilbert space with elements $\Psi, \Phi, \dots$, scalar product (Ψ, Φ) and norm $\|\Psi\|_H$, D a dense subspace of $\mathscr{H}$, and $\mathbb{T}$ an at most countable set of (multi-) indices $\Gamma, \Delta, \dots$. Then for every $f\in S(\mathbb{R}_{4L}, \mathbb{C})$ and $\Gamma\in\mathbb{T}$ there exists a linear operator $\varphi_\Gamma(f)$ with domain $D(\varphi(f))$ such that (a) $D\subseteq D(\varphi_\Gamma(f))$ and $\varphi_\Gamma(f)D\subseteq D$; (b) if $\varphi_{\Gamma^*}(\bar{f}) = \varphi^*_\Gamma(\bar{f}) := \varphi_\Gamma(f)^*$ denotes the adjoint operator of $\varphi_\Gamma(f)$, then $\Gamma^*\in\mathbb{T}$ for every $\Gamma\in\mathbb{T}$; and (c) for all $\Phi\in\mathscr{H}$, $\Psi\in D$ the mapping $f\to(\Phi, \varphi_\Gamma(f)\Psi)$ is a linear continuous functional on $S(\mathbb{R}_{4L}, \mathbb{C})$.

Plainly all field operators are closable and for the sake of notational simplicity we assume they are closed; i.e., $\varphi_{\Gamma^{**}} = \varphi_\Gamma$.

A.II: Metric operator and physical Hilbert space: There exists a linear, bounded, and Hermitian operator η with $\eta D\subseteq D$ that generates a nontrivial and nonpositive semidefinite sesquilinear form $\langle\cdot,\cdot\rangle := (\cdot,\eta\cdot)$ on $\mathscr{H}$ as well as a nontrivial and maximal linear subspace $H\subseteq\mathscr{H}$ such that for all $\Psi\in H$ $\langle\Psi,\Psi\rangle\geqslant 0$. If H_0 denotes the linear subspace of all $\Psi\in H$ with $\langle\Psi,\Psi\rangle = 0$, then the completion of the factor space H/H_0 (with elements $[\Psi] := \Psi + H_0$) in the natural scalar product $([\Psi],[\Phi])_H := \langle\Psi,\Phi\rangle$ is called the physical Hilbert space $\mathscr{H}_{\text{phys}}$.

Throughout this paper we mean by a representation R of the group $(G,\cdot)$ on a vector space E, a homomorphism R: $G\to\text{Aut}\,E, g\to R(g)$ of $(G,\cdot)$ into the automorphism group of E, and by a representation W of a Lie algebra $(\mathfrak{g}, +, \cdot, [\ ,\ ])$ on E a homomorphism $W: \mathfrak{g}\to\text{End}\,E, X\to W(X)$ of the vector space $(\mathfrak{g}, +, \cdot)$ into the vector space of the endomorphisms of E satisfying

$$W([X,Y]) = W(X)W(Y) - W(Y)W(X).$$

A.III: Translational invariance and the vacuum: There exists a representation T of the vector group of $\mathbb{R}_4$ on a dense subspace $D_T\supseteq D$ that leaves D invariant and has the further properties (a) (η-isometry), for all $y\in\mathbb{R}_4$ and Ψ, $\Phi\in D_T$

$$\langle T(y)\Psi, T(y)\Phi\rangle = \langle\Psi,\Phi\rangle,$$

and (b) (vacuum), there exists a unique state $\Psi_0\in H$ (called the vacuum) such that $\langle\Psi_0, \Phi_0\rangle = 1$ and for all $y\in\mathbb{R}_4$: $T(y)\Psi_0 = \Psi_0$.

A.IV: Completeness and η-stability: The vacuum Ψ_0 is cyclic with respect to the polynomial *-algebra $\mathscr{P}(\varphi)$ over $\mathbb{C}$ generated by the set $\{\text{id}_{\mathscr{H}}, \varphi_\Gamma(f)|\Gamma\in\mathbb{T}, f\in S(\mathbb{R}_4, \mathbb{C})\}$, $\mathscr{P}(\varphi)\Psi_0\cap H$ is dense in H and $\eta\,\mathscr{P}(\varphi)\Psi_0\subseteq\mathscr{P}(\varphi)\Psi_0$.

The completeness assumption means that the subspace (LH ~ linear hull)

$$\mathscr{D}_{\Pi} := \mathscr{P}(\varphi)\Psi_0 = \text{LH}\left\{\Psi_0, \prod_{i=1}^{n}\varphi_{\Gamma_i}(f_i)\Psi_0 \,\middle|\, f_i\in S(\mathbb{R}_4, \mathbb{C});\ \Gamma_i\in\mathbb{T}; n\in\mathbb{N}\right\} \quad (2.8)$$

is dense in $\mathscr{H}$. As a consequence $\mathscr{H}$ is separable, since $S(\mathbb{R}_4, \mathbb{C})$ is separable, and $\mathbb{T}$ countable. Let $[\bar{\otimes}^L_{i=1}\varphi_{\Gamma_i}](f)$, $f\in S(\mathbb{R}_{4L}, \mathbb{C})$ denote the linear operator (extended tensor product) defined on D by the limit (See Ref. 11 Sec. 2 and Ref. 27, Chap. II, §1)

$$\left[\bar{\otimes}^L_{i=1}\varphi_{\Gamma_i}\right](f)\Psi := \lim_{n\to\infty}\sum_{j=1}^{n}\prod_{i=1}^{L}\varphi_{\Gamma_i}(f^j_i)\Psi, \quad \Psi\in D, \quad (2.9)$$

for an arbitrary sequence $(\Sigma^n_{j=1} \otimes^L_{i=1} f^j_i)_{n\in\mathbb{N}}$ from the algebraic tensor product $\otimes^L S(\mathbb{R}_4, \mathbb{C})\subset S(\mathbb{R}_{4L}, \mathbb{C})$ converging in $S(\mathbb{R}_{4L}, \mathbb{C})$ to f. These operators inherit from the basic fields all domain and continuity properties of the latter. The dense subspace $\mathscr{D}_{\text{QL}}\subset\mathscr{H}$ of so called quasilocal states is defined by the linear hull

$$\mathscr{D}_{\text{QL}} := \text{LH}\left\{\Psi_0, \left[\bar{\otimes}^L_{i=1}\varphi_{\Gamma_i}\right](f)\Psi_0 \,\middle|\, f\in S(\mathbb{R}_{4L},\mathbb{C});\ \Gamma_1,\dots,\Gamma_L\in\mathbb{T}; L\in\mathbb{N}\right\}. \quad (2.10)$$

Plainly it belongs to the domain of every basic field operator $\varphi_\Gamma(f)$; i.e., $\mathscr{D}_{\text{QL}}\subseteq D$.

For the sake of completeness we only mention the two remaining axioms of a Wightman–Gårding theory without spelling them out in detail,[2,5] because we do not use them.

A. V: Spectrum condition.

A. VI: Locality (Einstein causality).

These two additional conditions, or course, would reduce the class of admissible test function spaces to the strictly localizable ones explicitly given at the beginning of this section[26] and the Fourier transforms of them.

The multi-indices Γ of the fields are fixed by finite-dimensional representations of one or more Lie groups, respectively their Lie algebras, according to which the fields transform due to the covariance condition below. Hence for a given Lie group G with Lie algebra $\mathfrak{g}$, the multi-index Γ is a pair $\Gamma = (\mu, \mathscr{A})$, where $\mathscr{A}$ describes a definite representation of G or $\mathfrak{g}$ together with all possible degeneracies or multiplicities, and μ counts the components within such a representation $\mathscr{A}$. The set composed of the subindices μ for a fixed $\mathscr{A}$ will be denoted by $\mathbb{T}_G(\mathscr{A})$, the number of its elements by $\mathring{\mathscr{A}} := |\mathbb{T}_G(\mathscr{A})|$, and the set of all subindices $\mathscr{A}$, $\mathscr{B}$,... by I_G. Thus the set $\mathbb{T}$ is the countable union of pairwise disjoint subsets $\mathbb{T} = \cup_{\mathscr{A}\in I_G} \mathbb{T}G(\mathscr{A}) \times \{\mathscr{A}\}$. It is of considerable advantage to introduce, instead of the one component fields $\varphi_{(\mu,\mathscr{A})}$, the Wightman–Gårding fields with respect to a symmetry group G defined by[5]

$$\phi^{\mathscr{A}}(f) := \sum_{\mu\in\mathbb{T}_G(\mathscr{A})} \varphi_{(\mu,\mathscr{A})}(f^{\mu}); \quad f\in S(\mathscr{A}) := S(\mathbb{R}_4, \mathbb{C}_{\mathring{\mathscr{A}}}). \tag{2.11}$$

Analogously for f from the completed tensor product (see Ref. 14, Appendix 2, Ref. 25, Part III, Chap. 50 ff, and Ref. 28, §41–44)

$$S(\mathscr{A}_1,...,\mathscr{A}_L) := \bar{\otimes}_{i=1}^{L} S(\mathbb{R}_4, \mathbb{C}_{\mathring{\mathscr{A}}_i}) = S(\mathbb{R}_{4L}, \bar{\otimes}_{i=1}^{L} \mathbb{C}_{\mathring{\mathscr{A}}_i}) \tag{2.12}$$

of the *nuclear* spaces $S(\mathbb{R}_4, \mathbb{C}_{\mathring{\mathscr{A}}})$, we define

$$\phi^{\mathscr{A}_1\cdots\mathscr{A}_L}(f) := \sum_{\mu_1\in\mathbb{T}_G(\mathscr{A}_1)} \cdots \sum_{\mu_L\in\mathbb{T}_G(\mathscr{A}_L)} \left[\bar{\otimes}_{i=1}^{L} \varphi_{(\mu_i,\mathscr{A}_i)}\right](f^{\mu_1\cdots\mu_L}). \tag{2.13}$$

Of course the matrix elements of the new field operators inherit from those of the original fields the properties of being linear continuous functionals, possessing finite order, etc.

Let us note in passing that a topology in $S(\mathscr{A}_1,...,\mathscr{A}_L)$ equivalent to its original ($\pi-$ or $\epsilon-$) one is given by a countable set of norms $|f|_p := \sqrt{[f,f]_p}$, $p\in\mathbb{N}^0$ derived from scalar products. For the case of the strictly localizable spaces they read

$$[f,h]^{L}_{\nu(m,\rho)} := \int d^{4L}p\, M^{L}_{m,\rho}(p)^2 \sum_{q\in R^{L}_{m}} \sum_{\mu_1\in\mathbb{T}_G(\mathscr{A}_1)} \cdots \sum_{\mu_L\in\mathbb{T}_G(\mathscr{A}_L)} \left[\overline{\partial_p^{|q|} f_F^{\mu_1\cdots\mu_L}(p)}\right] \partial_p^{|q|} h_F^{\mu_1\cdots\mu_L}(p). \tag{2.14}$$

According to Definition 2.1 and Sec. III, Theorem 3.2 in paper I,[11] a Lie group is called a (global) symmetry group iff there exist a decomposition of the index set $\mathbb{T}$ into a countable union $\mathbb{T} = \cup_{A\in I_G} \mathbb{T}_G(\mathscr{A}) \times \{\mathscr{A}\}$ of pairwise disjoint subsets with $\mathring{\mathscr{A}} = \mathring{\mathscr{A}}^*$, for every $\mathscr{A}\in I_G$ a continuous representation $R^{\mathscr{A}}: G \to \mathrm{Aut}\, S(\mathscr{A})$, $g \to R^{\mathscr{A}}(g)$ of G on $S(\mathscr{A})$ and a representation $U: G \to \mathrm{Aut}\, D_G$, $g \to U(g)$ of G in $\mathscr{H}$ with the following properties.

(U.1) $\mathscr{D}_{\mathrm{QL}} \subseteq D_G$.

(U.2a) (invariance of the vacuum): $\forall_{g\in G}$, $U(g)\Psi_0 = \Psi_0$.

(U.2b) (covariance): For all $f\in S(\mathscr{A}_1,...,\mathscr{A}_L)$, $L\in\mathbb{N}$, $g\in G$, and $\Psi\in\mathscr{D}_{\mathrm{QL}}$:

$$U(g)\phi^{\mathscr{A}_1\cdots\mathscr{A}_L}(f)U(g)^{-1}\Psi = \phi^{\mathscr{A}_1\cdots\mathscr{A}_L}(\bar{\otimes}_{i=1}^{L} R^{\mathscr{A}_i}(g) f)\Psi.$$

(U.3) (η-isometry): $\langle U(g)\Psi, U(g)\Phi\rangle = \langle\Psi,\Phi\rangle$; $g\in G$; $\Psi, \Phi\in D_G$.

The pair (G, U) is called a strict (global) symmetry if

$$\forall_{g\in G},\ U(g)(\mathscr{D}_{\Pi}\cap H) \subseteq \mathscr{D}_{\Pi}\cap H. \tag{2.15}$$

Here, $R^{\mathscr{A}_1\cdots\mathscr{A}_L}(g) := \bar{\otimes}_{i=1}^{L} R^{\mathscr{A}_i}(g)$ denotes the extended tensor product of the continuous operators $R^{\mathscr{A}_i}(g)$, $i = 1,..., L$; i.e., the continuous extension of their algebraic tensor product onto $S(\mathscr{A}_1,...,\mathscr{A}_L)$.

Besides a lot of other things about unbounded representations of symmetry groups, the author showed in Sec. V of paper I that the differential ∂U of every (unbounded) representation U of a symmetry group G with Lie algebra $\mathfrak{g}$ exists and is η-skew-symmetric. To be more definite, let $\sigma^\infty(\mathscr{A}_1,...,\mathscr{A}_L) \subset S(\mathscr{A}_1,...,\mathscr{A}_L)$ denote the dense subspace of $\mathscr{C}^\infty$-vectors for the continuous representations $R^{\mathscr{A}_1\cdots\mathscr{A}_L}$ (paper I, Theorem 3.1), $\partial R^{\mathscr{A}_1\cdots\mathscr{A}_L}$: $\mathfrak{g} \to \mathrm{End}\, \sigma^\infty(\mathscr{A}_1,...,\mathscr{A}_L)$, $X \to \partial R^{\mathscr{A}_1\cdots\mathscr{A}_L}(X)$ the differential of $R^{\mathscr{A}_1\cdots\mathscr{A}_L}$ (Ref. 14, pp. 252–254), and

$$\mathscr{D}^\infty(U) := \mathrm{LH}\{\Psi_0, \phi^{\mathscr{A}_1\cdots\mathscr{A}_L}(f)\Psi_0 \mid f\in\sigma^\infty(\mathscr{A}_1,...,\mathscr{A}_L); \mathscr{A}_1,...,\mathscr{A}_L \in I_G; L\in\mathbb{N}\} \tag{2.16}$$

the dense subspace of "C^∞-vectors for U." Then the strong limits

$$\partial U(X)\Psi := \operatorname*{s-lim}_{t\to 0} t^{-1}[U(\exp tX)\Psi - \Psi],$$

with $X\in\mathfrak{g}$, $\Psi\in\mathscr{D}^\infty(U)$ and exp: $\mathfrak{g}\to G$ the exponential mapping, define a η-skew-symmetric representation ∂U: $\mathfrak{g} \to \mathrm{End}\, \mathscr{D}^\infty(U)$, $X \mapsto \partial U(X)$ of $\mathfrak{g}$ in $\mathscr{H}$, which was called the differential of U.

In the following sections we investigate the inverse process, which we call integration.

III. INTEGRABILITY CONDITIONS AND RESULTS

In the first place, we have to make precise what we mean by a symmetry on the level of representations of the Lie algebra of a Lie group.

Definition 3.1 (infinitesimal symmetry): A connected Lie group G with Lie algebra $\mathfrak{g}$ is called an infinitesimal symmetry group iff (i) there exist a decomposition of the index set $\mathbb{T}$ into a countable union $\mathbb{T} = \cup_{\mathscr{A}\in I_{G_0}} \mathbb{T}_G(\mathscr{A}) \times \{\mathscr{A}\}$ of pairwise disjoint subsets with $\mathring{\mathscr{A}} = \mathscr{A}^*$, and for every $\mathscr{A}\in I_G$ a representation $\delta^{\mathscr{A}}$: $\mathfrak{g} \to \mathrm{End}\, E^{\mathscr{A}}$, $X \mapsto \delta^{\mathscr{A}}(X)$ of the Lie algebra $\mathfrak{g}$ on a dense subspace $E^{\mathscr{A}} \subseteq S(\mathscr{A}) = S(\mathbb{R}_4,$

$\mathbb{C}_{\mathscr{A}}$); (ii) the linear operators $\delta^{\mathscr{A}}(X)$ and $X \in \mathfrak{g}$ are closable with respect to every norm $|\cdots|_p, p \in \mathbb{N}^0$ [i.e., closable in every Hiblert space $H_p(\mathscr{A}) = H_p(\mathbb{R}_4, \mathbb{C}_{\mathscr{A}}), p \in \mathbb{N}^0$]; and (iii) the representation δU: $\mathfrak{g} \to \mathrm{End}\, \mathscr{D}_E$ of $\mathfrak{g}$ generated on the dense subspace (of $\mathscr{H}$)

$$\mathscr{D}_E := \mathrm{LH}\Big\{\Psi_0, \prod_{i=1}^{L} \phi^{\mathscr{A}_i}(f_i)\Psi_0 | f_i \in E^{\mathscr{A}_i}; \mathscr{A}_1, \ldots, \mathscr{A}_L \in I_G; L \in \mathbb{N}\Big\} \subseteq \mathscr{D}_\Pi \tag{3.1}$$

via (the covariance conditions)

$$\delta U(X)\Psi_0 := \mathscr{O} \quad (\mathscr{O} \text{ the zero vector in } \mathscr{H}),$$

$$\delta U(X) \prod_{i=1}^{L} \phi^{\mathscr{A}_i}(f_i)\Psi_0 := \sum_{i=1}^{L}\Big(\prod_{j=1}^{i-1} \phi^{\mathscr{A}_j}(f_j)\Big) \phi^{\mathscr{A}_i}(\delta^{\mathscr{A}_i}(X) f_i) \times \Big(\prod_{k=i+1}^{L} \phi^{\mathscr{A}_k}(f_k)\Big)\Psi_0 \tag{3.2}$$

is η-skew-symmetric on $\mathscr{D}_E$; i.e., for all $X \in \mathfrak{g}$ and

$$\Psi, \Phi \in \mathscr{D}_E: \ \langle \delta U(X)\Psi, \Phi\rangle = -\langle \Psi, \delta U(X)\Phi\rangle.$$

The pair $(\mathfrak{g}, \delta U)$ is called a *strict* infinitesimal symmetry if

$$\overline{\mathscr{D}_E \cap H} = H \ \wedge \ \forall_{X \in \mathfrak{g}}, \delta U(X)(\mathscr{D}_E \cap H) \subseteq \mathscr{D}_E \cap H. \tag{3.3}$$

Now the problem is under which additional conditions on $\delta^{\mathscr{A}}$ does the infinitesimal symmetry generate a global one in the sense of Sec. II. This means, there exists for every $\mathscr{A} \in I_G$ a unique continuous representation $R^{\mathscr{A}}$: $G \to \mathrm{Aut}\, S(\mathscr{A})$ as well as a (in general unbounded) representation U: $G \to D_G$ of G in $\mathscr{H}$, which satisfies, besides the conditions (U.1)–(U.3) in Sec. II, the following two conditions ($\restriction$ ~restriction).

(U.4) $E^{\mathscr{A}} \subseteq \sigma^\infty(\mathscr{A}) \wedge \mathscr{D}_E \subseteq \mathscr{D}^\infty(U)$.
(U.5) $\forall_{X \in \mathfrak{g}} \delta^{\mathscr{A}}(X) = \delta R^{\mathscr{A}}(X) \restriction E^{\mathscr{A}}$
$\wedge \delta U(X) = \delta U(X) \restriction \mathscr{D}_E$.

A major part of this problem is the integration of the representations $\delta^{\mathscr{A}}$, $\mathscr{A} \in I_G$ in the countably normed spaces $S(\mathscr{A})$. Since the latter are intersections of a countable number of Hilbert spaces $H_p(\mathscr{A}) = H_p(\mathbb{R}_4, \mathbb{C}_{\mathscr{A}})$, $p \in \mathbb{N}^0$, it is obvious that the integration in $S(\mathscr{A})$ is closely connected with the integration of $\delta^{\mathscr{A}}$ in every Hilbert space $H_p(\mathscr{A})$, $p \in \mathbb{N}^0$ (See Refs. 14–21). Integration in a Hilbert space is defined as follows.

Definition 3.2: Let G be a connected Lie group with Lie algebra $\mathfrak{g}$. A representation δ: $\mathfrak{g} \to \mathrm{End}\, E$ of $\mathfrak{g}$ on a dense $E \subset H_p$ is call integrable iff there exists a unique strongly continuous representation $R_p: G \to \mathrm{Aut}\, H_p$ such that $E \subseteq \sigma_p^\infty$ and $\forall_{X \in \mathfrak{g}}, \delta(X) = \partial R_p(X) \restriction E$.

Here σ_p^∞ and ∂R_p denote the subspace of all $\mathscr{C}^\infty$-vectors for R_p, respectively the differential of R_p.

The results obtained in paper I (especially in Sec. VI) imply that the following two conditions are *necessary* for the integrability of $\delta^{\mathscr{A}}$ in $S(\mathscr{A})$: (α) $\delta^{\mathscr{A}}$ is integrable in every Hilbert space $H_p(\mathscr{A})$, $p \in \mathbb{N}^0$; and (β) for all $p \in \mathbb{N}$ and $g \in G$: $R_p^{\mathscr{A}}(g) = R_{p-1}^{\mathscr{A}}(g) \restriction H_p(\mathscr{A})$.

Obviously these two conditions [but eventually not (α) alone] are also sufficient for the integration. However for practical reasons in applications one would prefer instead of (α) and (β), conditions that are directly imposed on the given representation $\delta^{\mathscr{A}}$. In case of (α) they are well known in the literature.[14–21] In the most important ones, one of the integrability conditions is that $\delta^{\mathscr{A}}(X)$, $X \in \mathfrak{g}$ are skew-symmetric operators in $H_p(\mathscr{A})$ (see Refs. 14–16, 18, and 20). It will be shown in Sec. IV that the symmetry of $\delta^{\mathscr{A}}(X_i)$, $i = 1, \ldots, n$, in $H_p(\mathscr{A})$ also implies condition (β) and hence together with the integrability in $H_p(\mathscr{A})$ also the integrability in $S(\mathscr{A})$.

Another condition for (β) is that the representation $\delta^{\mathscr{A}}$ is topologically irreducible in $H_p(\mathscr{A})$ [i.e., every nontrivial subspace of $E^{\mathscr{A}}$ that is invariant under $\delta^{\mathscr{A}}$ is dense in $H_p(\mathscr{A})$], and $E^{\mathscr{A}}$ contains at least one analytic vector for the representation $\delta^{\mathscr{A}}$. However, for the *strict* (infinitesimal) symmetry groups, i.e., the most important ones in GQFT, the irreducibility does not hold, because the invariance of $(\mathscr{D}_E \cap H)$ for δU leads to nontrivial closed invariant subspaces for $\delta^{\mathscr{A}}$ (see Sec. VI). A further one is that the subspace $E^{\mathscr{A}}$ is a core for $\partial R_p^{\mathscr{A}}$.

Besides the integration in $S(\mathscr{A})$, $\mathscr{A} \in I_G$, we have to prove the η-isometry of U in $\mathscr{H}$, and for this proof we will need the existence of what we call a p-entire subspace of analytic vectors for the representation $\delta^{\mathscr{A}}$ (see Definition 3.3 below). On the one hand, the integrability of $\delta^{\mathscr{A}}$ in $H_p(\mathscr{A})$ together with either the skew-symmetry or the irreducibility imply the existence of such a p-entire subspace in $H_p(\mathscr{A})$ (Lemma 4.1). On the other hand, the existence of a p-entire subspace in $H_p(\mathscr{A})$ implies both conditions (α) and (β) (see Ref. 14, Lemma 9.1 and Ref. 21, Proposition 2.3). Hence it seems to be a suitable alternative condition, especially for cases in which neither the skew-symmetry nor the irreducibility of $\delta^{\mathscr{A}}$ in $H_p(\mathscr{A})$ holds.

In the consideration of strict symmetries in Sec. VI we need, besides analytic vectors in $H_p(\mathscr{A})$, which means analyticity in one single norm $|\cdots|_p$, the concept of analytic vectors in the space $S(\mathscr{A})$, i.e., analyticity with respect to all countably infinite many norms. In order to avoid confusion we will call the former vectors p-analytic and the latter ∞-analytic.

In the following ${}^p\overline{T}$ denotes the closure of an operator T (or of a subspace T in the norm topology) in $H_p(\mathscr{A})$.

Definition 3.3:(i) An element of $f \in \cup_{n \in \mathbb{N}^0} H_n(\mathscr{A})$ is called *p-analytic for a representation* $\delta^{\mathscr{A}}$ of a Lie algebra $\mathfrak{g}$ with basis $X_1, \ldots, X_n$ if it is analytic for $\delta^{\mathscr{A}}$ in the Hilbert space $H_p(\mathscr{A})$, i.e., if there exists a number $M(f,p) \in \mathbb{R}^+$ such that for all $m \in \mathbb{N}$ and $i_1, \ldots, i_m \in \{1, \ldots, n\}$ we have

$$f \in \mathrm{Dom}\big({}^p\overline{\delta^{\mathscr{A}}(X_{i_1})} \cdot \ldots \cdot {}^p\overline{\delta^{\mathscr{A}}(X_{i_m})}\big), \tag{3.4}$$

$$\big|{}^p\overline{\delta^{\mathscr{A}}(X_{i_1})} \cdot \ldots \cdot {}^p\overline{\delta^{\mathscr{A}}(X_{i_m})} f\big|_p \leqslant M(f,p)^m m!. \tag{3.5}$$

(ii) A subspace $\Omega_p^s(A)$ of p-analytic vectors for $\delta^{\mathscr{A}}$ is called *p-entire for* $\delta^{\mathscr{A}}$ if it is dense in $H_p(\mathscr{A})$ and if in addition there exists a number $s \in \mathbb{R}^+$ and for every $f \in \Omega_p^s(\mathscr{A})$ a number $M(f, p) \in \mathbb{R}^+$ such that for all $m \in \mathbb{N}$ and $i_1, \ldots, i_m \in \{1, \ldots, n\}$ we have

$$\left| {}^p\overline{\delta^{\mathscr{A}}(X_{i_1})} \cdot \ldots \cdot {}^p\overline{\delta^{\mathscr{A}}(X_{i_m})} f \right|_p \leqslant M(f,p) s^m m! . \quad (3.6)$$

(iii) A vector $f \in \cap_{p\in \mathbb{N}^0} H_p(\mathscr{A})$ is called *∞-analytic for* $\delta^{\mathscr{A}}$ if it satisfies the conditions (3.4) and (3.5) for every $p \in \mathbb{N}^0$.

In the case of ∞-analytic vectors for $\delta^{\mathscr{A}}$, one should notice the p-dependence of the constants $M(f,p)$ in the inequality (3.5). If $\Omega_p^\omega(\mathscr{A})$ denotes the subspace of all p-analytic vectors for $\delta^{\mathscr{A}}$, then apparently we have $\Omega_p^\omega(\mathscr{A}) = \cup_{s\in \mathbb{R}^+} \Omega_p^s(\mathscr{A})$. In the special case of the differential $\partial R_p^{\mathscr{A}}(\partial R^{\mathscr{A}})$ of a continuous representation $R_p^{\mathscr{A}}(R^{\mathscr{A}})$ we will denote the subspace of all analytic (∞-analytic) vectors by $\sigma_p^\omega(\mathscr{A})$ $(\sigma^\omega(\mathscr{A}))$. It is well known (see Ref. 14, Theorem 4.4.5.7 and Ref. 15, Theorem 4) that $\sigma_p^\omega(\mathscr{A})$ is dense in $\sigma_p^\infty(\mathscr{A})$ and hence in $H_p(\mathscr{A})$. From an inspection of the proof of this result for Banach spaces (Ref. 14, Chap. 4, §4.4.5) and the observation of the relation $\sigma^\omega(\mathscr{A}) = \cap_{p\in\mathbb{N}^0} \sigma_p^\omega(\mathscr{A})$ it is easily seen (Appendix A) that the same holds also for the subspace of all ∞-analytic vectors for the differential $\partial R^{\mathscr{A}}$ in the complete countably normed space $S(\mathscr{A})$. Moreover $\sigma^\omega(\mathscr{A})$ is invariant under $\partial R^{\mathscr{A}}$ and $R^{\mathscr{A}}$:

$$\overline{\sigma^\omega(\mathscr{A})} = \overline{\sigma^\infty(\mathscr{A})} = S(\mathscr{A}) \wedge R^{\mathscr{A}}(g)\sigma^\omega(\mathscr{A}) \subseteq \sigma^\omega(\mathscr{A}). \quad (3.7)$$

With these preparations our main result reads as follows.

Theorem 3.1: Let G be simply connected Lie group and $\mathfrak{g}$ its Lie algebra with basis $\{X_1,\ldots,X_n\}$. Assume G is an infinitesimal symmetry group (in the sense of Definition 3.1) with the properties (I) for every $\mathscr{A} \in I_G$ and $p \in \mathbb{N}^0$ the representation $\delta^{\mathscr{A}}$: $\mathfrak{g} \to \mathrm{End}\, E^{\mathscr{A}}$ is integrable in the Hilbert space $H_p(\mathscr{A})$; and (II) for any $p\in\mathbb{N}^0$ either $\delta^{\mathscr{A}}(X_i)$; $i = 1,\ldots,n$ are skew-symmetric operators in $H_p(\mathscr{A})$ (or $\delta^{\mathscr{A}}$ is topologically irreducible and has at least one p-analytic vector in $E^{\mathscr{A}}$) or there exists a subspace $\Omega_p^{s(p)}(\mathscr{A}) \subset H_p(\mathscr{A})$ that is p-entire for $\delta^{\mathscr{A}}$. Then G is a global symmetry group; this means there exists for every $\mathscr{A}\in I_G$ a unique continuous representation $R^{\mathscr{A}}$: $G \to \mathrm{Aut}\, S(\mathscr{A})$ of G and in addition an η-isometric (in general unbounded) representation U: $G\to \mathrm{Aut}\,\mathscr{D}_{\mathrm{QL}}$ on the dense subspace $\mathscr{D}_{\mathrm{QL}} \subseteq \mathscr{H}$ satisfying the conditions (U.1)–(U.5). Moreover $U(g)$ is strongly continuous in g.

The proof of Theorem 3.1 is split into two parts [integration in $S(\mathscr{A}) = \cap_{p\in\mathbb{N}^0} H_p(\mathscr{A})$ and the remainder], which are presented in Secs. IV and V, respectively.

We remark in passing that at least in theories, in which either the vacuum Ψ_0 is an eigenvector of the metric operator η or η has an inverse, the η-isometry of U together with the conditions (U.1) and (U.2) imply that the linear operators $U(g)$, $g\in G$ are closable in $\mathscr{H}$ (see paper I, Theorem 4.1).

The fact that only the representations of *strict* global or infinitesimal symmetries lift to corresponding unitary, respectively skew-symmetric, representations in the physical Hilbert space $\mathscr{H}_{\mathrm{phys}} = \overline{H/H_0}$ creates a particular interest in the relation between them, i.e., the relation between the conditions (2.15) and (3.3) defining strict symmetries. The problem is if the differential ∂U of a strict global symmetry (G, U) is a strict infinitesimal one [i.e., $\mathscr{D}^\infty(U)$ and ∂U satisfy condition (3.3)], vice versa, if the integral U of a strict infinitesimal symmetry $(\mathfrak{g}, \delta U)$ is a strict global one [i.e., $\mathscr{D}_\Pi$ and $U(g) \upharpoonright \mathscr{D}_\Pi$ satisfy (2.15)].

It will be shown in Sec. VI (Theorem 6.1) that the pair $(\mathfrak{g}, \partial U)$ with ∂U the differential of a strict global symmetry (G, U) is always a strict infinitesimal symmetry. On the other hand, the integral U of a strict infinitesimal symmetry is in general *not* a strict global one. The reason is that even in the case of bounded representations the closure of a nontrivial and ∂U-invariant subspace F of $\mathscr{D}^\infty(U)$ does not need to be invariant under U (Ref. 14, the example on page 255) unless F consists of analytic vectors for ∂U (Ref. 14, Proposition 4.4.5.6 and Ref. 22, Proposition 1). Hence the best one can hope for is to find a suitable additional condition on ∂U that is both sufficient and necessary for the integral of a strict infinitesimal symmetry to be a strict global one. In Sec. VI (Theorem 6.3) such a condition is derived. It states that $\mathscr{D}^\infty(U)\cap H$ contains a dense subspace of analytic vectors for ∂U that is invariant under ∂U.

IV. INTEGRATION IN THE COUNTABLY NORMED SPACE

In the first step we show that the integrability of $\delta^{\mathscr{A}}$ in the Hilbert space $H_p(\mathscr{A})$, together with the first or second alternative in assumption (II) of Theorem 3.1, implies the third one. Hence with assumptions (I) and (II) of Theorem 3.1, there exists for every $p\in\mathbb{N}^0$ a p-entire subspace for $\delta^{\mathscr{A}}$.

Lemma 4.1: Let G be a simply connected Lie group, $\mathfrak{g}$ its Lie algebra with basis $\{X_1,\ldots, X_n\}$, and δ: $\mathfrak{g}\to\mathrm{End}\,E$ an integrable representation of $\mathfrak{g}$ on a dense subspace E of the Hilbert space H_p. Assume that δ shares at least one of the following properties: (A) $\delta(X_i)$, $i = 1,\ldots,n$ are skew-symmetric operators in H_p; (B) δ is topologically irreducible in H_p and has at least one (p-) analytic vector in E; or (C) the integral R_p of δ in H_p is topologically irreducible. Then there exists a subspace $\Omega_p^s \subseteq H_p$ that is p-entire for the differential δR_p and in case of (A) or (B) is also p-entire for δ. In addition, Ω_p^s is invariant under ∂R_p, respectively under ${}^p\bar\delta$.

Proof: (A) Since all $\delta(X_i)$, $i = 1,\ldots, n$ are skew-symmetric in H_p, the integral R_p of δ is a unitary representation. Then the Laplace operator $\Delta := \Sigma_{i=1}^n \partial R_p(X_i)^2$ is essentially self-adjoint on σ_p^∞ (Ref. 14, Theorem 4.4.4.3) and moreover we have (Ref. 14, Theorem 4.4.4.5)

$$\sigma_p^\infty = \bigcap_{m=1}^{\infty} \mathrm{Dom}(({}^p\bar\Delta)^m). \quad (4.1)$$

Due to results of Nelson (Ref. 14, Lemma 4.4.6.7 and Ref. 15, Lemmas 5.2 and 6.2), this implies the existence of a dense subspace $\Omega_p^s \subseteq \sigma_p^\infty$, $s>0$, such that for all $f\in\Omega_p^s$, all $m\in\mathbb{N}$ and $i_1,\ldots,i_m \in\{1,\ldots,n\}$, we have

$$|\partial R_p(X_{i_1})\cdot \cdots \cdot \partial R_p(X_{i_m}) f|_p \leqslant M(f) S^m m! . \quad (4.2)$$

Hence Ω_p^s is p-entire for ∂R_p. Moreover from Lemma 9.1 or Theorem 5 of Ref. 15 it follows that E is a core for all ${}^p\overline{\partial R_p(X_i)}$, $i = 1,\ldots,n$:

$${}^p\overline{\partial R_p(X_i)} = {}^p\overline{\partial R_p(X_i)\upharpoonright E} = {}^p\overline{\delta(X_i)}. \quad (4.3)$$

Therefore Ω_p^s is also p-entire for δ. We have to demonstrate that it is invariant under ${}^p\bar\delta$. Following an idea of Goodman[21] we define for $m\in\mathbf{N}$ and $f\in\sigma_p^\infty$,

$$\rho_m(f):=\sup\{|\partial R_p(X_{i_1})\cdot\cdots\cdot\partial R_p(X_{i_m})f|_p \mid i_1,\dots,i_m\in\{1,\dots,n\}\}, \tag{4.4}$$

and for $s>0$,

$$G_s:=\left\{f\in\sigma_p^\infty \,\middle|\, \sum_{i=1}^{\infty}(m!)^{-1}t^m\rho_m(f)<+\infty \text{ for all } t\in]0,s[\right\}. \tag{4.5}$$

By means of elementary calculations it follows at once that

$$\Omega_p^s=G_{1/s}. \tag{4.6}$$

For an arbitrary element $X=\Sigma_{i=1}^n\alpha_iX_i$, $\alpha_i\in\mathbf{R}$ we define a norm relative to the basis of $\mathfrak{g}$ by $|X|:=\Sigma_{i=1}^n|\alpha_i|$. Then Eq. (4.4) implies for all $k\in\mathbf{N}$

$$\rho_m(\partial R_p(X)^kf)\leqslant|X|^k\rho_{m+k}(f). \tag{4.7}$$

From this inequality it is obvious that G_s and thus Ω_p^s is invariant under $\partial R_p(X)$. In view of Eq. (4.3), the same is true for ${}^p\overline{\delta(X)}$.

(B) Let $h\in E$ be an analytic vector for δ. Then for some $s(h)>0$ and all $m\in\mathbf{N}$ and $i_1,\dots,i_m\in\{1,\dots,n\}$ we have

$$|\delta(X_{i_1})\cdot\cdots\cdot\delta(X_{i_m})f|_p\leqslant s(h)^m m!.$$

Hence the subspace $\Omega_p^{s(h)}$ of all vectors from E satisfying for all $m\in\mathbf{N}$ and $i_1,\dots,i_m\in\{1,\dots,n\}$ the bound

$$|\delta(X_{i_1})\cdot\cdots\cdot\delta(X_{i_m})f|_p\leqslant M(f)s(h)^m m! \tag{4.8}$$

is nontrivial. Its invariance under δ follows by literally the same arguments as in case (A) if we replace σ_p^∞ by E and ∂R_p by δ. Since δ is topologically irreducible, $\Omega_p^{s(h)}$ is dense in H_p and therefore p-entire for δ.

(C) Once again a result due to Nelson (Ref. 14, Theorem 4.4.5.7 and Ref. 15, Theorem 4), implies the existence of a dense subspace $\sigma_p^\omega\subseteq\sigma_p^\infty$ of (p-) analytic vectors for the differential ∂R_p of R_p. Taking one of them, say $h\in\sigma_p^\omega$, one constructs in exactly the same way as in case (B), respectively (A), a nontrivial subspace $\Omega_p^{s(h)\prime}\subseteq\sigma_p^\omega$, which is invariant under ∂R_p and the elements of which satisfy the inequality (4.2) [for $s=s(h)$]. Then Proposition 1 in Ref. 15 or Proposition 4.4.5.6 in Ref. 14 deliver the invariance of its closure ${}^p\overline{\Omega_p^{s(h)}}$ under R_p. Because R_p is topologically irreducible, this closure is equal to H_p and therefore $\Omega_p^{s(h)}$ p-entire for ∂R_p. □

The next theorem establishes the integrability of $\delta^{\mathscr{A}}$ in the countably normed space $S(\mathscr{A})=\cap_{p\in\mathbf{N}^0}H_p(\mathscr{A})$.

Theorem 4.1: Let G be a simply connected Lie group, $\mathfrak{g}$ its Lie algebra with basis $\{X_1,\dots,X_n\}$, and $\delta^{\mathscr{A}}:\mathfrak{g}\to\mathrm{End}\,E^{\mathscr{A}}$ a representation of $\mathfrak{g}$ on a dense subspace $E^{\mathscr{A}}$ of $S(\mathscr{A})=\cap_{p\in\mathbf{N}^0}H_p(\mathscr{A})$. Assume the linear operators $\delta^{\mathscr{A}}(X)$, $X\in\mathfrak{g}$, are closable and presumptions (I) and (II) of Theorem 3.1 hold for $\delta^{\mathscr{A}}$. Then there exists a unique continuous representation $R^{\mathscr{A}}:G\to\mathrm{Aut}\,S(\mathscr{A})$ with the properties

(1) $\sigma^\infty(\mathscr{A})=\bigcap_{p\in\mathbf{N}^0}\sigma_p^\infty(\mathscr{A})$,

(2) $E^{\mathscr{A}}\subseteq\sigma^\infty(\mathscr{A})$,

(3) $\forall_{X\in\mathfrak{g}},\quad \delta^{\mathscr{A}}(X)=\partial R^{\mathscr{A}}(X)\restriction E^{\mathscr{A}}$.

Proof: The main step in the proof is to show that $S(\mathscr{A})$ is invariant under the representations $R_p^{\mathscr{A}}$ of G in $H_p(\mathscr{A})$ and their restrictions $R_p^{\mathscr{A}}\restriction S(\mathscr{A})$ coincide for all $p\in\mathbf{N}^0$. For this it suffices to prove that for all $p\in\mathbf{N}$ and $g\in G$ we have

$$R_p^{\mathscr{A}}(g)=R_{p-1}^{\mathscr{A}}(g)\restriction H_p(\mathscr{A}). \tag{4.9}$$

The mapping

$$|\cdot|:\ \mathfrak{g}\to\mathbf{R},\quad X=\sum_{i=1}^n t_iX_i\to|X|:=\sum_{i=1}^n|t_i| \tag{4.10}$$

defines a norm on the $\mathbf{R}$-vector space $(\mathfrak{g},+,\cdot)$ and the exponential mapping exp: $\mathfrak{g}\to G$ maps some open neighborhood $N(\mathscr{O})$ of the zero element in $\mathfrak{g}$ diffeomorphically onto an open neighborhood $\mathscr{M}(e)$ of the unit element in G (see Ref. 29, Chap. VIII, §6, Proposition 15). According to assumption (II) (of Theorem 3.1) and Lemma 4.1, there exists a subspace $\Omega_p^s(\mathscr{A})\subseteq\sigma_p^\infty(\mathscr{A})$, $s>0$, which is p-entire for $\delta^{\mathscr{A}}$. Next we make use of the inequality

$$\forall_{h\in H_p(\mathscr{A})},\quad |h|_{p-1}\leqslant|h|_p,\quad p\in\mathbf{N}, \tag{4.11}$$

which implies

$$\forall_{X\in\mathfrak{g}},\quad {}^p\overline{\delta^{\mathscr{A}}(X)}\subseteq{}^{(p-1)}\overline{\delta^{\mathscr{A}}(X)}. \tag{4.12}$$

Since $H_p(\mathscr{A})$ is dense in $H_{p-1}(\mathscr{A})$, it follows from the last two relations that $\Omega_p^s(\mathscr{A})$ is also $(p-1)$-entire for $\delta^{\mathscr{A}}$. Finally we observe that due to Definition 3.2, $\delta^{\mathscr{A}}(X)=\partial R_p^{\mathscr{A}}(X)\restriction E^{\mathscr{A}}=\partial R_{p-1}^{\mathscr{A}}(X)\restriction E^{\mathscr{A}}$, for all $X\in\mathfrak{g}$. Hence for all X from the open neighborhood

$$N_s(\mathscr{O})=\{X\in N(\mathscr{O})\mid |X|<1/s\}$$

of the zero element in $\mathfrak{g}$ and all $f\in\Omega_p^s(\mathscr{A})$, the series

$$\sum_{m=1}^{\infty}\frac{1}{m!}\sum_{i_1,\dots,i_m\in\{1,\dots,n\}}t_{i_1}\cdot t_{i_2}\cdot\cdots\cdot t_{i_m}\times{}^p\overline{\delta^{\mathscr{A}}(X_{i_1})}\cdot\cdots\cdot{}^p\overline{\delta^{\mathscr{A}}(X_{i_m})}f \tag{4.13}$$

converges strongly in $H_p(\mathscr{A})$ to $R_p^{\mathscr{A}}(\exp X)f$ and due to (4.12) in $H_{p-1}(\mathscr{A})$ to $R_{p-1}^{\mathscr{A}}(\exp X)f$. Thus we obtain for all g from the open neighborhood

$$\mathscr{M}_s(e):=\{\exp X\mid X\in N_s(\mathscr{O})\} \tag{4.14}$$

of the unit element in G all $f\in\Omega_p^s(\mathscr{A})$

$$R_p^{\mathscr{A}}(g)f=R_{p-1}^{\mathscr{A}}(g)f. \tag{4.15}$$

For any $h\in H_p(\mathscr{A})$, there exists a sequence $(f_n)_{n\in\mathbf{N}}$ with $f_n\in\Omega_p^s(\mathscr{A})$ and $\lim_{n\to\infty}|f_n-h|_p=0$, and in view of the inequality (4.11) $\lim_{n\to\infty}|f_n-h|_{p-1}=0$. Next consider the norm

$$\begin{aligned}|R_p^{\mathscr{A}}(g)h-R_{p-1}^{\mathscr{A}}(g)h|_p&\leqslant|R_p^{\mathscr{A}}(g)h-R_p^{\mathscr{A}}(g)f_n|_p\\&\quad+|R_p^{\mathscr{A}}(g)f_n-R_{p-1}^{\mathscr{A}}(g)h|_p.\end{aligned} \tag{4.16}$$

Due to the continuity of $R_p^{\mathscr{A}}(g)$ [$g\in\mathscr{M}_s(e)$ fixed] in $H_p(\mathscr{A})$ the first term on the right-hand side of (4.16) converges to zero if n goes to infinity and furthermore the sequence $(R_p^{\mathscr{A}}(g)f_n-R_{p-1}^{\mathscr{A}}(g)h)_{n\in\mathbf{N}}$ is a Cauchy sequence in the norm $|\cdots|_p$. But then, due to the inequality (4.11), it is also a Cauchy sequence in the norm $|\cdots|_{p-1}$. In addition we have

$$\lim_{n\to\infty} |R_p^{\mathscr{A}}(g) f_n - R_{p-1}^{\mathscr{A}}(g) h|_{p-1}$$

$$= \lim_{n\to\infty} |R_{p-1}^{\mathscr{A}}(g) f_n - R_{p-1}^{\mathscr{A}}(g) h|_{p-1} = 0,$$

because $R_{p-1}^{\mathscr{A}}(g)$ is a continuous operator in $H_{p-1}(\mathscr{A})$. Since the norms defining the topology of $S(\mathscr{A})$ are pairwise compatible, the last three properties imply (Ref. 23, Chap. I)

$$\lim_{n\to\infty} |R_p^{\mathscr{A}}(g) f_n - R_{p-1}^{\mathscr{A}}(g) h|_p = 0.$$

This proves that Eq. (4.15) holds for all $h \in H_p(\mathscr{A})$ and all g from the open neighborhood $\mathscr{M}_s(e)$ of the unit element in G and therefore for all $g \in G$, because G is connected. Thus we have established Eq. (4.9) for all $p \in \mathbb{N}$ and $g \in G$, from which in turn we deduce

$$\forall_{p\in\mathbb{N}, g\in G}, \quad R_p^{\mathscr{A}}(g) \upharpoonright S(\mathscr{A}) = R_{p-1}^{\mathscr{A}}(g) \upharpoonright S(\mathscr{A}). \tag{4.17}$$

Now we define the representation $R^{\mathscr{A}}$ by

$$R^{\mathscr{A}}: G \to \mathrm{Aut}\, S(\mathscr{A}), \quad g \to R^{\mathscr{A}}(g) := R_p^{\mathscr{A}}(g) \upharpoonright S(\mathscr{A}). \tag{4.18}$$

In view of Eq. (4.17), it is continuous in the topology of $S(\mathscr{A}) = S(\mathbb{R}_4, \mathbb{C}_{\mathscr{A}})$.

Now Lemma 6.1 of paper I is applicable. Thus the property (1) is a direct consequence of Lemma 6.1(e).[11] Moreover the relation (4.12), respectively Lemma 6.1 of paper I, imply for all $p \in \mathbb{N}^0$ and $X \in \mathfrak{g}$

$$\sigma_{p+1}^\infty(\mathscr{A}) \subseteq \sigma_p^\infty(\mathscr{A}) \ \wedge \ \partial R_{p+1}^{\mathscr{A}}(X) = \partial R_p^{\mathscr{A}} \upharpoonright \sigma_{p+1}^\infty(\mathscr{A}). \tag{4.19}$$

Let us note in passing that these relations are direct consequences of Eq. (4.9) and the definitions of $\sigma_p^\infty(\mathscr{A})$ and $\partial R_p^{\mathscr{A}}(X)$ (see paper I Sec. VI). Now from presumption (I) and Definition 3.2, we obtain $E^{\mathscr{A}} \subseteq \sigma_p^\infty(\mathscr{A})$ for all $p \in \mathbb{N}^0$. Therefore property (2) follows immediately from (1) and the first part of the relation (4.19). Finally, from the definitions of the differentials $\partial R^{\mathscr{A}}$ and $\partial R_p^{\mathscr{A}}$, property (1), and relations (4.19), we deduce for every $p \in \mathbb{N}^0$ and $X \in \mathfrak{g}$

$$\partial R^{\mathscr{A}}(X) = \partial R_p^{\mathscr{A}}(X) \upharpoonright \sigma^\infty(\mathscr{A}). \tag{4.20}$$

But then property (2) in combination with presumption (I) and Definition 3.2 imply property (3).

It remains to prove the uniqueness of the representation $R^{\mathscr{A}}$. Assume there exists a further continuous representation $\tilde{R}^{\mathscr{A}}: G \to \mathrm{Aut}\, S(\mathscr{A})$ with properties (1), (2), and (3). Denote by $\tilde{R}_p^{\mathscr{A}}$ its continuous extension to the Hilbert space $H_p(\mathscr{A})$ and by $\tilde{\sigma}^\infty(\mathscr{A})$, $\tilde{\sigma}_p^\infty(\mathscr{A})$, $\partial\tilde{R}^{\mathscr{A}}$, and $\partial\tilde{R}_p^{\mathscr{A}}$ the corresponding subspaces of $\mathscr{C}^\infty$-vectors, respectively the differentials. Then according to Lemma 6.1 of paper I, relations (4.9) and (4.19) hold also for them. They in turn imply in combination with the properties (2) and (3), for all $p \in \mathbb{N}^0$,

$$E^{\mathscr{A}} \subseteq \tilde{\sigma}_p^\infty(\mathscr{A}) \ \wedge \ \forall_{X\in\mathfrak{g}}, \delta^{\mathscr{A}}(X) = \partial\tilde{R}_p^{\mathscr{A}}(X) \upharpoonright E^{\mathscr{A}}.$$

However, this is in contradiction to the fact that the integral of $\delta^{\mathscr{A}}$ in the Hilbert space $H_p(\mathscr{A})$ is unique. □

V. UNBOUNDED REPRESENTATIONS IN $\mathscr{H}$

For the construction of the (in general) unbounded representation of G in the unphysical Hilbert space, we need first of all some properties of the extended tensor products of the representations $R^{\mathscr{A}}$, their continuous extensions in a single norm $|\cdots|_p = \sqrt{[\cdots,\cdots]_p}$ onto the corresponding Hilbert spaces $H_p(\mathscr{A}_1,\ldots,\mathscr{A}_L)$, the differentials of all of them and, last but not least, the p-entire subspaces for the differentials.

On the completed tensor products $S(\mathscr{A}_1,\ldots,\mathscr{A}_L) = \bar{\otimes}_{i=1}^L S(\mathscr{A}_i)$ of the nuclear spaces $S(\mathscr{A})$ the extended tensor products $R^{\mathscr{A}_1\cdots\mathscr{A}_L}(g) := \bar{\otimes}_{i=1}^L R^{\mathscr{A}_i}(g)$ define a continuous representation $R^{\mathscr{A}_1\cdots\mathscr{A}_L}: G \to \mathrm{Aut}\, S(\mathscr{A}_1,\ldots,\mathscr{A}_L)$, $g \to R^{\mathscr{A}_1\cdots\mathscr{A}_L}(g)$ of G (paper I, Theorem 3.1[11]). Moreover the principle of uniform boundedness (Ref. 30, Theorem V.7) implies that $R^{\mathscr{A}_1\cdots\mathscr{A}_L}$ is uniformly bounded on some neighborhood of the unit element of G. Since $R^{\mathscr{A}_1\cdots\mathscr{A}_L}$ is a continuous representation, it possesses a dense invariant subspace $\sigma^\infty(\mathscr{A}_1,\ldots,\mathscr{A}_L)$ of $\mathscr{C}^\infty$-vectors and its differential defines a representation $\partial R^{\mathscr{A}_1\cdots\mathscr{A}_L}$: $\mathfrak{g} \to \mathrm{End}\, \sigma^\infty(\mathscr{A}_1,\ldots,\mathscr{A}_L)$ of the Lie algebra $\mathfrak{g}$ of G. For all $f \in \sigma^\infty(\mathscr{A}_1,\ldots,\mathscr{A}_L)$ and $h \in \sigma^\infty(\mathscr{B}_1,\ldots,\mathscr{B}_N)$ we have

$$\partial R^{\mathscr{A}_1\cdots\mathscr{A}_L\mathscr{B}_1\cdots\mathscr{B}_N}(x)(f\otimes h) = (\partial R^{\mathscr{A}_1\cdots\mathscr{A}_L}(X)f)\otimes h + f\otimes(\partial R^{\mathscr{B}_1\cdots\mathscr{B}_N}(X)h). \tag{5.1}$$

As already mentioned in Sec. II, in the nuclear spaces $S(\mathscr{A}_1,\ldots,\mathscr{A}_L)$ we have at least three equivalent topologies, the projective (π-) topology, the inductive (ϵ-) topology, and a further one defined by a countable set of norms $|\cdots|_p = \sqrt{[\cdots,\cdots]_p}$, which are derived from scalar products. If $H_p(\mathscr{A}_1,\ldots,\mathscr{A}_L) = {}^p\overline{S(\mathscr{A}_1,\ldots,\mathscr{A}_L)}$ denotes the completion of $S(\mathscr{A}_1,\ldots,\mathscr{A}_L)$ in the norm $|\cdots|_p$ [which should not be confused, for instance, with the completion in a projective (π-) norm], then $S(\mathscr{A}_1,\ldots,\mathscr{A}_L)$ is once again the intersection of the countable set of Hilbert spaces $H_p(\mathscr{A}_1,\ldots,\mathscr{A}_L)$, $p \in \mathbb{N}^0$ (see Sec. II). For every $g \in G$ there exist a unique continuous extension $R_p^{\mathscr{A}_1\cdots\mathscr{A}_L}(g)$ of $R^{\mathscr{A}_1\cdots\mathscr{A}_L}(g)$ onto $H_p(\mathscr{A}_1,\ldots,\mathscr{A}_L)$. These extensions define a strongly continuous representation $R_p^{\mathscr{A}_1\cdots\mathscr{A}_L}: G \to \mathrm{Aut}\, H_p(\mathscr{A}_1,\ldots,\mathscr{A}_L)$ of G. If $\sigma_p^\infty(\mathscr{A}_1,\ldots,\mathscr{A}_L)$ and $\partial R_p^{\mathscr{A}_1\cdots\mathscr{A}_L}$ denote the corresponding subspace of $\mathscr{C}^\infty$-vectors, respectively the differential of $R_p^{\mathscr{A}_1\cdots\mathscr{A}_L}$, then Eq. (4.1) holds also for them. Moreover from Lemma 6.1 of paper I, we know

$$\sigma_{p+1}^\infty(\cdots) \subseteq \sigma_p^\infty(\cdots) \ \wedge \ \sigma^\infty(\cdots) \bigcap_{n\in\mathbb{N}^0} \sigma_n^\infty(\cdots), \tag{5.2}$$

$$\partial R_{p+1}^{\cdots}(X) = \partial R_p^{\cdots}(X) \upharpoonright \sigma_{p+1}^\infty(\cdots), \tag{5.3}$$

$$\partial R^{\cdots}(X) = \partial R_p^{\cdots}(X) \upharpoonright \sigma^\infty(\cdots). \tag{5.4}$$

Let $\textcircled{\pi}_{i=1}^L H_p(\mathscr{A}_i)$ denote the projective tensor product of the normed spaces $H_p(\mathscr{A}) = H_p(\mathbb{R}_4, \mathbb{C}_{\mathscr{A}})$; i.e., the completion of the algebraic tensor product $\otimes_{i=1}^L H_p(\mathscr{A}_i)$ in the (projective) π-norm (Ref. 25, Chap. 43)

$$\|f\|_p^\pi := \inf\left\{\left|\sum_t \prod_{i=1}^L |h_i^t|_p \right| \sum_t \otimes_{i=1}^L h_i^t = f\right\}. \tag{5.5}$$

For all $f\in\otimes_{i=1}^{L}S(\mathscr{A}_i)\subset S(\mathscr{A}_1,...,\mathscr{A}_L)$ it plainly follows that

$$|f|_p\leq\|f\|_p^{\pi}. \tag{5.6}$$

Since $S(\mathscr{A})$ is dense in $H_p(\mathscr{A})$, the last two relations imply the inclusions

$$S(\mathscr{A}_1,...,\mathscr{A}_L)\subseteq\bigotimes_{i=1}^{L}H_p(\mathscr{A}_i)\subseteq H_p(\mathscr{A}_1,...,\mathscr{A}_L), \tag{5.7}$$

and every member of this chain is dense in its successor(s).

With these preparations it is not hard to prove the following lemma.

Lemma 5.1: Let G be a simply connected Lie group, $\mathfrak{g}$ its Lie algebra with basis $\{X_1,...,X_n\}$, and $\delta^{\mathscr{A}_j}$: $\mathfrak{g}\to\mathrm{End}\,E^{\mathscr{A}_j}$, $j=1,...,L$, representations of $\mathfrak{g}$ on the dense subspaces

$$E^{\mathscr{A}_j}\subseteq S(\mathscr{A}_j)=\bigcap_{n\in\mathbb{N}^0}H_n(\mathscr{A}_j).$$

Assume that for every $j\in\{1,...,n\}$ and $p\in\mathbb{N}^0$, $\delta^{\mathscr{A}_j}$ is integrable in $H_p(\mathscr{A}_j)$ and $H_p(\mathscr{A}_j)$ contains a subspace $\Omega_p^{s_j}(\mathscr{A}_j)$, $s_j=s_j(p)>0$, which is p-entire and invariant for $\delta^{\mathscr{A}_j}$. Then we have the following.

(i) The tensor product

$$\Omega_p^{s}(\mathscr{A}_1,...,\mathscr{A}_L):=\bigotimes_{j=1}^{L}\Omega_p^{s_j}(\mathscr{A}_j),$$

with $s=\Sigma_{j=1}^{L}s_j$ is contained in $\sigma_p^{\infty}(\mathscr{A}_1,...,\mathscr{A}_L)$ and p-entire for $\partial R_p^{\mathscr{A}_1\cdots\mathscr{A}_L}$; i.e., it is dense in $H_p(\mathscr{A}_1,...,\mathscr{A}_L)$ and all its elements satisfy, for every $m\in\mathbb{N}$ and $i_1,...,i_m\in\{1,...,n\}$, the bounds

$$\left|\partial R_p^{\mathscr{A}_1\cdots\mathscr{A}_L}(X_{i_1})\cdot\cdots\cdot\partial R_p^{\mathscr{A}_1\cdots\mathscr{A}_L}(X_{i_m})f\right|_p\leq M(f)s^m m!, \tag{5.8}$$

with $s=\Sigma_{j=1}^{n}s_j$ independent of f.

(ii) $\Omega_p^{s}(\mathscr{A}_1,...,\mathscr{A}_L)$ is invariant under $\partial R_p^{\mathscr{A}_1\cdots\mathscr{A}_L}$.

Proof: According to Ref. 21, Theorem 1.1 or Ref. 11, Theorem 6.1 it is true that

$$\sigma_p^{\infty}(\mathscr{A})=\bigcap_{m\in\mathbb{N}}\bigcap_{i\in\{1,...,n\}}\mathrm{Dom}((^{p}\,\overline{\partial R_p^{\mathscr{A}}(X_i)})^m). \tag{5.9}$$

Since $^{p}\,\overline{\delta^{\mathscr{A}}(X)}\subseteq{}^{p}\,\overline{\partial R_p^{\mathscr{A}}(X)}$, it follows from Eq. (5.9) and Definition 3.3 that $\Omega_p^{s}(\mathscr{A})\subseteq\sigma_p^{\infty}(\mathscr{A})$ and therefore

$$\bigotimes_{i=1}^{L}\Omega_p^{s_i}(\mathscr{A}_i)\subseteq\bigotimes_{i=1}^{L}\sigma_p^{\infty}(\mathscr{A}_i)\subset\sigma_p^{\infty}(\mathscr{A}_1,...,\mathscr{A}_L). \tag{5.10}$$

For the remainder of the proof we can restrict ourselves, without loss of generality, to the case $L=2$. The general case then easily follows via complete induction from the associativity of the tensor product and the literal repetition of the arguments below.

Consider an arbitrary element $f=\Sigma_{t=1}^{N}f_1^t\otimes f_2^t$ with $f_j^t\in\Omega_p^{s_j}(\mathscr{A}_j)$, $j=1,2$. Since $\Omega_p^{s_j}(\mathscr{A}_j)$ is p-entire for $\partial R_p^{\mathscr{A}_j}$ we have for all $m\in\mathbb{N}$; all $i_1,...,i_m\in\{1,...,n\}$, and some $\hat{M}(f_j^t)\in\mathbb{R}^+$:

$$\left|\left[\prod_{k=1}^{m}\partial R_p^{\mathscr{A}_j}(X_{i_k})\right]f_j^t\right|_p\leq\hat{M}(f_j^t)(s_j)^m m!. \tag{5.11}$$

From this inequality and the analog of Eq. (5.1) for $\partial R_p^{\mathscr{A}_1\mathscr{A}_2}$ and $\sigma_p^{\infty}(\mathscr{A}_1,\mathscr{A}_2)$ we deduce, by some elementary combinatorical gymnastics,

$$\begin{aligned}\left|\left[\prod_{k=1}^{m}\partial R_p^{\mathscr{A}_1\cdots\mathscr{A}_L}(X_{i_k})\right]f\right|_p&\leq\sum_{t=1}^{N}\Big\{\hat{M}(f_1^t)\,|f_2^t|_p(s_1)^m m!\\&\quad+\hat{M}(f_1^t)\hat{M}(f_2^t)\sum_{\nu=1}^{m-1}\binom{m}{\nu}(s_1)^{m-\nu}(s_2)^{\nu}(m-\nu)!\nu!\\&\quad+|f_1^t|_p\hat{M}(f_2^t)(s_2)^m m!\Big\}.\end{aligned} \tag{5.12}$$

This leads immediately to the inequality (5.8), if we define $s:=s_1+s_2$ and

$$M(f):=\sum_{t=1}^{N}\max\{\hat{M}(f_{2-r}^t)|f_r^t|_p,\ \hat{M}(f_1^t)\hat{M}(f_2^t)|r=1,2\}. \tag{5.13}$$

It remains to prove that $\Omega_p^{s}(\mathscr{A}_1,\mathscr{A}_2)$ is dense in $H_p(\mathscr{A}_1,\mathscr{A}_2)$. Due to relation (5.7), $H_p(\mathscr{A}_1)\otimes H_p(\mathscr{A}_2)$ is dense in $H_p(\mathscr{A}_1,\mathscr{A}_2)$. Consider an element $h=\Sigma_{t=1}^{N}h_1^t\otimes h_2^t\in H_p(\mathscr{A}_1)\otimes H_p(\mathscr{A}_2)$. Since $\Omega_p^{s_j}(\mathscr{A}_j)$ is dense in $H_p(\mathscr{A}_j)$, there exists for every $\epsilon>0$ an element $f_j^t\in\Omega_p^{s_j}(\mathscr{A}_j)$ with

$$|h_j^t-f_j^t|_p<\epsilon. \tag{5.14}$$

Now the element $f:=\Sigma_{t=1}^{N}f_1^t\otimes f_2^t$ is from $\Omega_p^{s}(\mathscr{A}_1,\mathscr{A}_2)$. By an elementary calculation we gain

$$\begin{aligned}h-f=\sum_{t=1}^{N}\{&(f_1^t-h_1^t)\otimes(h_2^t-f_2^t)\\&+h_1^t\otimes(h_2^t-f_2^t)+(h_1^t-f_1^t)\otimes h_2^t\}.\end{aligned} \tag{5.15}$$

The last two relations imply

$$|h-f|_p\leq\epsilon^2N+\epsilon\sum_{t=1}^{N}(|h_1^t|_p+|h_2^t|_p). \tag{5.16}$$

This proves part (i). Part (ii) follows directly from the corresponding property of the constituents of $\Omega_p^{s}(\mathscr{A}_1,...,\mathscr{A}_L)$. □

For the proof of the following theorem we need the antilinear and isometric (with respect to every norm $|\cdots|_p$, $p\in\mathbb{N}^0$) bijection

$$\mathscr{C}_L:\ S(\mathscr{A}_1,...,\mathscr{A}_L)\to S(\mathscr{A}_L^*,...,\mathscr{A}_1^*)\equiv S(\mathscr{A}_L,...,\mathscr{A}_1),$$

which is generated by

$$\mathscr{C}_L(\otimes_{i=1}^{L}f_i):=\otimes_{i=L}^{1}\bar{f}_i.$$

Here, $\mathscr{C}_L=\mathscr{C}_L(\mathscr{A}_1,...,\mathscr{A}_L)$, and possesses a unique isometric extension onto every Hilbert space $H_p(\mathscr{A}_1,...,\mathscr{A}_L)$, which we again denote by $\mathscr{C}_L$. By direct calculations, it is easy to verify the following equations:

$$\begin{aligned}&\mathscr{C}_L\sigma_p^{\infty}(\mathscr{A}_1,...,\mathscr{A}_L)=\sigma_p^{\infty}(\mathscr{A}_L,...,\mathscr{A}_1),\\&\mathscr{C}_L\Omega_p^{s}(\mathscr{A}_1,...,\mathscr{A}_L)=\Omega_p^{s}(\mathscr{A}_L,...,\mathscr{A}_1),\end{aligned} \tag{5.17}$$

$$\phi^{\mathscr{A}_1\cdots\mathscr{A}_L}(f)^*=\phi^{\mathscr{A}_L^*\cdots\mathscr{A}_1^*}(\mathscr{C}_Lf),\quad f\in S(\mathscr{A}_1,...,\mathscr{A}_L), \tag{5.18}$$

$$\mathscr{C}_L R_p^{\mathscr{A}_L\cdots\mathscr{A}_1}(g)\mathscr{C}_L^{-1} f = \overline{R_p^{\mathscr{A}_1\cdots\mathscr{A}_L}(g)\bar{f}},$$
$$f\in H_p(\mathscr{A}_1,\ldots\mathscr{A}_L),$$
$$\mathscr{C}_L \partial R_p^{\mathscr{A}_L\cdots\mathscr{A}_1}(X)\mathscr{C}_L^{-1} f = \overline{\partial R_p^{\mathscr{A}_1\cdots\mathscr{A}_L}(X)\bar{f}},$$
$$f\in\sigma_p(\mathscr{A}_1,\ldots,\mathscr{A}_L). \tag{5.19}$$

Theorem 5.1: Let G be a simply connected Lie group, $\mathfrak{g}$ its Lie algebra with basis $\{X_1,\ldots,X_n\}$ and $\delta^{\mathscr{A}}: \mathfrak{g}\to\text{End}\, E^{\mathscr{A}}$ a representation of $\mathfrak{g}$ on the dense subspace $E^{\mathscr{A}}\subseteq S(\mathscr{A})$, $\mathscr{A}\in I_G$. Assume that G is an infinitesimal symmetry group for which the conditions (I) and (II) of Theorem 3.1 hold. Then there exists an η-isometric (in general unbounded) representation U: $G\to\text{Aut}\,\mathscr{D}_{\text{QL}}$, $g\to U(g)$ with the properties (U.1)–(U.5) (in Sec. II resp. Sec. III). Moreover $U(g)$ is strongly continuous in the group elements g.

Proof: (a) Existence, continuity in g, and conditions (U.1) and (U.2). On the generating vectors of $\mathscr{D}_{\text{QL}}$ we define U by

$$U(g)\Psi_0 := \Psi_0,$$
$$U(g)\,\phi^{\mathscr{A}_1\cdots\mathscr{A}_L}(f)\Psi_0 := \phi^{\mathscr{A}_1\cdots\mathscr{A}_L}(R^{\mathscr{A}_1\cdots\mathscr{A}_L}(g)f)\Psi_0, \tag{5.20}$$
$$f\in S(\mathscr{A}_1,\ldots,\mathscr{A}_L),\quad \mathscr{A}_1,\ldots,\mathscr{A}_L\in I_G,\quad L\in\mathbb{N},$$

and on the remaining ones by linear extension. Then obviously U is a representation of G on $\mathscr{D}_{\text{QL}}$ since all $R^{\mathscr{A}_1\cdots\mathscr{A}_L}$ are representations of G on $S(\mathscr{A}_1,\ldots,\mathscr{A}_L)$. The domain and covariance conditions (U.1) and (U.2) (in Sec. II) are trivial consequences of these definitions. The strong continuity of $U(g)$ in g follows from Theorem 3.2(2) in paper I (to be more precise it follows by literal repetition of the proof of part (2) of Theorem 3.2).

(b) η-isometry (condition (U.3) in Sec. II). In order to establish the η-isometry of U on $\mathscr{D}_{\text{QL}}$ it suffices to prove that for all $f\in S(\mathscr{A}_1,\ldots,\mathscr{A}_L)$, all $h\in S(\mathscr{B}_1,\ldots,\mathscr{B}_N)$ and all g from an open neighborhood $\mathscr{M}(e)$ of the unit element $e\in G$ we have

$$\Big(\phi^{\mathscr{A}_1\cdots\mathscr{A}_L}(f)\Psi_0,\ \eta\,\phi^{\mathscr{B}_1\cdots\mathscr{B}_N}(R^{\mathscr{B}_1\cdots\mathscr{B}_N}(g^{-1})h)\Psi_0\Big) - \Big(\phi^{\mathscr{A}_1\cdots\mathscr{A}_L}(R^{\mathscr{A}_1\cdots\mathscr{A}_L}(g)f)\Psi_0,\ \eta\,\phi^{\mathscr{B}_1\cdots\mathscr{B}_N}(h)\Psi_0\Big) = 0. \tag{5.21}$$

Since $\mathscr{D}_{\text{QL}}$ is contained in D and D is invariant under η, the map

$$B:\ S(\mathscr{A}_L,\ldots,\mathscr{A}_1)\times S(\mathscr{B}_1,\ldots,\mathscr{B}_N)\to\mathbb{C},$$
$$(f,h)\to B(f,h) = (\Psi_0,\phi^{\mathscr{A}_L^\dagger\cdots\mathscr{A}_1^\dagger}(f)\eta\,\phi^{\mathscr{B}_1\cdots\mathscr{B}_N}(h)\Psi_0) \tag{5.22}$$

is a separately continuous bilinear functional and thus by virtue of Schwartz's nuclear theory it has a unique extension to a continuous linear functional $W^{\mathscr{A}_L^\dagger\cdots\mathscr{B}_N}$ on

$$S(\mathscr{A}_L^\dagger,\ldots,\mathscr{B}_N) = S(\mathscr{A}_L^\dagger,\ldots,\mathscr{A}_1^\dagger)\,\bar{\otimes}\,S(\mathscr{B}_1,\ldots,\mathscr{B}_N)$$

with $W^{\mathscr{A}_L^\dagger\cdots\mathscr{B}_N}(f\otimes h) = B(f,h)$. Now Eq. (5.21) is equivalent to the following one for all $f\in S(\mathscr{A}_1,\ldots,\mathscr{A}_L)$, $h\in S(\mathscr{B}_1,\ldots,\mathscr{B}_N)$, and $g\in\mathscr{M}(e)$:

$$W^{\mathscr{A}_L^\dagger\cdots\mathscr{B}_N}\Big((\mathscr{C}_L f)\otimes(R^{\mathscr{B}_1\cdots\mathscr{B}_N}(g^{-1})h) - (\mathscr{C}_L R^{\mathscr{A}_1\cdots\mathscr{A}_L}(g)f)\otimes h\Big) = 0. \tag{5.23}$$

Since every continuous linear functional is of finite order, there exists (Ref. 23, Chap. I, §41 and 42) a minimal $p_0\in\mathbb{N}^0$ and a unique continuous linear functional $W_{p_0}^{\mathscr{A}_L^\dagger\cdots\mathscr{B}_N}$ on the normed space $H_{p_0}(\mathscr{A}_L^\dagger,\ldots,\mathscr{B}_N)$ such that, for all $f\in S(\mathscr{A}_L^\dagger,\ldots,\mathscr{B}_N)$,

$$W_{p_0}^{\mathscr{A}_L^\dagger\cdots\mathscr{B}_N}(f) = W^{\mathscr{A}_L^\dagger\cdots\mathscr{B}_N}(f). \tag{5.24}$$

Let $|X|$, $X\in\mathfrak{g}$ denote the norm on $\mathfrak{g}$ relative to the basis $\{X_1,\ldots,X_n\}$ explicitly given in relation (4.10). According to Proposition 15 in Ref. 30, Chap. VIII, §6 and the comments to it, there exists an open neighborhood $Q(\mathscr{O})$ of the zero element in $\mathfrak{g}$ such that the exponential mapping $\exp\mathfrak{g}\to G$ maps $Q(\mathscr{O})$ diffeomorphically onto an open neighborhood $\mathscr{M}(e) = \exp[Q(\mathscr{O})]$ of the unit element in G with the property

$$g = \exp X\in\mathscr{M}(e)\Rightarrow g^{-1}\in\mathscr{M}(e)\ \wedge\ g^{-1} = \exp(-X). \tag{5.25}$$

For $s>0$ we denote by $Q_s(\mathscr{O})$ and $\mathscr{M}_s(e)$ the open (sub-) neighborhoods $Q_s(\mathscr{O}) := \{X\in Q(\mathscr{O})\mid |X|<1/s\}$, respectively $\mathscr{M}_s(e) := \exp[Q_s(\mathscr{O})]$.

In virtue of the presumptions of Theorem 5.1 (resp. Theorem 3.1) we can apply Lemma 4.1 and Lemma 5.1. Hence there exist subspaces

$$\Omega_{p_0}^s(\mathscr{A}_1,\ldots,\mathscr{A}_L) = \otimes_{j=1}^L \Omega_{p_0}^{s_j}(\mathscr{A}_j\subset H_{p_0}(\mathscr{A}_1,\ldots,\mathscr{A}_L),$$
$$s = \sum_{j=1}^L s_j > 0,$$
$$\Omega_{p_0}^t(\mathscr{B}_1,\ldots,\mathscr{B}_N) = \otimes_{j=1}^N \Omega_{p_0}^{t_j}(\mathscr{B}_j)\subset H_{p_0}(\mathscr{B}_1,\ldots,\mathscr{B}_N),$$
$$t = \sum_{j=1}^N t_j > 0, \tag{5.26}$$

which are p_0-entire for $\partial R_{p_0}^{\mathscr{A}_1\cdots\mathscr{A}_L}$, respectively $\partial R_{p_0}^{\mathscr{B}_1\cdots\mathscr{B}_N}$, and in addition invariant under these differentials. Furthermore relations (5.7) and (5.18) imply

$$(\mathscr{C}_L\Omega_{p_0}^s(\mathscr{A}_1,\ldots,\mathscr{A}_L))\otimes\Omega_{p_0}^t(\mathscr{B}_1,\ldots,\mathscr{B}_N)$$
$$\subset H_{p_0}(\mathscr{A}_L^\dagger,\ldots,\mathscr{A}_1^\dagger)\,\hat{\otimes}_\pi\, H_{p_0}(\mathscr{B}_1,\ldots,\mathscr{B}_N)$$
$$\subseteq H_{p_0}(\mathscr{A}_L^\dagger,\ldots,\mathscr{B}_N), \tag{5.27}$$

and every member of this chain is dense in its successor(s).

Let $g = \exp X$, $X = \Sigma_{i=1}^n t_i X_i$, be an arbitrary element of $\mathscr{M}_r(e)$ with $r := \max\{s,t\}$. Then for every $f\in\Omega_{p_0}^{\cdots}(\cdots)$, the infinite series

$$R_{p_0}^{\cdots}(\exp X)f = \sum_{m=1}^\infty \frac{1}{m!}\sum_{i_1,\ldots,i_m\in\{1,\ldots,n\}} t_{i_1}\cdot t_{i_2}\cdot\cdots\cdot t_{i_m} \times\partial R_{p_0}^{\cdots}(X_{i_1})\cdot\cdots\cdot\partial R_{p_0}^{\cdots}(X_{i_m})f \tag{5.28}$$

converges in the p_0-norm. This in turn implies by virtue of the continuity of the isometric operator $\mathscr{C}_L$ and the linear functional $W_{p_0}^{\cdots}$ that the following equation holds for all $f\in\Omega_{p_0}^s(\mathscr{A}_1,\ldots,\mathscr{A}_L)$, $h\in\Omega_{p_0}^t(\mathscr{B}_1,\ldots,\mathscr{B}_N)$, and $g = \exp X\in\mathscr{M}_r(e)$:

$$W_{p_0}^{\mathscr{A}\mathfrak{t}\cdots\mathscr{B}_N}\Big((\mathscr{C}_L f)\otimes\Big(R_{p_0}^{\mathscr{B}_1\cdots\mathscr{B}_N}(\exp(-X))h\Big)-(\mathscr{C}_L R^{\mathscr{A}_1\cdots\mathscr{A}_L}(\exp X)f)\otimes h\Big)$$
$$=-\lim_{\rho\to\infty}\sum_{m=1}^{\rho}\frac{1}{m!}\Big\{\sum_{i_1,\dots,i_m\in\{1,\dots,n\}} t_{i_1}\cdot t_{i_2}\cdot\cdots\cdot t_{i_m}\Big[W_{p_0}^{\mathscr{A}\mathfrak{t}\cdots\mathscr{B}_N}\Big(\mathscr{C}_L\Big[\prod_{j=1}^{m}\partial R_{p_0}^{\mathscr{A}_1\cdots\mathscr{A}_L}(X_{i_j})\Big]f\Big)\otimes h$$
$$-(-1)^m W_{p_0}^{\mathscr{A}\mathfrak{t}\cdots\mathscr{B}_N}\Big((\mathscr{C}_L f)\otimes\Big(\Big[\prod_{j=1}^{m}\partial R_{p_0}^{\mathscr{B}_1\cdots\mathscr{B}_N}(X_{i_j})\Big]h\Big)\Big)\Big]\Big\}. \tag{5.29}$$

We are going to show by means of the η-skew-symmetry of $\delta U(X)$ on $\mathscr{D}_E\subseteq\mathscr{D}_\Pi$ that the *wavy bracket* vanishes for every $m\in\mathbb{N}$. First of all, it follows from Eqs. (5.1), (5.2), and part (3) of Theorem 4.1 that for all $X\in\mathfrak{g}$ and $f_j\in E^{\mathscr{A}_j}$, $j=1,\dots,L$, we have

$$\partial R_{p_0}^{\mathscr{A}_1\cdots\mathscr{A}_L}(X)\Big(\otimes_{j=1}^{L} f_j\Big)=\sum_{i=1}^{L} f_1\otimes\cdots\otimes f_{i-1}\otimes\delta^{\mathscr{A}_i}(X)f_i\otimes f_{i+1}\otimes\cdots\otimes f_L. \tag{5.30}$$

The Eqs. (3.2) and the η-skew-symmetry of $\delta U(X)$, together with Eqs. (5.22) and (5.24) imply for all

$$f=\otimes_{j=1}^{L} f_j\in E(\mathscr{A}_1,\dots,\mathscr{A}_L):=\otimes_{j=1}^{L}E^{\mathscr{A}_j},$$

all

$$h=\otimes_{j=1}^{L} h_j\in E(\mathscr{B}_1,\dots,\mathscr{B}_N):=\otimes_{j=1}^{N}E^{\mathscr{B}_j}$$

and

$X_i\in\mathfrak{g}$; $i=1,\dots,n$;

$$W_{p_0}^{\mathscr{A}\mathfrak{t}\cdots\mathscr{B}_N}\big(\big[\mathscr{C}_L\partial R_{p_0}^{\mathscr{A}_1\cdots\mathscr{A}_L}(X_i)f\big]\otimes h-(\mathscr{C}_L f)\otimes\big[\partial R_{p_0}^{\mathscr{B}_1\cdots\mathscr{B}_N}(X_i)h\big]\big)=0. \tag{5.31}$$

In the next step we extend the validity of Eq. (5.31) to all elements of the form $a=\otimes_{i=1}^{L} a_i\in\Omega_{p_0}^{s}(\mathscr{A}_1,\dots,\mathscr{A}_L)$, respectively $b=\otimes_{j=1}^{N} b_j\in\Omega_{p_0}^{t}(\mathscr{B}_1,\dots,\mathscr{B}_N)$. Due to Definition 3.3 we have $\Omega_{p_0}^{s_j}(\mathscr{A}_j)\subseteq\mathrm{Dom}(\overline{{}^{p_0}\delta^{\mathscr{A}_j}(X_i)})$ for all $i=1,\dots,n$. Hence there exists [Ref. 31, Chap. III, §5(3)] a sequence $(f_j^k)_{k\in\mathbb{N}}$ with $f_j^k\in E^{\mathscr{A}_j}=\mathrm{Dom}(\delta^{\mathscr{A}_j}(X))$, $X\in\mathfrak{g}$ such that

$$\lim_{k\to\infty}|a_j-f_j^k|_{p_0}=0,$$
$$\lim_{k\to\infty}\Big|\overline{{}^{p_0}\delta^{\mathscr{A}_j}(X_i)}a_j-\delta^{\mathscr{A}_j}(X_i)f_j^k\Big|_{p_0}=0. \tag{5.32}$$

With the abbreviation $f^n:=\otimes_{i=1}^{L} f_j^n$ we deduce from Eqs. (5.30) and (5.32) via elementary calculations with tensor products

$$\lim_{k\to\infty}|a-f^k|_{p_0}=0\ \wedge\ \lim_{k\to\infty}\big|\partial R_{p_0}^{\mathscr{A}_1\cdots\mathscr{A}_L}(X_i)(a-f^k)\big|_{p_0}=0. \tag{5.33}$$

Denote by $(h^k)_{k\in\mathbb{N}}$, $h^k=\otimes_{i=1}^{N} h_i^k\in E(\mathscr{B}_1,\dots,\mathscr{B}_N)$ the corresponding sequence for $b=\otimes_{i=1}^{N} b_i$. By virtue of the linearity and continuity of the functional $W_{p_0}^{\cdots}$, the following inequality is true for every $k\in\mathbb{N}$ and some $\omega>0$:

$$|W_{p_0}^{\mathscr{A}\mathfrak{t}\cdots\mathscr{B}_N}(\big[\mathscr{C}_L\partial R_{p_0}^{\mathscr{A}_1\cdots\mathscr{A}_L}(X_i)a\big]\otimes b-(\mathscr{C}_L a)\otimes\big[\partial R_{p_0}^{\mathscr{B}_1\cdots\mathscr{B}_N}(X_i)b\big])|$$
$$\leqslant\omega\{|\big[\mathscr{C}_L\partial R_{p_0}^{\mathscr{A}_1\cdots\mathscr{A}_L}(X_i)a\big]\otimes b-\big[\mathscr{C}_L\partial R_{p_0}^{\mathscr{A}_1\cdots\mathscr{A}_L}(X_i)f^k\big]\otimes h^k|_{p_0}$$
$$+|(\mathscr{C}_L a)\otimes\big[\partial R_{p_0}^{\mathscr{B}_1\cdots\mathscr{B}_N}(X_i)b\big]-(\mathscr{C}_L f^k)\otimes\big[\partial R_{p_0}^{\mathscr{B}_1\cdots\mathscr{B}_N}(X_i)h^k\big]|_{p_0}\}. \tag{5.34}$$

Observing the isometry of the operator $\mathscr{C}_L$ we easily get the further estimate for the first term on the right-hand side:

$$|\big[\mathscr{C}_L\partial R_{p_0}^{\mathscr{A}_1\cdots\mathscr{A}_L}(X_i)a\big]\otimes b-\big[\mathscr{C}_L\partial R_{p_0}^{\mathscr{A}_1\cdots\mathscr{A}_L}(X_i)f^k\big]\otimes h^k|_{p_0}$$
$$\leqslant|\partial R_{p_0}^{\mathscr{A}_1\cdots\mathscr{A}_L}(X_i)(a-f^k)|_{p_0}\cdot|b-h^k|_{p_0}+|\partial R_{p_0}^{\mathscr{A}_1\cdots\mathscr{A}_L}(X_i)a|_{p_0}\cdot|b-h^k|_{p_0}$$
$$+|b|_{p_0}\cdot|\partial R_{p_0}^{\mathscr{A}_1\cdots\mathscr{A}_L}(X_i)(a-f^k)|_{p_0}, \tag{5.35}$$

and a similar one for the second term. Taking the limit $k\to\infty$ we deduce from relations (5.33)–(5.35) that Eq. (5.31) remains true for all elements of the form $a=\otimes_{j=1}^{L} a_j\in\Omega_{p_0}^{s}(\mathscr{A}_1,\dots,\mathscr{A}_L)$ and $b=\otimes_{j=1}^{N} b_j\in\Omega_{p_0}^{t}(\mathscr{B}_1,\dots,\mathscr{B}_N)$. However, then it follows via elementary calculations with tensor products using the linearity of $W_{p_0}^{\cdots}$ and the antilinearity of $\mathscr{C}_L$ that Eq. (5.31) is also true for all linear combinations of such elements, this means for all $a\in\Omega_{p_0}^{s}(\mathscr{A}_1,\dots,\mathscr{A}_L)$ and all $b\in\Omega_{p_0}^{t}(\mathscr{B}_1,\dots,\mathscr{B}_N)$. According to part (ii) of Lemma 5.1 these spaces are invariant under $\partial R_{p_0}^{\mathscr{A}_1\cdots\mathscr{A}_L}$, respectively $\partial R_{p_0}^{\mathscr{B}_1\cdots\mathscr{B}_N}$. Therefore by repeated applications of Eq. (5.31), we obtain for all $f\in\Omega_{p_0}^{s}(\mathscr{A}_1,\dots,\mathscr{A}_L)$, all $h\in\Omega_{p_0}^{t}(\mathscr{B}_1,\dots,\mathscr{B}_N)$, and all $m\in\mathbb{N}$,

$$W_{p_0}^{\mathscr{A}\mathfrak{t}\cdots\mathscr{B}_N}\Big(\Big[\mathscr{C}_L\Big(\prod_{j=1}^{m}\partial R_{p_0}^{\mathscr{A}_1\cdots\mathscr{A}_L}(X_{i_j})\Big)f\Big]\otimes h\Big)=(-1)^m W_{p_0}^{\mathscr{A}\mathfrak{t}\cdots\mathscr{B}_N}\Big((\mathscr{C}_L f)\otimes\Big[\Big(\prod_{j=m}^{1}\partial R_{p_0}^{\mathscr{B}_1\cdots\mathscr{B}_N}(X_{i_j})\Big)h\Big]\Big). \tag{5.36}$$

Feeding this result back into Eq. (5.29), we see that the *wavy brackets* on the right-hand side vanish for every $m\in\mathbb{N}$. Hence, the equation

$$W_{p_0}^{\mathscr{A}\mathfrak{t}\cdots\mathscr{B}_N}\Big((\mathscr{C}_L f)\otimes\big[R_{p_0}^{\mathscr{B}_1\cdots\mathscr{B}_N}(g^{-1})h\big]-\big[\mathscr{C}_L R_{p_0}^{\mathscr{A}_1\cdots\mathscr{A}_L}(g)f\big]\otimes h\Big)=0$$

holds for $g\in\mathscr{M}_r(e)$, all $f\in\Omega_{p_0}^{s}(\mathscr{A}_1,\dots,\mathscr{A}_L)$, and all $h\in\Omega_{p_0}^{t}(\mathscr{B}_1,\dots,\mathscr{B}_N)$. Because these spaces are dense in $H_{p_0}(\mathscr{A}_1,\dots,\mathscr{A}_L)$, respectively $H_{p_0}(\mathscr{B}_1,\dots,\mathscr{B}_N)$, we can use once again the continuity properties of $R^{\cdots}_{p_0}$, $\mathscr{C}_L$, and $W^{\cdots}_{p_0}$ to establish the validity of Eq. (5.37) for all $f\in H_{p_0}(\mathscr{A}_1,...,\mathscr{A}_L)$ and all $h\in H_{p_0}(\mathscr{B}_1,...,\mathscr{B}_N)$. Finally in view of the relations $S(\cdots)\subset H_{p_0}(\cdots)$ and $R^{\cdots}(g)=R^{\cdots}_{p_0}(g)\restriction S(\cdots)$, Eqs. (5.37) and (5.24) imply that conditions (5.23), or equivalenty (5.21) hold true for all g from the open neighborhood $\mathscr{M}_r(e)$. This proves the η-isometry of $U(g)$ on $\mathscr{D}_{\rm QL}$ for all $g\in\mathscr{M}_r(e)$, and therefore for all $g\in G$, because G is simply connected.

(c) Conditions (U.4) and (U.5). In view of the continuity of $U(g)$ in g, the properties (U.4) and (U.5) (explicitly stated in Sec. III) follow in a straightforward way from Theorem 4.1, and the results of Sec. V in paper I, especially from Theorem 5.1 and statements (I)–(IV)[11] on the structural properties of the differential $\partial R^{\mathscr{A}_1\cdots\mathscr{A}_L}$ of the continuous representation $R^{\mathscr{A}_1\cdots\mathscr{A}_L}$.

This completes the proof of Theorem 5.1. □

VI. STRICT GLOBAL AND INFINITESIMAL SYMMETRIES

In this final section we will investigate the interrelation between strict global and strict infinitesimal symmetries, which are singled out by the additional condition (2.15), respectively (3.3). We first show that the differential of every strict global symmetry is a strict infinitesimal one.

Theorem 6.1: Let G be a global symmetry (Lie) group with Lie algebra $\mathfrak{g}$, U its η-isometric representation on $\mathscr{D}_{\rm QL}\supset\mathscr{D}_{\Pi}$, $\mathscr{D}^{\infty}(U)$ the dense subspace of C^{∞}-vectors for U, and ∂U the differential of U. If the pair (G, U) is a strict global symmetry, i.e., $\forall_{g\in G}$, $U(g)(\mathscr{D}_{\Pi}\cap H)\subseteq\mathscr{D}_{\Pi}\cap H$, then the pair $(\mathfrak{g}, \partial U)$ is a strict infinitesimal symmetry, i.e., it satisfies the conditions

$$\overline{D^{\infty}(U)\cap H}=H\ \wedge\ \forall_{X\in\mathfrak{g}},\partial U(X)(D^{\infty}(U)\cap H)\subseteq\mathscr{D}^{\infty}(U)\cap H. \tag{6.1}$$

Proof: the subspace $\mathscr{D}_{\Pi}\cap H$ is dense in $\mathscr{D}_{\rm QL}\cap H$ and in view of the covariance condition (U.2) the subspace $\mathscr{D}_{\rm QL}$ is invariant under U. Then the invariance of $\mathscr{D}_{\Pi}\cap H$, the covariance condition (U.2b), the continuity of the representations $R^{\mathscr{A}_1\cdots\mathscr{A}_L}$, and the weak closedness of the Hilbert space H imply (see the proof of Corollary 3.2 in paper I) the invariance of $\mathscr{D}_{\rm QL}\cap H$,

$$\forall_{g\in G},\quad U(g)(\mathscr{D}_{\rm QL}\cap H)\subseteq(\mathscr{D}_{\rm QL}\cap H). \tag{6.2}$$

Now the second one of the conditions (6.1) is a direct consequence of the definition of $\mathscr{D}^{\infty}(U)$ in Eq. (2.16) and of the differential ∂U,

$$\partial U(X)=\operatorname*{s-lim}_{t\to 0}t^{-1}[U(\exp tX)\Psi-\Psi],\quad \Psi\in\mathscr{D}^{\infty}(U). \tag{6.3}$$

The subspace of $\mathscr{C}^{\infty}$-vectors $\mathscr{D}^{\infty}(U)\subseteq\mathscr{D}_{\rm QL}$ is invariant under $\partial U(X)$, $X\in\mathfrak{g}$ and $U(g)$, $g\in G$. Therefore if $\Psi\in\mathscr{D}^{\infty}(U)\cap H$, then for every $t\neq 0$ and $X\in\mathfrak{g}$ the vector $t^{-1}[U(\exp tX)\Psi-\Psi]$ is from $\mathscr{D}^{\infty}\cap H$. However this implies the same to be true for the limit because H is a closed subspace of $\mathscr{H}$.

From the linearity of the field operators $\phi^{\cdots}(f)$ it follows that the set

$$\tilde{S}(\mathscr{A}_1,...,\mathscr{A}_L):=\{f\in S(\mathscr{A}_1,...,\mathscr{A}_L)\mid\phi^{\mathscr{A}_1\cdots\mathscr{A}_L}(f)\Psi_0\in H\} \tag{6.4}$$

is a subspace of $S(\mathscr{A}_1,...,\mathscr{A}_L)$. Since H is a closed subspace of $\mathscr{H}$ the continuity of $\phi^{\cdots}(f)$ in f implies that $\tilde{S}(\mathscr{A}_1,...,\mathscr{A}_L)$ is a *complete, countably normed space*. Moreover due to relation (6.2) and the covariance condition (U.2), the space $\tilde{S}(\mathscr{A}_1,...,\mathscr{A}_L)$ is invariant under $R^{\mathscr{A}_1\cdots\mathscr{A}_L}$. Thus, the restrictions of $R^{\mathscr{A}_1\cdots\mathscr{A}_L}(g)$, $g\in G$, to $\tilde{S}(\mathscr{A}_1,...,\mathscr{A}_L)$ define a continuous representation

$$\tilde{R}^{\mathscr{A}_1\cdots\mathscr{A}_L}\colon G\to\operatorname{Aut}\tilde{S}(\mathscr{A}_1,...,\mathscr{A}_L),$$
$$g\to R^{\mathscr{A}_1\cdots\mathscr{A}_L}(g)\restriction\tilde{S}(\mathscr{A}_1,...,\mathscr{A}_L) \tag{6.5}$$

of G on the complete, countably normal space $\tilde{S}(\mathscr{A}_1,...,\mathscr{A}_L)$. $\tilde{R}^{\mathscr{A}_1\cdots\mathscr{A}_L}$ possesses a dense invariant subspace $\tilde{\sigma}^{\infty}(\mathscr{A}_1,...,\mathscr{A}_L)$ of $\mathscr{C}^{\infty}$-vectors (Ref. 14, Chap. 4, §4.4). Plainly we have $\tilde{\sigma}^{\infty}(\mathscr{A}_1,...,\mathscr{A}_L)\subset\sigma^{\infty}(\mathscr{A}_1,...,\mathscr{A}_L)$ and moreover

$$\mathscr{D}^{\infty}(U)\cap H=\mathrm{LH}\{\Psi_0,\phi^{\mathscr{A}_1\cdots\mathscr{A}_L}(f)\Psi_0\mid f\in\tilde{\sigma}^{\infty}(\mathscr{A}_1,...,\mathscr{A}_L);\ \mathscr{A}_1,...,\mathscr{A}_L\in I_G;\ L\in\mathbb{N}\},$$

respectively (6.6)

$$\mathscr{D}_{\rm QL}\cap H=\mathrm{LH}\{\Psi_0,\phi^{\mathscr{A}_1\cdots\mathscr{A}_L}(f)\Psi_0\mid f\in\tilde{S}(\mathscr{A}_1,...,\mathscr{A}_L);\ \mathscr{A}_1,...,\mathscr{A}_L\in I_G;\ L\in\mathbb{N}\}.$$

Now obviously $\mathscr{D}^{\infty}(U)\cap H$ is dense in $\mathscr{D}_{\rm QL}\cap H$ because $\tilde{\sigma}^{\infty}(\cdots)$ is dense in $\tilde{S}(\cdots)$; $\mathscr{D}_{\rm QL}\cap H$ contains $\mathscr{D}_{\Pi}\cap H$ and by assumption A.IV of Sec. II, the latter is dense in H. Therefore $\mathscr{D}^{\infty}(U)\cap H$ is dense in H. □

From the arguments given at the end of Sec. III it is clear that the converse of Theorem 6.1 cannot be true in general. These arguments also indicate that we have to look for some additional conditions in terms of analytic vectors. In the next step we prove that the integral of a strict infinitesimal symmetry is a strict global one if for every $p\in\mathbb{N}^0$ and $\mathscr{A}\in I_G$ the subspace

$$\tilde{\Omega}^{\omega}_p(\mathscr{A}):=\sigma^{\omega}_p(\mathscr{A})\cap\{f\in S(\mathscr{A})\mid f\in E^{\mathscr{A}}\ \wedge\ \phi^{\mathscr{A}}(f)\Psi_0\in H\} \tag{6.7}$$

of p-analytic vectors for $\delta^{\mathscr{A}}$ is dense in the subspace $\tilde{S}(\mathscr{A})=\{f\in S(\mathscr{A})\mid\phi^{\mathscr{A}}(f)\Psi_0\in H\}$. Roughly, this means that sufficiently many of the analytic vectors from $\mathscr{D}^{\infty}(U)$ lie in H. Due to the relation $f\in E^{\mathscr{A}}$ in Eq. (6.7), this condition will in general not be a necessary one. However, as will be seen below, a slight change leads to a necessary and sufficient condition. On the other hand, the above one seems to be more suitable for the applications.

Theorem 6.2: Let G be a simply connected Lie group with Lie algebra $\mathfrak{g}$, $\delta^{\mathscr{A}}$: $\mathfrak{g}\to\operatorname{End}E^{\mathscr{A}}$, $\mathscr{A}\in I_G$, and δU: $\mathfrak{g}\to\operatorname{End}\mathscr{D}_E$ representations of $\mathfrak{g}$ satisfying the conditions of Theorem 3.1, $R^{\mathscr{A}}_p$ the integral of $\delta^{\mathscr{A}}$ in $H_p(\mathscr{A})$ and $\sigma^{\omega}_p(\mathscr{A})$ the dense invariant subspace of analytic vectors for $\partial R^{\mathscr{A}}_p$. If the pair $(\mathfrak{g}, \delta U)$ is a strict infinitesimal symmetry and in addition for every $p\in\mathbb{N}^0$ and $\mathscr{A}\in I_G$ the subspace $\tilde{\Omega}^{\omega}_p$ of p-analytic vectors for $\delta^{\mathscr{A}}$ is dense in $\tilde{S}(\mathscr{A})$, then the pair $(G,$

U) with U the integral of δU in $\mathscr{H}$ is a strict global symmetry.

Proof: Let $\tilde{H}_p(\cdots) \subset H_p(\cdots)$ be the completion in the norm $|\cdots|_p$ of the complete countably normed space $\tilde{S}(\cdots)$ defined in Eq. (6.4). From the completeness of $\tilde{S}(\cdots)$ it follows (Ref. 23, Chap. I, §3.2) that

$$\tilde{S}(\mathscr{A}_1,...,\mathscr{A}_L) = \bigcap_{p_0 \in \mathbb{N}^0} \tilde{H}_p(\mathscr{A}_1,...,\mathscr{A}_L). \qquad (6.8)$$

With the abreviation

$$\tilde{E}^{\mathscr{A}} := E^{\mathscr{A}} \cap \tilde{S}(\mathscr{A}) = \{f \in S(\mathscr{A}) \mid f \in E^{\mathscr{A}} \ \wedge \ \phi^{\mathscr{A}}(f)\Psi_0 \in H\}, \qquad (6.9)$$

we obtain from Eq. (3.1)

$$\mathscr{D}_E \cap H = \mathrm{LH}\{\Psi_0, \phi^{\mathscr{A}_1 \cdots \mathscr{A}_L}(f)\Psi_0 \mid f \in \otimes_{i=1}^{L} \tilde{E}^{\mathscr{A}_i}; \ \mathscr{A}_1,...,\mathscr{A}_L \in I_G; \ L \in \mathbb{N}\}. \qquad (6.10)$$

In the first step we prove that $\tilde{H}_p(\mathscr{A})$ is invariant under $R_p^{\mathscr{A}}(g)$ for all $g \in G$. By virtue of the invariance condition

$$\forall_{X \in \mathfrak{g}}, \quad \delta U(X)(\mathscr{D}_E \cap H) \subseteq \mathscr{D}_E \cap H, \qquad (6.11)$$

the covariance condition (3.2), and Eq. (6.10), it follows that for every $f \in \tilde{E}^{\mathscr{A}}$ and $X \in \mathfrak{g}$ the vector

$$\delta U(X)\, \phi^{\mathscr{A}}(f)\Psi_0 = \phi^{\mathscr{A}}(\partial R^{\mathscr{A}}(X) f)\Psi_0 = \phi^{\mathscr{A}}(\partial R_p^{\mathscr{A}}(X) f)\Psi_0$$

is from $\mathscr{D}_E \cap H$ and therefore $\partial R_p^{\mathscr{A}}(X) f$ from $\tilde{E}^{\mathscr{A}}$. Thus $\tilde{E}^{\mathscr{A}}$ is invariant under $\partial R_p^{\mathscr{A}}$ and the same is then true for $\tilde{\Omega}_p^{\omega}(\mathscr{A}) = \sigma_p^{\omega}(\mathscr{A}) \cap \tilde{E}^{\mathscr{A}}$ because the subspace $\sigma_p^{\omega}(\mathscr{A})$ of all p-analytic vectors for $\partial R_p^{\mathscr{A}}$ is invariant under $\partial R_p^{\mathscr{A}}$. Now Proposition 1 in Ref. 22 (also Proposition 4.4.5.5 in Ref. 14) states that the closure ${}^p\overline{\tilde{\Omega}_p^{\omega}(\mathscr{A})} = {}^p\overline{\tilde{S}(\mathscr{A})} = \tilde{H}_p(\mathscr{A})$ is invariant under $R_p^{\mathscr{A}}(g)$ for all $g \in G$.

In the second step we generalize this result to the extended tensor products $R_p^{\mathscr{A}_1 \cdots \mathscr{A}_L}(g)$. Plainly the invariance of $\tilde{\Omega}_p^{\omega}(\mathscr{A}_j)$ under $\partial R_p^{\mathscr{A}_j}$ implies the invariance of the algebraic tensor product $\tilde{\Omega}_p^{\omega}(\mathscr{A}_1,...,\mathscr{A}_L) := \otimes_{j=1}^{L} \tilde{\Omega}_p^{\omega}(\mathscr{A}_j)$ under $\partial R_p^{\mathscr{A}_1 \cdots \mathscr{A}_L}$, because for $f = \otimes_{j=1}^{L} f_j$ we have

$$\partial R_p^{\mathscr{A}_1 \cdots \mathscr{A}_L}(X) f = \sum_{i=1}^{L} f_1 \otimes \cdots \otimes f_{i-1} \otimes \partial R_p^{\mathscr{A}_i}(X) f_i \otimes f_{i+1} \otimes \cdots \otimes f_L. \qquad (6.12)$$

The proof that $\tilde{\Omega}_p^{\omega}(\mathscr{A}_1,...,\mathscr{A}_L)$ is dense in $\tilde{H}_p(\mathscr{A}_1,...,\mathscr{A}_L)$ and consists only of analytic vectors for $\partial R_p^{\mathscr{A}_1 \cdots \mathscr{A}_L}$ is literally the same as the proof of Lemma 5.1. We just have to put $\hat{M} = 1$ and replace $\Omega_p^s(\cdots)$ by $\tilde{\Omega}_p^{\omega}(\cdots)$, $H_p(\cdots)$ by $\tilde{H}_p(\cdots)$, and s_j by $s_j(f,p)$. Now we can aply once again Proposition 1 of Ref. 22 (also Proposition 4.4.5.6 of Ref. 14) and obtain the invariance of $\tilde{H}_p(\mathscr{A}_1,...,\mathscr{A}_L)$ under $R_p^{\mathscr{A}_1 \cdots \mathscr{A}_L}$. But then Eq. (6.8) together with

$$R^{\mathscr{A}_1 \cdots \mathscr{A}_L}(g) = R_p^{\mathscr{A}_1 \cdots \mathscr{A}_L}(g) \restriction \big(\bigcap_{p \in \mathbb{N}} H_p(\mathscr{A}_1,...,\mathscr{A}_L)\big)$$

implies that the space $\tilde{S}(\mathscr{A}_1,...,\mathscr{A}_L)$ is invariant under $R_p^{\mathscr{A}_1 \cdots \mathscr{A}_L}$. In turn, from the second of Eqs. (6.6) and the covariance condition (U.2), we deduce that $\mathscr{D}_{\mathrm{QL}} \cap H$ is invariant under U. Since $\mathscr{D}_{\Pi}$ is a U-invariant subspace of $\mathscr{D}_{\mathrm{QL}}$ the same is true for $\mathscr{D}_{\Pi} \cap H$. □

At a first glance it seems that in the proof of Theorem 6.2 we did not use the first part of condition (3.3), which defines a strict infinitesimal symmetry. However this is not true, since it is a necessary condition for $\tilde{\Omega}_p^{\omega}(\mathscr{A})$ to be dense in $\tilde{S}(\mathscr{A})$.

As we have already remarked, due to the restriction $f \in E^{\mathscr{A}}$ in the right-hand side of Eq. (6.7) the density of $\tilde{\Omega}_p^{\omega}(\mathscr{A})$ in $\tilde{S}(\mathscr{A})$ will in general not be a necessary condition. A necessary and sufficient condition is obtained if we enlarge the subspace $\tilde{\Omega}_p^{\omega}(\mathscr{A})$ by dropping this restriction and demand that for every $p \in \mathbb{N}^0$ and $\mathscr{A} \in I_G$ the subspace $\tilde{\sigma}_p^{\omega}(\mathscr{A}) := \sigma_p^{\omega}(\mathscr{A}) \cap \tilde{S}(\mathscr{A})$ is dense in $\tilde{S}(\mathscr{A})$ and invariant under $\partial R_p^{\mathscr{A}}$. If $\sigma^{\omega}(\mathscr{A}) \subseteq \sigma^{\infty}(\mathscr{A})$ denotes the dense invariant subspace of ∞-analytic vectors for $\partial R^{\mathscr{A}}$ (see Appendix A), then by virtue of the equations

$$\sigma^{\omega}(\mathscr{A}) = \bigcap_{n \in \mathbb{N}^0} \sigma_n^{\omega}(\mathscr{A}) \ \wedge \partial R^{\mathscr{A}}(X) = \partial R_p^{\mathscr{A}}(X) \restriction \sigma^{\infty}(\mathscr{A}), \qquad (6.13)$$

we arrive at the equivalent condition, that for every $\mathscr{A} \in I_G$ the subspace $\tilde{\sigma}^{\omega}(\mathscr{A}) := \sigma^{\omega}(\mathscr{A}) \cap \tilde{S}(\mathscr{A})$ of ∞-analytic vectors for $\partial R^{\mathscr{A}}$ is dense in $\tilde{S}(\mathscr{A})$ and invariant under $\partial R^{\mathscr{A}}$. This again means that the subspace $\mathscr{D}^{\infty}(U) \cap H$ contains a dense ∂U-invariant subspace of analytic vectors.

Theorem 6.3: Let $(G, \mathfrak{g}, \delta^{\mathscr{A}}, \delta U)$ be as in Theorem 6.2, $R^{\mathscr{A}}$ the integral of $\delta^{\mathscr{A}}$ in $S(\mathscr{A}) = S(\mathbb{R}_4, \mathbb{C}_{\dot{\mathscr{A}}})$ and $\sigma^{\omega}(\mathscr{A})$ the dense invariant subspace of analytic vectors for $\partial R^{\mathscr{A}}$. Assume that the pair $(\mathfrak{g}, \delta U)$ is a strict infinitesimal symmetry. Then the pair (G, U) with U the integral δU in $\mathscr{H}$ is a strict global symmetry if and only if for every $\mathscr{A} \in I_G$ the subspace $\tilde{\sigma}^{\omega}(\mathscr{A}) := \sigma^{\omega}(\mathscr{A}) \cap \tilde{S}(\mathscr{A})$ of analtyic vectors for $\partial R^{\mathscr{A}}$ is invariant under $\partial R^{\mathscr{A}}$ and dense in $\tilde{S}(\mathscr{A})$.

Proof: (1)$\Leftarrow$: For every $p \in \mathbb{N}^0$, $\tilde{\sigma}_p^{\omega}(\mathscr{A})$ is an invariant subspace of p-analytic vectors for $\partial R_p^{\mathscr{A}}$, which is dense in $\tilde{H}_p(\mathscr{A}) = {}^p\overline{\tilde{S}(\mathscr{A})}$. Then Proposition 1 in Ref. 22 (or Proposition 4.4.5.6 in Ref. 14) tells us that its closure ${}^p\overline{\tilde{\sigma}_p^{\omega}(\mathscr{A})} = \tilde{H}_p(\mathscr{A})$ is invariant under $R_p^{\mathscr{A}}(g)$, $g \in G$. For the remainder of the proof jump directly to the second step in the proof of Theorem 6.2.

(2)$\Rightarrow$: Assume that the pair (G, U) is a strict global symmetry. In the proof of Theorem 6.1 it has been shown that $\tilde{S}(\mathscr{A})$ is a complete, countably normed subspace of $S(\mathscr{A})$ that is invariant under $R^{\mathscr{A}}$. Therefore, the restrictions of $R^{\mathscr{A}}(g)$, $g \in G$, onto $\tilde{S}(\mathscr{A})$ define a continuous representation $\tilde{R}^{\mathscr{A}}: G \to \mathrm{Aut}\, \tilde{S}(\mathscr{A})$, $g \mapsto \tilde{R}^{\mathscr{A}}(g) \restriction \tilde{S}(\mathscr{A})$ of G on a complete, countably normed space. Its differential $\partial \tilde{R}^{\mathscr{A}}$: $\mathfrak{g} \to \mathrm{End}\, \tilde{\sigma}^{\infty}(\mathscr{A})$ exists (Ref. 14, Chap. 4, §4.4) and shares the properties $\tilde{\sigma}^{\infty}(\mathscr{A}) = \delta^{\infty}(\mathscr{A}) \cap \tilde{S}(\mathscr{A})$ is dense in $\tilde{S}(\mathscr{A})$, and moreover

$$\forall_{X \in \mathfrak{g}}, \quad \partial \tilde{R}^{\mathscr{A}}(X) = \partial R^{\mathscr{A}}(X) \restriction \tilde{\sigma}^{\infty}(\mathscr{A}). \qquad (6.14)$$

According to Appendix A and Ref. 14, Chap. 4, §4.4, there exists a dense subspace $\tilde{\sigma}^{\omega}(\mathscr{A}) \subseteq \tilde{\sigma}^{\infty}(\mathscr{A})$ of ∞-analytic vectors for $\partial \tilde{R}^{\mathscr{A}}$ that is invariant under $\partial \tilde{R}^{\mathscr{A}}$. Due to Eq. (6.14) it is also an invariant subspace of ∞-analytic vectors for $\partial R^{\mathscr{A}}$.

ACKNOWLEDGMENTS

The author is deeply indebted to Dr. D. Schmersau for several valuable discussions and hints.

APPENDIX A: NELSON'S THEOREM

In the investigation of strict symmetries in Sec. VI we needed the generalization of a well-known result of Nelson (Ref. 15, Theorem 4) to complete, countably normed spaces. Nelson's theorem states that a continuous representation of a group G in a Banach space has a dense invariant subspace of vectors, which are analytic for its differential.

Theorem: Let $S = \cap_{p\in\mathbb{N}^0} H_p$ with $H_p = {}^p\bar{S}$ be a complete countably normed space with pairwise compatible norms $\|\cdots\|_p$, $p\in\mathbb{N}^0$, R a continuous representation of a connected Lie group G (with Lie algebra $\mathfrak{g}$) on S and R_p the continuous extension of R onto H_p. Denote by σ^∞ (σ_p^∞) and $\partial R(\partial R_p)$ the dense subspace of $\mathscr{C}^\infty$-vectors respectively the differential of $R(R_p)$. Then the subspace $\sigma^\omega = \cap_{p\in\mathbb{N}^0} \sigma_p^\omega \subseteq \sigma^\infty$ of all ∞-analytic vectors for ∂R is dense in σ^∞ and invariant under R and ∂R.

Proof: The proof follows directly from the Gårding–Nelson trick (Ref. 14, §4.4.5) by means of which the analogous result is obtained in a Banach space. With minor modifications we retain the definitions and notations used in Ref. 14, Chap. 4. Let $\mathscr{C}_c^\infty(G)$ denote the vector space of all infinitely often differentiable functions with compact support on G and d_G, the left invariant measure on G. Then the Gårding subspaces

$$H_p^\infty := \left\{\int_G d_G(g) f_c(g) R_p(g) h \mid f_c \in \mathscr{C}_c^\infty(G) \;\wedge\; h\in H_p\right\},$$

$$S_\infty := \left\{\int_G d_G(g) f_c(g) R(g) h \mid f_c \in \mathscr{C}_c^\infty(G) \;\wedge\; h\in S\right\} \tag{A1}$$

are dense invariant subspaces of σ_p^∞, respectively σ^∞ (Ref. 14, §4.4.1). Plainly they satisfy $S_\infty = \cap_{p\in\mathbb{N}^0} H_p^\infty$. Now the Gårding–Nelson trick consists of replacing f_c in the integrals by the kernels

$$K_t(g) = e^{t\Lambda(\Delta)} \cdot f_c(g), \quad t\in\mathbb{R}^+. \tag{A2}$$

If ∂L denotes the differential of the left regular representation L of G on $\mathscr{L}^2(G)$, then $\Lambda(\Delta)$ is the closure of the essentially self-adjoint operator $\partial L(\Sigma_{i=1}^n (X_i)^2)$ with $\{X_1,..., X_n\}$ a basis of $\mathfrak{g}$. According to Nelson and Gårding (Ref. 14, §4.4.5), for every $h\in H_p$, $f_c\in\mathscr{C}_c^\infty(G)$, and $t\in\mathbb{R}^+$, the vector

$$h_t := \int_G d_G(g) K_t(g) R_p(g) h \tag{A3}$$

is an element from σ_p^ω and moreover

$$\lim_{t\to 0}\left\| h_t - \int_G d_G(g) f_c(g) R_p(g) h \right\|_p = 0. \tag{A4}$$

The last equation implies that the subspace

$$H_p^\omega := \{h_t \mid h\in H_p,\; f_c\in\mathscr{C}_c^\infty(G),\; t\in\mathbb{R}^+\}$$

is dense in the Gårding space H_p^∞ and thus dense in σ_p^∞ and H_p. Now the generalization of this result to the space $S = \cap_{p\in\mathbb{N}^0} H_p$ simply rests on the fact that it holds in every Banach space H_p, $p\in\mathbb{N}^0$ with the kernels $K_t(g)$ being independent of p! Thus if h is from S then due to $R(g) = R_p(g) \restriction S$ and (A3) h_t is from σ_p^ω for every $p\in\mathbb{N}^0$ and therefore from $\sigma^\omega = \cap_{p\in\mathbb{N}^0} \sigma_p^\omega$. Furthermore since the limit relation (A4) holds for every $p\in\mathbb{N}^0$ the subspace $S_\omega := \{h_t \mid h\in S,\; f_c\in\mathscr{C}_c^\infty(G),\; t\in\mathbb{R}^+\}$ is dense in S_∞ and hence in S.

The invariance properties are a direct consequence of the equations

$$\sigma^\omega = \bigcap_{p\in\mathbb{N}^0} \sigma_p^\omega; \quad R(g) = R_p(g) \restriction S, \quad g\in G;$$
$$\partial R(X) = \partial R_p(X) \restriction \sigma^\infty, \quad X\in\mathfrak{g}$$

and the well-known corresponding properties of σ_p^ω under R_p (∂R_p) (See Ref. 14, §4.4.5). □

In the Schwartz space of strongly decreasing $\mathscr{C}^\infty$-functions a dense invariant subspace of ∞-analytic vectors for the Poincaré group is, for instance, generated by the Hermite functions.[32]

APPENDIX B: INTEGRABILITY CONDITIONS

In this appendix we present a list of the most important integrability conditions in a Hilbert or a Banach space, which has been collected from the literature. Any one of the following six conditions is sufficient for a representation δ: $\mathfrak{g}\to\operatorname{End} E$ of a Lie algebra with basis $\{X_1,..., X_n\}$ on a dense subspace E of a Hilbert or Banach space H to be integrable in the sense of Definition 3.2.

(1) (Nelson,[15] Corollary 9.1): For every $i = 1,..., n$, $\delta(X_i)$ is skew-symmetric in the Hilbert space H and the Laplacian

$$\delta(\Delta) := \delta(X_1)^2 + \cdots + \delta(X_n)^2$$

is essentially self-adjoint on E.

(2) (Nelson,[15] Lemma 9.1; Ref. 8, Proposition 1.3): For every $i = 1,...,n$, $\delta(X_i)$ is skew-symmetric in the Hilbert space H and there exists a dense invariant subspace $\Omega\subseteq E$, a number $s\in\mathbb{R}^+$ (independent of f) and for every $f\in\Omega$ a number $M(f)\in\mathbb{R}^+$ such that for all $m\in\mathbb{N}$ and $i_1,..., i_m\in\{1,..., n\}$, the following bound holds:

$$\|\delta(X_{i_1}) \cdot \cdots \cdot \delta(X_{i_m}) f\| \leqslant M(f) s^m m!\,.$$

Note that Ω is what we call an entire space for δ.

(3) (Moore,[17] Theorems 2 and 4): For every $X\in\mathfrak{g}$, the closure $\overline{\delta(X)}$ of $\delta(X)$ exists in the Banach space H and generates a strongly continuous one parameter group of operators $\rho_X: \mathbb{R}\to\operatorname{Aut} H$, $t\to\rho_X(t)$ with the following properties: (i) $\forall_{X\in\mathfrak{g}, t\in\mathbb{R}}, \rho_X(t) E\subseteq E$; and (ii) $\|\delta(y)\, \rho_X(t) f\|$ is locally bounded at $t = 0$ for every $f\in E$ and X, $Y\in\mathfrak{g}$.

(4) (Flato et al.,[18] Theorem 1): For all $i\in\{1,...,n\}$ the operators $\delta(X_i)$ are skew-symmetric in the Hilbert space H and they posses a *common* dense invariant subspace $\Omega\subseteq E$, the elements of which are analytic for *all* $\delta(X_i)$, $i = 1,...,n$.

(5) (Simon,[19] Corollary 1): Let $\{Y_i\in\mathfrak{g} \mid i = 1,...,m\}$ be a subset that generates $\mathfrak{g}$ via formation of commutators and linear combinations, and $\delta_i^\times := \delta(Y_i)^\times$ the (Banach space) adjoint operator of $\delta(Y_i)$. The adjoint operators $\delta_i^\times$, $i = 1,...,m$ possess a *common* dense invariant subspace $\Omega^\times \subset H^\times$ of analytic vectors and each $\delta_i^\times$ generates a strongly continuous one parameter group of operators

$$\rho_{\delta_i^\times}: \mathbb{R}\to\operatorname{Aut} H, \; t\to\rho_{\delta_i^\times}(t).$$

(6) (Fröhlich,[20] Theorem 3): For every $i\in\{1,...,n\}$ the operator $\delta(X_i)$ is skew-symmetric in the Hilbert space H with scalar product $[\cdots,\cdots]$. In addition there exists a self-adjoint positive operator $N>1$, a subspace $\Omega\subseteq E$, which is a core for N (i.e., $\overline{N\upharpoonright\Omega}=N$) and a positive real number K such that for all $f\in\Omega$ and $i\in\{1,...,n\}$, the following bounds hold:

$$\begin{aligned}&\|\delta(X_i)f\|\\ &\quad\leqslant\sqrt{2}K\,\|Nf\| \pm ([Nf,\delta(X_i)f]+[\delta(X_i)f,Nf])\\ &\quad\leqslant K\,\|N^{1/2}f\|^2=K\,[f,Nf].\end{aligned}$$

Remark: Conditions (1), (2), or (6) even imply that the domain E of the operators $\delta(X_i)$ is a core for the infinitesimal generators $dR(X_i):=\overline{\partial R(X_i)}$ (∂R the differential of the integral of δ); this means $\overline{\partial R(X_i)\upharpoonright E}=dR(X_i)$, $i=1,...,n$.

[1]F. Strocchi and A. S. Wightman, J. Math. Phys. **15**, 2198 (1974).
[2]F. Strocchi, Commun. Math. Phys. **56**, 57 (1977).
[3]F. Strocchi, Phys. Rev. D **17**, 2010 (1977).
[4]G. Morchio and F. Strocchi, Ann. Inst. H. Poincaré A XXXIII, 251 (1980).
[5]A. S. Wightman and L. Gårding, Ark. Fys. **28**, 129 (1964).
[6]R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics, and All That* (Benjamin, New York, 1964).
[7]R. Jost, *The General Theory of Quantized Fields* (Am. Math. Soc., Providence, 1965).
[8]H. J. Borchers, *In Statistical Mechanics and Field Theory* (Halsted, New York, 1972).
[9]L. V. Belvedere, J. A. Swieca, K. D. Rothe, and B. Schroer, Nucl. Phys. B **153**, 112 (1979).
[10]A. K. Raina and G. Wanders, Ann. Phys. NY **132**, 404 (1981).
[11]A. H. Völkel, J. Math. Phys. **26**, 2956 (1985).
[12]A. H. Völkel, J. Math. Phys. **26**, 1057 (1985).
[13]J. P. Antoine, Acta Univ. Wratislav. **519**, 226.
[14]G. Warner, *Harmonic Analysis on Semi-Simple Lie Groups I* (Springer, Berlin, 1972).
[15]E. Nelson, Ann. Math. **70**, 572 (1959).
[16]E. B. Davis, Commun. Math. Phys. **23**, 159 (1971).
[17]R. T. Moore, Bull. Am. Math. Soc. **71**, 903 (1965).
[18]M. Flato, J. Simon, H. Snellman, and D. Sternheimer, Ann. Sci. Ec. Norm. Sup. **5**, 423 (1972).
[19]J. Simon, Commun. Math. Phys. **28**, 39 (1972).
[20]J. Fröhlich, Commun. Math. Phys. **54**, 135 (1977).
[21]R. Goodman, Trans. Am. Math. Soc. **143**, 55 (1969).
[22]P. Cartier, *Représentations des groups de Lie* (Benjamin, New York, 1967).
[23]I. M. Gel'fand and G. E. Shilov, *Generalized Functions*, Vol. II (Academic, New York, 1968).
[24]I. M. Gel'fand and N. Ya. Vilenkin, *Generalized Functions*, Vol. IV (Academic, New York, 1964).
[25]F. Treves, *Topological Vector Spaces, Distributions and Kernels* (Academic, New York, 1967).
[26]A. M. Jaffe, Phys. Rev. **158**, 1454 (1967).
[27]A. H. Völkel, *Lecture Notes in Physics*, Vol. 66 (Springer, Berlin, 1977).
[28]G. Köthe, *Topological Vector Spaces*, Vol. II (Springer, New York, 1979).
[29]S. a. Gaal, *Linear Analysis and Representation Theory* (Springer, Berlin, 1973).
[30]M. Reed and B. Simon, *Methods of Modern Mathematical Physics*, Vol. I (Academic, New York, 1972).
[31]T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1966).
[32]H. Snellman, J. Math. Phys. **15**, 1054 (1974).

Gravitational anomalies in two dimensions: A cohomological approach

Giuseppe Bandelloni

Istituto di Fisica della Facolta di Ingegneria, Universita di Genova, Italy and Istituto Nazionale di Fisica Nucleare, Sezione di Genova, Italy

(Received 12 August 1985; accepted for publication 13 November 1985)

The spectral sequences method is employed to study the cohomology space of the Becchi–Rouet–Stora (BRS) operator, which describes the general coordinate transformations in a two-dimensional polynomial Lagrangian field theory. A pure external gravitational model is considered. In the Fadeev–Popov charge-one sector, two classes of elements are found: the first represents the ordinary trace anomalies, in the second the presence of the anomalies recently calculated by Bardeen and Zumino are pointed out.

I. INTRODUCTION

The modern approaches to quantum field theory (QFT) have raised the interest of physicists towards aspects of differential geometry, such as cohomology theory, homotopy theory, index theorems, and so on.

In particular, the cohomological approach to gauge field models[1,2] has been successful in performing the renormalization program without lengthy explicit Feynman diagrams calculations; however, it requires, for physical interest, a useful method for the calculation of the cohomology space of null-squared linear differential operators.

Several of these methods exist in the mathematical literature, but many of them are fitted to the structure of particular operators and are of no interest in the remaining cases.

Some years ago Dixon[3] introduced, in the QFT literature, the spectral sequences method.[4]

This mathematical procedure allows us to construct a space, isomorphic to the cohomology one, on a local polynomial space by successive approximations, given by an exact sequence of cohomology spaces.

The crucial point of this method is to reduce to an algebraic equation the coboundary condition satisfied by the anomalies; otherwise, the whole image space of the null-squared linear differential operator has to be explored to have an answer to our problem.

Nevertheless the mathematical beauty of this method strongly limits its applicability to physical problems, since the mathematical structure of the differential operators, which induce the cohomology sequences, requires a procedure that is too difficult for explicit calculations; finally the isomorphism, which relates the "physical" cohomology space with the one given by the method, is not evident *a priori*.

Hence the spectral sequences procedure, as it stands, is useful only to calculate the Betti numbers, which might sometimes be a poor result for so big a mathematical effort. Furthermore, in the QFT cohomological applications, we are interested not only in the local polynomial space cohomology, but, more generally, in the functional space cohomology: the problem becomes more difficult.

Recently[5] we have given a refinement of this method, which slightly modifies the Dixon approach, but strongly simplifies its handling for physical problems.

In this paper we shall apply this method to two-dimensional external gravity, and we shall put into evidence the cohomological birth of the Adler–Bardeen type gravitational anomalies recently discovered by Alvarez-Gaume and Witten,[6] explicitly calculated by Bardeen and Zumino[7] and discussed in Refs. 7–12. Moreover, the ordinary trace anomalies are recovered.

In Sec. II we describe the model and calculate the Becchi–Rouet–Stora (BRS) operator $\mathring{\delta}_L$, whose cohomology space is the object of this work.

In Sec. III we briefly sum up the spectral sequence method and calculate the local polynomials space cohomology of the operator $\mathring{\delta}_L$.

In Sec. IV we use the results of the previous section and make evident the subsidiary conditions that the elements of the local functional cohomology have to satisfy.

In Sec. V we discuss these subsidiary conditions, which split into two families: the first finds as more general solutions the ordinary trace anomalies[13]; among the solutions of the second, we underline the presence of the Bardeen–Zumino anomalies.

The Appendix is devoted to a brief presentation of an elegant approach to spectral sequences formalism, found by Zeeman, which is particularly useful for a pedagogical introduction of it; furthermore we shall prove an isomorphism used in Sec. III.

II. THE CLASSICAL MODEL

Let us consider a two-dimensional external gravitational model generated by the moving frame tensor $\tilde{\chi}^a_\mu(x)$ and the comoving frame tensor $\tilde{\chi}^\mu_a(x)$, which obey the Einstein transformations $\delta x^\mu = -\lambda^\mu(x)$

$$\delta\tilde{\chi}^a_\mu(x) = \lambda^\sigma(x)\partial_\sigma\tilde{\chi}^a_\mu(x) + \partial_\mu\lambda^\sigma(x)\tilde{\chi}^a_\sigma(x), \tag{2.1a}$$

$$\delta\tilde{\chi}^\mu_a(x) = \lambda^\sigma(x)\partial_\sigma\tilde{\chi}^\mu_a(x) - \partial_\sigma\lambda^\mu(x)\tilde{\chi}^\sigma_a(x). \tag{2.1b}$$

In order to avoid the presence of $\det[\tilde{\chi}^a_\mu]$ terms, we shall define the new field densities

$$\tilde{\chi}^a_\mu(x) = (\det[\tilde{\chi}^a_\mu])^{1/2}\chi^a_\mu(x), \tag{2.2a}$$

$$\tilde{\chi}^\mu_a(x) = (\det[\tilde{\chi}^a_\mu])^{-1/2}\chi^\mu_a, \tag{2.2b}$$

such that $\det[\chi^a_\mu] = 1$, and the model can be expanded in a polynomial basis, if the theory is scale invariant, as shown in Ref. 14.

In this approach, the Levi-Civita tensor $\epsilon^{\mu\nu}_{(x)}$ is defined in terms of the flat space antisymmetric tensor $\mathring{\epsilon}^{ab}$ as

 0022-2488/86/041128-11$02.50

$$\epsilon^{\mu\nu}(x) = \mathring{\epsilon}^{ab}\chi_a^\mu(x)\chi_b^\nu(x). \tag{2.3}$$

Its antisymmetry and its coordinate independence are easily verified; for example,

$$\epsilon^{11}(x) = \mathring{\epsilon}^{12}\chi_1^1(x)\chi_2^1(x) + \mathring{\epsilon}^{21}\chi_2^1(x)\chi_1^1(x) = 0, \tag{2.4a}$$

$$\epsilon^{12}(x) = \mathring{\epsilon}^{12}\chi_1^1(x)\chi_2^2(x) + \mathring{\epsilon}^{21}\chi_2^1(x)\chi_1^2(x) = \det\chi = 1 = -\epsilon^{21}(x). \tag{2.4b}$$

By introducing the ghost field $C^\mu(x)$ with Fadeev–Popov ($\Phi\Pi$) charge one, it is easy to construct the differential operator δ (with $\delta^2 = 0$), acting on the dimensionless fields $\chi_a^\mu(x)$ and $\chi_\mu^a(x)$ and the connection $\Gamma_{\mu\nu}^\rho(x)$ and $C_\mu(x)$, with the substitution operations

$$\delta\chi_a^\mu(x) = C^\lambda(x)\partial_\lambda\chi_a^\mu(x) + \tfrac{1}{2}\partial_\lambda C^\lambda(x)\chi_a^\mu(x) - \partial_\lambda C^\mu(x)\chi_a^\lambda(x), \tag{2.5a}$$

$$\delta\chi_\mu^a(x) = C^\lambda(x)\partial_\lambda\chi_\mu^a(x) - \tfrac{1}{2}\partial_\lambda C^\lambda(x)\chi_\mu^a(x) + \partial_\mu C^\lambda(x)\chi_\lambda^a(x), \tag{2.5b}$$

$$\delta\Gamma_{\mu\nu}^\rho(x) = C^\lambda(x)\partial_\lambda\Gamma_{\mu\nu}^\rho(x) + \partial_\mu C^\lambda(x)\Gamma_{\lambda\nu}^\rho(x) + \partial_\nu C^\lambda(x)\Gamma_{\mu\lambda}^\rho(x) - \partial_\lambda C^\rho(x)\Gamma_{\mu\nu}^\lambda(x) + \partial_\mu\partial_\nu C^\rho(x), \tag{2.5c}$$

$$\delta C^\mu(x) = C^\lambda(x)\partial_\lambda C^\mu(x). \tag{2.5d}$$

Using the $\Phi\Pi$ negative charged fields $\gamma_\mu^a(x)$, $\gamma_a^\mu(x)$, $\gamma_\rho^{\mu\nu}(x)$, and $\zeta_\mu(x)$, with $\Phi\Pi$ charge $Q_{\Phi\Pi}$,

$$Q_{\Phi\Pi}\gamma_\mu^a(x) = Q_{\Phi\Pi}\gamma_a^\mu(x) = Q_{\Phi\Pi}\gamma_\rho^{\mu\nu}(x) = -1; \quad Q_{\Phi\Pi}\zeta_\mu(x) = -2, \tag{2.6}$$

and uv dimensions (dim),

$$\dim\gamma_\mu^a(x) = \dim\gamma_a^\mu(x) = 1; \quad \dim\gamma_\rho^{\mu\nu}(x) = 0; \quad \dim\zeta_\mu(x) = 1, \tag{2.7}$$

respectively, it is possible to define the classical action functional

$$\Gamma^{\mathrm{cl}} = \int d^2x[\mathscr{S}_{inv}(\chi_\mu^a, \chi_b^\nu, \Gamma_{\mu\nu}^\rho)(x) + \gamma_\mu^a(x)\delta\chi_a^\mu(x) + \gamma_a^\mu(x)\delta\chi_\mu^a(x) + \gamma_\rho^{\mu\nu}(x)\delta\Gamma_{\mu\nu}^\rho(x) + \zeta_\mu(x)\delta C^\mu(x)] \tag{2.8}$$

as the most general functional invariant under the linear operator

$$\begin{aligned}
\mathring{\delta}_L = \int d^2x\Big[&\frac{\delta\Gamma^{\mathrm{cl}}}{\delta\chi_\mu^a(x)}\frac{\delta}{\delta\gamma_a^\mu(x)} + \frac{\delta\Gamma^{\mathrm{cl}}}{\delta\gamma_a^\mu(x)}\frac{\delta}{\delta\chi_\mu^a(x)} + \frac{\delta\Gamma^{\mathrm{cl}}}{\delta\chi_a^\mu(x)}\frac{\delta}{\delta\gamma_\mu^a(x)} + \frac{\delta\Gamma^{\mathrm{cl}}}{\delta\gamma_\mu^a(x)}\frac{\delta}{\delta\chi_a^\mu(x)} \\
&+ \frac{\delta\Gamma^{\mathrm{cl}}}{\delta\Gamma_{\mu\nu}^\rho(x)}\frac{\delta}{\delta\gamma_\rho^{\mu\nu}(x)} + \frac{\delta\Gamma^{\mathrm{cl}}}{\delta\gamma_\rho^{\mu\nu}(x)}\frac{\delta}{\delta\Gamma_{\mu\nu}^\rho(x)} + \frac{\delta\Gamma^{\mathrm{cl}}}{\delta C^\mu(x)}\frac{\delta}{\delta\zeta_\mu(x)} + \frac{\delta\Gamma^{\mathrm{cl}}}{\delta\zeta_\mu(x)}\frac{\delta}{\delta C^\mu(x)}\Big] \\
= \int d^2x\Big[&\Big(\frac{\delta\mathscr{S}^{inv}}{\delta\chi_\mu^a(x)} - \partial_\lambda(C^\lambda(x)\gamma_a^\mu(x)) - \frac{1}{2}\gamma_a^\mu(x)\partial_\lambda C^\lambda(x) + \gamma_a^\lambda(x)\partial_\lambda C^\mu(x)\Big)\frac{\delta}{\delta\gamma_a^\mu(x)} \\
&+ \Big(C_\lambda(x)\partial_\lambda\chi_\mu^a(x) - \frac{1}{2}\partial_\lambda C^\lambda(x)\chi_\mu^a(x) + \partial_\mu C^\lambda(x)\chi_\lambda^a(x)\Big)\frac{\delta}{\delta\chi_\mu^a(x)} \\
&+ \Big(\frac{\delta\mathscr{S}^{inv}}{\delta\chi_a^\mu(x)} - \partial_\lambda(C^\lambda(x)\gamma_\mu^a(x)) + \frac{1}{2}\gamma_\mu^a(x)\partial_\lambda C^\lambda(x) - \gamma_\lambda^a(x)\partial_\mu C^\lambda(x)\Big)\frac{\delta}{\delta\gamma_\mu^a(x)} \\
&+ \Big(C^\lambda(x)\partial_\lambda\chi_a^\mu(x) + \frac{1}{2}\partial_\lambda C^\lambda(x)\chi_a^\mu(x) - \partial_\lambda C^\mu(x)\chi_a^\lambda(x)\Big)\frac{\delta}{\delta\chi_a^\mu(x)} \\
&+ \Big(\frac{\delta\mathscr{S}^{inv}}{\delta\Gamma_{\mu\nu}^\rho} + \partial_\lambda(C^\lambda(x)\gamma_\rho^{\mu\nu}(x)) + \gamma_\rho^{\lambda\nu}(x)\partial_\lambda C^\mu(x) + \gamma_\rho^{\mu\lambda}(x)\partial_\lambda C^\nu(x) - \gamma_\lambda^{\mu\nu}(x)\partial_\rho C^\lambda(x)\Big)\frac{\delta}{\delta\gamma_\rho^{\mu\nu}(x)} \\
&+ (C^\lambda(x)\partial_\lambda\Gamma_{\mu\nu}^\rho(x) + \partial_\mu C^\lambda(x)\Gamma_{\lambda\nu}^\rho(x) + \partial_\nu C^\lambda(x)\Gamma_{\mu\lambda}^\rho(x) - \partial_\lambda C^\rho(x)\Gamma_{\mu\nu}^\lambda(x) + \partial_\mu\partial_\nu C^\rho(x))\frac{\delta}{\delta\Gamma_{\mu\nu}^\rho(x)} \\
&+ (\gamma_\lambda^a(x)\partial_\mu\chi_a^\lambda(x) - \tfrac{1}{2}\partial_\mu(\chi_a^\lambda(x)\gamma_a^\lambda(x)) + \partial_\lambda(\chi_a^\lambda(x)\gamma_\mu^a(x)) + \gamma_a^\lambda(x)\partial_\mu\chi_\lambda^a(x) + \tfrac{1}{2}\partial_\mu(\gamma_a^\lambda(x)\chi_\lambda^a(x)) \\
&- \partial_\lambda(\gamma_a^\lambda(x)\chi_\mu^a(x)) + \gamma_\rho^{\lambda\nu}(x)\partial_\mu\Gamma_{\lambda\nu}^\rho(x) - \partial_\lambda(\gamma_\rho^{\lambda\nu}(x)\Gamma_{\mu\nu}^\rho(x)) - \partial_\lambda(\gamma_\rho^{\nu\lambda}(x)\Gamma_{\nu\mu}^\rho(x)) + \partial_\sigma(\gamma_\mu^{\lambda\nu}(x)\Gamma_{\lambda\nu}^\sigma(x)) \\
&+ \partial_\nu\partial_\lambda\gamma_\mu^{\nu\lambda}(x) + \zeta_\lambda(x)\partial_\mu C^\lambda(x) - \partial_\lambda(\zeta_\mu(x)C^\lambda(x)))\frac{\delta}{\delta\zeta_\mu(x)} + C^\lambda(x)C^\mu(x)\frac{\delta}{\delta C^\mu(x)}\Big],
\end{aligned} \tag{2.9}$$

that is,

$$\mathring{\delta}_L\Gamma^{\mathrm{cl}} = 0. \tag{2.10}$$

If a quantum extension of the action functional

$$\Gamma = \sum_n \hbar^n\Gamma^n \quad \text{with} \quad \Gamma^0 = \Gamma^{\mathrm{cl}} \tag{2.11}$$

is implemented, Eq. (2.10) is, in general, violated at a generic mth order of $\hbar$ perturbation expansion, so we could get

$$\mathring{\delta}_L\Gamma = (\mathring{\delta}_L + O(\hbar))\Gamma = \hbar^m\Delta^1 + O(\hbar^{m+1}), \tag{2.12a}$$

where

$$\Delta^1 = \int d^2x \Delta^1(x) \equiv \int \Delta_2^1(x), \tag{2.12b}$$

and anomalies Δ^1 can occur.

The algebraic properties of the operator $\mathring{\delta}_L$ will lead, in the perturbation expansion, to the cocycle condition

$$\mathring{\delta}_L \Delta^1 = 0. \tag{2.13}$$

The above expression can be easily expressed in terms of local quantities; in fact, in the polynomial basis, the background space is isomorphic to R_2. Hence, we can apply the Poincaré lemma, using a technique introduced by Stora[15] in 1976, since the operator $\mathring{\delta}_L$ commutes with the derivative operator. We are thus led to the system (extended to a whatever $\Phi\Pi$ charge sector):

$$\mathring{\delta}_L \Delta_2^n(x) + d\Delta_1^{n+1}(x) = 0, \tag{2.14a}$$

$$\mathring{\delta}_L \Delta_1^{n+1}(x) + d\Delta_0^{n+2}(x) = 0, \tag{2.14b}$$

$$\mathring{\delta}_L \Delta_0^{n+2}(x) = 0. \tag{2.14c}$$

We shall use the spectral sequences method to solve Eq. (2.14c) and then we proceed backwards to derive $\Delta_2^1(x)$.

III. LOCAL POLYNOMIALS COHOMOLOGY

We shall introduce here the spectral sequences method; for a more complete mathematical review the reader is referred to Refs. 16–18. Here we outline the method, citing the theorems we shall use in the text; the Appendix will treat particular aspects of the problem.

First, we have to imbed our polynomial space into a Hilbert one: this was done by Dixon using a creation and destruction operators technique on the polynomial space considered as a Fock space.

This procedure defines an "*ad hoc*" very useful scalar product, and if, for example, we write the operator $\mathring{\delta}_L$ in a dyadic form, its adjoint is immediately defined.

Not to weigh the mathematical notation, (nor bore the reader), we shall use a derivative notation, and, as usual, adopt the dagger symbol to indicate the adjointness operation. We shall define, for example [if $f(x)$ is an element of the polynomial space F],

$$\left[\partial_\mu \chi_a^\rho(x) \frac{\partial}{\partial\partial_\nu\partial_\tau\partial_\sigma \chi_a^\lambda(x)}\right]^\dagger f(x) = \partial_\nu\partial_\tau\partial_\sigma \chi_a^\lambda(x) \frac{\partial f(x)}{\partial\partial_\mu \chi_a^\rho(x)}, \tag{3.1}$$

which finds an immediate twinning with the Dixon formalism.

So, according to the Hodge decomposition, extended by Dixon to this framework, the space F admits the direct decomposition

$$F = (\operatorname{Im} \mathring{\delta}_L) \oplus (\operatorname{Im} \mathring{\delta}_L^\dagger) \oplus (\operatorname{Ker} \mathring{\delta}_L \cap \operatorname{Ker} \mathring{\delta}_L^\dagger). \tag{3.2}$$

The local polynomials cohomology space is defined by the solutions of the system

$$\mathring{\delta}_L x = 0, \tag{3.3a}$$

$$\mathring{\delta}_L^\dagger x = 0. \tag{3.3b}$$

The problem can be solved directly if we find all the solutions of the above system. Obviously (due to the form of the operator $\mathring{\delta}_L$), this goes beyond our mathematical ability, and a more refined method has to be found.

We introduce now a self-adjoint counting operator ν; for our purposes we shall use

$$\begin{aligned}\nu = \sum_n \Big\{ & nD_{\alpha(n)} C^\mu(x) \frac{\partial}{\partial D_{\alpha(n)} C^\mu(x)} \\ & - 5\Big[D_{\alpha(n)}\, \gamma_\mu^a(x) \frac{\partial}{\partial D_{\alpha(n)}\, \gamma_\mu^a(x)} \\ & + D_{\alpha(n)}\, \gamma_a^\mu(x) \frac{\partial}{\partial D_{\alpha(n)}\, \gamma_a^\mu(x)} \\ & + D_{\alpha(n)}\, \gamma_\rho^{\mu\nu}(x) \frac{\partial}{\partial D_{\alpha(n)}\, \gamma_\rho^{\mu\nu}(x)} \\ & + 2D_{\alpha(n)} \zeta_\mu(x) \frac{\partial}{\partial D_{\alpha(n)} \zeta_\mu(x)} \Big]\Big\} \\ & (D_{\alpha(n)} C^\mu(x) \equiv \partial_{\alpha_1}\partial_{\alpha_2} \cdots \partial_{\alpha_n} C^\mu(x)),\end{aligned} \tag{3.4}$$

and filtrate the operator $\mathring{\delta}_L$ and the space F according to

$$\mathring{\delta}_L = \sum_p \mathring{\delta}_L(p), \tag{3.5}$$

where

$$p\mathring{\delta}_L(p) = [\nu, \mathring{\delta}_L(p)]_-, \tag{3.6}$$

and

$$F = \sum_p G_p, \tag{3.7}$$

such that $x_p \in G_p$ means $\nu x_p = p x_p$.

According to our scalar product, two spaces G_p and G_q with $p \neq q$ are orthogonal.

Particularly useful are the nested Hilbert spaces

$$F_p = \sum_{q>p} G_q, \quad \text{such that} \quad F_p \supset F_{p+1}, \tag{3.8a}$$

$$F^p = \sum_{q<p} G_q \quad \text{such that} \quad F^p \supset F^{p-1}, \tag{3.8b}$$

from which it is possible to build the quantities

$$E_r^p \equiv \frac{F_p \cap \mathring{\delta}_L^{-1} F_{p+r}}{\mathring{\delta}_L F_{p-r+1} \cap F_p + \mathring{\delta}_L^{-1} F_{p+r} \cap F_{p+1}} \tag{3.9}$$

(where $x \in \mathring{\delta}_L^{-1} F_p$ means $\mathring{\delta}_L x \in F_p$) and the related one

$$E_r \equiv \left\{ \bigcup_p E_r^p; -\infty < p < +\infty \right\}, \tag{3.10}$$

which have, as shown by Serre,[17] many beautiful properties. We summarize here in three steps.

(1) E_{r+1} are the cohomology spaces of differential operators d_r induced by $\mathring{\delta}_L$.

(2) E_∞ is isomorphic to the cohomology space of the operator $\mathring{\delta}_L$.

(3) If the operator ν will decompose the space F with a finite number m of filtrations, then E will coincide with E_∞.

The above results say that this procedure is of physical interest if (1) the operators d are evaluable in a simple way, (2) the isomorphism between E_∞ and our cohomology space can be easily realized, and (3) we have to choose a filtration operator that decomposes the whole space in a finite number of filtrations.

But, even if the third requirement can be satisfied with an oculate choice of the filtration operator,[19] mathematicians give no answers to the remaining points (1) and (2). For these purposes we proposed[5] an alternative point of view to get rid of point (1) and are hoping for a good idea with regard to point (2). In fact it is possible to show, using a technique introduced by Zeeman,[18] the isomorphism

$$E_{r+1}^{p} \cong \frac{E_r^p \cap (\mathring{\delta}_L^{-1} F_{p+r+1})}{E_r^p \cap (\mathring{\delta}_L F_{p-r})}, \tag{3.11}$$

where $\cong$ means isomorphism. The Appendix will treat mainly this aspect. The above formula is useful for selecting among the $x_p \in E_r^p$ all the ones that belong to E_{r+1}^p.

In fact, if $x_p \in E_r^p$, from Eq. (3.11), we shall get $x_p \in E_{r+1}^p$, if the following hold.

(1) $x_p \in \mathring{\delta}_L^{-1} F_{p+r+1}$, that is $\mathring{\delta}_L x_p \in F_{p+r+1}$, which will imply $(\mathring{\delta}_L x_p, F^{p+r}) = 0$. If we decompose $\mathring{\delta}_L x_p$ and F^{p+r} in their elements belonging to different ν eigenspaces, and use the orthogonality relations we get $(\mathring{\delta}_L(r)x_p, G_{p+r}) = 0$, from which it follows that $\mathring{\delta}_L(r)x_p = 0$.

(2) x_p has to be orthogonal to all the elements of the space $E_r^p \cap (\mathring{\delta}_L F_{p-r})$, thus we get $(x_p, \mathring{\delta}_L F_{p-r}) = 0$, which implies $(\mathring{\delta}_L^{\dagger} x_p, F_{p-r}) = 0$. After decomposition into orthogonal spaces we get $(\mathring{\delta}_L^{\dagger}(r)x_p, G_{p-r}) = 0$, from which it follows that $\mathring{\delta}_L^{\dagger}(r)x_p = 0$.

Hence, we have that if $x \in E_r$, then the same $x \in E_{r+1}$, if the system

$$\mathring{\delta}_L(r)x = 0, \tag{3.12a}$$

$$\mathring{\delta}_L^{\dagger}(r) = 0 \tag{3.12b}$$

is solved.

Starting from a generic $x \in E_0$ (which is the whole space) and iterating the above system for all r, we end up with a vector space that is isomorphic to our cohomology space. Our filtering operator ν, decomposes the operator $\mathring{\delta}_L$ (for the lowest eigenvalues) into

$$\begin{aligned} \mathring{\delta}_L(0) = {} & C^\mu(x) \sum_n \Big[D_{\alpha(n)} \partial_\mu \gamma_a^\sigma(x) \frac{\partial}{\partial D_{\alpha(n)} \gamma_a^\sigma(x)} + D_{\alpha(n)} \partial_\mu \gamma_\sigma^a(x) \frac{\partial}{\partial D_{\alpha(n)} \gamma_\sigma^a(x)} + D_{\alpha(n)} \partial_\mu \chi_a^\rho(x) \frac{\partial}{\partial D_{\alpha(n)} \chi_a^\rho(x)} \\ & + D_{\alpha(n)} \partial_\mu \chi_\rho^a(x) \frac{\partial}{\partial D_{\alpha(n)} \chi_\rho^a(x)} + D_{\alpha(n)} \partial_\mu \Gamma_{\rho\sigma}^\lambda(x) \frac{\partial}{\partial D_{\alpha(n)} \Gamma_{\rho\sigma}^\lambda(x)} \\ & + D_{\alpha(n)} \partial_\mu \gamma_\lambda^{\rho\sigma}(x) \frac{\partial}{\partial D_{\alpha(n)} \gamma_\lambda^{\rho\sigma}(x)} + D_{\alpha(n)} \partial_\mu \zeta_\rho(x) \frac{\partial}{\partial D_{\alpha(n)} \zeta_\rho(x)} \Big] \\ = {} & C^\mu(x) \sum_{n,i} D_{\alpha(n)} \partial_\mu \Phi_i(x) \frac{\partial}{\partial D_{\alpha(n)} \Phi_i(x)} \equiv C^\mu(x) \partial_\mu \Big|_{C^\mu = \text{const}} \end{aligned} \tag{3.13}$$

[where $\Phi_i(x)$ represents a column vector containing all the fields except $C^\mu(x)$] and so we get

$$\mathring{\delta}_L^{\dagger}(0) = \partial_\mu^{\dagger}\big|_{C^\mu = \text{const}} \frac{\partial}{\partial C^\mu(x)} \equiv \sum_{n,i} D_{\alpha(n)} \Phi_i(x) \frac{\partial}{\partial \partial_\mu D_{\alpha(n)} \Phi_i(x)} \frac{\partial}{\partial C^\mu(x)}, \tag{3.14}$$

going on

$$\mathring{\delta}_L(1) = \partial_\lambda C^\lambda(x) D(x) + \partial_\mu C^\nu(x) R_\nu^{\ \mu}(x) + \sum_n D_{\alpha(n)} (C^\rho(x) \partial_\rho C^\mu(x)) \frac{\partial}{\partial D_{\alpha(n)} C^\mu(x)}, \tag{3.15}$$

where

$$\begin{aligned} D(x) = {} & \sum_n \frac{1}{2} \Big[D_{\alpha(n)} \chi_a^\mu(x) \frac{\partial}{\partial D_{\alpha(n)} \chi_a^\mu(x)} - D_{\alpha(n)} \chi_\mu^a(x) \frac{\partial}{\partial D_{\alpha(n)} \chi_\mu^a(x)} + 2 D_{\alpha(n)} \zeta_\mu(x) \frac{\partial}{\partial D_{\alpha(n)} \zeta_\mu(x)} \\ & + D_{\alpha(n)} \gamma_\mu^a(x) \frac{\partial}{\partial D_{\alpha(n)} \gamma_\mu^a(x)} + 3 D_{\alpha(n)} \gamma_a^\mu(x) \frac{\partial}{\partial D_{\alpha(n)} \gamma_a^\mu(x)} + 2 D_{\alpha(n)} \gamma_\rho^{\mu\nu}(x) \frac{\partial}{\partial D_{\alpha(n)} \gamma_\rho^{\mu\nu}(x)} \Big], \end{aligned} \tag{3.16a}$$

$$\begin{aligned} R_\nu^{\ \mu}(x) = {} & \sum_{n,i} \Big[n \partial_\nu D_{\alpha(n-1)} \Phi_i(x) \frac{\partial}{\partial \partial_\mu D_{\alpha(n-1)} \Phi_i(x)} - D_{\alpha(n)} \chi_a^\mu(x) \frac{\partial}{\partial D_{\alpha(n)} \chi_a^\nu(x)} + D_{\alpha(n)} \chi_\nu^a(x) \frac{\partial}{\partial D_{\alpha(n)} \chi_\mu^a(x)} \\ & + D_{\alpha(n)} \gamma_\nu^a(x) \frac{\partial}{\partial D_{\alpha(n)} \gamma_\mu^a(x)} - D_{\alpha(n)} \gamma_a^\mu(x) \frac{\partial}{\partial D_{\alpha(n)} \gamma_a^\nu(x)} + D_{\alpha(n)} \Gamma_{\nu\rho}^\lambda(x) \frac{\partial}{\partial D_{\alpha(n)} \Gamma_{\mu\rho}^\lambda(x)} \\ & + D_{\alpha(n)} \Gamma_{\rho\nu}^\lambda(x) \frac{\partial}{\partial D_{\alpha(n)} \Gamma_{\rho\mu}^\lambda(x)} - D_{\alpha(n)} \Gamma_{\rho\sigma}^\mu(x) \frac{\partial}{\partial D_{\alpha(n)} \Gamma_{\rho\sigma}^\nu(x)} + D_{\alpha(n)} \gamma_\nu^{\rho\sigma}(x) \frac{\partial}{\partial D_{\alpha(n)} \gamma_\mu^{\rho\sigma}(x)} \\ & - D_{\alpha(n)} \gamma_\sigma^{\mu\rho}(x) \frac{\partial}{\partial D_{\alpha(n)} \gamma_\sigma^{\nu\rho}(x)} - D_{\alpha(n)} \gamma_\sigma^{\rho\mu}(x) \frac{\partial}{\partial D_{\alpha(n)} \gamma_\sigma^{\rho\nu}(x)} + D_{\alpha(n)} \zeta_\nu(x) \frac{\partial}{\partial D_{\alpha(n)} \zeta_\mu(x)} \Big], \end{aligned} \tag{3.16b}$$

hence

$$\mathring{\delta}_L^{\dagger}(1) = D(x) \frac{\partial}{\partial \partial_\lambda C^\lambda(x)} + R_\mu^{\ \nu}(x) \frac{\partial}{\partial \partial_\mu C^\nu(x)} + \sum_{n>p} \binom{n}{p} D_{\alpha(n)} C^\mu(x) \frac{\partial}{\partial D_{\alpha(p)} \partial_\rho C^\mu(x)} \frac{\partial}{\partial D_{\alpha(n-p)} C^\mu(x)}. \tag{3.17}$$

As regards the higher $\mathring{\delta}_L(r)$, $r>1$, they are very complex objects; they never act on the undifferentiated C^μ field.

The first step consists of looking for functions $\Delta(x)$ that solve the system

$$\mathring{\delta}_L(0)\Delta(x) \equiv C^\mu(x)\partial_\mu|_{C^\mu=\text{const}}\Delta(x)=0, \tag{3.18a}$$

$$\mathring{\delta}_L^\dagger(0)\Delta(x) \equiv \partial_\mu^\dagger|_{C^\mu=\text{const}}\frac{\partial\Delta(x)}{\partial C^\mu(x)}=0. \tag{3.18b}$$

Since the operator $\mathring{\delta}_L(0)$ mimics the ordinary differential operator in a polynomial space where the field $C_\mu(x)$ (and its derivatives) is considered as a constant, an *"ad hoc"* version of the Poincaré lemma shows that the most general solution takes the form

$$\Delta(x)=C^\mu(x)C^\nu(x)\Delta_{\mu\nu}(x)+F(D_{\alpha(n)}C^\mu(x)), \tag{3.19}$$

with the subsidiary condition

$$[\delta_\sigma^\mu C^\nu(x)-C^\mu(x)\delta_\sigma^\nu]\partial_\sigma|_{C^\mu=\text{const}}\Delta_{\mu\nu}(x)=0, \tag{3.20}$$

where $\Delta_{\mu\nu}(x)$ is an arbitrary function that cannot contain underived $C^\mu(x)$ fields, and $F(D_{\alpha(n)}C^\mu(x))$ is an arbitrary function of $C^\mu(x)$ and its derivatives.

At the successive step, we have to select the class of the previous functions $\Delta(x)$ that satisfy the system

$$\mathring{\delta}_L(1)\Delta(x)=0, \tag{3.21a}$$

$$\mathring{\delta}_L^\dagger(1)\Delta(x)=0. \tag{3.21b}$$

Now, since $\partial_\lambda(C^\lambda(x)C^\mu(x)C^\nu(x))=0$, we can write

$$C^\mu(x)C^\nu(x)[\mathring{\delta}_L(1)-\partial_\lambda C^\lambda(x)]\Delta_{\mu\nu}(x) + \mathring{\delta}_L F(D_nC^\mu(x))=0, \tag{3.22a}$$

$$C^\mu(x)C^\nu(x)[\mathring{\delta}_L(1)-\partial_\lambda C^\lambda(x)]^\dagger\Delta_{\mu\nu}(x) + \sum_n(\delta_\sigma^\mu C^\nu(x)-C^\mu(x)\delta_\sigma^\nu) \times D_{\alpha(n)}C^\rho(x)\frac{\partial\Delta_{\mu\nu}(x)}{\partial D_{\alpha(n)}\partial_\sigma C^\rho(x)} + \mathring{\delta}_L^\dagger(1)F(D_{\alpha(n)}C^\mu(x))=0. \tag{3.22b}$$

It is important to remark at this stage that in Eq. (3.19), the term $F(D_{\alpha(n)}C^\mu(x))$ always has to contain two undifferentiated $C_\mu(x)$ fields, otherwise we get

$$C^\mu(x)C^\nu(x)[\mathring{\delta}_L(1)-\partial_\lambda C^\lambda(x)]\Delta_{\mu\nu}(x) = \mathring{\delta}_L F(D_{\alpha(n)}C^\mu(x))=0. \tag{3.23}$$

So

$$F(D_{\alpha(n)}C^\mu(x))=\mathring{\delta}_L\hat{F}(D_{\alpha(n)}C^\mu(x))+F_\natural(D_{\alpha(n)}C^\mu(x)),$$

where $F_\natural(D_{\alpha(n)}C^\mu(x))$ belongs to the cohomology of the operator $\delta_1'=\int d^2xC^\rho(x)\partial_S C^\mu(x)[S/\delta C^\mu(x)]$ and the coboundary $\mathring{\delta}_L\hat{F}(D_{\alpha(n)}C^\mu(x))$ has to be discarded (otherwise Δ is not a minimal cohomology solution). Thus we must solve the system

$$\delta_1 F_\natural(x)=0, \tag{3.24a}$$

$$\delta_1^\dagger F_\natural(x)=0. \tag{3.24b}$$

Filtering now the above cohomology with the operator $\nu'=1-C^\mu(x)[\partial/\partial C^\mu(x)]$, as in Ref. 5, it is easy to find that

$$F_\natural(x)=C^\mu(x)C^\nu(x)F_{\natural\mu\nu}(x), \tag{3.25}$$

which satisfies analogous conditions to Eqs. (3.20) and (3.22), and so can be put into the term $C^\mu(x)C^\nu(x)\Delta_{\mu\nu}(x)$.

If we proceed into our filtration operation we have to restrict the classes of the functions $\Delta_{\mu\nu}(x)$ and F to the ones that satisfy the system

$$C^\mu(x)C^\nu(x)\mathring{\delta}_L(r)\Delta_{\mu\nu}(x)=0, \tag{3.26a}$$

$$C^\mu(x)C^\nu\mathring{\delta}_L^\dagger(r)\Delta_{\mu\nu}(x)+\mathring{\delta}_L^\dagger(r)\,F(D_{\alpha(n)}C^\mu(x))=0, \tag{3.26b}$$

for all r.

So we have to solve, *a priori*, a system of infinite equations.

Since the above equations hold, due to our decompositions, in mutually orthogonal spaces, we can sum Eqs. (3.22)–(3.26), and then we have to verify that no extraneous solutions exist that (due to spectacular cancellations) are solutions of the summed equations but not our primitive equations. So we get

$$C^\mu(x)C^\nu(x)S\Delta_{\mu\nu}(x)+\mathring{\delta}_L\,F(D_{\alpha(n)}C^\mu(x))=0, \tag{3.27a}$$

$$C^\mu(x)C^\nu(x)S^\dagger\Delta_{\mu\nu}(x)+\sum(\delta_\sigma^\mu C^\nu(x)-\delta_\sigma^\nu C^\mu(x)) \times D_{\alpha(n)}C^\rho(x)\frac{\partial\Delta_{\mu\nu}(x)}{\partial D_{\alpha(n)}\partial_\sigma C^\rho(x)} + \mathring{\delta}_L^\dagger\,F(D_{\alpha(n)}C^\mu(x))=0, \tag{3.27b}$$

where

$$S=\mathring{\delta}_L-C^\lambda(x)\partial_\lambda-\partial_\lambda C^\lambda(x), \tag{3.28}$$

and $S^2=0$.

It is easy to prove that our statement is correct. In fact, if we take as solutions

$$\Delta(x)=C^\mu(x)C^\nu(x)\Delta_{\mu\nu}(x)+F(D_{\alpha(n)}C^\mu(x)), \tag{3.29}$$

which satisfy Eq. (3.27) and the condition Eq. (3.20), it is not difficult to verify that

$$\mathring{\delta}_L\Delta(x)=\mathring{\delta}_L(C^\mu(x)C^\nu(x)\Delta_{\mu\nu}(x)+F(D_{\alpha(n)}C^\mu(x))) = \partial_\lambda(C^\lambda(x)C^\mu(x)C^\nu(x)\Delta_{\mu\nu}(x))=0, \tag{3.30a}$$

$$\begin{aligned}\mathring{\delta}_L^\dagger\Delta(x)&=\mathring{\delta}_L^\dagger(C^\mu(x)C^\nu(x)\Delta_{\mu\nu}(x)+F(D_{\alpha(n)}C^\mu(x)))\\ &=C^\mu(x)C^\nu(x)S^\dagger\Delta_{\mu\nu}(x)+\mathring{\delta}_L^\dagger\,F(D_{\alpha(n)}C^\mu(x))\\ &\quad+\sum_{n,i}(\delta_\sigma^\mu C^\nu(x)-C^\mu(x)\delta_\sigma^\nu)\\ &\quad\times\left[D_{\alpha(n)}C^\rho(x)\frac{\partial\Delta_{\mu\nu}(x)}{\partial D_{\alpha(n)}\partial_\sigma C^\rho(x)}+D_{\alpha(n)}\Phi_i(x)\frac{\partial\Delta_{\mu\nu}(x)}{\partial D_{\alpha(n)}\partial_\sigma\Phi_i(x)}\right]=0.\end{aligned} \tag{3.30b}$$

So our solutions satisfy the "true" cohomology system Eq. (3.3) by direct substitution and the isomorphism between the two procedures is one.

IV. LOCAL FUNCTIONALS COHOMOLOGY: GENERAL SOLUTION

The purpose of this section is to solve the system of equations [(2.14)] and so to calculate the cohomology of the operator $\mathring{\delta}_L$ in the functional space. In Sec. III we solved the cocycle equation $\mathring{\delta}_L \Delta_0^{n+2}(x) = 0$ and found

$$\Delta_0^{n+2}(x) = C^\mu(x) C^\nu(x) \Delta_{\mu\nu}^n(x) + F(D_{\alpha(n)} C^\mu(x)) + \mathring{\delta}_L \widehat{\Delta}(x), \tag{4.1}$$

for whatever $\widehat{\Delta}(x)$, and Eqs. (3.20) and (3.27) hold.

Now if we substitute the above solution into Eq. (2.14b) we have to find $\Delta_1^{n+1}(x)$, which satisfies

$$\mathring{\delta}_L \Delta_1^{n+1}(x) + \partial_\lambda (C^\mu(x) C^\nu(x) \Delta_{\mu\nu}^n(x)) dx^\lambda + d\, F(D_{\alpha(n)} C^\mu(x)) + d\mathring{\delta}_L \widehat{\Delta}(x) = 0. \tag{4.2}$$

The second term of Eq. (4.2) can be written in a different way, since in a two-dimensional space, the index λ of the derivative operator, always takes the value of one of the ΦΠ fields $C^\mu(x)$, so, for the antisymmetry properties of the $\Delta_{\mu\nu}(x)$ term we can derive

$$\partial_\lambda (C^\mu(x) C^\nu(x) \Delta_{\mu\nu}^n(x)) dx^\lambda = \partial_\lambda (C^\lambda(x) C^\nu(x) \Delta_{\mu\nu}^n(x)) dx^\mu. \tag{4.3}$$

In Sec. III we show that the term $F(D_{\alpha(n)} C^\mu(x))$ always contains two undifferentiated $C^\mu(x)$ fields (otherwise $F=0$) so

$$S\Delta_{\mu\nu}(x) = -\frac{\partial}{\partial C^\mu(x)} \frac{\partial}{\partial C^\nu(x)} \mathring{\delta}_L\, F(D_{\alpha(n)} C^\mu(x)). \tag{4.4}$$

So we can show that

$$\begin{aligned}\partial_\lambda (C^\lambda(x) C^\nu(x) \Delta_{\mu\nu}^n(x)) dx^\mu &= \mathring{\delta}_L (C^\nu(x) \Delta_{\mu\nu}^n(x)) dx^\mu - C^\lambda(x) \frac{\partial}{\partial C^\lambda(x)} \frac{\partial \mathring{\delta}_L}{\partial C^\mu(x)} F(D_{\alpha(n)} C^\mu(x)) \\ &= \mathring{\delta}_L (C^\nu(x) \Delta_{\mu\nu}^n(x)) dx^\mu - N_C \frac{\partial}{\partial C^\mu(x)} \mathring{\delta}_L\, F(D_{\alpha(n)} C^\mu(x)),\end{aligned} \tag{4.5}$$

where N_C is the undifferentiated $C^\mu(x)$ counting operator.

But for reasons already given, we have

$$N_C \frac{\partial}{\partial C^\mu(x)} \mathring{\delta}_L\, F(D_{\alpha(n)} C^\mu(x)) = \frac{\partial}{\partial C^\mu(x)} \mathring{\delta}_L\, F(D_{\alpha(n)} C^\mu(x)), \tag{4.6}$$

and since

$$\left\{ \mathring{\delta}_L, \frac{\partial}{\partial C^\mu(x)} \right\}_+ = \partial_\mu \tag{4.7}$$

we get

$$\begin{aligned}\partial_\lambda (C^\lambda(x) C^\nu(x) \Delta_{\mu\nu}^n(x)) dx^\mu &= \mathring{\delta}_L (C^\nu(x) \Delta_{\mu\nu}(x)) dx^\mu - \partial_\mu\, F(D_{\alpha(n)} C^\rho(x)) dx^\mu \\ &\quad + \mathring{\delta}_L \frac{\partial}{\partial C^\mu(x)} F(D_{\alpha(n)} C^\rho(x)) dx^\mu.\end{aligned} \tag{4.8}$$

So Eq. (4.2) can be written

$$\mathring{\delta}_L \left[\Delta_1^{n+1}(x) + C^\mu(x) \Delta_{\nu\mu}^n(x) dx^\nu + \frac{\partial}{\partial C^\mu(x)} F(D_{\alpha(n)} C^\rho(x)) + d\widehat{\Delta}(x) \right] = 0, \tag{4.9}$$

which is again a cocycle condition, so the most general solution takes the form

$$\begin{aligned}\Delta_1^{n+1}(x) &+ C^\mu(x) \Delta_{\mu\nu}^n(x) dx^\nu + \frac{\partial F}{\partial C^\mu(x)} (D_{\alpha(n)} C^\rho(x)) + d\widehat{\Delta}(x) \\ &= C^\mu(x) C^\nu(x) \sum_{\mu\nu,\sigma} (x) dx^\sigma + G_1 (D_{\alpha(n)} C^\mu(x)) + \mathring{\delta}_L \widehat{\Sigma}_1(x),\end{aligned} \tag{4.10}$$

with

$$C^\mu(x) C^\nu(x) S \sum_{\mu\nu,\sigma}^{n-1} (x) dx^\sigma + \mathring{\delta}_L G_1 (D_{\alpha(n)} C^\rho(x)) = 0, \tag{4.11}$$

and G_1 contains two undifferentiated ghosts $C^\mu(x)$ fields (otherwise $G=0$).

Now, we have to substitute the above solution into Eq. (2.14a), so we get

$$\begin{aligned}\mathring{\delta}_L \Delta_2^n(x) &- \partial_\lambda (C^\mu(x) \Delta_{\nu\mu}^n(x)) dx^\nu \wedge dx^\lambda - \partial_\nu \left[\frac{\partial F}{\partial C^\mu(x)} (D_{\alpha(n)} C^\rho(x)) \right] dx^\mu \wedge dx^\nu \\ &+ \partial_\lambda \left[C^\mu(x) C^\nu(x) \sum_{\mu\nu,\sigma}^{n-1} (x) \right] dx^\sigma \wedge dx^\lambda \\ &+ \partial_\mu G_1 (D_{\alpha(n)} C^\rho(x)) dx^\mu + \partial_\lambda \mathring{\delta}_L \widehat{\Sigma}_1(x) dx^\lambda = 0;\end{aligned} \tag{4.12}$$

but, repeating the above arguments we can derive the identities

$$\begin{aligned}\partial_\lambda (C^\nu(x) \Delta_{\mu\nu}^n(x)) dx^\mu \wedge dx^\lambda &= \partial_\lambda (C^\lambda(x) \Delta_{\mu\nu}^n(x)) dx^\mu \wedge dx^\nu, \\ \partial_\lambda \left(C^\mu(x) C^\nu(x) \sum_{\mu\nu,\sigma}^{n-1} (x) \right) dx^\sigma \wedge dx^\lambda &= \partial_\lambda \left(C^\lambda(x) C^\nu(x) \sum_{\mu\nu,\sigma}^{n-1} (x) \right) dx^\sigma \wedge dx^\nu,\end{aligned} \tag{4.13}$$

and

$$\begin{aligned}\partial_\lambda (C^\lambda(x) \Delta_{\mu\nu}^n(x)) &= \mathring{\delta}_L \Delta_{\mu\nu}^{n+2}(x) - \partial_\nu \frac{\partial F}{\partial C^\mu(x)} (D_{\alpha(n)} C^\mu(x)) + \mathring{\delta}_L \frac{\partial}{\partial C^\mu(x)} \frac{\partial F}{\partial C^\nu(x)} (D_{\alpha(n)} C^\mu(x)), \\ \partial_\lambda \left(C^\lambda(x) C^\nu(x) \sum_{\mu\nu,\sigma}^{n-1} (x) \right) dx^\sigma &= \mathring{\delta}_L \left(C^\nu(x) \sum_{\mu\nu,\sigma}^{n-1} (x) \right) dx^\sigma - \partial_\mu G_1 (D_{\alpha(n)} C^\sigma(x)) \\ &\quad + \mathring{\delta}_L \frac{\partial}{\partial C^\mu(x)} G_1 (D_{\alpha(n)} C^\sigma(x)).\end{aligned} \tag{4.14}$$

So Eq. (4.12) can be written

$$\mathring{\delta}_L\Big[\Delta_2^n(x)-\Delta_{\mu\nu}^n(x)dx^\mu\wedge dx^\nu - \frac{\partial}{\partial C^\mu(x)}\frac{\partial F}{\partial C^\nu(x)}(D_{\alpha(n)}C^\sigma(x))dx^\mu\wedge dx^\nu + C^\mu(x)\sum_{\nu\mu,\sigma}^{n-1}(x)dx^\sigma\wedge dx^\nu + \frac{\partial G_1}{\partial C^\mu(x)}(D_{\alpha(n)}C^\sigma(x))dx^\mu+\partial_\lambda\hat{\Sigma}_1(x)dx^\lambda\Big]=0, \tag{4.15}$$

and the most general solution $\Delta_2^n(x)$ of our system can be derived as

$$\begin{aligned}\Delta_2^n(x)={}&\Delta_{\mu\nu}^n(x)dx^\mu\wedge dx^\nu - C^\nu(x)\sum_{\nu\mu,\sigma}^{n-1}(x)dx^\sigma\wedge dx^\nu\\ &+C^\mu(x)C^\nu(x)\Lambda^{n-2}_{\mu\nu,\sigma\rho}dx^\sigma\wedge dx^\rho\\ &+H_2(D_{\alpha(n)}C^\sigma(x))-\frac{\partial G_1}{\partial C^\mu(x)}(D_{\alpha(n)}C^\sigma(x))dx^\mu\\ &+\frac{\partial}{\partial C^\mu(x)}\frac{\partial F}{\partial C^\nu(x)}(D_{\alpha(n)}C^\sigma(x))dx^\mu\wedge dx^\nu\\ &-\partial_\lambda\hat{\Sigma}_1(x)dx^\lambda+\mathring{\delta}_L\hat{\hat{\Delta}}(x),\end{aligned} \tag{4.16}$$

where the grading with respect the undifferentiated $C^\mu(x)$ is evident; as before $\Lambda_{\mu\nu,\rho\sigma}(x)$ and $H_2(D_{\alpha(n)}C^\sigma(x))$ will satisfy

$$C^\mu(x)C^\nu(x)S\Lambda_{\mu\nu,\rho\sigma}(x)dx^\rho\wedge dx^\sigma + \mathring{\delta}_L H_2(D_{\alpha(n)}C^\sigma(x))=0. \tag{4.17}$$

V. DISCUSSIONS OF THE SUBSIDIARY CONDITIONS

In this section we shall discuss the solutions we derived in Sec. IV.

Indeed, Eq. (4.16) is a very general decomposition [according to underived $C^\mu(x)$ content] for the solution $\Delta_2^n(x)$ $[\Sigma^{n+1}_{\mu\nu,\sigma},\ \Lambda^{n+2}_{\mu\nu,\rho\sigma}(x)]$ we have derived. The more important restrictions are provided by the subsidiary conditions [Eqs. (3.20)–(3.27)]. They strongly depend on the undifferentiated $C^\mu(x)$ content of the function $F(G)$. If the function F contains two undifferentiated $C^\mu(x)$ fields we have

$$C^\mu(x)C^\nu(x)S\Delta_{\mu\nu}(x)+\mathring{\delta}_L F(D_{\alpha(n)}C^\sigma(x))=0, \tag{5.1a}$$

$$C^\mu(x)C^\nu(x)S^\dagger\Delta_{\mu\nu}(x)+\sum_n(\delta^\mu_\sigma C^\nu(x)-C^\mu(x)\delta^\nu_\sigma)\times D_{\alpha(n)}C^\rho(x)\frac{\partial\Delta_{\mu\nu}(x)}{\partial D_{\alpha(n)}\partial_\sigma C^\rho(x)}+\mathring{\delta}_L F(D_{\alpha(n)}C^\sigma(x))=0 \tag{5.1b}$$

[and the condition Eq. (3.20) has to be satisfied]; if not, we have to solve their associated homogeneous equations

$$C^\mu(x)C^\nu(x)S\Delta_{\mu\nu}(x)=0, \tag{5.2a}$$

$$C^\mu(x)C^\nu(x)S^\dagger\Delta_{\mu\nu}(x)=0, \tag{5.2b}$$

$$\sum_n(\delta^\mu_\sigma C^\nu(x)-C^\mu(x)\delta^\nu_\sigma)D_{\alpha(n)}C^\rho(x)\times\frac{\partial\Delta_{\mu\nu}(x)}{\partial D_{\alpha(n)}\partial_\sigma C^\rho(x)}=0 \tag{5.2c}$$

[together with Eq. (3.20)] and the same will happen for the functions $\Sigma_{\mu\nu,\rho}(x)$, $\Lambda_{\mu\nu,\rho\sigma}(x)$, and G_1 (H_2). Since the function $\Delta_{\mu\nu}(x)(\Sigma_{\mu\nu,\rho}(x),\ \Lambda_{\mu\nu,\rho\sigma}(x))$ cannot contain undifferentiated $C^\mu(x)$ fields (and the same happens in the content of the operators S and $S^\dagger$), Eqs. (5.2a) and (5.2b) imply

$$S\Delta_{\mu\nu}(x)=S\Sigma_{\mu\nu,\rho}(x)=S\Lambda_{\mu\nu,\rho\sigma}(x)=0, \tag{5.3a}$$

$$S^\dagger\Delta_{\mu\nu}(x)=S^\dagger\Sigma_{\mu\nu,\rho}(x)=S^\dagger\Lambda_{\mu\nu,\rho\sigma}(x)=0, \tag{5.3b}$$

which identifies the local cohomology space of the operator S.

We shall call anomalies of the first kind the quantities that are solutions of Eq. (5.3), and anomalies of the second kind those that satisfy Eqs. (5.1).

The spectral sequences method is at our disposal to find solutions of Eq. (5.3), as in Ref. 5 in the four-dimensional case. On the other hand, we have not yet found the more general solution of Eqs. (5.1).

Here, where the first kind of anomaly is concerned, we shall briefly sum up and fit to the two-dimensional case the procedure of Ref. 5, and we shall find as only solutions the ordinary trace anomalies.

Furthermore we shall show that the gravitational anomalies found by Bardeen and Zumino[7] verify the subsidiary conditions [Eq. (5.1)].

A. Anomalies of the first kind

The solutions of Eqs. (5.3) can be found following the method shown in Sec. III; so if we filter the operator S with the self-adjoint operator ν'

$$\nu'=2-\partial_\mu C^\nu(x)\frac{\partial}{\partial\partial_\mu C^\nu(x)}, \tag{5.4}$$

we find at the lowest eigenvalue

$$S(1)=\partial_\mu C^\nu(x)[K_\nu^{\ \mu}(x)+\delta_\nu^\mu W(x)], \tag{5.5}$$

where

$$K_\nu^{\ \mu}(x)=R_\nu^{\ \mu}(x)+\sum_{n>1}(n\partial_\nu D_{\alpha(n-1)}C^\rho(x))\times\frac{\partial}{\partial\partial_\mu D_{\alpha(n-1)}C^\rho(x)}-D_{\alpha(n+1)}C^\mu(x)\frac{\partial}{\partial D_{\alpha(n+1)}C^\nu(x)}, \tag{5.6a}$$

$$W(x)=D(x)-1, \tag{5.6b}$$

with the properties

$$[K_\nu^{\ \mu}(x),K_\sigma^{\ \rho}(x)]_-=\delta^\mu_\sigma K_\nu^{\ \rho}(x)-\delta^\rho_\nu K_\sigma^{\ \mu}(x), \tag{5.7a}$$

$$[K_\nu^{\ \mu}(x),W(x)]=0, \tag{5.7b}$$

such that if we call

$$d_1=\partial_\mu C^\nu(x)K_\nu^{\ \mu}(x), \tag{5.8a}$$

$$d_2=\partial_\lambda C^\lambda(x)W(x), \tag{5.8b}$$

we find

$$d_1^2=d_2^2=\{d_1,d_2\}=0, \tag{5.9}$$

and so

$$S^2(1)=0. \tag{5.10}$$

Following the lines of Ref. 5, we can show that even in two dimensions the cohomology of the operator $S(1)$ is isomorphic to the intersections of the operators $d(1)$ and $d(2)$.

Moreover in the scalar sector the anomalies are scalar densities of Weyl weight equal to 1, and (as shown in Ref. 5) do not contain an underived $C^\mu(x)$ field; so the $\Sigma^{n-1}_{\mu\nu,\rho}(x)$, $\Lambda^{n-2}_{\mu\nu,\rho\sigma}(x)$ terms in Eq. (4.16) can be neglected.

The rescaling process we have adopted to redefine the model allows us to find a very interesting identity. In fact, if we introduce the operators (1) dim, the counting operator of the naive uv dimensions, (2) $Q_{\Phi\Pi}$, the Fadeev–Popov charge operator, and (3) $N_{\rm down}$ ($N_{\rm up}$), the counting operator of the lower (upper) indices, we find that

$$2W(x) = \dim - (2 + Q_{\Phi\Pi}) + N_{\rm up} - N_{\rm down}, \qquad (5.11)$$

so that if we study the local scalar anomalies in the $\Phi\Pi$ charge sectors 0 and 1, the dimensional analysis will simplify our research.

The cohomology of the operator can be easily calculated grading the anomalies with respect to the derivatives of the field $C^\mu(x)$; so among the anomalies of the type $\partial_\mu C^\nu(x)\Delta^\mu_\nu(x)$ we find, in the charge-one sector, only the ordinary trace anomalies

$$\Delta_{\nu\rho}(x) = \partial_\lambda C^\lambda(x)\hat{\Delta}(x)dx^\alpha \wedge dx^\beta \qquad (5.12)$$

(where $\hat{\Delta}(x)$ is a Lorentz-invariant density of weight 1), which comes from the consistency conditions

$$K^{\rho}_{\sigma}(x)\Delta^\mu_\nu(x) - K^{\mu}_{\nu}(x)\Delta^\rho_\sigma(x) = \delta^\rho_\nu\Delta^\mu_\sigma(x) - \delta^\mu_\sigma\Delta^\rho_\nu(x), \qquad (5.13a)$$

$$K^{\mu}_{\nu}(x)\Delta^\nu_\mu(x) = 0, \qquad (5.13b)$$

where Eq. (5.13a) is nothing other than the Weiss–Zumino consistency condition of the GL$(2,R)$ group, and the above system is solved with the aid of the commutation relations Eq. (5.7a).

The anomalies of the form $\partial_\mu\partial_\nu C^\rho(x)\Delta^{\mu\nu}_\rho(x)$ that satisfy Eq. (5.3a) can be investigated, in a two-dimensional space, in a very simple way. In fact, if the quantity $\Delta^{\mu\nu}_\rho(x)$ transforms as

$$\begin{aligned}\delta\Delta^{\mu\nu}_\rho(x) = {}& C^\lambda(x)\partial_\lambda\Delta^{\mu\nu}_\rho(x) + \partial_\lambda C^\lambda(x)\Delta^{\mu\nu}_\rho(x)\\ &+ \partial_\rho C^\lambda(x)\Delta^{\mu\nu}_\lambda(x) - \partial_\lambda C^\mu(x)\Delta^{\lambda\nu}_\rho(x)\\ &- \partial_\lambda C^\nu(x)\Delta^{\mu\lambda}_\rho(x) + X^{\mu\nu}_\rho(x), \qquad (5.14)\end{aligned}$$

then Eq. (5.3a) will lead to

$$\partial_\mu\partial_\nu C^\rho(x)X^{\mu\nu}_\rho(x) = 0, \qquad (5.15)$$

which is satisfied by the only two solutions

$$X^{\mu\nu}_\rho(x) = 0, \qquad (5.16a)$$

$$X^{\mu\nu}_\rho(x) = O^{\{\mu,\mu'\}_+\{\nu,\nu'\}_+}_{\{\rho,\rho'\}_+}(x)\partial_{\mu'}\partial_{\nu'}C^{\rho'}(x), \qquad (5.16b)$$

where $O^{\{\mu,\mu'\}_+\{\nu,\nu'\}_+}_{\{\rho,\rho'\}_+}(x)$ is an uv dimension-0 and Lorentz-invariant object with Weyl weight equal to 1, $\{\ \}_+$ means symmetrization, and in our model finds as its only solution:

$$O^{\{\mu,\mu'\}_+\{\nu,\nu'\}_+}_{\{\rho,\rho'\}_+}(x) = g^{\mu\mu'}(x)g^{\nu\nu'}(x)g_{\rho\rho'}(x). \qquad (5.17)$$

Thus it is easy to show that

$$\Delta_{\alpha\beta}(x) = S\Gamma^\rho_{\mu\nu}(x)\Delta^{\mu\nu}_\rho(x)dx^\alpha\wedge dx^\beta \qquad (5.18)$$

in the case of Eq. (5.16a) and

$$\Delta_{\alpha\beta}(x) = S\Gamma^\rho_{\mu\nu}(x)g^{\mu\mu'}(x)g^{\nu\nu'}(x)g_{\rho\rho'}(x)\Gamma^{\rho'}_{\mu'\nu'}(x) \qquad (5.19)$$

in the case of Eq. (5.16b).

The last possibility of the anomalies in the cases we consider here is

$$\Delta_{\alpha\beta}(x) = \partial_\mu\partial_\nu\partial_\rho C^\sigma(x)\Delta^{\mu\nu\rho}_\sigma(x)dx^\alpha\wedge dx^\beta, \qquad (5.20)$$

where $\Delta^{\mu\nu\rho}_\sigma(x)$ is a dimension-0 and -1 Weyl weight object. In our model the only solution is

$$\Delta^{\mu\nu\rho}_\sigma(x) = \tfrac{1}{3}(\delta^\mu_\sigma g^{\nu\rho}(x) + \delta^\nu_\sigma g^{\mu\rho}(x) + \delta^\rho_\sigma g^{\mu\nu}(x)), \qquad (5.21)$$

which is discarded, since it does not verify the cocycle condition Eq. (5.3a).

B. Anomalies of the second kind

As said before, we are not able to give a general solution of Eqs. (5.1). We shall show that the gravitational anomaly

$$\begin{aligned}&\Delta_{\alpha\beta}(x)dx^\alpha\wedge dx^\beta\\ &\quad = \partial_\mu\partial_\sigma C^\rho(x)\epsilon^{\mu\nu}(x)\Gamma^\sigma_{\nu\rho}\mathring{\epsilon}_{\alpha\beta}dx^\nu\, dx^\beta \qquad (5.22)\end{aligned}$$

(where $\mathring{\epsilon}_{\nu\varphi}$ is the flat antisymmetrized tensor) recently made evident by Bardeen and Zumino, will satisfy the above equations.

First of all it is easy to realize that Eq. (5.22) will contain a contribution from an anomaly of the first kind coming from the trace between the indices σ and ρ, or μ and ρ; thus we shall consider only the part that is traceless in these indices.

Indeed the Bardeen–Zumino anomaly satisfies Eq. (5.1a) for

$$\begin{aligned}F(D_{\alpha(n)}C^\rho(x)) = {}& \partial_\rho C^\sigma(x)\mathring{\delta}_L\partial_\sigma C^\rho(x)\\ &- \tfrac{2}{3}\partial_\rho C^\sigma(x)\partial_\sigma C^\mu(x)\partial_\mu C^\rho(x), \qquad (5.23)\end{aligned}$$

indeed

$$\begin{aligned}&C^\alpha(x)C^\beta(x)S\Delta_{\alpha\beta}(x)\\ &\quad = C^\alpha(x)C^\beta(x)\partial_\mu\partial_\rho C^\sigma(x)\epsilon^{\mu\nu}(x)\partial_\nu\partial_\sigma C^\rho(x)\mathring{\epsilon}_{\alpha\beta}\\ &\quad = -C^\mu(x)\partial_\mu\partial_\rho C^\sigma(x)C^\nu(x)\partial_\nu\partial_\sigma C^\rho(x)\\ &\quad = -[\mathring{\delta}_L\partial_\rho C^\sigma(x) - \partial_\rho C^\mu(x)\partial_\mu C^\sigma(x)]\\ &\qquad \times[\mathring{\delta}_L\partial_\sigma C^\rho(x) - \partial_\sigma C^\nu(x)\partial_\nu C^\rho(x)]\\ &\quad = -\mathring{\delta}_L[\partial_\rho C^\sigma(x)\mathring{\delta}_L\partial_\sigma C^\rho(x)\\ &\qquad - \tfrac{2}{3}\partial_\rho C^\sigma(x)\partial_\sigma C^\mu(x)\partial_\mu C^\rho(x)], \qquad (5.24)\end{aligned}$$

since any index of the tensor $\epsilon^{\mu\nu}(x)$ always takes the same value of one index of the tensor $\mathring{\epsilon}_{\alpha\beta}$; moreover we have used the identities

$$\begin{aligned}&C^\mu(x)\partial_\mu\partial_\rho C^\sigma(x)\\ &\quad = \mathring{\delta}_L\partial_\rho C^\sigma(x) - \partial_\rho C^\mu(x)\partial_\mu C^\sigma(x),\\ &\partial_\mu C^\rho(x)\partial_\rho C^\sigma(x)\partial_\sigma C^\nu(x)\partial_\nu C^\mu = 0.\end{aligned} \qquad (5.25)$$

It is now simple (but not straightforward) to realize that, in our case, Eq. (5.1b) is verified. In fact,

$$S^{\dagger}\Delta_{\alpha\beta}(x) = \mathring{\delta}_L \Delta_{\alpha\beta}(x)$$
$$= \left[\Gamma^{\sigma}_{\eta\rho} \frac{\partial}{\partial\partial_\eta \partial_\rho C^{\sigma}(x)} + \partial_\delta \Gamma^{\rho}_{\eta\sigma}(x) \frac{\partial}{\partial\partial_\delta\partial_\eta C^{\lambda}(x)} \frac{\partial}{\partial \Gamma^{\rho}_{\eta\sigma}(x)} + \partial_\delta \Gamma^{\rho}_{\eta\sigma}(x) \frac{\partial}{\partial\partial_\delta\partial_\sigma C^{\lambda}(x)} \frac{\partial}{\partial\Gamma^{\rho}_{\eta\lambda}(x)} \right.$$
$$\left. - \partial_\delta \Gamma^{\lambda}_{\eta\sigma}(x) \frac{\partial}{\partial\partial_\delta\partial_\lambda C^{\rho}(x)} \frac{\partial}{\partial\Gamma^{\lambda}_{\eta\sigma}(x)} \right] \epsilon^{\mu\nu}(x)\partial_\nu\partial_\gamma C^{\tau}(x)\Gamma^{\gamma}_{\mu\tau}(x)\mathring{\epsilon}_{\alpha\beta}$$
$$= \left[\Gamma^{\rho}_{\eta\sigma}(x)\epsilon^{\eta\tau}(x)\Gamma^{\sigma}_{\tau\rho}(x) + \partial_\delta \Gamma^{\rho}_{\eta\sigma}(x)\{\delta^{\delta\eta}_{\nu\gamma}\delta^{\tau}_{\lambda}\delta^{\lambda\beta}_{\mu\tau}\delta^{\gamma}_{\rho} + \delta^{\delta\sigma}_{\nu\gamma}\delta^{\tau}_{\lambda}\delta^{\eta\lambda}_{\mu\tau}\delta^{\gamma}_{\rho} - \delta^{\delta\lambda}_{\nu\gamma}\delta^{\tau}_{\rho}\delta^{\eta\sigma}_{\mu\tau}\delta^{\gamma}_{\lambda}\}\epsilon^{\mu\nu}(x) \right]\mathring{\epsilon}_{\alpha\beta}$$
$$= \partial_\delta \Gamma^{\rho}_{\eta\sigma}(x)\left[\delta^{\delta\eta}_{\nu\rho}\delta^{\tau\sigma}_{\mu\tau} + \delta^{\delta\sigma}_{\nu\rho}\delta^{\eta\tau}_{\mu\tau} - \delta^{\delta\gamma}_{\nu\gamma}\delta^{\eta\sigma}_{\mu\sigma}\right]\epsilon^{\mu\nu}(x)\mathring{\epsilon}_{\alpha\beta}$$
$$= \tfrac{3}{2}\partial_\sigma \Gamma^{\rho}_{\eta\sigma}(x)\left[\delta^{\delta\eta}_{\nu\rho}\delta^{\sigma}_{\mu} + \delta^{\delta\sigma}_{\nu\rho}\delta^{\eta}_{\mu} - \delta^{\eta\sigma}_{\mu\rho}\delta^{\delta}_{\nu}\right]\epsilon^{\mu\nu}(x)\mathring{\epsilon}^{\alpha\beta}$$
$$= \tfrac{3}{2}\partial_\rho \Gamma^{\rho}_{\mu\nu}\epsilon^{\mu\nu}(x)\mathring{\epsilon}_{\alpha\beta} = 0, \tag{5.26}$$

since our connection is torsionless [we have indicated $\delta^{\mu\nu}_{\rho\sigma} = \frac{1}{2}(\delta^{\mu}_{\rho}\delta^{\nu}_{\sigma} + \delta^{\mu}_{\sigma}\delta^{\nu}_{\rho})$] and

$$\mathring{\delta}^{\dagger}_L F(D_{\alpha(n)}C^{\rho}(x)) = \left[\partial_\sigma C^{\lambda}(x) \frac{\partial}{\partial\partial_\sigma C^{\rho}(x)} \frac{\partial}{\partial\partial_\rho C^{\lambda}(x)} + \partial_\rho C^{\sigma}(x) \frac{\partial}{\partial C^{\mu}(x)} \frac{\partial}{\partial\partial_\mu\partial_\rho C^{\sigma}(x)} \right]$$
$$\times \left[\mathring{\delta}_L \partial_\rho C^{\sigma}(x) - \tfrac{2}{3}\partial_\alpha C^{\beta}(x)\partial_\beta C^{\gamma}(x)\partial_\gamma C^{\alpha}(x) \right] = 0, \tag{5.27}$$

since it gives only terms of the form $\partial_\lambda C^{\lambda}(x)\partial_\sigma C^{\sigma}(x)$ and $\partial_\lambda C^{\sigma}(x)\partial_\sigma C^{\lambda}(x)$, which are zero for $\Phi\Pi$ charge reasons.

Furthermore,

$$(\delta^{\mu}_{\rho}C^{\nu}(x) - C^{\mu}(x)\delta^{\nu}_{\rho})\partial_\alpha C^{\lambda}(x) \frac{\partial}{\partial\partial_\alpha\partial_\rho C^{\lambda}(x)}$$
$$\times \left[\mathring{\epsilon}_{\mu\nu}\epsilon^{\tau\eta}(x)\Gamma^{\sigma}_{\tau\gamma}(x)\partial_\eta\partial_\sigma C^{\gamma}(x)\right]$$
$$= (\delta^{\mu}_{\rho}C^{\nu}(x) - C^{\mu}(x)\delta^{\nu}_{\rho})\partial_\alpha C^{\lambda}(x)$$
$$\times \left[\mathring{\epsilon}_{\mu\nu}\epsilon^{\tau\eta}(x)\Gamma^{\sigma}_{\tau\gamma}(x)\delta^{\alpha\rho}_{\eta\sigma}\delta^{\gamma}_{\lambda}\right]$$
$$= (\delta^{\mu}_{\rho}C^{\nu}(x) - C^{\mu}(x)\delta^{\nu}_{\rho})\partial_\alpha C^{\lambda}(x)$$
$$\times \left[\delta^{\tau}_{\mu}\delta^{\eta}_{\nu} - \delta^{\tau}_{\nu}\delta^{\eta}_{\mu}\right]\Gamma^{\sigma}_{\tau\gamma}(x)\delta^{\alpha\rho}_{\eta\sigma}\delta^{\gamma}_{\lambda}$$
$$= C^{\nu}(x)\partial_\nu C^{\gamma}(x)\Gamma^{\sigma}_{\gamma\sigma}(x) = 0, \tag{5.28}$$

since, as said previously, in the "pure" second kind anomaly the trace contributions have been taken away.

Finally the condition [Eq. (3.20)]

$$\partial^{\dagger}_{\rho}|_{C^{\mu}=\mathrm{const}}(\delta^{\mu}_{\rho}C^{\nu}(x) - C^{\mu}(x)\delta^{\nu}_{\rho})\Delta_{\mu\nu}(x) = 0 \tag{5.29}$$

is immediately verified since the operator $\partial^{\dagger}_{\rho}|_{C^{\mu}=\mathrm{const}}$ gives a nonzero contribution only on derived fields different from the $\Phi\Pi$ $C^{\mu}(x)$ ghost.

So, all the equations (5.1a)–(5.1c) have been satisfied.

VI. CONCLUSIONS

We have studied the cohomological birth of the gravitational anomalies (in particular of the Bardeen–Zumino one) within the spectral sequences mathematical framework.

It is well known that these pathologies can be swept away in a nonpolynomial phenomenological Lagrangian formulation of the gravitational interactions.

Where the second kind of anomaly is concerned, this is clearly shown in Ref. 7, to which we refer the reader for an exhaustive treatment.

On the other hand, the first kind of anomaly was interpreted, in Ref. 13, as representation instabilities in the Weyl sector; their cancellation was carried away through the introduction of a scalar field with zero dimensions leading to an infinite tower of polynomial terms in the invariant Lagrangian. This was shown in Ref. 9 to be equivalent to reformulating ours in a nonpolynomial language.

It would be useful to extend the spectral sequences method to nonpolynomial theories to verify the absence of anomalies in this framework; but, even if many mathematical tools (such as the continuous filtration procedure) and many important theorems are at our hand, the problem is, at present, too difficult for us.

ACKNOWLEDGMENTS

We are grateful to Professor M. Tonin and his coworkers for stimulating this problem and for helpful discussions. Conversations with Dr. P. Truini are gratefully acknowledged.

APPENDIX: SPECTRAL SEQUENCES

In this appendix we shall show a pictorial view, due to Zeeman,[18] which is very useful to physicists for a fast introduction to spectral sequences techniques.

With the filtration operator ν we have divided the space F in the sectors G, whose elements are eigenvectors of ν with eigenvalue p, and introduced the nested spaces F_p, such that

$$F \equiv F_0 \supset F_1 \supset F_2 \supset \cdots \supset F_m, \tag{A1}$$

and

$$\mathring{\delta}_L^{-1}F_0 \supset \mathring{\delta}_L^{-1}F_1 \supset \cdots \supset \mathring{\delta}_L^{-1}F_m \supset \mathring{\delta}_L^{-1}0$$
$$\supset \mathring{\delta}_L F \supset \mathring{\delta}_L F_1 \cdots \supset \mathring{\delta}_L F_m. \tag{A2}$$

In the Euclidean plane, let $[\pi, \rho]$ be the unit square given by $\pi - 1 < x < \pi, \rho - 1 < y < \rho$; where π and ρ are integers.

Let us define a two-dimensional lattice

$$\Delta \equiv \cup\{[\pi,\rho]; -(m-1) \leqslant \pi \leqslant 0; \pi - m - 1 \leqslant \rho \leqslant \pi\}, \tag{A3}$$

and

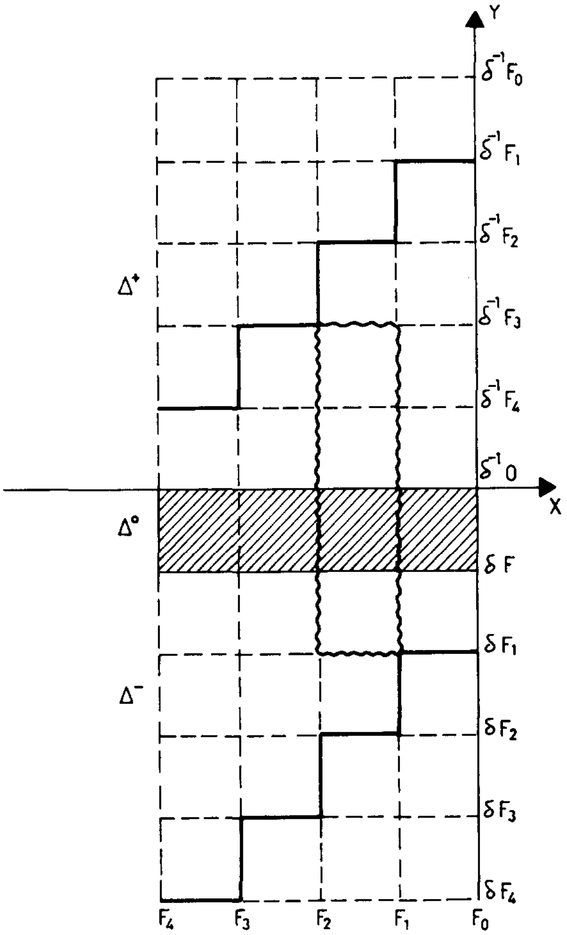

FIG. 1. Zeeman diagram.

$$\Delta^+ = \begin{cases} \Delta, & \text{in the half plane with } \rho \geqslant 0, \\ 0, & \text{otherwise,} \end{cases}$$

$$\Delta^0 = \begin{cases} \Delta, & \text{in the region where } \rho = 0, \\ 0, & \text{otherwise,} \end{cases}$$

$$\Delta^- \begin{cases} \Delta, & \text{in the region with } \rho < 0, \\ 0, & \text{otherwise.} \end{cases}$$

In this lattice the representatives of the quantities F_p, $\mathring{\delta}_L F_p$, and $\mathring{\delta}_L^{-1} F_p$ are defined, respectively, as

$$\lambda(F_p) = \Delta \quad \text{for} \quad \pi \leqslant -\text{p},$$

$$\lambda(\mathring{\delta}_L F_p) = \Delta \quad \text{for} \quad \rho \leqslant -(p+1),$$

$$\lambda(\mathring{\delta}_L^{-1} F_p) = \Delta \quad \text{for } \rho \leqslant m - p.$$

In Fig. 1 we represent the case for $m = 4$.

Using the language of the set theory we can make evident, by intersections of rectangles (which are made individual by the above representatives), the cohomology region

$$\lambda(H) = \lambda(\mathring{\delta}_L^{-1} O)/\lambda(\mathring{\delta}_L F) \tag{A4}$$

as the central shadowed rectangle in Fig. 1.

The regions $\lambda(E_r^p)$, where

$$E_r^p = \frac{F_p \cap \mathring{\delta}_L F_{p+r}}{\mathring{\delta}_L F_{p-r+1} \cap F_p + \mathring{\delta}_L F_{p+r} \cap F_{p+1}}, \tag{A5}$$

can be represented in the same way. In our example (Fig. 1), the region contained in the rectangle closed by the wiggly lines represents $\lambda\ (E_1^1)$.

If we calculate $\lambda\ (E_2^1)$ as before, the rectangle shrinks, loosing the top and the bottom squares.

In general, $\lambda(E_{r+1}^p)$ is derived from $\lambda(E_r^p)$ by omitting the top square, if it is in Δ^+, and the bottom square if it is in Δ^-: that is

$$\lambda\,[E_{r+1}^p] = \frac{\lambda\,[E_r^p] \cap \lambda\,[\mathring{\delta}_L^{-1} F_{p+r+1}]}{\lambda\,[E_r^p] \cap \lambda\,[\delta F_{p-r}]}, \tag{A6}$$

if $\lambda(\mathring{\delta}_L^{-1} F_{p+r+1}) \cap \Delta^+ \neq \Phi$, and $\lambda(\mathring{\delta}_L F_{p-r}) \cap \Delta^- \neq \Phi$. After at most m steps, we are left with a single square in Δ^0, so, summing on p, we reach the cohomology region: this is the statement of the theorem of Serre we quoted in Sec. III.

On the other hand it is evident that the same set of squares, as, for example, $\lambda(E_r^p)$, can be identified in several manners; for example,

$$\lambda\,[E_r^p] = \frac{(\lambda\,[F_p] \cap \lambda\,[\mathring{\delta}_L^{-1} F_{p+r}])/(\lambda\,[F_{p+1}] \cap \lambda\,[\mathring{\delta}_L^{-1} F_{p+r}])}{(\lambda\,[F_p] \cap \lambda\,[\mathring{\delta}_L^{-1} F_{p+r}])/(\lambda\,[F_{p+1}] \cap \lambda\,[\mathring{\delta}_L F_{p-r+1}])}. \tag{A7}$$

Zeeman has shown in Ref. 8 that the different ways one can make individual the same set of rectangles, will make individual the isomorphisms between the spaces whose representatives the same regions, for example,

$$E_r^p = \frac{(F_p \cap \mathring{\delta}_L^{-1} F_{p+r})/F_{p+1} \cap \mathring{\delta}_L^{-1} F_{p+r})}{(F_p \cap \mathring{\delta}_L F_{p+r})/(F_{p+1} \cap \mathring{\delta}_L F_{p-r+1})}. \tag{A8}$$

If we apply these results to Eq. (A6), the isomorphism Eq. (3.11) is now evident; in this philosophy all the isomorphisms derived by Dixon in Ref. 3 are easily derived.

These techniques give a pictorial view of many other theorems, which in the mathematical language will need (for a physicist not specialized in the sector) a lengthy introduction of the formalism.

Since we do not use them, we refer the interested reader to the paper of Zeeman for a higher level course on spectral sequences.

[1]C. Becchi, A. Rouet, and R. Stora, Ann. Phys. (NY) **98**, 287 (1976).
[2]B. W. Lee and S. Joglekar, Ann. Phys. (NY) **97**, 160 (1975).
[3]J. Dixon, "Cohomology and renormalization of gauge fields," I, II, and III, preprints, 1977–1978.
[4]H. Godement, *Les fascieaux fibrees* (Hermann, Paris, 1952).
[5]G. Bandelloni, Nuovo Cimento A **88**, 35 (1985).
[6]L. Alvarez-Gaume and E. Witten, Nucl. Phys. B **234**, 269 (1984).
[7]J. Bardeen and B. Zumino, Nucl. Phys. B **224**, 421 (1984).
[8]F. Langouche, T. Shucker, and R. Stora, Phys. Lett. B **145**, 342 (1984).
[9]L. Bonora, P. Pasti, and M. Tonin, "The anomaly structure of theories with external gravity," Padova preprint, 1985.

[10] L. Baulieu and J. Thierry-Mieg, Phys. Lett. B **145**, 53 (1984).
[11] F. Langouche, Phys. Lett. B **148**, 342 (1984).
[12] H. Leutwyler, "Gravitational anomalies: A soluble two-dimensional model," Univ. of Berne preprint, 1985.
[13] C. G. Callan, Phys. Rev. D **2**, 1541 (1970); K. Symanzik, Commun. Math. Phys. **19**, 227 (1970).
[14] G. Bandelloni, C. Becchi, A. Blasi, and R. Collina, Nucl. Phys. B **197**, 347 (1982); G. Bandelloni, Nuovo Cimento A **86**, 300 (1985).
[15] R. Stora, "Continuum gauge theory," in *New Developments in Quantum Field Theory and Statistical Mechanics: 1976 Cargese Lectures*, edited by M. Levy and P. Mitten (Plenum, New York, 1977).
[16] M. Grandis, Cah. Topo. Geom. Diff. **26**, 169 (1985).
[17] J. Serre, Ann. Math. **54**, 425 (1951).
[18] E. C. Zeeman, Ann. Math. **66**, 557 (1957).
[19] This is accomplished, in our case, by building the filtration operator only with the ΦΠ charged fields and their derivatives, since, as we shall show, we will be interested in anomalies with definite ΦΠ charge, uv dimensions, and Weyl weight.

Creating and annihilating Lie–Bäcklund transformations of the Federbush model

Paul H. M. Kersten
Department of Applied Mathematics, Post Office Box 217, 7500 AE Enschede, The Netherlands

(Received 3 September 1985; accepted for publication 26 November 1985)

Using software developed for symbolic integration, infinitesimal symmetries, conserved currents, and first- and second-order Lie–Bäcklund transformations for the Federbush model are established. Moreover four (x,t)-dependent Lie–Bäcklund transformations are constructed leading to infinite hierarchies of Lie–Bäcklund transformations.

I. INTRODUCTION AND GENERAL

In recent papers[1,2] we studied local and nonlocal Lie–Bäcklund transformations of the massive Thirring model. Ruijsenaars[3] discussed scattering theory for the Federbush,[4] the massless Thirring, and the continuum Ising model. In this paper we shall construct infinitesimal symmetries and conserved currents for the Federbush model. These infinitesimal symmetries have to satisfy the condition[5]

$$\mathscr{L}_V I \subset I, \tag{1.1}$$

where V is a vector field, $\mathscr{L}_V$ denotes the Lie derivative with respect to the vector field V, and I denotes a closed ideal of differential forms describing the Federbush model.

Introducing a Lagrangian for the Federbush model, we obtain conserved currents associated to infinitesimal symmetries, using Noether's theorem.[6,7] This will be done in Sec. II.

Using the local jet bundle formulation,[7] (1.1) can be generalized to Lie–Bäcklund transformations, which are formal vector fields defined on the infinite jet bundle J^∞ and which have to satisfy the condition

$$\mathscr{L}_V (D^\infty I) \subset D^\infty I, \tag{1.2}$$

where $D^\infty I$ is the infinite prolongation of the ideal I, and which amounts classically to the partial differential equation associated to I together with its differential consequences.[6]

From condition (1.2) we obtain the equivalent condition

$$\mathscr{L}_V I \subset D^\infty I, \tag{1.3}$$

due to the fact that the total derivative vector fields D_x and D_t commute. Moreover, by the assumption that the generating functions of the Lie–Bäcklund transformations only depend on a finite number of variables, condition (1.3) reduces to

$$\mathscr{L}_V I \subset D^k I, \quad \text{for some } k. \tag{1.4}$$

First- and second-order Lie–Bäcklund transformations are obtained in Sec. III, leading to eight (x,t)-independent and two (x,t)-dependent Lie–Bäcklund transformations. Furthermore, the Lie algebra structure is given.

In Sec. IV we construct four creating and annihilating Lie–Bäcklund transformations, which turn out to be *local* vector fields, in contrast with the two creating and annihilating Lie–Bäcklund transformations of the massive Thirring model,[2] which are *nonlocal* vector fields.

Finally, the Lie algebraic structure is discussed.

In the Appendix, one (of four) (x,t)-independent third-order Lie–Bäcklund transformation is given.

We carried through all calculations on a DEC system 20 computer using the symbolic language REDUCE 3.0 (see Ref. 8). We used software to do differential geometric computations developed by Gragert *et al.*[9,10] and software developed to solve overdetermined systems of partial differential equations.[11]

II. INFINITESIMAL SYMMETRIES AND CONSERVED CURRENTS FOR THE FEDERBUSH MODEL

In this section the Lie algebra of infinitesimal symmetries and conserved currents for the Federbush model are established. The Federbush model is described by

$$\begin{pmatrix} i(\partial_t + \partial_x) & -m(s) \\ -m(s) & i(\partial_t - \partial_x) \end{pmatrix} = 4s\pi\lambda \begin{pmatrix} |\psi_{-s,2}|^2 \psi_{s,1} \\ -|\psi_{-s,1}|^2 \psi_{s,2} \end{pmatrix},$$

$$s = \pm 1, \tag{2.1}$$

where $\psi_s(x,t)$ are two-component functions defined on $\mathbb{C}$ (see Ref. 4).

Suppressing the factor 4π from now on ($\lambda' = 4\pi\lambda$) and introducing the variables u_1, v_1, u_2, v_2, u_3, v_3, u_4, and v_4 by

$$\begin{aligned} &\psi_{+1,1} = u_1 + iv_1, \quad \psi_{+1,2} = u_2 + iv_2, \quad m(+1) = m_1, \\ &\psi_{-1,1} = u_3 + iv_3, \quad \psi_{-1,2} = u_4 + iv_4, \quad m(-1) = m_2, \end{aligned} \tag{2.2}$$

Eq. (2.1) is rewritten as a system of eight nonlinear partial differential equations for the component functions $u_1,\dots,v_4$, i.e.,

$$\begin{aligned}
u_{1t} + u_{1x} - m_1 v_2 &= \lambda(u_4^2 + v_4^2)v_1, \\
-v_{1t} - v_{1x} - m_1 u_2 &= \lambda(u_4^2 + v_4^2)u_1, \\
u_{2t} - u_{2x} - m_1 v_1 &= -\lambda(u_3^2 + v_3^2)v_2, \\
-v_{2t} + v_{2x} - m_1 u_1 &= -\lambda(u_3^2 + v_3^2)u_2, \\
u_{3t} + u_{3x} - m_2 v_4 &= -\lambda(u_2^2 + v_2^2)v_3, \\
-v_{3t} - v_{3x} - m_2 u_4 &= -\lambda(u_2^2 + v_2^2)u_3, \\
u_{4t} - u_{4x} - m_2 v_3 &= \lambda(u_1^2 + v_1^2)v_4, \\
-v_{4t} + v_{4x} - m_2 m_3 &= \lambda(u_1^2 + v_1^2)u_4.
\end{aligned} \tag{2.3}$$

The closed ideal I of differential forms defined on

$$\mathbf{R}^{18} = \{(x,t,u_1,\dots,v_4,u_{1x},\dots,v_{4x})\},$$

 0022-2488/86/041139-06$02.50

describing (2.3), is generated by the eight one-forms

$$
\begin{aligned}
\alpha_1 &= du_1 - u_{1x}\, dx - H(1,1)dt\,, \\
\alpha_2 &= dv_1 - v_{1x}\, dx - H(1,2)dt\,, \\
\alpha_3 &= du_2 - u_{2x}\, dx - H(1,3)dt\,, \\
\alpha_4 &= dv_2 - v_{2x}\, dx - H(1,4)dt\,, \\
\alpha_5 &= du_3 - u_{3x}\, dx - H(1,5)dt\,, \\
\alpha_6 &= dv_3 - v_{3x}\, dx - H(1,6)dt\,, \\
\alpha_7 &= du_4 - u_{4x}\, dx - H(1,7)dt\,, \\
\alpha_8 &= dv_4 - v_{4x}\, dx - H(1,8)dt\,,
\end{aligned}
\tag{2.4}
$$

and their exterior derivatives $d\alpha_1,\dots,d\alpha_8$.

The functions $H(1,i)$ $(i = 1,\dots,8)$ are obtained from (2.3) by solving these equations for $u_{1t},\dots,v_{4t}$.

Now, a vector field V defined on $\mathbb{R}^{18}$ is an infinitesimal symmetry of the ideal I (2.4), if

$$\mathscr{L}_V I \subset I\,. \tag{2.5}$$

Condition (2.5) leads to an overdetermined system of 16 partial differential equations for the components of this vector field.

Using the fact that the Lie algebra of infinitesimal symmetries of (2.4) only consists of Lie point symmetries,[6] we solved the resulting overdetermined system of partial differential equations using the developed software,[11] leading to a five-dimensional Lie algebra generated by

$$
\begin{aligned}
V_1 &= \partial_x, \quad V_2 = \partial_t\,, \\
V_3 &= t\partial_x + x\partial_t + \tfrac{1}{2}(u_1\partial_{u_1} + v_1\partial_{v_1} - u_2\partial_{u_2} - v_2\partial_{v_2} \\
&\quad + u_3\partial_{u_3} + v_3\partial_{v_3} - u_4\partial_{u_4} - v_4\partial_{v_4})\,, \\
V_4 &= -v_1\partial_{u_1} + u_1\partial_{v_1} - v_2\partial_{u_2} + u_2\partial_{v_2}\,, \\
V_5 &= -v_3\partial_{u_3} + u_3\partial_{v_3} - v_4\partial_{u_4} + u_4\partial_{v_4}\,.
\end{aligned}
\tag{2.6}
$$

In (2.6) only the components of vector fields on $J^0(\mathbb{R}^2,\mathbb{R}^8)$ are given; the other components are obtained by prolongation.[6,7]

We now introduce the Lagrangian L by

$$
\begin{aligned}
L &= u_{1x}v_1 - u_1v_{1x} - u_{2x}v_2 + u_2v_{2x} + u_{3x}v_3 - u_3v_{3x} \\
&\quad - u_{4x}v_4 + u_4v_{4x} + u_{1t}v_1 - u_1v_{1t} + u_{2t}v_2 \\
&\quad - u_2v_{2t} + u_{3t}v_3 - u_3v_{3t} + u_{4t}v_t - u_4v_{4t} \\
&\quad - 2m_1(u_1u_2 + v_1v_2) - 2m_2(u_3u_4 + v_3v_4) \\
&\quad - \lambda(u_1^2 + v_1^2)(u_4^2 + v_4^2) + \lambda(u_2^2 + v_2^2)(u_3^2 + v_3^2)\,.
\end{aligned}
\tag{2.7}
$$

A straightforward computation shows that the Euler–Lagrange equations associated to (2.7) are just (2.3).

Applying Noether's theorem[6] to the infinitesimal symmetries this leads to five conserved currents associated to the vector fields V_i $(i = 1,\dots,5)$ (2.6),

$$C^i = C_x^i\, dx + C_t^i\, dt \quad (i = 1,\dots,5)\,, \tag{2.8}$$

which satisfy

$$dC^i \in I\,.$$

In (2.8) the coefficients C_x^i and C_t^i $(i = 1,\dots,5)$ are given by

$$
\begin{aligned}
C_x^1 &= u_{1x}v_1 - u_1v_{1x} + u_{2x}v_2 - u_2v_{2x} \\
&\quad + u_{3x}v_3 - u_3v_{3x} + u_{4x}v_4 - u_4v_{4x}\,, \\
C_t^1 &= -u_{1x}v_1 + u_1v_{1x} + u_{2x}v_2 - u_2v_{2x} - u_{3x}v_3 \\
&\quad + u_3v_{3x} + u_{4x}v_4 - u_4v_{4x} + \lambda(R_1R_4 - R_2R_3)\,, \\
C_x^2 &= -u_{1x}v_1 + u_1v_{1x} + u_{2x}v_2 - u_2v_{2x} - u_{3x}v_3 + u_3v_{3x} \\
&\quad + u_{4x}v_4 - u_4v_{4x} + 2m_1(u_1u_2 + v_1v_2) \\
&\quad + 2m_2(u_3u_4 + v_3v_4) + \lambda(R_1R_4 - R_2R_3)\,, \\
C_t^2 &= u_{1x}v_1 - u_1v_{1x} + u_{2x}v_2 - u_2v_{2x} \\
&\quad + u_{3x}v_3 - u_3v_{3x} + u_{4x}v_4 - u_4v_{4x}\,, \\
C_x^3 &= x\,C_x^2 + t\,C_x^1, \quad C_t^3 = xC_t^2 + t\,C_t^1\,, \\
C_x^4 &= R_1 + R_2, \quad C_t^4 = -R_1 + R_2\,, \\
C_x^5 &= R_3 + R_4, \quad C_t^5 = -R_3 + R_4\,.
\end{aligned}
\tag{2.9}
$$

In (2.9) we introduced the notation

$$
\begin{aligned}
R_1 &= u_1^2 + u_1^2\,, \quad R_2 = u_2^2 + v_2^2\,, \\
R_3 &= u_3^2 + v_3^2\,, \quad R_4 = u_4^2 + v_4^2\,.
\end{aligned}
\tag{2.10}
$$

III. FIRST- AND SECOND-ORDER LIE–BÄCKLUND TRANSFORMATIONS OF THE FEDERBUSH MODEL

In this section we construct first- and second-order Lie–Bäcklund transformations of the Federbush model. In obtaining the results there is the remarkable fact that there exist Lie–Bäcklund transformations of first order, which are not equivalent to infinitesimal symmetries.[6] The Lie algebra structure of these Lie–Bäcklund transformations is given.

In order to derive first-order Lie–Bäcklund transformations of the Federbush model satisfying the condition (1.4),

$$\mathscr{L}_V I \subset DI\,, \tag{3.1}$$

we have to prolong the ideal I; i.e., DI is generated by $\alpha_1,\dots,\alpha_8$, $d\alpha_1,\dots,d\alpha_8$, as given in (2.4),

$$
\begin{aligned}
\alpha_9 &= du_{1x} - u_{1xx}\, dx - H(2,1)dt\,, \\
\alpha_{10} &= dv_{1x} - v_{1xx}\, dx - H(2,2)dt\,, \\
\alpha_{11} &= du_{2x} - u_{2xx}\, dx - H(2,3)dt\,, \\
\alpha_{12} &= dv_{2x} - v_{2xx}\, dx - H(2,4)dt\,, \\
\alpha_{13} &= du_{3x} - u_{3xx}\, dx - H(2,5)dt\,, \\
\alpha_{14} &= dv_{3x} - v_{3xx}\, dx - H(2,6)dt\,, \\
\alpha_{15} &= du_{4x} - u_{4xx}\, dx - H(2,7)dt\,, \\
\alpha_{16} &= dv_{2x} - v_{4xx}\, dx - H(2,8)dt\,,
\end{aligned}
\tag{3.2}
$$

and the exterior derivatives $d\alpha_9,\dots,d\alpha_{16}$.

In (3.2), $H(2,i)$ $(i = 1,\dots,8)$ is obtained from $H(1,i)$ $(i = 1,\dots,8)$ by total partial differentiation with respect to x. In the search for Lie–Bäcklund transformations we are interested only in vertical vector fields, i.e., the ∂_x- and ∂_t-components are zero, due to the fact that the total derivative vector fields D_x and $D_t^?$ satisfy (1.2) in a trivial way. Our first search was for (x,t)-independent vector fields V, the components of which are dependent on

$$u_1,\dots,v_4,u_{1x},\dots,v_{4x}\,, \tag{3.3}$$

which resulted in four Lie–Bäcklund transformations of first order, $X_1,\dots,X_4$:

$$\begin{aligned}X_1 &= \frac{\lambda}{2}v_1R_4\partial_{u_1} - \frac{\lambda}{2}u_1R_4\partial_{v_1} + \frac{\lambda}{2}v_2R_4\partial_{u_2} - \frac{\lambda}{2}u_2R_4\partial_{v_2}\\ &\quad + \tfrac{1}{2}m_2v_4\partial_{u_3} - \tfrac{1}{2}m_2u_4\partial_{v_3}\\ &\quad + \tfrac{1}{2}(2u_{4x} + m_2v_3 + \lambda v_4(R_1+R_2))\partial_{u_4}\\ &\quad + \tfrac{1}{2}(2v_{4x} - m_2u_3 - \lambda u_4(R_1+R_2))\partial_{v_4}\,,\\ X_2 &= \frac{\lambda}{2}v_1R_3\partial_{u_1} - \frac{\lambda}{2}u_1R_3\partial_{v_1} + \frac{\lambda}{2}v_2R_3\partial_{u_2} - \frac{\lambda}{2}u_2R_3\partial_{v_2}\\ &\quad + \tfrac{1}{2}(2u_{3x} - m_2v_4 + \lambda v_3(R_1+R_2))\partial_{u_3}\\ &\quad + \tfrac{1}{2}(2v_{3x} + m_2u_4 - \lambda u_3(R_1+R_2))\partial_{v_3}\\ &\quad - \tfrac{1}{2}m_2v_3\partial_{u_4} + \tfrac{1}{2}m_2u_3\partial_{v_4}\,, \qquad (3.4)\\ X_3 &= \tfrac{1}{2}m_1v_2\partial_{u_1} - \tfrac{1}{2}m_1u_2\partial_{v_1}\\ &\quad + \tfrac{1}{2}(2u_{2x} + m_1v_1 - \lambda v_2(R_3+R_4))\partial_{u_2}\\ &\quad + \tfrac{1}{2}(2v_{2x} - m_1u_1 + \lambda u_2(R_3+R_4))\partial_{v_2}\\ &\quad - \frac{\lambda}{2}v_3R_2\partial_{u_3} + \frac{\lambda}{2}u_3R_2\partial_{v_3}\\ &\quad - \frac{\lambda}{2}v_4R_2\partial_{u_4} + \frac{\lambda}{2}u_2R_2\partial_{v_4}\,,\\ X_4 &= \tfrac{1}{2}(2u_{1x} - m_1v_2 - \lambda v_1(R_3+R_4))\partial_{u_1}\\ &\quad + \tfrac{1}{2}(2v_{1x} + m_1u_2 + \lambda u_1(R_3+R_4))\partial_{v_1}\\ &\quad - \tfrac{1}{2}m_1v_1\partial_{u_2} + \tfrac{1}{2}m_1u_1\partial_{v_2}\\ &\quad - \frac{\lambda}{2}v_3R_1\partial_{u_3} + \frac{\lambda}{2}u_3R_1\partial_{v_3}\\ &\quad - v_4R_1\partial_{u_4} + \frac{\lambda}{2}u_4R_1\partial_{v_4}\,.\end{aligned}$$

Note that

$$\begin{aligned}&X_1+X_2+X_3+X_4 \doteq -\partial_x\,,\\ &X_1-X_2+X_3-X_4 \doteq -\partial_t\,,\\ &X_2+X_4 \doteq -\tfrac{1}{2}(\partial_x-\partial_t)\,,\quad X_1+X_3 \doteq -\tfrac{1}{2}(\partial_x-\partial_t)\,.\end{aligned} \tag{3.5}$$

Two vector fields X_a and X_b are equivalent (notation $X_a \doteq X_b$), if there exist functions $f,g \in C^\infty(J^\infty,\mathbf{R})$, such that

$$X_a = X_b + fD_x + gD_t\,, \tag{3.5'}$$

where D_x and D_t are the total partial derivative vector fields.

The Lie–Bäcklund transformations $X_1 \dots X_4$ of the Federbush model (2.3) were found using the following grading:

$$\begin{aligned}&\deg(x)=\deg(t)=-2,\quad \deg(\partial_x)=\deg(\partial_t)=2\,,\\ &\deg(u_1)=\cdots=\deg(v_4)=1\,,\\ &\deg(\partial_{u_1})=\cdots=\deg(\partial_{v_1})=-1\,,\\ &\deg(m_1)=\deg(m_2)=2\,.\end{aligned} \tag{3.6}$$

In order to find a first-order Lie–Bäcklund transformation equivalent to the vector field V_3 (2.6) we searched for a vertical vector field of the form

$$V = x\cdot LB_1 + t\cdot LB_2 + C\,, \tag{3.7}$$

where LB_1 and LB_2 are linear combinations of the vector fields V_4, V_5, $X_1,\dots,X_4$; C being a correction. From (3.7) and condition (3.1) we obtained two additional first-order Lie–Bäcklund transformations X_5 and X_6,

$$\begin{aligned}X_5 &= x(X_1-X_2) + t(X_1+X_2) + C_5\,,\\ X_6 &= x(X_3-X_4) + t(X_3+X_4) + C_6\,,\end{aligned} \tag{3.8a}$$

where

$$\begin{aligned}C_5 &= \tfrac{1}{2}(-u_3\partial_{u_3} - v_3\partial_{v_3} + u_4\partial_{u_4} + v_4\partial_{v_4})\,,\\ C_6 &= \tfrac{1}{2}(-u_1\partial_{u_1} - v_1\partial_{v_1} + u_2\partial_{u_2} + v_2\partial_{v_2})\,.\end{aligned} \tag{3.8b}$$

Note that

$$X_5 + X_6 \doteq -V_3\,. \tag{3.9}$$

Second-order Lie–Bäclund transformations of the Federbush model have to satisfy the condition

$$\mathscr{L}_V I \subset D^2 I\,, \tag{3.10}$$

where D^2I is the twice prolonged ideal I, defined on

$$\mathbf{R}^{26} = \{(x,t,u_1,\dots,v_4,u_{1x},\dots,v_{4x},u_{1xx},\dots,v_{4xx})\}\,. \tag{3.11}$$

We search for vertical vector fields, the generating functions $V^{u_1},\dots,V^{u_4}$ being dependent on $u_1,\dots,v_4,\dots,u_{1xx},\dots,v_{4xx}$.

Due to the grading (3.6) the coefficients are of the following structure

$$\begin{aligned}&[u]_{xx} + ([u]^2 + [m])[u]_x\\ &\quad + ([u]^5 + [m][u]^3 + [m]^2[u])\,,\end{aligned} \tag{3.12a}$$

whereas in (3.12a)

$$\begin{aligned}&[u] \text{ refers to } u_1,\dots,v_4\,,\\ &[u]_x \text{ refers to } u_{1x},\dots,v_{4x}\,,\\ &[u]_{xx} \text{ refers to } u_{1xx},\dots,v_{4xx}\,,\\ &[m] \text{ refers to } m_1,m_2\,.\end{aligned} \tag{3.12b}$$

Solving the overdetermined system of partial differential equations arising from (3.10), using (3.12) we obtained four second-order (x,t)-independent Lie–Bäcklund transformations $X_7,\dots,X_{10}$

$$\begin{aligned}&X_7^{u_1} = (\lambda/2)v_1K_7\,,\quad X_7^{v_1} = -(\lambda/2)u_1K_7\,,\\ &X_7^{u_2} = (\lambda/2)v_2K_7\,,\quad X_7^{v_2} = -(\lambda/2)u_2K_7\\ &X_7^{u_3} = \tfrac{1}{4}m_2\{2u_{4x} + \lambda v_4(R_1+R_2)\}\,,\\ &X_7^{v_3} = \tfrac{1}{4}m_2\{2v_{4x} - \lambda u_4(R_1+R_2)\}\,,\\ &X_7^{u_4} = \tfrac{1}{4}\{-4v_{4xx} + 2\lambda u_4(R_1+R_2)_x + 4\lambda u_{4x}(R_1+R_2)\\ &\qquad + 2m_2u_{3x} + \lambda m_2v_3(R_1+R_2) + \lambda^2v_4(R_1+R_2)^2\}\,,\\ &X_7^{v_4} = \tfrac{1}{4}\{+4u_{4xx} + 2\lambda v_4(R_1+R_2)_x + 4\lambda v_{4x}(R_1+R_2)\\ &\qquad + 2m_2v_{3x} - \lambda m_2u_3(R_1+R_2) - \lambda^2u_4(R_1+R_2)^2\}\,,\\ &X_8^{u_1} = (\lambda/2)v_1K_8\,,\quad X_8^{v_1} = -(\lambda/2)u_1K_8\,,\\ &X_8^{u_2} = (\lambda/2)v_2K_8\,,\quad X_8^{v_2} = -(\lambda/2)u_2K_8\,,\end{aligned}$$

$$
\begin{aligned}
X_8^{u_3} &= \tfrac14\{-4v_{3xx}+2\lambda u_3(R_1+R_2)_x+4\lambda u_{3x}(R_1+R_2)\\
&\quad -2m_2u_{4x}-\lambda m_2v_4(R_1+R_2)+\lambda^2 v_3(R_1+R_2)^2\},\\
X_8^{v_3} &= \tfrac14\{+4u_{3xx}+2\lambda v_3(R_1+R_2)_x+4\lambda v_{3x}(R_1+R_2)\\
&\quad -2m_2v_{4x}+\lambda m_2u_4(R_1+R_2)\\
&\quad -\lambda^2 u_3(R_1+R_2)^2\},\\
X_8^{u_4} &= \tfrac14 m_2\{-2u_{3x}-\lambda v_3(R_1+R_2)\},\\
X_8^{v_4} &= \tfrac14 m_2\{-2v_{3x}+\lambda u_3(R_1+R_2)\},\\
X_9^{u_1} &= \tfrac14 m_1\{+2u_{2x}-\lambda v_2(R_3+R_4)\},\\
X_9^{v_1} &= \tfrac14 m_1\{+2v_{2x}+\lambda u_2(R_3+R_4)\},\\
X_9^{u_2} &= \tfrac14\{-4v_{2xx}-2\lambda u_2(R_3+R_4)_x-4\lambda u_{2x}(R_3+R_4)\\
&\quad +2m_1u_{1x}-\lambda m_1v_1(R_3+R_4)+\lambda^2 v_2(R_3+R_4)^2\},\\
X_9^{v_2} &= \tfrac14\{+4u_{2xx}-2\lambda v_2(R_3+R_4)_x-4\lambda v_{2x}(R_3+R_4)\\
&\quad +2m_1v_{1x}+\lambda m_1u_1(R_3+R_4)-\lambda^2 u_2(R_3+R_4)^2\},\\
X_9^{u_3} &= (\lambda/2)v_3K_9,\quad X_9^{v_3}=-(\lambda/2)u_3K_9,\\
X_9^{u_4} &= (\lambda/2)v_4K_9,\quad X_9^{v_4}=-(\lambda/2)u_4K_9,\\
X_{10}^{u_1} &= \tfrac14\{-4v_{1xx}-2\lambda u_1(R_3+R_4)_x-4\lambda u_{1x}(R_3+R_4)\\
&\quad -2m_1u_{2x}+\lambda m_1v_2(R_3+R_4)+\lambda^2 v_1(R_3+R_4)^2\},\\
X_{10}^{v_1} &= \tfrac14\{+4u_{1xx}-2\lambda v_1(R_3+R_4)_x-4\lambda v_{1x}(R_3+R_4)\\
&\quad -2m_1v_{2x}-\lambda m_1u_2(R_3+R_4)-\lambda^2 u_1(R_3+R_4)^2\},\\
X_{10}^{u_2} &= \tfrac14 m_1\{-2u_{1x}+\lambda v_1(R_3+R_4)\},\\
X_{10}^{v_2} &= \tfrac14 m_1\{-2v_{1x}-\lambda u_1(R_3+R_4)\},\\
X_{10}^{u_3} &= (\lambda/2)v_3K_{10},\quad X_{10}^{v_3}=-(\lambda/2)u_3K_{10},\\
X_{10}^{u_4} &= (\lambda/2)v_4K_{10},\quad X_{10}^{v_4}=-(\lambda/2)u_4K_{10},
\end{aligned}
\tag{3.13}
$$

whereas in (3.13) $K_7,\ldots,K_{10}$ are defined by

$$
\begin{aligned}
K_7 &= +2u_{4x}v_4-2u_4v_{4x}\\
&\quad +m_2(u_3u_4+v_3v_4)+\lambda R_4(R_1+R_2),\\
K_8 &= +2u_{3x}v_3-2u_3v_{3x}\\
&\quad -m_2(u_3u_4+v_3v_4)+\lambda R_3(R_1+R_2),\\
K_9 &= -2u_{2x}v_2+2u_2v_{2x}\\
&\quad -m_1(u_1u_2+v_1v_2)+\lambda R_2(R_3+R_4),\\
K_{10} &= -2u_{1x}v_1+2u_1v_{1x}\\
&\quad +m_1(u_1u_2+v_1v_2)+\lambda R_1(R_3+R_4).
\end{aligned}
\tag{3.14}
$$

The Lie bracket for vertical vector fields V_i, defined by

$$V_i=V_i^{u_1}\partial_{u_1}+V_i^{v_1}\partial_{v_1}+\cdots+V_i^{u_4}\partial_{u_4}+V_i^{v_4}\partial_{v_4}, \tag{3.15}$$

is given by

$$[V_i,V_j]^k=V_i(V_j^k)-V_j(V_i^k)\quad (k=u_1,\ldots,v_4). \tag{3.16}$$

In (3.4), (3.8), and (3.14), only the $\partial_{u_1},\ldots,\partial_{v_4}$ components of the vertical vector fields are given; the other components are obtained by prolongation of the vector fields.[7]

Computation of the Lie bracket (3.16) of the vertical vector fields $V_4,V_5,X_1,\ldots,X_4,X_5,X_6,X_7,\ldots,X_{10}$, (2.6), (3.4), (3.8), and (3.12) leads to the following:

V_4 and V_5 are commuting with any other vector field,

$$[X_i,X_j]=0\quad (i,j=1,\ldots,4,7,\ldots,10),$$

while for X_5 and X_6 we derive

$$
\begin{aligned}
&[X_5,X_1]=+X_1,\quad [X_5,X_2]=-X_2,\\
&[X_5,X_3]=0,\quad [X_5,X_4]=0,\\
&[X_6,X_1]=0,\quad [X_6,X_2]=0,\\
&[X_6,X_3]=X_3,\quad [X_6,X_4]=-X_4,\\
&[X_5,X_7]=+2X_7-\tfrac12 m_2^2V_5,\\
&[X_5,X_8]=-2X_8+\tfrac12 m_2^2V_5,\\
&[X_5,X_9]=[X_5,X_{10}]=0,\quad [X_6,X_7]=[X_6X_8]=0,\\
&[X_6,X_9]=2X_9-\tfrac12 m_1^2V_4,\\
&[X_6,X_{10}]=-2X_{10}+\tfrac12 m_1^2V_4,
\end{aligned}
\tag{3.17}
$$

We now transform the vector fields by

$$
\begin{aligned}
&Y_0^+=V_4,\quad Y_0^-=V_5,\quad Z_0^+=X_6,\\
&Y_1^+=X_3,\quad Y_1^-=X_1,\quad Z_0^-=X_5,\\
&Y_{-1}^+=X_4,\quad Y_{-1}^-=X_2,\\
&Y_2^+=X_9-\tfrac14 m_1^2\,V_4,\quad Y_2^-=X_7-\tfrac14 m_2^2\,V_5,\\
&Y_{-2}^+=X_{10}-\tfrac14 m_1^2\,V_4,\quad Y_{-2}^-=X_8-\tfrac14 m_2^2\,V_5.
\end{aligned}
\tag{3.18}
$$

From (3.17) and (3.18) we obtain a direct sum of two Lie algebras, i.e., each " + " element commutes with each " − " element, and

$$[Z_0,Y_i]=i\,Y_i,\quad [Y_i,Y_k]=0\quad (i,k=-2,\ldots,2). \tag{3.19}$$

In (3.19), Z_0, Y_i $(i=-2,\ldots,2)$ are assumed to have the same upper sign.

IV. CREATING AND ANNIHILATING LIE–BÄCKLUND TRANSFORMATIONS OF THE FEDERBUSH MODEL

In this section we shall construct four (x,t)-dependent Lie–Bäcklund transformations that act as creating and annihilating operators on the (x,t)-independent vector fields $X_1,\ldots,X_4,X_7,\ldots,X_{10}$.

Motivated by the results obtained for the massive Thirring model[2] and the results obtained in Sec. III, probably leading to the direct sum of two Lie algebras, each of which has a similar structure to the Lie algebra for the massive Thirring model, we were forced to search for *nonlocal* Lie–Bäcklund transformations[2,12] including potential forms associated to the infinitesimal symmetries V_1, V_2 (2.6) (2.9).

Surprisingly, in carrying through the huge computation the nonlocal variables dropped off from our intermediate results, leading finally to *local*, (x,t)-dependent Lie–Bäcklund transformations.

So, for simplicity, we shall discuss the search for the creating and annihilating Lie–Bäcklund transformations assuming from the beginning that they are of *local* nature.

The Lie–Bäcklund transformations $X_1,\ldots,X_4$ are of degree 2, while the Lie–Bäcklund transformations $X_7,\ldots,X_{10}$ are of degree 4. We now search for an (x,t)-dependent Lie–

Bäcklund transformation of second order, linear in x and t, and of degree 2, i.e., a vector field V,

$$V = x\cdot LB_1 + t\cdot LB_2 + C^*, \tag{4.1}$$

where LB_1 and LB_2 are Lie–Bäcklund transformations of degree 4, and due to the fact that m_1 and m_2 are of degree 2, LB_1 and LB_2 are assumed to be linear combinations of $V_4, V_5, X_1,\dots, X_4, X_7,\dots, X_{10}$, while V has to satisfy

$$\mathscr{L}_V I \subset D^2 I. \tag{4.2}$$

From (4.2) and (4.1) we obtained, after a rather massive computation, four (x,t)-dependent Lie–Bäcklund transformations

$$\begin{aligned}
X_{11} &= x(-X_{10} + \tfrac{1}{2} m_1^2 V_4) + t(X_{10}) + C_{11},\\
X_{12} &= x(X_9 - \tfrac{1}{2} m_1^2 V_4) + t(X_9) + C_{12},\\
X_{13} &= x(-X_8 + \tfrac{1}{2} m_2^2 V_5) + t(X_8) + C_{13},\\
X_{14} &= x(X_7 - \tfrac{1}{2} m_2^2 V_5) + t(X_7) + C_{14},
\end{aligned} \tag{4.3}$$

where in (4.3) $C_{11},\dots,C_{14}$ are given by

$$\begin{aligned}
C_{11} &= \tfrac{1}{2}\{+2v_{1x} + m_1u_2 + \lambda u_1(R_3+R_4)\}\partial_{u_1}\\
&\quad + \tfrac{1}{2}\{-2u_{1x} + m_1v_2 + \lambda v_1(R_3+R_4)\}\partial_{v_1},\\
C_{12} &= \tfrac{1}{2}\{-2v_{2x} + m_1u_1 - \lambda u_2(R_3+R_4)\}\partial_{u_2}\\
&\quad + \tfrac{1}{2}\{+2u_{2x} + m_1v_1 - \lambda v_2(R_3+R_4)\}\partial_{v_2},\\
C_{13} &= \tfrac{1}{2}\{+2v_{3x} + m_2u_4 - \lambda u_3(R_1+R_2)\}\partial_{u_3}\\
&\quad + \tfrac{1}{2}\{-2u_{3x} + m_2v_4 - \lambda v_3(R_1+R_2)\}\partial_{v_3},\\
C_{14} &= \tfrac{1}{2}\{-2v_{4x} + m_2u_3 + \lambda u_4(R_1+R_2)\}\partial_{u_4}\\
&\quad + \tfrac{1}{2}\{+2u_{4x} + m_2v_3 - \lambda v_4(R_1+R_2)\}\partial_{v_4}.
\end{aligned} \tag{4.4}$$

From (4.3) and (3.18) we define

$$\begin{aligned}
Z^+_{-1} &= X_{11} = x(-Y^+_{-2} + \tfrac{1}{4} m_1^2 Y_0^+)\\
&\quad + t(+Y^+_{-2} + \tfrac{1}{4} m_1^2 Y_0^+) + C_{11},\\
Z_1^+ &= X_{12} = x(+Y_2^+ + \tfrac{1}{4} m_1^2 Y_0^+)\\
&\quad + t(+Y_2^+ + \tfrac{1}{4} m_1^2 Y_0^+) + C_{12},\\
Z^-_{-1} &= X_{13} = x(-Y^-_{-2} + \tfrac{1}{4} m_2^2 Y_0^-)\\
&\quad + t(+Y^-_{-2} + \tfrac{1}{4} m_2^2 Y_0^-) + C_{13},\\
Z_1^- &= X_{14} = x(+Y_2^- - \tfrac{1}{4} m_2^2 Y_0^-)\\
&\quad + t(+Y_2^- + \tfrac{1}{4} m_2^2 Y_0^-) + C_{14}.
\end{aligned} \tag{4.5}$$

Computation of the commutators of $Z^+_{-1}, Z_1^+, Z^-_{-1}$, and Z_1^- (4.5) and $Y_i^{\pm}$ ($i=-2,\dots,2$) leads to the following result:

$$\begin{aligned}
&[Z^+_{-1}, Y_2^+] = -\tfrac{1}{2} m_1^2 Y_1^+, \quad [Z_1^+, Y_2^+] = +Y_3^+,\\
&[Z^+_{-1}, Y_1^+] = +\tfrac{1}{4} m_1^2 Y_0^+, \quad [Z_1^+, Y_1^+] = +Y_2^+,\\
&[Z^+_{-1}, Y_0^+] = 0, \quad [Z_1^+, Y_0^+] = 0,\\
&[Z^+_{-1}, Y^+_{-1}] = -Y^+_{-2},\\
&[Z_1^+, Y^+_{-1}] = -\tfrac{1}{4} m_1^2 Y_0^+,\\
&[Z^+_{-1}, Y^+_{-2}] = +Y^+_{-3},\\
&[Z_1^+, Y^+_{-2}] = +\tfrac{1}{2} m_1^2 Y^+_{-1},
\end{aligned} \tag{4.6a}$$

$$\begin{aligned}
&[Z^-_{-1}, Y_2^-] = -\tfrac{1}{2} m_2^2 Y_1^-, \quad [Z_1^-, Y_2^-] = +Y_3^-,\\
&[Z^-_{-1}, Y_1^-] = +\tfrac{1}{4} m_2^2 Y_0^-, \quad [Z_1^-, Y_1^-] = +Y_2^-,\\
&[Z^-_{-1}, Y_0^-] = 0, \quad [Z_1^-, Y_0^-] = 0,\\
&[Z^-_{-1}, Y^-_{-1}] = -Y^-_{-2},\\
&[Z_1^-, Y^-_{-1}] = -\tfrac{1}{4} m_2^2 Y_0^-,\\
&[Z^-_{-1}, Y^-_{-2}] = +Y^-_{-3},\\
&[Z_1^-, Y^-_{-2}] = +\tfrac{1}{2} m_2^2 Y^-_{-1},
\end{aligned} \tag{4.6b}$$

while

$$\begin{aligned}
&[Z^+_{-1}, Z_1^+] = -\tfrac{1}{2} m_1^2 Z_0^+,\\
&[Z^-_{-1}, Z_1^-] = -\tfrac{1}{2} m_1^2 Z_0^-.
\end{aligned} \tag{4.6c}$$

All other commutators are zero. For completeness sake the vector field Y_3^+ is given in the Appendix.

V. CONCLUSION

By an extensive use of symbolic computation we succeeded in constructing four creating and annihilating Lie–Bäcklund transformations of the Federbush model, leading to infinite hierarchies of commuting Lie–Bäcklund transformations. We hope to give a formal proof of this in future work.

ACKNOWLEDGMENT

The author wishes to express his gratitude to Professor R. Martini for his stimulating interest in this research.

APPENDIX: THE LIE–BÄCKLUND TRANSFORMATION Y_3^+

The components of the vector field Y_3^+ are given by

$$\begin{aligned}
Y_3^{+,u_1} &= (m_1/4)\{-4v_{2xx} - 4\lambda R_{34}u_{2x} + 2m_1u_{1x}\\
&\quad - 4\lambda u_2(R_{34})_{(1)} + m_1^2 v_2\\
&\quad - \lambda m_1 R_{34} v_1 + \lambda^2 R_{34}^2 v_2\},\\
Y_3^{+,v_1} &= (m_1/4)\{+4u_{2xx} - 4\lambda R_{34}v_{2x} + 2m_1v_{1x}\\
&\quad - 4\lambda v_2(R_{34})_{(1)} - m_1^2 u_2\\
&\quad + \lambda m_1 R_{34} u_1 - \lambda^2 R_{34}^2 u_2\},\\
Y_3^{+,u_2} &= \tfrac{1}{4}\{-8u_{2xxx} - 4m_1v_{1xx} + 12\lambda R_{34}v_{2xx}\\
&\quad + 8\lambda v_2(R_{34})_{(2)} + 24\lambda v_{2x}(R_{34})_{(1)}\\
&\quad + 8\lambda v_2(R_{34})_{(1,1)}\\
&\quad + u_{2x}(4m_1^2 + 6\lambda^2 R_{34}^2) - 4\lambda m_1 R_{34}u_{1x}\\
&\quad + 12\lambda^2 u_2 R_{34}(R_{34})_{(1)}\\
&\quad - 4\lambda m_1 u_1(R_{34})_{(1)} + m_1^3 v_1 - 2\lambda m_1^2 R_{34} v_2\\
&\quad + \lambda^2 m_1 R_{34}^2 v_1 - \lambda^3 R_{34}^3 v_2\},\\
Y_3^{+,v_2} &= \tfrac{1}{4}\{-8v_{2xxx} + 4m_1u_{1xx} + 12\lambda R_{34}u_{2xx}\\
&\quad - 8\lambda u_2(R_{34})_{(2)} - 24\lambda u_{2x}(R_{34})_{(1)}\\
&\quad - 8\lambda u_2(R_{34})_{(1,1)}\\
&\quad + v_{2x}(4m_1^2 + 6\lambda^2 R_{34}^2) - 4\lambda m_1 R_{34}v_{1x}
\end{aligned} \tag{A1}$$

$$+ 12\lambda^2 v_2 R_{34}(R_{34})_{(1)}$$
$$- 4\lambda m_1 v_1 (R_{34})_{(1)} - m_1^3 u_1 + 2\lambda m_1^2 R_{34} u_2$$
$$- \lambda^2 m_1 R_{34}^2 u_1 + \lambda^3 R_{34}^3 u_2\} ,$$
$$Y_3^{+,u_3} = (\lambda/4) v_3 L , \quad Y_3^{+,v_3} = -(\lambda/4) u_3 L ,$$
$$Y_3^{+,u_4} = (\lambda/4) v_4 L , \quad Y_3^{+,v_4} = -(\lambda/4) u_4 L;$$

where in (A1)

$$\begin{aligned} R_{34} &= R_3 + R_4 , \\ (R_{34})_{(1)} &= u_3 u_{3x} + v_3 v_{3x} + u_4 u_{4x} + v_4 v_{4x} , \\ (R_{34})_{(2)} &= u_3 u_{3xx} + v_3 v_{3xx} + u_4 u_{4xx} + v_4 v_{4xx} , \\ (R_{34})_{(1,1)} &= u_{3x}^2 + v_{3x}^2 + u_{4x}^2 + v_{4x}^2 , \end{aligned} \qquad \text{(A2)}$$

and

$$\begin{aligned} L &= 8(u_2 u_{2xx} + v_2 v_{2xx}) - 4(u_{2x}^2 + v_{2x}^2) \\ &\quad + 12\lambda R_{34}(u_{2x} v_2 - v_{2x} u_2) \\ &\quad + 4m_1(u_1 v_{2x} - v_1 u_{2x} + u_2 v_{1x} - v_2 u_{1x}) \\ &\quad - m_1^2(2R_2 + R_1) + 4\lambda m_1 R_{34}(u_1 u_2 + v_1 v_2) \\ &\quad - 3\lambda^2 R_2 R_{34}^2 . \end{aligned} \qquad \text{(A3)}$$

[1]P. H. M. Kersten and R. Martini, J. Math. Phys. **26**, 822 (1985).
[2]P. H. M. Kersten and R. Martini, J. Math. Phys. **26**, 1775 (1985).
[3]S. N. M. Ruijsenaars, Commun. Math. Phys. **87**, 181 (1982).
[4]P. Federbush, Phys. Rev. **121**, 1247 (1961).
[5]B. K. Harrison and F. B. Estabrook, J. Math. Phys. **12**, 653 (1971).
[6]R. L. Anderson and N. H. Ibragimov, *Lie–Bäcklund Transformations in Applications* (SIAM, Philadelphia, 1979).
[7]F. A. E. Pirani, D. C. Robinson, and W. F. Shadwick "Local jet bundle formulation of Bäcklund transformations," in *Mathematical Physics Studies 1* (Reidel, Boston, 1979).
[8]A. C. Hearn, REDUCE *User's Manual (Version 3.0)* (The Rand Corporation, Santa Monica, 1983).
[9]P. K. H. Gragert, "Symbolic computations in prolongation theory," Ph.D. thesis, Twente University of Technology, Enschede, The Netherlands, 1981.
[10]P. K. H. Gragert, P. H. M. Kersten, and R. Martini, Acta Appl. Math. **1**, 43 (1983).
[11]P. H. M. Kersten, "Infinitesimal symmetries: A computational approach," Ph.D. thesis, Twente University of Technology, Enschede, The Netherlands, 1985.
[12]I. S. Krasilshchik and A. M. Vinogradov, Acta Appl. Math. **2**, 79 (1984).

An approximation formula in the inverse scattering problem

Yoshimi Saitō
Department of Mathematics, University of Alabama at Birmingham, Birmingham, Alabama 35294

(Received 22 August 1985; accepted for publication 5 December 1985)

An approximate formula is given by which the potential $Q(x)$ can be recovered with any accuracy if only the scattering data of the Schrödinger operator $-\Delta + Q(x)$ is known for some sufficiently high energy value k. Thus the scattering data for all k around $k = \infty$ is not needed in order to get a good approximate value for the potential. The main tools for the proof are an asymptotic formula for the S-matrix [Y. Saitō, J. Math. Phys. **25**, 3105 (1984)] and the spectral decomposition theorem for the Schrödinger operator $-\Delta + Q(x)$ based on the limiting absorption principle.

I. INTRODUCTION AND THE MAIN THEOREMS

One of the important questions in the inverse problem for the Schrödinger equation is to establish a method by which the potential may be recovered from the spectral measure or the scattering data. This problem is often called the reconstruction problem.

In the classical works of Gel'fand–Levitan[1] and Agranovich–Marchenko[2] for the Schrödinger equation in $\mathbf{R}^1$, roughly speaking, the potential was recovered by using the Fourier transform of the spectral measure or the scattering data. Newton[3] extended these results to the Schrödinger equation in $\mathbf{R}^3$. In all these methods we need the spectral measure on the whole real line or the scattering data for all energy numbers $k>0$ to recover the potential.

There are some other methods for recovering the potential that generally may be called high energy limit methods. In order to explain the method let us introduce the Schrödinger operator in $\mathbf{R}^3$ and the scattering matrix $S(k)$, $k>0$, associated with it. Let us consider the differential expression

$$h = -\Delta + Q(x) \quad (x\in\mathbf{R}^3), \tag{1.1}$$

where Δ is the Laplacian in $\mathbf{R}^3$ and the potential $Q(x)$ is a real-valued function that satisfies

$$|Q(x)| \leqslant C_0(1+|x|)^{-\beta} \quad (x\in\mathbf{R}^3), \tag{1.2}$$

with constants $C_0>0$ and $\beta>1$, i.e., $Q(x)$ is a short-range potential. Let H be a unique self-adjoint extension of h restricted to $C_0^\infty(\mathbf{R}^3)$. The scattering matrix $S(k)$, $k>0$, associated with H is defined as follows: The scattering operator S is defined by

$$W_+^* W_- = S \quad \Big(W_\pm = s-\lim_{t\to\pm\infty} e^{itH}e^{-itH_0}\Big). \tag{1.3}$$

Here H_0 is the (unique) self-adjoint extension of $-\Delta$ restricted to $C_0^\infty(\mathbf{R}^3)$ and W_+^* is the adjoint operator of W_+. Then it is well known that there exists a family $\{S(k)\}_{k>0}$ of unitary operators on $L_2(S^2)$ such that

$$(\mathscr{F}S\mathscr{F}^*G)(\xi) = \{S(|\xi|)G(|\xi|\,\cdot\,)\}(\tilde\xi)$$
$$[G\in C_0^\infty(\mathbf{R}_\xi^3),\ \ \xi\in\mathbf{R}^3,\ \ \tilde\xi=\xi/|\xi|], \tag{1.4}$$

where $\mathscr{F}$ and $\mathscr{F}^*$ are the usual Fourier transform and the inverse Fourier transform, respectively, i.e.,

$$(\mathscr{F}u)(\xi) = (2\pi)^{-3/2}\int_{\mathbf{R}^3} e^{-ix\xi}u(x)dx,$$
$$(\mathscr{F}^*v)(x) = (2\pi)^{-3/2}\int_{\mathbf{R}^3} e^{ix\xi}v(\xi)d\xi, \tag{1.5}$$

with the inner product $x\xi$ in $\mathbf{R}^3$. Set

$$F(k) = -2\pi i k^{-1}(S(k)-I), \tag{1.6}$$

with the identity operator I on $L_2(S^2)$. It is known that $F(k)$ is a Hilbert–Schmidt operator with its integral kernel $F(k,\omega,\omega')$, $k>0$, $\omega,\omega\in S^2$, if $Q(x)$ satisfies (1.2) with $\beta>2$.

Fadeev[4] showed the formula

$$\lim_{\substack{k\to\infty\\ \xi=k(\omega-\omega')}} F(k,\omega,\omega') = -\frac{1}{4\pi}\int_{\mathbf{R}^3} e^{-i\xi x}Q(x)dx$$
$$(\xi\in\mathbf{R}^3), \tag{1.7}$$

under the assumption that $\beta>3$ in (1.2). Here the limit is taken so that k goes to ∞ keeping the relation $\xi=k(\omega-\omega')$ for a given $\xi\in\mathbf{R}^3$. The potential $Q(x)$ can be recovered by the usual Fourier inversion formula. As for the extensions of Fadeev's result, see Newton,[3] Lemma 3.1 and Saitō,[5] Theorem 5.1. Another asymptotic formula was given by Saitō.[5] Set

$$f(x,k) = k^2(F(k)\psi_{x,k},\psi_{x,k})_{S^2}, \tag{1.8}$$

where $(\ ,\ )_{S_2}$ is the inner product of $L_2(S^2)$ and

$$\psi_{x,k} = \psi_{x,k}(\omega) = e^{-ikx\omega}\in L_2(S^2) \quad (\omega\in S^2), \tag{1.9}$$

with parameters $x\in\mathbf{R}^3$ and $k>0$. Here $x\omega$ is the inner product in $\mathbf{R}^3$. Then, under the condition that $\beta>1$ in (1.2), the limit

$$f(x,\infty) = \lim_{k\to\infty} f(x,k) = -2\pi\int_{\mathbf{R}^3}\frac{Q(y)}{|y-x|^2}dy \tag{1.10}$$

exists. The potential $Q(x)$ was recovered in Saitō[6] by solving (1.10) as an integral equation for $Q(x)$. In fact we have the inversion formula

$$Q(x) = -(4\pi^3)^{-1}\mathscr{F}^*(|\xi|\mathscr{F}f(\,\cdot\,,\infty))(x), \tag{1.11}$$

where the left-hand side of (1.11) is, in general, well-defined as an element of $\mathscr{S}'(\mathbf{R}^3)$, the dual space of all rapidly decreasing functions on $\mathbf{R}^3$. These results were extended to the

 0022-2488/86/041145-09$02.50

Schrödinger operator in $\mathbf{R}^N$ $(N \geqslant 2)$ with a short-range potential (Saitō[7]). In the works of Faddeev and Saitō we do not need any low energy scattering data but we do need the scattering data around $k = \infty$ in order to recover the potential.

In this paper we shall give an approximate formula by which the potential can be recovered with any accuracy if we know only the scattering data for some sufficiently high energy value. Thus we do not need the scattering data for k around $k = \infty$ in order to get a good approximate value for the potential. Here $Q(x)$ is assumed to satisfy the following.

Assumption 1.1: $Q(x)$ is a real-valued, C^2 function on $\mathbf{R}^3$. There exist constants $\beta > \frac{7}{4}, \rho > \frac{5}{2}$, and $C_0 > 0$ such that

$$|Q(x)| \leqslant C_0(1 + |x|)^{-\beta} \quad (x \in \mathbf{R}^3), \tag{1.12}$$

and

$$|(D^\alpha Q)(x)| \leqslant C_0(1 + |x|)^{-\rho} \quad (x \in \mathbf{R}^3, \quad |\alpha| = 1,2). \tag{1.13}$$

Here $\alpha = (\alpha_1,\alpha_2,\alpha_3)$ with nonnegative integers $\alpha_1,\alpha_2,\alpha_3$ is a multi-index, $|\alpha| = \alpha_1 + \alpha_2 + \alpha_3$ and

$$D^\alpha = D_1^{\alpha_1} D_2^{\alpha_2} D_3^{\alpha_3} \quad \left(D_j = \frac{\partial}{\partial x_j}, \quad x = (x_1,x_2,x_3)\right). \tag{1.14}$$

Here and in the sequel β and ρ are assumed (with no loss of generality) to satisfy

$$\tfrac{7}{4} < \beta \leqslant 2 \quad \text{and} \quad \tfrac{5}{2} < \rho < 3. \tag{1.15}$$

Let $f(x,k)$ and $f(x,\infty)$ be as in (1.8) and (1.10), respectively. Set

$$g(x,k) = f(x,k) - f(x,\infty). \tag{1.16}$$

The following two theorems are our main results.

Theorem 1.2: Let Assumption 1.1 be satisfied. Then for any $k > 0$, $f(x,k)$, $f(x,\infty)$, and $g(x,k)$ are C^1 functions of $x \in \mathbf{R}^3$. Let a be a positive number. Then there exist constants $b > \frac{3}{2}$ and $C = C(a,Q)$, which depend only on a and the potential $Q(x)$ such that

$$|D^\alpha g(x,k) \leqslant (C/k)(1 + |x|)^{-b}$$
$$(x \in \mathbf{R}^3, \quad k \geqslant a, \quad |\alpha| = 0,1), \tag{1.17}$$

$$|f(x,k)| \leqslant C(1 + |x|)^{-\min(\beta-1,b)},$$
$$|D_j f(x,k)| \leqslant C(1 + |x|)^{-b}$$
$$(x \in \mathbf{R}^3, \quad k \geqslant a, \quad j = 1,2,3). \tag{1.18}$$

It follows from the above theorem that $D^\alpha g(x,k)$ $(|\alpha| = 0,1)$ and $D_j f(x,k) = (\partial/\partial x_j) f(x,k)$ $(j = 1,2,3)$ belong to $L_2(\mathbf{R}^3)$. Let $h(x)$ be a function on $\mathbf{R}^3$. Then let us define $\mathscr{F}^*|\xi|\mathscr{F}h$ as a linear functional on $\mathscr{S}(\mathbf{R}^3)$ by

$$\langle \mathscr{F}^*|\xi|\mathscr{F}h, \phi\rangle = \int_{\mathbf{R}^3} h(x)\mathscr{F}^*(|\xi|\mathscr{F}\phi)(x)dx, \tag{1.19}$$

where $\mathscr{F}$ and $\mathscr{F}^*$ are the Fourier transform and the Fourier inverse transform, respectively, and $\mathscr{S}(\mathbf{R}^3)$ is all rapidly decreasing functions on $\mathbf{R}^3$. Let A_j $(j = 1,2,3)$ be defined by

$$A_j = \{\xi = (\xi_1,\xi_2,\xi_3) \in \mathbf{R}^3 / |\xi_j| > |\xi_k| \quad (k \neq j)\} \tag{1.20}$$

and let $\chi_j(\xi)$ be the characteristic function of A_j, i.e.,

$$\chi_j(\xi) = \begin{cases} 1 & (\xi \in A_j), \\ 0 & (\xi \notin A_j). \end{cases} \tag{1.21}$$

Theorem 1.3: Let Assumption 1.1 be satisfied.

(i) Then $\mathscr{F}^*|\xi|\mathscr{F}f(\cdot,k)$ is well defined for any $k > 0$ as an element of $\mathscr{S}'(\mathbf{R}^3)$. They belong to $L_2(\mathbf{R}^3)$, too. Let a be a positive number. Then there exists a constant $\widetilde{C} = \widetilde{C}(a,Q)$ depending only on a and $Q(x)$ such that

$$\|Q + (4\pi^3)^{-1}\mathscr{F}^*|\xi|\mathscr{F}f(\cdot,k)\|_{L_2} \leqslant \widetilde{C}/k \quad (k \geqslant a), \tag{1.22}$$

where $\| \ \|_{L_2}$ is the norm of $L_2(\mathbf{R}^3)$.

(ii) $\mathscr{F}^*|\xi|\mathscr{F}f(\cdot,k)$ can be expressed as

$$\mathscr{F}^*|\xi|\mathscr{F}f(\cdot,k) = -i\sum_{j=1}^{3}\mathscr{F}^*\left(\frac{|\xi|}{\xi_j}\chi_j(\xi)\mathscr{F}(D_j f(\cdot,k))\right), \tag{1.23}$$

where $\mathscr{F}(D_j f(\cdot,k))$ is the Fourier transform of $D_j f(x,k)$ in $L_2(\mathbf{R}^3)$ and $\mathscr{F}^*(|\xi|\xi_j^{-1}\chi_j(\xi)\mathscr{F}(D_j f(\cdot,k)))$ is the inverse Fourier transform of $|\xi|\xi_j^{-1}\chi_j(\xi)\mathscr{F}(D_j f(\cdot,k))$ in $L_2(\mathbf{R}^3_\xi)$.

Remark 1.4: As is shown in (ii) of Theorem 1.3, $\mathscr{F}^*|\xi|\mathscr{F}f(\cdot,k)$ is "computable" by (1.23) using $D_j f(\cdot,k)$, $j = 1,2,3$, which are in $L_2(\mathbf{R}^3)$.

Remark 1.5: Through the proof of Theorems 1.2 and 1.3 we can show that the constant $\widetilde{C}$ depends only on the constant C_0 in (1.12) and (1.13) and $\widetilde{C}$ is bounded when C_0 moves in a bounded set in $[0,\infty)$ if $a > 0$ is taken to be reasonably large. Therefore, when we know the range of the "size" C_0 of $Q(x)$ in advance, we can approximate the potential $Q(x)$ with any accuracy by choosing a sufficiently high energy number k in (1.15).

Theorems 1.2 and 1.3 can be extended to the general N-dimensional case $(N \geqslant 2)$. This will be discussed elsewhere.

In Sec. II we shall summarize some results on the operator $F(k)$ and the limiting absorption principle for H and H_0, which will be used to prove Theorems 1.2 and 1.3. Sections III and IV are devoted to showing the asymptotic behavior of $f(x,k)$ and $D_j f(x,k)$ $(j = 1,2,3)$ as $k \to \infty$. The proof of Theorems 1.2 and 1.3 will be given in Sec. V. Some technical lemmas are proved in the Appendices.

II. PRELIMINARIES

In this section we shall summarize some known results related to the operator $F(k)$ and the function $f(x,k)$ defined by (1.6) and (1.8), respectively.

(1) Let δ be a constant such that

$$\tfrac{1}{2} < \delta < \min(\beta - \tfrac{5}{4}, \beta/2), \tag{2.1}$$

where β is given in Assumption 1.1. For any $\tau \in \mathbf{R}$ the weighted Hilbert space $L_{2,\tau}(\mathbf{R}^3)$ is defined by

$$L_{2,\tau}(\mathbf{R}^3) = \{f(x); \textstyle\int_{\mathbf{R}^3}(1 + |x|)^{2\tau}|v(x)|^2\,dx < \infty\} \tag{2.2}$$

with its inner product $(\ ,\ )_\tau$ and norm $\| \ \|_\tau$,

$$(v_1,v_2)_\tau = \int_{\mathbf{R}^3}(1 + |x|)^{2\tau}v_1(x)\,\overline{v_2(x)}\,dx, \tag{2.3}$$

$$\|v\|_\tau = [(v,v)_\tau]^{1/2}. \tag{2.4}$$

Let H and H_0 be the perturbed and unperturbed Schrödinger operators defined in Sec. I, respectively. Then it follows from the limiting absorption principle (Saitō,[8] Ikebe–Saitō,[9] Lavine,[10] and Agmon[11]) that we have the limits

$$R_{\pm}(k) = \lim_{\epsilon \downarrow 0}(H - (k^2 \pm i\epsilon))^{-1}$$
$$R_{0\pm}(k) = \lim_{\epsilon \downarrow 0}(H_0 - (k^2 \pm i\epsilon))^{-1} \qquad (k>0), \qquad (2.5)$$

in $\mathbf{B}(L_{2,\delta}(\mathbf{R}^3), L_{2,-\delta}(\mathbf{R}^3))$, where $\mathbf{B}(X,Y)$ means the Banach space of all linear bounded operators from X to Y. As for the operator norms $\|R_{\pm}(k)\|$, $\|R_{0\pm}(k)\|$ in $\mathbf{B}(L_{2,\delta}(\mathbf{R}^3), L_{2,-\delta}(\mathbf{R}^3))$, we have the following estimates (Saitō[12,13]): for any $a>0$ there exists a constant $C_1 = C_1(a,Q)$ depending only on a and the potential $Q(x)$ such that

$$\|R_{\pm}(k)\| \leqslant C_1/|k|, \quad \|R_{0\pm}(k)\| \leqslant C_1/|k| \qquad (2.6)$$

hold for any $k \geqslant a$. Starting with the second resolvent equation

$$(H-z)^{-1} = (H_0 - z)^{-1} - (H-z)^{-1}Q(H_0 - z)^{-1},$$
$$(H-z)^{-1} = (H_0 - z)^{-1} - (H_0 - z)^{-1}Q(H-z)^{-1}$$
$$(I_m z \neq 0), \qquad (2.7)$$

we get the Lippman–Schwinger equations

$$R_{\pm}(k) = R_{0\pm}(k) - R_{\pm}(k)QR_{0\pm}(k),$$
$$R_{\pm}(k) = R_{0\pm}(k) - R_{0\pm}(k)QR_{\pm}(k), \qquad (2.8)$$

for $k>0$. As is well known, $R_{0\pm}(k)$ has the expression

$$\{R_{0\pm}(k)f\}(x) = \frac{1}{4\pi}\int_{\mathbf{R}^3} \frac{e^{\pm ik|x-y|}}{|x-y|} f(y)dy$$
$$(k>0, \quad f \in L_{2,\delta}(\mathbf{R}^3)). \qquad (2.9)$$

(2) Set

$$u(x, \pm k, f) = \{R_{\pm}(R)f\}(x),$$
$$u_0(x, \pm k, f) = \{R_{0\pm}(k)f\}(x), \qquad (2.10)$$

where $f \in L_{2,\delta}(\mathbf{R}^3)$ and $k>0$. Here, $u(x, \pm k, f)$ $[u_0(x, \pm k, f)]$ is a unique solution of the Schrödinger equation

$$(-\Delta + Q(x) - k^2)v = f[(-\Delta - k^2)v = f], \qquad (2.11)$$

with the radiation condition

$$\frac{\partial v}{\partial x_j} \mp ik\tilde{x}_j v \in L_{2,\delta-1}(\mathbf{R}^3) \quad (j = 1,2,3), \qquad (2.12)$$
$$v \in L_{2,-\delta}(\mathbf{R}^3), \qquad (2.13)$$

where $x = (x_1, x_2, x_3)$ and $\tilde{x}_j = x_j/|x|$. It follows from (2.6) that we have the estimates

$$\|u(\cdot, \pm k, f)\|_{-\delta} \leqslant (C_1/k)\|f\|_\delta$$
$$\|u_0(\cdot, \pm k, f)\|_{-\delta} \leqslant (C_1/k)\|f\|_\delta \qquad (k \geqslant a), \qquad (2.14)$$

where a and C_1 are as in (2.6). Further we have

$$\left\|\frac{\partial u(\cdot, \pm k, f)}{\partial x_j}\right\|_{-\delta} \leqslant C_2\|f\|_\delta$$
$$\left\|\frac{\partial u_0(\cdot, \pm k, f)}{\partial x_j}\right\|_{-\delta} \leqslant C_2\|f\|_\delta \qquad (j = 1,2,3, \quad k \geqslant a), \qquad (2.15)$$

with a positive constant $C_2 = C_2(a,Q)$ which depends only on a and $Q(x)$. Equations (2.15) follows from Eqs. (2.14) and the estimates for the radiation condition

$$\sup_{k \geqslant a}\left\|\frac{\partial_u(\cdot, \pm k, f)}{\partial x_j} - ik\tilde{x}_j u(\cdot, \pm k, f)\right\|_{\delta-1} \leqslant C_3\|f\|_\delta,$$
$$\sup_{k \geqslant a}\left\|\frac{\partial u_0(\cdot, \pm k, f)}{\partial x_j} - ik\tilde{x}_j u_0(\cdot, \pm k, f)\right\|_{\delta-1} \leqslant C_3\|f\|_\delta$$
$$(k \geqslant a, \quad f \in L_{2,\delta}(\mathbf{R}^3), \quad j = 1,2,3), \qquad (2.16)$$

with a constant $C_3 = C_3(a,Q)$ (Saitō[14,15]). Here we should note that $\delta - 1 > -\delta$ since $\delta > \frac{1}{2}$. From (2.8), we easily obtain

$$u(x, \pm k, f) = u_0(x, \pm k, f) - u(x, \pm k, Qu_0(\cdot, k.f)),$$
$$u(x, \pm k, f) = u_0(x, \pm k, f) - u_0(x, \pm k, Qu(\cdot, \pm k, f)), \qquad (2.17)$$

for $k>0$ and $f \in L_{2,\delta}(\mathbf{R}^3)$.

(3) For $r \geqslant 0$ and $k>0$ let us define the operator $\eta_0^*(r,k)$ and $\eta_+^*(r,k)$ on $L^2(S^2)$ by

$$\{\eta_0^*(r,k)\psi\}(\omega) = (2\pi)^{-3/2}\int_{S^2} e^{ik\omega\omega'}\psi(\omega')d\omega',$$
$$\{\eta_+^*(r,k)\psi\}(\omega) = \{\eta_0^*(r,k)\psi\}(\omega) - u(r\omega, -k, f_0), \qquad (2.18)$$

for $\psi \in L_2(S^2)$ with

$$f_0(x) = Q(x) \cdot \{\eta_0^*(|x|,k)\psi\}(\tilde{x}), \qquad (2.19)$$

respectively. They are the "eigenoperators" associated with H and H_0, respectively. Then it can be shown [Saitō,[14] and Saitō,[15] Theorem 10.6, (v) of Lemma 2.1, Theorem 2.6 and (1) of Remark 3.3] that

$$\{\eta_0^*(|x|,k)\psi\}(\tilde{x}) \in L_{2,-\delta}(\mathbf{R}^3),$$
$$\{\eta_+^*(|x|,k)\psi\}(\tilde{x}) \in L_{2,-\delta}(\mathbf{R}^3), \qquad (2.20)$$

for any $\psi \in L_2(S^2)$, where $\tilde{x} = x/|x|$. Then, $u(x) = \{\eta_+^*(|x|,k)\psi\}(\tilde{x})$ satisfies $(-\Delta + Q(x) - k^2)u = 0$ at least in the sense of distributions. Let $F(k)$ be as in (1.6). The following representation formula for $F(k)$ is well known [Ikebe,[16] Theorem 1; Agmon,[11] Theorem 7.2; and Saitō,[15] Theorem 3.2 and (i) of Remark 3.3]:

$$(F(k)\psi,\psi')_{S^2} = -2\pi^2\int_{\mathbf{R}^3} Q(y)\{\eta_0^*(|y|,k)\psi\}(\tilde{y}) \cdot \overline{\{\eta_+^*(|y|,k)\psi\}(\tilde{y})}dy, \qquad (2.21)$$

for $\psi, \psi' \in L_2(S^2)$, where $(\,,\,)_{S^2}$ denotes the inner product of $L_2(S^2)$. By the use of (2.18) we have from (2.21)

$$(F(k)\psi,\psi')_{S^2} = -2\pi^2\int_{\mathbf{R}^3} Q(y)\{\eta_0^*(|y|,k)\psi\}(\tilde{y}) \times \overline{\{\eta_0^*(|y|,k)\psi'\}(\tilde{y})}dy,$$
$$+ 2\pi^2\int_{\mathbf{R}^3} Q(y)\{\eta_0^*(|y|,k)\psi\}(\tilde{y})\,\overline{u(y, -k, f_0)}dy, \qquad (2.22)$$

with f_0 given in (2.19). Especially if

$$\psi(\omega) = \psi'(\omega) = \psi_{x,k}(\omega) = e^{-ikx\omega},$$

then, noting that

$$\{\eta_0^*(|y|,k)\psi_{k,x}\}(\tilde{y}) = (2\pi)^{-3/2}\int_{S^2} e^{ik(y-x)\omega}\,d\omega = \left(\frac{2}{\pi}\right)^{1/2}\frac{\sin(k|y-x|)}{k|y-x|}, \tag{2.23}$$

and using (2.22), we have

$$\begin{aligned} f(x,k) &= k^2(F(k)\psi_{k,x},\psi_{k,x})_{S^2} \\ &= -4\pi\int_{\mathbf{R}^3} Q(y)\{\phi(y-x)\}^2\,dy \\ &\quad + 4\pi\int_{\mathbf{R}^3} Q(y)\phi(y-x) \\ &\quad \times u(y,-k,Q(\cdot)\phi(\cdot-x))dy \\ &= f_1(x,k) + f_2(x,k), \end{aligned} \tag{2.24}$$

with

$$\phi(y-x) = \phi_k(y-x) = [\sin(k|y-x|)]/|y-x|, \tag{2.25}$$

where we should also note that

$$\begin{aligned} & u(y,-k,(\pi/2)^{1/2}k^{-1}Q(\cdot)\phi(\cdot-x)) \\ &= R(-k)\{(\pi/2)^{1/2}k^{-1}Q(\cdot)\phi(\cdot-x)\} \\ &= (\pi/2)^{-1}k^{-1}u(y,-k,Q(\cdot)\phi(\cdot-x)). \end{aligned} \tag{2.26}$$

(4) The following estimate will be used: Let s and t satisfy $0<s<3$, $t<3$, and $s+t>3$. Then there exists a positive constant $C=C(s,t)$ depending only on s and t such that

$$\int_{\mathbf{R}^3}\frac{(1+|y|)^{-t}}{|y-x|^s}\,dy \leqslant C(1+|x|)^{-(s+t-3)} \quad (x\in\mathbf{R}^3). \tag{2.27}$$

For the proof of (2.27) see Appendix A.

III. ESTIMATE FOR $f(x,k)$

Let $f(x,k) = f_1(x,k) + f_2(x,k)$ be given in (2.24). Set

$$f_1(x,k) = -2\pi\int_{\mathbf{R}^3}\frac{Q(y)}{|x-y|^2}\,dy + f_{12}(x,k), \tag{3.1}$$

with

$$\begin{aligned} f_{12}(x,k) &= 2\pi\int_{\mathbf{R}^3}\frac{Q(y)}{|x-y|^2}\cos(2k|x-y|)dy \\ &= 2\pi\int_{\mathbf{R}^3}\frac{Q(x+y)}{|y|^2}\cos(2k|y|)dy. \end{aligned} \tag{3.2}$$

Introducing the spherical coordinates and integrating by parts with respect to the radial coordinate r, we obtain

$$\begin{aligned} f_{12}(x,k) &= 2\pi\int_{S^2}\left\{\int_0^\infty Q(x+r\omega)\cos(2kr)dr\right\}d\omega \\ &= -\frac{\pi}{k}\int_{S^2}\int_0^\infty\frac{\partial Q(x+r\omega)}{\partial r}\sin(2kr)dr\,d\omega \\ &= -\frac{\pi}{k}\int_{\mathbf{R}^3}\left\{\sum_{j=1}^3 (D_jQ)(y)\left(\frac{y_j-x_j}{|y-x|}\right)\right\} \\ &\quad \times|x-y|^{-2}\sin(2k|x-y|)dy \\ &\quad \left(D_j = \frac{\partial}{\partial y_j}\right). \end{aligned} \tag{3.3}$$

Therefore we have

$$\begin{aligned} |f_{12}(x,k)| &\leqslant \frac{\pi}{|k|}\int_{\mathbf{R}^3}\frac{3C_0(1+|y|)^{-\rho}}{|x-y|^2}\,dy \\ &\leqslant (3\pi C_0C/k)(1+|x|)^{-(\rho-1)}, \end{aligned} \tag{3.4}$$

where we used (1.13) with $|\alpha|=1$ and (2.27) with $s=2$ and $t=\rho$, $C=C(\rho)$ is a constant depending only on ρ. Notice that $\rho<3$ by (1.15).

Let us next estimate the term

$$f_2(x,k) = 4\pi\int_{\mathbf{R}^3} Q(y)\phi(x-y)\,\overline{u(y,-k,Q(\cdot)\phi(\cdot-x))}dy. \tag{3.5}$$

Let $a>0$. It follows from the Schwarz inequality, (1.12), and (2.14) that

$$\begin{aligned} & |f_2(x,k)| \\ &\leqslant 4\pi\left\|\frac{Q}{|\cdot-x|}\right\|_\delta \|u(\cdot,-k,Q(\cdot)\phi(\cdot-x))\|_{-\delta} \\ &\leqslant \frac{4\pi C_1}{k}\left\|\frac{Q}{|\cdot-x|}\right\|_\delta^2 \\ &\leqslant \frac{4\pi C_1C_0}{k}\int_{\mathbf{R}^3}\frac{(1+|y|)^{-2(\beta-\delta)}}{|y-x|^2}\,dy \\ &\quad [k\geqslant 0,\quad C_1 = C_1(a,Q)]. \end{aligned} \tag{3.6}$$

Using (2.27) with $s=2$ and $t=2(\beta-\delta)$, which is less than 3 since $\beta\leqslant 2$ [(1.15)] and $\delta>\frac{1}{2}$, we obtain

$$|f_2(x,k)| \leqslant [4\pi C_1C_0(\beta,\delta)/k](1+|x|)^{-(2\beta-2\delta-1)}, \tag{3.7}$$

with a constant $C(\beta,\delta)$ depending only on β and δ. Since $\beta-\delta>\frac{3}{4}$ by (2.1), we have $2\beta-2\delta-1>\frac{1}{2}$. Thus, setting

$$b = \max(\rho-1, 2\beta-2\delta-1) > \tfrac{1}{2}, \tag{3.8}$$

we arrive at the following proposition.

Proposition 3.1: Let Assumption 1.1 be satisfied. Let a be a positive number. Then there exists a positive constant $C_4 = C_4(\beta,\rho,\delta,C_0,a)$ depending only on β,ρ,δ,C_0,a, where C_0 is given in (1.12) and (1.13), such that

$$\left|f(x,k) + 2\pi\int_{\mathbf{R}^3}\frac{Q(y)}{|x-y|^2}\,dy\right| \leqslant \frac{C_4}{k}(1+|x|)^{-b} \tag{3.9}$$

holds for any $k\geqslant a$ and any $x\in\mathbf{R}^3$ with $b<\frac{3}{2}$ given in (3.8). Therefore

$$\left\|f(\cdot,k) + 2\pi\int_{\mathbf{R}^3}\frac{Q(y)}{|\cdot-y|^2}\,dy\right\|_{L_2(\mathbf{R}^3)} \leqslant \frac{C_5}{k} \quad (k\geqslant a) \tag{3.10}$$

with a constant $C_5 = C_5(\beta,\rho,\delta,C_0,a)$.

IV. ESTIMATE FOR $D_j\,f(x,k)$

Let $f_1(x,k)$ and $f_2(x,k)$ be given in (2.24). Let us first estimate $D_j\,f_1(x,k)$. Setting, as in (3.1),

$$f_{12}(x,k) = f_1(x,k) + 2\pi \int_{\mathbf{R}^3} \frac{Q(y)}{|y-x|^2} dy, \tag{4.1}$$

we have, from (3.1) and (3.2),

$$\begin{aligned} D_j f_{12}(x,k) &= 2\pi \frac{\partial}{\partial x_j} \left\{ \int_{\mathbf{R}^3} \frac{Q(y)}{|x-y|^2} \cos(2k|x-y|) dy \right\} \\ &= 2\pi \frac{\partial}{\partial x_j} \left\{ \int_{\mathbf{R}^3} \frac{Q(x+y)}{|y|^2} \cos(2k|y|) dy \right\}. \end{aligned} \tag{4.2}$$

Introducing the spherical coordinates and integrating by parts as in Sec. III, we get

$$\begin{aligned} D_j f_{12}(x,k) &= -\frac{\pi}{k} \frac{\partial}{\partial x_j} \left\{ \int_{S^2} \left(\int_0^\infty \frac{\partial Q(xr\omega)}{\partial r} \sin(2kr) dr \right) d\omega \right\} \\ &= -\frac{\pi}{k} \int_{S^2} \int_0^\infty \left(\sum_{l=1}^3 (D_j D_l Q)(x + r\omega)\omega_l \right) \\ &\quad \times \sin(2kr) dr\, d\omega \\ &= -\frac{\pi}{k} \int_{\mathbf{R}^3} \sum_{l=1}^3 (D_j D_l Q)(y) \frac{y_l - x_l}{|y-x|} \\ &\quad \times \sin(2k|y-x|)|x-y|^{-2} dy. \end{aligned} \tag{4.3}$$

Thus, making use of (1.13) and (2.27) with $s = 2$ and $t = \rho$, we obtain the following estimate for $D_j f_{12}(x,k)$.

Proposition 4.1: Let Assumption 1.1 be satisfied. Then we have

$$\left| D_j \left\{ f_1(x,k) + 2\pi \int_{\mathbf{R}^3} \frac{Q(y)}{|y-x|^2} dy \right\} \right| \leqslant (3\pi C_0 C/k)(1+|x|)^{-(\rho-1)} \quad (k>0, \quad x \in \mathbf{R}^3), \tag{4.4}$$

where $C = C(\beta, C_0)$ is given in (3.4).

In order to estimate $f_2(x,k)$, we make use of the second relation of (2.17) to show

$$\begin{aligned} f_2(x,k) &= 4\pi \int_{\mathbf{R}^3} Q(y)\phi(y-x) \overline{u_0(y,-k,Q\phi(\cdot - x))} dy \\ &\quad - 4\pi \int_{\mathbf{R}^3} Q(y)\phi(y-x) \overline{u_0(y,-k,Qu)} dy \\ &= f_{21}(x,k) + f_{22}(x,k), \end{aligned} \tag{4.5}$$

with

$$f_{21}(x,k) = 4\pi \int_{\mathbf{R}^3} Q(y)\phi(y-x) u_0(y,-k,Q\phi(\cdot - x)) dy, \tag{4.6}$$

and

$$f_{22}(x,i) = -4\pi \int_{\mathbf{R}^3} Q(y)\phi(y-k) u_0(y,-k,Qu) dy, \tag{4.7}$$

where

$$Qu = Q(z)u(z) = Q(z)u(z,-k,Q(\cdot)\phi(\cdot - x)). \tag{4.8}$$

The next two lemmas will be used when get convenient expressions for $D_j f_{21}(x,k)$ and $D_j f_{22}(x,k)$ (Proposition 4.4).

Lemma 4.2: Let X be a normed space and let $p(t)$ be an X-valued function on an interval $I = (t_1,t_2)$. Let $q(t)$ be an X-valued, strongly continuous function on I such that

$$p(t+\Delta t) - p(t) = \int_t^{t+\Delta t} q(s) ds \tag{4.9}$$

for any t and Δt with $t, t+\Delta t \in I$.

(i) Then $p(t)$ is strongly differentiable in X at any $t \in I$ with its derivative $p'(t) = q(t)$.

(ii) Let Y be a normed space and let $T \in \mathbf{B}(X,Y)$, i.e., T is a bounded, linear operator from X into Y. Then $Tp(t)$ is strongly differentiable in Y at any $t \in I$ with its derivative $(Tp(t))' = Tq(t)$.

Proof: The proof is almost obvious by using the relation

$$\frac{p(t+\Delta t) - p(t)}{\Delta t} - q(t) = \frac{1}{\Delta t} \int_t^{t+\Delta t} (q(s) - q(t)) ds. \tag{4.10}$$

Q.E.D.

Lemma 4.3: Let $\phi(y)$ be a ($\mathbb{C}$-valued) continuous function on $\mathbf{R}^3 - \{0\}$ such that

$$\begin{aligned} |\phi(y)| &\leqslant c/|y|^\sigma \quad (|y| \leqslant 1), \\ |\phi(y)| &\leqslant c/|y|^\nu \quad (|y| \geqslant 1), \end{aligned} \tag{4.11}$$

with constants $c > 0$, $\sigma < \frac{3}{2}$, and $\nu \in \mathbf{R}$. Set

$$\Phi_x(y) = (1+|y|)^{-\mu} \phi(y-x), \tag{4.12}$$

with $\mu \in \mathbf{R}$. Then $\Phi_x(y)$ is an $L_{2,\gamma}(\mathbf{R}^3)$-valued strongly continuous function of $x \in \mathbf{R}^3$ if

$$\mu + \nu - \gamma > \tfrac{3}{2}. \tag{4.13}$$

The proof will be given in Appendix B because it is rather technical.

Proposition 4.4: Let Assumption 1.1 be satisfied. Let $f_{21}(x,k)$ and $f_{22}(x,k)$ be given in (4.6) and (4.7), respectively. Then, for each $k > 0$, $f_{21}(x,k)$ and $f_{22}(x,k)$ are C^1 functions on $\mathbf{R}^3$ with

$$\begin{aligned} D_j f_{21}(x,k) &= -4\pi \int_{\mathbf{R}^3} \{Q(y)((D_j\phi)(y-x)) \\ &\quad \times \overline{u_0(y,-k,Q(\cdot)\phi(\cdot - x))} + Q(y)\phi(y-x) \\ &\quad \times \overline{u_0(y,-k,Q(\cdot)(D_j\phi)(\cdot - x))}\} dy, \end{aligned} \tag{4.14}$$

$$\begin{aligned} D_j f_{22}(x,k) &= 4\pi \int_{\mathbf{R}^3} \{Q(y)((D_j\phi)(y-x)) \overline{u_0(y,-k,Qu)} \\ &\quad + Q(y)\phi(y-x) \\ &\quad \times \overline{u_0(y,k,Qu(\cdot,-k,Q(D_j\phi)(\cdot - x)))}\} dy, \end{aligned} \tag{4.15}$$

for $j = 1,2,3$.

Proof: Since

$$(D_j\phi)(y) = \frac{y_j}{|y|} \frac{k\cos(k|y|)}{|y|} - \frac{\sin(k|y|)}{|y|^2}, \tag{4.16}$$

$(\partial/\partial y_j)\,\phi(y)$ satisifes (4.11) with $\sigma = 1$ and $\nu = 1$. Noting that

$$\beta+1-\delta>(\tfrac{5}{4}+\delta)+1-\delta=\tfrac{9}{4}>\tfrac{3}{2}, \tag{4.17}$$

where we used (2.1), we can apply Lemma 4.3 to see that $Q(y)(D_j\phi)(y-x)$ is an $L_{2,\delta}(\mathbf{R}^3)$-valued, strongly continuous function of $x\in\mathbf{R}^3$. Set $j=1$ for the sake of simplicity. Then we have

$$Q(y)\phi(y-(x+(\Delta x_1,0,0)))-Q(y)\phi(y-x)$$
$$=-\int_{x_1}^{x_1+\Delta x_1} Q(y)(D_1\phi)(y-(s,x_2,x_3))ds, \tag{4.18}$$

for almost all $y\in\mathbf{R}^3$, and hence it follows from Lemma 4.2 (i) that $Q(y)\phi(y-x)$ is partially strongly differentiable in $L_{2,\delta}$ with respect to x_1 with the partial derivative

$$\frac{\partial}{\partial x_1}Q(y)\phi(y-x)=-Q(y)(D_1\phi)(y-x). \tag{4.19}$$

It also follows from Lemma 4.2 (ii) that

$$\begin{aligned}\frac{\partial}{\partial x_1}u_0(y,-kQ(\cdot)\phi(\cdot-x))&=\frac{\partial}{\partial x_1}\{R_{0-}(k)(Q(\cdot)\phi(\cdot-x))\}\\&=-R_{0-}(k)(Q(\cdot)(D_1\phi)(\cdot-x))\\&=-u_0(y,-k,Q(\cdot)(D_1\phi)(\cdot-x)),\end{aligned}$$
$$\tag{4.20}$$
$$\begin{aligned}\frac{\partial}{\partial x_1}u_0(y,-kQ(\cdot)u)&=\frac{\partial}{\partial x_1}\{R_{0-}(k)QR_-(k)(Q(\cdot)\phi(\cdot-x))\}\\&=-u_0(y,-k,Qu(\cdot,-k,Q(D_1\phi)(\cdot-x))),\end{aligned}$$

in $L_{2,\delta}(\mathbf{R}^3)$. The cases where $j=2,3$ can be treated in quite the same way. Thus, (4.14) and (4.15) are obtained from (4.19) and (4.20). Q.E.D.

Now we are in a position to give an estimate for $f_{22}(x,k)$.

Proposition 4.5: Let Assumption 1.1 be satisfied. Let a be a positive number. Then there exists a constant $C_6=C_6(\beta,\delta,C_0,a)$ depending only on β, δ, a, and the constant C_0 given in (1.12) and (1.13) such that

$$|D_jf_{22}(x,k)|\leqslant(C_6/k)(1+|x|)^{-(2\beta-2\delta-1)}$$
$$(x\in\mathbf{R}^3,\quad k\geqslant a,\quad j=1,2,3). \tag{4.21}$$

Proof: Since it follows from (4.16) that

$$|\phi(y-x)|\leqslant 1/|x-y|,$$
$$\tag{4.22}$$
$$|(D_j\phi)(y-x)|\leqslant kC_7/|y-x|\quad(k\geqslant1,\quad y\neq x),$$

with a constant $C_7>0$, we obtain, from the Schwarz inequality and (2.14),

$$\begin{aligned}&\left|\int_{\mathbf{R}^3}Q(y)((D_j\phi)(y-x))\overline{u_0(y,-k,Qu)}dy\right|\\&\leqslant\|Q(\cdot)((D_j\phi)(\cdot-x))\|_\delta\\&\quad\times\|u_0(\cdot,-k,Qu(\cdot,-k,Q\phi(\cdot-x))\|_{-\delta}\\&\leqslant kC_7\left|\left|\frac{Q}{\cdot-x}\right|\right|_\delta\frac{C_1}{k}\|Qu(\cdot,-k,Q\phi(\cdot-x))\|_\delta\\&[k\geqslant a,\quad C_1=C_1(a,Q)]\\&\leqslant C_1C_7\left|\left|\frac{Q}{|\cdot-x|}\right|\right|_\delta\|u(\cdot,-k,Q\phi(\cdot-x))\|_{-\delta}\\&\leqslant C_1C_7\left|\left|\frac{Q}{|\cdot-x|}\right|\right|_\delta\frac{C_1}{k}\left|\left|\frac{Q}{|\cdot-x|}\right|\right|_\delta\\&\leqslant\frac{C_1^2C_7}{k}\left|\left|\frac{Q}{|\cdot-x|}\right|\right|_\delta^2\\&\leqslant\frac{C_1^2C_7}{k}C(\beta,\delta)(1+|x|)^{-(2\beta-2\delta-1)},\end{aligned} \tag{4.23}$$

with a constant $C(\beta,\delta)$, where we have also used (2.27) with $s=2$ and $t=2(\beta-\delta)$, as in the proof of Proposition 3.1 [see (3.6) and (3.7)]. In the same manner we have

$$\begin{aligned}&\left|\int_{\mathbf{R}^3}Q(y)\phi(y-x)\right.\\&\qquad\left.\times u_0(y,k,Qu(\cdot,-k,Q(D_j\phi)(\cdot-x)))dy\right|\\&\leqslant[C_1^2C_7C(\beta,\delta)/k](1+|x|)^{-(2\beta-2\delta-1)}.\end{aligned} \tag{4.24}$$

We have (4.21) from (4.23), (4.24), and (4.15). Q.E.D.

Let us finally estimate $f_{21}(x,k)$.

Proposition 4.6: Let Assumption 1.1 be satisfied. Then we have

$$\begin{aligned}u_0(y,k,Q(\cdot)(D_j\phi)(\cdot-x))&=\frac{\partial}{\partial y_j}u_0(y,-k,Q\phi(\cdot-x))\\&\quad-u_0(y,-k,(D_jQ)\phi(\cdot-x)),\end{aligned} \tag{4.25}$$

in $L_{2,-\delta}(\mathbf{R}^3)$.

Proof: It follows from (2.9) and the relation

$$(D_j\phi)(z-x)=\frac{\partial\phi(z-x)}{\partial z_j} \tag{4.26}$$

that

$$u_0(y,-k,Q(\cdot)(D_j\phi)(\cdot-x))=\frac{1}{4\pi}\int_{\mathbf{R}^3}\frac{e^{-ik|z-y|}}{|z-y|}Q(z)\frac{\partial\phi(z-x)}{\partial z_j}dz. \tag{4.27}$$

By the use of partial integration and the relation

$$\frac{\partial}{\partial z_j}\left(\frac{e^{-ik|z-y|}}{|z-y|}\right)=-\frac{\partial}{\partial y_j}\left(\frac{e^{-ik|z-y|}}{|z-y|}\right), \tag{4.28}$$

we obtain, from (4.27),

$$\begin{aligned}&u_0(y,-k,Q(\cdot)(D_j\phi)(\cdot-x))\\&=-\frac{1}{4\pi}\int_{\mathbf{R}^3}\frac{\partial}{\partial z_j}\left\{\frac{e^{-ik|z-y|}}{|z-y|}Q(z)\right\}\phi(z-x)dz\\&=+\frac{1}{4\pi}\int_{\mathbf{R}^3}\frac{\partial}{\partial y_j}\left\{\frac{e^{-ik|z-y|}}{|z-y|}\right\}Q(z)\phi(z-x)dz\\&\quad-\frac{1}{4\pi}\int_{\mathbf{R}^3}\frac{e^{-ik|z-y|}}{|z-y|}(D_jQ)(z)\phi(z-x)dz\\&\equiv J_1+J_2,\end{aligned} \tag{4.29}$$

with $x\neq y$. Here, rigorously speaking, the partial integration should be applied in the region $B_{1/\epsilon}(0)-B_\epsilon(x)-B_\epsilon(y)$ with

$$B_\eta(z)=\{z\in\mathbf{R}^3/|z|<\eta\} \tag{4.30}$$

and ϵ should be taken to zero afterwards. But, when $x \neq y$, the singularities at $z = x$ and $z = y$ are of order $1/|z-x|$ and $1/|z-y|$, respectively, and the integrand goes to zero rapidly enough at infinity. Therefore (4.29) is justified. We have from (2.9) that

$$J_2 = -u_0(y, -k, (D_jQ)\,\phi(\cdot - x)). \tag{4.31}$$

Proceeding as in the proof of Lemma 4.3 (Appendix B), we can easily see that the integrand of J_1 is an $L_1(\mathbf{R}^3)$-valued, strongly continuous function of $y \in \mathbf{R}^3$. Therefore Lemma 4.2 can be applied to show that

$$J_1 = \frac{\partial}{\partial y_j}\left\{\frac{1}{4\pi}\int_{\mathbf{R}^3}\frac{e^{-ik|z-y|}}{|z-y|}Q(z)\phi(z-x)dz\right\}$$
$$= \frac{\partial}{\partial y_j}u_0(y, -k, Q(\cdot)\phi(\cdot - x)). \tag{4.32}$$

Since the right-hand side of (4.32) belongs to $L_{2,-\delta}(\mathbf{R}^3)$, (4.25) follows from (4.29), (4.31), and (4.32). Q.E.D.

Proposition 4.7: Let Assumption 1.1 be satisfied. Then we have

$$D_j f_{21}(x,k) = 4\pi\left[\int_{\mathbf{R}^3}(D_jQ)(y)\phi(y-x) \times \overline{u_0(y, -k, Q(\cdot)\phi(\cdot - x))}dy + \int_{\mathbf{R}^3}Q(y)\phi(y-x) \times \overline{u_0(y, -k, (D_jQ)\phi(\cdot - x))}dy\right]. \tag{4.33}$$

Proof: We have, from (4.14) and Proposition 4.6,

$$D_j f_{21}(x,k) = -4\pi\left[\int_{\mathbf{R}^3}Q(y)\frac{\partial\phi(y-x)}{\partial y_j} \times \overline{u_0(y, -k, Q(\cdot)\phi(\cdot - x))}dy + \int_{\mathbf{R}^3}Q(y)\phi(y-x) \times \frac{\partial}{\partial y_j}\overline{u_0(y, -k, Q(\cdot)\phi(\cdot - x))}dy - \int_{\mathbf{R}^3}Q(y)\phi(y-x) \times \overline{u_0(y, -k, (D_jQ)\phi(\cdot - x))}dy\right]. \tag{4.34}$$

By the use of integration by parts we have

$$\int_{\mathbf{R}^3}Q(y)\phi(y-x)\frac{\partial}{\partial y_j}\overline{u_0(y, -k, Q(\cdot)\phi(\cdot - x))}dy$$
$$= -\int_{\mathbf{R}^3}(D_jQ)(y)\phi(y-x) \times \overline{u_0(y, -k, Q(\cdot)\phi(\cdot - x))}dy - \int_{\mathbf{R}^3}Q(y)\frac{\partial\phi(y-x)}{\partial y_j} \times \overline{u_0(y, -k, Q(\cdot)\phi(\cdot - x))}dy. \tag{4.35}$$

As in the proof of Proposition 4.6, we have to be careful of the singularities $y = x$ and $y = \infty$. But, noting that $\phi(y-x)$ is bounded at $y = x$ and that $u_0(y)$ is locally L_2 and the integrand behaves reasonably well at $y = \infty$, we can easily justify the procedure of partial integration. Thus we get (4.33) from (4.34) and (4.35). Q.E.D.

Now we can finish estimating $D_j f(x,k)$.

Proposition 4.8: Let Assumption 1.1 be satisfied. Let a be a positive number. Then there exists a positive constant $C_7 = C_7(\beta, \rho, \delta, C_0, a)$ such that

$$\left|D_j\left\{f(x,k) + 2\pi\int_{\mathbf{R}^3}\frac{Q(y)}{|y-x|^2}dy\right\}\right| \leqslant (C_7/k)(1+|x|)^{-b} \quad (j = 1,2,3), \tag{4.36}$$

holds for any $k \geqslant a$ and any $x \in \mathbf{R}^3$ with b given in (3.8). Therefore

$$\left\|D_j\left\{f(\cdot,k) + 2\pi\int_{\mathbf{R}^3}\frac{Q(y)}{|y-\cdot|^2}dy\right\}\right\|_{L_2(\mathbf{R}^3)} \leqslant C_8/k \quad (j = 1,2,3), \tag{4.37}$$

with a constant $C_8 = C_8(\beta, \rho, \delta, C_0, a)$.

Proof: It follows from Propositions 4.1 and 4.5 that we have only to show the estimate

$$|D_j f_{21}(x,k)| \leqslant (C_9/k)(1+|x|)^{-(2\beta - 2\delta - 1)} \quad (j = 1,2,3), \tag{4.38}$$

for any $x \in \mathbf{R}^3$ and $k \geqslant a$. Proceeding as in the proof of Proposition 4.5, we can see that the first and second terms of the right-hand side of (4.33) are estimated by $\text{const} \times k^{-1}\|Q(\cdot)|\cdot - x|^{-1}\|_\delta \|(D_jQ)|\cdot - x|^{-1}\|_\delta$, whence we get (4.38) by using (2.27) with $s = 2$ and $t = 2\beta - 2\delta$, where we should note that $2\rho - 2\delta < 2\beta - 2\delta$ by (1.15).

V. PROOF OF THE MAIN THEOREMS

The proof of Theorems 1.2 and 1.3 will be divided into several steps.

(I) It follows from Propositions 3.1 and 4.8 that $g(x,k)$ is a C^1 function of $x \in \mathbf{R}^3$ for any $k > 0$ and satisfies the estimates

$$|D^\alpha g(x,k)| \leqslant (C_{10}/k)(1+|x|)^{-b} \quad (x \in \mathbf{R}^3, \quad k \geqslant a, \quad |\alpha| = 0,1), \tag{5.1}$$

where $C_{10} = C_{10}(\beta, \rho, \delta, C_0, a)$ is a constant depending only on $\beta, \rho, \delta, C_0, a$, with $a > 0$ and $b = \min(\rho - 1, 2\beta - 2\delta - 1)$ as in (3.8). Since we have, from (2.1) and the assumption $\rho > \frac{5}{2}$,

$$b > \tfrac{3}{2}, \tag{5.2}$$

we get (1.17) in Theorem 1.2 with $C = C_{10}$. Using (2.27), we obtain

$$\left|\int_{\mathbf{R}^3}\frac{Q(y)}{|y-x|^2}dy\right| \leqslant C_{11}(1+|x|)^{-(\beta-1)} \quad (x \in \mathbf{R}^3, \quad k > 0), \tag{5.3}$$

$$\left|\frac{\partial}{\partial x_j}\int_{\mathbf{R}^3}\frac{Q(y)}{|y-x|^2}dy\right| = \left|\int_{\mathbf{R}^3}\frac{(D_jQ)(y)}{|y-x|^2}dy\right| \leqslant C_{11}(1+|x|)^{-(\rho-1)} \quad (x \in \mathbf{R}^3, \quad k > 0, \quad j = 1,2,3), \tag{5.4}$$

with a constant $C_{11} = C_{11}(C_0)$. Thus (1.18) in Theorem 1.2 follows from (5.1), (5.3), and(5.4), which completes the proof of Theorem 1.2.

(II) Let $h(x)$ be a function on $\mathbb{R}^3$ such that

$$|h(x)| \leqslant C_{12}(1+|x|)^{-\tau} \quad (x \in \mathbb{R}^3), \tag{5.5}$$

with constants $C_{12} > 0$ and $\tau > 0$. Then, as is shown in Saitō,[6] Proposition 1.3 or Saitō,[7] Appendix A, $\mathscr{F}^*|\xi|\mathscr{F}h$ defined by (1.19) belongs to $\mathscr{S}'(\mathbb{R}^3)$. Thus we see that

$$\begin{aligned} &\mathscr{F}^*|\xi|\mathscr{F}f(\cdot,k) \in \mathscr{S}'(\mathbb{R}^3), \\ &\mathscr{F}^*|\xi|\mathscr{F}f(\cdot,\infty) \in \mathscr{S}'(\mathbb{R}^3), \\ &\mathscr{F}^*|\xi|\mathscr{F}g(\cdot,k) \in \mathscr{S}'(\mathbb{R}^3). \end{aligned} \tag{5.6}$$

By Ref. 6, Theorem 3.1 (or Ref. 7, Theorem 3.2), we have

$$Q(x) = -(4\pi^3)^{-1}\{\mathscr{F}^*|\xi|\mathscr{F}f(\cdot,\infty)\}(x), \tag{5.7}$$

which, together with the relation $f(x,k) - f(x,\infty) = g(x,k)$, gives

$$\begin{aligned} &-(4\pi^3)^{-1}\mathscr{F}^*|\xi|\mathscr{F}f(\cdot,k) - Q \\ &\quad = \mathscr{F}^*|\xi|\mathscr{F}g(\cdot,k) \quad (k>0). \end{aligned} \tag{5.8}$$

(III) Since $f(x,k)$ and $g(x,k)$ are small at infinity, they belong to $\mathscr{S}'(\mathbb{R}^3)$. Therefore we have

$$\begin{aligned} &\{\mathscr{F}f(\cdot,k)\}(\xi) = (1/i\xi_j)\mathscr{F}(D_j f(\cdot,k))(\xi), \\ &\{\mathscr{F}g(\cdot,k)\}(\xi) = (1/i\xi_j)\mathscr{F}(D_j g(\cdot,k))(\xi) \\ &\qquad [\xi = (\xi_1,\xi_2,\xi_3) \in \mathbb{R}^3, \quad j = 1,2,3,...]. \end{aligned} \tag{5.9}$$

The right-hand sides of (5.9) are not only elements of $\mathscr{S}'(\mathbb{R}^3)$ but also functions, because $D_j f(x,k)$ and $D_j g(x,k)$ are in $L_2(\mathbb{R}^3)$ by Theorem 1.2. Let A_j and $\chi_j(\xi)$ be as in (1.20) and (1.21), respectively. Then we have

$$\begin{aligned} &|\xi|\mathscr{F}f(\cdot,k) = \sum_{j=1}^{3} \frac{|\xi|}{i\xi_j}\chi_j(\xi)\mathscr{F}(D_j f(\cdot,k)), \\ &|\xi|\mathscr{F}g(\cdot,k) = \sum_{j=1}^{3} \frac{|\xi|}{i\xi_j}\chi_j(\xi)\mathscr{F}(D_j g(\cdot,k)). \end{aligned} \tag{5.10}$$

Noting that

$$|(|\xi|/i\xi_j)\chi_j(\xi)| \leqslant \sqrt{3}, \tag{5.11}$$

we see that each term of the right-hand sides of (5.10) belongs to $L_2(\mathbb{R}^3_\xi)$. Thus it follows that $\mathscr{F}^*|\xi|\mathscr{F}f(\cdot,k)$ and $\mathscr{F}^*|\xi|\mathscr{F}g(\cdot,k)$ are well defined as elements of $L_2(\mathbb{R}^3)$. Further, we obtain, from (5.1) and (5.11),

$$\begin{aligned} &\|\mathscr{F}^*|\xi|\mathscr{F}g(\cdot,k)\|_{L_2} \leqslant 3\sqrt{3}\sum_{j=1}^{3}\|(D_j g)(\cdot,k)\|_{L_2} \leqslant \frac{C_{12}}{k} \\ &\qquad [k \geqslant a, \quad C_{12} = C_{12}(\delta,Q,a), \quad a>0], \end{aligned} \tag{5.12}$$

which completes the proof of Theorem 1.3.

APPENDIX A: PROOF OF (2.27)

This estimate is given in Kuroda[17] without proof. Let

$$f(x) = \int_{\mathbb{R}^3} \frac{(1+|y|)^{-t}}{|y-x|^s}\,dy, \tag{A1}$$

with $0<s<3$, $t<3$, and $s+t>3$. It is sufficient to show the following: (a) $f(x)$ is a continuous function of $x \in \mathbb{R}^3$; and (b) there exists a constant $c_0 > 0$ such that

$$f(x) \leqslant c_0|x|^{-(t+s-3)} \quad (x \in \mathbb{R}^3 - \{0\}). \tag{A2}$$

Let us first show (a). Let $x_0 \in \mathbb{R}^3$. Taking $R>0$ such that $R>|x_0|$, we set

$$\begin{aligned} f(x) &= \int_{\mathbb{R}^3} \frac{1}{|y|^s(1+|y+x|)^t}\,dy \\ &= \int_{|y|>2R} + \int_{|y|<2R} = f_1(x) + f_2(x), \end{aligned} \tag{A3}$$

for $x \in \mathbb{R}^3$ with $|x| \leqslant R$. Then we have

$$\lim_{x\to x_0} f_1(x) = f_1(x_0), \tag{A4}$$

because since $|y+x| \geqslant |y|/2$ by $|y| \geqslant 2R \geqslant 2|x|$, the integrand of $f_1(x)$ is dominated by an integrable function which is independent of x and we can apply the Lebesgue convergence theorem. The continuity of $f_2(x)$ can be proved easily, since the integrand is dominated by $|y|^{-s}$. Let us next show (b). Let $x \neq 0$ and set

$$\begin{aligned} f(x) &= \int_{\mathbb{R}^3} \frac{C(1+|y|)^{-t}}{|y-x|^s}\,dy \\ &= \int_{|y|<|x|/2} + \int_{|y|>2|x|} + \int_{|x|/2<|y|<2|x|} \\ &\equiv g_1(x) + g_2(x) + g_3(x). \end{aligned} \tag{A5}$$

Since we have $|y-x| \geqslant 2^{-1}|x|$ from $|y| \leqslant 2^{-1}|x|$ and $t<3$, $g_1(x)$ is estimated as

$$g_1(x) \leqslant \int_{|y|<|x|/2} \left(\frac{1}{2}|x|\right)^{-s}|y|^{-t}\,dy \leqslant c_1|x|^{-(s+t-3)}, \tag{A6}$$

with a constant $c_1 > 0$. If $|y| \geqslant 2|x|$, we have $|y-x| \geqslant 2^{-1}|y|$. Therefore we get

$$g_2(x) \leqslant \int_{|y|>2|x|} \left(\frac{|y|}{2}\right)^{-s}|y|^{-t}\,dy \leqslant c_2|x|^{-(s+t-3)}. \tag{A7}$$

Since

$$\begin{aligned} g_3(x) &\leqslant \int_{|x|/2<|y|<2|x|} |y-x|^{-s}\left(\frac{|x|}{2}\right)^{-t}dy \\ &\leqslant \frac{2^t}{|x|^t}\int_{|x+z|<2|x|} |z|^{-s}\,dz, \end{aligned} \tag{A8}$$

noting that we have $|z| \leqslant 3|x|$ from $|x+z| \leqslant 2|x|$, we get

$$g_3(x) \leqslant \frac{2^t}{|x|^t}\int_{|z|<3|x|} |z|^{-s}\,dz \leqslant c_3|x|^{-(s+t-3)}. \tag{A9}$$

Thus (b) follows from (A6), (A7), and (A9).

APPENDIX B: PROOF OF LEMMA 4.3

Let $x,x_0 \in \mathbb{R}^3$ and set

$$\begin{aligned} h(x,x_0) &= \int_{\mathbb{R}^3} (1+|y|)^{2\gamma-2\mu} \\ &\qquad \times |\phi(y-x) - \phi(y-x_0)|^2\,dy \\ &= \int_{|y|<R} + \int_{|y|>R} \\ &= h_{1R}(x,x_0) + h_{2R}(x,x_0), \end{aligned} \tag{B1}$$

with $R>0$. We have only to show that $\lim_{x\to x_0} h(x,x_0) = 0$.

In order to show this, it is sufficient to show the following:

(a) $\lim_{R\to\infty} h_{2R}(x,x_0) = 0$ uniformly when $x \to x_0$,

(b) $\lim_{x\to x_0} h_{1R}(x,x_0) = 0$ for any fixed $R > 0$.

Since (a) is easy from the condition $\mu + \nu - \gamma > \frac{3}{2}$, let us prove (b). It is sufficient to show that

$$\tilde{h}(x,x_0) = \int_{|y-x_0|<1} |\phi(y-x) - \phi(y-x_0)|^2\, dy \to 0 \quad \text{as } x \to x_0, \tag{B2}$$

because of the continuity of $\phi(y)$ on $\mathbb{R}^3 - \{0\}$. Set

$$K_2 = \int_{|y-x_0|<1} |\phi(y-x)|^2\, dy,$$
$$K_3 = \int_{|y-x_0|<1} \phi(y-x_0)\,\overline{\phi(y-x)}\,dy. \tag{B3}$$

We can assume that $|x-x_0| < \frac{1}{2}$. Since

$$\begin{aligned}\{y/|y-x_0|<1\} &= \{y/|y-x|<1\} \cup \{y/|y-x| \geqslant 1,\ |y-x_0|<1\} \\ &\quad - \{y/|y-x|<1,\ |y-x_0|>1\} \\ &\equiv B_1 \cup B_2 - B_3,\end{aligned} \tag{B4}$$

we have

$$K_2 = \int_{B_1} + \int_{B_2} - \int_{B_3} = K_{21} + K_{22} + K_{23}, \tag{B5}$$

with

$$\begin{aligned}K_{21} &= \int_{|y|<1} |\phi(y)|^2 dy, \\ K_{22} &\leqslant \text{const} \times |B_2| \to 0 \quad \text{as } x \to x_0, \\ K_{23} &\leqslant \text{const} \times |B_3| \to 0 \quad \text{as } x \to x_0,\end{aligned} \tag{B6}$$

where $|B_j|$ is the Lebesgue measure of B_j and we have used the estimate

$$|y-x| \geqslant |y-x_0| - |x-x_0| > \tfrac{1}{2}, \tag{B7}$$

when $y \in B_3$. Thus we have

$$\lim_{x\to x_0} K_2 = \int_{|y|<1} |\phi(y)|^2\, dx. \tag{B8}$$

Let us consider K_3. Taking $\epsilon \in (0,\frac{1}{2})$, we have

$$\begin{aligned}K_3 &= \int_{|z|<1} \phi(z)\,\overline{\phi(z+x_0-x)}\,dz \\ &= \int_{\epsilon<|z|<1} + \int_{|z|<\epsilon} = K_{31} + K_{32}.\end{aligned} \tag{B9}$$

Assuming $|x-x_0| < \epsilon/2$, we have $|z+x-x_0| > \epsilon/2$ when $\epsilon < |z| < 1$. Thus $|\phi(z)\phi(z+x_0-x)|$ is bounded uniformly for x such that $|x-x_0| < \epsilon/2$. Therefore we have

$$\begin{aligned}K_{31} &= \int_{\epsilon<|z|<1} |\phi(z)|^2\, dz + o(1) \quad (x \to x_0) \\ &= \int_{|z|<1} |\phi(z)|^2\, dz + o(1) \quad (x \to x_0 \text{ and } \epsilon \downarrow 0).\end{aligned} \tag{B10}$$

On the other hand, we have

$$\begin{aligned}|K_{32}| &\leqslant \frac{1}{2}\int_{|z|<\epsilon} |\phi(z)|^2\, dz + \frac{1}{2}\int_{|z|<\epsilon} |\phi(z+x-x_0)|^2\, dz \\ &\leqslant \frac{1}{2}\left\{\int_{|z|<\epsilon} |\phi(z)|^2\, dz + \int_{|z|<3\epsilon/2} |\phi(z)|^2\, dz\right\} = o(1) \quad (\epsilon\downarrow 0),\end{aligned} \tag{B11}$$

where we should note that we get $|z| < 3\epsilon/2$ from $|z| < \epsilon$ and $|x-x_0| < \epsilon/2$. Thus we have

$$\lim_{x\to x_0} K_3 = \int_{|z|<1} |\phi(z)|^2\, dz. \tag{B12}$$

It follows from (B8) and (B12) that

$$\tilde{h}(x,x_0) = \int_{|z|<1} |\phi(z)|^2\, dz + K_2 - 2\,\mathrm{Re}\,K_3 \to 0 \tag{B13}$$

as $x \to x_0$, which completes the proof of Lemma 4.3.

[1] I. M. Gel'fand and B. M. Levitan, "On the determination of a differential equation by its spectral function," Izv. Akad. Nauk. SSSR Ser. Math. **15**, 309 (1951) [Am. Math. Soc. Transl. Ser. 2 **1**, 253 (1955)].

[2] Z. S. Agranovich and V. A. Marchenko, *The Inverse Problem of Scattering Theory* (Gordon and Breach, New York, 1963) (English translation).

[3] R. Newton, "Inverse scattering. II. Three dimensions," J. Math. Phys. **21**, 1698 (1980).

[4] L. D. Fadeev, "Uniqueness of the inverse scattering problem," Vestn. Leningr. Univ. **7**, 126 (1956).

[5] Y. Saitō, "Some properties of the scattering amplitude and the inverse scattering problem," Osaka J. Math. **19**, 527 (1982).

[6] Y. Saitō, "An inverse problem in potential theory and the inverse scattering problem," J. Math. Kyoto Univ. **22**, 315 (1982).

[7] Y. Saitō, "An asymptotic behavior of the *S*-matrix and the inverse scattering problem," J. Math. Phys. **25**, 3105 (1984).

[8] Y. Saitō, "The principle of limiting absorption for second-order differential equations with operator-valued coefficients," Publ. RIMS Kyoto Univ. **7**, 581 (1972).

[9] T. Ikebe and Y. Saitō, "Limiting absorption method and absolute continuity for the Schrödinger operator," J. Math. Kyoto Univ. **7**, 513 (1972).

[10] R. Lavine, "Absolute continuity of positive spectrum for Schrödinger operators with long-range potentials," J. Funct. Anal. **12**, 30 (1973).

[11] S. Agmon, "Spectral properties of Schrödinger operators and scattering theory," Ann. Scuola Nor. Sup. Pisa (4) **2**, 151 (1975).

[12] Y. Saitō, "The principle of limiting absorption for the non-selfadjoint Schrödinger operator in R^N ($N \neq 2$)," Publ. RIMS. Kyoto Univ. **9**, 397 (1974).

[13] Y. Saitō, "Extended limiting absorption method and analyticity of *S*-matrix," J. Reine Angew. Math. **243**, 1 (1983).

[14] Y. Saitō, "Spectral representations for Schrödinger operators with long-range potentials," in *Lecture Notes in Mathematics*, Vol. 727 (Springer, Berlin, 1979).

[15] Y. Saitō, "On the *S*-matrix for Schrödinger operators with long-range potentials," J. Reine Angew. Math. **314**, 99 (1980).

[16] T. Ikebe, "On the phase-shift formula for the scattering operator," Pacific J. Math. **15**, 511 (1965).

[17] S. T. Kuroda, "Construction of eigenfunction expansions by the perturbation method and its application to *n*-dimensional Schrödinger operators," MRC Technical Summary Report #744, 1967, Mathematical Research Center of University of Wisconsin, Madison.

A closed form solution of the s-wave Bethe–Goldstone equation with an infinite repulsive core

Michael K. Ong
Department of Mathematics, Bowdoin College, Brunswick, Maine 04011

(Received 7 November 1985; accepted for publication 11 December 1985)

A closed form s-wave solution of the Bethe–Goldstone equation for the interaction of two nucleons characterized by a potential with an infinite repulsive core is obtained in terms of angular prolate spheroidal wave functions that arise naturally in the analysis of band-limited functions. The asymptotic result for the case of small core radius shows excellent agreement with known results obtained via an approximate iterative procedure.

I. INTRODUCTION

In this paper we present a closed form s-wave solution of the Bethe–Goldstone equation for the interaction of two nucleons characterized by a potential with an infinite repulsive core. The Bethe–Goldstone equation in this case is simply the Schrödinger equation for a pair of interacting nucleons in the Fermi sea, omitting entirely the interaction of these particles with the rest of the Fermi sea (Brueckner's independent pair approximation). The force between two nucleons is singular. The s-wave solutions are the only ones that penetrate to small relative distances where the effects of the singular potential are strongest.

Bethe and Goldstone[1] considered the solution of the integral equation

$$\psi(\mathbf{x}) = e^{i\mathbf{k}\cdot\mathbf{x}} - \int d^3y\, v(y)\psi(\mathbf{y})G(\mathbf{x},\mathbf{y}), \tag{1.1}$$

where $\psi(\mathbf{x})$ is the wave function of the interacting pair, $\hbar\mathbf{k}$ is the relative momentum, $\mathbf{x}$ is the relative distance, and $\hbar\mathbf{k}_F$ is the Fermi momentum. The nonlocal operator G is defined by

$$G(\mathbf{x},\mathbf{y}) = \frac{1}{(2\pi)^3}\int_\Gamma \frac{d^3t}{e(t)}\, e^{i\mathbf{t}\cdot(\mathbf{x}-\mathbf{y})}, \tag{1.2}$$

where the integration is taken over the region outside the Fermi sphere, i.e., Γ: $|\frac{1}{2}\mathbf{P} \pm \mathbf{t}| > k_F$ and $\hbar\mathbf{P}$ is the momentum of the center of mass. In the effective mass approximation, $e(t)$ is taken to be a quadratic of the form

$$e(t) = (t^2 - k^2)/M^*,$$

where M^* is the effective mass.

Using the effective mass approximation, Eq. (1.1) can be transformed into an integrodifferential equation upon applying the operator $\nabla^2 + k^2$, viz.,

$$(\nabla^2 + k^2)\psi(\mathbf{x}) = v(x)\psi(\mathbf{x}) - \int d^3y\, v(y)\psi(\mathbf{y})\, g(\mathbf{x}-\mathbf{y}), \tag{1.3a}$$

where

$$g(\mathbf{x}-\mathbf{y}) = \int_{\overline{\Gamma}} \frac{d^3t}{(2\pi)^3}\, e^{i\mathbf{t}\cdot(\mathbf{x}-\mathbf{y})}. \tag{1.3b}$$

The region of integration $\overline{\Gamma}$ is the complement of the region Γ. The last term in Eq. (1.3a) is nonlocal and it represents the effect of the Pauli's exclusion principle.

For a spherically symmetric potential $v(x)$ and the simplest case when $\mathbf{P} = 0$, the angular components can be eliminated. If $\mathbf{P}\neq 0$, Eq. (1.3b) becomes quite complicated since we have to consider the coupling of different angular momenta. With the assumptions above, we look for the s-wave solution to Eq. (1.3a) in the form

$$\psi(x) = x^{-1}u(x). \tag{1.4}$$

In terms of $u(x)$, Eqs. (1.3) become

$$\left(\frac{d^2}{dx^2} + k^2\right)u(x) = v(x)u(x) - \int_0^\infty \chi(x,y)v(y)u(y)\,dy, \tag{1.5a}$$

where the kernel is

$$\chi(x,y) = \frac{2xy}{\pi}\int_0^{k_F} j_0(tx)\, j_0(ty)t^2\,dt = \frac{1}{\pi}\left[\frac{\sin k_F(x-y)}{x-y} - \frac{\sin k_F(x+y)}{x+y}\right]. \tag{1.5b}$$

Equation (1.5a) will be referred to, from now on, as the s-wave Bethe–Goldstone equation.

In general, the Bethe–Goldstone equation only can be solved numerically through iteration but the subtleties of the physics involved are usually lost in the process. In Ref. 1, Bethe and Goldstone originally demonstrated that the equation can be solved analytically for a hard-core potential. Their approximate iterative solution, however, is valid only for small core radius. In this paper, we present a closed form solution in terms of prolate spheroidal functions. Prolate spheroidal functions arise naturally in the analysis of band-limited functions.[2,3] As it turns out, the repulsive core introduces the band-limiting phenomenon on the wave function thereby allowing us to expand a part of the product $v(x)u(x)$ in a series of prolate spheroidal functions. For small core radius, our result reduces to the approximate solution obtained in Ref. 1.

This paper is organized as follows: In Sec. II, we present the Bethe–Goldstone equation for a repulsive core and derive a Fredholm integral equation with a symmetric kernel. Some known results concerning prolate spheroidal functions and their relationship to the band-limited solutions of the homogeneous integral equation are given in Sec. III. The solution of the s-wave Bethe–Goldstone equation is obtained in Sec. IV. Section V shows that the approximate solution for

 0022-2488/86/041154-05$02.50

a small core radius is in excellent agreement with known results. Finally, the normalization constant and the diagonal elements of the reaction matrix are derived in Secs. VI and VII, respectively.

II. THE s-WAVE EQUATION FOR A REPULSIVE HARD CORE POTENTIAL

For convenience, we introduce the dimensionless variables

$$r = k_F x, \quad r' = k_F y, \quad K = k/k_F,$$
$$v(r) \equiv v(x)/k_F, \quad u(r) \equiv k_F u(x), \tag{2.1}$$

so that Eqs. (1.5) become

$$\left(\frac{d^2}{dr^2} + K^2\right) u(r) = v(r)u(r) - \int_0^\infty \chi(r,r')v(r')u(r')\,dr', \tag{2.2a}$$

where the kernel is

$$\chi(r,r') = \frac{1}{\pi}\left[\frac{\sin(r-r')}{r-r'} - \frac{\sin(r+r')}{r+r'}\right]. \tag{2.2b}$$

Let the hard core potential extend to a distance (dimensionless) c. Since the wave function must vanish inside the infinite potential and be nonzero elsewhere, it must be continuous at $r=c$ but its slope may be discontinuous at the core boundary. The product $v(r)u(r)$ can then be taken as

$$v(r)u(r) = A\delta(r-c) + \omega(r)\theta(c-r), \tag{2.3}$$

where $\theta(c-r)$ is the Heaviside step function and the normalization constant A has to be determined by the condition that the wave function asymptotically goes over into the free-particle wave function so that there is no s-wave phase shift. This implies the condition

$$\lim_{r\to+\infty} [u(r)/r] \to j_0(Kr). \tag{2.4}$$

This is a consequence of the exclusion principle.

The first term in Eq. (2.3) gives the discontinuity of the slope of the wave function at the core boundary while the second term shows that the contribution of $\omega(r)$ is nonvanishing only inside the core. This is possible because of the limiting process $v \to \infty$, $u \to 0$ for $r<c$. This extra term must be determined so that

$$u'' + K^2 u = 0, \quad r<c.$$

Substituting Eq. (2.3) into Eq. (2.2a) and upon using the condition above, we obtain the Fredholm integral equation of the second kind

$$\omega(r) = A\chi(r,c) + \int_0^c \chi(r,r')\omega(r')\,dr', \quad r<c, \tag{2.5}$$

for the unknown function $\omega(r)$, which satisfies the condition

$$\omega(r) = \begin{cases} 0, & r>c, \\ \text{finite}, & r<c. \end{cases} \tag{2.6}$$

For small core radius c, Eq. (2.5) can be quite easily solved by iteration as was originally done by Bethe and Goldstone. For large c, this is no longer possible.

The condition (2.6) implies that the unknown function $\omega(r)$ is band limited. To facilitate our solution of the integral equation (2.5) via band-limited functions, we extend the region of integration to include $-c<r<0$. This requires the extension of $\omega(r)$ and $\chi(r,c)$ as odd functions of r in $-c<r<0$, i.e.,

$$\omega(r) = \begin{cases} \omega(r), & 0<r<c, \\ -\omega(-r), & -c<r<0; \end{cases} \tag{2.7a}$$

$$\chi(r,c) = \begin{cases} \chi(r,c), & 0<r<c, \\ -\chi(-r,c), & -c<r<0. \end{cases} \tag{2.7b}$$

With the adjustments above, Eq. (2.5) becomes

$$\omega(r) - \frac{1}{\pi}\int_{-c}^{c} \frac{\sin(r-r')}{r-r'}\omega(r')\,dr' = A\chi(r,c), \quad |r|<c, \tag{2.8}$$

subject to the condition

$$\omega(r) = \begin{cases} 0, & |r|>c, \\ \text{finite}, & |r|<c. \end{cases} \tag{2.9}$$

Equation (2.8) is an integral equation of the convolution type, which can, in principle, be solved by first reducing it to a Riemann–Hilbert boundary value problem.[4] However, in view of the fact that $\omega(r)$ is band limited, we are prompted to search for solutions that exhibit this behavior. The next section outlines some of the known facts about the eigenvalues and eigenfunctions of the homogeneous integral equation

$$\lambda f(t) = \int_{-1}^{1} f(s)\,\frac{\sin c(t-s)}{\pi(t-s)}\,ds, \quad |t|<1, \tag{2.10}$$

via prolate spheroidal functions.

III. SOME KNOWN RESULTS

The theory of prolate spheroidal functions can be found in numerous books[5–8] and articles.[2,3,9,10] To achieve consistency we will adopt the notation of Flammer.[6]

The integral equation

$$\lambda f(t) = \int_{-1}^{1} f(s)\,\frac{\sin c(t-s)}{\pi(t-s)}\,ds, \quad |t|<1,$$
$$f(t) \in L_2(-1,1), \tag{3.1}$$

with a symmetric L_2 kernel has a denumerable set of eigenvalues,

$$1 > \lambda_0 > \lambda_1 > \cdots > 0,$$

to which each λ_i is an associated real-valued eigenfunction $f_i(t)$, which forms a complete set in $L_2(-1,1)$ and which satisfies the orthogonal condition

$$\int_{-1}^{1} f_i(t) f_j(t)\,dt = \lambda_i \delta_{ij} \tag{3.2}$$

and the eigenvalue problem

$$\lambda_i f_i(t) = \int_{-1}^{1} f_i(s)\,\frac{\sin c(t-s)}{\pi(t-s)}\,ds, \quad |t|<1. \tag{3.3}$$

The eigenfunctions $f_i(t)$ are also bounded continuous solutions of the eigenvalue problem

$$\left\{\frac{d}{dt}\left[(t^2-1)\frac{d}{dt}\right] + c^2t^2\right\} f_i(t) = \chi_i f_i(t),$$
$$-\infty < t < \infty, \tag{3.4}$$

with the eigenvalues denumerated as follows:

$0<\chi_0<\chi_1<\cdots$.

The eigenfunctions $f_i(t)$ are band limited and are expressible in terms of the angular prolate spheroidal functions of zeroth order, $S_{0n}(c,t)$, $n=0,1,2,3,\ldots$. These prolate spheroidal functions satisfy both Eqs. (3.3) and (3.4), are real for real t, are continuous functions also of c for $c\geqslant 0$, and can be extended to be entire functions of the complex variable t. The functions $S_{0n}(c,t)$ are orthogonal in the interval $-1<t<1$ and each function has exactly n zeros in the same interval. They are even or odd functions of t according as n is even or odd. Both the eigenvalues $\lambda_n(c)$ of the integral equation (3.3) and the eigenvalues $\chi_n(c)$ of the differential equation (3.4) are continuous functions of c.

In the limit as $c\to 0$, we have $\chi_n(0)=n(n+1)$, $n=0,1,2,\ldots$, so that Eq. (3.4) becomes the familiar Legendre differential equation, and consequently

$$S_{0n}(c,t)\to P_n(t) \quad \text{as } c\to 0.$$

The eigenvalues $\lambda_n(c)$, on the other hand, can be expressed in terms of the radial prolate spheroidal functions, $R_{0n}^{(1)}(c,t)$, which differ from the angular functions by a real scale factor.

The following asymptotic expansions can be found in Refs. 11 and 12:

$$\lambda_n(c)=(2c/\pi)[R_{0n}^{(1)}(c,1)]^2, \quad \text{any } c.$$

For fixed n and small c,

$$\lambda_n(c)=\frac{2}{\pi}\left[\frac{2^{2n}(n!)^3}{(2n)!(2n+1)!}\right]^2 c^{2n+1} \times\exp\left\{-\frac{(2n+1)c^2}{(2n-1)^2(2n+3)^2}[1+O(c^4)]\right\}. \tag{3.5}$$

For fixed n and large c,

$$1-\lambda_n(c)=\frac{2^{3n+2}\sqrt{\pi}c^{n+1/2}e^{-2c}}{n!} \times\left[1-\frac{6n^2-2n+3}{32c}+O\left(\frac{1}{c^2}\right)\right]. \tag{3.6}$$

Some values have been tabulated in Ref. 11.

We will also need the asymptotic expansion of $S_{0n}(c,t)$ for fixed n and small c (see Ref. 12).

$$S_{0n}(c,t)=P_n(t)+c^2\left[\frac{n(n-1)}{2(2n-1)^2(2n+1)}P_{n-2}(t) -\frac{(n+1)(n+2)}{2(2n+3)^3(2n+1)}P_{n+2}(t)\right]+O(c^4), \tag{3.7}$$

where the P_n's are the Legendre polynomials.

IV. SOLUTION OF EQ. (2.8) AND EQ. (2.2)

We can change Eq. (2.8) by the substitutions

$$r=ct, \quad r'=cs, \qquad A^{-1}\omega(r)=f(t), \quad \chi(r,c)=g(t), \tag{4.1}$$

to

$$f(t)-\int_{-1}^{1} f(s)\frac{\sin c(t-s)}{\pi(t-s)}ds=g(t), \quad |t|<1. \tag{4.2}$$

We now seek a solution $f(t)\in L_2(-1,1)$ of Eq. (4.2) such that $f(t)\in B$, the class of band-limited functions in $|t|<1$.

The integral operator

$$\mathbf{K}f=\int_{-1}^{1}\frac{\sin c(t-s)}{\pi(t-s)}f(s)\,ds \tag{4.3}$$

has a continuous and symmetric kernel for all $-1<t,s<1$ so that $\mathbf{K}$ is a self-adjoint compact operator in $L_2(-1,1)$ (see Ref. 13). Furthermore, for any $\phi\in L_2(-1,1)$, we have the inner product

$$\langle \mathbf{K}\phi,\phi\rangle\geqslant 0,$$

so that $\mathbf{K}$ is positive. This guarantees the existence of a denumerable set of real and non-negative eigenvalues $\{\lambda_n\}_{n=1}^{\infty}$ and a corresponding set of orthogonal eigenfunctions $\{S_{0n}(c,t)\}_{n=1}^{\infty}$ so that by an application of Hilbert–Schmidt theorem[14] we can easily derive the solution of Eq. (4.1) in terms of an absolutely and uniformly convergent series as follows:

$$f(t)=g(t)+\sum_{n=1}^{\infty}\left(\frac{\lambda_n}{1-\lambda_n}\right)\langle g(t'),f_n(t')\rangle f_n(t), \tag{4.4}$$

where the eigenfunctions will be chosen as

$$f_n(t)=\left[\frac{\lambda_n(c)}{u_n(c)}\right]^{1/2}S_{0n}(c,t), \qquad u_n^2(c)=\int_{-1}^{1}[S_{0n}(c,t)]^2\,dt. \tag{4.5}$$

The eigenfunctions $f_n(t)$ are scaled as such so that the normalization condition (3.2) is satisfied.

The inner product in Eq. (4.4) is defined as

$$\langle g(t'),f_n(t')\rangle=\int_{-1}^{1}g(t')f_n(t')\,dt'. \tag{4.6}$$

Observe from Eq. (2.8) that $g(t)=\chi(r,c)$ is an odd function of t and since $f_n(-t)=(-1)^n f_n(t)$, Eq. (4.6) becomes

$$\begin{aligned}\langle g(t'),f_n(t')\rangle &=\begin{cases}0, & n=\text{even},\\ 2\int_0^1 g(t')f_n(t')\,dt', & n=\text{odd},\end{cases}\\ &=\frac{2}{c}\int_{-1}^{1}\frac{\sin c(t-1)}{\pi(t-1)}f_n(t)\,dt\\ &=(2/c)\,\lambda_n(c)f_n(1), \quad n=\text{odd}.\end{aligned} \tag{4.7}$$

The second equality was obtained by substituting the definition of $g(t')$ and utilizing the parity of $f_n(t')$ while the last equality results from a direct application of Eq. (3.3).

The complete solution of the integral equation (4.2) is, therefore,

$$f(t)=g(t)+\sum_{n\,\text{odd}}\frac{2\lambda_n^2}{c(1-\lambda_n)}f_n(1)f_n(t), \quad |t|<1, \tag{4.8}$$

or, in terms of angular prolate spheroidal functions,

$$f(t)=g(t)+\sum_{n\,\mathrm{odd}}\frac{2\lambda_n^3}{c(1-\lambda_n)u_n}S_{0n}(c,1)S_{0n}(c,t),$$
$$|t|<1. \tag{4.9}$$

The solution above now can be expanded to any desired order of c. This completes the determination of the extra contribution $\omega(r)$.

Since the wave function vanishes inside the core, the integrodifferential equation (2.2) becomes

$$\left(\frac{d^2}{dr^2}+K^2\right)u(r)$$
$$=A\,[\delta(r-c)-\chi(r,c)]-\int_0^c \chi(r,r')\omega(r')\,dr',$$
$$r>c. \tag{4.10}$$

Since $u(0)=0$, this equation can be solved easily to give

$$u(r)=\frac{1}{K}\int_0^r F(s)\sin K(r-s)\,ds,\quad r>c, \tag{4.11a}$$

where

$$F(s)=A\,[\delta(s-c)-\chi(s,c)]-\int_0^c \chi(s,s')\omega(s')\,ds'. \tag{4.11b}$$

This completes the solution of the (dimensionless) s-wave Bethe–Goldstone equation (2.2) apart from the normalization constant A.

V. APPROXIMATE SOLUTION FOR SMALL CORE RADIUS

From Eqs. (4.1) and (4.9) we have

$$A^{-1}\omega(r)=\chi(r,c)+\sum_{n\,\mathrm{odd}}B_n'S_{0n}\left(c,\frac{r}{c}\right),\quad 0<r<c \tag{5.1}$$

where

$$B_n'=[2\lambda_n^3/c(1-\lambda_n)u_n]\,S_{0n}(c,1),$$

and we have restricted ourselves back to $0<r<c$.

The integral in Eq. (4.11b) then becomes

$$A\int_0^c \chi(s,s')\left[\chi(s',c)+\sum_{n\,\mathrm{odd}}B_n'S_{0n}\left(c,\frac{s'}{c}\right)\right]ds'. \tag{5.2}$$

By the symmetry property of $\chi(s,s')$ and the parity of $S_{0n}(c,s'/c)$, the second integral becomes

$$\int_0^c \chi(s,s')\,S_{0n}\left(c,\frac{s'}{c}\right)ds'=\lambda_nS_{0n}\left(c,\frac{s}{c}\right), \tag{5.3}$$

so that

$$A^{-1}F(s)=[\delta(s-c)-\chi(s,c)]$$
$$-\int_0^c \chi(s,s')\chi(s',c)\,ds'$$
$$-\sum_{n\,\mathrm{odd}}B_nS_{0n}\left(c,\frac{s}{c}\right), \tag{5.4}$$

where $B_n\equiv\lambda_nB_n'$.

From Eq. (4.11a), the wave function becomes

$$u(r)=\frac{1}{K}\int_0^r F(s)\sin K(r-s)\,ds$$
$$=(A/K)(\mathrm{I}-\mathrm{II}-\mathrm{III}), \tag{5.5}$$

with

$$\mathrm{I}=\int_0^r[\delta(s-c)-\chi(s,c)]\sin K(r-s)\,ds,$$
$$\mathrm{II}=\int_0^r\left[\int_0^c \chi(s,s')\chi(s',c)\,ds'\right]\sin K(r-s)\,ds, \tag{5.6}$$
$$\mathrm{III}=\sum_{n\,\mathrm{odd}}B_n\int_0^c S_{0n}\left(c,\frac{s}{c}\right)\sin K(r-s)\,ds.$$

For small c,

$$\chi(x,y)\sim 2xy/3\pi,\quad x,y<c,$$

so that II can be integrated approximately to give

$$\mathrm{II}\sim\frac{4c^4}{27\pi^3}\left(\frac{r}{K}-\frac{\sin kr}{K^2}\right).$$

The expression I is precisely the approximate solution given in Ref. 1. Their result is quoted in the Appendix.

The last integral III is approximately

$$\mathrm{III}\cong\sum_{n\,\mathrm{odd}}B_n\int_0^c P_n\left(\frac{s}{c}\right)\sin K(r-s)\,ds$$
$$=B_1\int_0^c P_1\left(\frac{s}{c}\right)\sin K(r-s)\,ds+\cdots$$
$$\sim O(c^{10}), \tag{5.7}$$

where we have used the known results in Sec. III. This means that the contribution of III is negligible and the major contribution to the wave function comes from I and II. This agrees with the result obtained in Ref. 1.

VI. NORMALIZATION CONSTANT A

For completeness we shall determine the normalization constant A. From Eq. (5.5) we have

$$u(r)=\frac{1}{K}\int_0^r F(s)\sin K(r-s)\,ds$$
$$=\frac{\sin Kr}{K}\int_0^r F(s)\cos Ks\,ds$$
$$-\cos\frac{Kr}{r}\int_0^r F(s)\sin Ks\,ds. \tag{6.1}$$

The far-field scattering requirement (2.4) demands that the second integral above must vanish as $r\to+\infty$, i.e.,

$$\int_0^\infty F(s)\sin Ks\,ds=0. \tag{6.2}$$

Now consider

$$A^{-1}\int_0^\infty F(s)\sin Ks\,ds$$
$$=\int_0^\infty\Big\{[\delta(s-c)-\chi(s,c)]$$
$$-\int_0^c \chi(s,s')\chi(s',c)\,ds'$$

$$- \sum_{n\,\text{odd}} B_n S_{0n}\left(c, \frac{s}{c}\right)\Big\} \sin Ks\, ds$$
$$= (\text{IV} - \text{V}), \qquad (6.3)$$

with

$$\text{IV} = \int_0^\infty [\delta(s-c) - \chi(s,c)] \sin Ks\, ds$$
$$\text{V} = \int_0^\infty \Big[\int_0^c \chi(s,s')\chi(s',c)\, ds' + \sum_{n\,\text{odd}} B_n S_{0n}\left(c, \frac{s}{c}\right)\Big] \sin Ks\, ds.$$

If we rewrite

$$\delta(s-c) - \chi(s,c) = \frac{2sc}{\pi}\int_1^\infty j_0(ts)\, j_0(tc) t^2\, dt, \qquad (6.4)$$

then IV becomes

$$\text{IV} = c\int_1^\infty t\, j_0(tc)\delta(K-t)\, dt = 0.$$

The last integral vanishes since t lies outside the Fermi sphere while K lies inside.[15]

For small c, the first term in V is of $O(c^4)$ while the second term is of $O(c^{10})$ so that V is essentially zero. This proves the condition (6.2).

Finally, the condition (2.4) implies that

$$\int_0^\infty F(s) \cos Ks\, ds = 1, \qquad (6.5)$$

so that the normalization constant for small c is

$$A^{-1} = \int_0^\infty \Big\{ [\delta(s-c) - \chi(s,c)] - \Big[\left(\frac{4c^4}{27\pi^2}\right) - \sum_{n\,\text{odd}} B_n S_{0n}\left(c, \frac{s}{c}\right)\Big]\Big\} \cos Ks\, ds. \qquad (6.6)$$

The first square-bracket term is precisely the result obtained in Ref. 1 (see the Appendix) while the second square-bracket term constitutes additional contributions.

VII. THE REACTION MATRIX

The most important quantity of interest is the diagonal elements of the reaction matrix, which gives the potential energy. It is given by

$$\langle \mathbf{k}|G|\mathbf{k}\rangle = \int d^3x\, v(x)\psi(\mathbf{x})e^{-i\mathbf{k}\cdot\mathbf{x}}$$
$$= \frac{4\pi}{M^* k_F}\int v(r)u(r)\frac{\sin Kr}{K}\, dr. \qquad (7.1)$$

Upon substituting $v(r)u(r)$ from Eq. (2.3) and using the result Eq. (5.1), we can show that

$$\langle \mathbf{k}|G|\mathbf{k}\rangle = \frac{4\pi A}{M^* k}\Big\{ \sin Kc + \int_0^c \Big[\chi(r,c) + \sum_{n\,\text{odd}} B'_n S_{0n}\left(c, \frac{r}{c}\right)\Big] \sin Kr\, dr\Big\}$$
$$\cong \frac{4\pi A}{M^* k}\Big\{ \sin Kc + \frac{2c}{3\pi K^2}(\sin Kc - K\cos Kc) + O(c^{10})\Big\},$$

for small c. (7.2)

The first term above is precisely the result obtained in Ref. 1.

APPENDIX

The approximate solution from Ref. 1 is

$$u(r) = \frac{1}{K}\int_0^c f(s) \sin K(r-s)\, ds, \qquad (\text{A1})$$

where

$$f(r) \cong A\delta(r-c) - A\chi(r,c).$$

From Eqs. (A1), we obtain

$$u(r) \cong B \sin Kr + C\cos Kr, \qquad (\text{A2})$$

$$B = \frac{A}{2\pi K}\Big\{[2\pi - Si(1+K)2c - Si(1-K)2c] \cos Kc + \Big[\ln\frac{1+K}{1-K} + Ci(1+K)2c - Ci(1-K)2c\Big]\sin Kc\Big\},$$

$$C = \frac{A}{2\pi K}\Big\{\Big[\ln\frac{1+K}{1-K} - Ci(1+K)2c + Ci(1-K)2c\Big]\cos Kc - [Si(1+k)2c + Si(1-K)2c]\sin Kc\Big\},$$

where

$$Si(z) = \int_0^z \frac{\sin t}{t}\, dt,$$
$$Ci(z) = \gamma + \ln z + \int_0^z \frac{\cos t - 1}{t}\, dt \quad (|\arg z| < \pi),$$

$\gamma = 0.577\,216$ is Euler's constant.

The normalization constant is determined by the condition (2.4) so that

$$A^{-1} = \Big[1 - \frac{c}{\pi}\left(2 - K\ln\frac{1+K}{1-K}\right) - \cdots\Big]. \qquad (\text{A3})$$

[1]H. A. Bethe and J. Goldstone, Proc. R. Soc. London Ser. A **238**, 551 (1957).
[2]D. Slepian and H. O. Pollak, Bell System Tech. J. **40**, 43 (1961).
[3]H. J. Landau and H. O. Pollak, Bell System Tech. J. **40**, 65 (1961).
[4]P. P. Zabreyko, A. I. Koshelev, M. A. Krasnosel'skii, S. G. Mikhlin, L. S. Rakovshchik, and W. Ya. Stet'senko, *Integral Equations—A Reference Text* (Noordhoff, Leyden, 1975).
[5]J. Meixner and F. W. Schafke, *Mathieusche Funktionen und Spharoidfunktionen* (Springer, Berlin, 1954).
[6]C. Flammer, *Spheroidal Wave Functions* (Stanford U.P., Stanford, CA, 1957).
[7]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).
[8]J. A. Stratton, P. M. Morse, L. J. Chu, J. D. C. Little, and F. J. Corbató, *Spheroidal Wave Functions* (Wiley, New York, 1956).
[9]W. H. J. Fuchs, J. Math. Anal. Appl. **9**, 317 (1964).
[10]H. J. Landau and H. O. Pollak, Bell System Tech. J. **41**, 1295 (1962).
[11]D. Slepian and E. Sonnenblick, Bell System Tech. J. **44**, 1745 (1965).
[12]D. Slepian, J. Math. Phys. (Cambridge, Mass.) **44**, 99 (1965).
[13]H. Hochstadt, *Integral Equations* (Wiley, New York, 1973).
[14]S. G. Mikhlin, *Integral Equations* (Pergamon, New York, 1957).
[15]A. L. Fetter and J. D.Walecka, *Quantum Theory of Many-Particle Systems* (McGraw-Hill, New York, 1971).

Strong amplification of sidebands in a strongly driven three-level atomic system. II. Classical description of the laser field[a)]

Constantine Mavroyannis
Division of Chemistry, National Research Council of Canada, Ottawa, Ontario, Canada K1A 0R6

(Received 1 August 1985; accepted for publication 26 November 1985)

The fluorescent spectra arising from the interaction of a three-level atom with a strong pump field and a weak signal field have been studied simultaneously. The atom consists of an upper excited state $|2\rangle$ and two lower ground states $|3\rangle$ and $|1\rangle$, which arise by removing the degeneracy of the ground state by applying internal or external fields. The laser field depletes the metastable state $|3\rangle$ by bringing the electrons into the excited state $|2\rangle$ from where the electrons emit photons and decay into the lower states through the transitions $|2\rangle \rightarrow |3\rangle$ and $|2\rangle \rightarrow |1\rangle$, which are described by the signal field. Using a classical description of the laser field, where in the model Hamiltonian the laser–atom interaction is treated classically while the free and interacting electron and signal fields are quantized, the decay process $|2\rangle \leftrightarrow |1\rangle$ of the signal field is considered by evaluating the appropriate Green's function of the system. The spectral function for the $|2\rangle \leftrightarrow |1\rangle$ transition of the signal field describes one-photon, three-photon, and two-photon Raman processes, respectively. The one-photon spectra consist of the main peak at the signal frequency and a pair of sidebands, which are symmetrically located from the position of the main peak. The intensity of the main peak is positive while that of the sidebands is negative indicating that the signal is attenuated and is amplified at the corresponding frequencies, respectively. The three-photon and two-photon Raman spectra are described by a doublet, respectively, whose intensities are always negative, implying amplification of the signal field. The computed spectra are presented graphically and compared with those derived in a recent study, where the laser field is quantized and photon–photon correlations are taken into consideration in the limit of high photon densities of the laser field. A detailed discussion of both treatments is given for the processes under investigation.

I. INTRODUCTION

Several experimental[1–6] and theoretical[7–10] papers have studied the physical process of the optical amplification of sidebands arising in atomic systems, which interact with strong pump fields and weak signal fields simultaneously. We have recently considered,[10] hereafter referred to as I, the excitation spectra of a three-level atom interacting simultaneously with a strong laser field and a weak signal field. The atomic system, depicted in Fig. 1, consists of an upper excited state $|2\rangle$ and two lower states $|1\rangle$ and $|3\rangle$, where the degeneracy of the ground state has been lifted by internal process or by applying an external electric or a magnetic field. The strong laser field depletes the metastable state $|3\rangle$ by bringing the electrons into the excited state $|2\rangle$, $|3\rangle \rightarrow |2\rangle$, from where the electrons follow simultaneously to decay processes $|2\rangle \rightarrow |3\rangle$ and $|2\rangle \rightarrow |1\rangle$, which are described by the signal field.

For the system shown in Fig. 1, which is the three-level analog of Mollow's two-level system,[7] the spectral functions for the $|1\rangle \leftrightarrow |2\rangle$ transitions of the signal field have been calculated in I describing the stimulated processes of the one-photon, three-photon, and two-photon Raman, respectively. It is found that the intensities of the sidebands for the processes in question take negative values indicating that strong amplification of the signal field occurs at the corresponding frequencies.[10]

The excitation spectra, which describe the stimulated one-, two- and three-photon processes for the system depicted in Fig. 1, have been calculated in I by making use of the Green's function formalism in the limit of high photon densities of the laser field. In I, a model Hamiltonian is used, where all the free and interacting fields, namely, the electron, the signal (vacuum), and the laser fields, respectively, are quantized. The present study is an extension of I by considering the same processes as in I but using a classical description of the laser field while the electron and signal fields remain quantized as before.

The classical description of the laser field is used in Sec. II to calculate the Green's functions for the $|1\rangle \leftrightarrow |2\rangle$ transition describing the one- and three-photon processes, respectively. The excitation spectra for the one-photon and three-photon processes are considered in Sec. III, where the spectral functions for the processes in question are derived and compared with those derived in I while the computed spectra are presented graphically and discussed. The spectral function for the two-photon Raman spectra is calculated

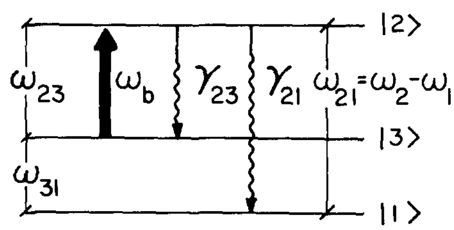

FIG. 1. Energy-level diagram of a three-level atom. The solid line indicates the laser field operating between the states $|2\rangle$ and $|3\rangle$, $|2\rangle \leftrightarrow |3\rangle$. Wiggly lines describe the radiative decays $|2\rangle \rightarrow |3\rangle$ and $|2\rangle \rightarrow |1\rangle$, respectively.

[a)] Issued as NRCC No. 25017.

 0022-2488/86/041159-05$02.50

in Sec. IV and compared with the corresponding ones obtained in I. A summary of the derived results is given in Sec. V.

II. CLASSICAL DESCRIPTION OF THE LASER FIELD

We shall calculate here the expression for the Green's function $G_{12,21}(\omega) = \langle\langle \alpha_1^\dagger \alpha_2; \alpha_2^\dagger \alpha_1 \rangle\rangle$ by using a classical description of the laser field, namely, the derivation of the classical counterpart of the expressions (29), (30), and (49) of I will be illustrated, which describe stimulated one-photon, three-photon, and two-photon processes, respectively. The Green's function $G_{12,21}(\omega)$ is defined by Eqs. (3)–(5) of I, where $\alpha_i^\dagger$ and α_i are the Fermi creation and annihilation operators describing the electron states $|i\rangle$ depicted in Fig. 1. To avoid repetitions, we adopt exactly the same system and notation as in I, hence, the reader is referred to I for details. Thus equations from I will be referred to I and will not be repeated here. Only new results and those differing from I will be quoted here.

To treat the laser field b classically, we consider the Hamiltonian (1) of I, where the term $\omega_b n_b$ is omitted (interaction representation for the laser field) while the fourth term describing the laser–atom interaction is replaced by

$$\tfrac{1}{2} i g_b (\alpha_3^\dagger \alpha_2 - \alpha_2^\dagger \alpha_3)(e^{i\omega_b t} + e^{-i\omega_b t}), \tag{1}$$

where g_b is the corresponding classical counterpart for the Rabi frequency Ω_b defined by Eq. (17) of I. Using this Hamiltonian, we derive the equation of motion for the Green's function $G_{12,21}(\omega)$ as

$$G_{12,21}(\omega) = [g_b^2/4d_{12}^2(\omega)][G_{13b}^c(\omega) + G_{13b^\dagger}^c(\omega)], \tag{2}$$

where

$$G_{13b}^c(\omega) = \langle\langle \alpha_1^\dagger \alpha_3 e^{i\omega_b t}; \alpha_3^\dagger \alpha_1 (e^{i\omega_b t'} + e^{-i\omega_b t'}) \rangle\rangle, \tag{3}$$

$$G_{13b^\dagger}^c(\omega) = \langle\langle \alpha_1^\dagger \alpha_3 e^{-i\omega_b t}; \alpha_3^\dagger \alpha_1 (e^{i\omega_b t'} + e^{-i\omega_b t'}) \rangle\rangle. \tag{4}$$

The Green's functions $G_{13b}^c(\omega)$ and $G_{13b^\dagger}^c(\omega)$ are the classical counterparts with respect to the laser field b to those of $G_{13b,31(b+b^\dagger)}(\omega)$ and $G_{13b^\dagger,31(b+b^\dagger)}(\omega)$ defined by Eqs. (15) and (16) of I, respectively.

To proceed further, we calculate the Green's functions $G_{13b}^c(\omega)$ and $G_{13b^\dagger}(\omega)$ in the rotating wave approximation (RWA) as has been done for $G_{13b,31(b+b^\dagger)}(\omega)$ and $G_{13b^\dagger,31(b+b^\dagger)}(\omega)$ in I to derive the result

$$G_{13b}^{c,\mathrm{RWA}}(\omega) = -\frac{(\bar{n}_3 - \bar{n}_1) d_{12}(\omega)}{2\pi[d_{12}(\omega) d_{13b}(\omega) - g_b^2/4]}, \tag{5}$$

$$G_{13b^\dagger}^{c,\mathrm{RWA}}(\omega) = -\frac{(\bar{n}_3 - \bar{n}_1) d_{12b^\dagger b^\dagger}(\omega)}{2\pi[d_{13b^\dagger}(\omega) d_{12b^\dagger b^\dagger}(\omega) - g_b^2/4]}, \tag{6}$$

where the propagators $d_{12}(\omega)$, $d_{13b}(\omega)$, $d_{13b^\dagger}(\omega)$, and $d_{12b^\dagger b^\dagger}(\omega)$ are defined by Eqs. (9) and (23)–(25) of I, respectively. Substituting Eqs. (5) and (6) into Eq. (2), we obtain

$$G_{12,21}(\omega) = \Pi_{12,21}^c(\omega) + S_{12,21}^c(\omega), \tag{7}$$

where

$$\Pi_{12,21}^c(\omega) = \frac{(\bar{n}_3 - \bar{n}_1)}{2\pi} \times \left[\frac{1}{d_{12}(\omega)} - \frac{d_{13b}(\omega)}{d_{12}(\omega) d_{13b}(\omega) - g_b^2/4}\right], \tag{8}$$

$$S_{12,21}^c(\omega) = -\frac{(\bar{n}_3 - \bar{n}_1) g_b^2 d_{12b^\dagger b^\dagger}(\omega)}{8\pi d_{12}^2(\omega)[d_{13b^\dagger}(\omega) d_{12b^\dagger b^\dagger}(\omega) - g_b^2/4]}. \tag{9}$$

Comparison between Eqs. (5) and (6) with the corresponding Eqs. (21) and (22) of I implies that Eqs. (5) and (6) are similar to those of (21) and (22) of I but their denominators differ by the numerical factor of $\frac{1}{2}$ provided that the correspondence principle is applicable, namely, when the interchange $\Omega_b \rightleftarrows g_b$ is considered. Thus only when the interchange $g_b^2 \rightleftarrows 2\Omega_b^2$ holds, then the classical expressions (5) and (6) become identical to Eqs. (21) and (22) of I, respectively, which are derived when the laser field is quantized and photon correlations are taken into account in the limit of high photon densities. As has been discussed elsewhere,[9–13] the factor of 2 arises when use is made of the decoupling approximation of the form $\beta_b^\dagger \beta_b \beta_b \to 2\langle \beta_b^\dagger \beta_b \rangle \beta_b$, which is applicable in the limit of high photon densities $\bar{n}_b \gg 1$ and describes the Bose character of the photons when photon–photon correlations are considered. Here, $\beta_b^\dagger$ and β_b are photon creation and annihilation operators of the laser field, which obey Bose statistics and $\bar{n}_b = \langle \beta_b^\dagger \beta_b \rangle$ is the average value of the photon number density operator.

The expression (8) for $\Pi_{12,21}^c(\omega)$ describes the one-photon process near the frequency of the signal field $\omega \approx \omega_{21}$ and it is larger by a factor of 2 than the corresponding expression given by Eq. (29) of I. The expression (9) for $S_{12,21}^c(\omega)$ is the classical counterpart of Eq. (30) of I and describes the three-photon processes near the frequency $\omega \approx \omega_{21} - 2\omega_b$. In the limit when $g_b \to 0$, the functions $\Pi_{12,21}^c(\omega) \to 0$ and $S_{12,21}^c(\omega) \to 0$ and, hence, the functions $\Pi_{12,21}^c(\omega)$ and $S_{12,21}^c(\omega)$, by analogy with the functions $\Pi_{12,21}(\omega)$, and $S_{12,21}(\omega)$ of I, behave as describing stimulated one-photon and three-photon processes, respectively.

III. EXCITATION SPECTRA

To discuss the excitation spectra, we introduce the notation

$$X = (\omega - \omega_{21})/\gamma_+^0 \qquad Y = (\omega - \omega_{21} + 2\omega_b)/\gamma_+^0, \tag{10a}$$

$$\nu_b = (\omega_{23} - \omega_b)/\gamma_+^0, \quad \xi_b = g_b/\gamma_+^0, \quad \xi^2 = \tfrac{1}{4}(\xi_b^2 + \nu_b^2), \tag{10b}$$

where X and Y are the reduced detuning frequencies of the signal field and of the three-photon process, respectively, ν_b and ξ_b are the relative detuning and the relative classical Rabi frequency, respectively, while ξ defines the total relative shift arising from the classical Rabi frequency and the detuning of the laser field. In Eqs. (10a) and (10b), $2\gamma_+^0 = \gamma_{21}^0 + \gamma_{23}^0$, where γ_{21}^0 and γ_{23}^0 are the spontaneous transition probabilities defined by Eq. (12) of I. Substituting Eqs. (10a) and (10b) into Eqs. (8) and (9) we have

$$\Pi^c_{12,21}(\omega) = \frac{(\bar n_3 - \bar n_1)}{4\pi\gamma^0_+}\left[\frac{2}{X+i} - \frac{1+\nu_b/2\xi - i/2\xi}{X+\frac{1}{2}\nu_b - \xi + i/2} - \frac{1-\nu_b/2\xi + i/2\xi}{X+\frac{1}{2}\nu_b + \xi + i/2}\right], \tag{11a}$$

$$S^c_{12,21}(\omega) = -\frac{(\bar n_3 - \bar n_1)g_b^2}{16\pi\gamma^0_+ d^2_{12}(\omega)}\left[\frac{1+\nu_b/2\xi - i/2\xi}{Y+\frac{1}{2}\nu_b - \xi + i/2} + \frac{1-\nu_b/2\xi + i/2\xi}{Y+\frac{1}{2}\nu_b + \xi + i/2}\right]. \tag{11b}$$

The excitation spectra are described by the spectral function determined by the imaginary part of the Green's function $G_{12,21}(\omega)$, which is equal to

$$\begin{aligned} P^c_{12}(\omega) &\equiv -2\,\mathrm{Im}\,G_{12,21}(\omega) \\ &= -2\,\mathrm{Im}[\Pi^c_{12,21}(\omega) + S^c_{12,21}(\omega)] \\ &= \frac{(\bar n_3 - \bar n_1)}{2\pi\gamma^0_+}\left[F^c_{12}(X) + \frac{g_b^2}{4d^2_{12}(\omega)}J^c_{12}(Y)\right], \end{aligned} \tag{12}$$

where

$$F^c_{12}(X) = 2/(x^2+1) - L^c(X,\nu_b,\xi) - L^c(X,\nu_b - \xi), \tag{13}$$

$$J^c_{12}(Y) = -L^c(Y,\nu_b,\xi) - L^c(Y,\nu_b,-\xi), \tag{14}$$

and the shape function $L^c(\lambda,\nu_b \pm \xi)$ is defined as

$$L^c(\lambda,\nu_b \pm \xi) = \frac{\frac{1}{2}(1 \pm \nu_b/2\xi) \pm (\lambda + \frac{1}{2}\nu_b \mp \xi)/2\xi}{(\lambda + \frac{1}{2}\nu_b \mp \xi)^2 + \frac{1}{4}}. \tag{15}$$

The expression $P^c_{12}(\omega)$ given by Eq. (12) is the classical counterpart of Eq. (36) of I and defines the absorption coefficient of the signal field describing the physical processes near the frequencies $\omega \approx \omega_{21}$ $(X \approx 0)$ and $\omega \approx \omega_{21} - 2\omega_b$ $(Y \approx 0)$, respectively. The shape functions $F^c_{12}(X)$ and $J^c_{12}(Y)$ given by Eqs. (13) and (14), respectively, are the classical counterparts of the corresponding expressions determined by Eqs. (37) and (38) of I.

Comparison between Eq. (13) for $F^c_{12}(X)$ and Eq. (37) of I for $F_{12}(X)$ implies that, apart from the difference in the expressions between η and ξ, the function $F^c_{12}(X)$ is larger by a factor of 2 than that of $F_{12}(X)$. We define by Eq. (32) of I, namely,

$$\eta^2 = \eta_b^2/2 + \nu_b^2/4, \tag{16}$$

and describes the total relative frequency shift arising from the Rabi frequency and the detuning while the corresponding one for ξ is defined by Eq. (10b). As an illustration, the functions $F_{12}(X)$ and $F^c_{12}(X)$ denoted as relative intensities are plotted in Fig. 2 versus the relative frequency X for the constant value of $\eta_b = \xi_b = 10$ and different values of detuning ν_b. The computed spectra described by the functions $F_{12}(X)$ and $F^c_{12}(X)$ are depicted in Fig. 2 by solid and dashed lines, respectively; the spectra for $F_{12}(X)$ (solid lines) are identical to those given in Fig. 2 of I. The spectra in Fig. 2 consist of the main peak at the frequency $X = 0$ of the signal field and a pair of sidebands that are peaked at $X = -\frac{1}{2}\nu_b \pm \eta$ (solid lines) and at $X = -\frac{1}{2}\nu_b \pm \xi$ (dashed lines), respectively. The frequency shifts between the positions of the sidebands described by $F_{12}(X)$ (solid lines) and those by $F^c_{12}(X)$ (dashed lines) are equal to $\pm(\eta - \xi)$, while the corresponding maximum relative intensities are determined by the expressions $i_\pm = (1 \pm \nu_b/2\eta)$ and $i^c_\pm = 2(1 \pm \nu_b/2\xi)$, respectively.

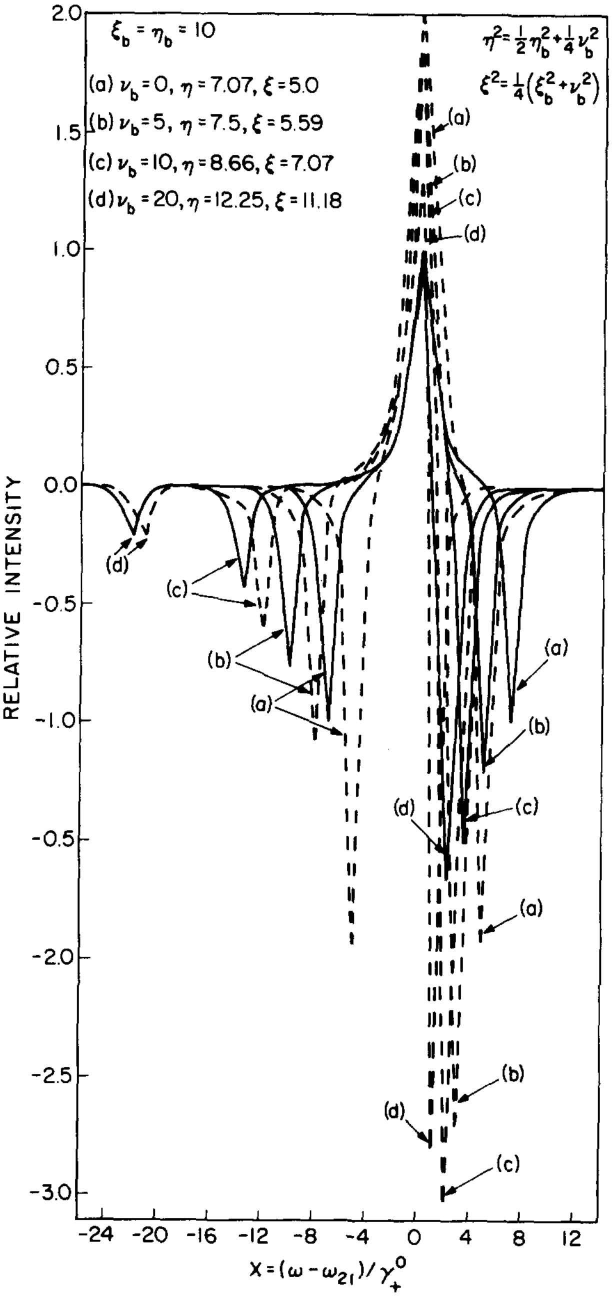

FIG. 2. One-photon spectra. The relative intensities $F_{12}(X)$ and $F^c_{12}(X)$ are computed from the rhs of Eq. (37) of I and Eq. (13), respectively, and are plotted versus the relative frequency $X = (\omega - \omega_{21})/\gamma^0_+$ for the relative Rabi frequency $\eta_b = 10$ and its classical counterpart $\xi_b = 10$ and various detunings. (a) $\nu_b = 0$, (b) $\nu_b = 5$, (c) $\nu_b = 10$, and (d) $\nu_b = 20$. Solid and dashed lines denote the spectra described by the shape functions $F_{12}(X)$ and its classical counterpart $F^c_{12}(X)$, respectively.

The spectral function $J^c_{12}(Y)$ defined by Eq. (14) for the three-photon process describes a pair of sidebands that are peaked at the frequencies $Y = -\frac{1}{2}\nu_b \pm \xi$, respectively. The frequency shifts between the peaks of the doublets described by the functions $J_{12}(Y)$ and $J^c_{12}(Y)$ are equal to $\pm(\eta - \xi)$, while the corresponding maximum relative intensities of the doublets are determined by the expressions $i_\pm = -2(1 \pm \nu_b/2\eta)$ and $i^c_\pm = -2(1 + \nu_b/2\xi)$, respectively. Hence, the spectra described by the function $J^c_{12}(Y)$ may be obtained from those computed from Eq. (38) of I and depicted in Fig. 4 of I if the positions of the peaks are shifted by $\pm(\eta - \xi)$ while their corresponding intensities by $\pm \nu_b(1/\eta - 1/\xi)$, respectively.

IV. TWO-PHOTON RAMAN SPECTRA

To consider the excitation spectrum near the Raman frequency $\omega = \omega_{31} \approx \omega_{21} - \omega_b$, we have to decouple the Green's functions that appear on the right-hand side (rhs) of Eq. (2) as has been done in I. In order to apply the decoupling approximation defined by Eq. (43) of I, we notice that the classical equivalent of the field-theoretical quantum mechanical expression quoted in I, namely,

$$\langle(\beta_b + \beta_b^\dagger)(\beta_b^\dagger + \beta_b)\rangle \approx 1 + 2\bar{n}_b \approx 2\bar{n}_b, \tag{17}$$

where $\bar{n}_b = \langle\beta_b^\dagger\beta_b\rangle \gg 1$ is the average value of the photon number operator of the laser field, is given by[13]

$$\langle(e^{i\omega_b t} + e^{-i\omega_b t})(e^{-i\omega_b t'} + e^{i\omega_b t'})\rangle_{t=t'} \approx 2. \tag{18}$$

Then using Eq. (18) and following the analogous procedure, which is used to derive Eq. (45) of I from Eq. (14) of I, we obtain

$$G_{12,21}(\omega) = \widetilde{G}_{12,21}(\omega) + R^c_{12}(\omega), \tag{19}$$

where

$$R^c_{12}(\omega) = [g_b^2/2d^2_{12}(\omega)]G_{13,31}(\omega), \tag{20}$$

$$G_{13,31}(\omega) = \langle\langle\alpha_1^\dagger\alpha_3;\alpha_3^\dagger\alpha_1\rangle\rangle. \tag{21}$$

The Green's function $\widetilde{G}_{12,21}(\omega)$ describes excitations near the frequencies $\omega \approx \pm\omega_b$ and will not be considered here, while the Green's function $G_{13,31}(\omega)$ describes excitations near the required Raman frequency $\omega \approx \omega_{31} \approx \omega_{21} - \omega_b$; Eq. (20) is the classical counterpart of Eq. (46) of I. Following I, we proceed to calculate the Green's function $G_{13,31}(\omega)$ in the RWA by making use of the Hamiltonian (1) of I but with the laser–atom interaction defined by Eq. (1). The result is

$$G^{\rm RWA}_{13,31}(\omega) = -\frac{(\bar{n}_3 - \bar{n}_1)d_{12b^\dagger}(\omega)}{2\pi[d_{13}(\omega)d_{12b^\dagger} - g_b^2/4]}, \tag{22}$$

where the propagators $d_{13}(\omega)$ and $d_{12b^\dagger}(\omega)$ are given by Eqs. (50) and (51) of I, respectively.

Comparison between Eq. (22) and Eq. (49) of I implies that Eq. (22) is similar but larger by a factor of 2 than the last term on the rhs of Eq. (49) of I, while the last term in the denominator of Eq. (22) is smaller by a factor of 2 than the corresponding one in Eq. (49) of I provided that the correspondence principle is applicable, namely, when $g_b \rightleftarrows \Omega_b$. The first term on the rhs of Eq. (49) of I, which is absent in the classical expression (22), describes the absence of spontaneous emission for atoms having a common upper and two different lower levels,[14–17] and it is quantum mechanical in nature. In the absence of both the quantized field and the classical field, namely, in the limit when $\Omega_b \to 0$ and $g_b \to 0$, then Eq. (22) and Eq. (49) of I become identical. This proves that the existing differences between Eq. (22) and Eq. (49) of I for $g_b \neq 0$ and $\Omega_b \neq 0$ are due entirely to the classical and to the quantized nature of the corresponding fields, respectively. A similar conclusion has been recently derived in the study[13] for the two-photon Raman spectra of a four-level atomic system interacting with a strong bichromatic field.

If we introduce the relative frequency Z as

$$Z = (\omega - \omega_{31})/\gamma^0_+ \tag{23}$$

and making use of Eqs. (20)–(22) and (10b), we have

$$R^c_{12}(\omega) = -\frac{(\bar{n}_3 - \bar{n}_1)g_b^2}{8\pi\gamma^0_+ d^2_{12}(\omega)}\left[\frac{1 - \nu_b/2\xi + i/2\xi}{Z - \frac{1}{2}\nu_b - \xi + i/2} + \frac{1 + \nu_b/2\xi - i/2\xi}{Z - \frac{1}{2}\nu_b + \xi + i/2}\right], \tag{24}$$

whose imaginary part is determined by

$$-2\,{\rm Im}\,R^c_{12}(\omega) = \frac{(\bar{n}_3 - \bar{n}_1)g_b^2}{4\pi\gamma^0_+ d^2_{12}(\omega)} I^c_{12}(Z), \tag{25}$$

where

$$I^c_{12}(Z) = -\frac{\frac{1}{2}(1 - \nu_b/2\xi) - (Z - \frac{1}{2}\nu_b - \xi)/2\xi}{(Z - \frac{1}{2}\nu_b - \xi)^2 + \frac{1}{4}} - \frac{\frac{1}{2}(1 + \nu_b/2\xi) + (Z - \frac{1}{2}\nu_b + \xi)/2\xi}{(Z - \frac{1}{2}\nu_b + \xi)^2 + \frac{1}{4}}. \tag{26}$$

The expressions (24)–(26) are the classical counterparts of Eqs. (53)–(55) of I, respectively. The shape function $I^c_{12}(Z)$ describes the two-photon Raman spectra near the frequency $Z \approx 0$ when the laser field is treated classically. The spectra consist of two sidebands that are peaked at the frequencies $Z = \frac{1}{2}\nu_b \pm \xi$ and have radiative widths of the order $\frac{1}{2}\gamma^0_+$. The intensities of the doublet take negative values, which indicate that amplification of the signal field is anticipated to take place at the frequency $Z = \frac{1}{2}\nu_b \pm \xi$, and their maximum values $I^c_\pm$ are determined by

$$I^c_\pm = -2(1 \mp \nu_b/2\xi), \tag{27}$$

which, apart from the differences between η and ξ, are larger by a factor of 2 (in absolute values) than the corresponding ones given by Eq. (36) of I. Hence, Eq. (26) for $I^c_{12}(Z)$ describes spectra similar to those described by Eq. (55) of I for $I_{12}(Z)$, which are depicted in Fig. 6 of I. The frequency shifts between the peaks of the doublets described by the functions $I_{12}(Z)$ (Fig. 6 of I) and $I^c_{12}(Z)$ are equal to $\pm(\eta - \xi)$ while the maximum relative intensities of the corresponding peaks are given by Eq. (56) of I and Eq. (27), respectively. Thus, the spectra described by $I^c_{12}(Z)$ can be easily obtained from those depicted in Fig. 6 of I for $I_{12}(Z)$ and they will not be shown here.

V. DISCUSSION

Using a classical description of the laser field, we have considered one-, three-, and two-photon Raman spectra arising from the $|1\rangle \leftrightarrow |2\rangle$ transition for the atomic system shown in Fig. 1. The calculation is for the same system considered in I, the only difference being that the laser–atom interaction is treated here classically.

The spectral function for the one-photon process is defined by the first term on the rhs of Eq. (12), where the function $F^c_{12}(X)$ as the relative intensity is determined by Eq. (13); $F^c_{12}(X)$ is the classical counterpart of the function $F_{12}(X)$ defined by Eq. (36) of I. Numerical results computed from the functions $F_{12}(X)$, Eq. (36) of I, and $F^c_{12}(X)$, Eq. (13), are presented graphically in Fig. 2 by solid and dashed lines, respectively. Resonance ($\nu_b = 0$) and off-resonance ($\nu_b \neq 0$) spectra are depicted in Fig. 2 for constant values of η_b and ξ_b, namely, for $\eta_b = \xi_b = 10$. The spectra in Fig. 2 describe (i) the main peak at the frequency $X = 0$ of the signal field, whose relative intensity is always positive and (ii) one pair of sidebands peaked at the frequencies $X = -\frac{1}{2}\nu_b \pm \eta$ (solid lines) and $X = -\frac{1}{2}\nu_b \pm \xi$ (dashed lines), respectively, whose relative intensities are always negative. The frequency shifts between the positions of the sidebands described by the function $F_{12}(X)$ (solid lines) and $F^c_{12}(X)$ (dashed lines) are equal to $\pm(\eta - \xi)$ while the corresponding maximum intensities are given by $i_\pm = -(1 \pm \nu_b/2\eta)$ and $i^c_\pm = -2(1 \pm \nu_b/2\xi)$, respectively. Figure 2a indicates that at resonance ($\nu_b = 0$) the maximum intensities of the peaks described by the classical expression $F^c_{12}(X)$ (dashed lines) are larger by a factor of 2 than those described by the function $F_{12}(X)$ (solid lines). However, in both treatments, the signal field is attentuated at the frequency $X = 0$ while it is strongly amplified at the frequencies $X = -\frac{1}{2}\nu_b \pm \eta$ and $X = -\frac{1}{2}\nu_b \pm \xi$, respectively.

The second term on the rhs of Eq. (12) describes three-photon processes near the frequency $Y \approx 0$ or $\omega \approx \omega_{21} - 2\omega_b$, where one photon of the signal field is absorbed while two photons of the laser field are emitted simultaneously. The shape function $J^c_{12}(Y)$ is determined by Eq. (14) and describes a doublet peaked at the frequencies $Y = -\frac{1}{2}\nu_b \pm \xi$ with maximum intensities equal to $i^c_\pm = -2(1 \pm \nu_b/2\xi)$, respectively, and spectral widths of the order of $\frac{1}{2}\gamma^0_+$. The shape function $J^c_{12}(Y)$ is the classical counterpart of the function $J_{12}(Y)$ given by Eq. (38) of I and they differ only as far as the definitions of ξ and η are concerned, which are determined by Eqs. (10b) and (16), respectively. Hence, the spectra described by the function $J^c_{12}(Y)$ can be obtained from those computed from $J_{12}(Y)$, which are shown in Fig. 4 of I, by appropriately changing the positions and the intensities of the doublet. Since the intensities of the doublet are always negative, the signal field is expected to be attenuated at the corresponding frequencies $Y = -\frac{1}{2}\nu_b \pm \xi$.

The spectral function $-2\,\mathrm{Im}\,R^c_{12}(\omega)$ describing the two-photon Raman process is determined by Eq. (25), which is the classical counterpart of the function $-2\,\mathrm{Im}\,R_{12}(\omega)$ defined by Eq. (54) of I. The expression (54) of I consists of two terms, one of which is the delta function $\delta(Z)$, while the other is the shape function $I_{12}(Z)$ given by Eq. (55) of I. The expression (25) consists only of the shape function $I^c_{12}(Z)$ defined by Eq. (26), which is the classical counterpart of Eq. (55) of I. The function $I^c_{12}(Z)$ describes a doublet peaked at the frequencies $Z = \frac{1}{2}\nu_b \pm \xi$ with maximum intensities equal to $I^c_\pm = -2(1 \mp \nu_b/2\xi)$, which are, apart from the different definitions of ξ and η, larger by a factor of 2 than the corresponding ones defined by Eq. (56) of I, namely, $I_\pm = -(1 \mp \nu_b/2\eta)$. Hence, the spectra described by the function $I^c_{12}(Z)$ can be easily obtained from those described by the function $I_{12}(Z)$, which are depicted in Fig. 6 of I. The presence and the absence of the delta function $\delta(Z)$ in Eq. (54) of I and Eq. (25), respectively, are due to the quantum nature of the quantized field and are attributed to the different way by which the quantized field and the classical field split the excitation spectrum. This is deduced from the comparison between the expressions for the Green's function $G_{13,31}(\omega)$ determined by Eq. (49) of I and its classical counterpart given by Eq. (22). Both expressions become identical only in the absence of the laser field, namely, when $\Omega_b \to 0$ and $g_b \to 0$.

The results of the present study, which are the classical counterparts of those derived in I, are consistent with those derived in recent studies,[9–13] where comparison has been made between the results obtained when the laser fields are quantized and photon–photon correlations are taken into account in the limit of high photon densities of the laser fields and when the laser fields are treated classically. As has been discussed[9–13] extensively before, both treatments have their own merits and should be used whenever they are appropriate. However, only experimental observations will reveal with certainty which of the two treatments provides a more appropriate description of the problem under investigation.

[1] B. Senitzky, G. Gould, and S. Cutler, Phys. Rev. **130**, 1460 (1963).
[2] B. Senitzky and S. Cutler, Microwave J. **1**, 62 (1964).
[3] A. M. Bonch-Bruevich, V. A. Khodovoi, and N. A. Chigir, Zh. Esp. Teor. Fiz. **67**, 2069 (1974) [Sov. Phys. JETP **40**, 1027 (1975)].
[4] H. M. Gibbs, S. L. McCall, and T. N. C. Venkatesam, Phys. Rev. Lett. **36**, 1135 (1976).
[5] A. C. Tam, Phys. Rev. A **19**, 1971 (1979).
[6] F. Y. Wu, S. Ezekiel, M. Ducloy, and B. R. Mollow, Phys. Rev. Lett. **38**, 1077 (1977).
[7] B. R. Mollow, Phys. Rev. A **5**, 2217 (1972).
[8] H. W. Galbraith, M. Dubs, and J. I. Steinfeld, Phys. Rev. A **26**, 1528 (1982).
[9] C. Mavroyannis, Phys. Rev. A **31**, 1563 (1985).
[10] C. Mavroyannis, J. Math. Phys. **26**, 1093 (1985).
[11] C. Mavroyannis, Phys. Rev. A **27**, 1414 (1983).
[12] C. Mavroyannis, J. Math. Phys. **25**, 2780 (1984).
[13] C. Mavroyannis, in *Advances in Chemical Reaction Dynamics, Proceedings of the NATO–ASI*, Iraklion, Crete, Greece, 25 August–7 September 1985, edited by P. M. Rentzepis and C. Capellos (Reidel, Boston, 1986).
[14] G. Breit, Rev. Mod. Phys. **5**, 91 (1933); P. W. Milonni, Phys. Rep. **25**, 53 (1976); A. Schenzle and R. G. Brewer, in *Laser Spectroscopy*, edited by S. Haroche, J. C. Pebay-Peyroula, T. W. Hänsch, and S. E. Harris (Springer, New York, 1975), p. 420.
[15] W. W. Chow, M. O. Scully, and J. O. Stoner, Phys. Rev. A **11**, 1380 (1975).
[16] R. M. Herman, H. Grotch, R. Kornblith, and J. H. Eberly, Phys. Rev. A **11**, 1389 (1975).
[17] C. Mavroyannis, Can. J. Phys. **58**, 957 (1980).

On inverse scattering in an elastic medium with vertical inhomogeneities

Philip M. Carrion
Geophysics Program, Henry Krumb School of Mines, Columbia University, New York, New York 10027

(Received 13 June 1984; accepted for publication 4 December 1985)

This paper treats the nonlinear inverse scattering problem for an elastic horizontally stratified medium with vertical inhomogeneities. It is shown that the vector wave equation can be presented as a set of two scalar wave equations; the first scalar equation describes the propagation of normally incident compressional waves (it is assumed that we deal with compressional incident plane waves impinging on an elastic half space). Using any inverse scattering technique (based, for example, on the Gelfand–Levitan theorem) applied to this equation, the acoustic impedance can be uniquely recovered from the recorded impulse response. The second equation is valid only in the region where incident compressional waves decay exponentially (evanescent waves) and propagate with complex angles, whereas mode-converted shear waves propagate with real angles. The main contribution in this region in the high-frequency approximation comes from mode-converted shear waves. We derived a new equation which describes the propagation of mode-converted shear waves in this region. Applying any inverse scattering technique to this equation, the rigidity modulus and the density can be removed separately. Knowing the acoustic impedance, the rigidity modulus, and the density, all Lame's parameters and the density can be recovered separately.

I. INTRODUCTION

The interpretation of seismic data requires a solution to the inverse scattering problem. The one-dimensional inverse scattering problem in the acoustic framework has been studied over the years. A comprehensive review of one-dimensional inverse scattering techniques was given by Newton.[1]

Recently several methods have been presented that treat the inverse scattering in elastic media with vertical inhomogeneities. Coen[2] was probably one of the first who considered the problem in an elastic horizontally stratified medium and recovered separately the elastic parameters. His method is based on separate experiments with vertically incident compressional and shear waves. In that case the problem is simplified since at normal incidence compressional and shear waves are uncoupled (there is no interconversion between P-waves and SV-waves at normal incidence). Aminzadeh,[3] Shiva and Mendel,[4] Clarke,[5] and Yagle and Levy[6] developed methods for estimation of all Lame's parameters and the density based on a layer-stripping technique. Recently Stickler[7] presented a now approach using the "trace formula." A method for determination of material density, compressional velocity based on a solution of a matrix Riccati equation was presented by Carazzone.[8] Meadows and Coen[9] considered exact and approximate algorithms for inversion of plane-layered isotropic and anisotropic elastic media.

All the methods mentioned above require separate experiments with incident compressional and shear waves. The primary concern of this paper is to develop a new technique that will provide a solution to the inverse scattering problem in elastic media with vertical inhomogeneities when only experiments with plane wave compressional waves are available. For example, in a marine environment when we try to estimate material properties of subbottom sediments the only sources available in the water are sources that generate only compressional waves.

In this paper we give an exact, formal answer giving the circumstances under which the inverse scattering problem in elastic stratified media can be solved without using shear-wave experiments.

We assume that the source that generates only compressional waves is placed in a liquid layer that covers an elastic horizontally homogeneous half space. We also assume that the time function of the source (source wavelet) is a Dirac's delta function. Incident compressional waves can be reflected, scattered, or mode converted in the elastic half space. The reflection response from the elastic half space is measured in the liquid layer in terms of negative stress (pressure) or the vertical component of particle velocity.

We propose a two-step process: The first step is the computation of the acoustic impedance from the wave equation that describes the experiment with compressional waves at normal incidence. The second step involves the derivation of a new equation that governs the propagation of evanescent compressional waves at complex angles of propagation and mode-converted shear waves at real angles of propagation. Under the assumptions of geometrical optics, evanescent waves propagating with complex angles decay very fast. Therefore in the region of rapidly decaying compressional waves that propagate with real angles (we certainly assume that such a region exists), only mode-converted shear waves occur. After solving the inverse scattering problem for this region using the derived equation, the rigidity modulus and the density can be computed separately. This allows for estimation of Lame's parameters as well as the density.

II. DECOMPOSITION OF THE VECTOR ELASTIC WAVE EQUATION

Let us consider the elastic wave equation, which can be written as

 0022-2488/86/041164-05$02.50

$$\nabla_i \cdot T_{ij} = \rho \frac{\partial^2 u_j}{\partial t^2}, \tag{2.1}$$

where u_j is the 3-D displacement vector, ρ is the density, and T_{ij} is a nine component stress tensor. In the case of a medium with vertical inhomogeneities, Eq. (2.1) can be reduced to

$$\mu(z)\nabla^2 u + [\lambda(z) + \mu(z)]\nabla\nabla\cdot u + \frac{\partial\lambda(z)}{\partial z}(\nabla\cdot u)e_z + \frac{\partial\mu(z)}{\partial z}\left[2\frac{\partial u}{\partial z} + e_z \times \nabla \times u\right] = \rho(z)\frac{\partial^2 u}{\partial t^2}, \tag{2.2}$$

where ρ, λ, μ are defined for $0 < z < \infty$ and take values in R; λ and μ are Lame's parameters; and e_z is a unit vector directed along the z axis.

Let us consider the cylindrical coordinate system and introduce a Hankel transform of order σ:

$$f_\sigma(k_r,\omega) = \int_0^\infty f(r,\omega)J_\sigma(k_r r)r\,dr = H_\sigma(f), \tag{2.3}$$

where $J_\sigma(\cdot)$ is a Bessel function of order σ and k_r is the radial wave number. Applying a Laplace transform to (2.1) and setting $s = i\omega$, where s is the argument of the Laplace transform, the following equation can be obtained after application of the Hankel transform of order zero ($\sigma = 0$) and after projection of (2.1) onto the z axis:

$$\frac{\partial^2 \hat{u}_z}{\partial z^2} + \left(\frac{1}{\alpha^2(z)} - p^2\right)\hat{u}_z + \frac{\partial}{\partial z}\ln[\lambda(z) + 2\mu(z)]\frac{\partial \hat{u}_z}{\partial z} + \frac{\omega p}{(\lambda(z) + 2\mu(z))}\left[\frac{\partial\lambda(z)}{\partial z} + (\lambda(z) + \mu(z))\frac{\partial}{\partial z}\right]H_1(u_r) + (\omega p)^2\frac{[\lambda(z) + \mu(z)]}{[\lambda(z) + 2\mu(z)]}\hat{u}_z = 0, \tag{2.4}$$

where

$$\hat{u}_z = H_0(u_z) \tag{2.5}$$

and u_r is the radial component of the displacement (Carrion and Hassanzadeh[10]). In Eq. (2.4), p is the ray parameter or horizontal slowness,

$$p = k_r/\omega, \tag{2.6}$$

and $\alpha(z)$ is the velocity of compressional waves as a function of z,

$$\alpha(z) = \sqrt{[\lambda(z) + 2\mu(z)]/p(z)}\,. \tag{2.7}$$

Equation (2.4) is the plane wave decomposition of the elastic wave equation (2.2) for the vertical component of the displacement; it describes the propagation of plane waves in an elastic medium with vertical inhomogeneities. Setting $p = 0$ in Eq. (2.4), we obtain the following equation:

$$\frac{\partial^2 \hat{u}_z}{\partial z^2} + \frac{\omega^2}{\alpha^2(z)}\hat{u}_z + \frac{\partial}{\partial z}\ln[\lambda(z) + 2\mu(z)]\frac{\partial \hat{u}_z}{\partial z} = 0. \tag{2.8}$$

Let us introduce a travel-time coordinate q:

$$q(z) = \int_0^\infty \frac{d\xi}{\alpha(\xi)}\,. \tag{2.9}$$

Using the definition of the travel-time coordinate $q(z)$, Eq. (2.8) can be presented as follows:

$$\left[\frac{\partial^2}{\partial q^2} + \frac{\partial \ln I(q)}{\partial q}\frac{\partial}{\partial q} + \omega^2\right]\hat{u}_z(\omega,q) = 0, \tag{2.10}$$

where

$$I(q) = \rho(q)\alpha(q)$$

is the acoustic impedance. Equation (2.10) is called the "reflectivity equation" since the second term in brackets on the left-hand side of this equation is proportional to the reflection coefficient in the single scattering approximation. This equation has been studied by several authors (see, for example, Ware and Aki,[11] Gray,[12] and Carrion *et al.*[13]).

In order to compute the acoustic impedance $I(q)$ by inversion of the differential operator in (2.10) several approaches can be used. Some of them will be discussed in the next paragraph.

III. COMPUTATION OF THE ACOUSTIC IMPEDANCE

Suppose that $I(q)$ is smooth enough so $I(q)\in C^2$. Then Eq. (2.10) can be transformed to the Schrödinger equation by the following substitutions:

$$G(q) = I^{-1/2}(q) \tag{3.1}$$

and

$$\psi(\omega,q) = u(\omega,q)G^{-1}(q). \tag{3.2}$$

This yields

$$\left[\frac{\partial^2}{\partial q^2} + \omega^2\right]\psi(\omega,q) = S(q)\psi(\omega,q), \tag{3.3}$$

where the scattering potential S is

$$S(q) = G(q)\frac{\partial^2}{\partial q^2}G^{-1}(q). \tag{3.4}$$

Now the Gel'fand–Levitan treatment can be applied as long as the boundary conditions are specified. Suppose that the boundary conditions are taken in terms of the observed data at the plane $z = 0$:

$d(t,r) = u(t,r,z = 0)$ is available for any $r\in[0,\infty)$.

A Laplace–Hankel transform applied to the data yields

$$\hat{u}(\omega,p) = H_0[d(t,r)]. \tag{3.5}$$

The boundary condition in the form of (3.5) with plane wave decomposed elastic wave equation (2.4) provides us with all ingredients needed for the inversion procedure. Setting $p = 0$ in Eq. (3.5) and using Eq. (3.3) the acoustic impedance $I(q)$ can be directly recovered using the Gel'fand–Levitan treatment. If we require that $I(q)\in C^1$, we can avoid the use of the Schrödinger operator and the acoustic impedance can be recovered from the reflectivity equation (2.10) using the algorithm proposed by Carrol and Santosa.[14]

Since the analyticity of the wave field in the lower half space is equivalent to the causality in the time domain, the principles of causality of the wave field can be used to construct different algorithms in the time domain. Burridge[15] extended the Gel'fand–Levitan theorem to the time variable wave fields and developed time domain algorithms for the recovery of the acoustic impedance using time domain Schrödinger operators.

An inverse Fourier transform of equation (1.10) yields

$$\left[\frac{\partial^2}{\partial q^2}+\frac{\partial \ln I(q)}{\partial q}\frac{\partial}{\partial q}-\frac{\partial^2}{\partial t^2}\right]u_z(t,q)=0. \quad (3.6)$$

This equation is equivalent to the following system of partial differential equations (PDE's):

$$\begin{vmatrix} \frac{\partial}{\partial t} & I(q)\frac{\partial}{\partial q} \\ \frac{\partial}{\partial q} & I(q)\frac{\partial}{\partial t}\end{vmatrix}\begin{vmatrix} P(t,q) \\ V(t,q)\end{vmatrix}=0, \quad (3.7)$$

where $V(t,q)=(\partial/\partial t)u_z(t,q)$ is the vertical component of particle velocity. Setting the initial conditions in the form

$$P(t,q)=V(t,q)=0, \quad t<0,$$

and the boundary conditions

$$V(t,0)=\dot{d}(t,q), \quad (3.8)$$

$$P(t,0)=\delta(t), \quad (3.9)$$

the acoustic impedance can be computed on one of the characteristics of a set of hyperbolic PDE's:

$$I(q)=P(q,q)/V(q,q). \quad (3.10)$$

The downward continuation algorithm similar to (3.7)–(3.10) has been discussed by Santosa and Schwetlick,[16] Bube and Burridge,[17] and Foster and Carrion.[18] The discrete implementation of these schemes is directly related to the Cholescky factorization of Toeplitz operators.

One of the advantages of such schemes is that they are computationally efficient and fast in comparison with algorithms based on a solution of integral equations. It also should be mentioned that in (3.10) the acoustic impedance or its derivatives can have a finite set of points where they have jump discontinuities.

IV. DERIVATION OF AN EQUATION FOR MODE-CONVERTED SHEAR WAVES

Let us consider formula (2.3) for $\sigma=0$. The inverse Hankel transform of order zero can be presented as

$$f(\omega,r)=\int_0^\infty f(\omega,k_r)J_0(k_r r)k_r\,dk_r. \quad (4.1)$$

Since $k_r=\omega p=\omega\sin\zeta/\alpha$, where ζ is the angle of propagation, (4.1) can be rewritten as

$$f(\omega,r)=\frac{\omega^2}{\alpha^2}\int_0^{\zeta^1} f(\omega,\zeta)J_0(\omega,\zeta)\sin\xi\cos\zeta\,d\zeta, \quad (4.2)$$

where $\zeta^1=\pi/2-i\infty$ (see Bath[19]). Formula (4.2) is a complete description of the point-source reflection response. The upper limit of integration in (4.2) is complex because the angle of propagation ζ can take on values which correspond to supercritical incidence (postcritical incidence). Complex values of the angular spectrum correspond to complex values of vertical number k_z and describe exponentially decay evanescent waves. It is important to investigate in which cases the contribution of the evanescent waves can be neglected. Since for evanescent waves the amplitudes behave as

$$\exp(-|k_z|z),$$

it is obvious that for $z>|k_z|^{-1}$ the contribution of the evanescent waves is negligibly small.

Let us now project Eq. (2.2) onto the z axis:

$$\mu(z)\nabla^2u+\left[(\lambda(z)+\mu(z))\frac{\partial}{\partial z}+\frac{\partial}{\partial z}\lambda(z)\right]\nabla\cdot u$$
$$+2\frac{\partial\mu(z)}{\partial z}\frac{\partial u}{\partial z}=\rho(z)\frac{\partial^2u}{\partial t^2}. \quad (4.3)$$

Let us now introduce two functions (potentials) φ and χ which generate the displacement via

$$u=\nabla\varphi+\nabla\times(0,\chi,0) \quad (4.4)$$

(see Richards and Frasier[20]).

The scalar potential φ can be presented as

$$\varphi(r,z)=A(r,z)\exp(i\omega\Omega), \quad (4.5)$$

where $A(r,z)$ is the amplitude and Ω describes the phase of the scalar potential. In the high-frequency approximation Ω satisfies the eikonal equation:

$$[\nabla\Omega]^2=\alpha^{-2}(z), \quad (4.6)$$

where $\alpha(z)$ is the local velocity for compressional waves.

The amplitude $A(r,z)$ satisfies the transport equation:

$$A\nabla^2\Omega+2\nabla A\cdot\nabla\Omega=0. \quad (4.7a)$$

Usually the eikonal equation (4.6) can be solved using ray tracing and then the transport equation (4.7a) can be reduced to the ordinary differential equation along the rays that are characteristics of the eikonal equation. In our derivation we are not interested in finding the amplitudes of the scalar and the vector potentials. They can be found from the continuity of traction and displacement across boundaries between each pair of adjacent layers.

Let us assume that for the incident compressional field

$$\Omega=pr+z\cos j/\alpha(z). \quad (4.7b)$$

It is easy to see that (4.7b) satisfies the eikonal equation. It is also easy to see that the transmitted compressional wave can be described in the approximation of geometrical optics by the phase of the scalar potential

$$\Omega_t=pr+\int_0^z\frac{\cos\xi}{\alpha(\xi)}d\xi, \quad (4.8)$$

which certainly also satisfies the eikonal equation (4.6). In the eikonal approximation the amplitude of transmitted P-waves can be estimated from a balance of energy flux toward and away from the interfaces between layers.

Since there are only two types of speeds for body waves in an elastic medium—compressional and shear—the eikonal equation for shear waves can be written as

$$(\nabla\Sigma)^2=\beta^{-2}(z), \quad (4.9)$$

where Σ is the phase of the vector potential χ. Suppose that a compressional wave was converted to shear wave at depth z and propagates as a shear wave. Then the reflected mode converted shear wave can be presented by the vector potential χ:

$$\chi(r,z)=pr+\int_z^0\frac{\cos\xi}{\beta(\xi)}d\xi. \quad (4.10)$$

Let us consider now the second term on the left-hand side of equation (4.3). It is quite obvious that only gradients of the scalar potential contribute to this term. Let us also consider the "postcritical" region for compressional waves, where

$$p\alpha(z) > 1, \quad \text{for all values of } z. \tag{4.11}$$

Then the scalar potential φ for transmitted compressional waves in the "postcritical" region can be expressed as follows:

$$\varphi_t(r,z) = A_t(r,z) \exp\left(i\omega pr - \omega \int_0^z \left|\frac{\cos\xi}{\alpha(\xi)}\right| d\xi\right), \tag{4.12}$$

where $A_t(r,z)$ is the amplitude of the transmitted compressional wave. Since the integrand in the exponential function is positive if condition (4.11) holds, then (4.12) describes compressional evanescent waves that exponentially decay with z. In the approximation of geometrical optics ($\omega \to \infty$), $\varphi_t \to 0$. This means that in the region described by (4.11), Eq. (4.3) can be approximately presented as

$$\mu(z)\nabla^2 u + 2\frac{\partial\mu(z)}{\partial z}\frac{\partial u}{\partial z} = -\rho(z)\omega^2 u. \tag{4.13}$$

Applying a Hankel transform to Eq. (4.13) yields

$$\frac{\partial^2 \hat{u}_z}{\partial z^2} + 2\frac{\partial \ln\mu(z)}{\partial z}\frac{\partial\hat{u}}{\partial z} + \omega^2\left(\frac{1}{\beta^2(z)} - p^2\right)\hat{u} = 0, \tag{4.14}$$

where $\beta^2 = \mu/\rho$ is the local speed of shear waves. Let us introduce a variable h similar to (2.9):

$$h(z) = \int_0^z \frac{\sqrt{1-p^2\beta^2(\xi)}}{\beta(\xi)}\, d\xi. \tag{4.15}$$

Using this coordinate whose physical meaning is the vertical travel time, Eq. (4.16) can be rewritten as

$$\left(\frac{\partial^2}{\partial h^2} + \frac{\partial}{\partial h}\ln\left[(\cos\zeta)K(h)\mu(h)\right]\frac{\partial}{\partial h} + \omega^2\right)\hat{u}(\omega,h) = 0, \tag{4.16}$$

where $\cos\zeta = \sqrt{1-p^2\beta^2}$ and $K(h)$ is the quantity which we call the "shear impedance,"

$$K(h) = \mu(h)\rho(h).$$

Equation (4.16) has an interesting property. This equation is similar to the acoustic impedance in the travel-time coordinate (2.10), which describes the propagation of the compressional waves. This remarkable property will allow us to obtain an algorithm that recovers the modulus and the density separately.

V. SEPARATE RECOVERY OF ALL LAME'S PARAMETERS AND THE DENSITY

Equation (4.16) is equivalent to the following set of PDE's in the time domain:

$$\frac{\partial W(t,h)}{\partial t} + K(h)\mu(h)\cos\zeta\frac{\partial V(t,h)}{\partial h} = 0, \tag{5.1}$$

$$\frac{\partial W(t,h)}{\partial h} + K(h)\mu(h)\cos\zeta\frac{\partial V(t,h)}{\partial t} = 0, \tag{5.2}$$

where $W(t,h)$ is an auxiliary function that satisfies the boundary condition

$$W(t,h=0) = \mu(0)\delta(t)/2, \tag{5.3}$$

and the initial condition

$$W(t,h) = 0, \quad \text{for all } t < 0. \tag{5.4}$$

Since $V(t,h=0)$ represents the observed data (boundary condition for the vertical component of particle velocity) the Gopinath–Sondhi-type integral operator can be written as follows:

$$f(h_i,t) + \frac{1}{2}\int_{-h_i}^{+h_i} V(|t-\tau|) f(\tau,h_i)\, d\tau = 1,$$
$$i = 1,2, \quad \tau \leqslant |h_i|, \tag{5.5}$$

which simply means that the (5.5) should be solved twice for any two experiments with plane compressional waves such that condition (4.11) is satisfied. Then we can estimate the quantity $B = K(h)\mu(h)\cos\zeta$ from

$$B_i = \frac{d}{dh_i}\int_0^{h_i} f(h_i,t)\, dt, \quad i = 1,2. \tag{5.6}$$

This means that the rigidity modulus and the density can be recovered separately by solving two equations (5.6) with two unknowns. Knowing the acoustic impedance (Sec. III) all Lame's parameters can be estimates. We should mention that the procedure [(5.5) and (5.6)] is similar to one proposed by Coen[21] for an acoustic medium using a Gel'fand–Levitan treatment.

VI. REMARKS

Sometimes it is important to calculate the parameters of a medium (acoustic or elastic) as functions of depth. As it was shown we calculate all parameters as function of travel times (for compressional and shear waves). For this reason, in order to rescale the computed parameters an integral equation similar to the one described by Howard[22] should be solved. Carrion[23] proposed a numerical recursive scheme for rescaling the computed parameters and recovering them directly as functions of depth.

It was shown (Carrion[24] *et al.* and Santosa and Symes[25]) that the accuracy of methods for joint reconstruction of the velocity and the density of an acoustic medium depends on the difference of the angles of incidence. It is important that the angular difference be large, otherwise the reconstruction problem fails. This means that the difference between two rays chosen for the determination of the rigidity modulus and the density should not be small.

VII. DISCUSSION

The inverse problem for an acoustic or an elastic medium can be treated as a special case of the generalized Riemann boundary value problem, which was extensively studied in recent years (see Chudnovsky[26]). In order to solve the Riemann problem the boundary conditions along with the incident wave field should be specified.

We demonstrated the possibility of recovering all Lame's parameters and the density using experiments with compressional plane waves only. (Until now, mathematically rigorous justifications have been given only when experiments with incident shear waves are available.) Although Eq. (4.14) is approximate (it was derived in the approximation of geometrical optics), actually it can be used for a much wider range of frequencies. For example, in exploration geophysics typical frequencies can be considered as satisfying the "high frequency" approximation and the impact of the evanescent waves will be very small.

It is important to mention that although we require that condition (4.11) holds, our formulation is valid when, below the depth $z = |k_z|^{-1}$, there are so-called low-velocity zones, where possibly $p\alpha(z)$ is less than 1. In general, this can cause the "tunneling effect" when an evanescent wave propagates with real angles. However, below a certain depth ($z > |k_z|^{-1}$), which is called the critical depth, the influence of the low-velocity zones is negligibly small since an evanescent wave arriving to this depth is characterized by negligibly small amplitude. In order to use Eq. (4.16) for the recovery of the rigidity modulus and the density in the region of compressional evanescent waves we should choose those values of Snell's parameter (ray parameter) p that satisfy $p\beta(z) < 1$ for all values of z. Let us consider the following example: Suppose compressional plane waves impinge on the ocean bottom. Taking $p^* = 1/\alpha_0$, where α_0 is the average sound speed in the water, we are assured that for any depth below the sea bottom, $p^*\alpha(z) > 1$. However, we also should be sure that $p^*\beta(z) < 1$. This means that we can recover only those shear velocities that satisfy $\beta(z) < 1/p^*$. If we want to recover higher shear velocities (if they exist) an additional approach will be discussed elsewhere.

ACKNOWLEDGMENTS

I would like to thank an anonymous reviewer whose suggestions and comments made the paper clearer.

This work was supported by the National Science Foundation No. EAR-850 59 22.

[1] R. Newton, Geophys. J. R. Astron. Soc. **76**, 233 (1981).
[2] S. Coen, J. Acoust. Soc. Am. **70**, 172 (1981).
[3] F. Aminzadeh, Geophys. Prospect. **33**, 422 (1985).
[4] M. Shiva and J. Mendel, Geophys. Prospect. **31**, 888 (1983).
[5] T. Clarke, Geophys. J. R. Astron. Soc. **77**, 120 (1983).
[6] A. Yagle and B. Levy, Geophys. **50**, 244 (1985).
[7] D. Stickler (personal communication, 1985).
[8] J. Carazzone, in *Proceedings of the 55th SEG Meeting* (1985), pp. 375–378.
[9] M. Meadows and S. Coen, in Ref. 8, pp. 336–339.
[10] P. Carrion and S. Hassanzedeh, Geophys. **50**, 530 (1985).
[11] J. Ware and K. Aki, J. Acoust. Soc. Am. **45**, 911 (1969).
[12] S. Gray, J. Math. Phys. **24**, 1148 (1969).
[13] P. Carrion, J. Kuo, and W. Patton, Geophys. J. R. Astron. Soc. **79**, 425 (1984).
[14] R. Carroll and F. Santosa, Math. Appl. Sci. **3**, 145 (1981).
[15] R. Burridge, Wave Motion **2**, 305 (1980).
[16] F. Santosa and R. Schwetlick, Wave Motion (1981).
[17] K. Bube and R. Burridge, SIAM Rev. **25**, 497 (1983).
[18] D. Foster and P. Carrion, Geophys. Prospect. (in press).
[19] M. Bath, *Mathematical Aspects of Seismology* (Elsevier, Amsterdam, 1968).
[20] P. Richards and C. Frasier, Geophys. **41**, 441 (1976).
[21] S. Coen, Geophys. **48**, 898 (1981).
[22] M. Howard, Geophys. **48**, 163 (1983).
[23] P. Carrion, J. Acoust. Soc. Am. **77**, 1370 (1985).
[24] P. Carrion, J. Kuo, and P. Stoffa, Geophys. Prospect. **32**, 375 (1984).
[25] F. Santosa and W. Symes, Geophys. J. R. Aston. Soc. **81**, 175 (1985).
[26] D. Chudnovsky, "Bifurcation phenomena in mathematical physics and related topics," in *Proceedings of the NATO Advanced Study Institute* (1979), pp. 384–447.

ERRATUM

Erratum: On a labeling for point group harmonics. I and II [J. Math. Phys. 26, 2413, 2441 (1985)]

Robert Conte
Service de Physique du Solide et de Résonance Magnétique, CEA-Saclay, 91191 Gif-sur-Yvette Cedex, France

Jacques Raynal
Service de Physique Théorique, CEA-Saclay, 91191 Gif-sur-Yvette Cedex, France

(Received 16 December 1985; accepted for publication 8 January 1986)

In the first paper, Table V, p. 2428, the last line of the second component of $|\Gamma_3 22b\rangle$, which begins with "(113)...," should be shifted one column entry to the right, so that (113) is under the heading $|2+\rangle$ and .43 under the heading $|22+\rangle$.

In the second paper, formula (30), p. 2446, the second index i_1 should be replaced by i_2.

 0022-2488/86/041169-01$02.50