Boson representations of the real symplectic group and their applications to the nuclear collective model

J. Deenen and C. Quesne[a)]

Service de Physique Théorique et Mathématique CP 229, Université Libre de Bruxelles, Bd. du Triomphe, B 1050 Brussels, Belgium

(Received 27 August 1984; accepted for publication 31 May 1985)

Both non-Hermitian Dyson and Hermitian Holstein–Primakoff representations of the Sp($2d,R$) algebra are obtained when the latter is restricted to a positive discrete series irreducible representation $\langle \lambda_d + n/2,\dots,\lambda_1 + n/2\rangle$. For such purposes, some results for boson representations, recently deduced from a study of the Sp($2d,R$) partially coherent states, are combined with some standard techniques of boson expansion theories. The introduction of Usui operators enables the establishment of useful relations between the various boson representations. Two Dyson representations of the Sp($2d,R$) algebra are obtained in compact form in terms of $\nu = d(d+1)/2$ pairs of boson creation and annihilation operators, and of an extra U(d) spin, characterized by the irreducible representation $[\lambda_1 \cdots \lambda_d]$. In contrast to what happens when $\lambda_1 = \cdots = \lambda_d = \lambda$, it is shown that the Holstein–Primakoff representation of the Sp($2d,R$) algebra cannot be written in such a compact form for a generic irreducible representation. Explicit expansions are, however, obtained by extending the Marumori, Yamamura, and Tokunaga method of boson expansion theories. The Holstein–Primakoff representation is then used to prove that, when restricted to the Sp($2d,R$) irreducible representation $\langle \lambda_d + n/2,\dots,\lambda_1 + n/2\rangle$, the dn-dimensional harmonic oscillator Hamiltonian has a U(ν)$\times$SU(d) symmetry group. Finally, the results are applied to the Sp($6,R$) nuclear collective model to demonstrate the existence of a hidden U(6)$\times$SU(3) symmetry in this model.

I. INTRODUCTION

During the last few years, the real symplectic group Sp($2d,R$) has played an ever-increasing role in physical applications. Its importance is largely due to the fact that it is the main component of the d-dimensional harmonic oscillator dynamical group.[1,2] The Sp($2d,R$) irreducible representations (irreps) encountered in applications are positive discrete series,[3,4] characterized by their lowest weight $\langle \lambda_d + n/2,\dots,\lambda_1 + n/2\rangle$, where $[\lambda_1\lambda_2\cdots\lambda_d]$ is a partition, and n is an integer greater than or equal to $2d$.

In various fields, such as the theory of collective motion in nuclei,[5–17] and the study of $1/N$ expansions in quantum mechanics and field theory,[18–20] it is important to obtain boson realizations (somewhat improperly called boson representations in the literature) of the Sp($2d,R$) algebra when it is restricted to a given irrep representation space. These boson realizations include both non-Hermitian Dyson[21] and Hermitian Holstein–Primakoff (HP)[22] representations.

The boson representations of a restricted Lie algebra are closely related to its realizations as an algebra of differential operators acting in a space of analytic functions in some complex variables.[23,24] The transition from the latter to the former can be achieved by simply replacing the complex variables z_i, and the corresponding differential operators $\partial/\partial z_i$, by boson creation and annihilation operators, $a_i^\dagger$ and a_i, respectively. However, in such a process, the Hermitian differential operator realizations are generally converted into non-Hermitian Dyson representations since a_i is the Hermitian conjugate of $a_i^\dagger$, whereas, except for the Bargmann representation,[23] $\partial/\partial z_i$ is not the Hermitian conjugate of z_i with respect to the scalar product defined in the space of analytic functions. Consequently, if it is rather easy to obtain one of the infinitely many Dyson representations, it is by far much more difficult to get the (up to a unitary transformation) uniquely defined HP representation.

For the case where $\lambda_1 = \cdots = \lambda_d = \lambda$, Rosensteel and Rowe[3] considered in 1977 a realization of the restricted Sp($2d,R$) algebra in a space of analytic functions on the Siegel half-plane.[25] In 1982, Kramer[16] constructed an essentially equivalent[26] realization in a space of analytic functions on the origin-centered unit ball, by using Perelomov generalization of coherent states (CS).[27] This same year, we proposed a realization in a space of analytic functions on the whole complex space,[6,28] shown later on by us[7] to be related to Barut–Girardello generalization of CS.[29] When translated in terms of boson representations as explained above, all these three realizations would give rise to Dyson representations, which are not Hermitian. By 1982, we obtained the corresponding Hermitian HP representation by extending to Sp($2d,R$) some results previously demonstrated for Sp($2,R$),[5,18,30,31] and Sp($4,R$).[18,32] Later on, we reformulated our derivation of the HP representation in terms of CS.[7]

For the more difficult case of a generic positive discrete series irrep, Rosensteel and Rowe[3] considered in 1977 Godement generalization[33] of the above-mentioned Siegel construction, valid when $\lambda_1 = \cdots = \lambda_d = \lambda$. However, they did not explicitly give the corresponding Sp($2d,R$) algebra realization. By 1983, we derived an explicit Sp($2d,R$) algebra realization[8] in a space of analytic functions, later on shown by us[9] to be connected with Barut–Girardello generalization

[a)] Maître de recherches F.N.R.S.

 0022-2488/85/112705-12$02.50

of CS. In terms of boson representations, both these realizations give again rise to Dyson representations.

To study the transition from the Dyson representations to the HP one, the theory of generalized CS did provide an appropriate framework. To extend to a generic positive discrete series irreps the CS considered in Ref. 7 for the $\lambda_1 = \cdots = \lambda_d = \lambda$ case, by 1984 we introduced a new concept, namely that of partially coherent states (PCS), characterized by both continuous and discrete labels.[9] We considered three classes of PCS, respectively, generalizing the Perelomov[27] and Barut–Girardello[29] CS, as well as the intermediate CS introduced in Ref. 7, then we related them to three boson representations of Dyson type in the first two cases, and of HP type in the last one.

A rather similar generalization of Perelomov CS was proposed independently by Rowe,[15] who also analyzed the transition from the corresponding Dyson representation to the HP one by a method close to that used in Ref. 6 for the $\lambda_1 = \cdots = \lambda_d = \lambda$ case. He did obtain an approximate HP representation, giving rise to a very accurate analytic expression for the Sp(6,R) generator matrix elements for those Sp(6,R) irreps relevant to the nuclear collective model.

The aim of the present paper is to further study the Sp(2d,R) boson representations along the guidelines of Ref. 9, then to apply them to the Sp(6,R) nuclear collective model. This approach parts from those of Refs. 6 and 15, but uses the standard techniques of boson expansion theories.[34–36] More specifically, we wish to show that the relations between the various boson representations can be displayed in a straightforward way by using Usui operators.[37] We also wish to prove that by applying the techniques of Marumori, Yamamura, and Tokunaga[34] the Sp(2d,R) generator HP representation can be written as an explicit expansion in terms of $\nu = d(d+1)/2$ pairs of boson creation and annihilation operators, and of an extra U(d) spin, characterized by the irrep $[\lambda_1\lambda_2\cdots\lambda_d]$. Since the two Dyson representations of the Sp(2d,R) generators, introduced in Ref. 9, were obtained in the same reference, this completes the determination of the Sp(2d,R) boson representations for a generic positive discrete series. There is, however, an essential difference with respect to the special case $\lambda_1 = \lambda_2 = \cdots = \lambda_d = \lambda$, in that, as a general rule, the HP representation cannot be written in a compact form similar to that obtained in Ref. 6 for this case.

This paper is organized as follows. In Sec. II, three Usui operators are defined, and their properties are listed. These operators are used in Sec. III to prove various properties of boson representations, which are then applied in Sec. IV to the case of the Sp(2d,R) generators. In Sec. V, explicit expansions are obtained for the HP representation of the latter. In Sec. VI, some properties demonstrated in the previous sections are employed to determine the dn-dimensional harmonic oscillator symmetry group, when the oscillator Hamiltonian is restricted to the Sp(2d,R) irrep $\langle \lambda_d + n/2,\ldots,\lambda_1 + n/2\rangle$. Finally, in Sec. VII, the results are applied to the Sp(6,R) nuclear collective model.

II. DEFINITION AND PROPERTIES OF USUI OPERATORS

In the present work, we shall use the following well-known realization of the Sp(2d,R) algebra[1]:

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

where

$$
C_{ij} = \sum_{s=1}^{n} \eta_{is}\xi_{js}, \quad i,j = 1,\ldots,d, \qquad (2.2)
$$

and η_{is}, ξ_{is}, $i = 1,\ldots,d$, $s = 1,\ldots,n$, denote dn pairs of boson creation and annihilation operators. The operators E_{ij} generate the U(d) subgroup of Sp(2d,R). In the following, we shall use instead the operators C_{ij}, which satisfy the same commutation relations. Since we wish to consider a generic positive discrete series irrep $\langle \lambda_d + n/2,\ldots,\lambda_1 + n/2\rangle$ of Sp(2d,R), we assume that n is greater than or equal to 2d. As we did show in Ref. 9, it is interesting to introduce various bases in this irrep representation space $\mathscr{F}$. For simplicity's sake, we shall restrict ourselves here to U(d)-uncoupled bases, and only briefly outline in Sec. IV the advantages that could result from the use of U(d)-coupled bases. It should be realized, however, that all the properties demonstrated in this section, as well as in Secs. III and V, can be extended to U(d)-coupled bases in a straightforward way.

The $\langle \lambda_d + n/2,\ldots,\lambda_1 + n/2\rangle$ lowest-weight state belongs to a Λ-dimensional irrep of the U(d) subgroup, characterized by $[\lambda_1\cdots\lambda_d]$. From the basis states $|(\lambda)\rangle$ of $[\lambda_1\cdots\lambda_d]$, where (λ) denotes a Gel'fand pattern,[38] all the states of $\mathscr{F}$ can be obtained by acting with the generators $D^\dagger_{ij}$. Basis states of $\mathscr{F}$ may therefore be defined by the relation

$$
|\mathbf{N};(\lambda)\rangle = F_{\mathbf{N}}(\mathbf{D}^\dagger)|(\lambda)\rangle, \qquad (2.3)
$$

where $F_{\mathbf{N}}(\mathbf{D}^\dagger)$ is the following function of the $d \times d$ matrix $\mathbf{D}^\dagger = \|D^\dagger_{ij}\|$:

$$
F_{\mathbf{N}}(\mathbf{D}^\dagger) = \prod_{i \leqslant j} (N_{ij}!)^{-1/2}[(1+\delta_{ij})^{-1/2}D^\dagger_{ij}]^{N_{ij}}, \qquad (2.4)
$$

and the quantum numbers N_{ij}, $1 \leqslant i \leqslant j \leqslant d$, run over all nonnegative integers. The states $|\mathbf{N};(\lambda)\rangle$ do not form an orthogonal basis. Let us denote their overlap matrix by $\mathbf{M}$

$$
M_{\mathbf{N}'(\lambda'),\mathbf{N}(\lambda)} = \langle \mathbf{N}';(\lambda')|\mathbf{N};(\lambda)\rangle. \qquad (2.5)
$$

Another interesting basis of $\mathscr{F}$ is the dual basis, whose states $|\mathbf{N};(\lambda))$ are defined by the relation

$$
(\mathbf{N}';(\lambda')|\mathbf{N};(\lambda)\rangle = \delta_{\mathbf{N}',\mathbf{N}}\delta_{(\lambda'),(\lambda)}, \qquad (2.6)
$$

and are given in terms of the states $|\mathbf{N}';(\lambda')\rangle$ by

$$
|\mathbf{N};(\lambda)) = \sum_{\mathbf{N}'(\lambda')} |\mathbf{N}';(\lambda')\rangle(\mathbf{M}^{-1})_{\mathbf{N}'(\lambda'),\mathbf{N}(\lambda)}, \qquad (2.7)
$$

where $\mathbf{M}^{-1}$ is the inverse of the overlap matrix. They satisfy the following unity resolution relation:

$$
\sum_{\mathbf{N}(\lambda)} |\mathbf{N};(\lambda))\langle \mathbf{N};(\lambda)| = I, \qquad (2.8)
$$

where I denotes the unit operator in $\mathscr{F}$.

Finally, we may also consider an orthonormal basis in

$\mathscr{F}$, whose states are defined by

$$|\mathbf{N};(\lambda)\} = \sum_{\mathbf{N}'(\lambda')} |\mathbf{N}';(\lambda')\rangle R_{\mathbf{N}'(\lambda'),\mathbf{N}(\lambda)}$$
$$= \sum_{\mathbf{N}'(\lambda')} |\mathbf{N}';(\lambda'))(\mathbf{MR})_{\mathbf{N}'(\lambda'),\mathbf{N}(\lambda)}. \tag{2.9}$$

From Eq. (2.6) and the orthonormality property of $|\mathbf{N};(\lambda)\}$, $\mathbf{R}$ must satisfy the following condition:

$$\mathbf{R}^\dagger\mathbf{MR} = \mathbf{I}, \tag{2.10}$$

or equivalently,

$$\mathbf{RR}^\dagger = \mathbf{M}^{-1}. \tag{2.11}$$

From Eq. (2.11), it follows that $\mathbf{R}$ differs from the square root of $\mathbf{M}^{-1}$ by some unitary matrix S, i.e.,

$$\mathbf{R} = \mathbf{M}^{-1/2}\mathbf{S}, \quad \text{where } \mathbf{SS}^\dagger = \mathbf{I}. \tag{2.12}$$

When choosing $\mathbf{S} = \mathbf{I}$, thence $\mathbf{R} = \mathbf{M}^{-1/2}$, we obtain the intermediate orthonormal basis considered in Ref. 9. In the present work, we shall, however, make no specific choice for S, and let it be arbitrary.

According to Ref. 9, to each of the three bases (2.3), (2.7), and (2.9), we can associate a set of PCS as follows:

$$|\mathbf{u};(\lambda)\rangle = \sum_{\mathbf{N}} F_{\mathbf{N}}(\mathbf{u}^*)|\mathbf{N};(\lambda)\rangle, \tag{2.13a}$$

$$|\mathbf{u};(\lambda)) = \sum_{\mathbf{N}} F_{\mathbf{N}}(\mathbf{u}^*)|\mathbf{N};(\lambda)), \tag{2.13b}$$

and

$$|\mathbf{u};(\lambda)\} = \sum_{\mathbf{N}} F_{\mathbf{N}}(\mathbf{u}^*)|\mathbf{N};(\lambda)\}, \tag{2.13c}$$

where

$$F_{\mathbf{N}}(\mathbf{u}^*) = \prod_{i\leqslant j} (N_{ij}!)^{-1/2}[(1+\delta_{ij})^{-1/2}u^*_{ij}]^{N_{ij}}. \tag{2.14}$$

These PCS are, respectively, referred to as Perelomov, Barut–Girardello, and intermediate PCS. Note that in contrast to Ref. 9, in Eq. (2.13) we use the same symbols u_{ij} to denote the complex variables the various PCS depend on, although their domain of variation is different. This does not matter here since the PCS will essentially be used as generating functions for the corresponding discrete basis states.

Let us now introduce the boson states

$$|\mathbf{N}] = \prod_{i\leqslant j} (N_{ij}!)^{-1/2}(a^\dagger_{ij})^{N_{ij}}|0], \tag{2.15}$$

built from $\nu = d(d+1)/2$ independent boson creation operators $a^\dagger_{ij} = a^\dagger_{ji}$, $i,j = 1,\ldots,d$, acting upon the vacuum state $|0]$. As in Ref. 9, we take their direct product with the U(d)-spin states $|(\lambda)]$,[39] characterized by the irrep $[\lambda_1\cdots\lambda_d]$, and obtain the extended boson states

$$|\mathbf{N};(\lambda)] = |\mathbf{N}] \otimes |(\lambda)]. \tag{2.16}$$

The latter are orthonormal and span a Hilbert space, which is the direct product of a standard boson space $\mathscr{B}$ with the Λ-dimensional representation space $\mathscr{S}$ of the U(d) irrep $[\lambda_1\cdots\lambda_d]$. It will prove convenient to use non-normalized boson creation operators $\bar{a}^\dagger_{ij} = \bar{a}^\dagger_{ji} = (1+\delta_{ij})^{1/2}a^\dagger_{ij}$, whose commutation relations with their corresponding annihilation operators $\bar{a}_{ij}$ are given by

$$[\bar{a}_{ij},\bar{a}^\dagger_{kl}] = \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}. \tag{2.17}$$

In their terms, the extended boson states can be rewritten as

$$|\mathbf{N};(\lambda)] = F_{\mathbf{N}}(\bar{\mathbf{a}}^\dagger)|\mathbf{0};(\lambda)], \tag{2.18}$$

where

$$|\mathbf{0};(\lambda)] = |0] \otimes |(\lambda)], \tag{2.19}$$

and $F_{\mathbf{N}}(\bar{\mathbf{a}}^\dagger)$ is defined by Eq. (2.4) with $\mathbf{D}^\dagger$ replaced by $\bar{\mathbf{a}}^\dagger$. Note the close similarity between Eqs. (2.3) and (2.18).

Extended Glauber CS are defined by

$$|\mathbf{u};(\lambda)] = \sum_{\mathbf{N}} F_{\mathbf{N}}(\mathbf{u}^*)|\mathbf{N};(\lambda)], \tag{2.20}$$

or

$$|\mathbf{u};(\lambda)] = |\mathbf{u}] \otimes |(\lambda)], \tag{2.21}$$

where $|\mathbf{u}]$ is a standard Glauber CS[40]

$$|\mathbf{u}] = \exp(\tfrac{1}{2}\,\mathrm{tr}\,\mathbf{u}^*\bar{\mathbf{a}}^\dagger)|0]. \tag{2.22}$$

Either of the three discrete bases (2.3), (2.7), and (2.9) of $\mathscr{F}$ can be mapped in a one-to-one fashion onto the basis (2.16) of $\mathscr{B} \otimes \mathscr{S}$. The operators performing these mappings are known as Usui operators[37] in boson expansion theories.[34–36] They are, respectively, given by

$$\hat{U} = \sum_{\mathbf{N}(\lambda)} |\mathbf{N};(\lambda)](\mathbf{N};(\lambda)|, \tag{2.23a}$$

$$U = \sum_{\mathbf{N}(\lambda)} |\mathbf{N};(\lambda)]\langle\mathbf{N};(\lambda)|, \tag{2.23b}$$

and

$$\check{U} = \sum_{\mathbf{N}(\lambda)} |\mathbf{N};(\lambda)]\{\mathbf{N};(\lambda)|. \tag{2.23c}$$

From Eq. (2.6), it is indeed obvious that $\hat{U}$ maps the basis states $|\mathbf{N};(\lambda)\rangle$ onto the extended boson states $|\mathbf{N};(\lambda)]$ as follows:

$$\hat{U}|\mathbf{N};(\lambda)\rangle = |\mathbf{N};(\lambda)]. \tag{2.24a}$$

In a similar way, it can be shown that

$$U|\mathbf{N};(\lambda)) = |\mathbf{N};(\lambda)] \tag{2.24b}$$

and

$$\check{U}|\mathbf{N};(\lambda)\} = |\mathbf{N};(\lambda)]. \tag{2.24c}$$

Let us also introduce the operators from $\mathscr{B} \otimes \mathscr{S}$ to $\mathscr{F}$ defined by

$$\hat{U}^\dagger = \sum_{\mathbf{N}(\lambda)} |\mathbf{N};(\lambda))[\mathbf{N};(\lambda)|, \tag{2.25a}$$

$$U^\dagger = \sum_{\mathbf{N}(\lambda)} |\mathbf{N};(\lambda)\rangle[\mathbf{N};(\lambda)|, \tag{2.25b}$$

and

$$\check{U}^\dagger = \sum_{\mathbf{N}(\lambda)} |\mathbf{N};(\lambda)\}[\mathbf{N};(\lambda)|. \tag{2.25c}$$

By using Eqs. (2.6) and (2.8), as well as similar relations for $|\mathbf{N};(\lambda)\}$ and $|\mathbf{N};(\lambda)]$, it is straightforward to prove the following equations:

$$I = \hat{U}^\dagger U \tag{2.26a}$$

$$= U^\dagger \hat{U} \tag{2.26b}$$
$$= \check{U}^\dagger \check{U} \tag{2.26c}$$

and

$$I_{\mathscr{B}} \otimes I_{\mathscr{S}},$$
$$= U\hat{U}^\dagger \tag{2.27a}$$
$$= \hat{U}U^\dagger \tag{2.27b}$$
$$= \check{U}\check{U}^\dagger \tag{2.27c}$$

where $I_{\mathscr{B}}$ and $I_{\mathscr{S}}$ are the unit operators in $\mathscr{B}$ and $\mathscr{S}$, respectively.

For subsequent purposes, it is convenient to introduce an operator T acting on $\mathscr{B} \otimes \mathscr{S}$, whose matrix elements with respect to the basis states $|\mathbf{N};(\lambda)]$ reproduce the overlaps of the corresponding $\mathscr{F}$ basis states $|\mathbf{N};(\lambda)\rangle$, i.e.,

$$[\mathbf{N}';(\lambda')|T|\mathbf{N};(\lambda)] = \langle \mathbf{N}';(\lambda')|\mathbf{N};(\lambda)\rangle. \tag{2.28}$$

This operator can be expressed in terms of the Usui operator U. By multiplying Eq. (2.24a) by $U^\dagger$ from the left, and by taking Eq. (2.26b) into account, $|\mathbf{N};(\lambda)\rangle$ can indeed be expressed in terms of $|\mathbf{N};(\lambda)]$ as follows:

$$|\mathbf{N};(\lambda)\rangle = U^\dagger|\mathbf{N};(\lambda)]. \tag{2.29}$$

When introducing Eq. (2.29) and its Hermitian conjugate into Eq. (2.28), we obtain

$$T = UU^\dagger, \tag{2.30}$$

which is a positive definite, Hermitian operator, as it should be. Owing to Eqs. (2.26a), (2.26b), (2.27a), and (2.27b), this operator has an inverse, given by

$$T^{-1} = \hat{U}\hat{U}^\dagger. \tag{2.31}$$

III. BOSON REPRESENTATIONS IN TERMS OF USUI OPERATORS

As shown in Ref. 9, the boson representations, in $\mathscr{B} \otimes \mathscr{S}$, of any operator X acting in $\mathscr{F}$ can be obtained from its PCS representations. We shall therefore start to summarize the relevant results of Ref. 9, and then reformulate them in terms of Usui operators.

In each of the three PCS representations, defined in Eq. (2.13), X is represented by a matrix differential operator, according to the following definitions:

$$\langle \mathbf{u};(\lambda)|X|\psi\rangle = \sum_{(\lambda')} \hat{X}_{(\lambda),(\lambda')} \langle \mathbf{u};(\lambda')|\psi\rangle, \tag{3.1a}$$

$$(\mathbf{u};(\lambda)|X|\psi\rangle = \sum_{(\lambda')} X_{(\lambda),(\lambda')} (\mathbf{u};(\lambda')|\psi\rangle, \tag{3.1b}$$

and

$$\{\mathbf{u};(\lambda)|X|\psi\rangle = \sum_{(\lambda')} \check{X}_{(\lambda),(\lambda')} \{\mathbf{u};(\lambda')|\psi\rangle, \tag{3.1c}$$

where $|\psi\rangle$ denotes any vector of $\mathscr{F}$. For instance, $\|\hat{X}_{(\lambda),(\lambda')}\|$ is the $\Lambda \times \Lambda$ matrix of differential operators representing X in the Perelomov PCS representation.

Each of the three PCS representations of X can be viewed as the representation of some operator acting in $\mathscr{B} \otimes \mathscr{S}$, in the extended Glauber CS representation defined in Eq. (2.21). For the first two PCS representations, the operators so obtained are Dyson representations of X,[21] that we, respectively, denote by $X_{\bar{\mathrm{D}}}$ and X_{D}, while for the last one, it is an HP representation,[22] that we call X_{HP}. Hence we have the following three definitions:

$$[\mathbf{u};(\lambda)|X_{\bar{\mathrm{D}}}|\psi] = \sum_{(\lambda')} \hat{X}_{(\lambda),(\lambda')} [\mathbf{u};(\lambda')|\psi], \tag{3.2a}$$

$$[\mathbf{u};(\lambda)|X_{\mathrm{D}}|\psi] = \sum_{(\lambda')} X_{(\lambda),(\lambda')} [\mathbf{u};(\lambda')|\psi], \tag{3.2b}$$

and

$$[\mathbf{u};(\lambda)|X_{\mathrm{HP}}|\psi] = \sum_{(\lambda')} \check{X}_{(\lambda),(\lambda')} [\mathbf{u};(\lambda')|\psi], \tag{3.2c}$$

where $|\psi]$ denotes any vector of $\mathscr{B} \otimes \mathscr{S}$. To explicitly obtain $X_{\bar{\mathrm{D}}}, X_{\mathrm{D}}$, and X_{HP} from the corresponding PCS representations, all we have to do is to replace (i) the complex variables u_{ij} and the corresponding differential operators $\Delta_{u_{ij}} = (1 + \delta_{ij})\partial/\partial u_{ij}$ by the non-normalized boson creation and annihilation operators $\bar{a}^\dagger_{ij}$ and $\bar{a}_{ij}$, respectively, and (ii) the $\Lambda \times \Lambda$ matrices in $\mathscr{S}$ by the operators they represent. This procedure will be illustrated for the Sp($2d,R$) generators in the next section.

Let us now express the boson representations of X, as defined above, in terms of the Usui operators introduced in the previous section. Starting with $X_{\bar{\mathrm{D}}}$, we first note that Eq. (2.29) implies for the CS the relation

$$|\mathbf{u};(\lambda)\rangle = U^\dagger|\mathbf{u};(\lambda)]. \tag{3.3}$$

Introducing then the Hermitian conjugate of Eq. (3.3) into Eq. (3.1a), and dropping $|\psi\rangle$ in the latter, we obtain

$$[\mathbf{u};(\lambda)|UX = \sum_{(\lambda')} \hat{X}_{(\lambda),(\lambda')} [\mathbf{u};(\lambda')|U. \tag{3.4}$$

Finally, by multiplying Eq. (3.4) by $\hat{U}^\dagger$ from the right, and by taking Eq. (2.27a) into account, we get the following equation:

$$[\mathbf{u};(\lambda)|UX\hat{U}^\dagger = \sum_{(\lambda')} \hat{X}_{(\lambda),(\lambda')} [\mathbf{u};(\lambda')|, \tag{3.5}$$

which, when compared with Eq. (3.2a), leads to the relation

$$X_{\bar{\mathrm{D}}} = UX\hat{U}^\dagger. \tag{3.6a}$$

Proceeding in the same way for the remaining two boson representations, we obtain

$$X_{\mathrm{D}} = \hat{U}XU^\dagger \tag{3.6b}$$

and

$$X_{\mathrm{HP}} = \check{U}X\check{U}^\dagger. \tag{3.6c}$$

Equation (3.6) enables us to easily derive the Hermiticity properties of the boson representations. By taking the Hermitian conjugate of Eq. (3.6b), we get the relation

$$(X_{\mathrm{D}})^\dagger = UX^\dagger\hat{U}^\dagger, \tag{3.7}$$

which is nothing else than the Dyson representation $(X^\dagger)_{\bar{\mathrm{D}}}$ of $X^\dagger$

$$(X_{\mathrm{D}})^\dagger = (X^\dagger)_{\bar{\mathrm{D}}}. \tag{3.8}$$

In the same way, we obtain

$$(X_{\mathrm{HP}})^\dagger = (X^\dagger)_{\mathrm{HP}}. \tag{3.9}$$

We therefore recover the well-known fact that the Dyson representations do not preserve the Hermiticity properties, while the HP representation does. However, Eq. (3.8) also tells us that both Dyson representations are intimately connected with one another through Hermitian conjugation.

Some important relations between the boson representations $X_{\bar{D}}$, X_D, and X_{HP} of the same operator X can also be inferred from Eq. (3.6). Let us start with the relation between X_D and $X_{\bar{D}}$. We note that by using Eq. (2.26b), Eq. (3.6b) can be inverted to express X in terms of X_D as follows:

$$X = U^\dagger X_D \hat{U}. \tag{3.10}$$

By introducing Eq. (3.10) into the right-hand side of Eq. (3.6a), we then obtain

$$X_{\bar{D}} = UU^\dagger X_D \hat{U}\hat{U}^\dagger, \tag{3.11}$$

or equivalently

$$X_{\bar{D}} = TX_D T^{-1}, \tag{3.12}$$

where use has been made of Eqs. (2.30) and (2.31).

Let us now establish a relation between X_D and X_{HP}. By introducing Eq. (3.10) into the right-hand side of Eq. (3.6c), the latter becomes

$$X_{HP} = \check{U}U^\dagger X_D \hat{U}\check{U}^\dagger, \tag{3.13}$$

or

$$X_{HP} = WX_D W^{-1}, \tag{3.14}$$

where the operator W, acting in $\mathscr{B} \otimes \mathscr{S}$, is defined by

$$W = \check{U}U^\dagger. \tag{3.15}$$

From Eqs. (2.26c) and (2.30), it follows that W satisfies the relation

$$W^\dagger W = T. \tag{3.16}$$

It can therefore be written as

$$W = VT^{1/2}, \tag{3.17}$$

where V is some unitary operator in $\mathscr{B} \otimes \mathscr{S}$, i.e.,

$$V^\dagger V = I_{\mathscr{B}} \otimes I_{\mathscr{S}}. \tag{3.18}$$

Equation (3.14) becomes

$$X_{HP} = VT^{1/2} X_D T^{-1/2} V^\dagger, \tag{3.19}$$

so that X_{HP} is determined by $T^{1/2} X_D T^{-1/2}$, up to some unitary transformation V.

We shall now proceed to show that V may be chosen to be the unit operator in $\mathscr{B} \otimes \mathscr{S}$. For such purposes, let us remember that X_{HP} is obtained by mapping the orthonormal basis $|\mathbf{N};(\lambda)\}$ of $\mathscr{F}$ onto the orthonormal basis $|\mathbf{N};(\lambda)]$ of $\mathscr{B} \otimes \mathscr{S}$. Hence $T^{1/2} X_D T^{-1/2}$ results from the mapping of $|\mathbf{N};(\lambda)\}$ onto another orthonormal basis of $\mathscr{B} \otimes \mathscr{S}$, whose states are defined by $V^\dagger|\mathbf{N};(\lambda)]$. Denoting by $V_{\mathscr{F}}$ the unitary operator in $\mathscr{F}$ which has the same action on $|\mathbf{N};(\lambda)\}$ as the operator V on $|\mathbf{N};(\lambda)]$, it is then obvious that the mapping of $|\mathbf{N};(\lambda)\}$ onto $V^\dagger|\mathbf{N};(\lambda)]$ is equivalent to that of $V_{\mathscr{F}}|\mathbf{N};(\lambda)\}$ onto $|\mathbf{N};(\lambda)]$. Neglecting V and $V^\dagger$ in Eq. (3.19) therefore merely reduces to a change of orthonormal basis in $\mathscr{F}$. Since the latter has been left arbitrary, we may incorporate $V_{\mathscr{F}}$ into the definition of $|\mathbf{N};(\lambda)\}$, which amounts to setting $V = I_{\mathscr{B}} \otimes I_{\mathscr{S}}$. Equation (3.19) finally becomes

$$X_{HP} = T^{1/2} X_D T^{-1/2}. \tag{3.20}$$

By combining it with Eq. (3.12), we also get the following relation between $X_{\bar{D}}$ and X_{HP}:

$$X_{HP} = T^{-1/2} X_{\bar{D}} T^{1/2}. \tag{3.21}$$

We conclude that the operator T, that relates both Dyson representations of a given operator, also determines, through its square root, the links between any of them and the HP representation of the same operator.

IV. BOSON REPRESENTATIONS OF THE Sp(2d,R) ALGEBRA

The purpose of the present section is to apply the results demonstrated in Sec. III to the case of the Sp(2d,R) generators.

The matrix differential operators representing the latter in both the Perelomov and Barut–Girardello PCS representations were determined previously, and are given in Eqs. (6.11) and (6.21) of Ref. 9. Proceeding as explained in the previous section, we obtain from them the following two Dyson representations of the Sp(2d,R) algebra:

$$(D^\dagger_{ij})_{\bar{D}} = \sum_k (\bar{a}^\dagger_{ik} \otimes \mathring{C}_{jk} + \bar{a}^\dagger_{jk} \otimes \mathring{C}_{ik}) + [(\bar{\mathbf{a}}^\dagger \bar{\mathbf{a}} + n - d - 1)\bar{\mathbf{a}}^\dagger]_{ij} \otimes I_{\mathscr{S}}, \tag{4.1a}$$

$$(D_{ij})_{\bar{D}} = \bar{a}_{ij} \otimes I_{\mathscr{S}}, \tag{4.1b}$$

$$(E_{ij})_{\bar{D}} = (C_{ij})_{\bar{D}} + (n/2)\delta_{ij} I_{\mathscr{B}} \otimes I_{\mathscr{S}},$$
$$(C_{ij})_{\bar{D}} = (\bar{\mathbf{a}}^\dagger \bar{\mathbf{a}})_{ij} \otimes I_{\mathscr{S}} + I_{\mathscr{B}} \otimes \mathring{C}_{ij}, \tag{4.1c}$$

and

$$(D^\dagger_{ij})_D = \bar{a}^\dagger_{ij} \otimes I_{\mathscr{S}}, \tag{4.2a}$$

$$(D_{ij})_D = \sum_k (\bar{a}_{ik} \otimes \mathring{C}_{kj} + \bar{a}_{jk} \otimes \mathring{C}_{ki}) + [\bar{\mathbf{a}}(\bar{\mathbf{a}}^\dagger \bar{\mathbf{a}} + n - d - 1)]_{ij} \otimes I_{\mathscr{S}}, \tag{4.2b}$$

$$(E_{ij})_D = (C_{ij})_D + (n/2)\delta_{ij} I_{\mathscr{B}} \otimes I_{\mathscr{S}},$$
$$(C_{ij})_D = (\bar{\mathbf{a}}^\dagger \bar{\mathbf{a}})_{ij} \otimes I_{\mathscr{S}} + I_{\mathscr{B}} \otimes \mathring{C}_{ij}, \tag{4.2c}$$

where we have explicitly exhibited the direct product structure of the operators. In Eqs. (4.1) and (4.2), $\mathring{C}_{ij}$ denotes the U(d)-spin operators, i.e., the U(d) generators restricted to $\mathscr{S}$. The operators $(\bar{\mathbf{a}}^\dagger \bar{\mathbf{a}})_{ij} \otimes I_{\mathscr{S}} + I_{\mathscr{B}} \otimes \mathring{C}_{ij}$, or in short $(\bar{\mathbf{a}}^\dagger \bar{\mathbf{a}})_{ij} + \mathring{C}_{ij}$, generate a U($d$) group acting in $\mathscr{B} \otimes \mathscr{S}$, and represent the U($d$) subgroup generators C_{ij} in both Dyson representations. Note that the Hermitian conjugation property Eq. (3.8) can be directly checked on Eqs. (4.1) and (4.2).

In contrast, the intermediate PCS representation of the Sp(2d,R) generators is not explicitly known, except in the case of the $\langle(\lambda + n/2)^d\rangle$ irreps.[6,28] We may, however, try to derive their HP representation for a generic Sp(2d,R) irrep from Eq. (3.20) or (3.21), i.e., by first determining the operator T, then calculating its square root.

To obtain the operator T, or equivalently its matrix elements, two procedures are at our disposal. We can start from the definition of T, given in Eq. (2.28), and determine its matrix elements by calculating the overlap matrix $\mathbf{M}$. For such purposes, the bases $|\mathbf{N};(\lambda)\rangle$ and $|\mathbf{N};(\lambda)]$ are, however, not quite appropriate. It is more convenient to use, in $\mathscr{F}$ and $\mathscr{B} \otimes \mathscr{S}$, basis states classified according to the U(d) groups generated by C_{ij} and $(\bar{\mathbf{a}}^\dagger\bar{\mathbf{a}})_{ij}+\mathring{C}_{ij}$, respectively. Let us denote them by

$$|(([l][\lambda])\alpha[h](h)\rangle = [P_{[l]}(\mathbf{D}^\dagger)\times|(\ )\rangle]^{\alpha[h]}_{(h)} = \sum_{(l)(\lambda)}\langle[l](l),[\lambda](\lambda)|\alpha[h](h)\rangle \times P_{[l](l)}(\mathbf{D}^\dagger)|(\lambda)\rangle \tag{4.3}$$

and

$$|(([l][\lambda])\alpha[h](h)] = [P_{[l]}(\bar{\mathbf{a}}^\dagger)\times|\mathbf{0};(\ )]]^{\alpha[h]}_{(h)} = \sum_{(l)(\lambda)}\langle[l](l),[\lambda](\lambda)|\alpha[h](h)\rangle \times P_{[l](l)}(\bar{\mathbf{a}}^\dagger)|\mathbf{0};(\lambda)], \tag{4.4}$$

where $P_{[l](l)}(\bar{\mathbf{a}}^\dagger)$ is a polynomial in the boson creation operators $\bar{a}^\dagger_{ij}$, characterized by a definite U(d) irrep $[l]=[l_1\cdots l_d]$, and a given row (l); $P_{[l](l)}(\mathbf{D}^\dagger)$ is the same polynomial function as $P_{[l](l)}(\bar{\mathbf{a}}^\dagger)$ but with $\bar{a}^\dagger_{ij}$ replaced by $D^\dagger_{ij}$; the symbol $\langle\ ,\ |\ \rangle$ is a U(d) Wigner coefficient; and α distinguishes between repeated irreps $[h]$ in the reduction of the product representation $[l]\times[\lambda]$. We assume that $P_{[l](l)}(\bar{\mathbf{a}}^\dagger)$ and α are chosen in such a way that the states defined in Eq. (4.4) are orthonormal. However, those of Eq. (4.3) are in general only orthogonal with respect to $[h]$ and (h). Let us denote by $\mathbf{M}^{[h]}$ the submatrix of their overlap matrix, corresponding to a given U(d) irrep $[h]$ and a given row (h), i.e.,

$$M^{[h]}_{[l']\alpha',[l]\alpha} = \langle(([l'][\lambda])\alpha'[h](h)|(([l][\lambda])\alpha[h](h)\rangle. \tag{4.5}$$

In terms of these U(d)-coupled states, the definition of T may be rewritten as

$$[(([l'][\lambda])\alpha'[h](h)|T|(([l][\lambda])\alpha[h](h)] = M^{[h]}_{[l']\alpha',[l]\alpha}. \tag{4.6}$$

The advantage of U(d)-coupled bases is therefore that T is already diagonal with respect to $[h]$ and (h), and that its matrix elements are independent of (h).

Alternatively, T can be determined from Eq. (3.12) relating the Dyson representations $X_{\rm D}$ and $X_{\bar{\rm D}}$ of the same operator X, and the fact that for the Sp($2d,R$) generators the latter are explicitly known. By taking $X=E_{ij}$ or $D^\dagger_{ij}$, we obtain the following two equations for T:

$$[T,\bar{\mathbf{a}}^\dagger\bar{\mathbf{a}}+\mathring{\mathbf{C}}]=0 \tag{4.7}$$

and

$$T\bar{\mathbf{a}}^\dagger T^{-1}=\bar{\mathbf{a}}^\dagger\tilde{\mathring{\mathbf{C}}}+\mathring{\mathbf{C}}\bar{\mathbf{a}}^\dagger+(\bar{\mathbf{a}}^\dagger\bar{\mathbf{a}}+n-d-1)\bar{\mathbf{a}}^\dagger, \tag{4.8}$$

where use is made of a compact matrix notation. Owing to the Hermiticity property (3.8), the remaining generators D_{ij} do not give rise to an independent condition for T. Equation (4.7) tells us that T is invariant under the U(d) group in $\mathscr{B}\otimes\mathscr{S}$, generated by the operators $(\bar{\mathbf{a}}^\dagger\bar{\mathbf{a}})_{ij}+\mathring{C}_{ij}$. To solve Eq. (4.8), it is therefore again convenient to use the U(d)-coupled basis states defined in Eq. (4.4). By taking the matrix elements of Eq. (4.8) between two such states, we obtain recursion relations for the matrix elements of T that can be solved in terms of one of them, the latter fixing the normalization of T. This was actually the procedure that we implemented in Ref. 6 to determine the HP representation of the Sp($2d,R$) algebra when $\lambda_1=\cdots=\lambda_d=\lambda$.[28] In such a case, $\mathring{C}_{ij}$ reduces to $\lambda\delta_{ij}I_{\mathscr{S}}$, so that Eq. (4.8) becomes

$$T\bar{\mathbf{a}}^\dagger T^{-1}=(\bar{\mathbf{a}}^\dagger\bar{\mathbf{a}}+n+2\lambda-d-1)\bar{\mathbf{a}}^\dagger, \tag{4.9}$$

which is indeed equivalent to Eq. (6.13) of Ref. 6. An extension of the latter work to a generic irrep $\langle\lambda_3+n/2,\lambda_2+n/2,\lambda_1+n/2\rangle$ of Sp($6,R$) was recently carried out by Rowe,[15] who derived an equation similar to Eq. (4.8) [cf. Eq. (2.13) of Ref. 15].

Once T has been determined by either of the two above-described methods, it remains to calculate its square root $T^{1/2}$ to be used in Eqs. (3.20) or (3.21). We first note that from Eq. (4.7), it results that

$$[T^{1/2},\bar{\mathbf{a}}^\dagger\bar{\mathbf{a}}+\mathring{\mathbf{C}}]=0. \tag{4.10}$$

Hence, the HP representation of E_{ij} coincides with its two Dyson representations

$$(E_{ij})_{\rm HP}=(E_{ij})_{\rm D}=(E_{ij})_{\bar{\rm D}} = [(\bar{\mathbf{a}}^\dagger\bar{\mathbf{a}})_{ij}+(n/2)\delta_{ij}I_{\mathscr{B}}]\otimes I_{\mathscr{S}}+I_{\mathscr{B}}\otimes\mathring{C}_{ij}. \tag{4.11}$$

Equations (3.20) or (3.21) will therefore be used only for $X=D^\dagger_{ij}$, since the HP representation of D_{ij} will then follow by Hermitian conjugation.

When $\lambda_1=\cdots=\lambda_d=\lambda$, the calculation of $T^{1/2}$ is an easy task since from Eq. (4.6), T is then diagonal in the U(d)-coupled basis (4.4). There is indeed only one irrep $[l]$ for any given $[h]$, namely that corresponding to $l_i=h_i-\lambda$, $i=1,\ldots,d$, and moreover there is no need for additional quantum numbers α, so that $\mathbf{M}^{[h]}$ is one dimensional, and therefore trivially diagonal. This explains why in Ref. 6 we were able to express the HP representation of the Sp($2d,R$) algebra in an analytic and compact form when $\lambda_1=\cdots=\lambda_d=\lambda$.[28]

In contrast, when $\lambda_1,\ldots,\lambda_d$ are not all equal, except in some very special cases, T is no longer diagonal in the U(d)-coupled basis. We have then first to diagonalize T, since from the equation

$$T=ADA^\dagger, \tag{4.12}$$

where D is diagonal and A unitary, we can then obtain $T^{1/2}$ as

$$T^{1/2}=AD^{1/2}A^\dagger. \tag{4.13}$$

The diagonalization of T, however, requires the solution of some algebraic equations, which in general are of a higher than four degree. Since the latter can only be performed numerically, we conclude that for a generic irrep $\langle\lambda_d+n/2,\ldots,\lambda_1+n/2\rangle$ of Sp($2d,R$), the HP representation of the Sp($2d,R$) algebra cannot be written in an analytic and compact form as was the case for $\langle(\lambda+n/2)^d\rangle$.

Nevertheless, as we shall proceed to show in the next section, explicit, though noncompact expressions can be obtained for the HP representation by resorting to a method well known in boson expansion theories, and due to Marumori, Yamamura, and Tokunaga (MYT).[34]

V. HOLSTEIN–PRIMAKOFF REPRESENTATION OF THE Sp($2d,R$) ALGEBRA

The purpose of the present section is to present the basic ideas of the MYT method in the simple case where uncoupled basis states are used, and to apply them to the determination of the Sp($2d,R$) algebra HP representation.

As shown in Sec. III, the HP representation $X_{\rm HP}$ of an operator X is given by Eq. (3.6c) in terms of the operators $\breve{U}$ and $\breve{U}^\dagger$, defined in Eqs. (2.23c) and (2.25c), respectively. Hence, it can be written as

$$X_{\rm HP} = \sum_{\mathbf{N}(\lambda)} \sum_{\mathbf{N}'(\lambda')} |\mathbf{N}';(\lambda')]\{\mathbf{N}';(\lambda')|X|\mathbf{N};(\lambda)\}[\mathbf{N};(\lambda)|, \tag{5.1}$$

or equivalently

$$X_{\rm HP} = \sum_{\mathbf{N}(\lambda)} \sum_{\mathbf{N}'(\lambda')} \{\mathbf{N}';(\lambda')|X|\mathbf{N};(\lambda)\} \times F_{\mathbf{N}'}(\bar{\mathbf{a}}^\dagger)|0]\,[0|F_{\mathbf{N}}(\bar{a}) \otimes |(\lambda')]\,[(\lambda)|, \tag{5.2}$$

where we have used Eqs. (2.18) and (2.19). The operator $X_{\rm HP}$ is therefore a linear combination of products of operators acting in $\mathscr{B}$, $F_{\mathbf{N}'}(\bar{\mathbf{a}}^\dagger)|0]\,[0|F_{\mathbf{N}}(\bar{\mathbf{a}})$, by operators acting in $\mathscr{S}$

$$P^{[\lambda]}_{(\lambda')(\lambda)} = |(\lambda')]\,[(\lambda)|. \tag{5.3}$$

We shall first find explicit expressions for both operators, then calculate the coefficients of the linear combinations, i.e., the matrix elements $\{\mathbf{N}';(\lambda')|X|\mathbf{N};(\lambda)\}$.

Let us start with the boson operator $F_{\mathbf{N}'}(\bar{\mathbf{a}}^\dagger)|0]\,[0|F_{\mathbf{N}}(\bar{\mathbf{a}})$. The key technique of the MYT method consists in expanding the vacuum projector $|0][0|$ into boson creation and annihilation operators. For such purposes, let us first establish the two following auxiliary equations:

$$F_{\mathbf{N}}(\bar{\mathbf{a}}^\dagger)F_{\mathbf{N}'}(\bar{\mathbf{a}}^\dagger) = \left[\prod_{i<j}\binom{N_{ij}+N'_{ij}}{N_{ij}}\right]^{1/2} F_{\mathbf{N}+\mathbf{N}'}(\bar{\mathbf{a}}^\dagger) \tag{5.4}$$

and

$$F_{\mathbf{N}'}(\bar{\mathbf{a}})F_{\mathbf{N}}(\bar{\mathbf{a}}^\dagger)|0] = \left[\prod_{i<j}\binom{N_{ij}}{N'_{ij}}\right]^{1/2} F_{\mathbf{N}-\mathbf{N}'}(\bar{\mathbf{a}}^\dagger)|0], \tag{5.5}$$

where $\binom{n}{k}$ denotes a binomial coefficient. Equation (5.4) directly follows from the definition of $F_{\mathbf{N}}$, given in Eq. (2.4). To prove Eq. (5.5), let us rewrite its left-hand side as follows:

$$F_{\mathbf{N}'}(\bar{\mathbf{a}})F_{\mathbf{N}}(\bar{\mathbf{a}}^\dagger)|0] = \sum_{\mathbf{N}''} F_{\mathbf{N}''}(\bar{\mathbf{a}}^\dagger)|0]\,[0|F_{\mathbf{N}''}(\bar{\mathbf{a}})F_{\mathbf{N}'}(\bar{\mathbf{a}})F_{\mathbf{N}}(\bar{\mathbf{a}}^\dagger)|0], \tag{5.6}$$

by using the unity resolution relation

$$\sum_{\mathbf{N}''} |\mathbf{N}'']\,[\mathbf{N}''| = I_{\mathscr{B}}. \tag{5.7}$$

By applying the Hermitian conjugate of Eq. (5.4) to $F_{\mathbf{N}''}(\bar{\mathbf{a}})F_{\mathbf{N}'}(\bar{\mathbf{a}})$, and the orthonormality of the states $|\mathbf{N}]$, the right-hand side of Eq. (5.6) is transformed into that of Eq. (5.5), thus completing the proof of the latter.

It is now straightforward to demonstrate that the vacuum projector can be expressed as

$$|0]\,[0| = \sum_{\mathbf{N}''} (-)^{\Sigma_{i<j}N''_{ij}} F_{\mathbf{N}''}(\bar{\mathbf{a}}^\dagger)F_{\mathbf{N}''}(\bar{\mathbf{a}}). \tag{5.8}$$

The proof of Eq. (5.8) is based on the fact that both sides of this relation have the same action on an arbitrary state $|\mathbf{N}]$ of $\mathscr{B}$. By successively using Eqs. (5.5) and (5.4), we indeed obtain the following result:

$$\sum_{\mathbf{N}''} (-)^{\Sigma_{i<j}N''_{ij}} F_{\mathbf{N}''}(\bar{\mathbf{a}}^\dagger)F_{\mathbf{N}''}(\bar{\mathbf{a}})|\mathbf{N}] = \left\{\prod_{i<j}\left[\sum_{N''_{ij}}(-)^{N''_{ij}}\binom{N_{ij}}{N''_{ij}}\right]\right\}F_{\mathbf{N}}(\bar{\mathbf{a}}^\dagger)|0]. \tag{5.9}$$

The latter can be rewritten as

$$\sum_{\mathbf{N}''} (-)^{\Sigma_{i<j}N''_{ij}} F_{\mathbf{N}''}(\bar{\mathbf{a}}^\dagger)F_{\mathbf{N}''}(\bar{\mathbf{a}})|\mathbf{N}] = \delta_{\mathbf{N},0}|0] = |0]\,[0|\mathbf{N}], \tag{5.10}$$

by taking the well-known combinatorial relation

$$\sum_k (-)^k\binom{n}{k} = \delta_{n,0} \tag{5.11}$$

into account.

When we replace the vacuum projector by the right-hand side of Eq. (5.8) in the boson operator $F_{\mathbf{N}'}(\bar{\mathbf{a}}^\dagger)|0]\,[0|F_{\mathbf{N}}(\bar{\mathbf{a}})$, and apply Eq. (5.4) as well as its Hermitian conjugate, we finally obtain for the boson operator the following expansion:

$$F_{\mathbf{N}'}(\bar{\mathbf{a}}^\dagger)|0]\,[0|F_{\mathbf{N}}(\bar{\mathbf{a}}) = \sum_{\mathbf{N}''}\left[\prod_{i<j}(-)^{N''_{ij}}\binom{N'_{ij}+N''_{ij}}{N''_{ij}}^{1/2}\binom{N_{ij}+N''_{ij}}{N''_{ij}}^{1/2}\right] \times F_{\mathbf{N}'+\mathbf{N}''}(\bar{\mathbf{a}}^\dagger)F_{\mathbf{N}+\mathbf{N}''}(\bar{\mathbf{a}}). \tag{5.12}$$

In the same way, the operators $P^{[\lambda]}_{(\lambda')(\lambda)}$, acting in $\mathscr{S}$, can be written in terms of the U(d)-spin operators $\mathring{C}_{ij}$. Detailed expressions for $P^{[\lambda]}_{(\lambda')(\lambda)}(\mathring{\mathbf{C}})$ are given in the Appendix. We conclude that $X_{\rm HP}$ can be expanded as follows:

$$X_{\rm HP} = \sum_{\mathbf{N}(\lambda)}\sum_{\mathbf{N}'(\lambda')}\left\{\sum_{\mathbf{N}''}\left[\prod_{i<j}(-)^{N''_{ij}}\binom{N'_{ij}}{N''_{ij}}^{1/2}\binom{N_{ij}}{N''_{ij}}^{1/2}\right] \times\{\mathbf{N}'-\mathbf{N}'';(\lambda')|X|\mathbf{N}-\mathbf{N}'';(\lambda)\}\right\} \times F_{\mathbf{N}'}(\bar{\mathbf{a}}^\dagger)F_{\mathbf{N}}(\bar{\mathbf{a}})\otimes P^{[\lambda]}_{(\lambda')(\lambda)}(\mathring{\mathbf{C}}). \tag{5.13}$$

It now remains to calculate the matrix elements of X between orthonormal basis states. From Eqs. (2.9) and (2.10), the latter can be expressed as

$$\{\mathbf{N}';(\lambda')|X|\mathbf{N};(\lambda)\} = \sum_{\bar{\mathbf{N}}(\bar{\lambda})}\sum_{\bar{\mathbf{N}}'(\bar{\lambda}')}(\mathbf{R}^{-1})_{\mathbf{N}'(\lambda'),\bar{\mathbf{N}}'(\bar{\lambda}')} \times\langle\bar{\mathbf{N}}';(\bar{\lambda}')|X|\bar{\mathbf{N}};(\bar{\lambda})\rangle R_{\bar{\mathbf{N}}(\bar{\lambda}),\mathbf{N}(\lambda)}, \tag{5.14}$$

in terms of the matrix $\mathbf{R}$ defining the orthonormal basis, and of its inverse $\mathbf{R}^{-1}$. The matrix elements of X between the nonorthogonal basis states $|\bar{\mathbf{N}};(\bar{\lambda})\rangle$ and their dual ones $|\bar{\mathbf{N}}';(\bar{\lambda}'))$ can be determined from the definition of X, and Eqs. (2.3) and (2.6). For the generators $D^\dagger_{ij}$, we get the following result:

$$(\bar{\mathbf{N}}';(\bar{\lambda}')|D^\dagger_{ij}|\bar{\mathbf{N}};(\bar{\lambda})\rangle = \left[\sum_{k<l}\delta_{(kl),(ij)}(\bar{N}_{kl}+1)^{1/2}\delta_{\bar{\mathbf{N}}',\bar{\mathbf{N}}+\mathbf{e}^{(kl)}}\right]\delta_{(\bar{\lambda}')(\bar{\lambda})}, \tag{5.15}$$

where

$$\delta_{(kl),(ij)} = (1+\delta_{kl})^{-1/2}[\delta_{ki}\delta_{lj}+\delta_{kj}\delta_{li}] \tag{5.16}$$

and

$$e^{(kl)}_{ij} = \delta_{ki}\delta_{lj}, \quad i\leqslant j. \tag{5.17}$$

The combination of Eq. (5.13) with Eqs. (5.14) and (5.15) finally leads to the sought-for HP representation of $D^\dagger_{ij}$

$$(D^\dagger_{ij})_{\mathrm{HP}} = \sum_{\mathbf{N}(\lambda)} \sum_{\mathbf{N}'(\lambda')} \Bigg\{ \sum_{\mathbf{N}''} \Bigg[\prod_{i<j} (-)^{N''_{ij}} \binom{N'_{ij}}{N''_{ij}}^{1/2} \binom{N_{ij}}{N''_{ij}}^{1/2} \Bigg]$$
$$\times \Bigg[\sum_{\overline{\mathbf{N}}(\bar\lambda)} \sum_{k<l} \delta_{(kl),(ij)} (\overline{N}_{kl} + 1)^{1/2}$$
$$\times (\mathbf{R}^{-1})_{\mathbf{N}' - \mathbf{N}''(\lambda'), \overline{\mathbf{N}} + \mathbf{e}^{(kl)}(\bar\lambda)} R_{\overline{\mathbf{N}}(\bar\lambda), \mathbf{N} - \mathbf{N}''(\lambda)} \Bigg] \Bigg\}$$
$$\times F_{\mathbf{N}'}(\bar{\mathbf{a}}^\dagger) F_{\mathbf{N}}(\bar{\mathbf{a}}) \otimes P^{[\lambda]}_{(\lambda')(\lambda)}(\mathring{\mathbf{C}}). \tag{5.18}$$

In conclusion, Eqs. (3.9), (4.11), and (5.18) show that the HP representation of any Sp($2d,R$) generator can be expressed as an explicit expansion in terms of ν pairs of boson creation and annihilation operators, and of the U(d)-spin operators. In such a formulation, the difficulties resulting from the nonorthogonality of the basis $|\mathbf{N};(\lambda)\rangle$ are hidden in the coefficients of the expansion, which in general will have to be calculated numerically. As a final point, let us note that the treatment presented in this section can be extended in a straightforward way to the case where U(d)-coupled states are used.

VI. THE RESTRICTED dn-DIMENSIONAL HARMONIC OSCILLATOR HAMILTONIAN AND ITS SYMMETRY GROUP

In some physical applications, such as the Sp($6,R$) nuclear collective model to be discussed in the next section, one has to deal with a dn-dimensional harmonic oscillator Hamiltonian restricted to the representation space $\mathscr{F}$ of a single Sp($2d,R$) irrep $\langle \lambda_d + n/2,...,\lambda_1 + n/2\rangle$. Taking units in which $\hbar$ and the oscillator frequency ω are equal to one, this harmonic oscillator Hamiltonian can be written in terms of the boson creation and annihilation operators η_{is} and ξ_{is}, $i = 1, ..., d$, $s = 1, ..., n$, introduced in Sec. II, as

$$H^{\mathrm{osc}} = \sum_{i=1}^{d} \sum_{s=1}^{n} \eta_{is}\xi_{is} + \frac{1}{2} dn. \tag{6.1}$$

When H^{osc} acts in the whole space of states built from the dn operators η_{is}, its symmetry group is well known to be U(dn).[41] However, when H^{osc} is restricted to $\mathscr{F}$, the degeneracy of its eigenvalues is considerably reduced, and the symmetry group responsible for this residual degeneracy is different from U(dn). The purpose of the present section is to determine this symmetry group. In the $d = 3$ case, a preliminary account of the following results was given in Ref. 8, and the same problem was also recently discussed by Castaños and Frank,[12] and Moshinsky.[13]

By comparing Eq. (6.1) with Eqs. (2.1) and (2.2), we note that

$$H^{\mathrm{osc}} = \sum_{i=1}^{d} E_{ii} = \sum_{i=1}^{d} C_{ii} + \frac{1}{2} dn \tag{6.2}$$

is just the first-order Casimir operator of the U(d) subgroup of Sp($2d,R$). All the basis states $|\mathbf{N};(\lambda)\rangle$ of $\mathscr{F}$, corresponding to a given value of $N = \Sigma_{i<j} N_{ij}$, are eigenstates of H^{osc} with the same eigenvalue $N + \frac{1}{2} dn$. Since they can be put into one-to-one correspondence with the extended boson states $|\mathbf{N};(\lambda)]$, for which $N = \Sigma_{i<j} N_{ij}$, the degeneracy of the eigenvalue $N + \frac{1}{2} dn$ of the restricted harmonic oscillator Hamiltonian is equal to the number of such extended boson states. The boson states $|\mathbf{N}]$, built from ν boson creation operators $\bar{a}^\dagger_{ij}$, and subject to the condition $N = \Sigma_{i<j} N_{ij}$, span the representation space of the symmetric irrep $[\mathrm{N}]$ of a U(ν) group. The residual degeneracy of the eigenvalue $N + \frac{1}{2} dn$ is therefore equal to the product of the dimension, dim$[N]$, of the U(ν) irrep $[N]$ by the dimension Λ of the U(d) irrep $[\lambda_1 \cdots \lambda_d]$. From this result, we may expect that the restricted harmonic oscillator symmetry group is the direct product group U(ν)$\times$U(d).

To prove this assertion, it remains to explicitly carry out the restriction of H^{osc} to $\mathscr{F}$. This can be done in an elementary way by considering the HP representation of H^{osc}. From Eq. (4.11), it is given by

$$H^{\mathrm{osc}}_{\mathrm{HP}} = [\mathrm{tr}(\bar{\mathbf{a}}^\dagger \bar{\mathbf{a}}) + \tfrac{1}{2} d (n + 2\lambda) I_{\mathscr{B}}] \otimes I_{\mathscr{S}}, \tag{6.3}$$

where

$$\lambda = d^{-1} \sum_{i=1}^{d} \lambda_i. \tag{6.4}$$

Let us now introduce in $\mathscr{B}$ a harmonic oscillator Hamiltonian, whose frequency is twice that of H^{osc},

$$H^{\mathrm{osc}}_{\mathscr{B}} = 2 \sum_{i<j} \left(a^\dagger_{ij} a_{ij} + \frac{1}{2} I_{\mathscr{B}} \right)$$
$$= \mathrm{tr}(\bar{\mathbf{a}}^\dagger \bar{\mathbf{a}}) + \nu I_{\mathscr{B}}. \tag{6.5}$$

Equation (6.3) can then be rewritten as

$$H^{\mathrm{osc}}_{\mathrm{HP}} = [H^{\mathrm{osc}}_{\mathscr{B}} + \tfrac{1}{2} d (n + 2\lambda - d - 1) I_{\mathscr{B}}] \otimes I_{\mathscr{S}}. \tag{6.6}$$

Apart from a constant term, the restricted harmonic oscillator Hamiltonian is therefore mapped onto the direct product of a ν-dimensional harmonic oscillator Hamiltonian in $\mathscr{B}$ by the unit operator in $\mathscr{S}$.

It is now obvious that the symmetry group of $H^{\mathrm{osc}}_{\mathrm{HP}}$ is the direct product of the symmetry group of $H^{\mathrm{osc}}_{\mathscr{B}}$ by the U(d)-spin group. The former is the U(ν) group generated by the operators $a^\dagger_{ij} a_{kl}$, $1 \leqslant i \leqslant j \leqslant d$, $1 \leqslant k \leqslant l \leqslant d$, while the latter is generated by $\mathring{C}_{ij}$, $i, j = 1, ..., d$. More precisely, if we eliminate the U(d)-spin group first-order Casimir operator, which reduces to the constant $d\lambda$, we obtain the direct product group U(ν)$\times$SU(d).

As it was discussed by Moshinsky for the $d = 3$ case in Ref. 13, it would be of great interest to obtain explicit expressions of the symmetry group generators, or equivalently of $a^\dagger_{ij}, a_{ij}$, and $\mathring{C}_{ij}$, in terms of the Sp($2d,R$) generators. This could be done in principle by inverting the HP representation. From Eqs. (3.6c) and (2.26c), the operator X in $\mathscr{F}$ corresponding to an operator $X_{\mathrm{HP}} = a^\dagger_{ij} \otimes I_{\mathscr{S}}, a_{ij} \otimes I_{\mathscr{S}}$, or $I_{\mathscr{B}} \otimes \mathring{C}_{ij}$ in $\mathscr{B} \otimes \mathscr{S}$, is given by the following equation:

$$X = \sum_{\mathbf{N}(\lambda)} \sum_{\mathbf{N}'(\lambda')} |\mathbf{N}';(\lambda')\}\, [\mathbf{N}';(\lambda')| X_{\mathrm{HP}} |\mathbf{N};(\lambda)]\, \{\mathbf{N};(\lambda)|, \tag{6.7}$$

in a U(d)-uncoupled basis, or a similar equation in a U(d)-coupled one. The obtaining of an explicit expression for X therefore amounts to rewriting $|\mathbf{N}';(\lambda')\}\{\mathbf{N};(\lambda)|$, or its counterpart in the U(d)-coupled basis, in terms of the Sp($2d,R$) generators. Due to the complicated form of the orthonormal states $\{N;(\lambda)\}$, as compared with that of the boson states $|N;(\lambda)]$, the inversion of the HP representation is much more difficult than the obtaining of the HP representation itself, and it still remains unsolved. In this connection, it is worth noting that the inversion of the Dyson representations, proposed by Moshinsky in Ref. 13, seems easier because such representations admit the compact forms (4.1) and (4.2). However, the functions of the $Sp(2d,R)$ generators so determined would have the right commutation relations, but not the required Hermiticity properties.

VII. APPLICATION TO THE Sp(6,R) NUCLEAR COLLECTIVE MODEL

In this concluding section, we wish to point out how the results of the previous sections can be applied to the $Sp(6,R)$ nuclear collective model.[42–44] In such a model, we have to deal with a translationally invariant A nucleon system, described in terms of its $3n$ relative Jacobi coordinates x_{is}, $i = 1, 2, 3$, $s = 1, ..., n = A - 1$, and their conjugate momenta $p_{is} = -i\partial/\partial x_{is}$. To these coordinates and momenta, we associate $3n$ boson creation and annihilation operators η_{is} and ξ_{is}, given in appropriate units by the following relations:

$$\eta_{is} = 1/\sqrt{2}\ (x_{is} - ip_{is}),\ \xi_{is} = 1/\sqrt{2}\ (x_{is} + ip_{is}). \tag{7.1}$$

In terms of the latter, the $Sp(6,R)$ generators are defined by Eq. (2.1), where we now set $d = 3$.

Much of the physical interest of the $Sp(6,R)$ collective model comes from the fact that it is a natural generalization of Elliott's SU(3) model,[45] including many major oscillator shell excitations. For light nuclei, the ground-state band of Elliott's model has a definite number N of quanta, and belongs to a definite SU(3) irrep $(\lambda\mu)$, or equivalently to a definite U(3) irrep $[\lambda_1\lambda_2\lambda_3]$. The representation space of the $Sp(6,R)$ irrep $\langle\lambda_3 + n/2, \lambda_2 + n/2, \lambda_1 + n/2\rangle$, based upon the lowest weight state of $[\lambda_1\lambda_2\lambda_3]$, then contains both the ground-state band and the bands obtained therefrom by applying the $2\hbar\omega$ collective excitation operators $D^\dagger_{ij}$ any number of times. For medium and heavy nuclei, various procedures, consistent with the Pauli principle, have been described to select the ground-state band, and consequently the $Sp(6,R)$ irrep based on it.[46–49] It turns out that for closed-shell nuclei, the $Sp(6,R)$ irrep so chosen is characterized by equal values of λ_1, λ_2, and λ_3, while for open-shell nuclei it corresponds to not all equal values of them.

The $Sp(6,R)$ collective model can also be formulated in terms of an $O(n)$ group if we realize that $Sp(6,R)$ is but a subgroup of the $Sp(6n,R)$ group generated by

$$\begin{aligned} D^\dagger_{is,jt} &= \eta_{is}\eta_{jt},\ D_{is,jt} = \xi_{is}\xi_{jt}, \\ E_{is,jt} &= \eta_{is}\xi_{jt} + \tfrac{1}{2}\delta_{ij}\delta_{st}, \end{aligned} \tag{7.2}$$

where $i,j = 1, 2, 3$, and $s, t = 1, ..., n$. For the latter, we may consider the following group chain[1]:

$$Sp(6n,R) \supset Sp(6,R) \times O(n), \tag{7.3}$$

where the $O(n)$ generators are given by

$$\Lambda_{st} = -i(C_{st} - C_{ts}), \tag{7.4}$$

in terms of the generators

$$C_{st} = \sum_{i=1}^{3} \eta_{is}\xi_{it} \tag{7.5}$$

of a $U(n)$ group. From the complementarity relationship between $Sp(6,R)$ and $O(n)$,[1,50,51] it results that the basis states of the $Sp(6,R)$ irrep $\langle\lambda_3 + n/2, \lambda_2 + n/2, \lambda_1 + n/2\rangle$ transform under a definite $O(n)$ irrep, characterized by $(\lambda_1\lambda_2\lambda_3)$. The $Sp(6,R)$ model, as defined above, is therefore completely equivalent[5–7,10,13] to the approach followed by various authors,[46,48] wherein collective states are constrained to a definite $O(n)$ irrep $(\lambda_1\lambda_2\lambda_3)$.

In view of the great phenomenological success encountered by both the Bohr–Mottelson model[52] and the interacting boson approximation (IBA) introduced by Arima and Iachello,[53] there have been many attempts to justify them microscopically. This led various authors to look for boson representations or approximations in the framework of the $Sp(6,R)$ model.[5–17]

As it is customary in nuclear physics, it is convenient to consider an oscillator shell model Hamiltonian $H^{\rm osc}$ [given in Eq. (6.1) wherein $d = 3$], which will provide us with a complete set of states.[10] In Ref. 6, we showed that, when it is restricted to an $Sp(6,R)$ irrep $\langle(\lambda + n/2)^3\rangle$,[28] as it is the case for closed-shell nuclei, $H^{\rm osc}$ can be mapped onto a six-dimensional boson oscillator Hamiltonian $H^{\rm osc}_{\mathscr{B}}$ with double frequency, and has therefore a U(6) symmetry group. Moreover, we were able to write down explicit and compact expressions for the boson operators

$$\bar{\mathbf{a}}^\dagger = [\mathbf{C} + (n-4)\mathbf{I}]^{-1/2}\mathbf{D}^\dagger,\ \bar{\mathbf{a}} = \mathbf{D}[\mathbf{C} + (n-4)\mathbf{I}]^{-1/2}, \tag{7.6}$$

in terms of the $Sp(6,R)$ generators, and hence for the symmetry group generators $a^\dagger_{ij}a_{kl}$.

When $H^{\rm osc}$ is restricted to an $Sp(6,R)$ irrep corresponding to not all equal λ_1, λ_2, λ_3 values, i.e., for open-shell nuclei, the results of Sec. VI, for the special case $d = 3$ and $\nu = 6$, show that $H^{\rm osc}$ can still be mapped onto $H^{\rm osc}_{\mathscr{B}}$, but that there appears an additional SU(3) symmetry group, the SU(3)-spin group. The full symmetry group is now the direct product group $U(6) \times SU(3)$. There is another essential difference between the open-shell case and the closed-shell one. Although, as shown in Secs. V and VI, in the former case it is possible to expand the $Sp(6,R)$ generators in terms of the boson operators and the SU(3)-spin operators, the converse is not true in general; *a fortiori*, there are no compact expressions similar to Eq. (7.6). The origin of these difficulties is to be found in the lack of a simple orthonormal basis in the $Sp(6,R)$ irrep representation space when $\lambda_1, \lambda_2, \lambda_3$ are not all equal. In the $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$ case, U(3)-coupled states form an orthonormal basis because the chain $Sp(6,R) \supset U(3)$ is then multiplicity-free. This property, however, fails to extend to the case where λ_1, λ_2, and λ_3 are not all equal.

In a recent paper,[12] Castaños and Frank determined the symmetry group of the restricted $3n$-dimensional harmonic oscillator, in the limit of a very large nucleon number A, claiming that it is the semidirect product group $U(6) \wedge SU(3)$. Since their conclusion seems to contradict our result, valid for arbitrary A, a few comments are needed. Comparison between both works is made difficult by the fact that Castaños and Frank use an entirely different, and less tractable approach, based upon the Dzublik–Zickendraht transformation[54,55] and a contraction of the Sp(6,R) generators. As in Eq. (4.11), their U(3) subgroup generators, in the $A \to \infty$ limit, separate into two terms. The first ones are the generators $(\bar{\mathbf{a}}^\dagger \bar{\mathbf{a}})_{ij}$ of the U(3) subgroup of U(6), as in the present work, while the second ones are rather complicated expressions, depending, among others, upon the vortex spin. In spite of this separation, they take the sums of the two terms for the generators of their U(3) symmetry group. As a consequence, the latter do not commute with the U(6) generators, hence both sets of generators give rise to a semidirect product group (after eliminating the U(3) first-order Casimir operator). Had we taken the sums $(\bar{\mathbf{a}}^\dagger \bar{\mathbf{a}})_{ij} + \mathring{C}_{ij}$ for our U(3) symmetry group generators, we should have arrived at the same conclusion. An interesting open question is whether the second terms in Castaños and Frank's U(3) generators can be related to the U(3)-spin operators $\mathring{C}_{ij}$ in the $A \to \infty$ limit.

As a final point, we would like to comment on the physical significance of the U(6) symmetry arising in the Sp(6,R) model. Although the search for such a microscopic hidden symmetry was partly motivated by the success of the IBA, its physical content is rather different from that of the IBA U(6) symmetry. In the present case, in the so-called U(3)-boson limit,[14] the s and d bosons indeed describe $2\hbar\omega$ collective excitations, i.e., giant resonances, instead of the low-energy collective excitations considered in the IBA. However, it should be appreciated that, in spite of this, some recent numerical calculations suggest that the high-energy s and d bosons of the Sp(6,R) model may have a rather great influence on low-energy spectra through the admixture of many major shell excited states.[49]

ACKNOWLEDGMENT

We would like to thank M. Moshinsky for many stimulating discussions.

APPENDIX: CONSTRUCTION OF THE OPERATOR $P^{[\lambda]}_{(\lambda')(\lambda)}(\mathring{\mathbf{C}})$

The purpose of this appendix is to find the explicit form of the operator $P^{[\lambda]}_{(\lambda')(\lambda)}(\mathring{\mathbf{C}})$, defined in Eq. (5.3), in terms of the U(d)-spin operators $\mathring{C}_{ij}$.

From its definition, it follows that $P^{[\lambda]}_{(\lambda')(\lambda)}$ can be entirely characterized by its transformation property under Hermitian conjugation

$$[P^{[\lambda]}_{(\lambda')(\lambda)}]^\dagger = P^{[\lambda]}_{(\lambda)(\lambda')}, \tag{A1}$$

and its action upon the basis states of $\mathscr{S}$

$$P^{[\lambda]}_{(\lambda')(\lambda)}|(\lambda'')] = \delta_{(\lambda),(\lambda'')}|(\lambda')]. \tag{A2}$$

Let us introduce the projection operator $P_{(\lambda)}$ onto the one-dimensional subspace $\mathscr{S}_{(\lambda)}$ of $\mathscr{S}$ spanned by $|(\lambda)]$. It is defined by the two following equations:

$$[P_{(\lambda)}]^\dagger = P_{(\lambda)} \tag{A3}$$

and

$$P_{(\lambda)}|(\lambda'')] = \delta_{(\lambda),(\lambda'')}|(\lambda)]. \tag{A4}$$

Let us also consider the transition operator $T^{[\lambda]}_{(\lambda')(\lambda)}$ from $\mathscr{S}_{(\lambda)}$ to $\mathscr{S}_{(\lambda')}$, specified by the two conditions

$$[T^{[\lambda]}_{(\lambda')(\lambda)}]^\dagger = T^{[\lambda]}_{(\lambda)(\lambda')} \tag{A5}$$

and

$$[(\lambda')|T^{[\lambda]}_{(\lambda')(\lambda)}|(\lambda)] = 1. \tag{A6}$$

From Eqs. (A3)–(A6), it is then obvious that both Eqs. (A1) and (A2) are satisfied if $P^{[\lambda]}_{(\lambda')(\lambda)}$ is written as follows:

$$P^{[\lambda]}_{(\lambda')(\lambda)} = P_{(\lambda')} T^{[\lambda]}_{(\lambda')(\lambda)} P_{(\lambda)}. \tag{A7}$$

The construction of $P^{[\lambda]}_{(\lambda')(\lambda)}$ therefore amounts to that of $P_{(\lambda)}$ and $T^{[\lambda]}_{(\lambda')(\lambda)}$.

Now $P_{(\lambda)}$ can be easily built by extending to U(d) the Löwdin–Shapiro construction of SU(2) projection operators.[56,57] Denoting by $P_{[\lambda_k]_k}$ the projection operator onto the representation space of the U(k) irrep $[\lambda_k]_k \equiv [\lambda_{1k}\lambda_{2k}\cdots\lambda_{kk}]$, we obtain for $P_{(\lambda)}$ the following result:

$$P_{(\lambda)} = \prod_{k=1}^{d-1} P_{[\lambda_k]_k}, \tag{A8}$$

where

$$P_{[\lambda_k]_k} = \prod_{j=1}^{k} \left(\prod_{\bar{\phi}_j^{(k)}} \frac{\Phi_j^{(k)} - \bar{\phi}_j^{(k)}}{\phi_j^{(k)} - \bar{\phi}_j^{(k)}} \right). \tag{A9}$$

Here $\Phi_j^{(k)}$ is the jth-order Casimir operator of the U(k) subgroup of the U(d)-spin group

$$\Phi_j^{(k)} = \mathrm{tr}_{(k)} \mathring{C}^j = \sum_{i_1,\ldots,i_j=1}^{k} \mathring{C}_{i_1 i_2} \mathring{C}_{i_2 i_3} \cdots \mathring{C}_{i_j i_1}, \tag{A10}$$

$\phi_j^{(k)}$ denotes its eigenvalue corresponding to the irrep $[\lambda_k]_k$, and $\bar{\phi}_j^{(k)}$ runs over the eigenvalues associated with the U(k) irreps appearing in the reduction of the U(d) irrep $[\lambda]$. Explicit expressions of $\phi_j^{(k)}$ can be found in Ref. 58. The ordering of the operators $P_{[\lambda_k]_k}$ in Eq. (A8) does not matter since they all commute with one another.

Let us now turn to the construction of the transition operator $T^{[\lambda]}_{(\lambda')(\lambda)}$, defined in Eqs. (A5) and (A6), and factorize it into a product of operators as follows:

$$T^{[\lambda]}_{(\lambda')(\lambda)} = \left[\prod_{k=d-1}^{2} T^{[\lambda'_k]_k}_{[\lambda'_k]_{k-1}[\lambda'_{k-1}]_{k-1}} \right]^\dagger T^{[\lambda]}_{[\lambda'_{d-1}]_{d-1}[\lambda_{d-1}]_{d-1}} \times \left[\prod_{k=d-1}^{2} T^{[\lambda_k]_k}_{[\lambda_k]_{k-1}[\lambda_{k-1}]_{k-1}} \right]. \tag{A11}$$

Here $[\lambda_k]_{k-1}$ denotes the U($k-1$) irrep $[\lambda_{1k}\lambda_{2k}\cdots\lambda_{k-1,k}]$, and the symbol $\Pi^2_{k=d-1}$ means that the operator corresponding to k is on the left of the operator corresponding to $k-1$. Equations (A5) and (A6) will be satisfied provided that

$$\left[T^{[\lambda]}_{[\lambda'_{d-1}]_{d-1}[\lambda_{d-1}]_{d-1}} \right]^\dagger = T^{[\lambda]}_{[\lambda_{d-1}]_{d-1}[\lambda'_{d-1}]_{d-1}}, \tag{A12}$$

$$T^{[\lambda_k]_k}_{[\lambda_k]_{k-1}[\lambda_{k-1}]_{k-1}} |(\lambda)_{k-1}] = |(\lambda)_k], \tag{A13}$$

and

$$\left[(\lambda')_{d-1} | T^{[\lambda]}_{[\lambda'_{d-1}]_{d-1}[\lambda_{d-1}]_{d-1}} | (\lambda)_{d-1} \right] = 1, \tag{A14}$$

where $(\lambda)_k$ denotes the Gel'fand pattern obtained from (λ) by replacing the U(1), U(2), ..., U($k-1$) irreps by the highest ones, i.e.,

$$(\lambda)_k = \begin{pmatrix} \lambda_{1,d-1} & \lambda_{2,d-1} & \cdots & \lambda_{d-1,d-1} \\ \vdots & \vdots & & \vdots \\ \lambda_{1,k+1} & \lambda_{2,k+1} & \cdots & \lambda_{k+1,k+1} \\ \lambda_{1k} & \lambda_{2k} & \cdots & \lambda_{kk} \\ \lambda_{1k} & \lambda_{2k} & \cdots & \lambda_{k-1,k} \\ \vdots & \vdots & & \vdots \\ & & \lambda_{1k} & \end{pmatrix}. \tag{A15}$$

Finally, explicit expressions of both the operators

$$T^{[\lambda_k]_k}_{[\lambda_k]_{k-1}[\lambda_{k-1}]_{k-1}}$$

and

$$T^{[\lambda]}_{[\lambda'_{d-1}]_{d-1}[\lambda_{d-1}]_{d-1}},$$

fulfilling Eqs. (A12)–(A14), can be constructed from the known matrix elements of the unitary group generators.[38,59,60] For the former, we obtain

$$T^{[\lambda_k]_k}_{[\lambda_k]_{k-1}[\lambda_{k-1}]_{k-1}} = \left\{ \prod_{i=1}^{k-1} [(\lambda_{ik} - \lambda_{i,k-1})!]^{-1} \prod_{j=i+1}^{k} (\lambda_{i,k-1} - \lambda_{jk} + j - i - 1)! [(\lambda_{ik} - \lambda_{jk} + j - i - 1)!]^{-1} \right.$$
$$\left. \times \prod_{l=i+1}^{k-1} (\lambda_{ik} - \lambda_{l,k-1} + l - i)! [(\lambda_{i,k-1} - \lambda_{l,k-1} + l - i)!]^{-1} \right\}^{1/2} \prod_{m=k-1}^{1} (\mathring{C}_{mk})^{\lambda_{mk} - \lambda_{m,k-1}}. \tag{A16}$$

For the latter, we first separate the set of differences $\lambda_{k,d-1} - \lambda'_{k,d-1}$, $k = 1, ..., d-1$, into three classes: (i) the positive ones corresponding to $k = i_1, i_2, ..., i_p$; (ii) the negative ones corresponding to $k = j_1, j_2, ..., j_q$; and (iii) those equal to zero. We then introduce the symbol $[\bar{\lambda}_{d-1}]_{d-1}$ to denote the U($d-1$) irrep which differs from $[\lambda'_{d-1}]_{d-1}$ by the replacement of $\lambda'_{i_1,d-1}$, ..., $\lambda'_{i_p d-1}$ by $\lambda_{i_1,d-1}, ..., \lambda_{i_p d-1}$, or equivalently the U($d-1$) irrep obtained from $[\lambda_{d-1}]_{d-1}$ by substituting $\lambda'_{j_1,d-1}, ..., \lambda'_{j_q d-1}$ for $\lambda_{j_1,d-1}, ..., \lambda_{j_q d-1}$. The explicit expression of $T^{[\lambda]}_{[\lambda'_{d-1}]_{d-1}[\lambda_{d-1}]_{d-1}}$, is given by

$$T^{[\lambda]}_{[\lambda'_{d-1}]_{d-1}[\lambda_{d-1}]_{d-1}} = T^{[\lambda]}_{[\lambda'_{d-1}]_{d-1}[\bar{\lambda}_{d-1}]_{d-1}} \, T^{[\lambda]}_{[\bar{\lambda}_{d-1}]_{d-1}[\lambda_{d-1}]_{d-1}}, \tag{A17}$$

where

$$T^{[\lambda]}_{[\bar{\lambda}_{d-1}]_{d-1}[\lambda_{d-1}]_{d-1}} = \prod_{r=q}^{1} \left\{ \prod_{k=1}^{j_r} (\lambda_k - \lambda'_{j_r d-1} + j_r - k)! [(\lambda_k - \lambda_{j_r d-1} + j_r - k)!]^{-1} \right.$$
$$\times \prod_{l=j_r+1}^{d} (\lambda_{j_r d-1} - \lambda_l + l - j_r - 1)! [(\lambda'_{j_r d-1} - \lambda_l + l - j_r - 1)!]^{-1}$$
$$\times \prod_{s=1}^{r-1} (\lambda'_{j_s d-1} - \lambda_{j_r d-1} + j_r - j_s)! [(\lambda'_{j_s d-1} - \lambda'_{j_r d-1} + j_r - j_s)!]^{-1}$$
$$\times \prod_{\substack{m=1 \\ \neq j_1, \ldots, j_{r-1}}}^{j_r - 1} (\lambda_{m,d-1} - \lambda_{j_r d-1} + j_r - m)! [(\lambda_{m,d-1} - \lambda'_{j_r d-1} + j_r - m)!]^{-1}$$
$$\left. \times \prod_{n=j_r+1}^{d-1} (\lambda'_{j_r d-1} - \lambda_{n,d-1} + n - j_r)! [(\lambda_{j_r d-1} - \lambda_{n,d-1} + n - j_r)!]^{-1} \right\}^{1/2}$$
$$\times (\mathring{C}_{j_r d})^{\lambda'_{j_r d-1} - \lambda_{j_r d-1}} \tag{A18}$$

and

$$T^{[\lambda]}_{[\lambda'_{d-1}]_{d-1}[\bar{\lambda}_{d-1}]_{d-1}} = \left[T^{[\lambda]}_{[\bar{\lambda}_{d-1}]_{d-1}[\lambda'_{d-1}]_{d-1}} \right]^{\dagger}. \tag{A19}$$

This completes the construction of $P^{[\lambda]}_{(\lambda')(\lambda)}$ as a function of the U(d)-spin operators $\mathring{C}_{ij}$.

[1] M. Moshinsky and C. Quesne, J. Math. Phys. **12**, 1772 (1971).
[2] B. G. Wybourne, *Classical Groups for Physicists* (Wiley, New York, 1974), p. 286.
[3] G. Rosensteel and D. J. Rowe, Int. J. Theor. Phys. **16**, 63 (1977).
[4] A. U. Klimyk, J. Math. Phys. **24**, 224 (1983).
[5] J. Deenen and C. Quesne, J. Math. Phys. **23**, 878 (1982).
[6] J. Deenen and C. Quesne, J. Math. Phys. **23**, 2004 (1982).
[7] J. Deenen and C. Quesne, J. Math. Phys. **25**, 1638 (1984).
[8] J. Deenen and C. Quesne, in *Group Theoretical Methods in Physics, Proceedings, Istanbul, Turkey, 1982, Lecture Notes in Physics*, No. 180, edited by M. Serdaroğlu and E. Inönü (Springer, Berlin, 1983), p. 444.
[9] J. Deenen and C. Quesne, J. Math. Phys. **25**, 2354 (1984).
[10] O. Castaños, A. Frank, E. Chacón, P. Hess, and M. Moshinsky, J. Math. Phys. **23**, 2537 (1982).
[11] M. Moshinsky, Physica A **114**, 322 (1982).
[12] O. Castaños and A. Frank, J. Math. Phys. **25**, 388 (1984).
[13] M. Moshinsky, Nucl. Phys. A **421**, 81c (1984).
[14] G. Rosensteel and D. J. Rowe, Phys. Rev. Lett. **47**, 223 (1981); D. J. Rowe and G. Rosensteel, Phys. Rev. C **25**, 3236 (1982).
[15] D. J. Rowe, J. Math. Phys. **25**, 2662 (1984).
[16] P. Kramer, Ann. Phys. (NY) **141**, 254, 269 (1982).
[17] P. Kramer, Z. Papadopolos, and W. Schweizer, Ann. Phys. (NY) **149**, 44 (1983).
[18] L. D. Mlodinow and N. Papanicolaou, Ann. Phys. (NY) **128**, 314 (1980); **131**, 1 (1981).
[19] N. Papanicolaou, Ann. Phys. (NY) **136**, 210 (1981).
[20] L. R. Mead and N. Papanicolaou, J. Phys. A: Math. Gen. **16**, 4135 (1983).
[21] F. J. Dyson, Phys. Rev. **102**, 1217 (1956).
[22] T. Holstein and H. Primakoff, Phys. Rev. **58**, 1098 (1940).
[23] V. Bargmann, Commun. Pure Appl. Math. **14**, 187 (1961).

[24]J. Dobaczewski, Nucl. Phys. A **369**, 213, 237 (1981); **380**, 1 (1982).
[25]C. L. Siegel, Am. J. Math. **65**, 1 (1943).
[26]The Siegel half-plane can be mapped onto the origin-centered unit ball through a homographic transformation.
[27]A. M. Perelomov, Commun. Math. Phys. **26**, 222 (1972); Usp. Fiz. Nauk **123**, 23 (1977) [Sov. Phys. Usp. **20**, 703 (1977)].
[28]Reference 6 actually deals with the case where $\lambda_1 = \cdots = \lambda_d = 0$, but it is straightforward to extend the results to the case where $\lambda_1 = \cdots = \lambda_d = \lambda \neq 0$ by replacing n by $n + 2\lambda$ in all the formulas.
[29]A. O. Barut and L. Girardello, Commun. Math. Phys. **21**, 41 (1971).
[30]A. O. Barut, Phys. Rev. **139**, B1433 (1965).
[31]P. A. Mello and M. Moshinsky, J. Math. Phys. **16**, 2017 (1975).
[32]M. Moshinsky and T. H. Seligman, J. Math. Phys. **22**, 1526 (1981).
[33]R. Godement, "Séminaire Cartan," Ecole Normale Supérieure, Paris, 1958.
[34]T. Marumori, M. Yamamura, and A. Tokunaga, Prog. Theor. Phys. **31**, 1009 (1964).
[35]D. Jansen, F. Dönau, S. Frauendorf, and R. V. Jolos, Nucl. Phys. A **172**, 145 (1971); F. Dönau and D. Jansen, Nucl. Phys. A **209**, 109 (1973).
[36]T. Tamura, Phys. Rev. C **28**, 2480 (1983).
[37]T. Usui, Prog. Theor. Phys. **23**, 787 (1960).
[38]I. M. Gel'fand and M. L. Tseitlin, Dokl. Akad. Nauk SSSR **71**, 825 (1950).
[39]Note the following change of notations with respect to Ref. 9: $\chi_{(\lambda)}$ is replaced by $|(\lambda)]$, while $|(\lambda)]$ is now denoted as $|0;(\lambda)]$.
[40]R. J. Glauber, Phys. Rev. **130**, 2529 (1963); **131**, 2766 (1963).
[41]J. M. Jauch and E. L. Hill, Phys. Rev. **57**, 641 (1940).
[42]S. Goshen and H. J. Lipkin, Ann. Phys. (NY) **6**, 301 (1959).
[43]O. L. Weaver, R. Y. Cusson, and L. C. Biedenharn, Ann. Phys. (NY) **102**, 493 (1976).
[44]G. Rosensteel and D. J. Rowe, Ann. Phys. (NY) **126**, 343 (1980).
[45]J. P. Elliott, Proc. R. Soc. London Ser. A **245**, 128, 562 (1958).
[46]V. Vanagas, *The Microscopic Nuclear Theory Within the Framework of the Restricted Dynamics*, in *Lecture Notes in Physics* (Univ. Toronto, Toronto, 1977); V. Vanagas, *The Microscopic Theory of the Collective Motion in Nuclei*, in *Group Theory and Its Applications in Physics—1980, AIP Conference Proceedings*, No. 71, edited by T. H. Seligman (AIP, New York, 1981).
[47]L. Sabaliauskas, Liet. Fiz. Rinkinys **19**, 5 (1979).
[48]G. F. Filippov, V. I. Ovcharenko, and Yu. F. Smirnov, *Microscopic Theory of Collective Excitations in Nuclei* (in Russian) (Naukova Dumka, Kiev, 1981).
[49]P. Park, J. Carvalho, M. Vassanji, D. J. Rowe, and G. Rosensteel, Nucl. Phys. A **414**, 93 (1984).
[50]E. Chacón, Ph.D. thesis, UNAM, Mexico, 1969.
[51]M. Moshinsky and C. Quesne, J. Math. Phys. **11**, 1631 (1970).
[52]A. Bohr and B. Mottelson, K. Dan. Vidensk. Selsk. Mat. Fys. Medd. **27**, No. 16 (1953).
[53]A. Arima and F. Iachello, Ann. Phys. (NY) **94**, 253 (1976); **111**, 201 (1978); **123**, 468 (1979).
[54]A. Ya. Dzublik, V. I. Ovcharenko, A. I. Steshenko, and G. F. Filippov, Yad. Fiz. **15**, 869 (1972) [Sov. J. Nucl. Phys. **15**, 487 (1972)].
[55]W. Zickendraht, J. Math. Phys. **10**, 30 (1969); **12**, 1663 (1971).
[56]P. O. Löwdin, Rev. Mod. Phys. **36**, 966 (1964).
[57]J. Shapiro, J. Math. Phys. **6**, 1680 (1965).
[58]A. M. Perelomov and V. S. Popov, Yad. Fiz. **3**, 924, 1127 (1966) [Sov. J. Nucl. Phys. **3**, 676, 819 (1966)]; V. S. Popov and A. M. Perelomov, Yad. Fiz. **5**, 693 (1967) [Sov. J. Nucl. Phys. **5**, 489 (1967)].
[59]G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).
[60]J. G. Nagel and M. Moshinsky, Rev. Mex. Fis. **14**, 29 (1965).

An inverse problem in crystallography

R. Rubinstein
Sverdrup Technology, Inc., Box 30650 Midpark Branch, Middleburg Heights, Ohio 44130
J. E. Fitzgerald
School of Civil Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

(Received 23 January 1985; accepted for publication 12 June 1985)

Coordinate-independent formulations are derived for the conditions satisfied by the elastic modulus tensor of materials with trigonal, hexagonal, and cubic symmetry. Analogous results for two-dimensional modulus tensors are also derived.

I. INTRODUCTION

The conditions satisfied by a material property tensor invariant under a crystallographic symmetry group are usually stated in a coordinate system adapted to the symmetry type in question; the coordinate planes and axes are chosen to coincide with the material planes and axes of symmetry. But it is not unreasonable to ask whether these conditions cannot be restated in a coordinate-independent fashion, in principle as algebraic relations among invariants of the material property tensor. This restatement would solve an inverse problem: given a material property tensor and a symmetry group, determine whether the tensor is invariant under some conjugate of the symmetry group. In this problem, the planes and axes of symmetry, if they exist, are unknown.

Apart from theoretical interest, such coordinate-independent symmetry formulations might have applications to the study of materials having randomly distributed flaws, voids, or filler particles; the mechanical properties of such materials could well be symmetric despite the lack of apparent structural symmetry. With this potential application in mind, this paper will treat the problem posed for the elastic modulus tensor; in any case, this is the simplest tensor for which the problem is not trivial. Three symmetry types will be characterized invariantly: trigonal symmetry; hexagonal symmetry, equivalent for modulus tensors to the existence of a threefold axis and a perpendicular plane of symmetry; and cubic symmetry. Analogous results for the much less difficult two-dimensional problem will also be obtained. Considerable use will be made of results found in the first part of Klein's *Lectures on the Icosahedron*,[1] especially in the characterization of cubic symmetry.

II. REDUCTION TO THE STUDY OF ALGEBRAIC FORMS

The elastic modulus tensor is a fourth-rank tensor having the index symmetries

$$c_{ijkl} = c_{jikl} = c_{ijlk}, \quad c_{ijkl} = c_{klij}. \tag{2.1}$$

The transformation of its components by coordinate rotations defines a 21-dimensional representation of the rotation group, the twice-symmetrized Kronecker product of the ordinary three-dimensional representation. The decomposition of this representation into irreducible subspaces, beginning with the tensors with vanishing pair traces, is routine.[2] The identity representation occurs twice, on

$$A_{ijkl} = (\delta_{ij}\delta_{kl}), \quad B_{ijkl} = (\delta_{ik}\delta_{jl}),$$

where the surrounding parentheses indicate symmetrization by the relations (2.1). The five-dimensional representation occurs twice, on tensors of the form

$$C^{(m)}_{ijkl} = (a^{(m)}_{ij}\delta_{kl}), \quad D^{(m)}_{ijkl} = (a^{(m)}_{ik}\delta_{jl}), \quad 1 \leqslant m \leqslant 5,$$

where the $a^{(m)}$ have trace zero. The quantities $C^{(m)}$ and $D^{(m)}$ all transform like second-order harmonic polynomials $P^{(m)}(x, y, z)$. The nine-dimensional representation occurs on the tensors

$$E^{(n)}_{ijkl} = \sum a^{(o)}_{ij} a^{(p)}_{kl}, \quad \sum a^{(o)}_{ik} a^{(p)}_{kl} = 0, \quad 1 \leqslant n \leqslant 9.$$

The $E^{(n)}$ transform like fourth-order harmonic polynomials $Q^{(n)}(x, y, z)$. Explicit expressions for the tensors $C^{(m)}$, $D^{(m)}$, and $E^{(n)}$, and for the polynomials $P^{(m)}$ and $Q^{(n)}$ appear in the Appendix.

Given a modulus tensor, express it as the sum

$$aA + bB + \sum c^{(m)} C^{(m)} + \sum d^{(m)} D^{(m)} + \sum e^{(n)} E^{(n)}.$$

The polynomials

$$p(x, y, z) = \sum c^{(m)} P^{(m)}(x, y, z),$$

$$q(x, y, z) = \sum d^{(m)} P^{(m)}(x, y, z),$$

$$r(x, y, z) = \sum e^{(n)} Q^{(n)}(x, y, z)$$

are covariants of the modulus tensor and are therefore invariant under the same symmetry operations.

It is convenient to introduce the two-valued complex representation of the rotation group by setting[3]

$$x = 2uv, \quad y = u^2 - v^2, \quad z = (u^2 + v^2)/i. \tag{2.2}$$

Then p, q, and r are replaced by binary forms $p(u, v)$, $q(u, v)$, and $r(u, v)$ of degrees 4, 4, and 8, respectively.

III. MODULUS TENSORS WITH A THREEFOLD AXIS

In the complex representation, the symmetry group contains the six operations (Ref. 1, p. 40)

$$u' = e^{ik\pi/3}u, \quad v' = e^{-ik\pi/3}v, \quad k = 0,1,\ldots,5. \tag{3.1}$$

The most general form which transforms into a scalar multiple of itself under these operations is (Ref. 1, p. 52)

$$u^m v^n \prod (a^{(i)} u^3 + b^{(i)} v^3).$$

The only invariant quartic is

$$F = u^2 v^2, \tag{3.2}$$

and the general invariant form of degree eight is

$$G = uv\left(8au^6 + \binom{8}{4} bu^3 v^3 + 8cv^6\right). \tag{3.3}$$

The general modulus tensor with trigonal symmetry contains seven arbitrary constants.[4] When such a tensor is written as a sum of irreducible representations, two constants multiply identity representations, two constants multiply a term from the five-dimensional representation, and three constants multiply terms from the nine-dimensional representation. Barring degeneracies, the appearance of these dimensions constitutes a simple necessary condition for trigonal symmetry, although it is not difficult to check that they also appear for tetragonal symmetry.

Compute the covariant of G, the sixth transvectant[5]

$$(G,G)^6 = -8!8!6\cdot 5\cdot 4\cdot 3(2ac + 25b^2)u^2 v^2.$$

The following conditions characterize trigonal symmetry invariantly: (i) the forms p, q, and $(r,r)^6$ are multiples of the same form which can be written as the product of two perfect squares U^2V^2; and (ii) when r is written in terms of U and V, it takes the form (3.3).

Return to the original variables by setting

$$X = 2UV, \quad Y = U^2 - V^2, \quad Z = (U^2 + V^2)/i. \tag{3.4}$$

The axis of symmetry is perpendicular to the plane $X = 0$.

Conditions on the invariants of a quartic equivalent to condition (i) could be stated; however, it is probably simpler to check the condition algebraically in each specific case.

IV. MODULUS TENSORS WITH A THREEFOLD AXIS AND A PERPENDICULAR PLANE OF SYMMETRY

The symmetry group contains 12 operations: the six of (3.1) and

$$u' = ie^{-ik\pi/3} v, \quad v' = ie^{ik\pi/3} u, \quad k = 0,\ldots,5$$

(Ref. 1, p. 40). The form (3.2) remains invariant, but the only invariant eighth-order form is u^4v^4. The general invariant modulus tensor contains five arbitrary constants, two in identity representations, two in five-dimensional representations, and one in the nine-dimensional representation. The invariant description of this symmetry type is (i) p and q are multiples of the same form U^2V^2; and (ii) r is a multiple of U^4V^4.

Returning to physical variables as in Sec. III, the symmetry plane is $X = 0$.

V. MODULUS TENSORS WITH CUBIC SYMMETRY

From Ref. 1, one finds that there are no invariant quartics, and only one invariant form of degree eight,

$$W = u^8 + 14u^4v^4 + v^8.$$

These results are also easily derived using physical variables. The general modulus tensor with cubic symmetry therefore contains three arbitrary constants: two in identity representations, and one in the nine-dimensional representation. The forms p and q vanish, and r must be transformable to a multiple of W. The possibility of this transformation can be stated invariantly using a sequence of root extractions. As in the Galois theory of algebraic equations, this sequence depends on the composition series of the cubic symmetry group. In the complex representation, this group has order 48 and the composition numbers, 2,3,2,2 (Ref. 1, p. 41). The Hessian of W,

$$H = (W,W)^2 = 8^2\cdot 7^2\cdot 3u^2v^2(u^4 - v^4)^2,$$

is a perfect square. Define t by

$$t = H^{1/2}/56\sqrt{3} = uv(u^4 - v^4).$$

This is Klein's octahedral form. From it and W, two perfect cubes can be derived, $(W,t)^1 \pm 6\sqrt{3}\, it^2$. Their cube roots are Klein's tetrahedral forms

$$\Phi = u^4 + 2i\sqrt{3}u^2v^2 + v^4,$$

$$\Psi = u^4 - 2i\sqrt{3}u^2v^2 + v^4.$$

Since $u^4 + v^4$ and u^2v^2 can be expressed in terms of these, $u^4 - v^4$ and uv can be found by extracting square roots. Once $u^4 \pm v^4$ are known, one more square-root extraction leads to u^2, v^2, and uv. Cubic symmetry can be characterized invariantly as follows: (i) the forms p and q vanish; (ii) for the form r, the sequence of root extractions described above is possible, and terminates with the forms U^2, V^2, and UV; and (iii) the variables X, Y, Z of (3.4) are related to x, y, z by a rotation.

VI. REFORMULATION OF THE CONDITIONS FOR TRIGONAL SYMMETRY

The invariant characterization of cubic symmetry just given emphasizes the algebraic structure of the symmetry group and is therefore more appealing than the somewhat *ad hoc* conditions given for trigonal symmetry. However, these can also be stated in terms of root extractions as follows. From G of (3.3), compute the transvectants

$$(G,G)^8 = 8!(-16ac + 70b^2),$$

$$((G,G)^6, (G,G)^6)^4 = (8!8!6!)^2(2ac + 25b^2)^2.$$

The combination ac and b can be obtained algebraically from these invariants; knowing them and uv, one can compute the perfect squares

$$au^6 \pm 2\sqrt{ac}\,u^3v^3 + cv^6.$$

Square-root extractions lead to $\sqrt{a}u^3 + \sqrt{c}v^3$, and cube roots to u and v. It is now a simple matter to characterize trigonal symmetry by the possibility of a sequence of root extractions.

The addition of a plane of symmetry to trigonal symmetry causes the modulus tensor to become transversely isotropic, that is, invariant under all rotations about the axis of symmetry. The symmetry group is continuous instead of discrete; therefore, a characterization based on Galois theory cannot be expected.

VII. EXAMPLE

The nonzero elastic constants of a material with cubic symmetry satisfy

$$c_{1111}=c_{2222}=c_{3333},$$
$$c_{1212}=c_{2323}=c_{3131}=\cdots, \quad (7.1)$$
$$c_{1122}=c_{2233}=c_{3311}=\cdots,$$

where additional terms found by applying the index symmetries (2.1) are indicated by dots (Ref. 4, p. 160). Using the notation of the Appendix, this tensor has the form

$$C_{ijkl}=C_1\delta_{ij}\delta_{kl}+C_2\delta_{ik}\delta_{jl}+C_3(E^{(4)}_{ijkl}+E^{(5)}_{ijkl}+E^{(6)}_{ijkl}), \quad (7.2)$$

where symmetrization by (2.1) is understood. These simple expressions arise because the coordinate planes are material symmetry planes. Suppose instead that the x-y plane is rotated by 45°. Then C takes the form

$$C'_{ijkl}=C_1\delta_{ij}\delta_{kl}+C_2\delta_{ik}\delta_{jl} + C'_3(3E^{(4)}_{ijkl}-2E^{(5)}_{ijkl}-2E^{(6)}_{ijkl}), \quad (7.3)$$

and the relations (7.1) are not valid. In the inverse problem, the form (7.3) is given; it is required to determine the symmetry type of C' and to find the rotation which reduces it to the form (7.2).

The first step in solving the inverse problem has already been completed: the tensor C' has been decomposed into terms belonging to irreducible representations. Because two isotropic terms, no terms from the five-dimensional representation, and one term from the nine-dimensional representation occur in the decomposition, C' may have cubic symmetry. The quartic covariant polynomial of C' is proportional to

$$r'(x,y,z)=x^4+y^4-4z^4-18x^2y^2+12z^2(x^2+y^2). \quad (7.4)$$

If a rotation can be found which transforms r' into a multiple of the corresponding polynomial derived from (7.2),

$$r(x,y,z)=x^4+y^4+z^4-3(x^2y^2+y^2z^2+z^2x^2), \quad (7.5)$$

then the cubic symmetry of C' will have been demonstrated.

In the complex representation, r' becomes the form

$$W'=3u^8+28u^6v^2-14u^4v^4+28u^2v^6+3v^8.$$

Performing the sequence of root extractions of Sec. V leads to the results

$$(W',W')^2=56\cdot56\cdot3(u^6-5u^4v^2-5u^2v^4+v^6)^2 = 56\cdot56\cdot3t'^2,$$

$$(W',t')^1/8+6i\sqrt{3}t'^2 = 6i\sqrt{3}(u^4-4iu^3v/\sqrt{3}+2u^2v^2+4iuv^3/\sqrt{3}+v^4)^3 = 6i\sqrt{3}\Phi'^3,$$

$$(W',t')^{1/8}-6i\sqrt{3}t'^2 = 6i\sqrt{3}(-u^4-4iu^3v/\sqrt{3}-2u^2v^2+4iuv^3/\sqrt{3}-v^4)^3 = 6i\sqrt{3}\Psi'^3.$$

Assume that there are variables U,V linearly related to u,v for which, in the notation of Sec. V,

$$-2^{1/3}\sqrt{3}i\Phi'(u,v)=\Phi(U,V), \quad (7.6)$$
$$-2^{1/3}\sqrt{3}i\Psi'(u,v)=\Psi(U,V).$$

According to a result of Klein (Ref. 1, p. 62), the existence of these variables is assured by easily verified equation $(t',t')^4=0$. The only remaining question is whether the physical variables X, Y, Z corresponding to U, V are related to x, y, z by a rotation. Equations (7.6) imply

$$-2^{1/3}(u^4+2u^2v^2+v^4)=2U^2V^2,$$
$$-2^{1/3}4uv(u^2-v^2)=U^4+V^4.$$

Forming linear combinations and taking square roots,

$$U^2+V^2=i2^{1/6}(u^2+2uv-v^2),$$
$$U^2-V^2=2^{1/6}(u^2-2uv-v^2).$$

Ignoring inessential multiplicative constants, the relations

$$x=(Y+Z)/\sqrt{2},\quad y=(-Y+Z)/\sqrt{2},\quad z=X$$

follow. One easily verifies that these equations define a rotation which reduces r' in Eq. (7.4) to a multiple of r in Eq. (7.5). This rotation therefore reduces C' in (7.3) to the canonical form (7.2). The solution is now complete.

VIII. THE TWO-DIMENSIONAL PROBLEM

The transformation by rotation of the components of a two-dimensional modulus tensor defines a six-dimensional representation of the circle group. As in the three-dimensional problem, the identity representation occurs twice. The representations on $e^{\pm 2i\theta}$ and $e^{\pm 4i\theta}$ occur once each. Using real irreducible representations instead, the covariant polynomials are

$$p=a(x^2-y^2)+2bxy, \quad (8.1)$$
$$q=c(x^4-6x^2y^2+y^4)+4d(x^3y-xy^3).$$

Only two types of symmetry are possible.

Orthotropic symmetry: There are two perpendicular lines of symmetry. Some rotation reduces both b and d of (8.1) to zero. The invariant condition is

$$2\arctan(b/a)=\arctan(d/c),$$

or equivalently

$$d(a^2-b^2)=2abc.$$

Tetragonal symmetry: This is the symmetry type of a square. It is invariantly characterized by the condition $a=b=0$.

APPENDIX: IRREDUCIBLE SUBSPACES FOR THE REPRESENTATION OF THE ROTATION GROUP ON ELASTIC MODULUS TENSORS

Let $M(Q)$ denote the matrix of the quadratic form Q. Symmetrization by (2.1) is always understood.

Five-dimensional spaces:

$$C^{(1)}_{ijkl}=M_{ij}(x^2-y^2)\delta_{kl},\quad C^{(2)}_{ijkl}=M_{ij}(y^2-z^2)\delta_{kl},$$
$$C^{(3)}_{ijkl}=M_{ij}(z^2-x^2)\delta_{kl},\quad C^{(4)}_{ijkl}=M_{ij}(2xy)\delta_{kl},$$
$$C^{(5)}_{ijkl}=M_{ij}(2yz)\delta_{kl},\quad C^{(6)}_{ijkl}=M_{ij}(2zx)\delta_{kl},$$
$$D^{(m)}_{ijkl}=C^{(m)}_{ikjl},$$
$$P^{(1)}=x^2-y^2,\quad P^{(2)}=y^2-z^2,\quad P^{(3)}=z^2-x^2,$$
$$P^{(4)}=2xy,\quad P^{(5)}=2yz,\quad P^{(6)}=2zx.$$

Nine-dimensional spaces:

$$E^{(1)}_{ijkl} = M_{ij}(x^2 - y^2)M_{kl}(2xy),$$
$$E^{(2)}_{ijkl} = M_{ij}(y^2 - z^2)M_{kl}(2yz),$$
$$E^{(3)}_{ijkl} = M_{ij}(z^2 - x^2)M_{kl}(2zx),$$
$$E^{(4)}_{ijkl} = M_{ij}(x^2 - y^2)M_{kl}(x^2 - y^2) - M_{ij}(2xy)M_{kl}(2xy),$$
$$E^{(5)}_{ijkl} = M_{ij}(y^2 - z^2)M_{kl}(y^2 - z^2) - M_{ij}(2yz)M_{kl}(2yz),$$
$$E^{(6)}_{ijkl} = M_{ij}(z^2 - x^2)M_{kl}(z^2 - x^2) - M_{ij}(2zx)M_{kl}(2zx),$$
$$E^{(7)}_{ijkl} = M_{ij}(2xy)M_{kl}(2yz) + M_{ij}(y^2 - z^2)M_{kl}(2zx),$$
$$E^{(8)}_{ijkl} = M_{ij}(2yz)M_{kl}(2zx) + M_{ij}(z^2 - x^2)M_{kl}(2xy),$$
$$E^{(9)}_{ijkl} = M_{ij}(2zx)M_{kl}(2xy) + M_{ij}(2yz)M_{kl}(x^2 - y^2),$$
$$Q^{(1)} = 2(x^3y - xy^3),\ \ Q^{(2)} = 2(y^3z - yz^3),$$
$$Q^{(3)} = 2(z^3x - zx^3),$$
$$Q^{(4)} = x^4 - 6x^2y^2 + y^4,\ \ Q^{(5)} = y^4 - 6y^2z^2 + z^4,$$
$$Q^{(6)} = z^4 - 6z^2x^2 + x^4,$$
$$Q^{(7)} = 6xy^2z - 2z^3x,\ \ Q^{(8)} = 6xyz^2 - 2x^3y,$$
$$Q^{(9)} = 6x^2yz - 2y^3z.$$

[1]F. Klein, *Lectures on the Icosahedron and the Solution of Equations of the Fifth Degree* (Dover, New York, 1956).

[2]H. Weyl, *The Classical Groups* (Princeton U. P., Princeton, NJ, 1946), p. 150.

[3]E. Cartan, *The Theory of Spinors* (Dover, New York, 1981), p. 41.

[4]A. E. H. Love, *A Treatise on the Mathematical Theory of Elasticity* (Dover, New York, 1944), p. 159.

[5]J. H. Grace and A. Young, *The Algebra of Invariants* (Chelsea, New York, 1952), p. 46.

On combinatoric determination of SU(n) weight multiplicities

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

(Received 24 October 1984; accepted for publication 17 May 1985)

A combinatoric method for evaluating the inner multiplicity of a central weight is given using results from the theory of distributions. A formula for special cases is derived, and a recursive algorithm is presented for the general problem.

I. INTRODUCTION

The determination of the multiplicity of a weight in a representation of a Lie algebra can be accomplished by several methods. These methods are either direct or recursive. For example, Kostant's formula[1]

$$m_\lambda(\mu) = \sum_{\sigma\in w} (-1)^{sn\sigma} P(\sigma(\lambda+\delta) - (\mu+\delta)) \tag{1.1}$$

is direct and involves a summation over the Weyl group and evaluation of Kostant's partition function. This is fine for algebras of rank 4 or smaller, but since $|W| \sim n!$ for rank n and no evaluation of $P(\cdots)$ is available, the formula is to be avoided for large rank. A useful recursive method is Freudenthal's formula[1]

$$\{(\lambda+\delta,\lambda+\delta) - (\mu+\delta,\mu+\delta)\} m_\lambda(\mu) = 2\sum_{\alpha>0}\sum_{i=1}^{\infty} m(\mu+i\alpha)(\mu+i\alpha,\alpha). \tag{1.2}$$

However, one must know all of the weights of a representation in order to use this method. In addition, the number of positive roots $\alpha > 0$ is on the order of n^2 for rank n so application is tedious. One also notices that this formula works from the periphery of the weight diagram inward, so the weights of highest multiplicity are the last to be be evaluated.

In this paper we shall explore the connections between the multiplicity problem and certain results from the theory of distributions.

Sections II and III outline the properties of the character function. In Sec. IV we give an integral formula and selection rules for multiplicities, and in Sec. V we evaluate the integral formula using combinatoric methods. We construct generating functions and recursion formulas in Sec. VI A and in VI B we develop a formula for the multiplicity of the zero weight in a tensor representation of rank equal to the rank of the underlying algebra plus one. Also in VI B we give generating functions for a distribution problem related to finding the multiplicity of a fundamental weight in a tensor representation. We restrict our attention to the groups SU(n).

II. CHARACTERS

Explicit expressions for the group character of a representation will be the basis of our development. Let g be a Lie group and $\lambda(g)$ a representation. Then the character is defined in the following way:

$$\chi_\lambda(g) = \mathrm{Tr}\,\lambda(g), \tag{2.1}$$

with the following properties:

$$\chi_{(\lambda\otimes\lambda')}(g) = \chi_\lambda(g)\cdot\chi_{\lambda'}(g),$$
$$\chi_{(\lambda\oplus\lambda')}(g) = \chi_\lambda(g) + \chi_{\lambda'}(g). \tag{2.2}$$

As an example, consider

$$\chi'_i = \sum_j u_{ij} x_j, \quad u \text{ a unitary map.} \tag{2.3}$$

Then $\exists V$, V unitary such that

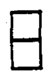
$$V^{-1}UV = \mathrm{diag}(\epsilon_1,\ldots,\epsilon_n), \tag{2.4}$$

so that there is a basis[2] x_i such that

$$u{:}x_i \to \epsilon_i x_i, \tag{2.5}$$

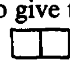
with no sum on i. The ϵ_i are clearly the eigenvalues of $u\in\mathrm{SU}(n)$. Now consider a tensor representation spanned by products of the form

$$x_i x_j, \quad i,j = 1,\ldots,n.$$

Let us impose the restriction $i<j$ which gives the irreducible representation of dimension $n(n-1)/2$ of SU(n), labeled in the Young scheme as .

Then $u{:}\ x_i x_j \to \epsilon_i\epsilon_j x_i x_j$

$$\mathrm{Tr}\,u = \chi_{\boxminus} = \sum_{i<j}^{n} \epsilon_i\epsilon_j. \tag{2.6}$$

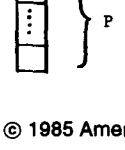
We recognize χ to be the second elementary symmetric function in n variables ϵ_i. Now let $i \leqslant j$ to give the $[n(n+1)/2]$-dimensional representation of SU(n), .

$$\mathrm{Tr}\,u = \chi_{\square\square} = \sum_{i\leqslant j}^{n} \epsilon_i\epsilon_j, \tag{2.7}$$

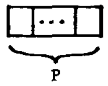
which we recognize as the second symmetric function on n variables ϵ_i. We shall adopt the notation $a_p(n)$ for elementary symmetric functions and $h_p(n)$ for totally symmetric functions. For example,

$$a_2(3) = \epsilon_1\epsilon_2 + \epsilon_1\epsilon_3 + \epsilon_2\epsilon_3,$$
$$h_2(3) = \epsilon_1^2 + \epsilon_2^2 + \epsilon_3^2 + \epsilon_1\epsilon_2 + \epsilon_1\epsilon_3 + \epsilon_2\epsilon_3.$$

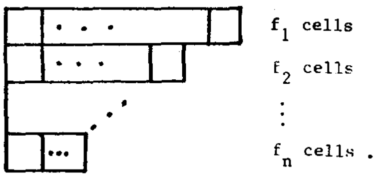
Let d_p be the pth fundamental representation labeled by the Young diagram



and d^p be the pth symmetric representation



Let $(f_1,f_2,...,f_n)$ designate the representation with diagram



Weyl's second formula[2] then reads

$$\chi_{(f_1,...,f_n)} = \det \Sigma, \ \Sigma_{ij} = \chi_{d^{i-j+f_i}}. \tag{2.8}$$

An example is, for SU(3),

$$\chi_{(21)} = \chi_{\begin{smallmatrix}\square\square\\ \square\end{smallmatrix}} = \begin{vmatrix} \chi_{\square\square} & 1 \\ \chi_{\square\square\square} & \chi_{\square} \end{vmatrix} = \begin{vmatrix} h_2(3) & h_0(3) \\ h_3(3) & h_1(3) \end{vmatrix}.$$

III. THE PARAMETERS ϵ_i

Recall $V^{-1}UV \to \text{diag}(\epsilon_1 \cdots \epsilon_n)$. Unitarity requires that

$$|\epsilon_i| = 1, \quad \epsilon_i = e^{2\pi i \phi_i}. \tag{3.1}$$

We also note that since SU(n) is composed of matrices of determinant 1,

$$\phi_1 + \phi_2 + \cdots + \phi_n = 0. \tag{3.2}$$

To discuss arbitrary weights in the future it would be convenient to relate fundamental weights λ_i to diagrams in the Young scheme.

Let us choose a Cartesian basis for the root system of the Lie algebra A_n in an $(n+1)$-dimensional space:

$$\text{for } A_n \to \text{SU}(n+1) \begin{cases} \alpha_1 = (1,-1,0,...,0), \\ \alpha_2 = (0,1,-1,0,...,0), \\ \vdots \\ \alpha_i = (0,..0,\underset{i}{1},\underset{i+1}{-1},0...0), \\ \alpha_n = (0,......0,\underset{n}{1},\underset{n+1}{-1}). \end{cases}$$

The fundamental weights λ_i are given in terms of the roots α_i by the inverse Cartan matrix[1]

$$\begin{aligned} \lambda_i = [1/(n+1)]\{&(n+1-i)\alpha_1 + 2(n+1-i)\alpha_2 \\ &+ \cdots + (i-1)(n+1-i)\alpha_{i-1} + i(n+1-i)\alpha_i \\ &+ i(n+1-(i+1))\alpha_{i+1} + \cdots + i\alpha_n\} \end{aligned} \tag{3.3}$$

$$= \frac{1}{n+1}\Big\{\underset{1}{n+1-i}, \underset{2}{n+1-i}, \underset{\cdots}{\ldots}, \underset{i}{n+1-i}, \underset{\cdots}{-i,-i,...,} \underset{n+1}{-i}\Big\}. \tag{3.4}$$

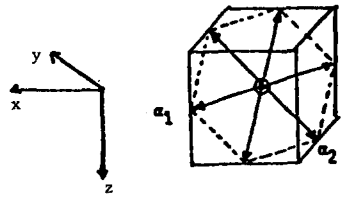
Note

$$\lambda_i + [i/(n+1)](1,1,...,1) = (1,1,...,1,0,...,0).$$

In any character formula one encounters such things as

$$\textstyle\sum_i m_i\phi_i = m\cdot\phi, \tag{3.5}$$

so one may always add any multiple of $\phi_1 + \cdots + \phi_{n+1} = 0$ to such an expression without changing its value. In other words, one may add any multiple of $(1,1,...,1)$ to a weight μ without changing $m\cdot\phi$. For example, $A_2 \to \text{SU}(3)$



,

$$\alpha_1 = (1,-1,0), \quad \alpha_2 = (0,1,-1),$$

$$C = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}, \quad C^{-1} = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix},$$

$$\lambda_1 = (\tfrac{2}{3}, -\tfrac{1}{3}, -\tfrac{1}{3}) \xrightarrow{\text{add } 1/3\,(111)} (1,0,0) \to \square\,,$$

$$\lambda_2 = (\tfrac{1}{3},\tfrac{1}{3},-\tfrac{2}{3}) \xrightarrow{\text{add } 2/3\,(111)} (1,1,0) \to \begin{smallmatrix}\square\\ \square\end{smallmatrix}\,.$$

IV. MULTIPLICITIES

In addition to our basic definition

$$\chi_\lambda = \text{Tr}\,\lambda(g), \tag{4.1}$$

we have the following relation:

$$\chi(\lambda,\phi) = \sum_{m\in V_\gamma} \gamma_\lambda(m) e^{im\cdot\phi}, \tag{4.2}$$

where $\gamma_\lambda(m)$ is the multiplicity of weight m in representation λ. Let m' be a weight of V_λ

$$\chi(\lambda,\phi)e^{-im'\cdot\phi} = \sum_{m\in V_\lambda} \gamma_\lambda(m) e^{i(m-m')\cdot\phi}, \tag{4.3}$$

$$\begin{aligned} \left(\frac{1}{2\pi}\right)^{n-1} \int_0^{2\pi} d\,\Phi\, \chi(\lambda,\phi) e^{-im'\cdot\phi} &= \sum_{m\in v_\lambda} \gamma_\lambda(m) \left(\frac{1}{2\pi}\right)^{n-1} \int_0^{2\pi} e^{i(m-m')\cdot\phi}\, d\,\Phi \\ &= \sum_{m\in V_\lambda} \gamma_\lambda(m)\delta(m-m') = \gamma_\lambda(m'), \end{aligned} \tag{4.4}$$

or

$$\gamma_\lambda(m) = \left(\frac{1}{2\pi}\right)^{n-1} \int_0^{2\pi} d\,\Phi\, \chi(\lambda,\phi) e^{-im\cdot\phi}. \tag{4.5}$$

This result is of little practical value since in general the character function $\chi(\lambda,\phi)$ is a complicated function of ϕ. However, one can use Weyl's second formula to express $\chi(\lambda,\phi)$ for $\lambda = (f_1,...,f_n)$ in terms of symmetric functions.

Example: For SU(3), $\lambda = (2,1)$ with diagram $\begin{smallmatrix}\square\square\\ \square\end{smallmatrix}$

$$\phi_1 + \phi_2 + \phi_3 = 0 \to \epsilon_3 = \epsilon_2^{-1}\epsilon_1^{-1},$$

$$\chi = \begin{vmatrix} h_2(3) & 1 \\ h_3(3) & h_1(3) \end{vmatrix}$$

$$= 2 + \epsilon_1^2\epsilon_2 + \epsilon_1\epsilon_2^{-1} + \epsilon_1\epsilon_2^2 + \epsilon_2\epsilon_1^{-1} + \epsilon_1^{-2}\epsilon_2^{-1} + \epsilon_1^{-1}\epsilon_2^{-2},$$

$$\gamma(0,0) = \left(\frac{1}{2\pi}\right)^2 \int_0^{2\pi} d\phi_1\, d\phi_2\, \chi = 2$$

$$= \text{multiplicity of center.}$$

An immediate consequence of Eq. (4.5) is a set of selection rules. We shall restrict ourselves to the discussion of positive weights in order to avoid anomalous Young diagrams. For n variables $\epsilon_1,...,\epsilon_n$ with $\epsilon_n = \epsilon_{n-1}^{-1} \cdots \epsilon_2^{-1}\epsilon_1^{-1}$ we observe the correspondence between positive weights and monomials:

weight	*monomial*
$(0,...,0)$	$(\epsilon_1,\epsilon_2 \cdots \epsilon_n)^m$
$(1,0,...,0)$	$\epsilon_1^{m+1}(\epsilon_2 \cdots \epsilon_n)^m$
$(1,1,0,...,0)$	$\epsilon_1^{m+1}\epsilon_2^{m+1}(\epsilon_3 \cdots \epsilon_n)^m.$

Consider a product of symmetric functions

$$h_x h_y \cdots h_z$$

in the character of a representation of SU(n). If

$$x + y + \cdots + z \neq mn, \quad m \in Z^+ U\{0\}, \tag{4.6}$$

there will be no monomials $(\epsilon_1\epsilon_2 \cdots \epsilon_n)^m$ in the product so the multiplicity of the center of the weight diagram of the representation will be zero. Continuing with this reasoning we obtain the following selection rules. For SU(n)

$$\begin{aligned}
&\gamma_{(f_1,...,f_n)}(0,...,0) = 0,\\
&\qquad \text{unless } f_1 + \cdots + f_n = mn,\\
&\gamma_{(f_1,...,f_n)}(1,0,...,0) = 0,\\
&\qquad \text{unless } f_1 + \cdots + f_n = mn + 1,\\
&\gamma_{(f_1,...,f_n)}(1,1,0,...,0) = 0,\\
&\qquad \text{unless } f_1 + \cdots + f_n = mn + 2,\\
&\gamma_{(f_1,...,f_n)}(p_1 \cdots p_n) = 0,\\
&\qquad \text{unless } f_1 + \cdots + f_n = mn + p_1 + \cdots + p_n.
\end{aligned} \tag{4.7}$$

V. CONNECTION WITH DISTRIBUTION THEORY: HAMMOND OPERATORS

In order to obtain a useful multiplicity formula for SU(n) we must circumvent the integration in Eq. (4.5). This may be done by introducing differential operators with a specified action on a product of symmetric functions.

The number of times that a given monomial appears in a product $h_{p_1}(n)h_{p_2}(n) \cdots$ is a classic problem in the theory of distributions. Consider a specific example. Let $n = 4$ and examine the product

$$h_4(\epsilon_1,\epsilon_2,\epsilon_3,\epsilon_4)h_3(\epsilon_1,\epsilon_2,\epsilon_3,\epsilon_4),$$

where h_4 has $\binom{4+4-1}{4} = 35$ terms, h_3 has $\binom{3+4-1}{3} = 20$ terms. What is the number of occurrences of

$$\epsilon_1^3\epsilon_2^2\epsilon_3\epsilon_4?$$

This monomial occurs eleven times

$$\epsilon_1^3\epsilon_2 \cdot \epsilon_2\epsilon_3\epsilon_4,\quad \epsilon_1^3\epsilon_3 \cdot \epsilon_2^2\epsilon_4,\quad \epsilon_1^3\epsilon_4 \cdot \epsilon_2^2\epsilon_3,$$
$$\epsilon_1^2\epsilon_2^2 \cdot \epsilon_1\epsilon_3\epsilon_4,\quad \epsilon_1^2\epsilon_2\epsilon_3 \cdot \epsilon_1\epsilon_2\epsilon_4,\quad \epsilon_1^2\epsilon_2\epsilon_4 \cdot \epsilon_1\epsilon_2\epsilon_3,$$
$$\epsilon_1^2\epsilon_3\epsilon_4 \cdot \epsilon_1\epsilon_2^2,\quad \epsilon_2^2\epsilon_1\epsilon_3 \cdot \epsilon_1^2\epsilon_4,\quad \epsilon_2^2\epsilon_1\epsilon_4 \cdot \epsilon_1^2\epsilon_3,$$
$$\epsilon_2^2\epsilon_3\epsilon_4 \cdot \epsilon_1^3,\quad \epsilon_1\epsilon_2\epsilon_3\epsilon_4 \cdot \epsilon_1^2\epsilon_2,$$

which corresponds to a distribution of seven objects

$$\epsilon_1,\epsilon_1,\epsilon_1,\epsilon_2,\epsilon_2,\epsilon_3,\epsilon_4$$

into seven parcels, four of one type and three of another.[3] One does not need to multiply the functions out and collect terms. Introduce instead the Hammond[3] operators D_i

$$D_i h_m = h_{m-i}, \quad D_i h_m h_n = \sum_{j=0}^{i} h_{m-j}h_{n-i+j}, \tag{5.1}$$

which acts as convolutions on products of the symmetric functions

$$\begin{aligned}
&D_3D_2D_1^2h_3h_4\\
&= D_2D_1^2(h_0h_4 + h_1h_3 + h_2h_2 + h_3h_1)\\
&= D_1^2(h_0h_2 + h_1h_1 + h_0h_2 + h_0h_2 + h_1h_1 + h_2h_0\\
&\quad + h_1h_1 + h_2h_0)\\
&= D_1(h_0h_1 + h_0h_1 + h_0h_1 + h_1h_0 + h_0h_1 + h_0h_1\\
&\quad + h_1h_0 + h_1h_0 + h_0h_1 + h_1h_0 + h_1h_0) = 11.
\end{aligned} \tag{5.2}$$

Using the Hammond operators we can write a very general multiplicity formula. Let $(f_1,...,f_n)$ denote a representation of SU(n). The monomial

$$\epsilon_1^{m_1}\epsilon_2^{m_2} \cdots \epsilon_n^{m_n}, \quad m_1 \geqslant m_2 \geqslant \cdots \geqslant m_n$$

corresponds to the weight

$$(m_1 - m_2)\lambda_1 + (m_2 - m_3)\lambda_2 + \cdots.$$

Then the multiplicity of this weight in $(f_1,...,f_n)$ is

$$D_{m_1}D_{m_2} \cdots D_{m_n} \quad \det\{h_{i-j+f_i}\}. \tag{5.3}$$

For an arbitrary weight this is no easier to evaluate than Kostant's formula. However, evaluation for weights near the center of the weight diagram is very simple.

VI. EVALUATION FOR CENTRAL WEIGHTS

A. A Special case

Consider first the zero weight. Due to the selection rules, we need only consider $f_1 + \cdots + f_n = mn$ and the formula reduces to

$$(D_m)^n \det h_{i-j+f_i}. \tag{6.1}$$

We would like to connect our problem with questions arising in the theory of distributions, so let us first examine

$$D_m^n h_y h_x = \sum_{q=0}^{mn} \Gamma^q(n,m) h_{y-q} h_{x-mn+q}. \tag{6.2}$$

Clearly $\Gamma^q(n,m) =$ the number of compositions of q into n integers, each no greater than m. It is equally clear that $\Gamma^q(n,m) =$ the number of ways of putting q objects in n boxes (distinguishable) with no more than m per box. This is generated by

$$
\begin{aligned}
&(1+t+\cdots+t^m)^n \\
&= \sum_{q=0}^{mn} t^q \Gamma^q(n,m) = \left(\frac{1-t^{m+1}}{1-t}\right)^n \\
&= (1-t^{m+1})^n \sum_q \binom{q+n-1}{q} t^q \\
&= \sum_{k=0}^{n} \binom{n}{k} (-t^{m+1})^k \sum_q \binom{q+n-1}{q} t^q \\
&= \sum_{k=0}^{n} \sum_q (-1)^k \binom{n}{k} \binom{q+n-1}{q} t^{mk+k+q} \\
&= \sum_q \left\{ \sum_{k=0} (-1)^k \binom{n}{k} \binom{q-mk-k+n-1}{q-k-mk} \right\} t^q. \qquad (6.3)
\end{aligned}
$$

These numbers have the following recursion relation readily derived from the generating function:

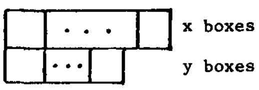
$$
\begin{aligned}
\Gamma^q(n,m) &= \Gamma^q(n-1,m) + \Gamma^{q-1}(n-1,m) \\
&\quad + \cdots + \Gamma^{q-m}(n-1,m). \qquad (6.4)
\end{aligned}
$$

As an example, $\Gamma^q(n,2)$ are

$$
\begin{array}{ccccccc}
 & & & 1 & & & \\
 & & 1 & 1 & 1 & & \\
 & 1 & 2 & 3 & 2 & 1 & . \\
1 & 3 & 6 & 7 & 6 & 3 & 1 \\
 & & \cdot & \cdot & \cdot & &
\end{array}
$$

Also note $\Gamma^q(n,1) = \binom{n}{q}$.

The product of two symmetric functions is the type of thing which arises in the character formula for representations of SU(n) whose Young diagrams are of the form



,

$(f_1, f_2, \cdots f_n) = (x,y,0,\ldots,0)$.

In this case

$$
\chi_{(x,y,0,\ldots,0)} = \begin{vmatrix} h_x(n) & h_{y-1}(n) \\ h_{x+1}(n) & h_y(n) \end{vmatrix}, \qquad (6.5)
$$

so that the multiplicity of the center is given by

$$
\begin{aligned}
&(D_m)^n (h_x(n) h_y(n) - h_{x+1}(n) h_{y-1}(n)) \\
&= \sum_{k=0} (-1)^k \binom{n}{k} \binom{y-mk+n-1-k}{y-k-mk} \\
&\quad - \sum_{k=0} (-1)^k \binom{n}{k} \binom{y-mk+n-2-k}{y-k-mk-1} \qquad (6.6) \\
&= \sum_{k=0} (-1)^k \binom{n}{k} \binom{y-mk-k+n-2}{n-2}, \\
&\qquad \text{for } x+y=mn, \text{ zero otherwise.}
\end{aligned}
$$

The multiplicity of the fundamental weights are also easy to obtain from the correspondences

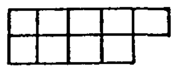
$$
\begin{aligned}
&(f_1 f_2, \ldots, f_n) \\
&= (\underbrace{1,\ldots,1}_{i},0,\ldots,0) \to (\epsilon_1 \cdots \epsilon_i)^{m+1} (\epsilon_{i+1} \cdots \epsilon_n)^m \\
&\qquad \to (D_{m+1})^i (D_m)^{n-i}, \qquad (6.7)
\end{aligned}
$$

$$
\begin{aligned}
&\gamma_{(x,y,0,\ldots,0)}(1,1,\ldots,1,0,\ldots,0) \\
&= (D_{m+1})^i (D_m)^{n-i} (h_x(n) h_y(n) - h_{x+1}(n) h_{y-1}(n)) \\
&= (D_{m+1})^i \left\{ \sum_{q=0}^{mn-mi} \Gamma^q(n-i,m) h_{y-q}(n) h_{x-m(n-i)+q}(n) \right. \\
&\quad \left. - \sum_{q=0}^{mn-mi} \Gamma^q(n-i,m) h_{y-1-q}(n) h_{x+1-m(n-i)+q}(n) \right\} \\
&= \sum_{p=0}^{mi+i} \sum_{q=0}^{mn-mi} \Gamma^q(n-i,m) \Gamma^p(i,m+1) \{ h_{y-q-p} \\
&\quad \times h_{x-(mn+i)+q+p} - h_{y-1-q-p} \\
&\quad \times h_{x+1-(mn+i)+q+p} \}.
\end{aligned}
\qquad (6.8)
$$

Only values of q and p which give terms proportional to $h_0 h_0$ do not vanish, therefore

$$
\begin{aligned}
&\gamma_{(x,y,0,\ldots,0)}(1,1,\ldots,1,0\ldots,0) \\
&= \sum_{p=0}^{mi+i} \{ \Gamma^{y-p}(n-i,m) - \Gamma^{y-1-p}(n-i,m) \} \\
&\quad \times \Gamma^p(i,m+1). \qquad (6.9)
\end{aligned}
$$

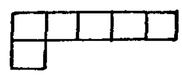
Examples: For SU(4) $n=4$ we have the following.

(i)  $=(5,4,0)$,

$x=5,\ y=4,\ i=1,\ m=2,$

since $5+4=2\cdot 4+1$,

$$
\begin{aligned}
\gamma^{(100)}_{(540)} &= \sum_{p=0}^{3} \{ \Gamma^{4-p}(3,2) - \Gamma^{3-p}(3,2) \} \Gamma^p(1,3) \\
&= \Gamma^4(3,2) - \Gamma^0(3,2) = 6 - 1 = 5.
\end{aligned}
$$

(ii)  $=(5,1,0)$,

$i=2,\ x=5,\ y=1,\ m=1,$

since $5+1=1\cdot 4+2$,

$$
\begin{aligned}
\gamma^{(110)}_{(510)} &= \sum_{p=0}^{4} \{ \Gamma^{1-p}(2,1) - \Gamma^{-p}(2,1) \} \Gamma^p(2,2) \\
&= \Gamma^1(2,1)\Gamma^0(2,2) - \Gamma^0(2,1)\Gamma^0(2,2) + \Gamma^0(2,1) \\
&\quad \times \Gamma^1(2,2) = 2 - 1 + 2 = 3.
\end{aligned}
$$

B. The general problem

It is clear from the preceding section that the solution to the general case requires knowledge of the action of some Hammond operator raised to a power on a product of an arbitrary number of symmetric functions:

$$
\begin{aligned}
&D_p^m h_x h_y \cdots h_z \\
&= D_p^{m-1} \sum_{i_1 \cdots l_1}^{i_1+j_1+\cdots+l_1=p} h_{x-i_1} h_{y-j_1} \cdots h_{z-l_1} \\
&= \sum_{i_1 \cdots l_1}^{i_1+\cdots+l_1=p} \cdots \sum_{i_k \cdots l_k}^{i_k+\cdots+l_k=p} h_{x-i_1\cdots i_k} h_{y-j_1\cdots j_k} \\
&\quad \cdots h_{z-l_1\cdots l_k} \\
&= \sum_{a,b\ldots,c} \Gamma_{b\,c}^{\,a}(p,m) h_{x-a} h_{y-b} \cdots h_{z-c}. \qquad (6.10)
\end{aligned}
$$

The coefficient $\Gamma\overset{a}{\underset{c}{b}}(p,m)$ expresses the number of ways to form $a,b,\ldots,c$ from k numbers, each no larger than p with the constraints

$$\begin{aligned} &i_1 + j_1 + \cdots + l_1 = p,\\ &i_2 + j_2 + \cdots + l_2 = p,\\ &\vdots \qquad\qquad\quad \vdots\\ &i_k + j_k + \cdots + l_k = p. \end{aligned} \tag{6.11}$$

To make the connection with combinatoric theory, one notes that $\Gamma\overset{a}{\underset{c}{\vdots}}(p,m)$ enumerates the solutions to the following problem. Consider the number of ways of putting integers (non-negative) into cells in a rectangular array such that each column totals p and the rows total $a,b,\ldots,c$, respectively.

In the present case

i_1	i_2	$\cdots$	i_k	a
j_1	j_2	$\cdots$	j_k	b
$\vdots$	$\vdots$		$\vdots$	$\vdots$
l_1	l_2	$\cdots$	l_k	c
p	p	$\cdots$	p	

The problem of finding the coefficients Γ is clearly related to the problem of enumerating Latin (or magic) rectangles.[3] The Latin rectangle problem can be solved using the inclusion/exclusion principle of combinatorics,[4] but we shall solve the problem at hand in a different manner by constructing generating functions and recursion relations.

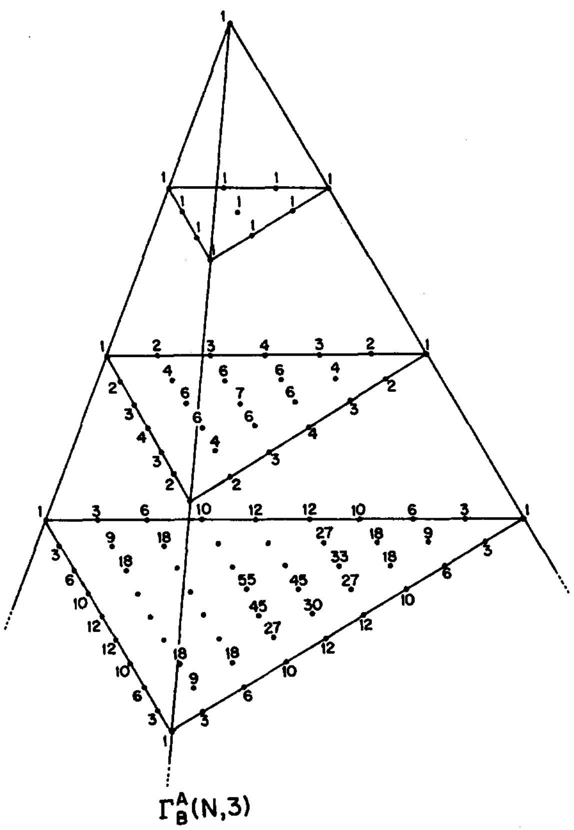


FIG. 1. Coefficients $\Gamma_B^A(N,3)$ in Eq. (6.13).

There is no loss in generality in taking

$$kp = a + b + \cdots + c. \tag{6.12}$$

This allows us to eliminate c and later will make the symmetries of the Γ coefficients obvious.

In order to write the generating function, we now equate the rectangle problem to a distribution problem. Suppose we have a objects of one type, b of another, ..., and c of another type. Suppose in addition that we have k distinguishable boxes. How many ways are there to place objects into boxes such that there will be p objects in each box? Obviously,

$$kp = a + b + \cdots + c.$$

If all of the objects are the same type, we already have the generating function, Eq. (6.5). The solution to the problem at hand is the most obvious generalization of this previous result. For clarity consider $p = 3$, and just three types of objects a,b,c; then $3k = a + b + c$. The generator is

$$\begin{aligned} &\{1 + (S+T) + (S^2 + ST + T^2)\\ &\quad + (S^3 + S^2T + ST^2 + T^3)\}^k\\ &\quad = \sum_{x,y}^{3k} \Gamma_y^x(k,3) S^x T^y. \end{aligned} \tag{6.13}$$

(See Fig. 1.) Note Γ has only two indices instead of three because of the constraint Eq. (6.12). For the number of rectangles

		$\cdots$		a	
		$\cdots$		b	
$\vdots$	$\vdots$		$\vdots$	$\vdots$	m,
		$\cdots$		d	
p_1	p_2	$\cdots$	p_k		

with $\Sigma_{i=1}^k p_i = a + b + \cdots + c + d$, the generator is the following. Let $x_1,\ldots,x_{m-1}$ be $m-1$ indeterminants. Then

$$\prod_{j=1}^{k} \sum_{i=0}^{p_j} h_i(m-1) = \sum_{\alpha_1\cdots\alpha_{m-1}} \Gamma \overset{\alpha_1}{\underset{\alpha_{m-1}}{\vdots}} X_1^{\alpha_1}\cdots X_{m-1}^{\alpha_{m-1}} \tag{6.14}$$

enumerates the possible solutions.

C. Application to central weights

A multiplicity formula in terms of factorials can be derived for the case $p_1 = p_2 = \cdots = p_k = 1$. Then we have

$$\begin{aligned} &\prod_{j=1}^{k} \sum_{i=0}^{1} h_i(m-1)\\ &\quad = (1 + X_1 + X_2 + \cdots + X_{m-1})^k\\ &\quad = \sum_{i_0\cdots i_{m-1}} \frac{k!}{i_0!\cdots i_{m-1}!} (1)^{i_0} X_1^{i_1}\cdots X_{m-1}^{i_{m-1}}, \end{aligned} \tag{6.15}$$

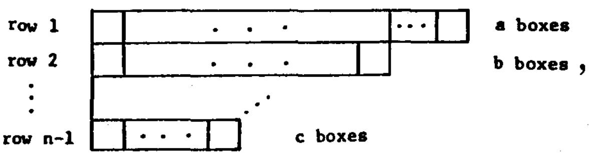
so

$$\Gamma \overset{i_1}{\underset{i_{m-1}}{\vdots}} (k,1) = k!/i_0! i_1! \cdots i_{m-1}!, \tag{6.16}$$

where

$$i_0 = k - i_1 - \cdots - i_{m-1}.$$

This result can be used to write the multiplicity of the zero weight in a respresentation with a boxes in row 1, b boxes in row 2,..., c boxes in row $n-1$;



for SU(n) with $a+b+\cdots+c=n=k$, namely,

$$\gamma^{(0,\ldots,0)}_{(a,b,\ldots,c)} = n!\begin{vmatrix} \frac{1}{a!} & \frac{1}{(b-1)!} & \cdots & \\ \frac{1}{(a+1)!} & \frac{1}{b!} & \cdots & \vdots \\ \vdots & & & \frac{1}{(c-1)!} \\ & & \cdots & \frac{1}{c!} \end{vmatrix}. \qquad (6.17)$$

Example: For SU(4), representation = (3,1,0),

$$\gamma^{(000)}_{(310)} = 4!\begin{vmatrix} 1/3! & 1/0! & 0 \\ 1/4! & 1/1! & 0 \\ 1/5! & 1/2! & 1/0! \end{vmatrix}$$

$$= 4!\left(\frac{1}{3!}\frac{1}{1!} - \frac{1}{4!}\frac{1}{0!}\right) = 3.$$

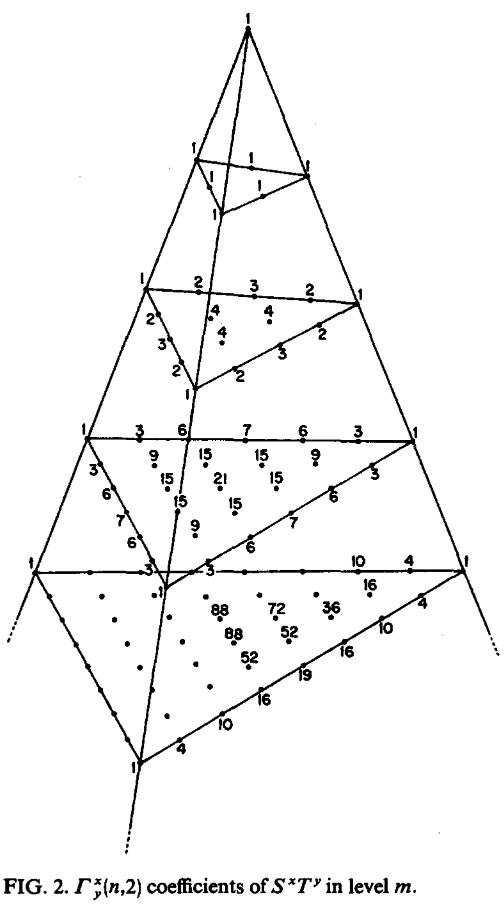


FIG. 2. $\Gamma^x_y(n,2)$ coefficients of S^xT^y in level m.

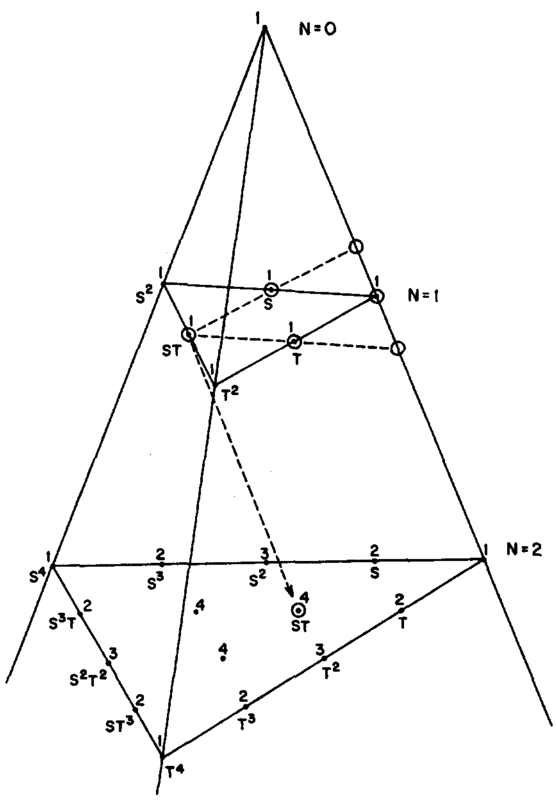


FIG. 3. $\Gamma^x_y(n,2)$ coefficients of S^xT^y in level N.

It is far more difficult to give a useful formula for cases in which $a+b\cdots+c=np, p>1$, but recursive methods for obtaining the Γ coefficients make graphical techniques possible. Consider

$$a+b+c=2n. \qquad (6.18)$$

The generating function is

$$(1+(X_1+X_2)+(X_1^2+X_1X_2+X_2^2))^n = \Sigma\Gamma^a_b(n,2)X_1^aX_2^b \qquad (6.19)$$

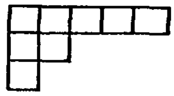
and note that one has

$$\Gamma^a_b(n,2) = \Gamma^a_b(n-1,2) + \Gamma^{a-1}_b(n-1,2) + \Gamma^{a-2}_b(n-1,2) + \Gamma^a_{b-1}(n-1,2) + \Gamma^a_{b-2}(n-1,2) + \Gamma^{a-1}_{b-1}(n-1,2). \qquad (6.20)$$

These coefficients are displayed in Figs. 2 and 3. One merely has to have the Γ's for $n=1$ and apply recursion to develop the series to any n. Note also the high degree of symmetry in the coefficients

$$\Gamma^a_b(n,2) = \Gamma^b_a(n,2) = \Gamma^{2n-a-b}_a(n,2) = \Gamma^b_{2n-a-b}(n,2), \quad \text{etc.} \qquad (6.21)$$

Let this be applied to



$= (5,2,1),$

of SU(4), zero weight.

Weyl's second formula gives

$$\chi(5,2,1) = \begin{vmatrix} h_5(4) & h_1(4) & 0 \\ h_6(4) & h_2(4) & h_0(4) \\ h_7(4) & h_3(4) & h_1(4) \end{vmatrix}$$

$$= [5,2,1] + [1,0,7] - [3,0,5] - [1,1,6].$$

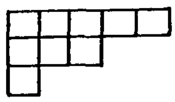
From Fig. 2, we have

$$\gamma_{(521)}^{(000)} = 36 + 4 - 16 - 16 = 8.$$

To obtain the multiplicity of the fundamental weight $(1,0,\ldots,0)$ we use the coefficients of

$$\{1 + (X_1 + X_2) + (X_1^2 + X_1X_2 + X_2^2) + (X_1^3 + X_1^2X_2 + X_1X_2^2 + X_2^3)\} \times \{1 + (X_1 + X_2) + (X_1^2 + X_1X_2 + X_2^2)\}^{n-1}, \quad (6.22)$$

for the case $a + b + c = 2n + 1$. (See, for example, Fig. 4.) The coefficients have the same recursion relation as those used in the previous example. For the SU(4) representation



$= (5,3,1)$,

Weyl's formula gives

$$\chi_{(5,3,1)} = [5,3,1] + [2,0,7] - [4,0,5] - [1,2,6],$$

$$\gamma_{(531)}^{(100)} = 58 + 10 - 22 - 37 = 9.$$

The generating function which gives the coefficients needed for the multiplicity of the fundamental weight

$$\lambda_i = (1,1,\ldots,\underset{i}{1},0,\ldots,0)$$

will be

$$(1 + h_1(m-1) + \cdots + h_{p+1}(m-1))^i \times (1 + h_1(m-1) + \cdots + h_p(m-1))^{n-i}, \quad (6.23)$$

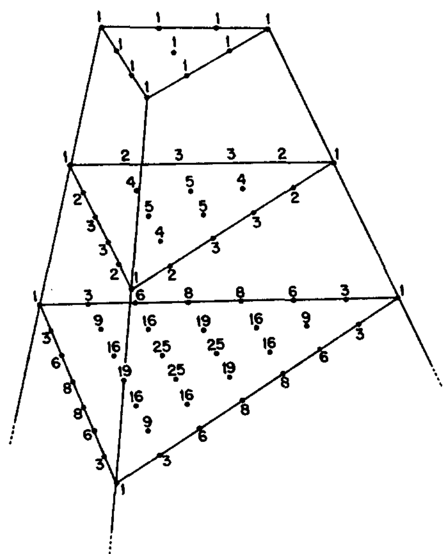


FIG. 4. Coefficients of $(1 + (S+T) + (S^2 + T^2 + ST) + (S^3 + T^3 + S^2T + ST^2))(1 + (S+T) + (S^2 + T^2 + ST))^{N-1}$.

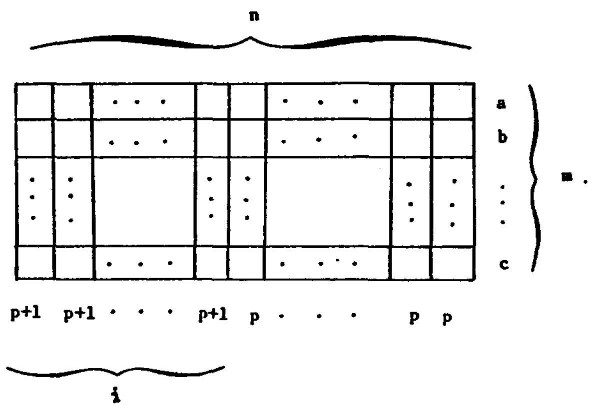
where

$$h_l(m-1) = h_l(X_1, X_2, \ldots, X_{m-1}).$$

This should be compared to Eq. (6.22) for the case $m = 3$, $p = 2$, $i = 1$. A suggested procedure is to find the coefficients in the expansion of

$$(1 + h_1(m-1) + \cdots + h_{p+1}(m-1))^i \quad (6.24)$$

by recursion and then to do the same for the coefficients of (6.23) in terms of those for (6.24). This may be done graphically up to $m = 4$. The generating function (6.23) enumerates solutions to the rectangle



Some points to note are the following.

(i) The number of indices $a,b,\ldots,c$ are the number of rows in the Young diagram. For four or more rows, the graphical method becomes cumbersome and recursion is the logical way to proceed.

(ii) Multiplicities are found by fixing a weight and varying the representation and the rank n of the group.

(iii) Multiplicities of weights near the center of the weight diagram are easy to find. Weights near the outer boundary are rather difficult by this method.

(iv) The number of terms[5] in $h_m(n)$ is $\binom{n+m-1}{m}$ so that

$$\dim(f_1,f_2,\ldots,f_n) = \det\left\{\binom{n-1+f_i+i-j}{f_i+i-j}\right\} \quad (6.25)$$

$$= \lim_{\substack{\epsilon_1 \to 1 \\ \vdots \\ \epsilon_n \to 1}} \chi(f_1,\ldots,f_n).$$

(v) Weyl's second formula expressed in terms of symmetric functions,[6] when expressed in expanded form, gives all weights and their multiplicities.

(vi) Formula (6.14) is also the generating function for simultaneous compositions of the numbers $a,b,\ldots,c$ into no more than k parts with restrictions.

VII. CONCLUSIONS

We have developed a direct combinatoric method for determining the multiplicity of a weight at or near the center of a weight diagram in a representation of SU(n). The method makes use of numbers which are solutions to a given combinatoric problem and are generated recursively. Actual multiplicity formulas may be obtained in specific cases. It would be amusing to find a similar construction for the groups Sp($2n$), SO($2n+1$), and SO($2n$). However, it is to be expected that any such method would be more complicated than the SU(n) case due to changes in Weyls' second formula and the Young labeling scheme. Any connection between Kostant's partition function and the Γ coefficients would also be interesting to find. However, the generator for the case of SU(4), being

$$[(1-x_1)(1-x_2)(1-x_3)(1-x_1x_2)(1-x_2x_3) \times(1-x_1x_3)(1-x_1x_2x_3)]^{-1}, \quad (7.1)$$

bears little relation to the Γ generator which is entirely in the numerator.

ACKNOWLEDGMENTS

We wish to express our thanks to Dr. Georgia Benkart and Dr. Steve Kass for helpful comments. Special thanks to Dr. Adam Bincer for assistance in the early stages of this work.

This research was supported in part by the University of Wisconsin Research Committee with funds granted by the Wisconsin Alumni Research Foundation, and in part by the United States Department of Energy under Contract No. De-AC02-76ER00881.

[1]J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory* (Springer, New York, 1972).
[2]H. Weyl, *The Classical Groups* (Princeton U. P., Princeton, NJ, 1973).
[3]P. A. MacMahon, *Combinatory Analysis* (Cambridge U. P., Cambridge, England, 1915).
[4]N. Ya. Vilenkin, *Combinatorics* (Academic, New York, 1971) (see, for example, problems 212 and 213, p. 186).
[5]A. B. Balantekin and I. Bars, J. Math. Phys. 25, 2028 (1984).
[6]E. D'Hoker, J. Math. Phys. 25, 1 (1984).

On the zeros of eigenfunctions of polynomial differential operators

E. Buendia, J. S. Dehesa, and M. A. Sanchez-Buendia
Departamento de Física Nuclear, Facultad de Ciencias, Universidad de Granada, Granada, Spain

(Received 23 February 1985; accepted for publication 10 May 1985)

It is shown that for polynomial eigenfunctions of an ordinary polynomial differential operator with coefficients depending only on the independent variable it is possible to determine the density of nodes around the mean without solving the corresponding eigenvalue problem. This is done by means of the first few moments, which can be directly expressed in terms of the above-mentioned coefficients. Also, very simple expressions for the asymptotic values (i.e., when the degree of the polynomial becomes very large) of these quantities are found. For illustration, these results are applied to various orthogonal polynomials, which satisfy ordinary differential equations of second, fourth, and/or sixth order.

I. INTRODUCTION

In many physical and mathematical problems we have to face the calculation of eigensolutions of operators (matrices, differential operators, integral operators). Generally, this is a difficult task. It would be desirable to obtain as much information as possible about the eigenvalues and eigenfunctions directly in terms of the parameters that characterize the operators, that is, without knowing the solution of the corresponding eigenvalue problem.

The collective or global properties of the eigenvalues and/or the nodes of the eigenfunctions of the operator are of special interest for mathematicians because of their novelty and for many-body physicists due to their relevant physical meaning. The most important properties of this type are the eigenvalue and/or node distribution densities or related quantities such as moments, cumulants, etc. Reference 1, for example, shows how the moments of the asymptotic eigenvalue density of a large class of Hamiltonian matrices can be calculated explicitly in terms of the matrix elements, without any diagonalization. Particular cases of this class of matrices are encountered in several branches of physics.[2] Similar problems are considered in Ref. 3.

The purpose of this paper is to obtain spectral properties of a collective nature of ordinary differential operators of any order without the need of solving the associated eigenvalue problem.

In 1980, Case[4] showed a simple method to find sum rules of zeros of polynomial solutions of certain ordinary polynomial differential operators in terms of their coefficients, and applied them to the classical orthogonal polynomials, which, as is well known, satisfy a second-order differential equation. The authors[5] have used the same method to investigate the distribution of the nodes of the eigensolutions of certain fourth-order differential operators that allowed them to obtain new properties of zeros of all the families of nonclassical orthogonal polynomials, which are eigensolutions of such operators.

Here it is our purpose to extend this investigation to ordinary polynomial differential operators (OPDO's in short) of *any* order. Precisely, we want to study the density of nodes of the eigenfunctions of the ordinary polynomial differential operators defined in Eq. (1) of Sec. II directly in terms of its coefficients. In particular we give explicit and recurrent expressions for the first four moments of the density of nodes. Also the corresponding quantities of the asymptotic limit of this density are shown in a simple manner. As a corollary we obtain the conditions to be fulfilled by the coefficients of the OPDO's so that the asymptotic density of the nodes will be Gaussian, rectangular, and semicircular around the mean. For the sake of illustration, the previous results are applied to systems of orthogonal polynomials, which fulfill an ordinary differential equation of second, fourth, or sixth order.

The paper is structured as follows: Section II is devoted to a sketch of the method used. Also, the relevant quantities are defined and some basic relations are given. Section III shows how the first four moments of the density of nodes of a polynomial eigenfunction are derived from these relations in a recurrent way. Section IV includes the derivation of the asymptotic values of these moments from the coefficients of Eq. (1) of the OPDO. In Sec. V the occurrence of some particular densities of nodes is discussed. Several applications to various systems of orthogonal polynomials are contained in Sec. VI and finally some concluding remarks are given.

II. METHODS AND BASIC TOOLS

Let us consider polynomials $P_N(x)$ satisfying a differential equation of the form

$$\sum_{i=0}^{n} g_i(x) P_N^{(i)}(x) = 0, \tag{1}$$

where $P_N^{(i)}(x)$ denotes the ith derivative of $P_N(x)$, and

$$g_i(x) = \sum_{j=0}^{i} a_j^{(i)} x^j \tag{2}$$

is a polynomial of degree no higher than i and such that the only coefficient $a_j^{(i)}$ that can be N-dependent is $a_0^{(0)}$. Further, we assume that all zeros of $P_N(x)$ are simple.

We are interested in the distribution of the zeros or nodes of the polynomial $P_N(x)$, which is characterized by the normalized-to-unity discrete density function defined by

$$\rho_N(x) = \frac{1}{N} \sum_{i=1}^{N} \delta(x - x_i). \tag{3}$$

The moments about the origin μ'_r and the central moments μ_r or moments around the mean of this function are given by

 0022-2488/85/112729-08$02.50

$$\mu_r'(N)=\frac{1}{N}\sum_{i=1}^{N}x_i^r=\frac{1}{N}y_r, \tag{4a}$$

$$\mu_r(N)=\frac{1}{N}\sum_{i=1}^{N}[x_i-\mu_1'(N)]^r$$
$$=\sum_{j=0}^{r}\binom{r}{j}\mu_{r-j}'(N)[-\mu_1'(N)]^j. \tag{4b}$$

Notice that $y_0=N$, and $\mu_r(N)=\mu_r'(N)$ if $\mu_1'=0$. It is possible to show[4] that these quantities can be calculated by the recursion relation

$$\sum_{i=2}^{n} i\sum_{j=0}^{i} a_j^{(i)}J_{r+j}^{(i)}=-a_0^{(1)}y_r-a_1^{(1)}y_{r+1}^{(1)},$$
$$r=0,1,2,..., \tag{5}$$

which allows us to calculate the moment of order $r+1$ in terms of the moments of lower orders. Indeed the $J_s^{(i)}$, $s=0,1,2,...$, are sum rules of the zeros of $P_N(x)$ defined by

$$J_r^{(i)}=\sum_{\neq}\frac{x_{l_1}^r}{(x_{l_1}-x_{l_2})(x_{l_1}-x_{l_3})\cdots(x_{l_1}-x_{l_i})} \tag{6}$$

($\Sigma_{\neq}$ means to sum over the l's subject to none of them being equal), which may be expressed in terms of the y_t with $t<s-i+1$. However this is a difficult problem, not yet solved for arbitrary values of s and i. Case[4] for $i=2$ and the present authors[5] for $i=3$ and 4 have given sum rules as functions of the y_t. Also the corresponding expression for $i=5$ is shown in Appendix A of this paper. Notice that in Eq. (5), the erroneous factorial $i!$ of Eq. (4) of Ref. 4 does not appear, as already pointed out in Ref. 5.

Since[4] $J_s^{(i)}=0$ for $0<s<i-2$, Eq. (5) reduces as follows:

$$\sum_{i=2}^{n} i\sum_{m=-1}^{r-1} a_{i+m+1-r}^{(i)}J_{i+m}^{(i)}=-a_0^{(1)}y_{r-1}-a_1^{(1)}y_r, \tag{7}$$

for $r=1,2,...$. From this equation one can make three important observations.

(i) The rth moment y_r only depends on the $r+1$ ordered sequences S_k of coefficients defined by

$$S_0=\{a_i^{(i)}\}_{i=1}^{n},\quad S_k=\{a_{i-k}^{(i)}\}_{i=k}^{n},\quad k=1,2,...,r. \tag{8}$$

(ii) This recurrent method cannot be used when all the members of sequences S_0 and S_1 are zero.

(iii) The evaluation of y_r previously required that one obtain the $J_{i+m}^{(i)}$, $m=-1,0,...,r-1$, in terms of the y_t with $t<m+1$. The latter is only known for $m=-1,0$, and 1. Indeed, Case[4] has found that

$$iJ_{i-1}^{(i)}=b_0^{(i)}, \tag{9a}$$
$$iJ_i^{(i)}=b_1^{(i)}y_1, \tag{9b}$$
$$iJ_{i+1}^{(i)}=[b_1^{(i)}-b_2^{(i)}]y_2+b_2^{(i)}y_1^2, \tag{9c}$$

where the convention

$$b_j^{(i)}=b_j^{(i)}(N)=\begin{cases}0, & i<j,\\ 1, & i=j,\\ \binom{i}{j}\prod_{t=j}^{i-1}(N-t), & i>j,\end{cases} \tag{10}$$

has been used. Appendix B contains the calculation of $J_{i+2}^{(i)}$ and $J_{i+3}^{(i)}$. In particular it is found that

$$iJ_{i+2}^{(i)}=(b_1^{(i)}-2b_2^{(i)}+2b_3^{(i)})y_3$$
$$+(2b_2^{(i)}-3b_3^{(i)})y_1y_2+b_3^{(i)}y_1^3, \tag{11a}$$
$$iJ_{i+3}^{(i)}=(b_1^{(i)}-3b_2^{(i)}+6b_3^{(i)}-6b_4^{(i)})y_4$$
$$+(2b_2^{(i)}-6b_3^{(i)}+8b_4^{(i)})y_1y_3$$
$$+(b_2^{(i)}-3b_3^{(i)}+3b_4^{(i)})y_2^2$$
$$+(3b_3^{(i)}-6b_4^{(i)})y_2y_1^2+b_4^{(i)}y_1^4. \tag{11b}$$

Equations (9) and (11) together with Eqs. (4) and (7) allow us to calculate the first four moments $\mu_r'(N)$, with $r=1,2,3,4$, of the density distribution $\rho_N(x)$. These quantities give a precise description of the density function around the mean and often allow us to obtain a good parametrization of this function. Furthermore these four quantities give the centroide, the variance $[\sigma^2=\mu_2(N)]$, the skewness [i.e., $\gamma_1=\mu_3(N)/\mu_2^{3/2}(N)$], and the excess or kurtosis [i.e., $\gamma_2=\mu_4(N)/\mu_2^2(N)-3$] of the density of zeros (nodes) of the polynomial eigenfunctions $P_N(x)$.

The centroide $\mu_1'(N)$ and the variance $\mu_2'(N)$ were implicitly calculated in Ref. 4 and for completeness we write them up in the next section. Also in Sec. III we show the derivation of the third and fourth moments.

III. THE FIRST FEW MOMENTS

A. The centroide $\mu_1'(N)=y_1/N$

Equations (7), (9a), and (9b) yield

$$y_1=-\frac{\Sigma_{i=1}^{n}a_{i-1}^{(i)}b_0^{(i)}}{\Sigma_{i=1}^{n}a_i^{(i)}b_1^{(i)}}. \tag{12}$$

B. The second moment $\mu_2'(N)=y_2/N$

Equations (7) and (9) give the following recurrence relation:

$$y_2=-\frac{\Sigma_{i=2}^{n}\{a_i^{(i)}b_2^{(i)}y_1^2+a_{i-2}^{(i)}b_0^{(i)}\}+\Sigma_{i=1}^{n}a_{i-1}^{(i)}b_1^{(i)}y_1}{\Sigma_{i=1}^{n}a_i^{(i)}(b_1^{(i)}-b_2^{(i)})}. \tag{13}$$

C. The third moment $\mu_3'(N)=y_3/N$

Putting $r=3$ in Eq. (7) we have

$$a_1^{(1)}y_3=-a_0^{(1)}y_2-\sum_{i=2}^{n}(a_{i-3}^{(i)}J_{i-1}^{(i)}+a_{i-2}^{(i)}J_i^{(i)}$$
$$+a_{i-1}^{(i)}J_{i+1}^{(i)}+a_i^{(i)}J_{i+2}^{(i)}).$$

By means of expressions (9) and (11a), this equation transforms as follows:

$$y_3=-\frac{A(y_1,y_2)}{\Sigma_{i=1}^{n}a_i^{(i)}(b_1^{(i)}-2b_2^{(i)}+2b_3^{(i)})}, \tag{14}$$

with

$$A(y_1,y_2)=\sum_{i=1}^{n}a_{i-1}^{(i)}(b_1^{(i)}-b_2^{(i)})y_2$$
$$+\sum_{i=2}^{n}[a_{i-3}^{(i)}b_0^{(i)}+a_{i-2}^{(i)}b_1^{(i)}y_1+a_{i-1}^{(i)}b_2^{(i)}y_1^2$$
$$+a_i^{(i)}\langle(2b_2^{(i)}-3b_3^{(i)})y_1y_2+b_3^{(i)}y_1^3\rangle].$$

D. The fourth moment $\mu_4'(N) = y_4/N$

Putting $r = 4$ in Eq. (7) and using the relations (9) and (11), it is straightforward to obtain the following recursion relation for y_4:

$$y_4 = -\frac{B(y_1, y_2, y_3)}{\sum_{i=1}^n a_i^{(i)}(b_1^{(i)} - 3b_2^{(i)} + 6b_3^{(i)} - 6b_4^{(i)})}, \tag{15}$$

with

$$\begin{aligned} B(y_1,y_2,y_3) = &\sum_{i=1}^n a_{i-1}^{(i)}(b_1^{(i)} - 2b_2^{(i)} + 2b_3^{(i)})y_3 \\ &+ \sum_{i=2}^n \langle a_{i-4}^{(i)} b_0^{(i)} + a_{i-3}^{(i)} b_1^{(i)} y_1 \\ &+ a_{i-2}^{(i)}[(b_1^{(i)} - b_2^{(i)})y_2 + b_2^{(i)} y_1^2] \\ &+ a_{i-1}^{(i)}[(2b_2^{(i)} - 3b_3^{(i)})y_1 y_2 + b_3^{(i)} y_1^3] \\ &+ a_i^{(i)}[(2b_2^{(i)} - 6b_3^{(i)} + 8b_4^{(i)})y_1 y_3 \\ &+ (b_2^{(i)} - 3b_3^{(i)} + 3b_4^{(i)})y_2^2 \\ &+ (3b_3^{(i)} - 6b_4^{(i)})y_1^2 y_2 + b_4^{(i)} y_1^4]\rangle. \end{aligned}$$

The convention $a_h^{(i)} = 0$ for negative values of h is used everywhere. Equations (12)–(15) confirm that once we know the five sets of parameters

$$S_k = \{a_{i-k}^{(i)}\}_{i=1}^n, \quad \text{with } k = 0,1,2,3,4, \tag{16}$$

the first four moments $\mu_r'(N)$ get fixed.

IV. ASYMPTOTIC MOMENTS

Here we want to show how to determine the asymptotic density of zeros $\rho(x)$ around the mean of the polynomial $P_N(x)$. We calculate precisely the asymptotic values of the first four moments about the origin of the density functin $\rho_N(x)$, that is,

$$\mu_r' = \lim_{N\to\infty} \mu_r'(N), \quad r = 1,2,3,4. \tag{17}$$

Due to Eq. (4a), to do that one has to study the asymptotic behavior of the corresponding y quantities. According to the previous sections, it is clear that this behavior gets fully determined by the five sequences S_k, $k = 0,1,2,3,4$. We assume that the elements in each sequence are ordered as shown by Eq. (16). We denote by $a_s^{(s)}$, $a_{r-1}^{(r)}$, $a_{t-2}^{(t)}$, $a_{u-3}^{(u)}$, $a_{v-4}^{(v)}$ the last nonvanishing element starting from the left in the sequences S_0, S_1, S_2, S_3, and S_4, respectively.

The results are summarized in the form of the following theorem.

Theorem 1: The asymptotic form of the first four y quantities is

$$y_1 = -(a_{r-1}^{(r)}/sa_s^{(s)})N^{r-s+1} + O(N^{k_1}), \tag{18a}$$

$$y_2 = -\left\langle \frac{s-2r-1}{2}\left[\frac{a_{r-1}^{(r)}}{sa_s^{(s)}}\right]^2 N^{2r-2s+1} + \frac{a_{t-2}^{(t)}}{sa_s^{(s)}} N^{t-s+1}\right\rangle + O(N^{k_2}), \tag{18b}$$

$$\begin{aligned} y_3 = -\Bigg\langle &\frac{a_{u-3}^{(u)}}{sa_s^{(s)}} N^{u-s+1} - (t+r-s+1)\frac{a_{r-1}^{(r)} a_{t-2}^{(t)}}{[sa_s^{(s)}]^2} N^{t+r-2s+1} \\ &- \left[\frac{r(s-3r)}{2} - \frac{(s-1)(2s-6r-1)}{6}\right]\left[\frac{a_{r-1}^{(r)}}{sa_s^{(s)}}\right]^3 N^{3r-3s+1}\Bigg\rangle + O(N^{k_3}), \end{aligned} \tag{18c}$$

$$\begin{aligned} y_4 = -\Bigg\langle &\frac{a_{v-4}^{(v)}}{sa_s^{(s)}} N^{v-s+1} - (u+r-s+1)\frac{a_{r-1}^{(r)} a_{u-3}^{(u)}}{[sa_s^{(s)}]^2} N^{u+r-2s+1} + \frac{s-2t-1}{2}\left[\frac{a_{t-2}^{(t)}}{sa_s^{(s)}}\right]^2 N^{2t-2s+1} \\ &- \frac{(2r+t)(s-t-2r) + (s-1)(2t+4r-2s+1)}{2}\,\frac{[a_{r-1}^{(r)}]^2 a_{t-2}^{(t)}}{[sa_s^{(s)}]^3} N^{2r+t-3s+1} \\ &- \frac{D(r,s)}{24}\left[\frac{a_{r-1}^{(r)}}{sa_s^{(s)}}\right]^4 N^{4r-4s+1}\Bigg\rangle + O(N^{k_4}), \end{aligned} \tag{18d}$$

where

$$\begin{aligned} D(r,s) = \langle &(s-r-1)[12r(s-3r) - 4(s-1)(2s-6r-1)] + 4r(r-1)(7r-3s+1) \\ &+ (s-1)[3(s-2r-1)(s+2r-3) - (s-2)(s-3)]\rangle. \end{aligned}$$

The values of k_i, for $i = 1,2,3,4$, in Eqs. (18) are equal to the maximal values of the sets E_i defined as follows:

$$\begin{aligned} E_1 &= \{r-s+1\}, \\ E_2 &= \{2r-2s+1, t-s+1\}, \\ E_3 &= \{u-s+1, t+r-2s+1, 3r-3s+1\}, \\ E_4 &= \{v-s+1, u+r-2s+1, 2t-2s+1, 2r \\ &\quad + t-3s+1, 4r-4s+1\}. \end{aligned}$$

Then the first four moments of the asymptotic density of zeros $\rho(x)$ easily follow from the expressions

$$\mu_r' = \lim_{N\to\infty} \frac{y_r}{N^{k_r+1}}, \quad r = 1,2,3,4. \tag{19}$$

This is the end of the theorem.

The proof of this theorm is fairly easy once one observes that, according to Eq. (10),

$$b_j^{(i)} = \binom{i}{j} N^{i-j} + O(N^{i-j-1}).$$

Taking this value in Eqs. (12)–(15), one has for the four y quantities a ratio of two polynomials in the variable. Then it suffices to divide the term of the highest N power of the polynomial in the numerator by the corresponding term of the polynomial in the denominator in order to obtain Eqs. (18) of our theorem.

V. SOME ASYMPTOTIC DENSITIES OF NODES

The theorem mentioned in the previous section allows us to know many things about the asymptotic density of the nodes of the eigenfunction $P_N(x)$ of the polynomial differential operator defined by Eqs. (1) and (2). Here we will mention only the following result.

Theorem 2: Assume that all members of the sequences S_1 and S_3 of the differential operator [Eqs. (1) and (2)] are zero and that the nonvanishing coefficients $a_s^{(s)}$ and $a_{t-2}^{(t)}$ exist. Then $\mu_1' = \mu_3' = 0$ and

$$\mu_4' = [(2t - s + 1)/2]\, \mu_2'^2. \tag{20}$$

These moments correspond to the first four moments of a density function of the Gaussian type, of the rectangular or uniform type, and of the semicircular type centered at the origin according to s being equal to $2t - 5$, $2t - \frac{18}{5}$, and $2t - 3$, respectively.

Even more, if $s = 2t - 1$ then the nodes of the eigenfunction $P_N(x)$ satisfying (1) and (2) are asymptotically piled up at exactly two points that are one standard deviation away from the mean.

Proof: Equation (20) easily follows from Eqs. (18b)–(18d). On the other hand, the Gaussian, uniform, and semicircular densities are given by

$$\rho(x) = (2\pi)^{-1/2} \exp(-\tfrac{1}{2}x^2), \quad -\infty < x < \infty,$$
$$\rho(x) = (2h)^{-1}, \quad -h \leqslant x \leqslant h,$$
$$\rho(x) = (2/\pi)(1 - x^2)^{1/2}, \quad -1 \leqslant x \leqslant 1.$$

Since all three densities are symmetric, the odd moments vanish, and the even moments are, respectively,

$$\mu_{2n}' = (2n-1)!!,$$
$$\mu_{2n}' = h^{2n}/(2n+1),$$
$$\mu_{2n}' = \frac{2^{-2n}}{n+1}\binom{2n}{n}.$$

These equations verify that $\mu_4' = 3\mu_2^2$, $\frac{9}{5}\mu_2'^2$, and $2\mu_2'^2$, also, respectively. Finally, Guttman[6] has shown that when $\mu_4' = \mu_2'^2$ the density function reduces to exactly two points, which are one standard deviation away from the mean. The last condition is fulfilled if $s = 2t - 1$ according to Eq. (20).

VI. APPLICATIONS

To see how the general results found in the previous sections work, let us specify the order n of the differential operator defined by (1).

A. Second-order differential operator (i.e., $n = 2$)

Here all we need are the three sequences

$$S_0 = (a_1^{(1)}, a_2^{(2)}), \quad S_1 = (a_0^{(1)}, a_1^{(2)}), \quad S_2 = (a_0^{(2)}).$$

Then the expressions of the first four moments given in Sec. III are strongly reduced as follows:

$$y_1 = -\frac{a_0^{(1)} b_0^{(1)} + a_1^{(2)} b_0^{(2)}}{a_1^{(1)} + a_2^{(2)} b_1^{(2)}}, \tag{21a}$$

$$y_2 = -\frac{a_0^{(1)} y_1 + a_2^{(2)} y_1^{(2)} + a_1^{(2)} b_1^{(2)} y_1 + a_0^{(2)} b_0^{(2)}}{a_1^{(1)} + a_2^{(2)} (b_1^{(2)} - 1)}, \tag{21b}$$

$$y_3 = -\frac{a_0^{(1)} y_2 + 2a_2^{(2)} y_1 y_2 + a_1^{(2)} [(b_1^{(2)} - 1) y_2 + y_1^2] + a_0^{(2)} b_1^{(2)} y_1}{a_1^{(1)} + a_2^{(2)} (b_1^{(2)} - 2)}, \tag{21c}$$

$$y_4 = -\frac{a_0^{(1)} y_3 + a_2^{(2)} (2y_1 y_3 + y_2^2) + a_1^{(2)} [(b_1^{(2)} - 2) y_3 + 2y_1 y_2] + a_0^{(2)} [(b_1^{(2)} - 1) y_2 + y_1^2]}{a_1^{(1)} + a_2^{(2)} (b_1^{(2)} - 3)}. \tag{21d}$$

Let us specialize to various classical orthogonal polynomials. In doing so, we find several results already obtained by Case.[4]

1. Hermite polynomials $H_N(x)$

Now $S_0 = \{-2, 0\}$, $S_1 = \{0, 0\}$, and $S_2 = \{1\}$. Since all memebers of S_1 are zero, Eqs. (21a) and (21c) give $\mu_1'(N) = \mu_3'(N) = 0$. Also, Eqs. (21b) and (21c) yield

$$\mu_2'(N) = \tfrac{1}{2}(N-1), \quad \mu_4'(N) = \tfrac{1}{2}(N-1)(N - \tfrac{3}{2}). \tag{22}$$

Notice that $\mu_2'(N) = N/2 + O(N^{-1})$ and $\mu_4'(N) = N^2/2 + O(N)$. These asymptotic values could have been obtained directly from Theorem 1. For completeness let us point out that in a context of random matrices, Bronk[7] has shown that for large N the density $\rho_N(x)$ is given by

$$\rho_N(x) = \begin{cases} (1/\pi)(2N - x^2)^{1/2}, & |x| \leqslant 2N, \\ 0, & |x| \geqslant 2N. \end{cases}$$

2. Laguerre polynomials $L_N^{(\alpha)}(x)$

Here $S_0 = \{-1, 0\}$, $S_1 = \{\alpha + 1, 1\}$, and $S_2 = \{0\}$. Equations (21) allow us to find the values

$$\mu_1'(N) = N + \alpha, \tag{23a}$$

$$\mu_2'(N) = (N + \alpha)(2N + \alpha - 1), \tag{23b}$$

$$\mu_3'(N) = (N+\alpha)[N(N+\alpha) + (2N+\alpha-2)(2N+\alpha-1)], \tag{23c}$$

$$\mu_4'(N) = (N+\alpha)[N(N+\alpha)(6N+3\alpha-5) + (2N+\alpha-3)(2N+\alpha-2)(2N+\alpha-1)]. \tag{23d}$$

On the other hand, the asymptotic values of these quantities can be obtained in a straightforward manner either from Eqs. (23) or directly from Theorem 1. The values of these asymptotic moments are

$$\mu_1'(N) = N + O(N^0), \quad \mu_2'(N) = 2N^2 + O(N),$$
$$\mu_3'(N) = 5N^3 + O(N^2), \quad \mu_4'(N) = 14N^4 + O(N^3).$$

Bronk[7] has also shown that for large N the density $\rho_N(x)$ has the following form:

$$\rho_N(x) = \begin{cases} (4\pi x)^{-1}[-x^2 + (4N+2\alpha-2)x - (\alpha-1)^2]^{1/2}, \\ x_0 < x < x_1, \\ 0, \quad \text{otherwise,} \end{cases}$$

where x_0 and x_1 are the roots of the radicand.

3. Chebyshev polynomials $C_N(x)$

Here $S_0 = \{-1,-1\}$, $S_1 = \{0,0\}$, and $S_2 = \{1\}$. Here since $s = 2$ and $t = 2$, Eqs. (21) produce the following known moments:

$$\mu_1'(N) = \mu_3'(N) = 0, \quad \mu_2'(N) = \tfrac{1}{2} + O(N^{-1}),$$
$$\mu_4'(N) = \tfrac{3}{8} + O(N^{-1}).$$

Bronk[7] also found that the density $\rho_N(x)$ is given for large N as follows:

$$\rho_N(x) = \begin{cases} N\pi^{-1}(1-x^2)^{1/2}, & \text{for } |x| \leqslant 1, \\ 0, & \text{otherwise.} \end{cases}$$

4. Jacobi polynomials $P_N^{(\alpha,\beta)}(x)$

Now $S_0 = \{-\alpha-\beta-2,-1\}$, $S_1 = \{\beta-\alpha,0\}$, and $S_2 = \{1\}$. Equations (21) yield

$$\mu_1'(N) = \frac{\beta-\alpha}{2N+\alpha+\beta},$$

$$\mu_2'(N) = \frac{(2N+\alpha+\beta)^2(N-1) + (N+\alpha+\beta)(\beta-\alpha)^2}{(2N+\alpha+\beta)^2(2N+\alpha+\beta-1)},$$

$$\mu_3'(N) = \frac{(\beta-\alpha)y_2 - 2y_1y_2 + 2(N-1)y_1}{N(2N+\alpha+\beta-2)},$$

$$\mu_4'(N) = \frac{(\beta-\alpha)y_3 - 2y_1y_3 - y_2^2 + (2N-3)y_2 + y_1^2}{N(2N+\alpha+\beta-3)}.$$

Asymptotically these expressions reduce as follows:

$$\mu_1'(N) = \tfrac{1}{2}(\beta-\alpha)N^{-1} + O(N^{-2}),$$
$$\mu_2'(N) = \tfrac{1}{2} + O(N^{-1}),$$
$$\mu_3'(N) = \tfrac{1}{2}(\beta-\alpha)N^{-1} + O(N^{-2}),$$
$$\mu_4'(N) = \tfrac{3}{8} + O(N^{-1}).$$

These asymptotic moments are an immediate consequence of Eqs. (18).

B. Fourth-order differential operator (i.e., $n = 4$)

In this case the polynomials $P_N(x)$ are described by the five following sequences: $S_0 = \{a_i^{(i)}\}_{i=1}^4$, $S_1 = \{a_{i-1}^{(i)}\}_{i=1}^4$, $S_2 = \{a_{i-2}^{(i)}\}_{i=2}^4$, $S_3 = \{a_{i-3}^{(i)}\}_{i=3}^4$, and $S_4 = \{a_0^{(4)}\}$. The only systems of orthogonal polynomials, apart from the classical ones, which satisfy the differential equation [(1) and (2)] are the so-called Krall–Legendre, Krall–Laguerre, and Krall–Jacobi polynomials.[8]

In Ref. 5 we have given explicit expressions for the first four moments of the density of zeros of the Krall polynomials. These same expressions can be obtained in a much more simple and straightforward way from Eqs. (12)–(15). Here we will not write such expressions again but only calculate their asymptotic values as a further illustration of the usefulness of Theorem 1 found in Sec. IV.

1. Krall–Legendre polynomials $\mathscr{K}_N^{(\alpha)}(x)$

Here $S_0 = \{8\alpha, 4\alpha+12, 8, 1\}$, $S_1 = \{0,0,0,0\}$, $S_2 = \{-4\alpha-12,-8,-2\}$, $S_3 = \{0,0\}$, and $S_4 = \{1\}$. Since all the elements of S_1 and S_3 vanish, Theorem 1 yields $\mu_1' = \mu_3' = 0$. Besides, since $s = 4$, $t = 4$, and $v = 4$, we obtain for the sets E_i, $i = 2,4$, the following values:

$$E_2 = (-7,1), \quad E_4 = (1,-7,1,-7,1).$$

Then according to Eqs. (18b) and (18d) the values of the asymptotic moments are

$$\mu_2' = -a_2^{(4)}/4a_4^{(4)} + O(N^{-1}) = \tfrac{1}{2} + O(N^{-1}),$$
$$\mu_4' = -a_0^{(4)}/4a_4^{(4)} + a_2^{(4)2}/4a_4^{(4)} + O(N^{-1}), = \tfrac{3}{8} + O(N^{-1}).$$

Let us finally say that it is possible to prove that

$$\mu_{2k}' = 2^{-2k}\binom{2k}{k} \quad \text{and} \quad \mu_{2k-1}' = 0,$$

for $k = 1,2,3,\ldots$.

2. Krall–Laguerre polynomials $\mathscr{L}_N^{(R)}(x)$

Now $S_0 = \{2R+2,1,0,0\}$, $S_1 = \{-2R,-2R-6,-2,0\}$, $S_2 = \{0,4,1\}$, $S_3 = \{0,0\}$, and $S_4 = \{0\}$. Here $s = 2$, $r = 3$, and $t = 4$, since the elements of the E sets are $E_1 = \{2\}$, $E_2 = \{3,3\}$, $E_3 = \{-1,4,4\}$, and $E_4 = \{-1,0,5,5,5\}$. Then the asymptotic moments are, according to Theorem 1,

$$\mu_1' = -\frac{a_2^{(3)}}{2a_2^{(2)}} = 1,$$

$$\mu_2' = -\frac{5}{2}\frac{a_2^{(3)2}}{2a_2^{(2)}} + \frac{a_2^{(4)}}{2a_2^{(2)}} = 2,$$

$$\mu_3' = -6\frac{a_2^{(3)}a_2^{(4)}}{2a_2^{(2)2}} + 8\frac{a_2^{(3)3}}{2a_2^{(2)}} = 5,$$

$$\mu_4' = -\frac{63}{2}\frac{a_2^{(3)2}a_2^{(4)}}{2a_2^{(2)3}} - \frac{693}{24}\frac{a_2^{(3)4}}{2a_2^{(2)}} = 14.$$

It is easy to prove that

$$\mu_r' = \frac{1}{r+1}\binom{2r}{r}, \quad r = 0,1,2,\ldots.$$

3. Krall–Jacobi polynomials $\mathscr{J}_N^{(\alpha,M)}(x)$

In this case the S sets are

$$S_0 = \{(\alpha+2)(2\alpha+2+2M), \alpha^2 + 9\alpha + 14 + 2M, 2\alpha+8, 1\},$$

$$S_1 = \{-2M, -6\alpha - 12 - 2M, -2\alpha - 12, -2\},$$
$$S_2 = \{0,4,1\}, \quad S_3 = \{0,0\}, \quad S_4 = \{0\}.$$

Here $s = 4$, $r = 4$, and $t = 4$. Then the corresponding E sets are $E_1 = (1)$, $E_2 = (1,1)$, $E_3 = (-3,1,1)$, and $E_4 = (-3,-3,1,1,1)$. In this case Theorem 1 produces the values

$$\mu_1' = \tfrac{1}{2}, \quad \mu_2' = \tfrac{3}{8}, \quad \mu_3' = \tfrac{5}{16}, \quad \mu_4' = \tfrac{35}{128}.$$

In fact it is possible to prove that

$$\mu_r' = 2^{-2r}\binom{2r}{r}, \quad r = 0,1,2,\ldots.$$

The distribution of zeros of the Krall polynomials is studied in detail in Ref. 5 from their differential equations and in Ref. 9 from their explicit expressions.

C. Sixth-order differential operator (i.e., $n = 6$)

Here the density of zeros of the polynomial $P_N(x)$ is fully characterized by the seven following sets of parameters $S_0 = \{a_i^{(i)}\}_{i=1}^6$, and $S_k = \{a_{i-k}^{(i)}\}_{i=k}^6$, for $k = 1,2,3,4,5,6$. Recently a new set of polynomials $\mathbb{L}_N^{(A,B,C)}(x)$ has been found. The polynomials are solutions[10] of a sixth-order differential equation and are orthogonal on $[-1,1]$ with respect to a weight distribution. These polynomials, named the Krall–Littlejohn polynomials, are generalizations of the Legendre and Krall–Legendre polynomials and satisfy many properties shared by the classical orthogonal polynomials of Jacobi, Laguerre, and Hermite. They are defined by the sequences

$$\begin{aligned}
S_0 = \{&24ABC^2 + 12AC + 12BC, 12ABC^2 + 42AC \\
&+ 42BC + 72, 24AC + 24BC \\
&+ 168, 3AC + 3BC + 96, 18, 1\},\\
S_1 = \{&12BC - 12AC, 12BC - 12AC, 0,0,0,0\},\\
S_2 = \{&-(12ABC + 30AC + 30BC + 72),\\
&-(24AC + 24BC + 168),\\
&-(6AC + 6BC + 132), -36, -3\},\\
S_3 = \{&0,0,0,0\}, \quad S_4 = \{3AC + 3BC + 36, 18, 3\},\\
S_5 = \{&0,0\},\\
S_6 = \{&-1\}.
\end{aligned}$$

Equations (10)–(13) give the first few moments of the wanted density of zeros. We will not write them explicitly here. What we will do is to apply the results of Sec. IV to investigate the asymptotic density of zeros of $\mathbb{L}_N^{(A,B,c)}(x)$. From the given sequences we observe that $s = 6$, $r = 2$, $t = 6$, and $v = 6$. Then the E sets are $E_1 = \{3\}$, $E_2 = \{-7,1\}$, $E_3 = \{-5,-3,-11\}$, and $E_4 = \{1,-9,-7,1,-17\}$. The corresponding asymptotic moments turn out to be given by

$$\mu_1' = \tfrac{1}{60}(AC - BC), \quad \mu_2' = \tfrac{1}{2},$$
$$\mu_3' = \tfrac{1}{60}(AC - BC), \quad \mu_4' = \tfrac{3}{8}.$$

In fact, starting from the three-recurrence relation verified by these polynomials it is possible to prove that the asymptotic moments of even order are given by

$$\mu_{2k}' = 2^{-2k}\binom{2k}{k}, \quad k = 0,1,2,\ldots.$$

It is interesting to remark that these quantities coincide with the corresponding moments of even order of the Krall–Legendre polynomials and also with the first k moments of the Jacobi and Krall–Jacobi polynomials.

VII. CONCLUSION

It has been shown that for polynomials satisfying ordinary polynomial differential equations with coefficients depending only on the independent variable, the normalized density of zeros around the mean can be found. This is done by means of the first four moments, which are given explicitly. Also it is found that the asymptotic values of these quantities are particularly easy to obtain directly from the coefficients which characterize the differential equations.

For illustration, the previous results are applied to all the orthogonal polynomials satisfying a second- and/or fourth-order differential equation of the type of Eqs. (1) and (2). Also the (asymptotic) moments of the density of the zeros of a new system of orthogonal polynomials verifying a sixth-order differential equation are given. It is striking to point out that those moments of even order are equal to the even moments of the Krall-Legendre polynomials and also to the moments of the Jacobi and Krall–Jacobi polynomials.

ACKNOWLEDGMENT

We are grateful for the partial support of the Spanish Comision Asesora de Investigacion Cientifica y Técnica (CAICYT).

APPENDIX A: DERIVATION OF $J_r^{(5)}$

By definition

$$J_r^{(5)} = \sum_{\neq} \frac{x_{l_1}^r}{(x_{l_1} - x_{l_2})(x_{l_1} - x_{l_3})(x_{l_1} - x_{l_4})(x_{l_1} - x_{l_5})}. \quad \text{(A1)}$$

Here we want to express $J_r^{(5)}$ in terms of the quantities y_t with $t<r-4$. First of all we transform Eq. (A1) into the form[4]

$$\begin{aligned}
J_r^{(5)} = &\frac{4}{5}\sum_{\neq} \frac{x_{l_1}^r - x_{l_2}^r}{(x_{l_2} - x_{l_3})(x_{l_2} - x_{l_4})(x_{l_1} - x_{l_5})(x_{l_1} - x_{l_2})}\\
&- \frac{2}{5}\sum_{\neq} \frac{x_{l_1}^r - x_{l_3}^r}{(x_{l_2} - x_{l_3})(x_{l_2} - x_{l_4})(x_{l_3} - x_{l_5})(x_{l_1} - x_{l_3})}\\
&+ \frac{2}{5}\sum_{\neq} \frac{x_{l_1}^r - x_{l_5}^r}{(x_{l_2} - x_{l_3})(x_{l_2} - x_{l_4})(x_{l_1} - x_{l_5})(x_{l_3} - x_{l_5})}.
\end{aligned}$$

The use of the relation

$$\frac{x_l^r - x_m^r}{x_l - x_m} = \sum_{i=0}^{r-1} x_l^{r-1-i} x_m^i$$

and adequate algebraic manipulations allow us to write $J_r^{(5)}$ as follows:

$$J_r^{(5)} = \frac{1}{5}\sum_{s=3}^{r-1}\sum_{t=2}^{s-1}\sum_{u=1}^{t-1}\sum_{v=0}^{u-1} \times B_5(r-1-s, s-1-t, t-1-u, u-1-v, v). \quad \text{(A2)}$$

The B function is defined by

$$B_5(t_1,t_2,t_3,t_4,t_5)=\sum_{\neq} x_{l_1}^{t_1} x_{l_2}^{t_2} x_{l_3}^{t_3} x_{l_4}^{t_4} x_{l_5}^{t_5}.$$

The same procedure described in Appendices A and B of Ref. 5 yields

$$\begin{aligned}B_5(t_1,t_2,t_3,t_4,t_5)&\\ &= y_{t_1} y_{t_2} y_{t_3} y_{t_4} y_{t_5} - 10 y_{t_1+t_2} y_{t_3} y_{t_4} y_{t_5}\\ &\quad + 20 y_{t_1+t_2+t_3} y_{t_4} y_{t_5} + 15 y_{t_1+t_2} y_{t_3+t_4} y_{t_5}\\ &\quad - 30 y_{t_1+t_2+t_3+t_4} y_{t_5} - 20 y_{t_1+t_2+t_3} y_{t_4+t_5}\\ &\quad + 24 y_{t_1+t_2+t_3+t_4+t_5}. \qquad \text{(A3)}\end{aligned}$$

Taking this expression with $t_1=r-1-s$, $t_2=s-1-t$, $t_3=t-1-u$, $t_4=u-1-v$, $t_5=v$, and $\Sigma_{i=1}^{5} t_i = r-4$ and using Eq. (A2), we obtain

$$\begin{aligned}J_r^{(5)}=\frac{1}{5}\Big\{&\langle 5N^4-10rN^3+[10[(r-2)(r-3)+4]+15(2r-5)]N^2\\ &-[5[(r-1)(r-2)(r-3)+6]+10r(r-3)]N+r(r-1)(r-2)(r-3)\rangle y_{r-4}\\ &+\sum_{s=4}^{r-2}\langle 10N^3-30(s-1)N^2+[15s(r-s)+20[(s-1)(s-2)+1]]N\\ &-[10r(s-2)(r-2s+1)+15s(s-1)(s-2)]\rangle y_{r-1-s}y_{s-3}(1-\delta_{r,5})\\ &+\sum_{s=5}^{r-2}\sum_{t=3}^{s-2}\langle 10N^2-10(t-2)N+[10t(t-1)+15(r-s)(s-t)]\rangle y_{r-1-s}y_{s-1-t}y_{t-2}\prod_{i=5}^{6}(1-\delta_{r,i})\\ &+\sum_{s=6}^{r-2}\sum_{t=4}^{s-2}\sum_{u=2}^{t-2}\langle 5N-10u\rangle y_{r-1-s}y_{s-1-t}y_{t-1-u}y_{u-1}\prod_{i=5}^{7}(1-\delta_{r,i})\\ &+\sum_{s=7}^{r-2}\sum_{t=5}^{s-2}\sum_{u=3}^{t-2}\sum_{v=1}^{u-2} y_{r-1-s}y_{s-1-t}y_{t-1-u}y_{u-1-v}y_v\prod_{i=5}^{8}(1-\delta_{r,i})\Big\},\end{aligned}$$

valid for $r\geqslant 5$.

APPENDIX B: DERIVATION OF $J_{i+2}^{(i)}$ AND $J_{i+3}^{(i)}$

From Eqs. (29) and (31) of Ref. 5 and Eq. (A2) one notes that there exists a close relation between the quantities $J_r^{(i)}$ and $B_i(t_1,t_2,\ldots,t_i)$. Even more one realizes that

$$k\equiv\sum_{j=1}^{i} t_j = r-i+1.$$

Besides, the B quantities are fully symmetric. Then we can assume without any loss of generality that the nonvanishing arguments, if they exist, are the first ones. One easily has

$$\begin{aligned}&B_i(0,0,\ldots,0)=\prod_{s=0}^{i-1}(N-s),\\ &B_i(t_1,t_2,\ldots,t_j,0,0,\ldots,0)\\ &\quad=\Big\{\prod_{s=j}^{i-1}(N-s)\Big\}B_j(t_1,t_2,\ldots,t_j),\end{aligned}$$

for $i\geqslant 1$.

To evaluate $J_{i+2}^{(i)}$ and $J_{i+3}^{(i)}$ one needs to know the explicit expressions of the B quantities for $k=3,4$. These expressions can be obtained directly from the explicit forms of B_3, B_4 already shown in Appendices A and B of Ref. 5, respectively. The final results depend on the partitions of the numbers 3 and 4. They are as follows.

(a) $k=3$: The partitions of number 3 are $\{(3),(2,1),(1,1,1)\}$. The result in this case is

$$B_i(t_1,t_2,\ldots,t_i)=\begin{cases}\prod_{s=1}^{i-1}(N-s)\,y_3, & \text{for } (3),\\ \prod_{s=2}^{i-1}(N-s)(y_1y_2-y_3), & \text{for } (2,1),\\ \prod_{s=3}^{i-1}(N-s)(y_1^3-3y_1y_2+2y_3), & \text{for } (1,1,1).\end{cases}$$

(b) $k=4$: The partitions of the number 4 are $\{(4),(3,1),(2,2),(2,1,1),(1,1,1,1)\}$. Then

$$B_i(t_1,t_2,\ldots,t_i)=\begin{cases}\prod_{s=1}^{i-1}(N-s)\,y_4, & \text{for } (4),\\ \prod_{s=2}^{i-1}(N-s)(y_1y_3-y_4), & \text{for } (2,1,1),\\ \prod_{s=2}^{i-1}(N-s)(y_2^2-y_4), & \text{for } (2,2),\\ \prod_{s=3}^{i-1}(N-s)(y_2y_1^2-2y_3y_1-y_2^2+2y_4), & \text{for } (2,1,1),\\ \prod_{s=4}^{i-1}(N-s)(y_1^4-6y_1^2y_2+8y_1y_3+3y_2^2-6y_4), & \text{for } (1,1,1,1).\end{cases}$$

Taking this expression in the forms contained in Appendices A and B of Ref. 5 and Appendix A of this paper for $J_r^{(i)}$, with $i=2,3,4,5$, we obtain the wanted values for $J_{i+m}^{(i)}$, with $m=2,3$ and $i=2,3,4,5$, which allow a straightforward generalization for any i. The final results are Eqs. (11).

[1]J. S. Dehesa, J. Phys. A: Math Gen. **11**, L223 (1978); M. C. Bosca and J. S. Dehesa, J. Phys. A: Math. Gen. **17**, 3487 (1984).
[2]J. P. Gaspard and F. Cyrot-Lackman, J. Phys. C: Sol. State Phys. **6**, 3077 (1973); D. de Brucq and E. Tirapegui, Nuovo Cimento A **67**, 225 (1970).
[3]B. Simon, J. Phys. A: Math. Gen. **15**, 2981 (1982); W. Craig and B. Simon, Commun. Math. Phys. (to be published). See also references therein.
[4]K. M. Case, J. Math. Phys. **21**, 702 (1980).
[5]J. S. Dehesa, E. Buendia, and M. A. Sanchez-Buendia, J. Math. Phys. **26**, 1547 (1985).
[6]L. Guttman, Ann. Math. Statist. **19**, 277 (1948).
[7]B. V. Bronk, J. Math. Phys. **5**, 1661 (1964).
[8]A. M. Krall, Proceed. R. Soc. Edinburgh Ser. A **87**, 271 (1981). See also references therein.
[9]J. S. Dehesa and F. J. Galvez, preprint 1985.
[10]L. L. Littlejohn, Question. Math. **5**, 255 (1982); L. L. Littlejohn and M. A. Krall, Lecture Notes Math. **964**, 435 (1983).

Solution of an extremum problem pertaining to analytic extrapolation techniques

G. Auberson and S. Ciulli
Département de Physique Mathématique,[a] *Université des Sciences et Techniques du Languedoc, 34060 Montpellier Cedex, France*

(Received 23 November 1984; accepted for publication 29 March 1985)

A preliminary problem appearing in analytic extrapolation procedures when the "physical data" are given on a line contained in the analyticity domain and their errors are measured through the norm of the supremum is solved. As an interesting by-product, Blaschke functions are constructed, which, in some sense, generalize the usual Chebyshev polynomials.

I. INTRODUCTION

Analytic extrapolation is a useful mathematical tool in theoretical physics, especially in elementary particle theory where analyticity can be viewed as an information conveyor.[1] Among the various problems arising in this context, the following one has an interest of its own. A function F analytic in some domain $\mathscr{D}$ is known to be bounded (in modulus) by a given function on the *boundary* of $\mathscr{D}$ (physical conditions on the "timelike cut"). Further, the values of F are given on some *interior* set Γ, possibly within some error channel (physical data in the "spacelike region"). One looks for the range of possible values of F at some interior point z_0 not on Γ. A number of variants of the problem can be conceived, according to the choice of the set Γ and the way both the boundedness condition and the errors of the physical data are expressed (choice of norm). When Γ simply consists in a finite set of points and the errors are assumed to vanish, standard Hardy space techniques[2] readily yield the solution, e.g., in the L^2 and the L^∞ norms[3,4] (for the latter choice, the Nevanlinna–Pick–Schur construction[1,5] is used as an essential ingredient). In a further step, the solution can be adapted to account for nonzero errors.[6] When Γ is a continuum, nonvanishing errors have to be included from the very beginning in order to avoid obvious triviality or inconsistency. Such a problem has been solved in the case of a real segment Γ and for various L^2 norms.[7] We wish to consider here the same problem within the L^∞ norm. In more distinct terms, it can be formulated as follows: There are given a domain $\mathscr{D}\subset\mathbb{C}$, a simple arc $\Gamma\subset\mathscr{D}$, a point $z_0\in\mathscr{D}$ not on Γ [Fig. 1(a)], a positive function $M(z)$ defined on the boundary $\partial\mathscr{D}$, a complex function $R(z)$, and a positive function $\eta(z)$ defined on Γ. One is asked to find the range of $F(z_0)$ when F goes through the set of functions holomorphic in $\mathscr{D}$ and is subjected to the constraints

$$|F(z)|\leqslant M(z),\quad z\in\partial\mathscr{D},$$
$$|F(z)-R(z)|\leqslant\eta(z),\quad z\in\Gamma.$$

After mapping the domain $\mathscr{D}$ onto the unit disk (in such a way that z_0 is sent onto the center), and by using an appropriate outer function[1,2,4] to get rid of $M(z)$, one can restate the problem in the standard Hardy space H^∞ as

$$\text{range of } f(0)=?,$$
$$f\in H^\infty,$$
$$\|f\|_\infty\equiv\text{ess sup}|f(e^{i\theta})|\leqslant 1, \tag{1.1}$$
$$|f(z)-r(z)|\leqslant\epsilon(z),\quad z\in\gamma\ (=\text{image of }\Gamma).$$

For arbitrary functions $r(z)$ and $\epsilon(z)$, there is clearly little hope to obtain a solution in a closed form, or even to get a significant insight into the structure of the corresponding extremal functions f. In this paper, we shall rather address ourselves to a simplified version, namely the particular case where γ is the real interval $[a,b]$ $(0<a<b<1)$ [Fig. 1(b)], $r(z)\equiv 0$, and $\epsilon(z)=\text{const}=\epsilon$ (nontriviality obviously requires that $\epsilon<1$). That is to say, we intend to solve

$$\hat{f}\equiv\sup|f(0)|=?,$$
$$f\in H^\infty,\quad \|f\|_\infty\leqslant 1,\quad |f(x)|\leqslant\epsilon,\quad x\in\gamma. \tag{1.2}$$

By "solving" it we mean something more constructive than just settling the question of existence and uniqueness of an extremal function (which certainly could be established by more abstract arguments). In fact, its structural properties will be described in detail, and the problem brought to a point where the implementation of numerical methods could be contemplated. In view of physical application, the main assumption to be removed in a further step would be of course $r(z)\equiv 0$.

We shall first show that in (1.2) the functions f can be restricted to the class of pure Blaschke products with positive zeros (Sec. II). Then, an auxiliary problem needs to be solved, which is interesting by itself: it leads to the construction of a particular class of Blaschke products playing a role analogous to that of the Chebyshev polynomials in the usual context of best polynomial approximation (Sec. III). The properties of the final solution are described in Secs. IV and V, and some limiting cases are explicitly given in Sec. VI. In

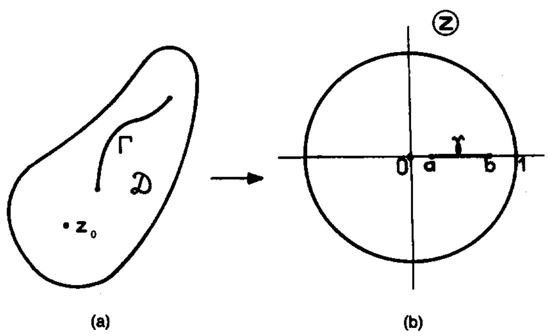


FIG. 1. Geometrical setting of the problem. (a) general; (b) after conformal mapping and restriction on Γ.

[a] Physique Mathématique et Théorique, Equipe de Recherche Associée au C.N.R.S.

 0022-2488/85/112737-09$02.50

particular, it will be seen (maybe rather surprisingly) that *not necessarily all* the zeros of the extremal functions $f_0(z)$ choose to concentrate on γ in order to minimize $|f_0(x)|$ there in an optimal way. Sec. VII contains some concluding comments.

II. REMOVAL OF IRRELEVANT FACTORS

We start with the canonical factorization theorem in H^∞ (see Ref. 2):

$$f\in H^\infty \quad (f\not\equiv 0) \Leftrightarrow f(z)=B(z)C(z)S(z), \tag{2.1}$$

where $B(z)$ is a Blaschke product

$$B(z)=e^{i\omega}z^m \prod_n \frac{|\alpha_n|}{\alpha_n}\frac{\alpha_n - z}{1-\alpha_n^* z},$$

$$0<|\alpha_n|<1, \quad \sum_n (1-|\alpha_n|)<\infty, \tag{2.2}$$

$C(z)$ is the "outer" function

$$C(z)=\exp\int_{-\pi}^{\pi}\frac{d\theta}{2\pi}\ln\chi(\theta)\frac{e^{i\theta}+z}{e^{i\theta}-z}, \tag{2.3}$$

with

$$\chi(\theta)=\lim_{r\uparrow 1}|f(re^{i\theta})| \text{ a.e.},$$

$$\|f\|_\infty = \text{ess}\sup_{-\pi<\theta<\pi}\chi(\theta),$$

and $S(z)$ is the "singular inner" function

$$S(z)=\exp\left[-\int_{-\pi}^{\pi}d\nu(\theta)\frac{e^{i\theta}+z}{e^{i\theta}-z}\right], \tag{2.4}$$

with $d\nu(\theta)$ a positive measure, singular with respect to the Lebesgue measure.

A. Removal of the "outer" and the "singular inner" functions

Our first contention is that in (1.1) the function f can be restricted to the subset of *pure Blaschke products* (2.2). Noticing that $\|f\|_\infty<1$ implies $\ln\chi(\theta)\leqslant 0$ a.e., we see that the product $D(z)\equiv C(z)S(z)$ admits the representation

$$D(z)=\exp\left[-\int_{-\pi}^{\pi}d\mu(\theta)\frac{e^{i\theta}+z}{e^{i\theta}-z}\right], \tag{2.5}$$

where $d\mu(\theta)=-(1/2\pi)\ln\chi(\theta)d\theta+d\nu(\theta)$ is a *positive* measure for the functions f relevant to the problem (1.1). Hence, for real $x^2<1$, we have $|D(x)|=\exp[-E(x)]$ with

$$E(x)=(1-x^2)\int_{-\pi}^{\pi}d\mu(\theta)\frac{1}{1-2\cos\theta x+x^2}\geqslant 0. \tag{2.6}$$

Since $0<|D(x)|\leqslant 1$, we can define, for all $x\in]-1,1[$,

$$d(x)=\frac{|D(x)|+x}{1+|D(x)|x}. \tag{2.7}$$

Lemma 1: $d(x)$ is a nondecreasing function on $]-1,1[$.

Proof: $d(x)$ is obviously differentiable on $]-1,1[$ and

$$d'(x)=\frac{2\sinh E(x)-(1-x^2)E'(x)}{|D(x)|(1+|D(x)|x)^2}$$
$$\geqslant\frac{2E(x)-(1-x^2)E'(x)}{|D(x)|(1+|D(x)|x)^2}. \tag{2.8}$$

But a simple computation using the representation (2.6) gives

$$2E(x)-(1-x^2)E'(x)$$
$$=2(1-x^2)(1+x)^2\int_{-\pi}^{\pi}d\mu(\theta)$$
$$\times\frac{1-\cos\theta}{(1-2\cos\theta x+x^2)^2}\geqslant 0. \tag{2.9}$$

Q.E.D.

Thus we have $d(x)\geqslant d(0)$ for $0\leqslant x<1$, which can be rewritten as

$$|D(x)|\geqslant[|D(0)|-x]/[1-|D(0)|x]. \tag{2.10}$$

As it stands, this inequality is useless for our purpose, because its right-hand side (rhs) is not necessarily positive on the whole interval $\gamma=[a,b]$. However, the same reasoning applied to the function $[D(z)]^{1/p}$ clearly leads to the inequalities

$$|D(z)|^{1/p}\geqslant[|D(0)|^{1/p}-x]/[1-|D(0)|^{1/p}x] \quad (0\leqslant x<1), \tag{2.11}$$

valid for all $p>0$. Choosing p as an integer large enough to secure $|D(0)|^{1/p}>b$, we are allowed to rewrite Eq. (2.11) as

$$|D(x)|\geqslant\left[\frac{|D(0)|^{1/p}-x}{1-|D(0)|^{1/p}x}\right]^p>0, \quad \text{for } 0\leqslant x\leqslant b. \tag{2.12}$$

Therefore, in the search of the supremum $\hat{f}$, it is always "advantageous" to replace the factor $D(z)=C(z)S(z)$ in each function $f(z)$ by the Blaschke factor $b(z)\equiv[(|D(0)|^{1/p}-z)/(1-|D(0)|^{1/p}z)]^p$, since $|b(0)|=|D(0)|$ and $|b(x)|\leqslant|D(x)|$ on γ. This means that

$$\hat{f}=\sup|f(0)|, \quad f\in\mathscr{B}, \quad |f(x)|\leqslant\epsilon, \quad x\in\gamma, \tag{2.13}$$

where $\mathscr{B}$ is now the class of Blaschke products (2.2) (notice that the constraint $\|f\|_\infty\leqslant 1$ is then automatic; in fact $\|f\|_\infty=1$ for $f\in\mathscr{B}$).

B. Restrictions on the Blaschke factors

We now show that in Eq. (2.13), only Blaschke products with *positive* zeros need to be considered. First of all, we can immediately restrict the class $\mathscr{B}$ by setting $\omega=m=0$ in Eq. (2.2): the phase factor $e^{i\omega}$ plays no role in the problem except for generating a trivial family of equivalent extremal functions $e^{i\omega}f_0(z)$ from one of them, and the power factor z^m can be omitted at once since evidently $\hat{f}\neq 0$. Next, we observe that it is "advantageous" to replace each factor of the form

$$B_\alpha(z) \quad (|\alpha|/\alpha)[(\alpha-z)/(1-\alpha^* z)] \tag{2.14}$$

appearing in the representation (2.2) of the function $f\in\mathscr{B}$ by

$$B_{|\alpha|}(z)=(|\alpha|-z)/(1-|\alpha|z). \tag{2.15}$$

Actually, $B_{|\alpha|}(0)=B_\alpha(0)=|\alpha|$, and elementary algebra shows that

$$|B_{|\alpha|}(x)|\leqslant|B_\alpha(x)|, \quad \forall x\geqslant 0. \tag{2.16}$$

As in Sec. II A, we conclude that

$$\hat{f}=\sup_{\substack{f\in\mathscr{B}_+\\ |f(x)|\leqslant\epsilon,\ x\in\gamma}}|f(0)|=\sup_{\substack{f\in\mathscr{B}_+\\ |f(x)|\leqslant\epsilon,\ x\in\gamma}}f(0), \tag{2.17}$$

where $\mathscr{B}_+$ is the class of Blaschke productes with positive zeros

$$f(z)=\prod_n \frac{\alpha_n - z}{1-\alpha_n z}, \quad 0<\alpha_n<1, \quad \forall n \tag{2.18}$$

[the last equality in Eq. (2.17) stems from the fact that $f(0)=\Pi_n \alpha_n > 0$].

III. SOLVING AN AUXILIARY PROBLEM: AN EXTENSION OF CHEBYSHEV POLYNOMIALS

Let $\mathscr{B}^N_+$ be the subclass in $\mathscr{B}_+$ of Blaschke products with a *finite* number N of factors

$$f(z)=\prod_{n=1}^{N} \frac{\alpha_n - z}{1-\alpha_n z}, \quad 0<\alpha_n<1 \quad (n=1,\ldots,N), \tag{3.1}$$

and let us consider the problem (2.17) restricted to $\mathscr{B}^N_+$

$$\hat{f}_N = \sup_{\substack{f\in \mathscr{B}^N_+ \\ |f(x)|<\epsilon, \ x\in\gamma}} f(0). \tag{3.2}$$

It is immediately realized that, given ϵ and N, the conditions $f\in \mathscr{B}^N_+$ and $|f(x)|<\epsilon(x\in\gamma)$ are not necessarily compatible. If, e.g., $N=1$, one easily finds that

$$\inf_{0<\alpha_1<1}\left[\sup_{a<x<b}\left|\frac{\alpha_1 - x}{1-\alpha_1 x}\right|\right] = \frac{1-ab}{b-a} - \left[\left(\frac{1-ab}{b-a}\right)^2 - 1\right]^{1/2}, \tag{3.3}$$

so that ϵ must not be less than this value. On the other hand, given ϵ, there are clearly functions f in $\mathscr{B}^N_+$ such that $|f(x)|<\epsilon$ on γ if N is large enough. This remark leads us, in a preliminary step, to look for the critical value of ϵ corresponding to a given N, i.e., for a solution of the minimax problem

$$\epsilon_N(\gamma) = \inf_{g\in\mathscr{B}^N_+} \sup_{a<x<b} |g(x)|. \tag{3.4}$$

Observing that

$$(\alpha - x)/(1-\alpha x) > (b-x)/(1-bx), \quad \forall x\in[a,b], \tag{3.5}$$

when $b<\alpha<1$, and that

$$|(\alpha - x)/(1-\alpha x)| > (a-x)/(1-ax), \quad \forall x\in[a,b], \tag{3.6}$$

when $0<\alpha<a$, we immediately conclude, through the "replacement argument" repeatedly used in Sec. II that Eq. (3.4) can be written as

$$\epsilon_N(\gamma) = \inf_{g\in\mathscr{B}^N_\gamma} \sup_{a<x<b} |g(x)|, \tag{3.7}$$

where $\mathscr{B}^N_\gamma \subset \mathscr{B}^N_+$ contains only the Blaschke products, the zeros of which belong to $[a,b]$. Moreover, since the Blaschke class is invariant under the transformation $z\to z' = (z-\beta)/(1-\beta z)$, $-1<\beta<1$, it is always possible to send $\gamma=[a,b]$ onto a symmetrical interval $[-c,c]$ by such a transformation. One finds

$$c = \frac{1-ab}{b-a} - \left[\left(\frac{1-ab}{b-a}\right)^2 - 1\right]^{1/2}. \tag{3.8}$$

This means that $\epsilon_N(\gamma)$ depends on the interval γ through the only parameter c [or, equivalently, the "Blaschke invariant" $(1-ab)/(b-a)$]. Thus, we shall rewrite Eq. (3.7) in the canonical form

$$\epsilon_N(c) = \inf_{g\in\mathscr{B}^N_c} \sup_{-c<x<c} |g(x)|, \tag{3.9}$$

with

$$\mathscr{B}^N_c = \left\{ g \,\middle|\, g(z) = \prod_{n=1}^{N} \frac{\alpha_n - z}{1-\alpha_n z}, \quad -c<\alpha_n<c \right\}. \tag{3.10}$$

In order to solve the problem (3.9), it is convenient to map conformally the unit disk $|z|<1$ cut along the interval $[-c,c]$ onto an annulus. This is achieved by the mapping

$$z\to w(z) = i\exp\left[-\frac{i\pi}{2k(c^2)} F\left(\frac{z}{c}\middle| c^2\right)\right], \tag{3.11}$$

where $F(u|k)$ is the elliptic integral of the first kind

$$F(u|k) = \int_0^u \frac{dt}{[(1-t^2)(1-k^2t^2)]^{1/2}} \tag{3.12}$$

and $K(k)\equiv F(1|k)$ is the corresponding complete elliptic integral.[8] Then the cut disk is sent onto the annulus $1<|w|<\rho(c)$, the images of the cut $[-c,c]$ and of the circle $|z|=1$ being the circles $|w|=1$ and $|w|=\rho(c)$, respectively (Fig. 2). Here

$$\rho(c) = w(1) = e^{\pi K'(c^2)/4K(c^2)}, \tag{3.13}$$

with the usual meaning of $K'(k)$:

$$K'(k) = \int_1^{1/k} \frac{dt}{[(t^2-1)(1-k^2t^2)]^{1/2}}. \tag{3.14}$$

When deriving Eq. (3.13), use has been made of the property[9]

$$\int_1^{1/\sqrt{k}} \frac{dt}{[(t^2-1)(1-k^2t^2)]^{1/2}} = \frac{1}{2}K'(k). \tag{3.15}$$

The inverse mapping of (3.11) is

$$w\to z(w) = c\,\mathrm{sn}\left(\frac{2iK(c^2)}{\pi}\ln\frac{w}{i}\,\middle|\, c^2\right), \tag{3.16}$$

where $\mathrm{sn}(\cdot|k)$ is the Jacobian elliptic sine [the inverse of the function $F(\cdot|k)$].

Let us now set $G(w)\equiv g(z(w))$ for any $g\in\mathscr{B}^N_c$. Since $|g(e^{i\varphi})|=1$, one sees that $G(w)$ is holomorphic in the annulus $1<|w|<\rho(c)$ and that $|G(\rho(c)e^{i\theta})|=1$. Furthermore, the analyticity of $g(z)$ on the segment $[-c,c]$ entails the relation $G(w)=G(1/w)$ for $|w|=1$, which immediately extends to the enlarged annulus A: $1/\rho(c)<|w|<\rho(c)$ (Fig. 2). Thus, for any $g(z)$ in $\mathscr{B}^N_c$, the corresponding function $G(w)$ is holomorphic in A and enjoys the properties

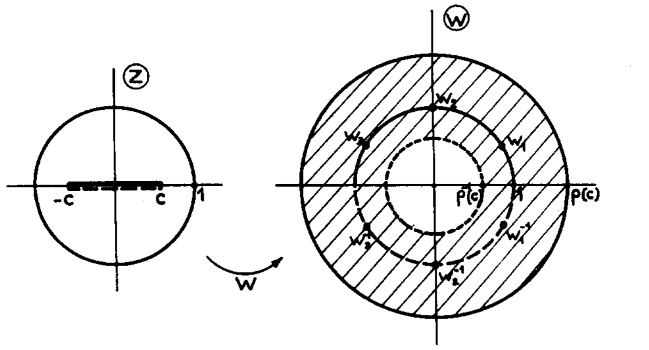


FIG. 2. Conformal mapping of the cut unit disk onto the annulus $1<|w|<\rho(c)$, and the extended annulus A (hatched area). The location of the zeros w_n, $1/w_n$ of the function $G_N(w)$ is also shown in the case $N=3$.

$$G(w) = G(1/w) \quad \text{in } A, \tag{3.17}$$

$$|G(w)| = 1 \text{ on the boundary circles } |w| = \rho(c) \text{ and } |w| = 1/\rho(c), \tag{3.18}$$

$$G(w) \text{ real on the circle } |w| = 1. \tag{3.19}$$

The N zeros α_n of $g(z)$ becomes $2N$ zeros of $G(w)$ located on the circle $|w| = 1$: $w_n = w(\alpha_n)$ and $1/w_n$.

The advantage of the transformation $z \to w$ is that the problem posed now appears to admit an almost rotational symmetry [only broken by the constraint (3.17)] that allows us to guess the solution. Actually, in view of this symmetry, it seems natural that the extremum (3.9) will be reached by distributing uniformly the $2N$ zeros of $G(w)$ on the unit circle [without violating Eq. (3.17)], i.e., by setting (see Fig. 2)

$$w_n = e^{i(\pi/2N)(2n-1)}, \quad n = 1,\ldots,N. \tag{3.20}$$

We are going to show that the function $g_N(z)$ so defined, i.e., via

$$G_N(w) = \prod_{n=1}^{N} \frac{z(w_n) - z(w)}{1 - z(w_n)z(w)}, \tag{3.21}$$

realizes the extremum (3.9) indeed.

Lemma 2: The function $G_N(w)$ defined by Eqs. (3.20) and (3.21) has the property

$$G_N(we^{i\pi/N}) = -G_N(w), \quad \forall\, w \in A. \tag{3.22}$$

Proof: The function $H(w) \equiv G_N(we^{i\pi/N})/G_N(w)$ is holomorphic on A and does not vanish there, since the numerator and the denominator have the same set of (simple) zeros. On the other hand, $|H(w)| = 1$ on the boundary of A according to Eq. (3.18). As a result, $H(w) = \text{const} = e^{i\delta}$. Because of (3.19), the only two possibilities are $H(w) = \pm 1$. But the case $H(w) = 1$ would imply a constant sign of $G_N(w)$ on $|w| = 1$, in contradiction with the fact that the zeros w_n and $1/w_n$ are simple. Hence $H(w) = -1$. Q.E.D.

We deduce from Eq. (3.22) the angular periodicity property

$$G_N(we^{2i\pi/N}) = G_N(w), \quad \forall\, w \in A \tag{3.23}$$

(which has nothing to do with the periodicities of the function sn), and also, together with Eq. (3.17),

$$G_N(w) = G_N(w_{n+1}w_n/w), \quad \forall\, n, \quad \forall\, w \in A, \tag{3.24}$$

which entails the *symmetry* of the function $G_N(e^{i\theta})$ with respect to the "middle points" $\sqrt{w_{n+1}w_n} = e^{i(\pi/N)n}$. This last property, together with Eq. (3.22) and the real analyticity of $G_N(e^{i\theta})$, implies that $G_N(e^{i\theta})$ necessarily has alternate, local maxima and minima at the points $\theta = \pi n/N$ $(n = 0,1,\ldots,2N-1)$ with the *same* absolute value λ:

$$\lambda = G_N(-1) = \prod_{n=1}^{N} \frac{z(w_n) + c}{1 + z(w_n)c}. \tag{3.25}$$

We now claim that these extrema are actually *absolute* extrema of $G_N(e^{i\theta})$ or, equivalently, that

$$\lambda = \sup_{-c \leqslant x \leqslant c} |g_N(x)|, \tag{3.26}$$

where

$$g_N(z) = \prod_{n=1}^{N} \frac{z(w_n) - z}{1 - z(w_n)z}. \tag{3.27}$$

Indeed, we have already identified N roots of each equation

$$g_N(z) = \lambda \quad (\text{resp. } -\lambda) \tag{3.28}$$

located in the interval $[-c,c]$ (notice that the roots $z = \pm c$ are simple whereas the others are double). Should we have $g_N(x) > \lambda$ (resp. $< -\lambda$) somewhere on $[-c,c]$, this would necessarily mean that Eqs. (3.28) have supplementary roots in this interval. But this is impossible since they are Nth-degree algebraic equations.

Finally, let us show that $g_N(z)$ solves the problem (3.9). To this end, consider any function $\bar{g}$ in $\mathscr{B}_c^N$ such that $\sup_{-c\leqslant x \leqslant c} |\bar{g}(x)| \leqslant \lambda$. Then, because of the properties of $g_N(x)$ just derived ($N+1$ alternate maxima and minima with the same absolute magnitude λ), we immediately infer that the equation

$$\bar{g}(x) = g_N(x) \tag{3.29}$$

has *at least* N roots on $[-c,c]$ (accounting for possible multiplicities). But, as we know from the Nevanlinna–Pick–Schur construction,[5] an Nth-order Blaschke product normalized as in Eq. (3.1) is uniquely determined by its values at N points. Hence $\bar{g}(z) \equiv g_N(z)$, which is then the *unique* extremal function associated with the problem (3.9). It follows that $\epsilon_N(c) = \lambda$, or, according to Eqs. (3.16), (3.20), and (3.25),

$$\epsilon_N(c) = c^N \prod_{n=1}^{N} \frac{\text{sn}((K(c^2)(N+1-2n)/N)|c^2) + 1}{1 + c^2\, \text{sn}((K(c^2)(N+1-2n)/N)|c^2)}. \tag{3.30}$$

In the cases $N = 1$ and 2, this expression reduces to elementary functions of c:

$$\epsilon_1(c) = c \tag{3.31}$$

[one recovers Eq. (3.3) on account of Eq. (3.8)],

$$\epsilon_2(c) = c^2/(1 + \sqrt{1-c^4}) \tag{3.32}$$

[by using[9] $\text{sn}(K(k)/2|k) = (1 + \sqrt{1-k^2})^{-1/2}$]. This is no longer true for $N \geqslant 3$, but it is possible to bring the expression (3.30), as well as the formula (3.27) for the extremal function, into more compact forms. This is achieved by introducing first the "nome"[9]

$$q(k) \equiv e^{-\pi[K'(k)/K(k)]}, \tag{3.33}$$

and by eliminating the parameter k in favor of q [through an implicit inversion of Eq. (3.33)]. We shall write accordingly

$$K[q] \equiv K(k(q)), \quad \text{sn}[v|q] \equiv \text{sn}(v|k(q)). \tag{3.34}$$

Then

$$\epsilon_N(c) = \sqrt{k(q^N)}, \tag{3.35}$$

$$g_N(z) = (-1)^N \epsilon_N(c)\, \text{cd}\left[N \frac{K[q^N]}{K[q]}\, \text{cd}^{-1}\left[\frac{z}{c}\Big|q\right] \Big| q^N\right]. \tag{3.36}$$

In these formulas, q is written for $q(c^2) = \exp[-\pi K'(c^2)/K(c^2)]$, and the function $\text{cd}[\cdot|q]$, one of the Jacobian cosines,[9] is simply defined as

$$\text{cd}[v|q] \equiv \text{sn}[v + K[q]|q]. \tag{3.37}$$

The easiest proof of Eqs. (3.35) and (3.36) does not proceed directly from Eqs. (3.30) and (3.27), but as follows. One uses the variable $v = F((z/c)|c^2) = \text{sn}^{-1}[(z/c)|q]$ in place of z (or w). Then, from the mere definitions and the property (3.23), it is not difficult to show that $q_N(z)$ is meromorphic and doubly periodic in the variable v, with periods $4K(c^2)/N$ and $2iK'(c^2)$. By inspecting its zeros and its poles through Eq. (3.27), one discovers that it is a second-order elliptic function.[10] This in turn is readily identified with the appropriate Jacobian function, apart from a normalization, which is fixed by the condition $g_N(z=-1)=1$. Finally $\epsilon_N(c)$ is given by $g_N(z=-c)$. The details are left to the reader.

Let us point out the striking similarity of the expression (3.36) of the "hidden" Blaschke product $g_N(z)$ with the familiar expression of the Chebyshev polynomials in terms of trigonometric functions

$$T_N(z)=\cos(N\cos^{-1}z). \tag{3.38}$$

Of course, the similarity is not fortuitous: While Eqs. (3.35) and (3.36) solve the problem (3.9), the Chebyshev polynomials solve the classical problem

$$\eta_N=\inf_{h\in\mathscr{P}^N}\ \sup_{-1<y<1}|h(y)|, \tag{3.39}$$

where $\mathscr{P}^N$ is the class of N th-degree polynomials of the form $h(z)=\Pi_{n+1}^N(\beta_n-z)$. The solution for η_N and the extremal polynomial h_N are well known:

$$\eta_N=1/2^{N-1},\quad h_N(z)=(-1)^N\eta_N T_N(z). \tag{3.40}$$

The relationship between the "Blaschke minimax" (3.9) and the polynomial minimax (3.39) is made precise by noticing that the unit circle relevant to the first problem is repelled to infinity if one makes the change of variable $x=cy$ and takes the limit $c\to0$. Therefore (setting $\alpha_n=c\beta_n$), one expects the following limits to be true:

$$\lim_{c\to0}[\epsilon_N(c)/c^N]=\eta_N,\quad \lim_{c\to0}[g_N(cy)/c^N]=h_N(y). \tag{3.41}$$

Indeed, these relations are easily derived from the behaviors[9]

$$q(k)\underset{k\to0}{=}k^2/16+O(k^4),$$
$$K[q]\underset{q\to0}{=}\pi/2+O(q), \tag{3.42}$$
$$\mathrm{cd}[y|q]\underset{q\to0}{=}\cos y+O(q).$$

Thus, *our extremal Blaschke products $g_N(z)$ appear as natural extensions of the Chebyshev polynomials.*

Finally, for future use, we shall record the form of Eqs. (3.35) and (3.36) in the asymptotic limit $N\to\infty$. From Eqs. (3.33) and (3.42), we immediately get

$$\epsilon_N(c)\underset{N\to\infty}{\simeq}2\exp\left(-\frac{N\pi}{4}\frac{K'(c^2)}{K(c^2)}\right) \tag{3.43}$$

and

$$g_N(z)\underset{N\to\infty}{\simeq}(-1)^N2\exp\left(\frac{-N\pi}{4}\frac{K'(c^2)}{K(c^2)}\right)\times\cos\left(\frac{N\pi}{2K(c^2)}\mathrm{cd}^{-1}\left[\frac{z}{c}\,|q\right]\right). \tag{3.44}$$

Let us make Eq. (3.44) more explicit in the case $-1<z<-c$. From the relations [see Eqs. (3.12) and (3.37)] $\mathrm{cd}^{-1}(u|k)=\mathrm{sn}^{-1}(u|k)-K(k)=F(u|k)-K(k)$,

$$F(u|k)=-K(k)+i\int_1^{-u}dt\frac{1}{[(t^2-1)(1-k^2t^2)]^{1/2}} \tag{3.45}$$
$$(-1/k<u<-1),$$

we obtain

$$\begin{aligned}\cos&\left(\frac{N\pi}{2K(c^2)}\mathrm{cd}^{-1}\left[\frac{z}{c}\,|q\right]\right)\\&=(-1)^N\cosh\frac{N\pi}{2K(c^2)}\int_1^{-z/c}\frac{dt}{[(t^2-1)(1-c^4t^2)]^{1/2}}\\&\underset{N\to\infty}{\simeq}(-1)^N\frac{1}{2}\exp\left[\frac{N\pi}{2K(c^2)}\times\int_1^{-z/c}\frac{dt}{[(t^2-1)(1-c^4t^2)]^{1/2}}\right],\end{aligned} \tag{3.46}$$

so that, on account of Eq. (3.15),

$$g_N(z)\underset{N\to\infty}{\simeq}\exp\left[-\frac{N\pi}{2K(c^2)}\int_{-z/c}^{1/c}\frac{dt}{[(t^2-1)(1-c^4t^2)]^{1/2}}\right] \tag{3.47}$$
$$(-1<z<-c).$$

IV. STRUCTURAL PROPERTIES OF THE SOLUTION IN $\mathscr{B}_+^{N_0}$

Given $\epsilon(<1)$, we define N_0 as the smallest integer such that the two conditions $f\in\mathscr{B}_+^{N_0}$ and $|f(x)|\leqslant\epsilon$ on γ are compatible. According to the discussion of Sec. III, N_0 is obtained by first computing c through Eq. (3.8) and then solving Eq. (3.35) for N, which gives

$$N_0=\left\{\frac{K(c^2)}{K'(c^2)}\frac{K'(\epsilon^2)}{K(\epsilon^2)}\right\}, \tag{4.1}$$

where $\{x\}\equiv$smallest integer $\geqslant x$. Considering now the problem (2.17) restricted to $\mathscr{B}_+^{N_0}$:

$$\hat{f}_{N_0}=\sup_{\substack{f\in\mathscr{B}_+^{N_0}\\|f(x)|\leqslant\epsilon\quad x\in\gamma}}f(0). \tag{4.2}$$

We are in a position to show very simply that the supremum is attained. Actually, Eq. (4.2) can be recast in the form

$$\begin{aligned}\hat{f}_{N_0}&=\sup_{\substack{\left|\prod_{n=1}^{N_0}\frac{\alpha_n-x}{1-\alpha_nx}\right|\leqslant\epsilon,\ a<x<b\\0<\alpha_n<1}}\left(\prod_{n=1}^{N_0}\alpha_n\right)\\&=\sup_{\substack{\left|\prod_{n=1}^{N_0}\frac{\alpha_n-x}{1-\alpha_nx}\right|\leqslant\epsilon,\ a<x<b\\0\leqslant\alpha_n\leqslant1}}\left(\prod_{n=1}^{N_0}\alpha_n\right).\end{aligned} \tag{4.3}$$

The replacement of $\alpha_n>0$ by $\alpha_n\geqslant0$ is obviously allowed. As for the replacement of $\alpha_n<1$ by $\alpha_n\leqslant1$, it is allowed too since the first constraint implies $\alpha_n<1$ anyway (if not, N_0 would not be the smallest N compatible with the value of ϵ). Then $\hat{f}_{N_0}$, which appears as the supremum of the continuous function $\Pi_{n=1}^{N_0}\alpha_n$ on a compact subset of $\mathbb{R}^{N_0}$, is necessarily attained.

We shall denote by $f_0(z)$ the associated extremal function [the notation anticipates the fact, to be demonstrated in Sec. V, that the extremal function is unique and actually solves the initial problem (2.17)], and order its zeros α_n according to the index n:

$$0<\alpha_1\leqslant\alpha_2\leqslant\cdots\leqslant\alpha_{N_0}<1. \tag{4.4}$$

By performing suitable variations of the simple and double Blaschke subproducts contained in $f_0(z)$, we shall establish the following properties of this function:

(i) the N_0 zeros α_n are all simple (and thus distinct);
(ii) $\alpha_1>a$, $\alpha_{N_0-1}<b$ [thus, at most one zero (α_{N_0}) lies outside[11] the interval γ];
(iii) $f_0(a)=\epsilon$;
(iv) between any two adjacent zeros α_n and α_{n+1}, $f_0(x)$ assumes exactly once the value $(-1)^n\epsilon$, for some $x\in\gamma$.
(4.5)

Let us show first that $\alpha_1\geqslant a$. Indeed, $\alpha_1<a$ would imply [in the notation of Eq. (2.14)]

$$B_a(0)>B_{\alpha_1}(0) \quad \text{(trivially)}, \tag{4.6}$$

$$|B_a(x)|<|B_{\alpha_1}(x)|, \quad \text{for } a\leqslant x<1 \quad \text{(a simple check)}, \tag{4.7}$$

so that the substitution of $B_a(z)$ for $B_{\alpha_1}(z)$ into $f_0(z)$ would increase $f_0(0)$ without violating the constraint, in contradiction with the fact that $f_0(z)$ is an extremal function.

Then, we can prove (iii) by first noticing that $f_0(x)$ is a positive, decreasing function on $]-1,\alpha_1[$, as are each of its Blaschke factors. Let us assume that $f_0(a)<\epsilon$, and consider the function $\tilde{f}_0(z)$ obtained by substituting $B_{\alpha_1'}(z)$ for $B_{\alpha_1}(z)$ into $f_0(z)$, with $\alpha_1<\alpha_1'<\alpha_2$ (we are supposing first that α_1 is a simple zero). Then, for α_1' close enough to α_1, $\tilde{f}_0(a)<\epsilon$ (by continuity), and

$$0\leqslant\tilde{f}_0(x)\leqslant\tilde{f}_0(a), \quad \text{for } a\leqslant x\leqslant\alpha_1', \tag{4.8}$$

while, as above [Eqs. (4.7) and (4.6)]

$$|B_{\alpha_1'}(x)|<|B_{\alpha_1}(x)|, \quad \text{for } \alpha_1'\leqslant x<1, \tag{4.9}$$

$$B_{\alpha_1'}(0)>B_{\alpha_1}(0). \tag{4.10}$$

Equations (4.8)–(4.10) show again that $f_0(z)$ cannot be extremal, which disproves our assumption $f_0(a)<\epsilon$. The same reasoning readily extends to the case where α_1 is a multiple zero. We conclude that $f_0(a)=\epsilon$ and that $\alpha_1>a$.

We have thus established (iii) and the first part of (ii). To proceed further (when $N_0\geqslant 2$), we need the following lemma.

Lemma 3: Consider the product $B_\alpha(z)B_\beta(z)$ with $a<\alpha\leqslant\beta<1$. Let α' be such that $a<\alpha'<\alpha$ and $\beta'(<1)$ fixed by the condition

$$B_{\alpha'}(a)B_{\beta'}(a)=B_\alpha(a)B_\beta(a) \quad (\text{notice that } \beta'>\beta). \tag{4.11}$$

Then

$$\begin{aligned} &B_{\alpha'}(x)B_{\beta'}(x)\leqslant B_\alpha(x)B_\beta(x), && \text{for } a\leqslant x<1,\\ &B_{\alpha'}(x)B_{\beta'}(x)>B_\alpha(x)B_\beta(x), && \text{for } -1<x<a. \end{aligned} \tag{4.12}$$

Proof: It is enough to show that

$$(x-a)\frac{\partial}{\partial\alpha'}[B_{\alpha'}(x)B_{\beta'}(x)]\geqslant 0,$$
$$\text{for } a<\alpha'<\beta'<1, \quad -1<x<1 \tag{4.13}$$

under the constraint (4.11). But

$$\frac{\partial}{\partial\alpha'}[B_{\alpha'}(x)B_{\beta'}(x)] =\frac{\partial B_{\alpha'}(x)}{\partial\alpha'}B_{\beta'}(x)+B_{\alpha'}(x)\frac{\partial B_{\beta'}(x)}{\partial\beta'}\frac{d\beta'}{d\alpha'}, \tag{4.14}$$

where $d\beta'/d\alpha'$, drawn from Eq. (4.11), is

$$\frac{d\beta'}{d\alpha'}=-\frac{(\beta'-a)(1-\beta'a)}{(\alpha'-a)(1-\alpha'a)}. \tag{4.15}$$

Then, an explicit calculation gives, after some rearrangements:

$$\frac{\partial}{\partial\alpha'}[B_{\alpha'}(x)B_{\beta'}(x)] =\frac{(1-x^2)(x-a)(1-ax)(\beta'-\alpha')(1-\alpha'\beta')}{(1-\alpha'x)^2(1-\beta'x)^2(\alpha'-a)(1-\alpha'a)}, \tag{4.16}$$

which entails Eq. (4.13). Q.E.D.

The second part of (ii) is proved again "*per absurdo*." One assumes that $\alpha_{N_0-1}\geqslant b$, and one introduces the function $\tilde{f}_0(z)$ obtained by substituting $B_{\alpha'_{N_0-1}}(z)B_{\alpha'_{N_0}}(z)$ for $B_{\alpha_{N_0-1}}(z)B_{\alpha_{N_0}}(z)$ into $f_0(z)$, with $\alpha'_{N_0-1}<\alpha_{N_0-1}$ and α'_{N_0} fixed by $B_{\alpha'_{N_0-1}}(a)B_{\alpha'_{N_0}}(a)=B_{\alpha_{N_0-1}}(a)B_{\alpha_{N_0}}(a)$. For α'_{N_0-1} close enough to α_{N_0-1} one has

$$|\tilde{f}_0(x)|<\epsilon, \quad \text{for } \alpha'_{N_0-1}\leqslant x\leqslant b \quad (\text{if } \alpha'_{N_0-1}<b) \tag{4.17}$$

by continuity and

$$|\tilde{f}_0(x)|<|f_0(x)|, \quad \text{for } a\leqslant x\leqslant\alpha'_{N_0-1} \tag{4.18}$$

from the first Eq. (4.12) and by noticing that $B_{\alpha'_{N_0-1}}(x)B_{\alpha'_{N_0}}(x)\geqslant 0$ in the interval $[a,\alpha'_{N_0-1}]$. On the other hand, the second Eq. (4.12) gives

$$\tilde{f}_0(0)>f_0(0), \tag{4.19}$$

which shows that our assumption cannot be right, and completes the proof of (ii).

Let us turn to (iv). That $f_0(x)$ assumes *at least* once the value $+\epsilon$ or $-\epsilon$ on [a,b] between two adjacent zeros α_n and α_{n+1} is deduced in a completely similar way from the previous lemma by supposing $|f(x)|<\epsilon$ on the interval $[\alpha_n,\alpha_{n+1}]\cap[a,b]$, and changing (α_n,α_{n+1}) into $(\alpha_n',\alpha_{n+1}')$. The property (i) also follows, as $\alpha_n=\alpha_n+1$ is but a particular case of the just dismissed assumption. Finally, to complete the proof of (iv), it remains to show that $f_0(x)$ assumes *at most* once the value $+\epsilon$ or $-\epsilon$ on $[a,b]$ between α_n and α_n+1. But the converse would imply that for $\eta>0$ small enough, the equation

$$f_0(x)=\epsilon-\eta \quad \text{or} \quad f_0(x)=-\epsilon+\eta \tag{4.20}$$

has at least N_0+2 (real) roots, which is impossible since Eqs. (4.20) are N_0th-degree algebraic equations.

The general behavior of $f_0(x)$, when ϵ varies between two critical values $\epsilon_{N_0-1}(c)$ and $\epsilon_{N_0}(c)$, is illustrated in Fig. 3.

V. THE SOLUTION IN $\mathscr{B}_+^{N_0}$ *IS* THE SOLUTION IN H^∞

So far, we have no insurance that the extremum (2.17) we are looking for is really attained within the "minimal" subclass $\mathscr{B}_+^{N_0}$ compatible with a given value of ϵ. *A priori*, one cannot discard the possibility that Blaschke products in $\mathscr{B}_+$ with a larger (even an infinite) number N of factors may lead to a larger value $f(0)$. It turns out, however, that the

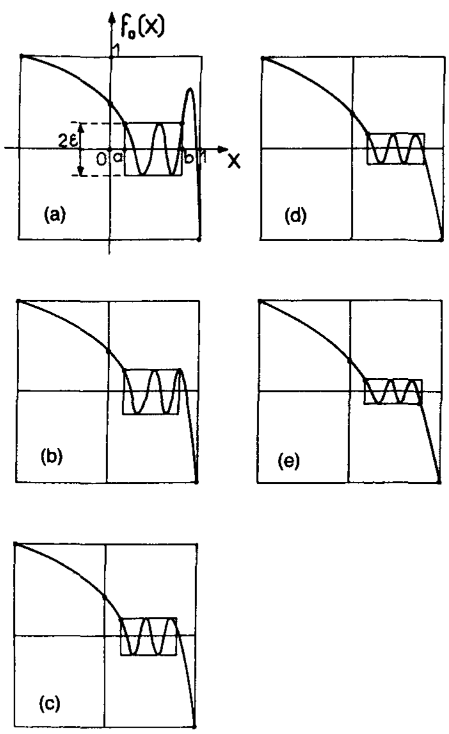


FIG. 3. Qualitative behavior of the extremal function $f_0(x)$ for decreasing values of ϵ in the interval $[\epsilon_5(c),\epsilon_4(c)[$, for $N_0 = 5$. (a) and (b) $f_0(b) = \epsilon$; (c) $0 < f_0(b) < \epsilon$; (d) $-\epsilon < f_0(b) < 0$; (e) [case $\epsilon = \epsilon_5(c)$]: $f_0(b) = -\epsilon$ and $f_0(x)$ is explicit [Eqs. (3.36) and (6.5) – (6.8)]. In the cases (a)–(c) the largest zero of $f_0(x)$ lies outside the interval $[a,b]$.

information previously gathered on the extremal function f_0 associated with the problem (4.2) is sufficient to show that f_0 actually solves the problem (2.17), and consequently the original problem (1.1). That is to say, we are going to prove that

$$\hat{f} = \hat{f}_{N_0} \equiv f_0(0). \tag{5.1}$$

To this end, let us consider an arbitrary function $h(z) \in \mathscr{B}_+$ such that

$$|h(x)| < \epsilon, \quad x \in \gamma, \tag{5.2}$$

$$h(0) > f_0(0). \tag{5.3}$$

Then, from the properties (4.5) of $f_0(x)$, it is easy to see (Fig. 4) that the equality $h(x) = f_0(x)$ holds for *at least* N_0 *positive*

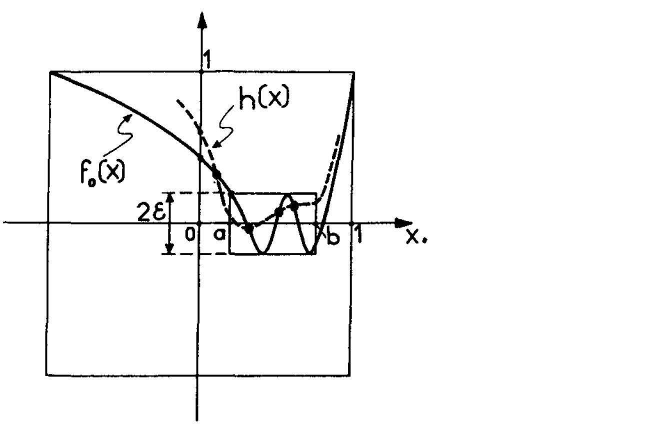


FIG. 4. Comparison of an hypothetical Blaschke product $h(x)$ subjected to the constraints (5.2) and (5.3) with the extremal function $f_0(z)$ solving the problem (4.2) (case $N_0 = 4$).

values $x_i < 1$ (accounting for possible multiplicities):

$$h(x_i) = f_0(x_i), \quad i = 1,2,\ldots,N_0. \tag{5.4}$$

Now, we can obviously write

$$h(0) \leqslant \sup_{\substack{f \in \mathscr{B}_+ \\ f(x_i) = f_0(x_i),\ \forall i}} f(0) \leqslant \sup_{\substack{f \in H^\infty \text{ and real analytic} \\ \|f\|_\infty < 1 \\ f(x_i) = f_0(x_i),\ \forall i}} f(0). \tag{5.5}$$

As it is well known, the last extremum problem in Eq. (5.5) is readily solved by appealing to the Nevanlinna–Pick–Schur construction,[1,5] with a Blaschke product as the resulting extremal function. Actually, according to this procedure, any real analytic $f \in H^\infty$ with $\|f\|_\infty \leqslant 1$ and with prescribed (and mutually compatible) values $f_0(x_i)$ at the x_i's, admits the descending recursive representation

$$g_{N_0}(z) = \text{arbitrary real analytic function in } H^\infty$$
$$\text{such that } \|g_{N_0}\|_\infty \leqslant 1, \tag{5.6}$$

$$g_{i-1}(z) = \frac{g_{i-1,i} + [(x_i - z)/(1 - x_i z)]g_i(z)}{1 + g_{i-1,i}[(x_i - z)/(1 - x_i z)]g_i(z)}, \tag{5.7}$$
$$i = N_0, N_0 - 1, \ldots, 1,$$

$$f(z) = g_0(z), \tag{5.8}$$

where the $g_{i-1,i}$ are themselves defined by the double ascending recursion[12]

$$g_{0,k} = f_0(x_k), \quad k = 1,\ldots,N_0,$$
$$g_{i,k} = \frac{g_{i-1,k} - g_{i-1,i}}{1 - g_{i-1,k}g_{i-1,i}} \frac{1 - x_i x_k}{x_i - x_k}, \tag{5.9}$$
$$i = 1,\ldots,N_0 - 1, \quad k = i+1,\ldots,N_0,$$

and satisfy

$$-1 < g_{i-1,i} < 1, \quad i = 1,\ldots,N_0. \tag{5.10}$$

Then Eq. (5.7) gives

$$g_{i-1}(0) = \frac{g_{i-1,i} + x_i g_i(0)}{1 + g_{i-1,i} x_i g_i(0)}. \tag{5.11}$$

On account of Eq. (5.10) and $x_i > 0$, we see that the rhs of Eq. (5.11) is an increasing function of $g_i(0)$. As a result, the maximum of $f(0) = g_0(0)$ will be reached by choosing $g_{N_0}(0) = +1$, i.e., $g_{N_0}(z) \equiv 1$ in Eq. (5.6). Then clearly the corresponding extremal function $g_0(z)$ deduced from Eq. (5.7) is a N_0th-order Blaschke product $f_+(z)$. Thus

$$\sup_{\substack{f \in H^\infty \text{ and real analytic} \\ \|f\|_\infty < 1 \\ f(x_i) = f_0(x_i),\ \forall i}} f(0) = f_+(0). \tag{5.12}$$

On the other hand, Eqs. (5.6)–(5.8) show that there are only *two* real analytic N_0th-order Blaschke products taking on the prescribed values $f_0(x_i)$ at the x_i's, namely $f_+(z)$ just obtained and $f_-(z)$ obtained by setting $g_{N_0}(z) \equiv -1$ in Eq. (5.6). Therefore, our previous extremal function $f_0(z)$ must be one of them. To decide which one, it is enough to check the normalization at $z = -1$. From Eqs. (5.6)–(5.8) we get $f_\pm(-1) = \pm 1$, whereas $f_0(-1) = 1$ according to Eq. (3.1). Hence $f_+(z) = f_0(z)$ and finally, from Eqs. (5.5) and (5.12),

$$h(0) \leqslant f_0(0). \tag{5.13}$$

This means that Eqs. (5.2) and (5.3) cannot be true simultaneously, and establishes the announced result (5.1).

A similar reasoning based on Eqs. (5.6)–(5.8) shows that the extremal function $f_0(z)$ is *unique* (up to the inessential phase factor mentioned in Sec. II B).

VI. EXPLICIT SOLUTION FOR LARGE AND SMALL VALUES OF ϵ

When the "error parameter" ϵ is either (i) sufficiently large $[\epsilon_2(c)\leqslant\epsilon<1]$ or (ii) asymptotically small ($\epsilon\to0$), it is possible to work out explicit formulas.

(i) If $N_0<2$, the supremum $\hat f=\hat f_{N_0}$ can be computed by elementary means. The solution is trivial for $N_0=1$, but already requires some tedious algebra in the case $N_0=2$. The results are as follows.

(a) For $\epsilon_1(c)\leqslant\epsilon<1$ $(N_0=1)$,

$$\hat f=(a+\epsilon)/(1+a\epsilon). \tag{6.1}$$

The zero α of the extremal function $f_0(z)$ lies outside the interval γ $(\alpha>b)$ only if $\epsilon>r$, where r is the Blaschke invariant

$$r=(b-a)/(1-ab). \tag{6.2}$$

(b) For $r^2/(1+\sqrt{1-r^4})\leqslant\epsilon<\epsilon_1(c)$ $(N_0=2)$,

$$\hat f=\frac{ab+(b+a)[(1-ab)/(b-a)]\epsilon+\epsilon^2}{1+(b+a)[(1-ab)/(b-a)]\epsilon+ab\epsilon^2}. \tag{6.3}$$

The larger zero α_2 of $f_0(z)$ is always $>b$.

(c) For $\epsilon_2(c)\leqslant\epsilon<r^2/(1+\sqrt{1-r^4})$ $(N_0=2)$,

$$\hat f=\frac{a^2+2a[2\epsilon(1+\epsilon^2)]^{1/2}/(1+\epsilon)+\epsilon}{1+2a[2\epsilon(1+\epsilon^2)]^{1/2}/(1+\epsilon)+a^2\epsilon}. \tag{6.4}$$

Here $\alpha_2>b$ only for ϵ larger than some (quite complicated) expression depending only on r.

Let us recall that $\epsilon_1(c)$ and $\epsilon_2(c)$ are given by Eqs. (3.31) and (3.32) with $c=r/(1+\sqrt{1-r^2})$ [Eq. (3.8)].

(ii) If $N_0\to\infty$ ($\epsilon\to0$), the asymptotic form of $\hat f$ in terms of ϵ can be found. Actually, when ϵ tends to zero, it goes through the increasingly dense sequence of critical values $\epsilon_N(c)$ [Eq. (3.35)] for which the extremal function $f_0(z)$ is completely known. Since $\hat f$ is a decreasing function of ϵ, it is clearly sufficient to stick to these critical values in order to derive the form of $\hat f$ in the asymptotic limit $\epsilon\to0$.

For $\epsilon=\epsilon_N(c)$, according to the result of Sec. III, the only function $g\in\mathscr{B}_c^N$ which respects the constraint $|g(x)|\leqslant\epsilon(-c\leqslant x\leqslant c)$ is g_N. Therefore, in this case,

$$f_0(z)=g_N(z'), \tag{6.5}$$

where $z\to z'$ is the proper Blaschke transformation which sends the interval $[a,b]$ onto $[-c,c]$, namely,

$$z'=(z-\beta)/(1-\beta z) \tag{6.6}$$

with

$$\beta=(1+ab)/(b+a)-\{[(1+ab)/(b+a)]^2-1\}^{1/2} \tag{6.7}$$

(notice that β, contrary to c, is *not* a Blaschke invariant). Hence

$$\hat f=f_0(0)=g_N(-\beta). \tag{6.8}$$

We can now rely on the asymptotic expression (3.47) to get

$$\hat f\underset{N\to\infty}{\simeq}\exp\left[-\frac{N\pi}{2K(c^2)}\int_{\beta/c}^{1/c}\frac{dt}{[(t^2-1)(1-c^4t^2)]^{1/2}}\right], \tag{6.9}$$

and eliminate N in favor of ϵ with the help of Eq. (3.43). Using Eq. (3.15), we finally obtain

$$\hat f\underset{\epsilon\to0}{\simeq}(\epsilon/2)^{1-p}, \tag{6.10}$$

where

$$p=\frac{\int_1^{\beta/c}dt/\sqrt{(t^2-1)(1-c^4t^2)}}{\int_1^{1/c}dt/\sqrt{(t^2-1)(1-c^4t^2)}}. \tag{6.11}$$

It is worth noticing that in the limit $a\to0$, Eq. (6.11) gives $p\to0$ and Eq. (6.10) would yield $\hat f\simeq\epsilon/2$, whereas the exact value of the supremum in this case is evidently $\hat f=\epsilon$. This apparent discrepancy is merely a sign of the noncommutativity of the two limits $\epsilon\to0$ and $a\to0$, which can be traced back to Eq. (3.46).

VII. CONCLUDING REMARKS

We have provided a constructive solution to the extremum problem posed in (1.1) by showing first that the extremal function $f_0(z)$ is nothing but a *finite* Blaschke product, and by displaying then its distinctive features. In particular, we wish to emphasize the fact that the number N_0 of its Blaschke factors can be computed through *explicit* formulas, namely Eqs. (3.8) and (4.1). Moreover, for an infinite sequence of "critical values" of the error parameter ϵ accumulating to 0, the solution is exactly known [Eqs. (3.36) and (6.5)–(6.8)]. It involves a remarkable class of "Chebyshev–Blaschke products" which (at least to our knowledge) have not shown up yet in the literature.

For practical purposes, it is interesting to compare our supremum $\hat f$ with nonoptimal bounds resulting simply from the maximum modulus principle. Such a bound is readily derived by using once more the variable $w=w(z')$ defined through Eqs. (6.6) and (3.11). Indeed, within the notation of Sec. III, to any function $f(z)$ in H^∞ corresponds a function $G(w)$ analytic in the annulus A. Hence the constraints $\|f\|_\infty\leqslant1$ and $|f(x)|\leqslant\epsilon$ $(a\leqslant x\leqslant b)$ become, respectively,

$$|G(\rho(c)e^{i\theta})|\leqslant1\quad(0\leqslant\theta<2\pi), \tag{7.1}$$

$$|G(e^{i\theta})|\leqslant\epsilon\quad(0\leqslant\theta<2\pi). \tag{7.2}$$

Moreover,

$$f(0)=G(w_0), \tag{7.3}$$

where

$$w_0=w(-\beta)=-\exp\left[\frac{\pi}{2K(c^2)}\int_1^{\beta/c}\frac{dt}{[(t^2-1)(1-c^4t^2)]^{1/2}}\right]. \tag{7.4}$$

Then the Hadamard's "three-circle theorem"[13] applied to the circles $|w|=1$, $|w|=|w_0|$, and $|w|=\rho(c)$ tells us that

$$|G(w_0)|\leqslant\epsilon|w_0|^{\ln(1/\epsilon)/\ln\rho(c)}, \tag{7.5}$$

or, thanks to Eqs. (7.3), (7.9), (3.13), and (3.15),

$$|f(0)|\leqslant\epsilon^{1-p}, \tag{7.6}$$

where p is given again by Eq. (6.11). Although the bound (7.6) is valid for *all* $\epsilon<1$ [contrary to the asymptotic form (6.10)], it is of course not an optimal one because the function $G_1(w)$ saturating the inequality (7.5) fails to meet the condition $G_1(w) = G_1(1/w)$ on A, and also because $G_1(w)$ is a multivalued function, except for particular values of ϵ.[14] We see that for small values of ϵ the supremum (6.10) improves the bound (7.6) by an ϵ-independent factor $(1/2)^{1-p}$, which, however, is never below 1/2 ($p \to 0$ only in the limit $a \to 0$).

Regarding now the possible extensions of our work, the following remarks can be made. If one considers first the problem (1.1) modified by simply letting the point of interest z_0 ($=0$ there) to acquire a nonzero imaginary part, then the extremal function is already not easy to describe because its zeros no longer align on the interval γ. If on the other hand γ is allowed to be an "arbitrary" arc (or even set) not touching the unit circle, the first challenging question to settle is whether the "singular inner" and possibly "outer" functions are out of play again. The answer is unknown to us, but we conjecture this to be true, at least for sufficiently regular sets γ. As for the full problem stated in (1.0), it is doubtful that any progress towards its solution can be achieved without severe restrictions on both the data function $r(z)$ and the error function $\epsilon(z)$.

[1]S. Ciulli, C. Pomponiu, and I. Sabba-Stefanescu, Phys. Rep. C **17**, 134 (1975).

[2]P. L. Düren, *Theory of H^p Spaces* (Academic, New York, 1970).

[3]See, e.g., Refs. 1 and 7.

[4]G. Auberson, L. Epele, G. Mahoux, and F. R. A. Simão, Ann. Inst. H. Poincaré A **XXII**, 317 (1975).

[5]J. L. Walsh, *Interpolation and Approximation* (Am. Math. Soc., Ann Arbor, MI, 1956), Chap. 10.

[6]M. Ciulli and S. Ciulli, Univ. of Groningen, 1977 (unpublished); see also I. Sabba Stefanescu, J. Math. Phys. **21**. 175 (1980); I. Sabba Stefanescu, preprint TKP 85-2, Karlsruhe, 1985.

[7]S. Ciulli and T. D. Spearman, Nuovo Cimento A **83**, 352 (1984); M. Ciulli, S. Ciulli, and T. D. Spearman, J. Math. Phys. **25**, 3194 (1984).

[8]The determination of the multivalued function $F(u|k)$ is chosen in the rectangle $(-2K(k), 2K(k)] \times [0, 2K'(k))$.

[9]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1970).

[10]See, e.g., E. T. Whittaker and G. N. Watson, *A Course in Modern Analysis* (Cambridge U. P., Cambridge, 1958), Chap. XX.

[11]The fact that both cases $\alpha_{N_0} < b$ and $\alpha_{N_0} > b$ can occur (according to the precise value of ϵ) is already apparent for $N_0 = 1$ (see Sec. V).

[12]Equations (5.9) are valid only if all the x_i's are distinct. Otherwise some (obvious) modifications are necessary (see Ref. 1).

[13]See, e.g., W. Rudin, *Real and Complex Analysis* (McGraw–Hill, New York, 1974), Chap. 12.

[14]See, e.g., L. Ahlfors, *Conformal Invariants* (McGraw–Hill, New York, 1973), Chap. 3, Sec. 4.

Classical SU(2) Yang–Mills–Higgs system: Time-dependent solutions by similarity method

K. Babu Joseph and B. V. Baby
High Energy Physics Group, Department of Physics, Cochin University, Cochin 682 022, India

(Received 15 January 1985; accepted for publication 26 April 1985)

A similarity analysis of the Wu–Yang–'t Hooft–Julia–Zee-ansatz-reduced system of nonlinear differential equations of classical SU(2) Yang–Mills–Higgs theory is presented. This yields the similarity group $\mathscr{G}$ of the equations. Considering $\mathscr{G}$ and one of its subgroups denoted $\mathscr{G}_1$, some previously known time-dependent solutions in the Prasad–Sommerfield limit are generated. Two new time-dependent solutions are also reported.

I. INTRODUCTION

Following 't Hooft's[1] proposal that there might exist static monopole solutions for the classical SU(2) Yang–Mills–Higgs system, Prasad and Sommerfield[2] found, by a process of guesswork, two explicit (static) analytic solutions in the limit of vanishing Higgs coupling, one of which describes a monopole and the other, a dyon. Subsequently, an ansatz-based search for new solutions was vigorously undertaken by several workers,[3–6] who came up with a class of time-dependent solutions exhibiting singularities.

The similarity method of analysis of differential equations has been widely used to study the existence of continuous symmetries,[7] and to derive particular solutions.[8] In this paper we carry out a similarity analysis of the nonlinear coupled differential equations associated with a spontaneously broken SU(2) Yang–Mills theory. This gives the invariance group $\mathscr{G}$ of the system of equations, and we use this information to generate some of the previously known time-dependent solutions in the Prasad–Sommerfield (PS) limit. We also report two new time-dependent solutions possessing surface singularities.

II. THE SIMILARITY METHOD

In this section we shall summarize the essentials of the similarity method.[9] Let us consider a system of second-order partial differential equations with two dependent variables $u^\alpha(\alpha = 1,2)$ and two independent variables $x_i (i = 1,2)$

$$\mathscr{H}(x_i, u^\alpha, u^\alpha_{x_i}, u^\alpha_{x_i x_j}) = 0. \tag{2.1}$$

Under a one-parameter family of infinitesimal transformations,

$$x_i' = x_i + \epsilon X_i(x_i, u^\alpha) + O(\epsilon^2), \tag{2.2}$$

$$u^{\alpha'} = u^\alpha + \epsilon U^\alpha(x_i, u^\alpha) + O(\epsilon^2), \tag{2.3}$$

where the X_i and U^α are the infinitesimals, the derivatives of the u^α also transform according to

$$u^{\alpha'}_{x_i'} = u^\alpha_{x_i} + \epsilon [U^\alpha_{x_i}] + O(\epsilon^2), \tag{2.4}$$

$$u^{\alpha'}_{x_i' x_j'} = u^\alpha_{x_i x_j} + \epsilon [U^\alpha_{x_i x_j}] + O(\epsilon^2). \tag{2.5}$$

The infinitesimals of the derivative, denoted by the symbols $[\cdot]$, are called the extensions,[9] the one occurring in (2.4) being called the first extension and that in (2.5) the second extension. The invariance requirement of (2.1) under the above transformations leads to the "invariant surface condition"

$$X_i \frac{\partial H}{\partial x_i} + U^\alpha \frac{\partial H}{\partial u^\alpha} + [U^\alpha_{x_i}] \frac{\partial H}{\partial u^\alpha_{x_i}} + [U^\alpha_{x_i x_j}] \frac{\partial H}{\partial u^\alpha_{x_i x_j}} = 0, \tag{2.6}$$

where repeated indices are summed over. On solving (2.6) the infinitesimals X_i and U^α can be uniquely determined which yield the similarity group $\mathscr{G}$ under which the system (2.1) is invariant.

By the infinitesimal transformation (2.3) we shall have

$$u^\alpha(x_i + \epsilon X_i + O(\epsilon^2)) = u^\alpha + \epsilon U^\alpha + O(\epsilon^2). \tag{2.7}$$

On expanding and equating the $O(\epsilon)$ terms,

$$X_i \frac{\partial u^\alpha}{\partial x_i} = U^\alpha. \tag{2.8}$$

The solutions of (2.8) are obtained by solving the characteristic equations

$$\frac{dx_i}{X_i} = \frac{du^\alpha}{U^\alpha}. \tag{2.9}$$

These give the solutions in the form

$$x_1 = x_1(x_2, c_1, c_2), \tag{2.10}$$

$$u^\alpha = u^\alpha(x_2, c_1, c_2), \tag{2.11}$$

where c_1 and c_2 are arbitrary constants of integration, c_1 playing the role of an independent variable called the similarity variable ψ and c_2 that of a dependent variable called the similarity solution F, such that $F = F(\psi)$. Thus we have

$$u^\alpha(x_1, x_2) = F^\alpha(\psi(c_1)). \tag{2.12}$$

On substituting these relations in (2.1) the latter takes the form of an ordinary differential equation

$$\mathscr{K}(\psi, F^\alpha, F^{\alpha'}, F^{\alpha''}) = 0, \tag{2.13}$$

where the prime denotes differentiation with respect to the similarity variable ψ. Equation (2.13) is called the similarity-reduced equation.

III. SIMILARITY GROUP OF SU(2) YANG–MILLS–HIGGS SYSTEM

Using the notation of Refs. 2 and 3, the Lagrangian for the SU(2) Yang–Mills–Higgs system is written

$$L = -\tfrac{1}{4} F^{\mu\nu a} F^{a}_{\mu\nu} - \tfrac{1}{2} \Pi^{\mu a} \Pi^{a}_{\mu} + \tfrac{1}{2}\mu^2(\phi^a\phi^a) - \lambda/4(\phi^a\phi^a)^2, \tag{3.1}$$

where

$$\Pi^a_\mu = \partial_\mu \phi^a + e\epsilon^{abc} A^b_\mu \phi^c \tag{3.2}$$

and

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + e\epsilon^{abc} A^b_\mu A^c_\nu. \tag{3.3}$$

For the Higgs potential

$$V(\phi) = -\tfrac{1}{2}\mu^2(\phi^a\phi^a) + (\lambda/4)(\phi^a\phi^a)^2, \tag{3.4}$$

the equations of motion are

$$\partial_\mu \Pi^{\mu a} + e\epsilon^{abc} A^b_\mu \Pi^{\mu c} - \frac{\partial V(\phi)}{\partial \phi^a} = 0, \tag{3.5}$$

$$\partial_\mu F^{\mu\nu a} - e\epsilon^{abc} F^{\mu\nu b} A^c_\mu + e\epsilon^{abc} \Pi^{\nu b} \phi^c = 0. \tag{3.6}$$

If we apply the Wu–Yang–'t Hooft–Julia–Zee (WYHJZ) ansatz[10,11]

$$\phi^a = \hat{r}^a H(r,t)/er, \quad A^a_0 = \hat{r}^a J(r,t)/er,$$
$$A^a_i = \epsilon_{aij}\hat{r}_j(1 - K(r,t))/er, \tag{3.7}$$

where

$$\hat{r}^a = r^a/r, \tag{3.8}$$

then (3.5) takes the form

$$r^2(H_{,rr} - H_{,tt}) = 2HK^2 + (\lambda/e^2)(H^3 - C^2 r^2 H), \tag{3.9}$$

where

$$C = \mu e/\sqrt{\lambda} \tag{3.10}$$

and

$$H_{,r} = \frac{\partial H(r,t)}{\partial r}, \quad \text{etc.} \tag{3.11}$$

Henceforth we shall confine ourselves to the PS limit $\lambda \to 0$, C finite.

For $\nu = 0$, (3.6) gives

$$r^2 J_{,rr} = 2JK^2, \tag{3.12}$$

and for $\nu = 1,2,3$, from (3.6) we find

$$r^2(K_{,rr} - K_{,tt}) = K(K^2 - 1) + K(H^2 - J^2), \tag{3.13}$$

$$r^2 J_{,tr} = J_{,t}, \tag{3.14}$$

$$J_{,t}K + 2K_{,t}J = 0. \tag{3.15}$$

In subsequent analysis we shall assume $J = 0$. This leads to the following pair of nonlinear coupled partial differential equations in Minkowski space:

$$r^2(K_{,rr} - K_{,tt}) - K(K^2 - 1 + H^2) = 0, \tag{3.16}$$

$$r^2(H_{,rr} - H_{,tt}) - 2HK^2 = 0. \tag{3.17}$$

To study the similarity group $\mathscr{G}$ of this system which will be referred to as that of the SU(2) Yang–Mills–Higgs system,[12] we define a generic dependent variable $u^\alpha (\alpha = 1,2)$ such that $u^1 = K$ and $u^2 = H$, and consider a one-parameter family of infinitesimal transformations as in (2.2) and (2.3), defined by

$$r' = r + \epsilon R(r,t,u^1,u^2) + O(\epsilon^2),$$
$$t' = t + \epsilon T(r,t,u^1,u^2) + O(\epsilon^2), \tag{3.18}$$
$$u^{\alpha'} = u^\alpha + \epsilon U^\alpha(r,t,u^1,u^2) + O(\epsilon^2).$$

The infinitesimals R, T, and U^α must ensure the invariance of (3.16) and (3.17) under the transformations (3.18). The derivatives $u^\alpha_{,rr}$ and $u^\alpha_{,tt}$ transform according to

$$u^{\alpha'}_{,r'r'} = u^\alpha_{,rr} + \epsilon[U^\alpha_{,rr}] + O(\epsilon^2),$$
$$u^{\alpha'}_{,t't'} = u^\alpha_{,tt} + \epsilon[U^\alpha_{,tt}] + O(\epsilon^2), \tag{3.19}$$

where $[U^\alpha_{,rr}]$ and $[U^\alpha_{,tt}]$ are the second extensions introduced in the preceding section.

When Eqs. (3.18) are substituted into the transformed system corresponding to (3.16) and (3.17) and coefficients of first order in ϵ are equated to zero, we find

$$r^2[U^1_{,rr}] - r^2[U^1_{,tt}] + 2rX(K_{,rr} - K_{,tt}) + U^1(1 - 3K^2 - H^2) - 2U^2KH = 0, \tag{3.20}$$

$$r^2[U^2_{,rr}] - r^2[U^2_{,tt}] + 2rX(H_{,rr} - H_{,tt}) - 2K^2U^2 - 4HKU^2 = 0. \tag{3.21}$$

The second extension $[U^\alpha_{,xx}]$ for $x = r, t$ is given by

$$[U^\alpha_{,xx}] = \frac{\partial^2 U^\alpha}{\partial x^2} + 2\frac{\partial^2 U^\alpha}{\partial x\, \partial u^\beta} u^\beta_{,x} - \frac{\partial^2 X^i}{\partial x^2} u^\alpha_{,i} + \frac{\partial U^\alpha}{\partial u^\beta} u^\beta_{,xx} - 2\frac{\partial X^i}{\partial x} u^\alpha_{,xi} + \frac{\partial^2 U^\alpha}{\partial u^\beta \partial u^\gamma} u^\beta_{,x} u^\gamma_{,x} - 2\frac{\partial^2 X^i}{\partial x\, \partial u^\beta} u^\beta_{,x} u^\alpha_{,i} - \frac{\partial X^i}{\partial u^\beta}(u^\alpha_{,i} u^\beta_{,xx} + 2u^\beta_{,x} u^\alpha_{,ix}) - \frac{\partial^2 X^i}{\partial u^\beta \partial u^\gamma} u^\beta_{,x} u^\gamma_{,x} u^\alpha_{,i}, \tag{3.22}$$

where

$$X^1 = R, \quad X^2 = T. \tag{3.23}$$

Solving (3.21) for the infinitesimals,

$$R = 2\lambda rt + \kappa r, \tag{3.24}$$

$$T = \lambda(r^2 + t^2) + \kappa t + \sigma, \tag{3.25}$$

$$U^\alpha = 0, \tag{3.26}$$

where λ, κ, and σ are constants. Equation (3.26) states the invariance of any solution of the SU(2) Yang–Mills–Higgs system reduced by the ansatz (3.7), under the similarity transformations (3.18).

The occurrence of three independent parameters λ, κ, and σ above permits us to define three generators $G_a (a = 1,2,3)$

$$G_1 = 2rt\frac{\partial}{\partial r} + (r^2 + t^2)\frac{\partial}{\partial t},$$
$$G_2 = r\frac{\partial}{\partial r} + t\frac{\partial}{\partial t}, \quad G_3 = \frac{\partial}{\partial t}, \tag{3.27}$$

which satisfy the Lie algebra

$$[G_1, G_2] = -G_1, \quad [G_2, G_3] = -G_3, \quad [G_3, G_1] = 2G_2. \tag{3.28}$$

We identify this as the Lie algebra associated with the similarity group $\mathscr{G}$ of Eqs. (3.16) and (3.17).

IV. TIME-DEPENDENT SOLUTIONS

In Ref. 8 we developed a method of constructing particular solutions of nonlinear Klein–Gordon equations under various subgroups of the similarity group. That procedure may be extended to the SU(2) Yang–Mills–Higgs system reduced by the WYHJZ ansatz. The idea is to consider different subgroups of the similarity group $\mathcal{G}$, define a similarity variable for each subgroup, set up the corresponding similarity-reduced equations, and solve them. Solutions are obtained in cases where the reduced equations are of the PS type. The different cases are discussed below.

A. Full group $\mathcal{G}$: $\lambda \neq 0, \kappa \neq 0, \sigma = \kappa^2/4\lambda$

Equations (3.24) and (3.25) yield the similarity variable

$$\chi = r/(t^2 - r^2 + \kappa t/\lambda + \kappa^2/4\lambda). \tag{4.1}$$

The similarity-reduced system of equations is

$$\chi^2 \frac{d^2K(\chi)}{d\chi^2} = K(\chi)(K^2(\chi) - 1) + K(\chi)H(\chi), \tag{4.2}$$

$$\chi^2 \frac{d^2H(\chi)}{d\chi^2} = 2H(\chi)K^2(\chi). \tag{4.3}$$

This is of the same form as the equation considered by Prasad and Sommerfield for the static case.[2] A solution of (4.2) and (4.3) is

$$K(\chi) = C\chi/\sinh(C\chi), \tag{4.4}$$

$$H(\chi) = C\chi \coth(C\chi) - 1, \tag{4.5}$$

where C has been defined in (3.10). This solution coincides with that reported in Ref. 3.

However, a new solution can be obtained by replacing r in the static solution reported in Ref. 13. Thus we are led to the solution

$$K(\chi) = \chi/(A + \chi), \tag{4.6}$$

$$H(\chi) = A/(A + \chi), \tag{4.7}$$

where A is a nonzero arbitrary constant. Both $K(\chi)$ and $H(\chi)$ are singular on the surface $(A + \chi) = 0$.

B. Subgroup $\mathcal{G}_1$: $\kappa = \sigma = 0$

Under the subgroup $\mathcal{G}_1 \subset \mathcal{G}$ specified by $\kappa = \sigma = 0$, the infinitesimals read

$$R = 2\lambda rt, \tag{4.8}$$

$$T = \lambda(r^2 + t^2). \tag{4.9}$$

With a similarity variable

$$\eta = r/(t^2 - r^2), \tag{4.10}$$

the reduced system assumes the form of (4.2) and (4.3) (with the replacement $\chi \to \eta$). We note that there exist two families of solutions just as in the case of the full group $\mathcal{G}$ mentioned above. They are

$$K(\eta) = C\eta/\sinh(C\eta), \tag{4.11}$$

$$H(\eta) = C\eta \coth(C\eta) - 1, \tag{4.12}$$

as found in Ref. 3, and

$$K(\eta) = \eta/(A + \eta), \tag{4.13}$$

$$H(\eta) = A/(A + \eta) \quad (A \neq 0), \tag{4.14}$$

which constitutes a new time-dependent solution. The latter is singular on the surface $(A + \eta) = 0$.

C. Subgroup $\mathcal{G}_2$: $\lambda = 0, \kappa \neq 0, \sigma \neq 0$

For the subgroup $\mathcal{G}_2 \subset \mathcal{G}$, $\lambda = 0$, and the infinitesimals are

$$R = \kappa r, \tag{4.15}$$

$$T = \kappa t + \sigma. \tag{4.16}$$

The corresponding similarity variable is

$$\zeta = r/(t + a), \tag{4.17}$$

where a is a nonzero arbitrary constant. The similarity equations read

$$(\zeta^2 - \zeta^4)K'' - 2\zeta^3 K' = K(K^2 - 1 + H^2), \tag{4.18}$$

$$(\zeta^2 - \zeta^4)H'' - 2\zeta^3 H' = 2HK^2, \tag{4.19}$$

where a prime denotes differentiation with respect to ζ. It has not been possible for us to find nontrivial exact solutions for this system.

V. DISCUSSION

The similarity method of analysis of the nonlinear coupled differential equations equivalent to the classical SU(2) Yang–Mills–Higgs system gives the similarity group $\mathcal{G}$, which is evidently dependent on the ansatz employed. There is an explicit time-dependent similarity variable for each subgroup of $\mathcal{G}$. Under the full group $\mathcal{G}$ as well as under one of its subgroups $\mathcal{G}_1$, time-dependent solutions arise as generalizations of the well-known static solutions of Refs. 2 and 13. This indicates the possibility of transforming any static solution of (3.16) and (3.17) into a nontrivial time-dependent form. The two new solutions herein obtained as well as those reported earlier in the literature can be continued to the Euclidean space.

The complex solutions considered in Ref. 14 have been shown[15] to follow by the replacement $H \to iH$ in the ansatz (3.7). Consequently, pursuing the similarity route herein explored, a time-dependent version of such complex solutions can be arrived at quite trivially.

ACKNOWLEDGMENTS

One of us (BVB) thanks Fr. Joseph Palathingal for continued support.

One of us (BVB) also thanks the University Grants Commission, New Delhi for financial support under Project No. 1174.

[1] G. 't Hooft, Nucl. Phys. B 79, 276 (1974).
[2] M. K. Prasad and C. M. Sommerfield, Phys. Rev. Lett. 35, 760 (1975).
[3] W. Mecklenburg and D. P. O'Brien, Phys. Rev. D 18, 1327 (1978).
[4] H. Boutaleb-Joutei, A. Chakrabarti, and A. Comtet, Phys. Rev. D 20, 1884 (1979).
[5] M. Sabir and C. M. Ajithkumar, J. Phys. G 9, 1469 (1983).
[6] D. Ray, Phys. Rev. D 22, 2100 (1980).
[7] M. Lakshmanan and P. Kaliappan, J. Math. Phys. 24, 795 (1983).
[8] K. Babu Joseph and B. V. Baby, J. Math. Phys. (submitted).
[9] G. M. Bluman and J. D. Cole, *Similarity Methods for Differential Equations* (Springer, Berlin, 1974).
[10] T. T. Wu and C. N. Yang, *Properties of Matter Under Unusual Conditions*, edited by H. Mark and S. Fernbach (Interscience, New York, 1968).
[11] B. Julia and A. Zee, Phys. Rev. D 11, 2227 (1975).
[12] In a strict sense this is not true; the group $\mathcal{G}$ depends on the ansatz employed to reduce the original set of Yang–Mills–Higgs equations.
[13] I. Ju, Phys. Rev. D 17, 1637 (1978).
[14] J. P. Hsu and E. Mac, J. Math. Phys. 18, 100 (1977).
[15] T. T. Wu and C. N. Yang, Phys. Rev. D 13, 3233 (1976).

On the existence of global integral forms on supermanifolds

Alice Rogers
Department of Mathematics, King's College, Strand, London WC2R 2LS, England

(Received 3 October 1984; accepted for publication 18 April 1985)

It is shown that the Berezin approach to integration on supermanifolds can be applied to cases where the supermanifold is a twisted extension of a real manifold. This is done by showing that supermanifolds admit a subatlas of coordinate charts with transition functions of a quite restricted kind.

I. INTRODUCTION

The integration of functions of commuting and anticommuting variables has proved an extremely useful technique in quantum field theory; it is, for instance, vital for path integration over Bose and Fermi fields, and it is a valuable tool in constructing and quantizing supersymmetric quantum field theories. The notion of "supermanifold" enables one to extend nontrivial conventional manifolds (with commuting coordinates, real or complex) to include anticommuting coordinates as well. Various approaches to supermanifolds have been taken, which are broadly equivalent but differ in the range of possibilities which they allow. (A good recent review is given by Batchelor.[1]) This paper investigates the possibility of integration over nontrivial supermanifolds—nontrivial in that the supermanifold may be a twisted product of the underlying real manifold with its nilpotent even and odd Grassmann extension (further details are in Sec. II). On conventional manifolds, an m form is integrated over an m-dimensional manifold; on supermanifolds, Berezin[2] has defined (m,n) superforms which are integrated over (m,n)-dimensional supermanifolds. [An (m,n)-dimensional supermanifold has m even and n odd coordinates.] However, not all (m,n) superforms, but only those which Berezin calls "integral" superforms, can be integrated in a consistent manner. Berezin gives a criterion for deciding whether a form is integral; this criterion is expressed in terms of local coordinates, and it is the purpose of this paper to investigate the global existence of integral forms, and thus to investigate the possibility of integration over nontrivial supermanifolds. Section II summarizes the necessary supermanifold terminology. In Sec. III conditions satisfied by supermanifolds which admit global integral forms are considered. In Sec. IV it is proved that, although the conditions established in Sec. III look quite restrictive, the various supermanifolds of most obvious use to physicists all allow a global integral (m,n) form to be patched together (provided that the underlying m-dimensional manifold admits an m form). The results of this section are closely related to Batchelor's theorem[3] that a graded manifold can always be realized as the sheaf of cross sections of a vector bundle. Section V discusses the implications of these results.

II. A BRIEF SUMMARY OF SUPERMANIFOLD GEOMETRY

Supermanifolds have coordinates which take values in the even and odd parts of a Grassmann algebra. In this paper attention is restricted to real Grassmann algebras with a finite number of generators; use of infinite Grassmann algebras involves analytic rather than C^∞ functions, and thus has more in common with complex supermanifold integration, to be considered in another paper. Here, B_L denotes the real Grassmann algebra with L odd generators, B_{L0} and B_{L1} denote the even and odd parts of B_L, respectively, while $B_L^{m,n}$ denotes the Cartesian product of m copies of B_{L0} and n copies of B_{L1}. The augmentation map $\epsilon: B_L \to \mathbb{R}$ is the projection onto the zero length part of an element of B_L. One also has $\epsilon_{m,n}: B_L^{m,n} \to \mathbb{R}^m$, defined by

$$\epsilon(a^1,\ldots,a^m;\ b^1,\ldots,b^n) := (\epsilon(a^1),\ldots,\epsilon(a^m)). \tag{2.1}$$

Two topologies are defined on B_L (and hence on $B_L^{m,n}$). One, the fine topology, is the usual topology of B_L as a finite-dimensional vector space, while in the second coarser topology, the De Witt topology,[4] a set $V \subset B_L$ is open in B_L if and only if $V = \epsilon^{-1}(U)$ for some open U in $\mathbb{R}$. Similarly, a set $V \subset B_L^{m,n}$ is open in the De Witt topology if and only if $V = \epsilon_{m,n}^{-1}(U)$ for some open set U in $\mathbb{R}^m$. A notion of "superdifferentiability" of B_L-valued functions is also required[4–8]; for B_{L0} it is simple to define an l_1 norm on B_L and then define differentiation rather as in ordinary analysis but using the Grassmann algebra, while for functions of B_{L1} it is simplest to require a power series expansion, with odd derivatives taking the expected form, although other approaches are possible.

Some difficulties are glossed over here, a good treatment is given by Rothstein in Ref. 9. In order to obtain a graded Leibnitz rule and hence a good tangent space one must really use the tensor product of G^∞ $(B_L^{m,0})$ with the algebra of polynomials in the odd coordinate functions. Infinitely differentiable functions of $B_L^{m,n}$ into B_L will be referred to as G^∞. A useful technique is the extension of a C^∞ B_L-valued function of $\mathbb{R}^m$ to a G^∞ function of $B_L^{m,n}$ by a Taylor expansion in nilpotent elements.

Definition 2.1: The mapping $Z: C^\infty(U) \to G^\infty(\epsilon_{m,n}^{-1}(U))$, where U is open in $\mathbb{R}^m$, is defined by

$$Z(f)((a^1,\ldots,a^m;\ b^1,\ldots,b^n)) := \sum_{i_1=0\cdots i_m=0}^{L} \frac{1}{i_1!\cdots i_m!}\,\partial_1^{i_1}\cdots\partial_m^{i_m} f(\epsilon(a^1),\ldots,(a^m)) \times (a^1-(a^1))^{i_1}\cdots(a^m-\epsilon(a^m))^{i_m}. \tag{2.2}$$

[The superdifferentiability of $Z(f)$ is proved in Ref. 6.]

The definition of supermanifold to be used in this paper is now given. A more general definition is possible (by using the finer topology on B_L)[6] but new supermanifolds intro-

 0022-2488/85/112749-05$02.50

duced by this definition require a more complicated integration theory.[10,11]

Definition 2.2: Let Y be a paracompact Hausdorff topological space. (a) An (m,n) open chart on Y over B_L is a pair (V,ψ) with V an open subset of Y and ψ a homeomorphism of V onto an open subset of $B_L^{m,n}$ (using the fine topology). And additionally V is in fact open in $B_L^{m,n}$ with the coarser De Witt topolgoy.

(b) An $(m,n)G^\infty$ structure on Y over B_L is a collection $\{(V_\alpha,\psi_\alpha)|\alpha\in A\}$ of open charts such that (i) $Y=\bigcup_{\alpha\in A}V_\alpha$, (ii) for each pair α,β in Λ, the mapping $\psi_\beta\circ\psi_\alpha^{-1}$ is a G^∞ mapping of $\psi_\alpha(V_\alpha\cap V_\beta)$ onto $\psi_\beta(V_\alpha\cap V_\beta)$, and (iii) the collection $\{(V_\alpha,\psi_\alpha)|\alpha\in A\}$ is a maximal collection of open charts satisfying (i) and (ii).

(c) An (m,n)-dimensional SM-G^∞ supermanifold over B_L is a paracompact Hausdorff topological space Y together with an (m,n) G^∞ structure over B_L. (This is a more restricted definition than that of Ref. 6.)

(d) A subcollection of charts $\{(V_\alpha,\psi_\alpha)|\alpha\in\Gamma\subset A\}$ satisfying (b) (i) and (ii) is called a subatlas of the supermanifold.

Equipped with this definition of supermanifold, the tangent space, frame bundle, and tensor fields can be defined much as on conventional manifolds.[4–8] One particular type of tensor field turns out to be useful for integration, the so-called (p,q) superforms (with $p\leqslant m,q\leqslant n$). To anyone familiar with the super-Jacobian transformation rule of Berezin integration,[12] the definition is quite natural. [For the rest of this paper Y will denote an (m,n)-dimensional SM-G^∞ supermanifold over B_L with G^∞ structure $\{(V_\alpha,\psi_\alpha)|\alpha\in A\}$.]

Definition 2.3 [2]*:* Let V be open in Y, and $G^\infty(V)$, $D^1(V)$, and $D_1(V)$ denote the spaces of G^∞ functions, vector fields, and one-forms on V, respectively. Then, given positive integers p and q, a (p,q) superform on V is a map $\phi:(D^1(V))^p\times(D_1(V)^q)\to G^\infty(V)$

$$(X_1,\dots,X_p;\ \omega_1,\dots,\omega_q)\rightsquigarrow\langle\phi\,|X_1,\dots,X_p;\ \omega_1,\dots,\omega_q\rangle,$$

which is a graded linear map of $G^\infty(U)$ modules together with the mixed graded symmetric and antisymmetric properties

$$\begin{aligned}&\langle\phi\,|X_1,\dots,X_i,X_{i+1}\dots,X_p;\ \omega_1,\dots,\omega_q\rangle\\&\quad=(-1)^{1+|X_i||X_{i+1}|}\langle\phi\,|X_1,\dots,X_{i+1},X_i,\dots,X_p;\\&\qquad\qquad\omega_1,\dots,\omega_q\rangle,\end{aligned}\tag{2.3}$$

$$\begin{aligned}&\langle\phi\,|X_1,\dots,X_p;\ \omega_1,\dots,\omega_j,\omega_{j+1},\dots,\omega_q\rangle\\&\quad=(-1)^{|\omega_j||\omega_{j+1}|}\langle\phi\,|X_1,\dots,X_p;\\&\qquad\qquad\omega_1,\dots,\omega_{j+1},\omega_j,\dots,\omega_q\rangle.\end{aligned}\tag{2.4}$$

($|X_i|$ is the Grassmann degree of X_i, etc.)

The exterior product of two such forms can be defined in the expected manner, and on a local coordinate neighborhood V_α, ϕ can be expanded in term of the coordinate differentials dx^i,dy^j and derivatives $\partial/\partial x^i,\partial/\partial y^j$. [Here x^i $(i=1,\dots,m)$ and y^j $(j=1,\dots,n)$ are the even and odd coordinates on V_α.] One has

$$\begin{aligned}&dx^i\wedge dx^k=-dx^k\wedge dx^i,\\&dy^j\wedge dy^l=dy^l\wedge dy^j,\\&\frac{\partial}{\partial x^i}\wedge\frac{\partial}{\partial x^k}=\frac{\partial}{\partial x^k}\wedge\frac{\partial}{\partial x^i},\\&\frac{\partial}{\partial y^j}\wedge\frac{\partial}{\partial y^l}=-\frac{\partial}{\partial y^l}\wedge\frac{\partial}{\partial y^j}.\end{aligned}\tag{2.5}$$

This means there does exist a quasi-"top form" in that an (m,n) superform ϕ on an (m,n)-dimensional supermanifold will split, in a given set of local coordinates, into

$$\phi=\phi_0+\phi_1,\tag{2.6}$$

with

$$\phi_0=f(x,y)dx^1\wedge\cdots\wedge dx^m\wedge\frac{\partial}{\partial y^1}\wedge\cdots\wedge\frac{\partial}{\partial y^n},\tag{2.7}$$

while each term in ϕ_1 contains at least one of the $\partial/\partial x^i$ or at least one of the dy^j, or both. Berezin defines an integral superform to be a superform which has exactly $\phi=\phi_0$ in some local coordinate system. Of course, in other coordinate systems, ϕ_1 terms will in general appear, but a consistent integration theory may be built up ignoring these terms; by combining the usual technique for integration over nontrivial manifolds (using patching with partitions of unity) with Berezin integration for the odd variables, a consistent integral of a global integral form may be constructed.[2,13] In fact, it would be quite possible to define an (m,n) superform at a point p on Y as an equivalence class of triples $(V_\alpha,\psi_\alpha,f_\alpha)$, where (V_α,ψ_α) is a coordinate chart on Y, f_α is an element of $G^\infty(\psi_\alpha(V_\alpha))$, and the equivalence relation is $(V_\alpha,\psi_\alpha,f_\alpha)\sim(V_\beta,\psi_\beta,f_\beta)$ if and only if f_α is equal to the product of f_β with the super-Jacobian of $\psi_\alpha\circ\psi_\beta^{-1}$. (A similar formulation of the tangent space of a conventional manifold is given by Lang.[14]) This approach is effectively taken in supergravity, where the supervolume form "ber $E_M{}^A\,dx\,d\theta$" is frequently used.[15] It gives valid results, but there is no equivalent (p,q) form with $p\neq m$ or $q\neq n$ in this approach. Now it is highly desirable to have such forms in order to be able to include integration over submanifolds into the formalism (and such results as Stokes' theorem), and so the use of the full (p,q) form, complete with possible undesirable $\partial/\partial x^i$ and dy^j parts seems an essential starting point.

To end this brief survey of supermanifold geometry, the construction of the underlying real manifold (by patching and the augmentation map) of a supermanifold is described.

It has been shown by Batchelor[5] that an m-dimensional C^∞ manifold is constructed from the supermanifold Y by the following process.

(i) Define on Y the relation $\sim$ by $p\sim q$ if and only if there exists $\alpha\in\Lambda$ such that $p,q\in V_\alpha$, and $\epsilon\circ\psi_\alpha(p)=\epsilon\circ\psi_\alpha(q)$. It may be shown that $\sim$ is an equivalence relation.

(ii) Let

$$U_\alpha=\{\bar p|p\in V_\alpha\}.\tag{2.8}$$

(Here $\bar p$ denotes the equivalence class of p.)

(iii) Define $\phi_\alpha:U_\alpha\to\mathbb{R}^m$ by

$$\phi_\alpha(\bar p)=\epsilon\circ\psi_\alpha(p).\tag{2.9}$$

It is clear that ϕ_α is well defined, and relatively straightforward to prove that $\{(U_\alpha,\phi_\alpha)|\alpha\in\Lambda\}$ is an m-dimensional C^∞ stucture on $X:=Y/\sim$. X is called the underlying manifold of Y. The canonical projection of Y onto X is denoted ϵ_Y.

A useful result is that, if $(t^1,\ldots t^m)$ and $(t^{1'},\ldots,t^{m'})$ denote coordinates in U_α, U_β, respectively, and

$$x^{i'}(x^1,\ldots,x^m;\ y^1,\ldots,y^n)$$
$$=x^{i'}_\Omega\,(x^1,\ldots,x^m)+\text{terms in } y, \tag{2.10}$$

then

$$Z\left(\frac{\partial t'^i}{\partial t^j}\right)=\frac{\partial x'^i_\Omega}{\partial x^j}. \tag{2.11}$$

Also, if $x^{i'}$ is independent of the y^j,

$$P_i\circ\psi_\beta\circ\psi_\alpha^{-1}=Z(p_i\circ\phi_\alpha\circ\phi_\beta^{-1}). \tag{2.12}$$

Since Y is paracompact, the body X is also paracompact. This means that X has a subatlas with a countable number of charts; this subatlas can be used to construct a subatlas of Y with a countable number of charts.

III. CRITERIA FOR THE EXISTENCE OF A GLOBAL INTEGRAL FORM ON A SUPERMANIFOLD

In this section, the possibility of having global integral forms on a supermanifold is related to the existence of subatlases of the supermanifold whose transition functions $\psi_\beta\circ\psi_\alpha^{-1}$ take a quite severely restricted form. (In the next section it is shown that in fact all the supermanifolds considered in this paper have subatlases of the required form, and thus that this requirement is much less restrictive than it appears to be.) The definition of a global integral form is straightforward.

Definition 3.1: An (m,n) superform on Y is said to be a global integral form if at each point p on Y there is a coordinate chart containing p on which ϕ takes the form

$$\phi=f(x,y)dx^1\wedge\cdots\wedge dx^m\wedge\frac{\partial}{\partial y^1}\wedge\cdots\wedge\frac{\partial}{\partial y^n}. \tag{3.1}$$

The first observation is that under a change of local coordinates $\psi_\beta\circ\psi_\alpha^{-1}$ with

$$(x'^1,\ldots,x'^m;\ y^{1'},\ldots,y'^n)=\psi_\beta\circ\psi_\alpha^{-1}(x^1,\ldots,x^m;\ y^1,\ldots,y^n), \tag{3.2}$$

$$dx'^i=\sum_{k=1}^m\frac{\partial x'^i}{\partial k^k}dx^k+\sum_{j=1}^n\frac{\partial x'^i}{\partial y^j}dy^j\quad(i=1,\ldots,m), \tag{3.3}$$

and

$$\frac{\partial}{\partial y'^j}=\sum_{l=1}^n\frac{\partial y^l}{\partial y'^j}\frac{\partial}{\partial y^l}+\sum_{i=1}^m\frac{\partial x^i}{\partial y'^j}\frac{\partial}{\partial x^i}\quad(j=1,\ldots,n). \tag{3.4}$$

Thus one has immediately the following proposition.

Proposition 3.2: A necessary condition for the existence of a nontrivial global integral superform on Y is that Y has a subatlas $\{(V_\alpha,\psi_\alpha)|\alpha\in\Gamma\subset\Lambda\}$ such that for each α, β in Γ, the transition function $\psi_\beta\circ\psi_\alpha^{-1}$ satisfies [with the notation of (3.2)]

$$\frac{\partial x'^i}{\partial y^j}=0,\quad i=1,\ldots,m,\quad j=1,\ldots,n. \tag{3.5}$$

The use of atlases with this property for integration on supermanifolds was first considered by Picken and Sundermeyer.[16]

The next proposition shows that sufficient conditions for the supermanifold Y to admit a global integral form are that the underlying m-dimensional real manifold should admit a global m form and that Y should have a subatlas satisfying (3.5) but additionally having the y'^j depend only linearly on the y^k. (In Sec. IV it is shown that this is always the case, a fact which corresponds in the geometric approach to supermanifolds to Batchelor's theorem in the algebraic approach that all graded manifolds can be realized as the sheaf of cross sections of the exterior bundle of a vector bundle.[3])

Proposition 3.3: Suppose that the supermanifold Y has a subatlas $\{(V_\alpha,\psi_\alpha)|\alpha\in\Gamma\subset\Lambda\}$ with each transition function satisfying (3.5) and also

$$y^{j'}=\sum_{l=1}^n f_l^j\,(x^1,\ldots,x^m)y^l\quad(j=1,\ldots,n), \tag{3.6}$$

with each $f_l^j\ G^\infty$. (Such a subatlas will be referred to as a restricted subatlas.) Also suppose that the underlying m-dimensional real manifold of Y admits a nontrivial global m form ω, then Y admits a global integral form.

Proof (with the notation of Sec. II for the underlying manifold X of Y): Suppose that in the chart (U_α,ϕ_α) on X with $\alpha\in\Gamma$, the global m form ω satisfies

$$\omega=f_\alpha\,dt^1\wedge\cdots\wedge dt^m, \tag{3.7}$$

where

$$f_\alpha\in C^\infty(U_\alpha).$$

Then

$$\phi=Z(f_\alpha\circ\phi_\alpha^{-1})\circ\psi_\alpha y^1\cdots y^n$$
$$\times dx^1\wedge\cdots\wedge dx^m\wedge\frac{\partial}{\partial y^1}\wedge\cdots\wedge\frac{\partial}{\partial y^n} \tag{3.8}$$

defines a global integral form on Y. To prove this, it must be shown that the formula (3.8) for ϕ is consistent on the overlap $V_\alpha\cap V_\beta$ of coordinate neighorhoods for all α, $\beta\in\Gamma$. [The formula (3.8) is of course only correct for charts in the restricted subatlas, but this is quite sufficient to define ϕ globally.]

Let $p\in V_\alpha\cap V_\beta$. Then

$$y^{1'}y^{2'}\cdots y^{n'}=\det(f_j^i)\,y^1y^2\cdots y^n, \tag{3.9}$$

where the entries in the matrix (f_j^i) are defined in Eq. (3.6). Also,

$$\frac{\partial}{\partial y'^1}\wedge\cdots\wedge\frac{\partial}{\partial y'^n}=(\det(f_j^i))^{-1}\frac{\partial}{\partial y^1}\wedge\cdots\wedge\frac{\partial}{\partial y^n}.$$

Thus

$$Z(f_\beta\circ\phi_\beta^{-1})\circ\psi_\beta y'^1\cdots y'^n$$
$$\times dx'^1\wedge dx'^2\wedge\cdots\wedge dx'^m\wedge\frac{\partial}{\partial y'^1}\wedge\frac{\partial}{\partial y'^2}\wedge\cdots\wedge\frac{\partial}{\partial y'^n} \tag{3.10}$$
$$=Z(f_\beta\circ\phi_\beta^{-1})\circ\psi_\beta y^1\cdots y^n$$
$$\times dx'^1\wedge\cdots\wedge dx'^m\wedge\frac{\partial}{\partial y^1}\wedge\cdots\wedge\frac{\partial}{\partial y^n}.$$

Now, since ω is an m form on X, we have

$$f_\beta = \det(g_i^k) f_\alpha,$$

where $g_i^k = \partial t^k / \partial t^i$. Now (using the approach at the end of Sec. II) one finds

$$Z(g_i^k \circ \phi_\alpha^{-1}) \circ \psi_\alpha = \frac{\partial x^k}{\partial x'^i}, \tag{3.11}$$

and thus

$$\begin{aligned}
Z(f_\beta \circ \phi_\beta^{-1}) \circ \psi_\beta
&= Z(((\det g_i^k) \circ \phi_\beta^{-1})(f_\alpha \circ \phi_\beta^{-1})) \circ \psi_\beta \\
&= (Z((\det g_i^k) \circ \phi_\beta^{-1}) \circ \psi_\beta)(Z(f_\alpha \circ \phi_\beta^{-1}) \circ \psi_\beta) \\
&= (Z((\det g_i^k) \circ (\phi_\alpha^{-1} \circ \phi_\alpha \circ \phi_\beta^{-1}) \circ \psi_\beta)) \\
&\quad \times (Z(f_\alpha \circ \phi_\alpha^{-1}) \circ Z(\phi_\alpha \circ \phi_\beta^{-1}) \circ \psi_\beta) \\
&= (Z((\det g_i^k \circ \phi_\alpha^{-1})) \circ \psi_\alpha)(Z(f_\alpha \circ \phi_\alpha^{-1}) \circ \psi_\alpha)
\end{aligned} \tag{3.12}$$

[using (2.12)]. Hence

$$\begin{aligned}
&Z(f_\beta \circ \phi_\beta^{-1}) \circ \psi_\beta y'^1 \cdots y'^n \, dx'^1 \wedge \cdots \wedge dx'^m \\
&\quad \wedge \frac{\partial}{\partial y'^1} \wedge \cdots \wedge \frac{\partial}{\partial y'^n} \\
&= \det\left(\frac{\partial x^k}{\partial x'^i}\right) Z(f_\alpha \circ \phi_\alpha^{-1}) \circ \psi_\alpha y^1 \cdots y^n \, dx'^1 \\
&\quad \wedge \cdots \wedge dx'^m \wedge \frac{\partial}{\partial y^1} \wedge \cdots \wedge \frac{\partial}{\partial y^n} \\
&= Z(f_\alpha \circ \phi_\alpha^{-1}) \circ \psi_\alpha y^1 \cdots y^n dx^1 \wedge \cdots \wedge dx^m \\
&\quad \wedge \frac{\partial}{\partial v^1} \wedge \cdots \wedge \frac{\partial}{\partial y^n},
\end{aligned} \tag{3.13}$$

as required.

□

IV. A PROOF THAT A SUPERMANIFOLD ALWAYS HAS A RESTRICTED SUBATLAS

In this section it is proved that any supermanifold (according to Definition 2.2) has a restricted subatlas so that by the results of the previous section, global integration of *something* is possible provided that the underlying manifold admits a top form. It also allows one to fit the supervolume form used in superspace supergravity[15] into the Berezin superform framework. The result proved in this section is the equivalent in the geometric approach to supermanifolds, of Batchelor's result[3] in the algebraic approach, that all graded manifolds[17,18] can be realized as the sheaf of cross sections of the exterior bundle of a vector bundle of the underlying real manifold. Just as the isomorphism between the graded manifold (which is a sheaf of graded commutative algebras) with the sheaf of vector bundle cross sections is not canonical, the restricted subatlas whose existence is proved in this section is far from canonical. The geometric version of Batchelor's theorem which this section contains is of direct use to integration theory, and also gives a clear geometric insight into the redundancy in the full definition of supermanifold.

Proposition 4.1: Any supermanifold Y has a restricted subatlas. That is, recapping Eqs. (3.5) and (3.6), and using the notation of that section, the transition functions satisfy

$$\frac{\partial x'^i}{\partial y^j} = 0, \quad i = 1,\dots,m, \quad j = 1,\dots,n \tag{4.1}$$

and

$$y'^j = \sum_{i=1}^{n} f_k^j (x^1,\dots,x^m) y^i, \quad j = 1,\dots,n. \tag{4.2}$$

Outline of proof: Starting with an arbitrary subatlas, with a countable number of charts, a modified subatlas is constructed in two stages, first so that (4.1) is satisfied, and then so that (4.2) is also satisfied. Suppose that $\{(V_s,\sigma_s) | s \in Z^+\}$ is an arbitrary subatlas with a countable number of charts. (Here Z^+ denotes the set of positive integers.) Then, starting from σ_1, (which is unchanged) one may alter σ_2, σ_3,...in turn so that one has a subatlas $\{(W_s,\rho_s | s \in Z^+\}$ with

$$p^i \circ \rho_s \circ \rho_t^{-1} := Z(\epsilon_{m,n} \circ p^i \circ (\sigma_s \circ \sigma_t^{-1})), \quad i = 1,\dots,m \tag{4.3}$$

and

$$p^{j+m} \circ \rho_s \circ \rho_t^{-1} = p^{j+m} \circ \sigma_s \circ \sigma_t^{-1}, \quad j = 1,\dots,n, \tag{4.4}$$

for each s, t in Z^+.

Here the cover $\{W_s | s \in Z^+\}$ is a refinement of $\{V_s | s \in Z^+\}$ with $\overline{W}_s \subset V_s$ for each s in Z^+. This subatlas has transition functions satisfying (4.1). Next one may modify ρ_2, ρ_3,...so that another subatlas $\{(T_s,\psi_s) | s \in Z^+\}$ is obtained with $\overline{T}_s \subset W_s$ and

$$p^i \circ \psi_s \circ \psi_t^{-1} = p^i \circ \rho_s \circ \rho_t^{-1}, \tag{4.5}$$

$$\begin{aligned}
&p^{j+m} \circ \psi_s \circ \psi_t^{-1} (u^1,\dots,u^m;\ v^1,\dots,v^n) \\
&\quad = \sum_{l=1}^{n} (v^l \partial_{l+m} (p^{j+m} \circ \rho_s \circ \rho_t^{-1})) \\
&\quad\quad \times (u^1,\dots,u^m;\ 0,\dots,0).
\end{aligned} \tag{4.6}$$

This final subatlas satisfies both (4.1) and (4.2) and is the required restricted subatlas.

Obviously the subatlas is not canonical; the choice of first chart (V_1,σ_1) is quite arbitrary, as well as the choice of subatlas $\{(V_s,\sigma_s) | s \in Z^+\}$.

V. CONCLUSIONS

The main result of this paper is simple—the coordinate patching of supermanifolds can be sufficiently unraveled to allow the construction of global integral superforms (provided that the underlying manifold admits a top form). More general supermanifolds, with nontrivial topology in the odd sector, are possible if one uses the fine topology on $B_L^{m,n}$ instead of the De Witt topology.[6,19] Here the theory of integration becomes more difficult, and it is clear that the contour approach (already essential for a fully consistent theory of integration on the even part of a Grassmann algebra,[4,9] must be extended to the odd variables. This problem is currently under investigation.[10] Some ideas in this line have been given by Rabin,[11,20] whose initial motivation was the need to find a discrete process whose limit was the Berezin integral (analogous to the Riemann sum approach to conventional integration), in the context of lattice supersymmetry.

Returning to the type of supermanifold considered in this paper, the restricted subatlas discovered here should be useful for developing a good theory of integration on subsupermanifolds, which would in turn have application to supersymmetric quantum field theories where integrals over the full superspace often have the wrong dimension for constructing actions. A good theory of integration on subsupermanifolds would also be useful for handling topological results (such as index theory) for supersymmetric theories. Even a full Stokes' theorem has not been established for supermanifolds.

The proof of the existence of a restricted subatlas can quite easily be adapted to cover the case of Grassmann-type algebras with an infinite number of generators,[6] but, as mentioned before, having an infinite number of generators relates to analytic manifolds, and is more naturally considered in the context of complex supermanifolds.

ACKNOWLEDGMENT

This research was supported by the Science and Engineering Research Council of Great Britain.

[1]M. Batchelor, lectures given at the NATO Workshop on the mathematical aspects of supergravity held in Hamburg, July 1983 (Reidel, Dordrecht, Netherlands, 1984).
[2]F. A. Berezin, Sov. J. Nucl. Phys. **30**, 605 (1979).
[3]M. Batchelor, Trans. Am. Math. Soc. **253**, 329 (1979).
[4]B. S. De Witt, *Supermanifolds* (Cambridge U. P., London, 1984).
[5]M. Batchelor, Trans. Am. Math. Soc. **258**, 257 (1980).
[6]A. Rogers, J. Math. Phys. **21**, 1352 (1980).
[7]A. Jadzyck and K. Pilch, Commun. Math. Phys. **78**, 373 (1981).
[8]C. Boyer and S. Gitler, Trans. Am. Math. Soc. **285**, 241 (1984).
[9]M. Rothstein, "On the resolution of some difficulties in the theory of G^∞ supermanifolds," Washington University preprint.
[10]A. Rogers, "Realising the Berezin integral as a superspace contour integral," King's College preprint, March 1985.
[11]J. M. Rabin, "Integration on general fermionic supermanifolds," Enrico University Preprint EFI 85-33, May 1985.
[12]F. A. Berezin, *The Mathematics of Second Quantisation* (Academic, New York, 1966).
[13]A. Rogers, J. Math. Phys. **26**, 385 (1985).
[14]S. Lang, *Differential Manifolds* (Addison–Wesley, Reading, MA, 1972).
[15]R. Arnowitt, P. Nath, and B. Zumino, Phys. Lett. B **56**, 1218 (1975).
[16]R. Picken and K. Sundermeyer (private communication).
[17]F. A. Berezin and D. A. Leites, Sov. Math. Dokl. **16**, 1218 (1975).
[18]B. Kostant, "Graded manifolds, graded Lie theory, and prequantisation," in *Differential Geometric Methods in Mathematical Physics, Lecture Notes in Mathematics*, Vol. 570 (Springer, Berlin, 1977).
[19]A. Rogers, J. Math. Phys. **22**, 443 (1981).
[20]J. M. Rabin, "The Berezin integral as a contour integral," lecture given at the conference "Supersymmetry in physics," Los Alamos National Laboratory, December 1983.

Mathematical aspects of quantum fluids. I. Generalized two-cycles of ^{4}He type

B. A. Kupershmidt
The University of Tennessee Space Institute, Tullahoma, Tennessee 37388 and Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

(Received 3 April 1985; accepted for publication 12 June 1985)

It is shown that the two-cocycle involved in the Hamiltonian description of the superfluid ^{4}He, both nonrotating and rotating, is a particular case of generalized symplectic two-cocycles on semidirect product Lie algebras.

I. INTRODUCTION

Macroscopic description of quantum fluids provides a deeply nontrivial generalization of classical fluid dynamics: namely, various two-fluid systems, including superfluid ^{4}He and ^{3}He. It is superfluid ^{4}He which has in its Hamiltonian description a new feature not encountered in elementary hydrodynamics: a (generalized) two-cocycle. This fact was found for the nonrotating ^{4}He in Refs. 1 (in physical language) and 2, and for the rotating ^{4}He in Ref. 2. The purpose of this paper is to explain the origin of this particular two-cocycle by an examination of two-cocycles on general semidirect product Lie algebras. In subsequent papers I shall analyze the interaction of generalized symplectic two-cocycles (introduced in Sec. II of this paper) with commutative Clebsch representations (for ^{4}He), and also the nature of noncommutative Clebsch representations (for anisotropic ^{3}He–A with spin); the change of order parameters (for ^{3}He–A) has been analyzed in detail in Ref. 3.

To set the scene for our discussion, here is the subject requiring an interpretation: the Poisson bracket for the nonrotating ^{4}He (formula (5) in Ref. 2; the rotating ^{4}He has a similar structure [formula (14) in Ref. 2] as far as two-cocycles are concerned):

$$\{H,F\}\sim\left\{\frac{\delta F}{\delta M_k}\left[(M_l\,\partial_k+\partial_l\,M_k)\left(\frac{\delta H}{\delta M_l}\right)+\rho\,\partial_k\left(\frac{\delta H}{\delta\rho}\right)-\alpha_{,k}\,\frac{\delta H}{\delta\alpha}+\sigma\,\partial_k\left(\frac{\delta H}{\delta\sigma}\right)\right]+\left[\frac{\delta F}{\delta\rho}\,\partial_l\rho+\frac{\delta F}{\delta\alpha}\,\alpha_{,l}+\frac{\delta F}{\delta\sigma}\,\partial_l\sigma\right]\left(\frac{\delta H}{\delta M_l}\right)\right\} \tag{1.1a}$$

$$+\left(\frac{\delta F}{\delta\alpha}\frac{\delta H}{\delta\rho}-\frac{\delta F}{\delta\rho}\frac{\delta H}{\delta\alpha}\right). \tag{1.1b}$$

The notation here is $\partial_l=\partial/\partial x_l$, where $(x_1,\dots,x_n)$ are coordinates in $\mathbb{R}^n$ ($n=3$ if you are fond of three dimensions); $(\cdot)_{,l}=\partial(\cdot)/\partial x_l$; $1\leqslant k,\ l\leqslant n$, and sum is taken over repeated indices; $\mathbf{M}=(M_1,\dots,M_n)$ is the total momentum density (of the normal flow); ρ is the mass density; σ is the entropy density; α is the condensate phase which defines the curl-free superfluid velocity $\mathbf{v}^s$ as $\mathbf{v}^s=\nabla\alpha$; $\delta H/\delta(\cdot)$ denotes the variational derivative of H with respect to $(\cdot)$; and $\sim$ means equality modulo total derivatives (or "divergences").

The curly bracket part (1.1a) of the Poisson bracket (1.1) is the natural bracket associated [via the formula (2.19) below] to the dual space of the semidirect product Lie algebra

$$\mathfrak{g}(^4\mathrm{He})=D_n\ltimes(\Lambda^0\oplus\Lambda^n\oplus\Lambda^0), \tag{1.2}$$

with the commutator

$$[(X;f;\beta;a),\ (\bar X;\bar f;\bar\beta;\bar a)]=([X,\bar X];X(\bar f)-\bar X(f);X(\bar\beta)-\bar X(\beta);X(\bar a)-\bar X(a)), \tag{1.3}$$

where D_n is the Lie algebra of vector fields on $\mathbb{R}^n$; $\Lambda^k=\Lambda^k(\mathbb{R}^n)$ is the $C^\infty(\mathbb{R}^n)$-module of differential k forms on $\mathbb{R}^n$; $X,\bar X\in D_n$; $f,a,\bar f,\bar a\in\Lambda^0$; $\beta,\bar\beta\in\Lambda^n$; the Lie derivative action of D_n on Λ^k is denoted $X(\cdot)$ for $X\in D_n$ and $(\cdot)\in\Lambda^k$; and the dual coordinates on $(\mathfrak{g}(^4\mathrm{He}))^*$ are M_k to $\partial_k\in D_n$, ρ to $1\in\Lambda^0$, α to $dx_1\wedge\cdots\wedge dx_n\in\Lambda^n$, and σ to $1\in\Lambda^0$. [The usual adiabatic fluid dynamics is recovered from (1.1) for H, F independent of α; equivalently, when Λ^n is absent in (1.2) and $\beta,\bar\beta$ are absent in (1.3).]

The part (1.1b) of the Poisson bracket (1.1) represents the following (generalized) two-cocycle on the Lie algebra $\mathfrak{g}(^4\mathrm{He})$ (1.2):

$$\omega((X;f;\beta;a),(\bar X;\bar f;\bar\beta;\bar a))=-f\bar\beta+\beta\bar f. \tag{1.4}$$

Now we can formulate more precisely the main problem addressed in this paper: what is the origin of the formula (1.4), and why does this skew-symmetric form turn out to be a two-cocycle on the Lie algebra (1.2)? This question will be answered, from various points of view, by Proposition 3.3, Theorem 3.1, Proposition 3.6, Proposition 4.1, and Theorem 4.2.

The plan of the presentation is as follows. In the next section we set up the machinery of Lie algebras over function rings, define generalized two-cocycles on such algebras, and establish the one-to-one correspondence between two-cocycles on Lie algebras on one hand and affine Hamiltonian operators on the other hand, explaining the relation between formulas (1.1) and (1.2)–(1.4). In Sec. III we describe a large class of two-cocycles on semidirect product Lie algebras associated to a pair of mutually adjoint representations (Theorem 3.1), and specialize this result to the case of symplectic two-cocycles of the form (1.4). In the last section we derive defining equations [(4.6), (4.7)] satisfied by two-cocycles on semidirect product Lie algebras of general form.

II. AFFINE HAMILTONIAN STRUCTURES AND GENERALIZED TWO-COCYCLES ON DIFFERENTIAL-DIFFERENCE LIE ALGEBRAS

In this section we recall the one-to-one correspondence between affine Hamiltonian operators and (generalized) two-cocycles on functional Lie algebras. In particular, this will

 0022-2488/85/112754-05$02.50

explain how formula (1.1) is read off the Lie algebra (1.2), (1.3) and the two-cocycle (1.4) on it, and vice versa. The proofs can be found in (Ref. 4, Chap. VIII, Sec. 5).

Let K be a commutative algebra. Let $\partial_1, ..., \partial_n: K \to K$ be n commuting derivations. Let G be a discrete group acting by automorphisms on K, and suppose that the actions of G and ∂'s commute. (The presence of G, which was taken to be $\mathbf{Z}^r$ in Ref. 4, is motivated by the need to cover possible discretizations and numerical models of continuous systems.) Let N be a natural number or ∞. K^N consists of column vectors with only finite number of nonzero components.

A differential-difference algebra structure on K^N is a map $K^N \times K^N \to K^N$ of the form

$$[X, Y]_k = \sum c^k_{i,g,\mu|j,h,\nu}\, \hat{g}\, \partial^\mu(X_i)\cdot \hat{h}\, \partial^\nu(Y_j), \tag{2.1}$$

where the sum in (2.1) is finite for each k; $c^k_{...} \in K$; $g, h \in G$; $\partial^\mu = \partial_1^{\mu_1} \cdots \partial_n^{\mu_n}$ for $\mu = (\mu_1, ..., \mu_n) \in \mathbf{Z}^n_+$; and $\hat{g}(\cdot)$ denotes the image of $(\cdot)$ under the action of $g \in G$. The algebra (2.1) is called a Lie algebra if the commutator (2.1) satisfies the following conditions:

$$\text{(i)}\,[X, Y] = -[Y, X] \quad \text{(skew symmetry)}; \tag{2.2a}$$

$$\text{(ii)}\,[X, [Y, Z]] + \text{cp} = 0 \quad \text{(Jacobi identity)}, \tag{2.2b}$$

where "cp" stands for "cyclic permutation"; and (iii) The properties (i) and (ii) remain true under any (differential-difference) extension $K' \supset K$, i.e., for any extension K' on which G and ∂'s act in a manner compatible with their action on K. This property is called "stability," and it means simply that skew symmetry and the Jacobi identity are properties of the structure constants $c^k_{...}$ themselves, and not of the particular choice of K.

Trivial elements in K are defined as elements of $\operatorname{Im} \mathscr{D} := \Sigma_{s=1}^n \operatorname{Im} \partial_s + \Sigma_{g \in G} \operatorname{Im}(\hat{g} - \hat{e})$, where e is the unit element of G; we write $a \sim b$ if $(a - b)$ is trivial.

A bilinear form on K^N is a map $K^N \times K^N \to K$ of the form

$$\omega(X, Y) = \sum \omega_{i,g,\mu|j,h,\nu}\, \hat{g}\, \partial^\mu(X_i)\cdot \hat{h}\, \partial^\nu(Y_j), \quad \omega_{...} \in K. \tag{2.3}$$

To each bilinear form ω one uniquely associates an operator $b_\omega: K^N \to K^N$ acting by the rule

$$\omega(X, Y) \sim X^t b_\omega(Y), \tag{2.4}$$

where "t" stands for "transpose," so that ("integrating by parts")

$$(b_\omega)_{ij} = \sum \hat{g}^{-1}(-\partial)^\mu\, \omega_{i,g,\mu|j,h,\nu}\, \hat{h}\, \partial^\nu, \tag{2.5}$$

where $(-\partial)^\nu = (-\partial_1)^{\nu_1} \cdots (-\partial_n)^{\nu_n}$. The form ω is called symmetric (resp., skew symmetric), if $\omega(X, Y) \sim \omega(Y, X)$ [resp., $\omega(X, Y) \sim -\omega(Y, X)$]. The form ω is symmetric (resp., skew symmetric) if and only if the corresponding operator b_ω is symmetric: $(b_\omega)^\dagger = b_\omega$ [resp., b_ω is skew symmetric: $(b_\omega)^\dagger = -b_\omega$], where, for an operator $T: K^N \to K^M$, the adjoint operator $T^\dagger: K^M \to K^N$ is uniquely defined by the equation

$$v^t T(u) \sim [T^\dagger(v)]^t u, \quad u \in K^N, \quad v \in K^M, \tag{2.6}$$

so that

$$(T^\dagger)_{ij} = (T_{ji})^\dagger, \tag{2.7}$$

and

$$(a\hat{g}\,\partial^\nu)^\dagger = (-\partial)^\nu \hat{g}^{-1} a, \quad a \in K. \tag{2.8}$$

By an operator we shall always understand a map $K'^N \to K'^M$, each of whose matrix elements is a finite sum of the expressions of the form $a\hat{g}\,\partial^\nu$, $a \in K'$.

A skew-symmetric form ω on a Lie algebra $\mathfrak{g} = K^N$ is called a (generalized) two-cocycle if

$$\omega(X, [Y, Z]) + \text{cp} \sim 0, \quad \forall X, Y, Z \in \mathfrak{g}. \tag{2.9}$$

We now turn to the Hamiltonian formalism. Let I be a countable set. Set $C = K[q_i^{(g|\nu)}]$, $i \in I$, $g \in G$, $\nu \in \mathbf{Z}^n_+$, and extend G and ∂'s to act on C by the rule

$$\hat{h}(q_i^{(g|\nu)}) = q_i^{(hg|\nu)}, \quad \partial^\mu(q_i^{(g|\nu)}) = q_i^{(g|\mu+\nu)}. \tag{2.10}$$

Set $N = |I|$. The Euler–Lagrange map $\delta = \delta/\delta\bar{q}: C \to C^N$, defined by the formula

$$\left(\frac{\delta H}{\delta \bar{q}}\right)_i = \frac{\delta H}{\delta q_i} = \sum \hat{g}^{-1}(-\partial)^\nu \left(\frac{\partial H}{\partial q_i^{(g|\nu)}}\right), \tag{2.11}$$

annihilates $\operatorname{Im} \mathscr{D}$ in C:

$$\operatorname{Ker} \delta = \operatorname{Im} \mathscr{D} + K. \tag{2.12}$$

Here, $\delta H/\delta q_i$ is called the variational derivative of H with respect to q_i.

A derivation X of C over K is called evolutionary if it commutes with the actions of G and ∂'s, so that

$$X = \sum \hat{g}\, \partial^\nu(X_i) \cdot \frac{\partial}{\partial q_i^{(g|\nu)}}, \quad X_i := X(q_i^{(e|0)}). \tag{2.13}$$

The set of all evolution derivations is a Lie algebra denoted $D^{ev}(C)$. For $X \in D^{ev}(C)$, $H \in C$,

$$X(H) \sim \bar{X}^t \frac{\delta H}{\delta \bar{q}} \quad \text{("formula for the first variation")}, \tag{2.14}$$

where

$$(\bar{X})_i = X_i. \tag{2.15}$$

A map $\Gamma: C \to D^{ev}(C)$, $H \mapsto X_H$, is called Hamiltonian, if there exists an operator $B: C^N \to C^N$ such that

$$\bar{X}_H = B\left(\frac{\delta H}{\delta \bar{q}}\right); \tag{2.16}$$

$$\{H, F\} \sim -\{F, H\} \quad \text{(skew symmetry)}, \tag{2.17}$$

where the Poisson bracket $\{H, F\}$ is defined as $X_H(F)$; and

$$X_{\{H,F\}} = [X_H, X_F], \tag{2.18}$$

or, equivalently,

$$\{H, \{F, S\}\} + \text{cp} \sim 0, \tag{2.18'}$$

for any $H, S, F \in C' = K'(q_i^{(g|\nu)})$, with arbitrary extension $K' \supset K$. The property (2.17) is equivalent to B being skew symmetric, while (2.18) can be reduced to a set of quadratic equations on the matrix elements of B.

An operator $a\hat{g}\,\partial^\nu: C \to C$ is called q independent [resp., linear (in q)] if $a \in K$ [resp., if $a = \Sigma\, a_{i,h,\mu}\, q_i^{(h|\mu)}$, $a_{...} \in K$]. An operator is affine if it is a sum of a q-independent and a linear operator. The same terminology applies to sums of operators, and to matrix operators.

Let $B = B^l + b$ be an affine operator: $C^N \to C^N$, with B^l being linear and b being q independent. We make K^N into a (differential-difference) algebra, setting

$$\bar{q}^t[X, Y] \sim X^t B^l(Y), \quad (\bar{q})_i := q_i^{(e|0)}, \quad X, Y \in K^N. \quad (2.19)$$

Conversely, given an algebra structure on K^N, (2.19) defines a linear operator B^l.

Theorem 2.1: Affine Hamiltonian operators are in one-to-one correspondence with generalized two-cocycles on differential-difference Lie algebras. This correspondence is given by the formulas (2.19) and $b_\omega = b$. (For the case when $G = \{e\}$, $n = 1$, and ∂_1 acts trivially on K, this theorem was proven in Ref. 5).

Applying Theorem 2.1 to the Poisson bracket (1.1), and using formulas (2.14)–(2.16), we recover the Lie algebra (1.1), (1.3) together with the two-cocycle (1.4).

Now we are prepared to handle the problem of the origin of the two-cocycle (1.4).

III. θ-ADJOINT REPRESENTATIONS AND GENERALIZED SYMPLECTIC TWO-COCYCLES

Denote $\mathrm{Diff}(V)$ the associative algebra of operators acting from V to V, where $V = K^M$ for some M. We will also denote by $\mathrm{Diff}(V)$ the corresponding Lie algebra.

A representation of a Lie algebra $\mathfrak{g} = K^N$ is an operator $\pi:\mathfrak{g} \to \mathrm{Diff}(V)$, which is a Lie algebra homomorphism, and which remains a Lie algebra homomorphism for any extension $K' \supset K$. Two representations $\pi_i:\mathfrak{g} \to \mathrm{Diff}(V_i)$, $i = 1, 2$, are called θ adjoint, with respect to a bilinear operator $\theta: V_1 \times V_2 \to K$, if

$$\theta(\pi_1(X)(v_1), v_2) \sim -\theta(v_1, \pi_2(X)(v_2)), \quad v_i \in V_i, \quad X \in \mathfrak{g}. \quad (3.1)$$

Recall that if $\pi:\mathfrak{g} \to \mathrm{Diff}(V)$ is a representation then $\mathfrak{g} \times_\pi V = \mathfrak{g} \ltimes V$, called the semidirect product of $\mathfrak{g}$ and V, is a Lie algebra with the commutator

$$\left[\begin{pmatrix} X \\ v \end{pmatrix}, \begin{pmatrix} Y \\ w \end{pmatrix}\right] = \begin{pmatrix} [X, Y] \\ \pi(X)(w) - \pi(Y)(v) \end{pmatrix}, \quad X, Y \in \mathfrak{g}, \quad v, w \in V. \quad (3.2)$$

Theorem 3.1: Let $\pi_i:\mathfrak{g} \to \mathrm{Diff}(V_i)$ be two θ-adjoint representations. Then, on the semidirect product Lie algebra $\mathfrak{g} \ltimes (V_1 \oplus V_2)$, the following bilinear form θ^s is a two-cocycle:

$$\theta^s\left(\begin{pmatrix} X \\ u_1 \\ u_2 \end{pmatrix}, \begin{pmatrix} Y \\ v_1 \\ v_2 \end{pmatrix}\right) = \theta(u_1, v_2) - \theta(v_1, u_2),$$
$$X, Y \in \mathfrak{g}, \quad u_1, v_1 \in V_1, \quad u_2, v_2 \in V_2. \quad (3.3)$$

(The form θ^s is called generalized symplectic two-cocycle. The reason for this name will be clear from Proposition 3.3 below.)

Proof: We have, by (2.9)

$$\theta^s\left(\begin{pmatrix} X \\ u_1 \\ u_2 \end{pmatrix}, \left[\begin{pmatrix} Y \\ v_1 \\ v_2 \end{pmatrix}, \begin{pmatrix} Z \\ w_1 \\ w_2 \end{pmatrix}\right]\right) + \mathrm{cp} \quad [\text{by } (3.2)]$$
$$= \theta^s\left(\begin{pmatrix} X \\ u_1 \\ u_2 \end{pmatrix}, \begin{pmatrix} [Y, Z] \\ \pi_1(Y)(w_1) - \pi_1(Z)(v_1) \\ \pi_2(Y)(w_2) - \pi_2(Z)(v_2) \end{pmatrix}\right) + \mathrm{cp} \quad [\text{by } (3.4)]$$
$$= \theta(u_1, \pi_2(Y)(w_2) - \pi_2(Z)(v_2)) - \theta(\pi_1(Y)(w_1) - \pi_1(Z)(v_1), u_2) + \mathrm{cp}$$
$$= \theta(u_1, \pi_2(Y)(w_2)) + \mathrm{cp} - \theta(w_1, \pi_2(Y)(u_2)) + \mathrm{cp} - \theta(\pi_1(Y)(w_1), u_2) + \mathrm{cp} + \theta(\pi_1(Y)(u_1), w_2) + \mathrm{cp}$$
$$= [\theta(u_1, \pi_2(Y)(w_2)) + \theta(\pi_1(Y)(u_1), w_2) + \mathrm{cp}] - [\theta(w_1, \pi_2(Y)(u_2)) + \theta(\pi_1(Y)(\omega_1), u_2) + \mathrm{cp}] \quad [\text{by } (3.1)] \sim 0. \quad \blacksquare$$

Remark 3.2: If we have an additional representation $\pi_3: \mathfrak{g} \to \mathrm{Diff}(V_3)$, then the same formula (3.3), extended on the left-hand side to include u_3 and v_3, provides a two-cocycle on the Lie algebra $\mathfrak{g} \ltimes (V_1 \oplus V_2 \oplus V_3)$, since the proof above does not depend upon the presence of elements in V_3.

The two-cocycle (1.4) can now be explained by using the following important specialization of Theorem 3.1. Notice that given a representation π_1 and a bilinear operator θ, the θ-adjoint representation π_2 does not, in general, exist. However, it does exist for the case when $\dim(V_2) = \dim(V_1)$ and

$$\theta(v_1, v_2) = v_1^t v_2. \quad (3.4)$$

Proposition 3.3: In this case, π_2 is the dual representation, so that V_2 can be identified with the dual space V_1^* to V_1, thanks to Remark 3.2; and θ^s is the symplectic form on $V_1 \oplus V_1^*$.

Proof: From (3.1) and Proposition 3.3 we have, for $u \in V_1$, $v \in V_2$

$$0 \sim [\pi_1(X)(u)]^t v + u^t[\pi_2(X)(v)] \quad [\text{by } (2.6)]$$
$$\sim u^t[\pi_1(X)^\dagger(v) + \pi_2(X)(v)].$$

Hence (by Lemma VII 1.12 in Ref. 4),

$$[\pi_1(X)^\dagger + \pi_2(X)](v) = 0,$$

so that π_2, if it exists, must be defined by the rule

$$\pi_2(X) = -\pi_1(X)^\dagger, \quad (3.5)$$

i.e., π_2 must be the dual representation. Let us see that π_2, given by (3.5), is indeed a representation. Using the formula $(RS)^\dagger = S^\dagger R^\dagger$, we obtain

$$\pi_2([X, Y]) = -\{\pi_1([X, Y])\}^\dagger = -[\pi_1(X), \pi_1(Y)]^\dagger$$
$$= [\pi_1(X)^\dagger, \pi_1(Y)^\dagger] = [\pi_2(X), \pi_2(Y)].$$

Finally, since θ^s in (3.3) is the pullback of a bilinear form on $V_1 \oplus V_2$, we have, for the θ given by (3.4)

$$\theta^s\left(\begin{pmatrix}u_1\\u_2\end{pmatrix}, \begin{pmatrix}v_1\\v_2\end{pmatrix}\right) = u_1^t\, v_2 - v_1^t\, u_2 = \langle u_1, v_2\rangle - \langle v_1, u_2\rangle, \tag{3.6}$$

in familiar notation. ■

Remark 3.4: The proof above shows that, given a representation π: $\mathfrak{g} \to \mathrm{Diff}(V)$, there exists and is unique the dual representation, $\mathfrak{g} \to \mathrm{Diff}(V^*)$, given by (3.5).

Comparing now formulas (1.4) and (3.6), we see that the two-cocycle in (1.4) is of the symplectic type, provided that the action of D_n on Λ^n is dual to the action of D_n on Λ°. This is indeed the case.

Proposition 3.5: The actions of D_n on Λ^k and Λ^{n-k} are dual to each other, for any k: $0 \leqslant k \leqslant n$.

Proof: Both Λ^k and Λ^{n-k} are free $K = C^\infty$ $(\mathbb{R}^n)$-modules of dimension $m = \binom{n}{k}$. To identify Λ^k and Λ^{n-k} with K^m, we choose a basis of Λ^k consisting of the forms $dx_J = dx_{j(1)} \wedge \cdots \wedge dx_{j(k)}$, $1 \leqslant j(1) < \cdots < j(k) = n$, $J = (j(1), ..., j(k))$, and a basis in Λ^{n-k} consisting of the forms $dx_{J'} = dx_{j'(1)} \wedge \cdots \wedge dx_{j'(n-k)}$, such that $J \cup J' = \{1, 2, ..., n\}$ and $dx_J \wedge dx_{J'} = d^n x := dx_1 \wedge \ldots \wedge dx_n$. Then (3.4) becomes

$$\theta(v, u) d^n x = v \wedge u, \quad v \in \Lambda^k, \quad u \in \Lambda^{n-k}. \tag{3.7}$$

Therefore, for $X = \Sigma X_i\, \partial_i$, we get

$$\begin{aligned}&[\theta(X(v), u) + \theta(v, X(u)) d^n x]\\ &\quad = X(v) \wedge u + v \wedge X(u) = X(v \wedge u) = X(\theta(v,u) d^n x)\\ &\quad = \left[\sum \partial_i(\theta(v, u) X_i)\right] d^n x,\end{aligned}$$

so that

$$\theta(X(v), u) + \theta(v, X(u)) \sim 0,$$

and (3.1) is, thus, satisfied. ■

Theorem 3.1 can be viewed from a slightly more general perspective. Suppose π:$\mathfrak{g} \to \mathrm{Diff}(V)$ is a representation, ω is a skew-symmetric form on V, and V is not necessarily of the form $V_1 \oplus V_2$. Extend ω on $\mathfrak{g} \ltimes V$ via the natural projection $\mathfrak{g} \ltimes V \to V$. We want to know when ω is a two-cocycle on $\mathfrak{g} \ltimes V$.

Proposition 3.6: ω is a two-cocycle on $\mathfrak{g} \ltimes V$ if and only if

$$b_\omega \pi(X) + \pi(X)^\dagger b_\omega = 0, \quad \forall X \in \mathfrak{g}. \tag{3.8}$$

Proof: We have

$$\begin{aligned}&\omega\left(\left[\begin{pmatrix}X\\u\end{pmatrix}, \begin{pmatrix}Y\\v\end{pmatrix}\right], \begin{pmatrix}Z\\w\end{pmatrix}\right) + \mathrm{cp}\\ &\quad = \omega\left(\begin{pmatrix}[X, Y]\\ \pi(X)(v) - \pi(Y)(u)\end{pmatrix}, \begin{pmatrix}Z\\w\end{pmatrix}\right) + \mathrm{cp}\\ &\quad = \omega(\pi(X)(v) - \pi(Y)(u), w) + \mathrm{cp}\\ &\quad = \omega(\pi(X)(v), w) + \mathrm{cp} - \omega(\pi(X)(w), v) + \mathrm{cp}\\ &\quad \sim [\omega(\pi(X)(v), w) + \omega(v, \pi(X)(w))] + \mathrm{cp}\\ &\quad \sim \{[\pi(X)(v)]^t b_\omega(w) + v^t b_\omega \pi(X)(w)\} + \mathrm{cp}\\ &\quad \sim v^t\, [\pi(X)^\dagger b_\omega + b_\omega \pi(X)](w) + \mathrm{cp},\end{aligned}$$

and since v and w are arbitrary, (3.8) is equivalent to ω being a two-cocycle. ■

From Proposition 3.6, Theorem 3.1 and Proposition 3.3 are recovered as follows. Firstly, from (3.3) we have

$$b_{\theta^s} = \begin{pmatrix}0 & b_\theta\\ -b_\theta^\dagger & 0\end{pmatrix}, \tag{3.9}$$

and since

$$\pi(X) = \begin{pmatrix}\pi_1(X) & 0\\ 0 & \pi_2(X)\end{pmatrix},$$

(3.8) becomes

$$\begin{aligned}0 &= b_{\theta^s} \pi(X) + \pi(X)^\dagger (b_{\theta^s})^\dagger\\ &= \begin{pmatrix}0 & b_\theta \pi_2(X)\\ -b_\theta^\dagger \pi_1(X) & 0\end{pmatrix}\\ &\quad + \begin{pmatrix}0 & \pi_1(X)^\dagger b_\theta\\ -\pi_2(X)^\dagger\, b_\theta^\dagger & 0\end{pmatrix},\end{aligned}$$

which is equivalent to the pair of equations

$$b_\theta \pi_2(X) + \pi_1(X)^\dagger b_\theta = 0, \quad \pi_2(X)\, b_\theta^\dagger + b_\theta^\dagger\, \pi_1(X) = 0, \tag{3.10}$$

each one of them saying that π_1 and π_2 are θ adjoint. Secondly, letting $b_\theta = 1$ in (3.9) makes it into (3.6), and makes (3.10) into (3.5).

IV. TWO-COCYCLES ON GENERAL SEMIDIRECT PRODUCT LIE ALGEBRAS

In this section we consider equations defining the most general two-cocycles on semidirect product Lie algebras. The speciality of the results in the preceding section is twofold. Firstly, V was considered to be an Abelian Lie algebra. Secondly, we considered only those two-cocycles on $\mathfrak{g} \ltimes V$ that vanish when one of their arguments is in $\mathfrak{g}$ and another is in V. To improve the treatment, we start by removing first the requirement that V is Abelian.

If $L = K^M$ is a Lie algebra then $\mathrm{Der}(L)$ denotes a Lie subalgebra in $\mathrm{Diff}(L)$ consisting of derivations of L. Recall that if π: $\mathfrak{g} \to \mathrm{Der}(L)$ is a Lie algebra homomorphism then $\mathfrak{g} \times_\pi L = \mathfrak{g} \ltimes L$, called the semidirect product of $\mathfrak{g}$ and L, is a Lie algebra with the commutator

$$\left[\begin{pmatrix}X\\u\end{pmatrix}, \begin{pmatrix}Y\\v\end{pmatrix}\right] = \begin{pmatrix}[X, Y]\\ \pi(X)(v) - \pi(Y)(u) + [u, v]\end{pmatrix}, \quad X, Y \in \mathfrak{g}, \quad u, v \in L. \tag{4.1}$$

Proposition 4.1: Let ω be a skew-symmetric form on L extended on $\mathfrak{g} \ltimes L$ via the projection $\mathfrak{g} \ltimes L \mapsto L$. Then ω is a two-cocycle on $\mathfrak{g} \ltimes L$ if and only if ω is a two-cocycle on L and

$$b_\omega \pi(X) + \pi(X)^\dagger b_\omega = 0, \quad \forall X \in \mathfrak{g}. \tag{4.2}$$

Proof: Since L is isomorphic to the subalgebra in $\mathfrak{g} \ltimes L$ consisting of the elements of the form $\binom{0}{u}$, ω must be a two-cocycle on L. Granted that, we have

$$\begin{aligned}&\omega\left(\left[\begin{pmatrix}X\\u\end{pmatrix}, \begin{pmatrix}Y\\v\end{pmatrix}\right], \begin{pmatrix}Z\\w\end{pmatrix}\right) + \mathrm{cp}\\ &\quad = \omega\left(\begin{pmatrix}[X, Y]\\ \pi(X)(v) - \pi(Y)(u) + [u, v]\end{pmatrix}, \begin{pmatrix}Z\\w\end{pmatrix}\right) + \mathrm{cp}\\ &\quad = \omega(\pi(X)(v) - \pi(Y)(u), w) + \mathrm{cp} + \omega([u, v], w) + \mathrm{cp}\end{aligned}$$

(since ω is a two-cocycle on L)

$$\sim \omega(\pi(X)(v) - \pi(Y)(u), w) + \mathrm{cp}$$

(as in the proof of Proposition 3.6)

$$\sim v^t[\pi(X)^\dagger b_\omega + b_\omega \pi(X)](w) + \text{cp},$$

so that ω is a two-cocycle on $\mathfrak{g}\ltimes L$ iff (4.2) is satisfied. ■

Finally, we consider the general case. Suppose Ω is a two-cocycle on $\mathfrak{g}\ltimes L$. Let $\Omega_{\mathfrak{g}}$ be the restriction of Ω on $\mathfrak{g}$ considered as a subalgebra in $\mathfrak{g}\ltimes L$ consisting of the elements of the form $\binom{X}{0}$. Then, obviously, $\Omega_{\mathfrak{g}}$ is a two-cocycle on $\mathfrak{g}$. Conversely, any two-cocycle on $\mathfrak{g}$ can be pulled back to become a two-cocycle on $\mathfrak{g}\ltimes L$ via the projection $\mathfrak{g}\ltimes L \to \mathfrak{g}$. Thus, we assume from now on that $\Omega_{\mathfrak{g}} = 0$ (by changing Ω into $\Omega - \Omega_{\mathfrak{g}}$). Now, since

$$\Omega\left(\binom{X}{u}, \binom{Y}{v}\right) \sim \Omega\left(\binom{0}{u}, \binom{0}{v}\right) + \Omega\left(\binom{X}{0}, \binom{0}{v}\right) - \Omega\left(\binom{Y}{0}, \binom{0}{u}\right), \tag{4.3}$$

we see that to define Ω we need two objects: a skew-symmetric form ω on L, and a bilinear form ν: $\mathfrak{g}\times L \to K$:

$$\omega(u, v) = \Omega\left(\binom{0}{u}, \binom{0}{v}\right), \quad \nu(X, v) = \Omega\left(\binom{X}{0}, \binom{0}{v}\right), \tag{4.4}$$

$$\Omega\left(\binom{X}{u}, \binom{Y}{v}\right) \sim \omega(u, v) + \nu(X, v) - \nu(Y, u). \tag{4.5}$$

For Ω to be a two-cocycle on $\mathfrak{g}\ltimes L$, ω must be a two-cocycle on L. Granted that, we have to find out when formula (4.5) defines a two-cocycle on $\mathfrak{g}\ltimes L$.

Theorem 4.2: Let ω be a two-cocycle on L, ν: $\mathfrak{g}\times L \to K$ be a bilinear form. Then Ω, given by formula (4.5), is a two-cocycle on $\mathfrak{g}\ltimes L$ iff

$$\omega(\pi(X)(v), w) + \omega(v, \pi(X)(w)) \sim \nu(X, [v, w]), \quad \forall X\in\mathfrak{g}, \quad \forall v, w\in L, \tag{4.6}$$

$$b^\nu([X, Y]) = \pi(Y)^\dagger b^\nu(X) - \pi(X)^\dagger b^\nu(Y), \quad \forall X, Y\in\mathfrak{g}, \tag{4.7}$$

where the operator $b^\nu: \mathfrak{g} \to L^*$ is defined by the relation

$$\nu(X, w) \sim w^t b^\nu(X), \quad \forall X\in\mathfrak{g}, \quad \forall w\in L. \tag{4.8}$$

Proof: We have

$$\Omega\left(\left[\binom{X}{u}, \binom{Y}{v}\right], \binom{Z}{w}\right) + \text{cp}$$
$$= \Omega\left(\binom{[X, Y]}{\pi(X)(v) - \pi(Y)(u) + [u, v]}, \binom{Z}{w}\right) + \text{cp}$$
$$\sim \omega(\pi(X)(v) - \pi(Y)(u) + [u, v], w) + \text{cp} + \nu([X, Y], w) + \text{cp} - \nu(Z, \pi(X)(v) - \pi(Y)(u) + [u, v]) + \text{cp}$$

[since ω is a two-cocycle on L]

$$\sim \{\omega(\pi(X)(v) - \pi(Y)(u), w) - \nu(Z, [u, v])\} + \text{cp} \tag{4.9a}$$
$$+ \{\nu([X, Y], w) - \nu(Z, \pi(X)(v) - \pi(Y)(u))\} + \text{cp}. \tag{4.9b}$$

Since (4.9a) is linear on $\mathfrak{g}$ and bilinear on L, while (4.9b) is bilinear on $\mathfrak{g}$ and linear on L, Ω is a two-cocycle iff (4.9a) and (4.9b) are separately trivial. Transforming (4.9a) we obtain

$$\{\omega(\pi(X)(v), w) - \omega(\pi(X)(w), v) - \nu(X, [v, w])\} + \text{cp},$$

so that (4.6) follows. For (4.9b), we get

$$\{\nu([X, Y], w) - \nu(X, \pi(Y)(w)) + \nu(Y, \pi(X)(w))\} + \text{cp}$$
$$\sim w^t\{b^\nu([X, Y]) - \pi(Y)^\dagger b^\nu(X) + \pi(X)^\dagger b^\nu(Y)\},$$

and (4.7) follows.

Remark 4.3: For $\nu = 0$, Theorem 4.2 reduces to Proposition 4.1.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation and by the U. S. Department of Energy.

[1] I. E. Dzyaloshinskii and G. E. Volovick, Ann. Phys. (NY) **125**, 67 (1980).
[2] D. D. Holm and B. A. Kupershmidt, Phys. Lett. A **91**, 425 (1982).
[3] B. A. Kupershmidt and T. Ratiu, Commun. Math. Phys. **90**, 235 (1983).
[4] B. A. Kupershmidt, *Discrete Lax Equations and Differential-Difference Calculus* (Revue Asterisque, Paris, 1985), Vol. 123.
[5] I. M. Gel'fand and I. Ya. Dorfman, Funkcional. Anal. Prilozen. **15**, 23 (1981); [Funct. Anal. Appl. **15**, 173 (1981)].

On a class of differential equations derived from plasma statistics

Richard L. Liboff
Schools of Applied Physics and Electrical Engineering, Cornell University, Ithaca, New York 14853

(Received 9 August 1984; accepted for publication 17 May 1985)

A class of second-order differential equations stemming from an equation of motion for the two-particle spatial correlation function in one-component plasmas is studied. These equations contain an irregular singularity of varying order at the origin. The general form of solution is obtained which, together with the construction of asymptotic series, demonstrates that both solutions to all equations in the class are singular at the origin. Behavior removed from the origin is shown to be oscillatory or exponential depending on specifics of the equations.

I. INTRODUCTION

In studies of the two-particle correlation function for one-component plasmas,[1–6] a differential equation emerges for the "total" correlation function[7,8] $h(x)$, given by

$$h''(x) + (2/x - k/x^2)h'(x) \pm h = 0. \tag{1}$$

Interparticle displacement is represented by x, k is a constant, and the sign of the last term depends on conditions of the plasma. This equation was recently studied by the author[9] and in the present work these findings are extended to the generalized equation

$$h''(x) + (N/x - k/x^N)h'(x) + Q(x)h(x) = 0, \tag{2}$$

where $Q(x)$ is an arbitrary function and the number $N \geqslant 2$. The latter equation is seen to have an $(N-1)$-order irregular singularity at the origin.[10]

II. ANALYSIS

It is convenient to rewrite (2) in the canonical form[11]

$$h''(x) + P(x)h'(x) + Q(x)h(x) = 0. \tag{3}$$

The integrating factor

$$\phi \equiv \exp \int P\, dx$$

permits (3) to be rewritten as

$$(h'\phi)' + \phi Q h = 0. \tag{4}$$

In the present study,

$$\phi = x^N \exp[k/(N-1)x^{N-1}]. \tag{5}$$

To construct the general form of the solution to (2) we introduce the function

$$\tilde{h}' \equiv \phi^{-1}, \tag{6}$$

which has the property

$$(\tilde{h}'\phi)' = 0. \tag{7}$$

In the present study, $\tilde{h}$ is given by

$$\tilde{h} = \int \frac{dx}{\phi} = \frac{1}{k} \exp \frac{k}{(1-N)x^{N-1}}, \tag{8}$$

so that

$$\tilde{h}\phi = x^N/k. \tag{8a}$$

Note also that

$$\tilde{h}\phi' = \tilde{h}(e^{\int P})' = P\tilde{h}\phi. \tag{8b}$$

The property (7) with reference to (4) suggests a solution in the product form

$$h = \tilde{h} f. \tag{9}$$

Substitution into (4) gives

$$f'' + f'[P + 2/\tilde{h}\phi] + Qf = 0.$$

With (8a) we find

$$f'' + f'[N/x + k/x^N] + Qf = 0. \tag{10}$$

Comparison with (2) indicates that if $f(x,k)$ is a solution to this equation, then

$$k^{-1} \exp[-k/(N-1)x^{N-1}] f(x,-k)$$

is also a solution.

Thus we find that the general solution to (2) is given by

$$h(x) = Af(x,k) + \frac{B}{k} f(x,-k) \exp\left[-\frac{k}{(N-1)x^{N-1}}\right], \tag{11}$$

where A and B are arbitrarily constants. For integer $N \geqslant 2$, (11) implies that at most, only one solution of (2) is regular at the origin. That is, suppose f is regular at the origin. Then the product term in (11) is evidently irregular.

A. Asymptotic series

Following a procedure described by Ince,[10] the nonsingular quality of a solution near a singular point of a differential equation may be examined by Taylor series expanding the solution about this point. In this calculation we assume that Q is regular at the origin and to facilitate calculation set $Q = 1$.

Substitution of the series

$$f(x,k) = \sum_{n=0}^{\infty} a_n x^n \tag{12}$$

into (2) then gives the recurrence relation ($n \geqslant N+1$, $N \geqslant 2$)

$$kna_n = a_{n+1-N}\, n(n+1-N) + a_{n-N-1}, \tag{13}$$

$$a_1 = a_2 = 0, \quad a_3 = a_4 = \cdots = a_N = 0. \tag{13a}$$

Substituting back into (12) gives

$$f(x,k) = a_0\left[1 + \frac{1}{k(N+1)} x^{N+1} + \frac{1}{a_0} \sum_{q=N+2}^{\infty} a_q x^q\right], \tag{14}$$

 0022-2488/85/112759-02$02.50

which is seen to be divergent, as is evident from (13). Thus we find that Taylor series expansion gives a divergent series and we may conclude that both solutions to the class of equations (2) are irregular at the origin. Although divergent, (14) is still useful in the asymptotic sense.[12–14] That is, any finite sum of leading terms is a solution to (2) to arbitrary accuracy for sufficiently small x. For example, substitution of the first two terms of (14) into (2) leaves a remainder of order x^{N-1}.

Combining (14) with (11) gives the following general solution to (2) (with $Q = 1$), relevant to small values of x:

$$h(x) = A\left[1 + \frac{x^{N+1}}{k(N+1)} + \cdots\right] + \frac{B}{k}\exp\left(-\frac{k}{(N-1)x^{N-1}}\right) \times [1 - x^{N+1}/k(N+1) + \cdots]. \tag{15}$$

In passing, it is interesting to note that with $A = -B/k = -1$, the leading terms of (15) return the canonical expression for $h(x)$ near the origin[8]

$$h(x) \simeq -1 + \exp(-V_N(x)/k_B T). \tag{15a}$$

This expression for $h(x)$ is relevant to a fluid at temperature T with interparticle potential $V_N(x)$. Here we have made the identification

$$\frac{V_N(x)}{k_B T} = -\int \frac{k}{x^N}\,dx = \frac{k}{(N-1)x^{N-1}}.$$

B. Oscillatory and exponential behavior

Returning to the mainstream of the analysis, we note that the transformation[11]

$$h = e^{-\int(P/2)}u \tag{16}$$

removes the first-derivative term from (2), leaving the Schrödinger-like equation[15,16]

$$u'' + u[Q - \tfrac{1}{4}P^2 - P'/2] = 0, \tag{17}$$

in which Q plays the role of energy and

$$\tfrac{1}{4}(P^2 + 2P')$$

the role of potential.[17]

In the present study, (17) has the explicit form[18]

$$u'' + u[Q - [(N/4x^2)(N-2) + (k/2x^N)^2]] = 0. \tag{18}$$

Thus we may conclude that solutions of (2) oscillate for

$$Q > (N/4x^2)(N-2) + (k/2x^N)^2. \tag{19}$$

If in addition, Q is bounded

$$0 < Q < M, \tag{20}$$

then (19) indicates that oscillation will occur for sufficiently large x whereas nonoscillatory behavior will occur for sufficiently small x.

Consider the case that Q approaches a positive constant a^2 (or very slowly varying function) for large x. Then in this same domain, (18) reduces to

$$u'' + a^2 u + 0. \tag{21}$$

Combining this result with (16) gives the asymptotic solution

$$h \sim x^{-N/2}\exp\left(-\frac{k}{2(N-1)x^{N-1}}\right)[A\sin ax + B\cos ax], \tag{22}$$

where A and B are arbitrary constants.

In the special case that $N = 2$, (22) reduces to

$$h \sim e^{-k/2x}[\overline{A}j_0(ax) + \overline{B}n_0(ax)], \tag{23}$$

where j_0 and n_0 are spherical Bessel and Neumann functions, respectively.

III. CONCLUSIONS

We have studied a class of differential equations which are of common form with varying order of irregular singularities at the origin. The general structure of the solutions to this class of equations was found and it was concluded that, at most, one solution is regular at the origin. Subsequent series expansion established that both solutions are irregular at the origin for all equations in the class. This series expansion was shown to be asymptotic in the sense that it gives an accurate estimate of the solution for sufficiently small argument.

Finally, it was demonstrated that in different domains of the independent variable, solutions to these differential equations are exponential or oscillatory depending on specifics of coefficients in the equations. The solution to the first member of the class of equations was found to be asymptotic to spherical Bessel functions.

ACKNOWLEDGMENTS

I am grateful to my colleagues Kenneth Gardner, Chris Near, and Brian Jones for fruitful discussions on these topics and their careful reading of this manuscript.

[1] E. H. Lieb and H. Narnhoffer, J. Stat. Phys. **12**, 291 (1975).
[2] R. L. Liboff, Am. J. Phys. **48**, 532 (1980).
[3] M. A. Guillen and R. L. Liboff, J. Plasma Phys. **32**, 81 (1984).
[4] E. A. Frieman and D. L. Book, Phys. Fluids **6**, 1700 (1963).
[5] S. Ishimaru, Rev. Mod. Phys. **54**, 1017 (1982).
[6] C. Deutsch, Y. Forutani, and M. Gombart, Phys. Rev. A **13**, 2244 (1976).
[7] D. L. Goodstein, *States of Matter* (Prentice–Hall, Englewood Cliffs, NJ, 1975).
[8] D. A. McQuarrie, *Statistical Mechanics* (Harper and Row, New York, 1976).
[9] R. L. Liboff, Phys. Rev. A **31**, 1883 (1985). In this paper, a closely allied equation to (2) was found for a neutral fluid with Newtonian attraction and point repulsion between particles.
[10] E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956).
[11] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge U. P., Cambridge, England, 1952).
[12] A. Erdélyi, *Asymptotic Expansions* (Dover, New York, 1956).
[13] C. Bender and S. Orszag, *Advanced Mathematics for Scientists and Engineers* (McGraw–Hill, New York, 1978).
[14] Asymptotic series solutions to singular differential equations were first discussed by L. W. Thomé, J. Math. **76**, 273 (1873).
[15] R. L. Liboff, *Introductory Quantum Mechanics* (Holden–Day, Oakland, CA, 1981).
[16] Singular potentials in the Schrödinger equation are discussed by W. Frank and D. Land, Rev. Mod. Phys. **43**, 36 (1971).
[17] This relation comprises an algorithm which transforms the time-independent Schrödinger equation to the canonical form (3).
[18] Solutions of equations of this form in various limiting situations may be found in E. Kamke, *Differentialgleichungen* (Academische Verlagsgesellschaft, Leipzig, 1943).

Reciprocal relations for effective conductivities of anisotropic media

John Nevard
Department of Mathematics, Stanford University, Stanford, California 94305
Joseph B. Keller
Departments of Mathematics and Mechanical Engineering, Stanford University, Stanford, California 94305

(Received 4 March 1985; accepted for publication 11 June 1985)

Any pair of two-dimensional anisotropic media with local conductivity tensors that are functions of position and that are related to one another in a certain reciprocal way are considered. It is proved that their effective conductivity tensors are related to each other in the same way for both spatially periodic media and statistically stationary random media. An inequality involving the effective conductivity tensors of two three-dimensional media that are reciprocally related is also proved. These results extend the corresponding results for locally isotropic media obtained by Keller, Mendelsohn, Hansen, Schulgasser, and Kohler and Papanicolau. They also yield a relation satisfied by the effective conductivity tensor of a medium reciprocal to a translated or rotated copy of itself.

I. INTRODUCTION

We shall consider the effective conductivity tensor $\Sigma(\sigma)$ of an anisotropic medium with the local conductivity tensor $\sigma(\mathbf{r})$. Our goal is to extend to anisotropic media the reciprocal theorems and reciprocal inequality concerning $\Sigma(\sigma)$, which were proved for locally isotropic media by Keller,[1,2] Mendelsohn,[3] Hansen,[4] Schulgasser,[5] and Kohler and Papanicolau.[6] We shall use the terminology appropriate to electrical conductivity, although the results also apply to thermal conductivity, electrical permittivity, magnetic permeability, diffusivity, etc.

Following the previous authors, we consider two types of media: (a) periodic, in which $\sigma(\mathbf{r})$ is a spatially periodic function of $\mathbf{r}$, and (b) random, in which $\sigma(\mathbf{r})$ is a stationary random function of $\mathbf{r}$. In both cases, $\sigma(\mathbf{r})$ is symmetric and positive definite. In case (a), $\langle f(\mathbf{r})\rangle$ will denote the average of a periodic function over a period cell and in case (b) it will denote the stochastic average of a stationary random function.

In order to define $\Sigma(\sigma)$ for smooth $\sigma(\mathbf{r})$, we consider the potential $\varphi(\mathbf{r})$ in the medium when the average electric field is a given constant $\mathbf{E}$. Then φ satisfies the equations

$$\nabla\cdot[\sigma(\mathbf{r})\nabla\varphi(\mathbf{r})]=0\,, \tag{1.1}$$

$$\varphi(\mathbf{r}+\mathbf{a})-\varphi(\mathbf{r})=\mathbf{E}\cdot\mathbf{a}\,, \tag{1.2a}$$

$$\langle\nabla\varphi(\mathbf{r})\rangle=\mathbf{E}\,. \tag{1.2b}$$

In (1.2a), which holds in case (a), $\mathbf{a}$ is any period of σ, i.e., any vector such that $\sigma(\mathbf{r}+\mathbf{a})=\sigma(\mathbf{r})$. Henceforth (1.2) will mean (1.2a) in case (a) and (1.2b) in case (b). In both cases the problem (1.1) and (1.2) yields a solution φ, which is unique up to an additive constant, so $\nabla\varphi$ is uniquely determined. Furthermore $\nabla\varphi$ is linear in $\mathbf{E}$, and then so is $\langle\sigma\nabla\varphi\rangle$. Therefore we can define $\Sigma(\sigma)$ by

$$\langle\sigma\nabla\varphi\rangle=\Sigma(\sigma)\mathbf{E}\,. \tag{1.3}$$

We will now deduce a few facts abut Σ, which we will need later, beginning with case (a). First we consider any periodic function $\psi(\mathbf{r})$ and integrate $\nabla\cdot(\psi\sigma\nabla\varphi)$ over a period cell. Upon using (1.1) we find that the integrand becomes $\nabla\psi\sigma\nabla\varphi$, while from the divergence theorem and the periodicity of the functions we see that the integral vanishes. Thus in case (a) for any periodic ψ we get

$$\langle\nabla\psi\sigma\nabla\varphi\rangle=0\,. \tag{1.4}$$

Next we denote by φ' the solution of (1.1) and (1.2a) with $\mathbf{E}$ replaced by $\mathbf{E}'$ and we write it in the form $\varphi'=\mathbf{E}'\cdot\mathbf{r}+\psi(\mathbf{r})$. Then we see from (1.2a) that ψ is periodic. Now we take the scalar product of $\mathbf{E}'$ with (1.3) and add the result to (1.4) to get

$$\langle\mathbf{E}'\sigma\nabla\varphi\rangle+\langle\nabla\psi\sigma\nabla\varphi\rangle=\mathbf{E}'\Sigma(\sigma)\mathbf{E}\,. \tag{1.5}$$

Rewriting the left side of (1.5) yields

$$\langle\nabla\varphi'\sigma\nabla\varphi\rangle=\mathbf{E}'\Sigma(\sigma)\mathbf{E}\,. \tag{1.6}$$

Because $\nabla\varphi'$ is linear in $\mathbf{E}$ and $\nabla\varphi$ is linear in $\mathbf{E}$, the left side of (1.6) is a symmetric bilinear form in $\mathbf{E}'$ and $\mathbf{E}$, which is positive definite when $\mathbf{E}'=\mathbf{E}$. Therefore the right side has these same properties, so $\Sigma(\sigma)$ is symmetric and positive definite. We also note that $\nabla\varphi'=\mathbf{E}'+\nabla\psi$, and that $\langle\nabla\psi\rangle=0$ because ψ is periodic. Therefore $\langle\nabla\varphi'\rangle=\mathbf{E}'$, which shows, when the prime is omitted, that (1.2b) holds in case (a) also.

Finally we recall that among all functions $\Phi(\mathbf{r})$ satisfying (1.2a), the solution φ of (1.1) minimizes the quadratic form representing the rate of energy dissipation:

$$\langle\nabla\varphi\sigma\nabla\varphi\rangle\leqslant\langle\nabla\Phi\sigma\nabla\Phi\rangle\,. \tag{1.7}$$

Upon using (1.6) with $\varphi'=\varphi$ to replace the left side of (1.7), we can write

$$\mathbf{E}\Sigma(\sigma)\mathbf{E}=\min_{\Phi}\langle\nabla\Phi\sigma\nabla\Phi\rangle\,. \tag{1.8}$$

In (1.8), Φ must satisfy (1.2a).

In case (b), (1.4) holds when $\nabla\psi$ is stationary, as we show in the Appendix. Then (1.6) follows as before because $\nabla\psi=\nabla\varphi'-\mathbf{E}'$ is stationary. Therefore $\Sigma(\sigma)$ is also symmetric and positive definite in case (b). Furthermore (1.7) and (1.8) also hold provided that $\nabla\Phi$ satisfies (1.2b). Finally in case (b) we will need the variational principle dual to (1.8), which is proved in the Appendix. It is expressed in terms of a constant current $\mathbf{I}$ in the form

$$\mathbf{I}\Sigma^{-1}(\sigma)\mathbf{I} = \min_{\mathbf{J}} \langle \mathbf{J}\sigma^{-1}\mathbf{J}\rangle . \tag{1.9}$$

The minimum is over all statistically stationary functions $\mathbf{J}(\mathbf{r})$ satisfying $\langle \mathbf{J}\rangle = \mathbf{I}$ and $\nabla \cdot \mathbf{J} = 0$.

When $\sigma(\mathbf{r})$ is not smooth, the variational principles (1.8) and (1.9) can be used to define $\Sigma(\sigma)$. Alternatively, if $\sigma(\mathbf{r})$ is piecewise smooth, (1.1) can be replaced by its weak form, and then (1.3) can still be used. From (1.8) it follows that $\Sigma(\sigma)$ is a continuous function of σ in the L_1 norm. Therefore properties of $\Sigma(\sigma)$ that hold for smooth $\sigma(\mathbf{r})$ also hold for any discontinuous $\sigma(\mathbf{r})$, which can be approximated arbitrarily closely by smooth σ.

II. RECIPROCAL RELATIONS IN TWO DIMENSIONS

We shall now state and prove the following reciprocal theorem.

Theorem 1: Let $\sigma(\mathbf{r})$ be the piecewise smooth conductivity tensor of a two-dimensional medium and let $\tau(\mathbf{r}) = k\sigma(\mathbf{r})/\det \sigma(\mathbf{r})$ be that of a second medium, where k is a positive constant scalar. Suppose that $\sigma(\mathbf{r})$ is either a periodic function or a stationary random function. Then

$$\Sigma(\tau) = k\Sigma(\sigma)/\det \Sigma(\sigma) . \tag{2.1}$$

Since σ is symmetric it can be diagonalized by choosing its principal directions as coordinate directions at each point. Then $\sigma = \mathrm{diag}(\sigma_{11},\sigma_{22})$ and the hypothesis of the theorem states that $\tau = \mathrm{diag}(k/\sigma_{22}, k/\sigma_{11})$. The conclusion of the theorem can be rewritten by choosing the principal directions of $\Sigma(\sigma)$ as coordinate directions. Then $\Sigma(\sigma) = \mathrm{diag}[\Sigma_{11}(\sigma),\Sigma_{22}(\sigma)]$ and the theorem states that $\Sigma(\tau) = \mathrm{diag}[k/\Sigma_{22}(\sigma), k/\Sigma_{11}(\sigma)]$.

Proof: We shall prove this theorem for smooth σ, which suffices in view of the last sentence of Sec. I. We shall also prove it for $k = 1$, since $\Sigma(k\tau) = k\Sigma(\tau)$, so the general case follows at once from this case. We begin by introducing the constant antisymmetric tensor

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

and noting that $\tau = \rho\sigma^{-1}\rho^{-1}$. Then if φ is the solution of (1.1), it follows from (1.1) that $\nabla \times (\rho\sigma\nabla\varphi) = 0$. Therefore there is a scalar function $\psi(\mathbf{r})$ such that

$$\nabla\psi = \rho\sigma\nabla\varphi . \tag{2.2}$$

This equation determines $\psi(\mathbf{r})$ up to an additive constant. Furthermore, from (2.2) we get $\nabla \cdot (\tau\nabla\psi) = \nabla \cdot (\rho\nabla\varphi) = \nabla\times\nabla\varphi = 0$. Thus

$$\nabla \cdot (\tau\nabla\psi) = 0 . \tag{2.3}$$

Now (2.3) shows that ψ satisfies (1.1) with σ replaced by τ. In addition (2.2) shows that $\nabla\psi$ is periodic in case (a) and stationary in case (b). Therefore ψ satisfies (1.2a) in case (a) and (1.2b) in case (b), with some constant field $\mathbf{E}'$. To find $\mathbf{E}'$ we just use (2.2) for $\nabla\psi$ in (1.2b), since we have shown that (1.2b) holds in case (a) also. In this way we get, by using (1.3),

$$\mathbf{E}' = \langle \rho\sigma\nabla\varphi \rangle = \rho\Sigma(\sigma)\mathbf{E} . \tag{2.4}$$

On the other hand, from (2.2) and the definition of τ, we have $\nabla\varphi = \sigma^{-1}\rho^{-1}\nabla\psi = \rho^{-1}\tau\nabla\psi$. We now average this relation and note that $\langle\nabla\varphi\rangle = \mathbf{E}$ by (1.2b). We also use (1.3) in the form $\langle\tau\nabla\psi\rangle = \Sigma(\tau)\mathbf{E}'$. In this way we obtain

$$\mathbf{E} = \rho^{-1}\Sigma(\tau)\mathbf{E}' . \tag{2.5}$$

Finally we use (2.5) in (2.4), which yields $\mathbf{E}' = \rho\Sigma(\sigma)\rho^{-1}\Sigma(\tau)\mathbf{E}'$. From this identity we conclude that

$$\Sigma(\tau) = \rho\Sigma^{-1}(\sigma)\rho^{-1} = \Sigma(\sigma)/\det \Sigma(\sigma) . \tag{2.6}$$

This result is just (2.1), which proves the theorem.

As a first application of Theorem 1, we shall consider a periodic or stationary random two-dimensional medium composed of two materials with the constant conductivity tensors σ_1 and σ_2. This means that $\sigma(r)$ is piecewise constant, taking the two values σ_1 and σ_2. We suppose further that σ_1 and σ_2 are related by

$$\sigma_1 = (k/\det \sigma_2)\,\sigma_2 , \quad \sigma_2 = (k/\det \sigma_1)\,\sigma_1 . \tag{2.7}$$

Each of these relations implies the other, and also yields

$$k = (\det \sigma_1 \cdot \det \sigma_2)^{1/2} . \tag{2.8}$$

Let $\tau(r)$ be the conductivity tensor of the medium obtained by interchanging σ_1 and σ_2. Then in view of (2.7), τ and σ are related as in the hypothesis of Theorem 1, so the conclusion (2.1) holds with k given by (2.8).

We can state this result as follows, by writing $\Sigma(\sigma) = \Sigma(\sigma_1,\sigma_2)$ and $\Sigma(\tau) = \Sigma(\sigma_2,\sigma_1)$.

Interchange Corollary: Let $\Sigma(\sigma_1,\sigma_2)$ be the effective conductivity tensor of a periodic or stationary random two-dimensional medium composed of two materials with conductivities σ_1 and σ_2 satisfying (2.7). Let $\Sigma(\sigma_2,\sigma_1)$ be the effective conductivity of the medium obtained by interchanging the two materials. Then

$$\Sigma(\sigma_2,\sigma_1) = \frac{(\det \sigma_1 \cdot \det \sigma_2)^{1/2}}{\det \Sigma(\sigma_1,\sigma_2)}\,\Sigma(\sigma_1,\sigma_2) . \tag{2.9}$$

Taking determinants in (2.9) yields

$$\det \Sigma(\sigma_1,\sigma_2) \cdot \det \Sigma(\sigma_2,\sigma_1) = \det \sigma_1 \cdot \det \sigma_2 . \tag{2.10}$$

As a second application, we shall consider a periodic or stationary random two-dimensional medium for which there is a rigid body motion $r \to Rr + c$ and a scalar constant k such that

$$\sigma(Rr + c) = [k/\det \sigma(r)]\,\sigma(r) . \tag{2.11}$$

Here R is a matrix representing a rotation, reflection, or combination of them and c is a vector representing a translation. By choosing $\tau(r) = \sigma(Rr + c)$, we see that the hypothesis of Theorem 1 holds, so (2.1) applies and yields

$$\Sigma[\sigma(Rr + c)] = \{k/\det \Sigma[\sigma(r)]\}\,\Sigma[\sigma(r)] . \tag{2.12}$$

The translation does not alter $\Sigma[\sigma(r)]$, but the rotation and/or reflection transform it to

$$\Sigma[\sigma(Rr + c)] = R\Sigma[\sigma(r)]\,R^{-1} . \tag{2.13}$$

Upon using (2.13) in (2.12) we obtain

$$R\Sigma[\sigma]\,R^{-1} = [k/\det \Sigma(\sigma)]\,\Sigma(\sigma) . \tag{2.14}$$

By equating the determinants of the two sides of (2.14), we get

$$\det \Sigma(\sigma) = k . \tag{2.15}$$

Then (2.14) simplifies to

$$R\Sigma(\sigma)\,R^{-1} = \Sigma(\sigma) . \tag{2.16}$$

We can summarize these results in the following corollary.

Rigid Motion Corollary: Let $\Sigma(\sigma)$ be the effective conductivity tensor of a periodic or stationary random two-dimensional medium satisfying (2.11) for some rotation and/or reflection matrix R and some translation vector c. Then $\Sigma(\sigma)$ satisfies (2.15) and (2.16). Furthermore from (2.16) it follows that if R is a rotation through an angle other than zero or π, Σ must be a scalar multiple of the identity I. If R is a reflection about the x or y axis, then Σ must be diagonal when expressed in the x, y basis.

As an application of this corollary, we shall consider a medium composed of alternate squares, or other identical space-filling patches, with conductivities σ_1 and σ_2 satisfying (2.7). Then the medium resembles a checkerboard, and there is a translation c that moves each patch of conductivity σ_1 onto a region originally occupied by a patch of conductivity σ_2, and vice versa. Consequently (2.11) holds with this value of c, with $R = I$, and with k given by (2.8). Then the corollary applies, and from (2.15) and (2.8) we get

$$\det \Sigma(\sigma_1,\sigma_2) = (\det \sigma_1 \cdot \det \sigma_2)^{1/2}, \tag{2.17}$$

If the medium has some additional symmetry that makes Σ a scalar, then (2.17) determines it completely. This is the case, for example, when σ_1 and σ_2 are scalars and the medium is actually a checkerboard. Then (2.17) gives $\Sigma(\sigma_1,\sigma_2) = (\sigma_1\sigma_2)^{1/2}$.

III. RECIPROCAL INEQUALITY IN THREE DIMENSIONS

For both the periodic and statistically stationary cases, a generalization of Schulgasser's inequality[5] in three dimensions holds for media whose conductivity tensors are of the form

$$\sigma(\mathbf{r}) = \begin{pmatrix} \sigma_{11} & \sigma_{12} & 0 \\ \sigma_{12} & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{zz} \end{pmatrix}. \tag{3.1}$$

Theorem 2: Let σ be a piecewise smooth positive definite symmetric matrix given by (3.1); let

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

and let $\tau = k\rho\sigma^{-1}\rho^{-1}$, where k is a positive constant. Suppose σ is statistically stationary, or periodic in the 1,2 plane and periodic in z. Then for any two orthogonal directions x, y in the 1,2 plane,

$$\Sigma_{xx}(\sigma)\Sigma_{yy}(\tau) \geqslant k. \tag{3.2}$$

Proof of periodic case: Let D be a period cell of the medium with $|z| \leqslant a/2$, where a is the period in z. We denote its cross section at z by D_z and its volume by $|D|$. Let $\sigma^{(t)}(x,y,z) = \sigma(x,y,t)$. Then $\sigma^{(t)}$ is the conductivity of a medium that does not vary with z, so it is essentially two dimensional. Let P be the set of potentials ψ defined by

$$P = \{\psi\colon\ \psi(x,y,-a/2) = \psi(x,y,a/2),$$
$$\psi \text{ satisfies (1.2a) with } \mathbf{E} = \hat{\imath}\}.$$

We now use the relation (1.8) with $\mathbf{E} = \hat{\imath}$, minimizing over the set of potentials P. Then the left side of (1.8) is $\Sigma_{xx}(\sigma)$. We transform the right side as follows:

$$\begin{aligned}
\Sigma_{xx}(\sigma) &= \min_{\psi\in P} \frac{1}{|D|} \int_D \sigma \nabla\psi\nabla\psi\, dV \\
&\geqslant \frac{1}{a}\int_{-a/2}^{a/2} \min_{\psi\in P} \left[\frac{1}{|D_z|}\int_{D_z} \sigma^{(z)}(x,y,\xi) \right. \\
&\qquad \left. \times \nabla\psi\nabla\psi(x,y,z) dx\, dy\right] dz \\
&= \frac{1}{a}\int_{-a/2}^{a/2} \left[\min_{\psi\in P} \frac{1}{|D|}\int_D \sigma^{(z)}(x,y,\xi) \right. \\
&\qquad \left. \times \nabla\psi(x,y,\xi)\cdot\nabla\psi(x,y,\xi) dx\, dy\, d\xi\right] dz \\
&= \frac{1}{a}\int_{-a/2}^{a/2} \Sigma_{xx}(\sigma^{(z)})\, dz.
\end{aligned}$$

By using brackets to denote the average over a period in z, we can write this result as

$$\Sigma_{xx}(\sigma) \geqslant \langle \Sigma_{xx}(\sigma^{(z)})\rangle. \tag{3.3}$$

Similarly, by choosing $\mathbf{E} = \hat{\jmath}$ in the definition of P we obtain

$$\Sigma_{yy}(\tau) \geqslant \langle \Sigma_{yy}(\tau^{(z)})\rangle. \tag{3.4}$$

Multiplying corresponding sides of (3.3) and (3.4), and then using the Cauchy–Schwartz inequality yields

$$\begin{aligned}
\Sigma_{xx}(\sigma)\Sigma_{yy}(\tau) &\geqslant \langle \Sigma_{xx}(\sigma^{(z)})\rangle\, \langle \Sigma_{yy}(\tau^{(z)})\rangle \\
&\geqslant \langle [\Sigma_{xx}(\sigma^{(z)})\Sigma_{yy}(\tau^{(z)})]^{1/2}\rangle^2.
\end{aligned} \tag{3.5}$$

For each z, $\sigma^{(z)}$ and $\tau^{(z)}$ are essentially two-dimensional conductivities so Theorem 1 holds. By using this theorem for $\Sigma_{yy}(\tau^{(z)})$, and then using the symmetry and positivity of $\Sigma(\sigma^{(z)})$, we obtain

$$\Sigma_{xx}(\sigma^{(z)})\Sigma_{yy}(\tau^{(z)}) = \frac{k\Sigma_{xx}(\sigma^{(z)})\Sigma_{yy}(\sigma^{(z)})}{\Sigma_{xx}(\sigma^{(z)})\Sigma_{yy}(\sigma^{(z)}) - \Sigma_{xy}^2(\sigma^{(z)})} \geqslant k. \tag{3.6}$$

Finally by using (3.6) in (3.5) we get

$$\Sigma_{xx}(\sigma)\Sigma_{yy}(\tau) \geqslant k. \tag{3.7}$$

Proof of statistically stationary case: The proof is different here, for the intermediate inequality (3.3) is in spatial terms, and has no immediate analog in this case. We follow closely the proof of Kohler and Papanicolau,[6] who treated statistically stationary isotropic media. We start with the definition $\Sigma_{xx}(\sigma) = \hat{\imath}^T\Sigma(\sigma)\hat{\imath}$ and use (1.8) in the statistically stationary case. Then we can write

$$\Sigma_{xx}(\sigma) = \min_{\Phi} \langle \sigma\nabla\Phi\cdot\nabla\Phi\rangle, \tag{3.8}$$

where we minimize over all stationary Φ such that $\langle\nabla\Phi\rangle = \hat{\imath}$. Now we consider the class $\mathscr{E}$ of stationary fields defined by $\mathscr{E} = \{\mathbf{E}\colon \langle\mathbf{E}\rangle = \hat{\imath},\ \partial_1 E_2 = \partial_2 E_1\}$. This set includes all $\nabla\Phi$ considered above, so that

$$\min_{\Phi}\langle\sigma\nabla\Phi\cdot\nabla\Phi\rangle \geqslant \min_{\mathbf{E}\in\mathscr{E}}\langle\sigma\mathbf{E}\cdot\mathbf{E}\rangle. \tag{3.9}$$

If $\mathbf{E} = (E_1,E_2,E_3)$ is in $\mathscr{E}$, then so is $\mathbf{E}' = (E_1,E_2,0)$. Furthermore $\sigma\mathbf{E}'\cdot\mathbf{E}' \leqslant \sigma\mathbf{E}\cdot\mathbf{E}$. Therefore the minimum of the right side of (3.9) is achieved for a vector with $E_3 = 0$:

$$\min_{\mathbf{E}\in\mathscr{E}}\langle\sigma\mathbf{E}\cdot\mathbf{E}\rangle = \min_{\substack{\mathbf{E}\in\mathscr{E}\\ E_3=0}}\langle\sigma\mathbf{E}\cdot\mathbf{E}\rangle . \tag{3.10}$$

Upon combining (3.8)–(3.10), we get

$$\Sigma_{xx} = \min_{\substack{\mathbf{E}'\in\mathscr{E}\\ E_3=0}}\langle\sigma\mathbf{E}\cdot\mathbf{E}\rangle . \tag{3.11}$$

Next we introduce the set

$$\mathscr{J} = \{\mathbf{J}: \mathbf{J} \text{ is stationary}, \nabla\cdot\mathbf{J}=0, \langle\mathbf{J}\rangle = \hat{j}\},$$

and note that $\{\mathbf{E}: \mathbf{E}\in\mathscr{E}, E_3=0\} = \{\rho^{-1}\mathbf{J}: \mathbf{J}\in\mathscr{J}, J_3=0\}$. Then,

$$\min_{\substack{\mathbf{E}\in\mathscr{E}\\ E_3=0}}\langle\sigma\mathbf{E}\cdot\mathbf{E}\rangle = \min_{\substack{\mathbf{J}\in\mathscr{J}\\ J_3=0}}\langle\sigma\rho^{-1}\mathbf{J}\cdot\rho^{-1}\mathbf{J}\rangle . \tag{3.12}$$

We remove the condition $J_3=0$, decreasing the right-hand side, and bring ρ through the inner product to obtain

$$\min_{\substack{\mathbf{E}=\mathscr{E}\\ E_3=0}}\langle\sigma\mathbf{E}\cdot\mathbf{E}\rangle \geqslant \min_{\mathbf{J}\in\mathscr{J}}\langle\rho\sigma\rho^{-1}\mathbf{J}\cdot\mathbf{J}\rangle . \tag{3.13}$$

Recalling that $\rho\sigma\rho^{-1} = k\tau^{-1}$, we can use (1.9) to write

$$\min_{\substack{\mathbf{E}\in\mathscr{E}\\ E_3=0}}\langle\sigma\mathbf{E}\cdot\mathbf{E}\rangle \geqslant k\hat{j}^T\Sigma^{-1}(\tau)\hat{j} . \tag{3.14}$$

By using (3.14) in (3.11) we get

$$\Sigma_{xx}(\sigma) \geqslant k\hat{j}^T\Sigma^{-1}(\tau)\hat{j} . \tag{3.15}$$

To evaluate the right side of (3.15) we note that Σ is symmetric and positive definite. Therefore the quadratic form $\mathbf{v}\Sigma^{-1}\mathbf{w}$ is an inner product and satisfies the Cauchy–Schwartz inequality. Thus we have

$$[\hat{j}^T\Sigma^{-1}(\tau)\hat{j}]\,[(\Sigma(\tau)\hat{j})^T\Sigma^{-1}(\tau)(\Sigma(\tau)\hat{j})] \geqslant [\hat{j}^T\Sigma^{-1}(\tau)\Sigma(\tau)\hat{j}]^2 = 1 .$$

Simplifying the left side of this inequality yields

$$[\hat{j}^T\Sigma^{-1}(\tau)\hat{j}]\,[\hat{j}^T\Sigma(\tau)\hat{j}] \geqslant 1 . \tag{3.16}$$

Substituting the definition $\Sigma_{yy}(\tau) = \hat{j}^T\Sigma(\tau)\hat{j}$ into (3.16) and using (3.15) yield

$$\Sigma_{xx}(\sigma)\Sigma_{yy}(\tau) \geqslant k , \tag{3.17}$$

as was to be shown.

ACKNOWLEDGMENTS

This research was supported by the Office of Naval Research, the Air Force Office of Scientific Research, and the National Science Foundation.

APPENDIX: STATISTICAL VARIATIONAL PRINCIPLES

We prove that

$$\langle\nabla\psi\sigma\nabla\varphi\rangle = 0 \tag{1.4}$$

in the statistically stationary case, where $\nabla\psi$ is stationary and $\langle\nabla\psi\rangle = 0$. ($\nabla\varphi$ is assumed to be statistically stationary.)

To begin with, ψ itself is stationary since $\nabla\psi$ is stationary and $\langle\nabla\psi\rangle = 0$. We introduce the vector field $\mathbf{T}(\mathbf{r})$, defined by

$$T_i(\mathbf{r}) = \left\langle \psi(\mathbf{x}+\mathbf{r})\sigma_{ij}(\mathbf{x})\frac{\partial\varphi(\mathbf{x})}{\partial x_j}\right\rangle . \tag{A1}$$

(The summation convention is used.) Since everything within the brackets is stationary, we can also write

$$\mathbf{T}_i(\mathbf{r}) = \left\langle \psi(\mathbf{x})\sigma_{ij}(\mathbf{x}-\mathbf{r})\frac{\partial\varphi(\mathbf{x}-\mathbf{r})}{\partial x_j}\right\rangle . \tag{A2}$$

To establish (1.4) we evaluate $\nabla\cdot\mathbf{T}$ at $\mathbf{r}=0$ in two different ways. First, using (A1)

$$\left.\frac{\partial T_i(\mathbf{r})}{\partial r_i}\right|_{\mathbf{r}=0} = \left.\left\langle\sigma_{ij}(\mathbf{x})\frac{\partial\varphi(\mathbf{x})}{\partial x_j}\frac{\partial\psi(\mathbf{x}+\mathbf{r})}{\partial r_i}\right\rangle\right|_{\mathbf{r}=0} .$$

This is just

$$\left.\frac{\partial T_i(\mathbf{r})}{\partial r_i}\right|_{\mathbf{r}=0} = \langle\sigma\nabla\varphi\cdot\nabla\psi\rangle . \tag{A3}$$

On the other hand, using (A2)

$$\left.\frac{\partial T_i(\mathbf{r})}{\partial r_i}\right|_{\mathbf{r}=0} = -\left.\left\langle\psi(\mathbf{x})\frac{\partial}{\partial r_i}\left[\sigma_{ij}(\mathbf{x}-\mathbf{r})\frac{\partial\varphi(\mathbf{x}-\mathbf{r})}{\partial r_j}\right]\right\rangle\right|_{\mathbf{r}=0} . \tag{A4}$$

However, the right-hand side of (A4) is 0 since φ is a solution to (1.1). Thus, (A3) and (A4) imply (1.4). The preceding argument is due to Molyneaux.[7]

The dual variational principle

$$\mathbf{I}\Sigma^{-1}(\sigma)\mathbf{I} = \min_{\mathbf{J}}\langle\mathbf{J}\sigma^{-1}\mathbf{J}\rangle , \tag{1.9}$$

where the minimum is over all stationary functions $\mathbf{J}(\mathbf{r})$ satisfying $\langle\mathbf{J}\rangle = \mathbf{I}$ and $\nabla\cdot\mathbf{J} = 0$, is established as follows: Let $\mathbf{J}_0$ be stationary and divergence-free and let $\langle\mathbf{J}_0\rangle = 0$. We can write

$$\langle\nabla\varphi\cdot\mathbf{J}_0\rangle = \left\langle\frac{\partial\varphi}{\partial x_i}\delta_{ijk}\frac{\partial F_k}{\partial x_j}\right\rangle , \tag{A5}$$

where δ_{ijk} is the Levi-Civita tensor, i.e., $J_0 = \operatorname{curl}\mathbf{F}$ for some field $\mathbf{F}$. As above, since $\langle\mathbf{J}_0\rangle = 0$ and is stationary, $\mathbf{F}$ is also stationary. Analogously to our introduction of T, we define $S_j(\mathbf{r})$ by

$$S_j(\mathbf{r}) = \left\langle\frac{\partial\varphi(\mathbf{x})}{\partial x_i}\delta_{ijk}\,\mathbf{F}_k(\mathbf{x}+\mathbf{r})\right\rangle \tag{A6}$$

and note that

$$S_j(\mathbf{r}) = \left\langle\frac{\partial\varphi(\mathbf{x}-\mathbf{r})}{\partial x_i}\delta_{ijk}\,\mathbf{F}_k(\mathbf{x})\right\rangle . \tag{A7}$$

Now, computing $\nabla\cdot S$ at $\mathbf{r}=0$, first using (A6), we get

$$\left.\frac{\partial S_j(\mathbf{r})}{\partial r_j}\right|_{\mathbf{r}=0} = \left.\left\langle\frac{\partial\varphi(\mathbf{x})}{\partial x_i}\delta_{ijk}\frac{\partial\mathbf{F}_k(\mathbf{x}+\mathbf{r})}{\partial r_j}\right\rangle\right|_{\mathbf{r}=0} ,$$

which, by (A5), is just

$$\left.\frac{\partial S_j(\mathbf{r})}{\partial r_j}\right|_{\mathbf{r}=0} = \langle\nabla\varphi\cdot\mathbf{J}_0\rangle . \tag{A8}$$

Using (A7),

$$\left.\frac{\partial S_j(\mathbf{r})}{\partial r_j}\right|_{\mathbf{r}=0} = -\left.\left\langle\frac{\partial^2\varphi(\mathbf{x}-\mathbf{r})}{\partial r_j\partial r_i}\delta_{ijk}F_k(\mathbf{x})\right\rangle\right|_{\mathbf{r}=0} . \tag{A9}$$

Here, the right-hand side is 0 because $\nabla\times\nabla\varphi = 0$. It follows that $\langle\nabla\varphi\cdot\mathbf{J}_0\rangle = 0$.

Let $\mathbf{J}_s = \sigma\nabla\varphi$, the actual current. By the above, $\langle\sigma^{-1}\mathbf{J}_s\cdot\mathbf{J}_0\rangle = 0$, hence

$$\langle\sigma^{-1}(\mathbf{J}_s\cdot\mathbf{J}_0)\cdot(\mathbf{J}_s+\mathbf{J}_0)\rangle = \langle\sigma^{-1}\mathbf{J}_s\cdot\mathbf{J}_s\rangle + \langle\sigma^{-1}\mathbf{J}_0\cdot\mathbf{J}_0\rangle \geqslant \langle\sigma^{-1}\mathbf{J}_s\cdot\mathbf{J}_s\rangle . \tag{A10}$$

Since any admissible field $\mathbf{J}$ may be written as $\mathbf{J}_s + \mathbf{J}_0$, for a suitable $\mathbf{J}_0$, we see that choosing $\mathbf{J} = \mathbf{J}_s$ minimizes $\langle \sigma^{-1} + \mathbf{J} \cdot \mathbf{J} \rangle$. Also, $\langle \sigma^{-1} \mathbf{J}_s \cdot \mathbf{J}_s \rangle = \langle \nabla\varphi \cdot \nabla\varphi \rangle = \mathbf{E}^T \Sigma(\sigma) \mathbf{E}$. Thus, (A10) implies

$$\mathbf{E}^T \Sigma(\sigma) \mathbf{E} = \min_{\mathbf{J}} \langle \sigma^{-1} + \mathbf{J} \cdot \mathbf{J} \rangle . \tag{A11}$$

Finally, the average current $\mathbf{I} = \Sigma(\sigma)\mathbf{E}$, and we may substitute $\mathbf{E} = \Sigma^{-1}(\sigma) I$ into (A11) to obtain (1.9).

[1]J. B. Keller, J. Appl. Phys. **34**, 991 (1963).
[2]J. B. Keller, J. Math. Phys. **5**, 548 (1964).
[3]K. Mendelsohn, J. Appl. Phys. **46**, 917 (1975).
[4]K. Schulgasser, J. Math. Phys. **17**, 378 (1976).
[5]E. B. Hansen, "Generalization of a theorem on the conductivity of a composite medium," unpublished.
[6]W. Kohler and G. C. Papanicolau, Springer Lect. Notes Phys. **154**, 111 (1981).
[7]J. E. Molyneaux, J. Math. Phys. **11**, 1172 (1970).

The structures of generalized noncommutative Toda lattices

B. A. Kupershmidt
The University of Tennessee Space Institute, Tullahoma, Tennessee 37388

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

Non-Abelian Toda lattices with a finite number of "above-diagonal" variables are related to appropriate Lie algebras of operators and a certain two-cocycle.

I. INTRODUCTION

Among various generalizations of the classical Toda lattice, one can discern four separate ideas pertinent to the infinite (or periodic) case. If one uses the symbolic notations of Ref. 1 then the equations of the Toda hierarchy are the discrete Lax equations

$$\partial_P(L) = [P_+, L] = [-P_-, L], \tag{1.1}$$

with

$$L = \xi + q_0 + \xi^{-1} q_1, \tag{1.2}$$

and the foregoing generalizations concentrate on the following points.

(a) Consider q_0 and q_1 in (1.2) as r by r matrices; this is called "the non-Abelian Toda lattice" in the physics literature.

(b) Consider the universal case[1]

$$L = \xi + \sum_{j\geqslant 0} \xi^{-j} q_j, \tag{1.3}$$

from which (1.2) can be obtained by the specialization $\{q_j = 0 \mid j \geqslant 2\}$.

(c) Allow L in (1.3) to contain a finite number of "above-diagonal" variables[2,3]

$$L = \xi^{\beta}\Big(1 + \sum_{j\geqslant 0} \xi^{-j-1} q_j\Big), \tag{1.4}$$

with some $\beta \geqslant 2$, $\beta \in \mathbf{Z}_+$.

(d) Specialize L in (1.4) by restricting the surviving flows of (1.1) to the "submanifold" $\{q_j = 0 \mid j \not\equiv 0 \pmod{\alpha}\}$, $\alpha \geqslant 2$, $\alpha \in \mathbf{Z}_+$.

In the scalar situation, that is, when all the variables q_j commute with each other, the cases (b)–(d) were analyzed in considerable detail in Refs. 1 and 2. The result of this analysis can be summarized as follows: the universal case (1.3), while making the theory more transparent, presents no new features compared with the proper Toda lattice (1.2), except for the problem of the existence of the third Hamiltonian structure. The associated Lie algebra, the Lie algebra $g_{\geqslant 0}$ of non-negative-order Volterra operators, is an analog of the corresponding Lie algebra in the differential case[4] (this and other results mentioned in the Introduction will be explained in the sections below). Allowing the above-diagonal variables (1.4) results in the appearance of a generalized two-cocycle ω on the Lie algebra $g_{<0}$ of negative-order Volterra operators. Also, positive and negative variables split in the Hamiltonian form; specialization (d) presents some technical problems but no new features.

In this paper we analyze the cumulative effect of noncommutativity (a), universality (b), and above-diagonality (c). In other words, we consider the Lax equations (1.1) for the Lax operator (1.4) with variables q_j being r by r matrices. It turns out that important ingredients of the scalar case, namely, Lie algebras $g_{\geqslant 0}$ and $g_{<0}$ and a two-cocycle ω on $g_{<0}$, still appear in the noncommutative situation, though in a generalized manner.

The plan of the paper is as follows. In Sec. II we establish a proper setup for Eqs. (1.1), and then cast them in a variational form in Sec. III. Finally, in Sec. IV we interpret the resulting equations from the point of view of appropriate Volterra Lie algebras.

II. THE LAX EQUATIONS

Let k be a field of characteristic zero and let $C = \mathrm{k}[q^{(m)}_{j,\mu\nu}]$, $m \in \mathbf{Z}$, $j \in \mathbf{Z}_+$, $1 \leqslant \mu, \nu \leqslant r$, be the algebra of polynomials in free variables $q^{(m)}_{j,\mu\nu}$. We make C into an algebra with an automorphism Δ by defining $\Delta(q^{(m)}_{j,\mu\nu}) = q^{(m+1)}_{j,\mu\nu}$, $\Delta(c) = c$, $\forall c \in \mathrm{k}$. Let $\mathrm{Mat}_r(C)$ be the algebra of r by r matrices over C with Δ naturally extended to it, and let $C' = \mathrm{Mat}_r(C)((\xi^{-1}))$ be the associative algebra of the Laurent series, with relations $\xi^s g = \Delta^s(g)\xi^s$, $\forall g \in \mathrm{Mat}_r(C)$. We consider matrices q_j in (1.4) having matrix elements $q_{j,\mu\nu}$. Thus L in (1.4) belongs to C' and for every $n \in \mathbf{Z}_+$ we can form the Lax equations (1.1) with $P = L^n$. The resulting derivations (or "flows") all commute and have a common set of integrals

$$H_n = n^{-1}\,\mathrm{Tr}\,\mathrm{Res}\; L^n, \tag{2.1}$$

where Res singles out the ξ^0 coefficient (see Refs. 2 and 3).

Denote $Q_s = q_{\beta-1+s}$, $s \geqslant 0$; $R_i = q_{\beta-2-i}$, $0 \leqslant i \leqslant \beta - 2$ (for $\beta = 1$ no R is introduced). Let

$$L^n = \sum_s p_s(n)\xi^s, \quad p_s(n) \in \mathrm{Mat}_r(C). \tag{2.2}$$

Rewriting in longhand the first equality in (1.1), $\partial_P(L) = [P_+, L]$, we get

$$\partial_P(Q_s) = \sum_{k\geqslant 0} [\Delta^s(p_k(n))Q_{k+s} - \Delta^{-k}(Q_{k+s}p_k(n))], \tag{2.3}$$

while the second equality in (1.1), $\partial_P(L) = [-P_-, L]$, results in

$$\partial_P(R_i) = (1 - \Delta^{-\beta})\Delta^{\beta-1-i} p_{i+1-\beta}(n) + \sum_{s<0} [\Delta^{-s}(R_{i-s}p_s(n)) - \Delta^{-i-1}(p_s(n))R_{i-s}], \tag{2.4}$$

where we agree, in (2.4) and below, to drop those terms in the sums which do not make sense.

 0022-2488/85/112766-03$02.50

III. HAMILTONIAN FORM

The Hamiltonian formalism is a device for reexpressing equations of "motion" in terms of associated conservation laws. For Lax equations, one uses for this purpose the natural extension of the formal calculus of variations from the ring C into the C'-bimodule of differential forms $\Omega^1(C')$ (see, e.g., Refs. 1, 2, and 5). The calculus then yields

$$d\,\mathrm{Tr}\,\mathrm{Res}\; L^n \sim n\,\mathrm{Tr}\,\mathrm{Res}(L^{n-1}\,dL)\,, \tag{3.1}$$

where $a\sim b$ means $(a-b)\in\mathrm{Im}(\Delta-1)$. Using (2.1) and (2.2) we get from (3.1)

$$\begin{aligned} dH_{n+1} &= d\,[1/(n+1)]\mathrm{Tr}\,\mathrm{Res}\; L^{n+1}\\ &\sim \mathrm{Tr}\,\mathrm{Res}(L^n/dL)\\ &= \mathrm{Tr}\,\mathrm{Res}\Big(\sum_{s,j} p_s(n)\xi^s\xi^{\beta-j-1}\,dq_j\Big)\\ &= \mathrm{Tr}\sum_j p_{j+1-\beta}(n)dq_j\,. \end{aligned}$$

Therefore,

$$p_{j+1-\beta}(n)=\frac{\delta H}{\delta q_j^t}\,,\quad H=H_{n+1}\,, \tag{3.2}$$

where t stands for transpose

$$\left(\frac{\delta H}{\delta q_j^t}\right)_{\mu\nu}=\frac{\delta H}{\delta q_{j,\nu\mu}}\,.$$

Substituting (3.2) into (2.3) and (2.4), we get

$$\partial_P(Q_s)=\sum_{k>0}\left[\Delta^s\left(\frac{\delta H}{\delta Q_k^t}\right)Q_{k+s}-\Delta^{-k}\left(Q_{k+s}\frac{\delta H}{\delta Q_k^t}\right)\right], \tag{3.3}$$

$$\partial_P(R_i)=(1-\Delta^{-\beta})\Delta^{\beta-1-i}\frac{\delta H}{\delta R^t_{\beta-2-i}} \tag{3.4a}$$

$$+\sum_{s>0}\left[\Delta^{1+s}\left(R_{i+s+1}\frac{\delta H}{\delta R_s^t}\right)-\Delta^{-i-1}\left(\frac{\delta H}{\delta R_s^t}\right)R_{i+s+1}\right]. \tag{3.4b}$$

Equations (3.3) and (3.4) give us the Hamiltonian form of the Lax equations (1.1). Notice that the variables Q and R are *split* in this form. To show that Eqs. (3.3) and (3.4) are indeed Hamiltonian, and not merely expressions involving variational derivatives of $H=H_{n+1}$, we identify them with relevant objects on dual spaces of appropriate Lie algebras.

IV. LIE ALGEBRAS $g_{>0}$ AND $g_{<0}$

Let K be a k-algebra with an automorphism Δ over k, let K_r denote $\mathrm{Mat}_r(K)$, and let $g_{>0}$ and $g_{<0}$ be the Lie algebras generated by the associative algebras $K_r[[\xi]]$ and $\xi^{-1}\times K_r[[\xi^{-1}]]$, respectively, with relations $\xi^s g=\Delta^s(g)\xi^s$, $g\in K_r$. Let B^+ and B^- be the matrices of the (formal) Kirillov form on the dual spaces of the Lie algebras $g_{>0}$ and $g_{<0}$, respectively (see Refs. 2, 4, and 6). Let us compute $B^+=(B^+_{s,\sigma\rho|k,\mu\nu})$. If $X=\sum X_s\xi^s$, $Y=\sum Y_k\xi^k\in g_{>0}$, then, by definition, $\sum X_{s,\rho\sigma}B^+_{s,\sigma\rho|k,\mu\nu}Y_{k,\nu\mu}\sim\sum Q_{s,\mu\nu}(\mathrm{Res}[X,Y]\times\xi^{-s})_{\nu\mu}$, where $Q_{s,\mu\nu}$ are dual coordinates on $g^*_{>0}$. Thus

$$B^+_{s,\sigma\rho|k,\mu\nu}=Q_{k+s,\mu\rho}\Delta^s\delta^\sigma_\nu-\Delta^{-k}Q_{k+s,\sigma\nu}\delta^\rho_\mu\,. \tag{4.1}$$

Since (3.3) can be rewritten as

$$\partial_P(Q_{s,\sigma\rho})=\sum B^+_{s,\sigma\rho|k,\mu\nu}\left(\frac{\delta H}{\delta Q_{k,\mu\nu}}\right), \tag{4.2}$$

we obtain the following.

Proposition 4.1: The Q part (3.3) of the generalized Toda lattices is associated with the dual space of the Lie algebra $g_{>0}$.

Remark 4.2: In the case $\beta=1$ of the non-Abelian Toda lattice proper, there are no R variables involved.

The same calculation can be used for $g_{<0}$. Let $X=\sum X_s\xi^{-s-1}$, $Y=\sum Y_k\xi^{-k-1}\in g_{<0}$, and let $R_{s,\mu\nu}$ be the dual coordinates on $g^*_{<0}$. Then

$$\sum X_{s,\sigma\rho}B^-_{s,\rho\sigma|k,\mu\nu}Y_{k,\nu\mu}\sim\sum R_{s,\mu\nu}(\mathrm{Res}[X,Y]\xi^{s+1})_{\nu\mu}\,,$$

so

$$B^-_{s,\rho\sigma|k,\mu\nu}=\Delta^{1+k}R_{s+k+1,\rho\nu}\delta^\sigma_\mu-R_{s+k+1,\mu\sigma}\Delta^{-s-1}\delta^\rho_\nu\,. \tag{4.3}$$

Let us consider the following bilinear form on $g_{<0}$:

$$\omega(X,Y)=\mathrm{Tr}\,\mathrm{Res}[X(1-\Delta^\beta)(Y)\xi^\beta]\,. \tag{4.4}$$

It is easy to check that $\omega(X,Y)\sim-\omega(Y,X)$ and

$$\{\omega([X,Y],Z)+\omega([Z,X],Y)+\omega([Y,Z],X)\}\sim 0\,,\quad \forall X,Y,Z\in g_{<0}.$$

In other words, ω is a (generalized) two-cocycle on $g_{<0}$ (see, e.g., Chap. VIII in Ref. 2). The matrix of the corresponding operator $b{:}g_{<0}\to g^*_{<0}$ is defined by

$$\omega(X,Y)\sim X_{s,\sigma\rho}b_{s,\rho\sigma|k,\mu\nu}Y_{k,\nu\mu}\,,$$

hence

$$b^-_{s,\rho\sigma|k,\mu\nu}=\delta^{s+k}_{\beta-2}\delta^\rho_\nu\delta^\sigma_\mu(\Delta^{-s-1}-\Delta^{1+k})\,. \tag{4.5}$$

Let I_β be the ideal in $g_{<0}$ consisting of elements $\sum X_s\xi^{-s-1}$ with $X_s=0$ for $s>\beta-2$. Since $\omega(I_\beta,g_{<0})\sim 0$ we have a well-defined two-cocycle on the factor algebra $g_\beta:=g_{<0}/I_\beta$. Formulas (4.3) and (4.5) will remain unchanged with the exception that the variables R_{s+k+1} in (4.3) must be dropped for $s+k+1>\beta-2$. With this understanding the evolution equations on g^*_β, with the Hamiltonian matrix $-(B^-+b)$, become

$$\partial_P(R_{s,\rho\sigma})=-(B^-_{s,\rho\sigma|k,\mu\nu}+b_{s,\rho\sigma|k,\mu\nu})\frac{\delta H}{\delta R_{k,\mu\nu}}\,,$$

which is exactly (3.4). The result of this reasoning we collect in the following.

Theorem 4.1: The Hamiltonian form of the generalized Toda lattice equations (3.3) and (3.4) is generated by the Poisson structure on the dual to the Lie algebra $g_{<0}\oplus g_\beta$ together with the two-cocycle ω_β on g_β.

Remark 4.3: If the highest term $\xi^\beta 1$ of L in (1.4) is taken to be $\xi^{\beta}c$, $c = \operatorname{diag}(c_1,\ldots c_r)$, the corresponding form ω ceases to be a two-cocycle when c is not a scalar matrix.

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation Grant No. DMS-8403200.

[1] B. A. Kupershmidt, Lett. Math. Phys. **6**, 85 (1982).

[2] B. A. Kupershmidt, *Discrete Lax Equations and Differential-Difference Calculus* (Revue Asterisque, Paris, 1985), Vol. 123.

[3] P. van Moerbeke and D. Mumford, Acta Math. **143**, 93 (1979).

[4] D. R. Lebedev and Yu. I. Manin, Funct. Anal. Appl. **13**, 40 (1979).

[5] Yu. I. Manin, J. Sov. Math. **11**, 1 (1979).

[6] V. Guillemin and S. Sternberg, Ann. Phys. **127**, 220 (1980).

Formulation of Noether's theorem for Fokker-type variational principles

Warren N. Herman
Department of Mathematics and Physics, Spring Garden College, Philadelphia, Pennsylvania 19119

(Received 23 April 1985; accepted for publication 7 June 1985)

A formulation of Noether's theorem is given for Fokker-type variational principles describing directly interacting particles. Many-body as well as two-body interactions depending at most on the particle positions and velocities are considered. Invariance up to a divergence of the action integral under infinitesimal transformations, as usual, leads to divergences that equal linear combinations of the Lagrangian derivatives. Conservation laws can be obtained when the Lagrangian derivatives vanish. The use of the formulation, which is independent of any specific transformations, is illustrated by rederiving the form of the conserved quantities following from the invariance of general two-body Fokker-type variational principles under the infinitesimal transformations of the Lorentz group and of the Galilei group; such conservation laws were previously derived using a method that, although exploiting the symmetries of the action integral, did not directly connect the divergence of the conserved quantities with linear combinations of the Lagrangian derivatives. Other applications and extensions are discussed.

I. INTRODUCTION

Noether's theorem[1] is well known in connection with the generation of conserved quantities that result from the invariance of variational principles under infinitesimal transformations. Detailed formulations exist for field theories,[1–3] where the action is an integral over a volume element, and for Newtonian action-at-a-distance theories,[2–5] where the action is an integral over a universal time parameter. Lacking in the literature, however, is a formulation for theories of directly interacting particles whose equations of motion can be derived from variational principles of the type first introduced by Fokker[6] in electrodyamics and generalized by Havas[4] to include general two-body Galilei- and Lorentz-invariant interactions. In such principles the action involves a sum of multiple integrals over timelike parameters, with the number of parameters equaling the number of particles involved in the interaction. The distinguishing feature of this type of variational principle is that the resulting equations of motion are integrodifferential equations involving many parameters, rather than ordinary or partial differential equations. No general methods exist for solving such equations exactly and their initial value problem is not yet understood.

Noether's theorem is not a statement about conservation laws *per se*, but rather a statement about the existence of linear combinations of Lagrangian derivatives that are divergences; conservation laws are obtained when the equations of motion are satisfied, i.e., when these Lagrangian derivatives vanish. It should be noted, however, that Noether's theorem does not determine the total number of divergence-free (conserved) quantities following from invariance under a group of infinitesimal transformations. It only determines the number of quantities that become divergence-free by virtue of the vanishing of the Lagrangian derivatives.[5]

Although an explicit formulation of Noether's theorem for Fokker-type variational principles is lacking, conservation laws have been derived from invariance properties of such principles under specific transformations. In the case of electrodynamics, Dettman and Schild[7] have derived the ten conservation laws following from the invariance of the action under the infinitesimal transformations of the ten-parameter Lorentz group. Included in their conservation laws was the conservation of energy-momentum, obtained earlier without the aid of symmetry invariance arguments.[6,8] Following closely the method of Dettman and Schild, Havas[4] obtained the conservation laws that follow from the invariance of Fokker-type action principles under the infinitesimal transformations of the Lorentz and Galilei groups; the interactions considered were general two-body interactions depending at most on the four-dimensional positions and velocities of the particles.

It is the purpose of this paper to give, independently of any specific transformations, a formulation of Noether's theorem for Fokker-type variational principles that depend at most on positions and velocities. A formulation for two-body interactions is given in Sec. II and generalized to include n-body interactions in Sec. III. The formulation is used to rederive the conserved quantities associated with the Lorentz and Galilei groups in Sec. IV. This formulation, however, is applicable to all groups, such as the conformal group and conformal extensions of the Galilei group. (These groups are being studied by P. Havas and J. Plebański with the results to be published shortly.) Section V contains a discussion of the results.

II. TWO-BODY INTERACTIONS

Fokker-type variational principles describing a system of N particles with two-body interactions have the form

$$\delta\mathscr{I} = 0, \quad \mathscr{I} \equiv \mathscr{I}_K + \mathscr{I}^{(2)}, \tag{1a}$$

$$\mathscr{I}_K \equiv \sum_i \int_{-\infty}^{\infty} dT_i\, \Lambda_i\,, \tag{1b}$$

$$\mathscr{I}^{(2)} \equiv \sum_{i<j}\sum \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dT_i\, dT_j\, \Lambda_{ij}\,, \tag{1c}$$

 0022-2488/85/112769-08$02.50

where Λ_i depends on the ith particle's coordinates,

$$z_i^\mu(T_i)\equiv(z_i^0, z_i^1, z_i^2, z_i^3) = (t_i(T_i), x_i(T_i), y_i(T_i), z_i(T_i)), \tag{2}$$

and their derivatives with respect to an arbitrary parameter T_i,

$$u_i^\mu \equiv \frac{dz_i^\mu}{dT_i}, \tag{3}$$

while Λ_{ij} depends on the variables of both particle i and particle j; in particular,

$$\Lambda_i = \Lambda_i(z_i^\mu, u_i^\mu),$$
$$\Lambda_{ij} = \Lambda_{ij}(z_i^\mu, u_i^\mu, z_j^\mu, u_j^\mu). \tag{4}$$

The equations of motion, which follow from performing the variation $z_i^\mu \to z_i^\mu + \delta z_i^\mu$, with $\delta z_i^\mu \to 0$ as $T_i \to \pm\infty$, are

$$L_{i\mu} \equiv \mathscr{L}_{i\mu}(\Lambda_i - V_i^{(2)}) = 0, \tag{5}$$

where the Lagrangian operator $\mathscr{L}_{i\mu}$ is defined by

$$\mathscr{L}_{i\mu} \equiv \frac{\partial}{\partial z_i^\mu} - \frac{d}{dT_i}\frac{\partial}{\partial u_i^\mu}, \tag{6}$$

and $V_i^{(2)}$ is the generalized two-body potential

$$V_i^{(2)} \equiv -\sum_{j>i}\int_{-\infty}^{\infty} dT_j\, \Lambda_{ij} - \sum_{j<i}\int_{-\infty}^{\infty} dT_j\, \Lambda_{ji}. \tag{7}$$

It is well known that the Lagrangian operator $\mathscr{L}_{i\mu}$ identically annihilates any function that is a total T_i derivative. Thus, if

$$\overline{\Lambda}_i = \Lambda_i + \frac{dC_i^{(1)}}{dT_i}, \quad \overline{V}_i^{(2)} = V_i^{(2)} + \frac{dC_i^{(2)}}{dT_i}, \tag{8}$$

where $C_i^{(1)}$ and $C_i^{(2)}$ are arbitrary functions of z_i^μ, then $\Lambda_i - V_i^{(2)}$ and $\overline{\Lambda}_i - \overline{V}_i^{(2)}$ both yield the same equations of motion (5). Since integrating $V_i^{(2)}$ in Eq. (7) from $T_i = -\infty$ to $T_i = +\infty$ and summing over all i yields $-2\mathscr{I}^{(2)}$, Eq. (1) can be written as

$$\mathscr{I} = \sum_i \int_{-\infty}^{\infty} dT_i(\Lambda_i - \tfrac{1}{2} V_i^{(2)}). \tag{9}$$

Writing $\overline{\mathscr{I}}$ also in this form, subtracting Eq. (9), and making use of Eq. (8), we obtain

$$\overline{\mathscr{I}} - \mathscr{I} = \sum_i \int_{-\infty}^{\infty} dT_i \frac{dC_i^{(1)}}{dT_i} - \frac{1}{2}\sum_i \int_{-\infty}^{\infty} dT_i \frac{dC_i^{(2)}}{dT_i} = \int_{(-\infty)}^{(\infty)} dC, \tag{10}$$

where

$$C(T_1, T_2, \ldots, T_N) \equiv \sum_i C_i^{(1)}(T_i) - \tfrac{1}{2}\sum_i C_i^{(2)}(T_i), \tag{11}$$

and the parentheses around a limit indicates a complete set of T's, viz. $(T_1, T_2, \ldots, T_N)$. Thus, $\overline{\mathscr{I}}$ and $\mathscr{I}$ differ by the integral of a total differential. More generally, any two action integrals that differ by the integral of a total differential yield the same equations of motion.

We now consider a p-parameter group G_p of infinitesimal transformations

$$\overline{z}_i^\mu = z_i^\mu + \Delta z_i^\mu, \quad \overline{T}_i = T_i + \Delta T_i, \tag{12}$$

where Δz_i^μ and ΔT_i are functions of z_i^μ, u_i^μ, and T_i, depending linearly on the infinitesimal parameters of G_p. Since we are concerned with the effect of these transformations on the functional form of the action integral, we will consider the inexact differential

$$đ\mathscr{I} \equiv \sum_i dT_i(\Lambda_i - \tfrac{1}{2} V_i^{(2)}). \tag{13}$$

Equation (9) can be obtained from (13) by performing the integration

$$\int_{(T^*)}^{(T^{**})} đ\mathscr{I}, \tag{14}$$

followed by letting $T_i^* \to -\infty$, $T_i^{**} \to +\infty$, $i = 1, 2, \ldots, N$.

For infinitesimal transformations that leave $đ\mathscr{I}$ functionally invariant up to a total T differential, we have

$$\Delta đ\mathscr{I} \equiv đ\overline{\mathscr{I}} - đ\mathscr{I} = dC, \tag{15}$$

where C depends linearly on the parameters of G_p. Functional invariance, as usual, means that the $\overline{\Lambda}$'s are the same functions of the $\overline{z}^\mu$'s and $\overline{u}^\mu$'s as the Λ's are of the z^μ's and u^μ's. The equations of motion in the new coordinates will then have the same form as in the old coordinates.

It will be convenient to introduce the variations

$$\delta z_i^\mu \equiv \overline{z}_i^\mu(T_i) - z_i^\mu(T_i) = \Delta z_i^\mu - u_i^\mu\, \Delta T_i, \tag{16}$$

$$\delta u_i^\mu \equiv \overline{u}_i^\mu(T_i) - u_i^\mu(T_i) = \Delta u_i^\mu - \frac{du_i^\mu}{dT_i}\Delta T_i, \tag{17}$$

where

$$\Delta u_i^\mu \equiv \overline{u}_i^\mu(\overline{T}_i) - u_i^\mu(T_i) = \frac{d\Delta z_i^\mu}{dT_i} - u_i^\mu \frac{d\Delta T_i}{dT_i}, \tag{18}$$

Unlike the Δz_i^μ and Δu_i^μ, the variations possess the useful property

$$\delta u_i^\mu = \frac{d\delta z_i^\mu}{dT_i}. \tag{19}$$

In the above notation, Noether's theorem for Fokker-type variational principles states the following: "If $đ\mathscr{I}$ is invariant under the infinitesimal transformations of G_p up to a divergence, there exist precisely p linearly independent combinations of the Lagrangian derivatives $L_{i\mu}$ that are divergences. Conversely, if p such combinations are divergences, there exists a set of p linearly independent infinitesimal transformations that leave $đ\mathscr{I}$ invariant up to a divergence; these transformations generate a G_p provided the δz_i^μ depend at most linearly on the u_i^μ."[9] Here divergence means total T differential.

To obtain the form of these divergences, we begin by writing $\Delta đ\mathscr{I}$ in terms of the variations (16) and (17). We obtain

$$\Delta d\mathscr{I} = \sum_i dT_i\left[\delta\Lambda_i + \frac{d}{dT_i}(\Lambda_i\Delta T_i)\right] + \frac{1}{2}\sum_i dT_i\left\{\sum_{j>i}\int_{-\infty}^{\infty} dT_j\left[\delta\Lambda_{ij} + \frac{d}{dT_i}(\Lambda_{ij}\Delta T_i) + \frac{d}{dT_j}(\Lambda_{ij}\Delta T_j)\right]\right.$$
$$\left. + \sum_{j<i}\int_{-\infty}^{\infty} dT_j\left[\delta\Lambda_{ji} + \frac{d}{dT_i}(\Lambda_{ji}\Delta T_i) + \frac{d}{dT_j}(\Lambda_{ji}\Delta T_j)\right]\right\}, \tag{20}$$

when use is made of the definitions (13) and (7). By making the substitution

$$\frac{1}{2}\frac{d}{dT_i} = \frac{d}{dT_i} - \frac{1}{2}\frac{d}{dT_i} \tag{21}$$

in the double sum, interchanging indices, and rearranging terms, Eq. (20) can be written in the form

$$\Delta d\mathscr{I} = \delta d\mathscr{I} + d\left[\sum_i(\Lambda_i - V_i^{(2)})\Delta T_i\right] + \frac{1}{2}\sum_{i<j}\sum\left(dT_i\int_{-\infty}^{\infty}dT_j - dT_j\int_{-\infty}^{\infty}dT_i\right)\times\left[\frac{d}{dT_j}(\Lambda_{ij}\Delta T_j) - \frac{d}{dT_i}(\Lambda_{ij}\Delta T_i)\right]. \tag{22}$$

The operator identity

$$dT_j\int_{-\infty}^{\infty}dT_i - dT_i\int_{-\infty}^{\infty}dT_j = d\left[\left(\int_{T_i}^{\infty}\int_{-\infty}^{T_j} - \int_{-\infty}^{T_i}\int_{T_j}^{\infty}\right)dT_i dT_j\right] \tag{23}$$

can be used to further simplify this to

$$\Delta d\mathscr{I} = \delta d\mathscr{I} + dF, \tag{24a}$$

$$F = \sum_i(\Lambda_i - V_i^{(2)})\,\Delta T_i + \frac{1}{2}\sum_{i<j}\sum\left(\int_{T_i}^{\infty}\int_{-\infty}^{T_j} - \int_{-\infty}^{T_i}\int_{T_j}^{\infty}\right)dT_i\,dT_j\times\left[\frac{d}{dT_i}(\Lambda_{ij}\Delta T_i) - \frac{d}{dT_j}(\Lambda_{ij}\Delta T_j)\right]. \tag{24b}$$

To evaluate $\delta d\mathscr{I}$ it is useful to define the single particle variation

$$\delta_i \equiv \delta z_i^\mu\frac{\partial}{\partial z_i^\mu} + \delta u_i^\mu\frac{\partial}{\partial u_i^\mu}. \tag{25}$$

Then

$$\delta d\mathscr{I} = \sum_i dT_i\,(\delta_i\Lambda_i) + \frac{1}{2}\sum_i dT_i\left[\sum_{j>i}\int_{-\infty}^{\infty}dT_j\,(\delta_i+\delta_j)\Lambda_{ij} + \sum_{j<i}\int_{-\infty}^{\infty}dT_j(\delta_i+\delta_j)\Lambda_{ji}\right]. \tag{26}$$

The substitution

$$\tfrac{1}{2}\delta_i = \delta_i - \tfrac{1}{2}\delta_i\,, \tag{27}$$

followed by an interchange of indices and rearrangement of terms, yields

$$\delta d\mathscr{I} = \sum_i dT_i\delta_i(\Lambda_i - V_i^{(2)}) + \frac{1}{2}\sum_{i<j}\sum\left(dT_i\int_{-\infty}^{\infty}dT_j - dT_j\int_{-\infty}^{\infty}dT_j\right)\times(\delta_j - \delta_i)\Lambda_{ij}\,. \tag{28}$$

This in turn can be reduced to the form

$$\delta d\mathscr{I} = \sum_i dT_i\,\delta z_i^\mu\,L_{i\mu} + dA, \tag{29a}$$

$$A \equiv \sum_i \delta z_i^\mu\frac{\partial}{\partial u_i^\mu}(\Lambda_i - V_i^{(2)}) + \frac{1}{2}\sum_{i<j}\sum\left(\int_{T_i}^{\infty}\int_{-\infty}^{T_j} - \int_{-\infty}^{T_i}\int_{T_j}^{\infty}\right)\times dT_i dT_j(\delta_i - \delta_j)\Lambda_{ij}\,, \tag{29b}$$

by use of the operator identity (23) and the relation

$$\delta_i = \delta z_i^\mu\,\mathscr{L}_{i\mu} + \frac{d}{dT_i}\left(\delta z_i^\mu\frac{\partial}{\partial u_i^\mu}\right), \tag{30}$$

which follows from Eqs. (6) and (25).

Thus, Eqs. (24) and (29) combined with (15) result in the identity

$$\sum_i dT_i\,\delta z_i^\mu\,L_{i\mu} = dB,$$
$$B \equiv C - F - A = C - \sum_i(\Lambda_i - V_i^{(2)})\,\Delta T_i - \sum_i \delta z_i^\mu\frac{\partial}{\partial u_i^\mu}(\Lambda_i - V_i^{(2)}) - \frac{1}{2}\sum_{i<j}\sum\left(\int_{T_i}^{\infty}\int_{-\infty}^{T_j} - \int_{-\infty}^{T_i}\int_{T_j}^{\infty}\right)dT_i\,dT_j\times\left[(\delta_i - \delta_j)\Lambda_{ij} + \frac{d}{dT_i}(\Lambda_{ij}\Delta T_i) - \frac{d}{dT_j}(\Lambda_{ij}\Delta T_j)\right], \tag{31}$$

which establishes Noether's first theorem. The converse follows by the arguments given in Refs. 1 and 2. When the equations of motion $L_{i\mu} = 0$ are satisfied, the vanishing of the total T differential of B gives rise to p conserved quantities, since B depends linearly on the p independent parameters of the infinitesimal transformation.

III. n-BODY INTERACTIONS

The derivation of the form of the "divergences" following from the application of Noether's theorem to Fokker-type principles with many-body interactions parallels the case of two-body interactions dealt with in Sec. II, except that the total differentials involved are considerably more complicated. In the development that follows, the two-body terms will not be explicitly written since they are the same as those given in Sec. II, while three-body terms will be written out completely to help clarify the notation for the general n-body term.

The n-body variational principle has the form

$$\delta\mathscr{I}=0,\quad \mathscr{I}\equiv\mathscr{I}_K+\mathscr{I}_I,\quad \mathscr{I}_I\equiv\sum_{q=2}^{n}\mathscr{I}^{(q)},\tag{32a}$$

where $\mathscr{I}_K$ is given by Eq. (1b), $\mathscr{I}^{(2)}$ by Eq. (1c), and the general term $\mathscr{I}^{(q)}$ by

$$\mathscr{I}^{(q)}\equiv\sum_{i<j<\cdots<l}\sum\cdots\sum\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}dT_i dT_j\cdots dT_l \times\Lambda_{ij\ldots l}(z_i^\mu,z_j^\mu,\cdots,z_l^\mu,u_i^\mu,u_j^\mu,\cdots,u_l^\mu);\tag{32b}$$

note that the $\Lambda_{ij\ldots l}$ in $\mathscr{I}^{(q)}$ has q indices. The resulting equations of motion have the form

$$\mathscr{L}_{i\mu}(\Lambda_i-V_i)\equiv L_{i\mu}=0,\tag{33}$$

where the generalized potential is now

$$V_i\equiv\sum_{q=2}^{n}V_i^{(q)},\tag{34}$$

with $V_i^{(2)}$ given by Eq. (7), and

$$\begin{aligned}V_i^{(3)}&\equiv-\sum_{j>i}\sum_{k>j}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}dT_j\,dT_k\,\Lambda_{ijk}-\sum_{k<i}\sum_{j>i}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}dT_j\,dT_k\,\Lambda_{kij}-\sum_{j<k}\sum_{k<i}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}dT_j\,dT_k\,\Lambda_{jki},\\ &\ \vdots\\ V_i^{(n)}&\equiv-\sum_{(j\,k\,\cdots\,l)}\sum\cdots\sum\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}dT_j\,dT_k\cdots dT_l\,\Lambda_{(ijk\ldots l)}.\end{aligned}\tag{35}$$

The parentheses around the indices indicates a sum of q such terms consisting of the even permutations of $ijk\cdots l$. The sum in each term is over all indices except i and is subject to the restriction that a given index must be larger than any to the left of that index and smaller than any index to its right; this is because, for example, Λ_{1369} completely describes the four-body contribution to the interaction between particles 1, 3, 6, and 9 and therefore a kernel such as Λ_{3196} is not needed, does not appear in $\mathscr{I}^{(4)}$, and is undefined.

The imperfect differential that gives $\mathscr{I}$ when integrated from $(-\infty)$ to (∞) is

$$đ\mathscr{I}=\sum_i dT_i\Big(\Lambda_i-\sum_{q=2}^{n}\frac{1}{q}V_i^{(q)}\Big).\tag{36}$$

When this is subjected to the infinitesimal transformations (12), there results

$$\begin{aligned}\Delta đ\mathscr{I}=\delta đ\mathscr{I}&+d\Big[\sum_i(\Lambda_i-V_i)\Delta T_i\Big]+\text{two-body terms}\\ &+\frac{1}{3}\sum_{i<j<k}\sum\sum\bigg\{\Big(dT_i\int_{-\infty}^{\infty}dT_j-dT_j\int_{-\infty}^{\infty}dT_i\Big)\int_{-\infty}^{\infty}dT_k\Big[\frac{d}{dT_j}(\Lambda_{ijk}\Delta T_j)-\frac{d}{dT_i}(\Lambda_{ijk}\Delta T_i)\Big]\\ &\qquad+\Big(dT_k\int_{-\infty}^{\infty}dT_i-dT_i\int_{-\infty}^{\infty}dT_k\Big)\int_{-\infty}^{\infty}dT_j\Big[\frac{d}{dT_i}(\Lambda_{ijk}\Delta T_i)-\frac{d}{dT_k}(\Lambda_{ijk}\Delta T_k)\Big]\\ &\qquad+\Big(dT_j\int_{-\infty}^{\infty}dT_k-dT_k\int_{-\infty}^{\infty}dT_j\Big)\int_{-\infty}^{\infty}dT_i\Big[\frac{d}{dT_k}(\Lambda_{ijk}\Delta T_k)-\frac{d}{dT_j}(\Lambda_{ijk}\Delta T_j)\Big]\bigg\}\\ &+\cdots+n\text{-body terms},\end{aligned}\tag{37}$$

where, for the three-body terms, the identity

$$\frac{1}{3}\frac{d}{dT_i}=\frac{d}{dT_i}-\frac{1}{3}\frac{d}{dT_i}-\frac{1}{3}\frac{d}{dT_i}\tag{38}$$

has been used to split off terms making up $d(\sum_i V_i^{(3)}\Delta T_i)$, indices have been interchanged and the terms with common integrands grouped together; similar operations were performed on the other many-body terms. Using the operator identity (23) then gives

$$\Delta đ\mathscr{I}=\delta đ\mathscr{I}+dF,\tag{39a}$$

$$\begin{aligned}F\equiv&\sum_i(\Lambda_i-V_i)\Delta T_i+\text{two-body terms}\\ &+\frac{1}{3}\sum_{i<j<k}\sum\sum\bigg\{\Big(\int_{T_i}^{\infty}\int_{-\infty}^{T_j}-\int_{-\infty}^{T_i}\int_{T_j}^{\infty}\Big)dT_i\,dT_j\int_{-\infty}^{\infty}dT_k\Big[\frac{d}{dT_i}(\Lambda_{ijk}\Delta T_i)-\frac{d}{dT_j}(\Lambda_{ijk}\Delta T_j)\Big]\end{aligned}$$

$$+\left(\int_{T_j}^{\infty}\int_{-\infty}^{T_k}-\int_{-\infty}^{T_j}\int_{T_k}^{\infty}\right)dT_j\,dT_k\int_{-\infty}^{\infty}dT_i\left[\frac{d}{dT_j}(\Lambda_{ijk}\Delta T_j)-\frac{d}{dT_k}(\Lambda_{ijk}\Delta T_k)\right]$$

$$+\left(\int_{T_k}^{\infty}\int_{-\infty}^{T_i}-\int_{-\infty}^{T_k}\int_{T_i}^{\infty}\right)dT_k dT_i\int_{-\infty}^{\infty}dT_j\left[\frac{d}{dT_k}(\Lambda_{ijk}\Delta T_k)-\frac{d}{dT_i}(\Lambda_{ijk}\Delta T_i)\right]\Bigg\}$$

$$+\cdots+\frac{1}{n}\sum_{i<j<k<\cdots<l}\sum\sum\cdots\sum\left\{\left(\int_{T_i}^{\infty}\int_{-\infty}^{T_j}-\int_{-\infty}^{T_i}\int_{T_j}^{\infty}\right)dT_i\,dT_j\int_{-\infty}^{\infty}dT_k\cdots\int_{-\infty}^{\infty}dT_l\right.$$

$$\left.\times\left[\frac{d}{dT_i}(\Delta T_i-\frac{d}{dT_j}(\Delta T_j\right]\right\}_c\Lambda_{ijk\ldots l}). \tag{39b}$$

The curly brackets with the subscript c in an n-body term mean that the quantity enclosed is a sum of n parameters, each of which is an even cyclic permutation of the indicated operator; note that the T derivatives act on the product of $\Lambda_{ijk\ldots l}$ and the appropriate ΔT.

To evaluate $\delta d\mathscr{I}$, the identity

$$\frac{1}{n}\delta=\delta_i+\frac{1}{n}(\delta_j-\delta_i+\delta_k-\delta_i+\cdots+\delta_l-\delta_i), \tag{40}$$

which is valid when δ acts on an n-particle function, is used to split off terms contributing to $V_i^{(n)}$. An interchange of indices followed by a regrouping of terms yields

$$\delta d\mathscr{I}=\sum_i dT_i\delta_i(\Lambda_i-V_i)+\text{two-body terms}$$

$$+\frac{1}{3}\sum_{i<j<k}\sum\sum\left\{\left(dT_i\int_{-\infty}^{\infty}dT_j-dT_j\int_{-\infty}^{\infty}dT_i\right)\int_{-\infty}^{\infty}dT_k(\delta_j-\delta_i)\Lambda_{ijk}\right.$$

$$+\left(dT_j\int_{-\infty}^{\infty}dT_k-dT_k\int_{-\infty}^{\infty}dT_j\right)\int_{-\infty}^{\infty}dT_i(\delta_k-\delta_j)\Lambda_{ijk}$$

$$\left.+\left(dT_k\int_{-\infty}^{\infty}dT_i-dT_i\int_{-\infty}^{\infty}dT_k\right)\int_{-\infty}^{\infty}dT_j(\delta_i-\delta_k)\Lambda_{ijk}\right\}$$

$$+\cdots+\frac{1}{n}\sum_{i<j<k<\cdots<l}\sum\sum\cdots\sum\left\{\left(dT_i\int_{-\infty}^{\infty}dT_j-dT_j\int_{-\infty}^{\infty}dT_i\right)\int_{-\infty}^{\infty}dT_k\cdots\int_{-\infty}^{\infty}dT_l(\delta_j-\delta_i)\right\}_c\Lambda_{ijk\ldots l}. \tag{41}$$

Use of the operator identity (23) then gives

$$\delta d\mathscr{I}=\sum_i dT_i\delta z_i^\mu L_{i\mu}+dA, \tag{42a}$$

$$A\equiv\sum_i\delta z_i^\mu\frac{\partial}{\partial u_i^\mu}(\Lambda_i-V_i)+\text{two-body terms} \tag{42b}$$

$$+\frac{1}{3}\sum_{i<j<k}\sum\sum\left\{\left(\int_{T_i}^{\infty}\int_{-\infty}^{T_j}-\int_{-\infty}^{T_i}\int_{T_j}^{\infty}\right)dT_i\,dT_j\int_{-\infty}^{\infty}dT_k(\delta_i-\delta_j)\right\}_c\Lambda_{ijk}$$

$$+\cdots+\frac{1}{n}\sum_{i<j<k<\cdots<l}\sum\sum\cdots\sum\left\{\left(\int_{T_i}^{\infty}\int_{-\infty}^{T_j}-\int_{-\infty}^{T_i}\int_{T_j}^{\infty}\right)dT_i\,dT_j\int_{-\infty}^{\infty}dT_k\cdots\int_{-\infty}^{\infty}dT_l\,(\delta_i-\delta_j)\right\}_c\Lambda_{ijk\ldots l}.$$

Thus, if $d\mathscr{I}$ is invariant up to a total differential dC under the infinitesimal transformation (12), Eqs. (39) and (42) give

$$\sum_i dT_i\,\delta z_i^\mu L_{i\mu}=dB,\quad B\equiv C-A-F. \tag{43}$$

As in the two-body case, conservation laws result when the equations of motion $L_{i\mu}=0$ are satisfied.

IV. GALILEI- AND LORENTZ-INVARIANT VARIATIONAL PRINCIPLES

The conservation laws resulting from the invariance of Fokker-type variational principles under the infinitesimal transformations of both the Galilei and Lorentz groups have been obtained by Havas,[4] who used the method of Dettman and Schild[7] for each subgroup (translations, rotations, boosts) separately. Here we obtain the same conserved quantities using the explicit formulation of Noether's theorem developed in Sec. II for two-body interactions. In this section, equation numbers with a subscript L or G identify equations that apply only to the Lorentz case or only to the Galilei case, respectively; equations numbered with no subscript apply to both cases.

In the Lorentz case, we choose the nonsingular metric tensor $\eta_{\mu\nu}$ and its inverse $\eta^{\mu\nu}$ to be

$$\eta_{\mu\nu} \equiv \mathrm{diag}\,(1, -1/c^2, -1/c^2, -1/c^2), \tag{44$_\mathrm{L}$}$$

$$\eta^{\mu\nu} \equiv \mathrm{diag}\,(1, -c^2, -c^2, -c^2), \tag{45$_\mathrm{L}$}$$

which satisfy

$$\eta_{\mu\rho}\eta^{\rho\nu} = \delta^{\nu}_{\mu}\,. \tag{46$_\mathrm{L}$}$$

In the Galilei case, the metric tensor is singular[10] and, following Ref. 4, can be chosen as

$$g_{\mu\nu} \equiv \mathrm{diag}\,(1,0,0,0). \tag{44$_\mathrm{G}$}$$

In addition, it is useful to define another tensor

$$h^{\mu\nu} \equiv \mathrm{diag}\,(0, -1, -1, -1). \tag{45$_\mathrm{G}$}$$

The tensors $g_{\mu\nu}$ and $h^{\mu\nu}$ correspond to the limits as $c \to \infty$ of $\eta_{\mu\nu}$ and $\eta^{\mu\nu}/c^2$ and satisfy

$$g_{\mu\rho}h^{\rho\nu} = 0. \tag{46$_\mathrm{G}$}$$

Introducing the notation

$$G_{\mu\nu} = \eta_{\mu\nu}, \tag{47$_\mathrm{L}$}$$

$$G_{\mu\nu} = g_{\mu\nu}\,, \tag{47$_\mathrm{G}$}$$

the proper time τ_i of the ith particle is defined by

$$d\tau_i \equiv (G_{\mu\nu}\, dz_i^\mu\, dz_i^\nu)^{1/2}, \tag{48}$$

for both cases. Note that in the Galilei case, the proper time τ_i equals the coordinate time z_i^0 up to an additive constant. If the proper time is chosen to parametrize the particles' worldlines (i.e., $T_i \to \tau_i$), then it follows from Eq. (48) that

$$G_{\mu\nu}\, v_i^\mu v_i^\nu = 1, \quad G_{\mu\nu}\, v_i^\mu a_i^\nu = 0, \tag{49}$$

where

$$v_i^\mu \equiv \frac{dz_i^\mu}{d\tau_i}, \quad a_i^\mu \equiv \frac{dv_i^\mu}{d\tau_i}. \tag{50}$$

Unlike the Lorentz case, in the Galilei case it is not possible to lower and raise indices reversibly with $g_{\mu\nu}$ and $h^{\mu\nu}$. Clearly since

$$g_{\mu\nu}a^\nu = 0\,, \tag{51$_\mathrm{G}$}$$

it is not possible to regain a^σ by raising with $h^{\mu\sigma}$. On the other hand, one can introduce a covariant vector

$$a_{i\mu} \equiv (\mathbf{a}_i \cdot \mathbf{v}_i, -\mathbf{a}_i), \tag{52$_\mathrm{G}$}$$

which has the property that contraction with $h^{\mu\nu}$ yields a_i^ν; here $\mathbf{v}_i$ and $\mathbf{a}_i$ are the ordinary three-velocity and three-acceleration, respectively. If we also define

$$\mathfrak{w}_{i\mu} \equiv \eta_{\mu\nu}\, v_i^\nu = v_{i\mu}\,, \tag{53$_\mathrm{L}$}$$

$$\mathfrak{w}_{i\mu} \equiv (\tfrac{1}{2}\,\mathbf{v}_i^2, -\mathbf{v}_i), \tag{53$_\mathrm{G}$}$$

then for both cases,

$$a_{i\mu} = \frac{d\mathfrak{w}_{i\mu}}{d\tau_i}. \tag{54}$$

In the Galilei case, however, $\mathfrak{w}_{i\mu}$ does not transform like a covariant vector under Galilei velocity boosts and v_i^μ cannot be obtained from $\mathfrak{w}_{i\mu}$ by raising indices with $h^{\mu\nu}$.

The variational principles for the Lorentz and Galilei cases can both be written in the parameter-invariant form

$$\delta\mathscr{I} = 0, \quad \mathscr{I} \equiv \mathscr{I}_K + \mathscr{I}^{(2)}, \tag{55a}$$

$$\mathscr{I}_K \equiv -\sum_i \int_{-\infty}^{\infty} dT_i\, \frac{m_i \mathfrak{v}_{i\rho} u_i^\rho}{u_i}, \tag{55b}$$

$$\mathscr{I}^{(2)} \equiv -\sum_{i<j}\sum \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dT_i\, dT_j \times u_i u_j U_{ij}\left(z_i^\mu, z_j^\mu, \frac{u_i^\mu}{u_i}, \frac{u_j^\mu}{u_j}\right), \tag{55c}$$

where m_i is the mass of particle i,

$$u_i \equiv (G_{\mu\nu}\, u_i^\mu u_i^\nu)^{1/2}, \tag{56}$$

and

$$\mathfrak{v}_{i\rho} \equiv \eta_{\rho\sigma}\, u_i^\sigma = u_{i\rho}\,, \tag{57$_\mathrm{L}$}$$

$$\mathfrak{v}_{i\rho} \equiv \left(\frac{\mathbf{u}_i^2}{2u_i}, -\mathbf{u}_i\right), \quad \mathbf{u}_i = (u_i^1, u_i^2, u_i^3)\,. \tag{57$_\mathrm{G}$}$$

The equations of motion are obtained by performing the variation with arbitrary parametrization T_i; proper time parametrization ($T_i \to \tau_i$) can be chosen after the variation is completed.[12]

Comparison of Eqs. (1) and (55) yields the indentifications

$$\Lambda_i = -m_i \mathfrak{v}_{i\rho}\, u_i^\rho/u_i, \quad \Lambda_{ij} = -u_i u_j U_{ij}\,. \tag{58}$$

Note that in the Galilei case, $u_i = u_i^0$, and using Eq. (57$_\mathrm{G}$),

$$\Lambda_i = +\frac{1}{2} m_i \frac{\mathbf{u}_i^2}{u_i^0}, \tag{59$_\mathrm{G}$}$$

which is just the Newtonian kinetic energy when $T_i \to \tau_i$ (in which case $u_i^0 \to 1$). From Eqs. (59$_\mathrm{G}$) and (57$_\mathrm{G}$), it follows that

$$\frac{\partial \Lambda_i}{\partial u_i^\mu} = -m_i \frac{\mathfrak{v}_{i\mu}}{u_i} \tag{60$_\mathrm{G}$}$$

holds for the Galilei case just as for the Lorentz case.

To simplify the subsequent calculations, we will make use of the following relations, valid for an arbitrary parameter-invariant function $f(z_i^\mu, u_i^\mu/u_i, \ldots)$:

$$\frac{\partial f}{\partial u_i^\mu} = \frac{\partial f}{\partial (u_i^\rho/u_i)}\left[\frac{\delta_\mu^\rho}{u_i} - \frac{G_{\mu\sigma} u_i^\sigma u_i^\rho}{(u_i)^3}\right], \tag{61a}$$

$$u_i^\mu \frac{\partial f}{\partial u_i^\mu} = 0, \quad u_i^\mu \frac{\partial (u_i f)}{\partial u_i^\mu} = u_i f. \tag{61b}$$

With the aid of these relations and the definitions (16) and (17) of δz_i^μ and δu_i^μ, it can then be shown that

$$\delta_i(u_i f) = \Delta z_i^\mu \frac{\partial(u_i f)}{\partial z_i^\mu} + \frac{d(\Delta z_i^\mu)}{dT_i}\frac{\partial(u_i f)}{\partial u_i^\mu} - \frac{d}{dT_i}(u_i f \Delta T_i). \quad (62)$$

This expression together with Eqs. (61) and the identifications (58) turn Eq. (31) into

$$\sum_i dT_i\, \delta z_i^\mu L_{i\mu} = dB,$$

$$\begin{aligned} B = C &+ \sum_i \Delta z_i^\mu \left[\frac{m_i \mathfrak{v}_{i\mu}}{u_i} + \frac{\partial V_i^{(2)}}{\partial(u_i^\mu/u_i)} + G_{\mu\rho}\frac{u_i^\rho}{u_i}\left(V_i^{(2)} - \frac{u_i^\sigma}{u_i}\frac{\partial V_i^{(2)}}{\partial(u_i^\sigma/u_i)}\right)\right] \\ &+ \frac{1}{2}\sum_{i<j}\sum\left(\int_{T_i}^{\infty}\int_{-\infty}^{T_j} - \int_{-\infty}^{T_i}\int_{T_j}^{\infty}\right) dT_i\, dT_j \\ &\times\left\{\Delta z_i^\mu \frac{\partial}{\partial z_i^\mu}(u_i u_j U_{ij}) - \Delta z_j^\mu \frac{\partial}{\partial z_j^\mu}(u_i u_j U_{ij}) \right. \\ &+ u_j \frac{d\Delta z_i^\mu}{dT_i}\left[\frac{\partial U_{ij}}{\partial(u_i^\mu/u_i)} + G_{\mu\rho}\frac{u_i^\rho}{u_i}\left(U_{ij} - \frac{u_i^\sigma}{u_i}\frac{\partial U_{ij}}{\partial(u_i^\sigma/u_i)}\right)\right] \\ &\left. - u_i \frac{d\Delta z_j^\mu}{dT_j}\left[\frac{\partial U_{ij}}{\partial(u_j^\mu/u_j)} + G_{\mu\rho}\frac{u_j^\rho}{u_j}\left(U_{ij} - \frac{u_j^\sigma}{u_j}\frac{\partial U_{ij}}{\partial(u_j^\sigma/u_j)}\right)\right]\right\}, \end{aligned} \quad (63)$$

where, in parameter-invariant form, the potential $V_i^{(2)}$ defined in Eq. (7) goes over into $u_i V_i^{(2)}$, so that

$$V_i^{(2)} = \sum_{j>i}\int_{-\infty}^{\infty} dT_j\, u_j U_{ij} + \sum_{j<i}\int_{-\infty}^{\infty} dT_j\, u_j U_{ji}. \quad (64)$$

We can now obtain the conservation laws following from invariance (or invariance up to a divergence) under the Lorentz and Galilei groups. We here assume that the interaction U_{ij} is invariant under the respective groups; U_{ij} can be made manifestly invariant by assuming dependence on the independent two-body invariants given in Ref. 4.

Invariance of the action (55) under the infinitesimal space-time translations

$$\Delta z_i^\mu = \epsilon^\mu \quad (65)$$

leads to a law of conservation of energy-momentum

$$dP_\mu(\tau_1,\tau_2,\ldots,\tau_N) = 0,$$

$$\begin{aligned} P_\mu \equiv \sum_i &\left[m_i \mathfrak{w}_{i\mu} + \frac{\partial V_i^{(2)}}{\partial v_i^\mu} + G_{\mu\sigma} v_i^\sigma\left(V_i^{(2)} - v_i^\rho \frac{\partial V_i^{(2)}}{\partial v_i^\rho}\right)\right] \\ &+ \sum_{i<j}\sum\left(\int_{\tau_i}^{\infty}\int_{-\infty}^{\tau_j} - \int_{-\infty}^{\tau_i}\int_{\tau_j}^{\infty}\right) d\tau_i\, d\tau_j \frac{\partial U_{ij}}{\partial s_{ij}^\mu}, \end{aligned} \quad (66)$$

where

$$s_{ij}^\mu \equiv z_i^\mu - z_j^\mu, \quad (67)$$

we have put $B = P_\mu \epsilon^\mu$, and proper-time parametrization has been chosen.

The infinitesimal spatial rotations and velocity boosts for both groups can be expressed in the form

$$\Delta z_i^\rho = \mathrm{H}^{\rho\mu}\epsilon_{\mu\nu}\, z_i^\nu, \quad \epsilon_{\mu\nu} = -\epsilon_{\nu\mu}, \quad (68)$$

where the $\epsilon_{\mu\nu}$ are arbitrary constants, and

$$H^{\mu\nu} \equiv \eta^{\mu\nu}, \quad (69_{\mathrm{L}})$$

$$H^{\mu\nu} \equiv h^{\mu\nu}. \quad (69_{\mathrm{G}})$$

In the Lorentz case, both $\mathscr{I}_K$ and, by assumption, $\mathscr{I}^{(2)}$ are invariant under the infinitesimal transformations (68). In the Galilei case, however, $\mathscr{I}_K$ is not invariant since $\mathfrak{v}_{i\rho}$ is not a covariant vector. Instead, we have

$$\Delta d\mathscr{I} = \Delta d\mathscr{I}_K = dC,$$

$$C = \sum_i m_i \epsilon_{\mu 0}\, z_i^\mu. \quad (70)$$

In order to treat both cases at once, it is helpful to put C in the four-dimensional form

$$C = \sum_i m_i\left[\epsilon_{\mu\nu}\frac{u_i^\mu}{u_i} z_i^\nu - \epsilon_{\mu\nu} H^{\mu\rho}\frac{\mathfrak{v}_{i\rho}}{u_i} z_i^\nu\right], \quad (71)$$

which vanishes in the Lorentz case, and reduces to the form (70) in the Galilei case.

Using Eqs. (71) and (68) in (63), with $L_{i\mu} = 0$, we obtain

$$dB = 0,$$

$$\begin{aligned} B = \sum_i &m_i\left[\epsilon_{\mu\nu}\, v_i^\mu z_i^\nu - \epsilon_{\mu\nu} H^{\mu\rho}\mathfrak{w}_{i\rho} z_i^\nu\right] \\ &+ \sum_i H^{\mu\rho}\epsilon_{\rho\sigma}\, z_i^\sigma\left[m_i \mathfrak{w}_{i\mu} + \frac{\partial V_i^{(2)}}{\partial v_i^\mu} + G_{\mu\nu}\, v_i^\nu\left(V_i^{(2)} - v_i^\lambda \frac{\partial V_i^{(2)}}{\partial v_i^\lambda}\right)\right] \\ &+ \frac{1}{2}\sum_{i<j}\sum\left(\int_{\tau_i}^{\infty}\int_{-\infty}^{\tau_j} - \int_{-\infty}^{\tau_i}\int_{\tau_j}^{\infty}\right) d\tau_i\, d\tau_j \\ &\times\left\{H^{\mu\rho}\epsilon_{\rho\sigma}\, z_i^\sigma \frac{\partial U_{ij}}{\partial z_i^\mu} - H^{\mu\rho}\epsilon_{\rho\sigma}\, z_j^\sigma \frac{\partial U_{ij}}{\partial z_j^\mu} \right. \\ &+ H^{\mu\rho}\epsilon_{\rho\sigma} v_i^\sigma\left[G_{\mu\lambda} v_i^\lambda\left(U_{ij} - v_i^\nu \frac{\partial U_{ij}}{\partial v_i^\nu}\right) + \frac{\partial U_{ij}}{\partial v_i^\mu}\right] \\ &\left. - H^{\mu\rho}\epsilon_{\rho\sigma} v_j^\sigma\left[G_{\mu\lambda} v_j^\lambda\left(U_{ij} - v_j^\nu \frac{\partial U_{ij}}{\partial v_j^\nu}\right) + \frac{\partial U_{ij}}{\partial v_j^\mu}\right]\right\}. \end{aligned} \quad (72)$$

In the Galilei case $H^{\mu\rho} G_{\mu\lambda} = 0$, whereas in the Lorentz case $H^{\mu\rho}\epsilon_{\rho\sigma} v_j^\sigma G_{\mu\lambda} v_j^\lambda = \epsilon_{\rho\sigma} v_j^\rho v_j^\sigma = 0$, by antisymmetry of $\epsilon_{\rho\sigma}$. Then defining $B = L^{\mu\nu}\epsilon_{\mu\nu}/2$, and simplifying, Eq. (72) reduces to the law of conservation of angular momentum and the center-of-mass theorem

$$dL^{\mu\nu}(\tau_1,\tau_2,\ldots,\tau_N)=0, \tag{73}$$

$$\begin{aligned}
L^{\mu\nu} = \sum_i \Bigg\{ \Bigg[m_i v_i^\mu + \Bigg(\frac{\partial V_i^{(2)}}{\partial v_i^\rho} &+ G_{\rho\lambda} v_i^\lambda \Bigg\{ V_i^{(2)} - v_i^\sigma \frac{\partial V_i^{(2)}}{\partial v_i^\sigma} \Bigg\} \Bigg) H^{\rho\mu} \Bigg] z_i^\nu \\
- \Bigg[m_i v_i^\nu + \Bigg(\frac{\partial V_i^{(2)}}{\partial v_i^\rho} &+ G_{\rho\lambda} v_i^\lambda \Bigg\{ V_i^{(2)} - v_i^\sigma \frac{\partial V_i^{(2)}}{\partial v_i^\sigma} \Bigg\} \Bigg) H^{\rho\nu} \Bigg] z_i^\mu \Bigg\} \\
&+ \frac{1}{2} \sum_{i<j}\sum \Bigg(\int_{\tau_i}^{\infty} \int_{-\infty}^{\tau_j} - \int_{-\infty}^{\tau_i} \int_{\tau_j}^{\infty} \Bigg) d\tau_i \, d\tau_j \\
&\times \Bigg[H^{\lambda\mu} (z_i^\nu + z_j^\nu) \frac{\partial U_{ij}}{\partial s_{ij}^\lambda} - H^{\lambda\nu} (z_i^\mu + z_j^\mu) \frac{\partial U_{ij}}{\partial s_{ij}^\lambda} \\
&+ \frac{\partial U_{ij}}{\partial v_i^\lambda} (H^{\lambda\mu} v_i^\nu - H^{\lambda\nu} v_i^\mu) \\
&- \frac{\partial U_{ij}}{\partial v_j^\lambda} (H^{\lambda\mu} v_j^\nu - H^{\lambda\nu} v_j^\mu) \Bigg] .
\end{aligned}$$

V. DISCUSSION

In Sec. II a formulation of Noether's theorem was obtained for Fokker-type variational principles depending at most on the positions and velocities of particles interacting via direct two-body interactions; the generalization to n-body interactions was presented in Sec. III. Equation (31) gives the form of the divergence (here a total T differential) that equals a linear combination of Lagrangian derivatives for the two-body case, and Eq. (43) gives the corresponding result for the n-body case. As usual, conservation laws can be obtained from these equations when the Lagrangian derivatives vanish. The formulation is independent of any specific transformations. In the past,[4,7] conservation laws for Fokker-type variational principles were derived from the invariance properties of the action integral by choosing specific transformations and applying the method of Dettman and Schild.[7] This method requires a lengthy calculation for each transformation and does not in any obvious way indicate the connection of the "divergence" of the resulting conserved quantities to a linear combination of the Lagrangian derivatives—a connection which lies at the heart of Noether's theorem.

To illustrate the use of the formulation given here, the form of the conserved quantities following from the invariance of two-body Fokker-type variational principles under the infinitesimal transformations of the Lorentz group and of the Galilei group were obtained in Sec. IV. These conservation laws are in agreement with those obtained earlier by Havas,[4] who used the method of Dettman and Schild but did not present the details of the calculations due to their lengthy nature. While in the Lorentz case no direct connection between the conservation laws and Noether's theorem has heretofore been demonstrated, it is worth mentioning that the approximately relativistic conserved quantities obtained by expanding these exact quantities in powers of $1/c$ to order $1/c^2$ have been shown to be obtainable by applying Noether's theorem (formulated for a single independent time parameter) to the approximately relativistic Lagrangians derived from the exact variational principles by a similar expansion in powers of $1/c$ to the same order.[14]

As mentioned earlier, the formulation is not restricted to any particular group and can be used, for example, for the conformal group and conformal extensions[15] of the Galilei group, which are currently under investigation by P. Havas and J. Plebański. The formulation of Noether's theorem given here can be extended to include additional dependent variables, such as the classical analog of spin; the result of such an extension has already been used for the classical analog of isospin.[16] The formulation can also readily be generalized to include derivatives of arbitrary order.

ACKNOWLEDGMENTS

Part of this paper was written while on a Sabbatical leave granted by Spring Garden College. I am indebted to Professor Peter Havas (Temple University) for helpful discussions on this work and suggestions on the manuscript.

[1]E. Noether, Nachr. Akad. Wiss. Göttingen, Math.-Phys. Kl., 235 (1918) [English translation in Transport Theory Statistical Phys. **1**, 186 (1971)].

[2]E. Bessel-Hagen, Math. Ann. **84**, 258 (1921). Noether's theorem as stated in this paper is slightly more general than that given in Ref. 1; it is also credited to Noether. Following Refs. 4 and 5, by "Noether's theorem" we shall mean the generalized version.

[3]E. L. Hill, Rev. Mod. Phys. **23**, 253 (1951).

[4]P. Havas, in *Problems in the Foundations of Physics*, edited by M. Bunge (Springer, New York, 1971), p. 31. The following misprints in Eq. (60) of this article have been communicated by P. Havas: the two occurrences of s_{ij}^μ in the next to last line should be replaced by $z_i^\mu + z_j^\mu$; the plus sign in the last line should be a minus sign; and the factor $\frac{1}{2} g_i g_j$ should be inserted just inside the double sum.

[5]P. Havas and J. Stachel, Phys. Rev. **185**, 1636 (1969).

[6]A. D. Fokker, Z. Phys. **58**, 386 (1929). The generalizations of this action principle as given in Ref. 4 will here be referred to as Fokker-type principles.

[7]J. W. Dettman and A. Schild, Phys. Rev. **95**, 1057 (1954).

[8]J. A. Wheeler and R. P. Feynman, Rev. Mod. Phys. **21**, 425 (1949). These authors also discuss this energy-momentum tensor in the context of an "adjunct" field theory. An alternative form of this tensor in such a theory was given by J. Frenkel, Z. Phys. **58**, 518 (1925); see also J. L. Synge, Trans. R. Soc. Can. **34**, 1 (1940).

[9]The part pertaining to the converse in the statement of Noether's first theorem given in Refs. 4 and 5 is correct according to the notation of Ref. 2, where δx represents the substantial variation $\Delta x - v\Delta t$, but should have δx replaced by $\delta x - v\delta t$ in their notation.

[10]For a discussion of the four-dimensional representation of the Galilei case, see Refs. 11 and 4.

[11]P. Havas, Rev. Mod. Phys. **36**, 938 (1964).

[12]An alternative to a parameter-invariant approach is to use proper time parametrization and introduce Lagrange multipliers treating the first of Eqs. (46) as constraints. See, for example, Ref. 4 and 13.

[13]H. W. Woodcock and P. Havas, Phys. Rev. D **6**, 3422 (1972).

[14]H. W. Woodcock, Ph. D. thesis, Temple University, 1972 (unpublished), and Ref. 13. A similar result has been shown to hold for the generalization of these variational principles to include classical isospin in W. Herman, Ph. D. thesis, Temple University, 1976 (unpublished), and Ref. 16.

[15]For the most general conformal extension of the Galilei group, see P. Havas and J. Plebański, J. Math. Phys. **19**, 482 (1978).

[16]W. N. Herman and P. Havas, Phys. Rev. D **17**, 1985 (1978).

On nonlocal point interactions in one, two, and three dimensions

L. Dabrowski[a)]
SISSA, International School for Advanced Studies, I-34014 Trieste, Italy
H. Grosse
Institute für Theoretische Physik, Universität Wien, A-1090 Vienna, Austria

(Received 30 January 1985; accepted for publication 26 April 1985)

Three characterizations of all self-adjoint extensions of the Laplacian in one, two, and three dimensions are discussed.

I. INTRODUCTION

Starting from the self-adjoint (sa) d-dimensional Laplacian

$$-\Delta = -\sum_{j=1}^{d} \frac{\partial^2}{\partial y_j^2}, \quad d = 1,2,3, \tag{1.1}$$

defined on the dense domain

$$D = \left\{ \psi \in \mathscr{H} \,\middle|\, \frac{\partial \psi}{\partial y_j} \text{ absolutely continuous}, \quad \Delta\psi \in \mathscr{H} \right\} \tag{1.2}$$

in $\mathscr{H} = L^2(\mathbf{R}^d, d^d y)$, we shall restrict $-\Delta$ to the set of functions vanishing at n different $\mathbf{x}_j \in \mathbf{R}^d$, $j = 1, \ldots, n$, and obtain a closed symmetric operator H, which is given in a momentum representation by $H = \mathbf{p}^2$ defined on the domain

$$D(H) = \left\{ \psi \in L^2(\mathbf{R}^d, d^d p) \,\middle|\, \int d^d p \, |\mathbf{p}^2 \psi(\mathbf{p})|^2 < \infty, \right.$$

$$\left. \int d^d p \, \psi(\mathbf{p}) e^{i\mathbf{p}\mathbf{x}_j} = 0, \quad j = 1, \ldots, n \right\} \tag{1.3}$$

and related to the Laplacian $-\Delta$ via Fourier transformation.

From the general theory of operator extensions it is well known that H with deficiency indices (n,n) admits an n^2-parameter family of sa extensions. A particular n-parameter subfamily of extensions corresponding to local δ-like point interactions is discussed in the literature[1–11]

In this paper we investigate to which situation the other extensions correspond. In Sec. II all extensions of H for $d = 1,2,3$ are given, following the general mathematical theory[12,13]. In Sec. III it is shown that these extensions can be obtained as the norm resolvent limit of separable potentials. In the one-dimensional case the connection to nonlocal δ-like interactions, which can be defined by special boundary conditions [see Eq. (3.12)] is indicated. Finally in Sec. IV a third characterization following Ref. 1 is given. It is shown that scaled separable potentials converge in the norm resolvent sense. Like in the case of local potentials one has to distinguish between the cases where there are zero energy resonance states present or where there are not.

II. EXTENSIONS OF SYMMETRIC OPERATORS

Here we give a short account of theory of sa extensions of a closed symmetric operator and apply the abstract construction to the operator H. In the general theory one makes use of the Cayley transform, which provides a correspondence between symmetric and isometric operators.

Let A be a closed symmetric operator and z a nonreal number. The *Cayley transform* V of A is then defined on $D(V) = \mathrm{Ran}(A - \bar{z})$ by

$$Vf = -(A - z)(A - \bar{z})^{-1} f, \quad \forall f \in D(V), \tag{2.1}$$

and is isometric. The A can be recovered from V by

$$Ah = (z + \bar{z}V)(1 + V)^{-1} h, \quad \forall h \in D(A). \tag{2.2}$$

The *deficiency indices* (m,n) of A are given by

$$m = \dim N_{z_1}, \quad n = \dim N_{z_2}, \quad N_{z_i} = \mathrm{Ran}(A - z_i)^{\perp}, \tag{2.3}$$

where $\mathrm{Im}\, z_1 < 0$ and $\mathrm{Im}\, z_2 > 0$; these for V are similarly defined with $|z_1| < 1$ and $|z_2| > 1$. The deficiency indices do not depend on the chosen point in the appropriate half-planes and are identical for A and V.

If $\tilde{V}$ is an isometric extension of V then $\tilde{V}$ maps a subspace of $D(V)^{\perp} = H \ominus D(V)$ of the Hilbert space $\mathscr{H}$ onto a subspace of $(\mathrm{Ran}\, V)^{\perp} = \mathscr{H} \ominus \mathrm{Ran}(V)$ of the same dimension. The V is maximal (i.e., has no proper extensions) iff $\mathrm{Min}(m,n) = 0$, unitary iff $m = n = 0$, and can be extended to a unitary operator $\tilde{V}$ iff $m = n$; then $\tilde{V} = V \oplus U$ with U: $D(V)^{\perp} \to \mathrm{Ran}(V)^{\perp}$. The inverse Cayley transform (2.2) gives then all extensions $\tilde{A}$ of A.

It follows that all sa extensions are parametrized by a family of unitary operators U. The domain of an sa extension A^U of A consists of vectors f:

$$f = f_0 + g_z + Ug_z, \quad f_0 \in D(A), \quad g_z \in N_{\bar{z}}, \quad \mathrm{Im}\, z > 0, \tag{2.4}$$

where U is a unitary mapping U: $N_{\bar{z}} \to N_z$; the action of A^U is given by

$$A^U f = A f_0 + z g_z + \bar{z} U g_z. \tag{2.5}$$

Let R_z^U and R_z^W be the resolvents of two sa extensions A^U and A^W of A. Denote by $g_z^k, k = 1, \ldots, n$ a basis of $N_{\bar{z}}$. The difference of the two resolvents satisfies *Krein's formula*

$$R_z^U = R_z^W + \sum_{k,l=1}^{n} g_z^k M_{kl}(z) \langle g_{\bar{z}}^l, \cdot \rangle, \tag{2.6}$$

where the matrix function m obeys

$$M(z) - M(z_0) = (z - z_0) M(z) S(\bar{z}, z_0) M(z_0) \tag{2.7}$$

and the matrix function $S(\bar{z}, z_0)$ has elements

$$S_{lm}(\bar{z}, z_0) = \langle g_{\bar{z}}^l, g_{z_0}^m \rangle, \quad 1 < l,m < n. \tag{2.8}$$

The vectors g_z^k can be determined as regular analytic function of z with the help of the one-to-one mapping

[a)] On leave of absence from the Institute of Theoretical Physics, Wroclaw University, Wroclaw, Poland.

 0022-2488/85/112777-04$02.50

$$g_z^k = \{1 + (z - z_0)R_z^W\}g_{z_0}^k, \tag{2.9}$$

where the $g_{z_0}^k$ are fixed vectors.

Next we apply the above construction to the densely defined, symmetric, and closed operator H given by Eq. (1.3). The deficiency subspace N_{-i} is spanned by n linearly independent functions.

$$g_i^l(\mathbf{p}) = \frac{1}{(2\pi)^{d/2}} \frac{e^{-i\mathbf{p}\mathbf{x}_l}}{\mathbf{p}^2 - i}. \tag{2.10}$$

Next we use the Fourier transform of $-\Delta$ (call it H_0) as an extension of H. Here H_0 plays the role of A^W in the general construction. The analog of R_z^W is now the resolvent

$$R_z(\mathbf{p}) = 1/(\mathbf{p}^2 - z), \tag{2.11}$$

and according to (2.9) we set

$$g_z^l(\mathbf{p}) = \frac{1}{(2\pi)^{d/2}} \frac{e^{-i\mathbf{p}\mathbf{x}_l}}{\mathbf{p}^2 - z}. \tag{2.12}$$

Note that $g_z^l \notin L^2(\mathbf{R}^d, d^d p)$ for $d \geqslant 4$; this indicates that H is already self-adjoint; therefore δ-like interactions cannot be defined in this way.[11]

The domain of the extension H^U is now given by

$$D(H^U) = D(H) + \sum_{l=1}^n \alpha_l (g_i^l + U g_i^l), \quad \alpha_l \in \mathbf{C}, \quad l = 1,...,n, \tag{2.13}$$

where the unitary operator $U: N_{-i} \to N_i$ can be written as

$$U g_i^l = \sum_{m=1}^n U_{lm} g_{-i}^m. \tag{2.14}$$

The $n \times n$ matrix U_{lm} satisfies

$$U^* S(-i, -i) U^T = S(i,i) = S(-i, -i), \tag{2.15}$$

where $S(i,i)$ is a special case of the matrix function (2.8)

$$S_{lm}(\bar{z}, z_0) = \int \frac{d^d p}{(2\pi)^d} \frac{e^{i\mathbf{p}(\mathbf{x}_l - \mathbf{x}_m)}}{(\mathbf{p}^2 - z)(\mathbf{p}^2 - z_0)}. \tag{2.16}$$

H^U acts on $\psi \in D(H^U)$ like

$$(H^U \psi)(\mathbf{p}) = (\mathbf{p}^2 \psi)(\mathbf{p}) + \sum_{l=1}^n \alpha_l \Big(i g_i^l(\mathbf{p}) - i \sum_{m=1}^n U_{lm} g_{-i}^m(\mathbf{p}) \Big), \quad \psi \in D(H), \tag{2.17}$$

therefore $U_{lm} = -\delta_{lm} =: (U_0)_{lm}$ corresponds to the extension H_0.

In order to use Krein's formula we have to determine $M(z)$ entering into (2.6). Since the resolvents of H^U and H_0 are known for $z = -i$ we get

$$2i(R_{-i}^U - R_{-i}) = V \oplus U - V \oplus U_0 = \sum_{j,k,l=1}^n (U_{jk} + \delta_{jk}) S_{jl}^{-1}(i,i) g_{-i}^k \langle g_i^l | \cdot \rangle, \tag{2.18}$$

where V denotes the Cayley transform of H. Thus

$$M(-i) = (1/2i)(U^T + 1) S^{-1}(i,i), \tag{2.19}$$

and for $\det(U^T + 1) \neq 0$ we find from (2.7)

$$M^{-1}(z) = 2iS(i,i)(U^T + 1)^{-1} - (z + i) S(\bar{z}, -i). \tag{2.20}$$

This is a special case of Ref. 12, p. 372, Eq. (12), and holds for $z \in \rho(H^U) \cap \rho(H_0)$, where $\rho(A)$ denotes the resolvent set of an operator A. Equations (2.6) and (2.20) imply that the spectrum of H^U is given by the absolutely continuous spectrum of H_0 and the pure point spectrum is determined by

$$\sigma_p(H^U) = \{ z \in \mathbf{C} | \det M^{-1}(z) = 0 \}. \tag{2.21}$$

III. FIRST LIMITING PROCEDURE

Here we define a family of operators H_N and show that the sa extensions constructed in Sec. II are obtained as norm resolvent limits from H_N for $N \to \infty$. We start by defining

$$(H_N \psi)(\mathbf{p}) = (H_0 \psi)(\mathbf{p}) + \sum_{l,m=1}^n \Lambda_{lm}^N E_l^N(\mathbf{p}) \langle E_m^N \psi \rangle, \quad \psi \in \mathscr{H} = L^2(\mathbf{R}^d, d^d p), \tag{3.1}$$

where

$$E_l^N(\mathbf{p}) = \begin{cases} e_l(\mathbf{p}), & \text{for } |\mathbf{p}| \leqslant N, \\ 0, & \text{for } |\mathbf{p}| > N, \end{cases} \quad E_l(\mathbf{p}) = \frac{e^{i\mathbf{p} \cdot \mathbf{x}_l}}{(2\pi)^d}, \tag{3.2}$$

and $\langle E_m^N, \psi \rangle$ denotes the scalar product in $\mathscr{H}$. Λ^N is a Hermitian matrix, which we shall relate to the unitary matrix U of (2.17) later on.

Since the potential in (3.1) is bounded (for N finite) we can write down explicitly the resolvent $R_z^N = (H_N - z)^{-1}$ in terms of $R_z = (H_0 - z)^{-1}$:

$$R_z^N = R_z + R_z \sum_{l,m=1}^n E_l^N M_{lm}^N(z) \langle R_{\bar{z}} E_m^N, \cdot \rangle, \tag{3.3}$$

$$-(M^N(z))_{lm}^{-1} = (\Lambda^N)_{lm}^{-1} + G_{lm}^N(z), \quad G_{lm}^N(z) = \langle E_l^N, R_z E_m^N \rangle. \tag{3.4}$$

Since

$$\lim_{N \to \infty} \| R_z (E_l^N - E_l) \| = 0, \tag{3.5}$$

we conclude that (3.3) converges in norm to R_z^U of Eq. (2.6) iff

$$\lim_{N \to \infty} M_{lm}^N(z) = M_{lm}(z), \tag{3.6}$$

with $M(z)$ given by (2.24). Next we have to choose the matrices Λ^N and discuss separately the cases of one, two, and three dimensions.

d=1: $G^N(z)$ from Eq. (3.4) converges as $N \to \infty$ to $G(z)$,

$$G_{lm}(z) = \int \frac{dp}{2\pi} \frac{e^{ip(x_l - x_m)}}{p^2 - z}. \tag{3.7}$$

Note that (2.24) can be rewritten in terms of $G(z)$ as

$$M^{-1}(z) = [G(i) - G(-i)](U^T + 1)^{-1} - G(z) + G(-i). \tag{3.8}$$

If we therefore choose Λ to be independent of N and equal to

$$-\Lambda^{-1} = -(\Lambda^N)^{-1} = (G(i) - G(-i))(U^T + 1)^{-1} + G(-i), \tag{3.9a}$$

$$\Lambda(G(i) + G(-i)U^T) = 1 + U^T, \tag{3.9b}$$

we obtain the relationship between the Hermitian matrix Λ and the unitary matrix U. Equation (3.6) is trivially fulfilled, Hermiticity of Λ easily follows from (2.15).

There is an interesting interpretation in terms of boundary conditions for functions $f\in D(H^U)$. Denote the difference between the right and left derivative of a function ψ by $\delta'(\psi)$,

$$\delta'(\psi)(x) = \lim_{\epsilon\downarrow 0}[\psi'(x+\epsilon) - \psi'(x-\epsilon)], \tag{3.10}$$

and observe that we have

$$\delta'(\tilde{g}_i^l)(x_k) = -\delta_{lk}, \tag{3.11}$$

where $\tilde{g}_i^l$ denotes the inverse Fourier transform of g_i^l. Calculating δ' for an element of $\psi\in D(H^U)$ and using (3.11) shows that the domain of the sa extension may be characterized by the *nonlocal* condition

$$\delta'(\psi)(x_l) = \sum_{m=1}^{n} \Lambda_{lm}\psi(x_m), \tag{3.12}$$

where Λ represents coupling constants in front of nonlocal δ-function potentials.

d=2 and 3: Here we may choose Λ^N of Eq. (3.4) as

$$(\Lambda^N)^{-1} = 2iS(i,i)(U^T+1)^{-1} + G^N(-i), \tag{3.13}$$

with $G^N(-i)$ being defined by (3.4). Note that the diagonal elements of this matrix diverge logarithmically (resp. linearly) in $d=2$ (resp. $d=3$) dimensions. Equation (3.13) generalizes the result of Ref. 11.

Example: We have nonlocal interaction with two centers ($n=2$) at $x=x_1$ and $x=x_2\neq x_1$ for $d=1$. Let

$$\Lambda = \frac{2}{l}\begin{pmatrix} a & \gamma \\ \bar{\gamma} & b \end{pmatrix}, \quad l=|x_1-x_2|, \quad a,b\in\mathbb{R}, \quad \gamma\in\mathbb{C}. \tag{3.14}$$

The pure point spectrum is determined by (2.21) or more explicitly by imposing

$$\det\begin{pmatrix} a+\gamma e^{ik}-ik & ae^{ik}+\gamma \\ be^{ik}+\bar{\gamma} & \bar{\gamma}e^{-ik}+b-ik \end{pmatrix} = 0, \tag{3.15}$$

where $k=\sqrt{z}\cdot l$. Solving this transcendental equation for $k=x+iy$, x,y real, gives bound states for $x=0$, $y>0$ and virtual states for $x=0$, $y\leqslant 0$. Resonances correspond to solutions of (3.15) with $x\neq 0$. The case $a=b$ and $\gamma=0$ has been discussed in Ref. 1. Let us therefore concentrate on the other extreme case of nonlocal δ interactions with $a=b=0$ and γ real. Equation (3.15) means that

$$\gamma e^{ik} - ik = \pm\gamma. \tag{3.16}$$

Bound states are given as solutions of

$$\gamma e^{-y} + y = \pm\gamma, \tag{3.17}$$

with $y>0$. For $0<\gamma<1$ there is no such solution, for $\gamma>1$ or $\gamma<0$ one bound state appears. Resonances are determined by solving simultaneously

$$x - \gamma\sin x e^{-y} = 0, \tag{3.18a}$$

$$y + \gamma\cos x e^{-y} = \pm\gamma. \tag{3.18b}$$

Because of the $x\to -x$ symmetry it suffices to concentrate on the case $x>0$. It can be seen from (3.18a) that there are no resonances in strips

$$\begin{aligned} &x\in[2k\pi,(2k+1)\pi], \quad k=1,2,..., \quad \text{if } \gamma<0, \\ &x\in[(2k-1)\pi,2k\pi], \quad k=1,2,..., \quad \text{if } \gamma>0. \end{aligned} \tag{3.19}$$

In all other strips there are exactly two solutions, with the exception of $x\in[0,\pi]$, where there is only one solution. This can be seen by eliminating y from Eqs. (3.18), getting

$$\ln\left(\frac{\gamma\sin x}{x}\right) + x\cot x = \pm\gamma, \tag{3.20}$$

and studying the asymptotic behavior of both sides of (3.20) as $x\to k\pi$. As a function of γ these resonances move in the appropriate strip.

IV. SECOND PROCEDURE: SCALING LIMIT

Since the local point interactions can be obtained in a nice way with the help of a suitable scaling limit[1–5], it is natural to try to obtain the nonlocal point interaction also in that way. For the case of one center, the starting point is the separable interaction

$$H = -\Delta + \lambda\Phi\langle\Phi,\cdot\rangle. \tag{4.1}$$

The unitary dilatation group on $L^2(\mathbb{R}^d)$ is defined by

$$U^\epsilon(\Phi)(\mathbf{x}) = (1/\epsilon^{d/2})\Phi(\mathbf{x}/\epsilon) = \Phi^\epsilon(\mathbf{x}) \tag{4.2}$$

and a scaled Hamiltonian by

$$\widetilde{H}^\epsilon = (1/\epsilon^2)U_\epsilon HU_\epsilon^{-1} = -\Delta + (\lambda/\epsilon^2)\Phi^\epsilon\langle\Phi^\epsilon,\cdot\rangle. \tag{4.3}$$

Since we intend to discuss nonlocal point interactions with n discrete centers we start with the Hamiltonian

$$H^\epsilon = -\Delta + \frac{1}{\epsilon^2}\sum_{l,m=1}^{n}\Phi_l^\epsilon C_{lm}(\epsilon)\langle\Phi_m^\epsilon,\cdot\rangle, \tag{4.4}$$

where $C(\epsilon)$ denotes a Hermitian matrix and Φ_l^ϵ is centered around $x=x_l$, e.g.,

$$\Phi_l^\epsilon(\mathbf{x}) = (1/\epsilon^{d/2})\Phi_l((\mathbf{x}-\mathbf{x}_l)/\epsilon) \tag{4.5}$$

for some function $\Phi_l\in L^1(\mathbb{R}^d)\cap L^2(\mathbb{R}^d)$. The ϵ dependence of $C(\epsilon)$ will be specified for $d=1,2,3$ below.

The separable case is particularly simple since the resolvent $R_z^\epsilon = (H^\epsilon - z)^{-1}$ can be written down in closed form in terms of R_z; similar to (3.3),

$$R_z^\epsilon = R_z + \delta R_z^\epsilon, \quad \delta R_z^\epsilon = \sum_{l,m=1}^{n} R_z\Phi_l^\epsilon D_{lm}(\epsilon)\langle R_{\bar{z}}\Phi_m^\epsilon,\cdot\rangle, \tag{4.6}$$

with

$$-D_{lm}^{-1}(\epsilon) = \epsilon^2 C_{lm}^{-1}(\epsilon) + \langle\Phi_l^\epsilon, R_z\Phi_m^\epsilon\rangle. \tag{4.7}$$

Next we observe that there is norm convergence of

$$\delta R_z^\epsilon \xrightarrow[\epsilon\to 0]{\|\cdot\|} \sum_{l,m=1}^{n} g_z^l M_{lm}(z)\langle g_{\bar{z}}^m,\cdot\rangle, \quad \operatorname{Im} z>0, \tag{4.8}$$

if $\epsilon^{-d}D^{-1}(\epsilon)$ converges towards $M^{-1}(z)$ and $\int d^dx\Phi_l(x) = 1\neq 0$; this allows us to adjust the ϵ dependence of $D(\epsilon)$.

d=1: We may take a universal dependence of $C_{lm}(\epsilon)$ on ϵ: $C_{lm}^{-1}(\epsilon) = (\epsilon^{-1} + O(1))\widetilde{C}_{lm}^{-1}$; a study of the limit $\lim_{\epsilon\to 0}\epsilon^{-d}D^{-1}(\epsilon) = M^{-1}(z)$ gives

$$-\widetilde{C}^{-1} = (G(i) - G(-i))(U^T+1)^{-1} + G(-i). \tag{4.9}$$

Again $\widetilde{C}$ corresponds to coupling constants of nonlocal point interaction.

d=2: Since the diagonal elements of $C^{-1}(\epsilon)$ have to compensate a logarithmic divergence of $\langle\Phi_l^\epsilon, R_z\Phi_l^\epsilon\rangle$ as $\epsilon\to 0$, we take

$$C_{lm}^{-1}(\epsilon) = -\delta_{lm} \ln \epsilon I_l(\Phi) - J_{lm}(\Phi) + \widetilde{C}_{lm}^{-1}, \quad (4.10)$$

where I_l and J_{lm} are defined by expanding

$$\epsilon^{-2}\langle \Phi_l^\epsilon, R_{-i} \Phi_m^\epsilon \rangle = \delta_{lm} \ln \epsilon I_l(\Phi) + J_{lm}(\Phi) + O(\epsilon) \quad (4.11)$$

and $\widetilde{C}^{-1}$ will be determined immediately in order to get

$$\lim_{\epsilon \downarrow 0} \{ C_{lm}^{-1}(\epsilon) + \epsilon^{-2}\langle \Phi_l^\epsilon, R_z \Phi_m^\epsilon \rangle \} = -M^{-1}(z). \quad (4.12)$$

Therefore we identify

$$\widetilde{C}^{-1} = -2iS(i,i)(U^T + 1)^{-1} + (z+i)S(\bar{z}, -i) \quad (4.13)$$

and the correct z dependence results in (4.12).

$d=3$: This time assume an expansion

$$C_{lm}^{-1}(\epsilon) = C_{lm}^{-1} + \epsilon \widetilde{C}_{lm}^{-1} + O(\epsilon^2). \quad (4.14)$$

Expanding $\langle \Phi_l^\epsilon, R_z \Phi_m^\epsilon \rangle$ one notes again a different behavior of diagonal and off-diagonal matrix elements:

$$\epsilon^{-2}\langle \Phi_l^\epsilon, R_z \Phi_l^\epsilon \rangle \underset{\epsilon \to 0}{\to} \int d^3x \int d^3y \, \Phi_l^*(\mathbf{x} - \mathbf{x}_l) \frac{1}{4\pi|\mathbf{x} - \mathbf{y}|} \Phi_l(\mathbf{y} - \mathbf{x}_l) + \frac{ik}{4\pi}\epsilon + O(\epsilon^2), \quad k = \sqrt{z}$$

$$\epsilon^{-2}\langle \Phi_l^\epsilon, R_z \Phi_m^\epsilon \rangle \underset{\epsilon \to 0}{\to} \epsilon \frac{e^{ik|\mathbf{x}_l - \mathbf{x}_m|}}{4\pi|\mathbf{x}_l - \mathbf{x}_m|}, \quad l \neq m. \quad (4.15)$$

In order to obtain a nontrivial extension a cancellation of terms

$$C_{lm}^{-1} + \int d^3x \int d^3y \, \Phi_l^*(\mathbf{x} - \mathbf{x}_l) \times \frac{1}{4\pi|\mathbf{x} - \mathbf{y}|} \Phi_m(\mathbf{y} - \mathbf{x}_m) = 0 \quad (4.16)$$

is necessary.

Note, that the self-energy expressions in (4.15) and (4.16) are finite, since $\|\Phi\|_{6/5} < \infty$ under our assumptions. Equation (4.16) implies that there exist n resonance functions for the operator

$$-\Delta + \sum_{l,m=1}^{n} \Phi_l C_{lm} \langle \Phi_m, \cdot \rangle, \quad (4.17)$$

given by $R_0 \Phi_k$, which do not belong to the Hilbert space.

The terms of $O(\epsilon)$ in (4.15) allow us to determine $\widetilde{C}^{-1}$ via

$$\widetilde{C}^{-1} + H(z) = -M^{-1}(z),$$

$$H_{lm}(z) = \begin{cases} ik/4\pi, & \text{for } l = m, \\ e^{ik|\mathbf{x}_l - \mathbf{x}_m|}/4\pi|\mathbf{x}_l - \mathbf{x}_m|, & \text{for } l \neq m, \end{cases} \quad (4.18)$$

$$k = \sqrt{z}.$$

The z dependence of both sides in (4.18) is identical; $\widetilde{C}$ is related to the matrix U characterizing the extension through Eqs. (4.18) and (2.20).

Let us finally remark that the second limiting procedure is conceptually not so different from the first one, but shows also the connection of the scaling limit with the low energy expansion for separable potentials.

ACKNOWLEDGMENT

One of us (L.D.) would like to thank Professor W. Thirring for the kind hospitality at the University of Vienna. The other (H. G.) thanks F. Gesztesy for discussions.

[1]S. Albeverio and R. Hoegh-Krohn, J. Operator Theory **6**, 313 (1981).
[2]S. Albeverio and R. Hoegh-Krohn, J. Math. Anal. Appl. **101**, 491 (1984).
[3]S. Albeverio, F. Gesztesy, and R. Hoegh-Krohn, Ann. Inst. H. Poincaré, Sec. A **37**, 1 (1982).
[4]S. Albeverio, F. Gesztesy, R. Hoegh-Krohn, and W. Kirsch, J. Operator Theory **12**, 101 (1984).
[5]S. Albeverio, F. Gesztesy, H. Holten, and R. Hoegh-Krohn, *Solvable Models in Quantum Mechanics* (book in preparation).
[6]M. Breitenecker and H. R. Grümm, Commun. Math. Phys. **15**, 337 (1969).
[7]A. Grossmann, R. Hoegh-Krohn, and M. Mebkhout, J. Math. Phys. **21**, 2376 (1980).
[8]A. Grossmann, R. Hoegh-Krohn, and M. Mebkhout, Commun. Math. Phys. **77**, 87 (1980).
[9]W. Kirsch and F. Martinelli, Commun. Math. Phys. **85**, 329 (1982).
[10]L. E. Thomas, J. Math. Phys. **20**, 1848 (1979).
[11]J. Zorbas, J. Math. Phys. **21**, 840 (1980).
[12]N. Akhiezer and I. M. Glazman, *Theory of Linear Operators in Hilbert Space* (Ungar, New York, 1981), Vol. II.
[13]G. Flamand, in *Applications of Mathematics to Problems in Theoretical Physics*, edited by F. Lurcat (Gordon and Breach, New York, 1967).

Uncertainty relations in stochastic mechanics

S. Golin
Fakultät für Physik, Theoretische Physik, Universität Bielefeld, and Forschungszentrum Bielefeld-Bochum-Stochastik, Postfach 8640, D-4800 Bielefeld 1, Federal Republic of Germany

(Received 4 February 1985; accepted for publication 10 May 1985)

Position–momentum uncertainty in Nelson's stochastic mechanics [Phys. Rev. **150**, 1079 (1966)] has previously been investigated by de la Peña-Auerbach and Cetto [Phys. Lett. A **39**, 65 (1972)]. In this paper their result is generalized, and full equivalence between the uncertainty relations in stochastic mechanics and conventional quantum mechanics is established. Force–momentum uncertainty is also considered.

I. INTRODUCTION

Nelson's stochastic mechanics[1–3] allows for an alternate description of quantum systems. In this framework, the Heisenberg position–momentum uncertainty relation can easily be rediscovered.[4] It corresponds to a purely kinematical fact about diffusions, namely the nondifferentiability of their sample paths.

In fact, stochastic mechanics yields a stronger result than the usual Heisenberg uncertainty relation. Schrödinger[5] was the first to recognize that in quantum mechanics, too, the familiar uncertainty relations can be given a stronger form.

In this paper, it will be shown that Schrödinger's version of the position–momentum uncertainty relation is fully equivalent to the result in stochastic mechanics. This also applies to force–momentum uncertainty.

II. UNCERTAINTY RELATIONS *À LA* SCHRÖDINGER

In this section Schrödinger's derivation of uncertainty relations will be recalled. Since the consideration of domains of operators is irrelevant for our purposes, the domains will not be made mention of in the sequel.

We start by noting that, in general, the product of two Hermitian operators A and B is not Hermitian anymore. But it can be split in a way reminiscent of the decomposition of a complex number into its real and imaginary parts:

$$AB = \tfrac{1}{2}(AB + BA) + \tfrac{1}{2}[A,B].$$

Of course, $AB + BA$ is Hermitian, whereas $[A,B]$ is an anti-Hermitian operator. Let $\langle \cdot \rangle$ denote expectation. Then

$$\mathrm{Re}\,\langle AB \rangle = \tfrac{1}{2}\langle AB + AB \rangle,$$

$$\mathrm{Im}\,\langle AB \rangle = (1/2i)\langle [A,B] \rangle.$$

Define, as usual $\mathrm{Var}\, A := \langle A^2 \rangle - \langle A \rangle^2$. We also introduce the notion of covariance, taking care of the noncommutativity of the operators

$$\mathrm{Cov}(A,B) := \tfrac{1}{2}\langle AB + BA \rangle - \langle A \rangle \langle B \rangle.$$

The covariance of observables has been discussed by Margenau and Hill.[6] They conclude that it has physically attractive features.

For the sake of brevity set $a := A - \langle A \rangle$, $b := B - \langle B \rangle$. By the Schwarz inequality, $\langle a^2 \rangle \langle b^2 \rangle \geqslant |\langle ab \rangle|^2$. According to the above remarks we have $|\langle ab \rangle|^2 = \frac{1}{4}\langle ab + ba \rangle^2 + \frac{1}{4}|\langle [a,b] \rangle|^2$. Therefore,

$$\mathrm{Var}\, A\; \mathrm{Var}\, B \geqslant \mathrm{Cov}^2(A,B) + \tfrac{1}{4}|\langle [A,B] \rangle|^2.$$

This is Schrödinger's version of the uncertainty relations. It differs from the usual form by the additional first term on the right-hand side (rhs). It really gives a stronger bound on the uncertainty of the observables, since $\mathrm{Cov}(A,B)$ does not vanish in general. For instance, let us consider the position and momentum operators in one dimension

$$\begin{aligned}\frac{1}{2}\langle XP - PX \rangle &= \frac{\hbar}{2i}\int dx\, \psi^*\Big[x\frac{\partial}{\partial x}\psi + \frac{\partial}{\partial x}(x\psi)\Big]\\ &= \frac{\hbar}{2i}\int dx\, x\Big[\psi^*\frac{\partial}{\partial x}\psi - \psi\frac{\partial}{\partial x}\psi^*\Big]\\ &= \frac{\hbar}{2i}\int dx\, x\, |\psi|^2 \frac{\partial}{\partial x}\ln\frac{\psi}{\psi^*},\end{aligned}$$

$$\begin{aligned}\langle P \rangle &= \frac{\hbar}{i}\int dx\, \psi^*\frac{\partial}{\partial x}\psi\\ &= \frac{\hbar}{2i}\int dx\Big[\psi^*\frac{\partial}{\partial x}\psi - \psi\frac{\partial}{\partial x}\psi^*\Big]\\ &= \frac{\hbar}{2i}\int dx\, |\psi|^2\frac{\partial}{\partial x}\ln\frac{\psi}{\psi^*},\end{aligned}$$

$$\begin{aligned}\mathrm{Cov}(X,P) = \hbar\Big[&\int dx\, x\, |\psi|^2\frac{\partial\varphi}{\partial x}\\ &- \int dx'\, x'\, |\psi|^2 \int dx\, |\psi|^2 \frac{\partial\varphi}{\partial x}\Big],\end{aligned}$$

where we have set $\psi = |\psi| e^{i\varphi}$. Now it is obvious that $\mathrm{Cov}(X,P) \not\equiv 0$; for example, if $|\psi|^2$ is an even function and $\partial\varphi/\partial x$ is odd, then the rhs need not vanish. Further examples are discussed in Ref. 6.

The reason why the uncertainty relations *à la* Schrödinger are not particularly well-known is that one normally makes use of the uncertainty relations in the interpretation of the noncommutativity of observables: noncommuting observables cannot simultaneously be measured within arbitrary accuracy. And for this statement, of course, the usual uncertainty relations suffice.

III. POSITION–MOMENTUM AND FORCE–MOMENTUM UNCERTAINTY

Nelson's theory of stochastic mechanics provides a different mathematical—and possibly physical—representation of quantum mechanics. The main object in this scheme is a diffusion process $\xi(t)$ associated to the quantum-mechanical system. For simplicity we restrict the following exposition to a particle in one dimension, the generalization to higher dimensions being trivial.

The diffusion is determined by a stochastic differential equation

$$d\xi(t) = b(\xi(t), t)dt + dw(t),$$

where $w(t)$ denotes the Wiener process with variance 2ν. The probability density $\rho(x,t)$ of the process connects the forward drift $b(x,t)$ to the backward drift $b_*(x,t)$ (cf., e.g., Ref. 2)

$$b_*(x,t) = b(x,t) - 2\nu \frac{\partial}{\partial x} \ln \rho(x,t).$$

In fact, these drifts represent nothing but the mean forward and backward velocities of the process $\xi(t)$. The osmotic velocity $u(x,t)$ and the current velocity $v(x,t)$ are defined by

$$u(x,t) := \frac{1}{2}(b(x,t) - b_*(x,t)) = \nu \frac{\partial}{\partial x} \ln \rho(x,t),$$

$$v(x,t) := \tfrac{1}{2}(b(x,t) + b_*(x,t)).$$

We now turn to the uncertainty relations in stochastic mechanics which are due to de la Peña-Auerbach and Cetto (see also Ref. 7).

To simplify the notation, quantities as $\operatorname{Var} u(\xi(t),t)$ will be abbreviated by $\operatorname{Var} u$. By use of the Schwarz inequality,

$$\begin{aligned}\operatorname{Var}\xi \operatorname{Var} u &= E[(\xi - E\xi)^2]E[(u - Eu)^2]\\ &\geqslant |E[(\xi - E\xi)(u - Eu)]|^2\\ &= \operatorname{Cov}^2(\xi,u).\end{aligned}$$

On the other hand (provided the density falls off sufficiently fast),

$$Eu = \nu \int dx \frac{\partial}{\partial x}\rho = 0,$$

$$E[\xi u] = \nu \int dx\, x \frac{\partial}{\partial x}\rho = -\nu,$$

since $\int dx\, \rho(x,t) = 1$. So we can conclude that

$$\operatorname{Cov}^2(\xi u) = \nu^2.$$

Similarly, $\operatorname{Var}\xi \operatorname{Var} v \geqslant \operatorname{Cov}^2(\xi,v)$. Therefore the uncertainty relations in stochastic mechanics assume the form

$$\operatorname{Var}\xi(\operatorname{Var} u + \operatorname{Var} v) \geqslant \operatorname{Cov}^2(\xi,v) + \nu^2.$$

We shall now interpret this result in the language of conventional quantum mechanics. So far we have only considered the kinematical aspects of the diffusion. In order to relate the stochastic point of view to ordinary quantum mechanics, we have to add the dynamics. This can be accomplished, e.g., by the Guerra/Morato variational principle.[3,8]

As a result of this, the diffusion coefficient ν is then given in terms of $\hbar$, Planck's constant divided by 2π, and the particle mass m,

$$\nu = \hbar/2m.$$

The probability density $\rho(x,t)$ of the process and the (normalized) solution $\psi(x,t)$ of the Schrödinger equation are related by

$$\rho(x,t) = |\psi(x,t)|^2.$$

Moreover,

$$u(x,t) = \frac{\hbar}{m} \operatorname{Re} \frac{\partial}{\partial x} \ln \psi(x,t),$$

$$v(x,t) = \frac{\hbar}{m} \operatorname{Im} \frac{\partial}{\partial x} \ln \psi(x,t).$$

Clearly, $\operatorname{Var} Q = \operatorname{Var}\xi$. Also,

$$\begin{aligned}\langle P\rangle &= \frac{\hbar}{i}\int dx\, \psi^* \frac{\partial}{\partial x}\psi = \frac{m}{i}\int dx\, \rho(u + iv)\\ &= E[mv],\\ \langle P^2\rangle &= -\hbar^2 \int dx\, \psi^* \frac{\partial^2}{\partial x^2}\psi\\ &= -\hbar^2 \int dx \left|\frac{\partial}{\partial x}\psi\right|^2\\ &= m^2 \int dx\, \rho(u^2 + v^2)\\ &= E[m^2u^2] + E[m^2v^2] = \operatorname{Var}[mu] + E[m^2v^2],\end{aligned}$$

and therefore

$$\operatorname{Var} P = \operatorname{Var}[mu] + \operatorname{Var}[mv].$$

Likewise,

$$\begin{aligned}\tfrac{1}{2}\langle XP + PX\rangle &= \operatorname{Re}\langle XP\rangle\\ &= \operatorname{Re}\int dx\, x\psi^* \frac{\hbar}{i}\frac{\partial}{\partial x}\psi\\ &= \hbar \operatorname{Im}\int dx\, x\psi^* \frac{\partial}{\partial x}\psi\\ &= \hbar \int dx\, x\rho \operatorname{Im}\frac{\partial}{\partial x}\ln\psi\\ &= m\int dx\, x\rho v = E[\xi(mv)].\end{aligned}$$

Hence

$$\operatorname{Cov}(X,P) = \operatorname{Cov}(\xi,mv).$$

Now the stochastic uncertainty relation can be rewritten in quantum-mechanical terms as

$$\operatorname{Var} X \operatorname{Var} P \geqslant \operatorname{Cov}^2(X,P) + \hbar^2/4,$$

and this is exactly the quantum-mechanical position–momentum uncertainty relation in Schrödinger's formulation. We also note that the Heisenberg uncertainty principle follows already from

$$\operatorname{Var}\xi \operatorname{Var}[mu] \geqslant \hbar^2/4,$$

i.e., it can be traced back to the nondifferentiability of the trajectories of $\xi(t)$, which shows up in $b \neq b_*$ or $u \neq 0$.

What other uncertainty relations could be thought of in the framework of stochastic mechanics? Rather than taking position we might consider a function of it, e.g., we could take the force $ma = (\partial/\partial x)V$, where a denotes acceleration and V is the potential. The uncertainty relations in stochastic mechanics are easily obtained:

$$\mathrm{Var}\, a(\mathrm{Var}\, u + \mathrm{Var}\, v) \geqslant \mathrm{Cov}^2(a,v) + (E\,[au])^2,$$

since $Eu = 0$.

Let F denote the operator corresponding to force [i.e., F is multiplication by $(d/dx)V$]. The computations for position and momentum can be mimicked and one obtains

$$\mathrm{Var}\, F = \mathrm{Var}[ma],$$

$$\mathrm{Cov}(F,P) = \mathrm{Cov}(ma,mv),$$

$$\langle [F,P] \rangle = -2iE\,[(ma)(mu)],$$

i.e.,

$$\mathrm{Var}\, F\, \mathrm{Var}\, P \geqslant \mathrm{Cov}^2(P,F) + \tfrac{1}{4}|\langle [F,P]\rangle|^2.$$

Again, there is full correspondence between stochastic mechanics and conventional quantum mechanics.

IV. CONCLUSION

We have seen that the position–momentum and force–momentum uncertainty relations gained in stochastic mechanics are equivalent to those in quantum mechanics. In the stochastic approach the basic underlying fact for these relations is the nondifferentiability of the sample paths.

There are, however, more observables that can be considered in this scheme (some of them cannot even be formulated in conventional quantum mechanics). Can we establish uncertainty relations for them, too? The possibility of further uncertainty relations is under present investigation.

ACKNOWLEDGMENTS

I am indebted to Professor Ph. Blanchard for comments on this paper and the benefit of many conversations. Discussions with Professor S. Albeverio are gratefully acknowledged. Thanks are also due to Professor I. Białynicki-Birula for pointing out Ref. 5 to me.

[1]E. Nelson, "Derivation of the Schrödinger equation from Newtonian mechanics," Phys. Rev. **150**, 1079 (1966).
[2]E. Nelson, *Dynamical Theories of Brownian Motion* (Princeton U. P., Princeton, NJ, 1967).
[3]E. Nelson, *Quantum Fluctuations* (Princeton U. P., Princeton, NJ, to be published).
[4]L. de la Peña-Auerbach and A. M. Cetto, "Stronger form for the position–momentum uncertainty relation," Phys. Lett. A **39**, 65 (1972).
[5]E. Schrödinger, "Zum Heisenbergschen Unschärfeprinzip," Sitzungsber. Preuss. Akad. Wiss. Phys. Math. Kl., 296 (1930).
[6]H. Margenau and R. N. Hill, "Correlation between measurements in quantum theory," Prog. Theor. Phys. **26**, 722 (1961).
[7]D. de Falco, S. De Martino, and S. De Siena, "Position–momentum uncertainty in stochastic mechanics," Phys. Rev. Lett. **49**, 181 (1982).
[8]F. Guerra and L. M. Morato, "Quantization of dynamical systems and stochastic control theory," Phys. Rev. D **27**, 1774 (1983).

Solution of the Schrödinger equation for a particle in an equilateral triangle

Wai-Kee Li
Department of Chemistry, Chinese University of Hong Kong, Shatin, N. T., Hong Kong
S. M. Blinder
Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109

(Received 25 April 1985; accepted for publication 12 June 1985)

The complete solution for the quantum-mechanical problem of a particle in an equilateral triangle is derived. By use of projection operators, eigenfunctions belonging explicitly to each of the irreducible representations of the symmetry group C_{3V} are constructed. The most natural definition of the quantum numbers p and q includes not only integers but also nonintegers of the class $\frac{1}{3}$ and $\frac{2}{3}$ modulo 1. Some relevant features relating to symmetry and degeneracy are also discussed.

I. INTRODUCTION

The two-dimensional Schrödinger equation for a particle confined within an equilateral triangle has been considered by several authors.[1-4] Mathews and Walker[1] derived a solution in the form of a double Fourier series after generating a periodic lattice by successive reflections and rotations of the triangle. Krishnamurthy *et al.*[2] applied an ingenious transformation of the solution for three fermions in a one-dimensional segment into that for a single particle in a triangle. Shaw[3] reduced the Schrödinger equation to a quasi-one-dimensional form involving a complex coordinate $z = x + iy$. However, he obtained only those eigenstates transforming as the A_1 and A_2 representations of the symmetry group C_{3V}. The corresponding problem in a classical context was solved by Lamé[4] a very long time ago.

The various solutions of the problem result in functional forms and energy expressions of rather different appearance. In common with the problem of the isosceles right triangle, recently solved by one of us,[5] the Schrödinger equation for the equilateral triangle is *not* soluble by separation of variables. Recently, analogous nonseparable solutions for tetrahedral boxes also have been obtained.[2,6]

II. METHOD OF SOLUTION

We seek solutions of the Schrödinger equation

$$-\left(\frac{\hbar^2}{2m}\right)\left[\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right]\Psi(x,y)=E\Psi(x,y), \quad (1)$$

such that $\Psi(x,y) = 0$ on the three sides of an equilateral triangle of side a situated as shown in Fig. 1(a). It is convenient to introduce the altitude of the triangle, given by $A = (\sqrt{3}/2)a$. The three boundary conditions thus require that

$$\Psi(x,y)=0, \quad \text{when} \begin{cases} y=0, \\ y=\sqrt{3}x, \\ y=\sqrt{3}(a-x)=2A-\sqrt{3}x. \end{cases} \quad (2)$$

It will be expedient to introduce three auxiliary variables

$$u=(2\pi/A)y, \quad v=(2\pi/A)(-y/2+\sqrt{3}x/2),$$
$$w=(2\pi/A)(-y/2-\sqrt{3}x/2)+2\pi. \quad (3)$$

These are proportional to the perpendicular distances from an interior point to the three sides of the triangle, as shown in Fig. 1(b). The sum of these perpendiculars equals the altitude of the triangle and thus

$$u+v+w=2\pi. \quad (4)$$

The boundary conditions (2) now assume the more symmetrical form

$$\Psi=0, \quad \text{when} \begin{cases} u=0, & v=2\pi-w, \\ v=0, & w=2\pi-u, \\ w=0, & u=2\pi-v. \end{cases} \quad (5)$$

The equilateral triangle problem is invariant under the point group C_{3V}. Equivalently, the sides (or vertices) can be permuted according to the symmetric group S_3, isomorphic with C_{3V}. Let the variables u,v,w transform under S_3 as follows:

$$C_3\colon u\leftarrow v\leftarrow w\leftarrow u, \quad C_3^2\colon u\rightarrow v\rightarrow w\rightarrow u,$$
$$\sigma_1\colon u\leftrightarrow v, \quad \sigma_2\colon w\leftrightarrow u, \quad \sigma_3\colon v\leftrightarrow w. \quad (6)$$

Thus the vector (u,v,w) generates the following 3×3 reducible representation of S_3 or C_{3V}:

$$E=\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}, \quad C_3=\begin{pmatrix}0&1&0\\0&0&1\\1&0&0\end{pmatrix},$$
$$C_3^2=\begin{pmatrix}0&0&1\\1&0&0\\0&1&0\end{pmatrix}, \quad \sigma_1=\begin{pmatrix}0&1&0\\1&0&0\\0&0&1\end{pmatrix}, \quad (7)$$
$$\sigma_2=\begin{pmatrix}0&0&1\\0&1&0\\1&0&0\end{pmatrix}, \quad \sigma_3=\begin{pmatrix}1&0&0\\0&0&1\\0&1&0\end{pmatrix}.$$

Boundary conditions aside, the free-particle Schrödinger equation (1) admits solutions of the form $f(c_1x + c_2y)$, in which $f(z)$ is a harmonic function such as $\sin z$, $\cos z$, or $\exp(\pm iz)$. With the use of Eqs. (3), let this function be expressed in the form $f(pu - qv)$, in which p and q are constants.

 0022-2488/85/112784-03$02.50

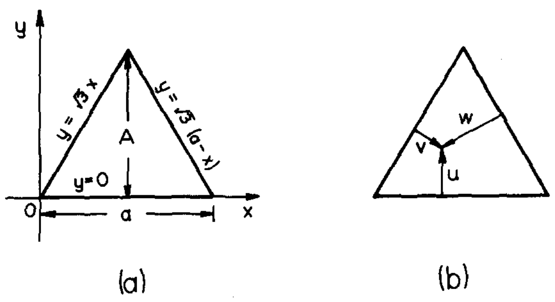


FIG. 1. (a) The coordinate system for an equilateral triangle showing boundary conditions. (b) The auxiliary variables u,v,w.

III. A_1 SOLUTIONS

We shall next construct C_{3v}-symmetry adapted functions by application of projection operators. Recall that C_{3v} admits of three irreducible representations: A_1 and A_2, which are nondegenerate, and E, which is doubly degenerate. The A_1 projection operator

$$\mathscr{P}(A_1)=E+C_3+C_3^2+\sigma_1+\sigma_2+\sigma_3, \tag{8}$$

applied to the "basis function" $f(pu-qv)$, with the use of (6), gives

$$\Psi_{p,q}(A_1)=f(pu-qv)+f(pv-qw)+f(pw-qu)$$
$$+f(pv-qu)+f(pw-qv)+f(pu-qw). \tag{9}$$

It is readily shown that the boundary conditions (5) can be fulfilled only if $f=\sin$ and p,q are integers. We find further that

$$\begin{aligned}\Psi_{q,p}&=-\Psi_{p,q},\\ \Psi_{-p,-q}&=-\Psi_{p,q},\\ \Psi_{p+q,-p}&=-\Psi_{p,q}.\end{aligned} \tag{10}$$

Thus, without loss of generality, the quantum numbers p,q can be restricted such that $p>q\geqslant 0$, with p and q integral. The eigenfunctions (9) can be reduced to more compact trigonometric forms as follows:

$$\Psi_{p,q}(A_1)=\cos[q\sqrt{3}\pi x/A]\sin[(2p+q)\pi y/A]$$
$$-\cos[p\sqrt{3}\pi x/A]\sin[(2q+p)\pi y/A]$$
$$-\cos[(p+q)\sqrt{3}\pi x/A]\sin[(p-q)\pi y/A],$$
$$q=0,1,2,\ldots,\quad p=q+1,q+2,\ldots. \tag{11}$$

Specifically, for $q=0$,

$$\Psi_{p,0}(A_1)=\sin(2p\pi y/A)-2\sin(p\pi y/A)$$
$$\times\cos(p\sqrt{3}\pi x/A),\quad p=1,2,3\ldots. \tag{12}$$

Note that the above functions are not normalized. These agree with the specific cases listed by Shaw.[3] The energy eigenvalues corresponding to (11) and (12) are given by

$$\begin{aligned}E_{p,q}&=(p^2+pq+q^2)E_0,\\ E_0&\equiv h^2/2mA^2=E_{1,0}.\end{aligned} \tag{13}$$

IV. A_2 SOLUTIONS

For the A_2 representations, the projection operator

$$\mathscr{P}(A_2)=E+C_3+C_3^2-\sigma_1-\sigma_2-\sigma_3 \tag{14}$$

applied to $f(pu-qv)$ results in

$$\Psi_{p,q}(A_2)=f(pu-qv)+f(pv-qw)+f(pw-qu)$$
$$-f(pv-qu)-f(pw-qv)-f(pu-qw). \tag{15}$$

These fulfill the boundary conditions with $f=\cos$ and, again, for integral p,q. In analogy with (10), we find for the A_2 functions,

$$\begin{aligned}\Psi_{q,p}&=-\Psi_{p,q},\\ \Psi_{-p,-q}&=\Psi_{p,q},\\ \Psi_{p+q,-q}&=-\Psi_{p,q}.\end{aligned} \tag{16}$$

The last relation shows that $\Psi_{p,q}=0$ if $q=0$. Otherwise the same spectrum as the A_1 functions is obtained, with $p>q>0$, p and q integral. The A_2 eigenfunctions in trigonometric form analogous to (11) are given by

$$\Psi_{p,q}(A_2)=\sin[q\sqrt{3}\pi x/A]\sin[(2p+q)\pi y/A]$$
$$-\sin[p\sqrt{3}\pi x/A]\sin[(2q+p)\pi y/A]$$
$$+\sin[(p+q)\sqrt{3}\pi x/A]\sin[(p-q)\pi y/A],$$
$$q=1,2,3,\ldots,\quad p=q+1,q+2,\ldots. \tag{17}$$

The eigenvalues are again given by (13), except that $q=0$ is missing. Remarkably, every A_2 eigenstate is degenerate with an A_1 eigenstate carrying the same quantum numbers. The only nondegenerate eigenstates are the A_1 with $q=0$. A similar situation arises for a particle in a square, as discussed by Shaw,[3] in which there occur degenerate pairs of A_1+B_1 species and again of A_2+B_2 species.

V. E SOLUTIONS

Finally, for the E representation, we make use of the projection operator

$$\mathscr{P}(E)=E+\epsilon C_3+\epsilon^* C_3^2-\sigma_1\mathscr{C}-\epsilon\sigma_2\mathscr{C}-\epsilon^*\sigma_3\mathscr{C}, \tag{18}$$

where $\epsilon=\exp(2\pi i/3)$ and $\mathscr{C}$ represents the operation of complex conjugation. Applying (18) to $f(pu-qv)$ we obtain

$$\Psi_{p,q}(E)=f(pu-qv)+\epsilon f(pv-qw)$$
$$+\epsilon^* f(pw-qu)-f^*(pv-qu)$$
$$-\epsilon f^*(pw-qv)-\epsilon^* f^*(pu-qw). \tag{19}$$

The boundary conditions are satisfied with the function $f(z)=\exp(+iz)$, but now with $p,q=n+\frac{1}{3}$ ($n=$ integer). The complex conjugate of (19) gives the partner in this degenerate representation. One can alternatively apply (18) with $\epsilon=\exp(-2\pi i/3)$. This generates a second class of E eigenfunctions (19) with $p,q=n+\frac{2}{3}$. The following relationships among the E functions can be demonstrated:

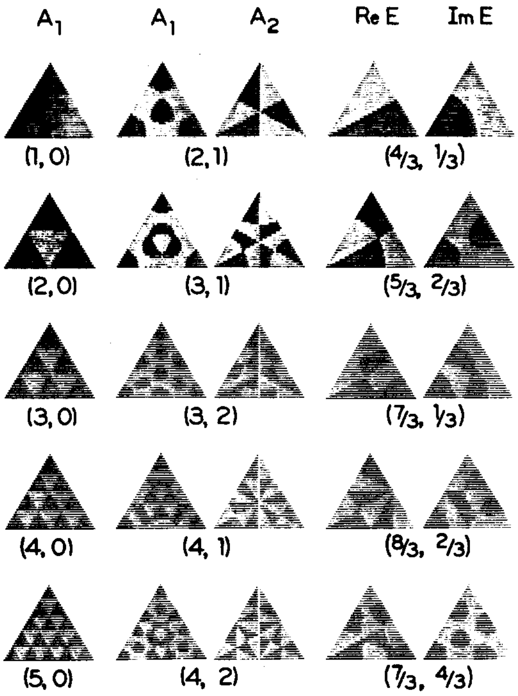


FIG. 2. The graphical representation of some lower eigenstates (p,q) of each symmetry type. For visual simplicity, only the sign ($+$ or $-$) of the wave function is plotted.

$$\begin{aligned} \Psi_{q,p} &= -\Psi_{p,q}, \\ \Psi_{-p,-q} &= \Psi^*_{p+1/3,q+1/3}, \\ \Psi_{p+q,-p} &= -\Psi^*_{p+1/3,q+1/3}. \end{aligned} \tag{20}$$

Thus, E states can be labeled by the quantum numbers $q=\frac{1}{3},\frac{2}{3},\frac{4}{3},\frac{5}{3},\ldots$, $p=q+1,q+2,\ldots$. The real and imaginary parts of $\Psi_{p,q}(E)$ turn out to have the same forms as (17) and (11), respectively, but with p,q now equal to $\frac{1}{3}$ or $\frac{2}{3}$ modulo 1, viz.,

$$\begin{aligned} \operatorname{Re}\Psi_{p,q}(E) &= \Psi_{p,q}(A_2), \\ \operatorname{Im}\Psi_{p,q}(E) &= \Psi_{p,q}(A_1), \\ q=\tfrac{1}{3},\tfrac{2}{3},\tfrac{4}{3},\tfrac{5}{3},\ldots, &\quad p=q+1,q+2,\ldots. \end{aligned} \tag{21}$$

VI. SUMMARY

The Schrödinger equation (1) subject to the boundary conditions (2) has solutions $\Psi_{p,q}$. The A_1 eigenfunctions are given by Eq. (9) or Eq. (11) [Eq. (12) if $q=0$], the A_2 eigenfunctions by Eq. (15) or Eq. (17), and the E eigenfunctions by Eq. (19) or Eq. (21). Figure 2 represents some of the lower-energy eigenfunctions of each symmetry species. For visual simplicity, only the sign of the wave function ($+$ or $-$) is plotted. The energy eigenvalues are given by the formula

$$E_{p,q} = (p^2+pq+q^2)E_0,$$

$$q = \begin{cases} 0,1,2,\ldots, & \text{for } A_1, \\ 1,2,3,\ldots, & \text{for } A_2, \\ \frac{1}{3},\frac{2}{3},\frac{4}{3},\frac{5}{3},\ldots, & \text{for } E, \end{cases}$$

$$p = q+1, q+2,\ldots.$$

As discussed in Refs. 3 and 7, systems of high symmetry often exhibit "accidental" degeneracies beyond those implied by the purely geometrical symmetry of the Hamiltonian. Thus, in the equilateral triangle problem, $E=49$ (in units of E_0) represents a threefold-degenerate level compounded of an A_1 state with a A_1-A_2 pair, corresponding to the $(p,q)=(7,0)$ and $(5,3)$. This is the first of an infinite number of such combinations. The first fourfold degeneracy from two coinciding A_1-A_2 pairs occurs for $E=91$, with $(p,q)=(6,5)$ and $(9,1)$. We eventually encounter a sixfold degeneracy at $E=1519$ with states (23,22), (33,10), (35,7) and an eightfold degeneracy at $E=1729$ with states (25,23), (32,15), (37,8), (40,3). Degeneracies also arise from coincident E levels. Thus $E=30\frac{1}{3}$ is fourfold degenerate with states $(\frac{11}{3},\frac{8}{3})$ and $(\frac{16}{3},\frac{1}{3})$; $E=212\frac{1}{3}$ is sixfold degenerate with states $(\frac{31}{3},\frac{19}{3})$, $(\frac{35}{3},\frac{14}{3})$ and $(\frac{41}{3},\frac{5}{3})$. Such nongeometrical degeneracies can often be enumerated by applying results from number theory. For example, the number of integer combinations (m,n) such that m^2+mn+n^2 equals a particular integer is calculable.[8]

An amusing correspondence can be drawn between equilateral triangle eigenstates and families of leptons and quarks. The doubly degenerate levels with the quantum numbers $n+\frac{1}{3}$ and $n+\frac{2}{3}$ are quite suggestive of pairs of quarks (right and left handed) with charge $+\frac{1}{3}$ and $-\frac{2}{3}$, respectively. Similarly, the degenerate A_1,A_2 states might correspond to pairs of (right and left) leptons such as electrons or muons. Finally, the nondegenerate A_1's with $q=0$ suggest left-handed neutrinos (the right-handed partners being nonexistent).

[1] J. Mathews and R. L. Walker, *Mathematical Methods for Physicists* (Benjamin, New York, 1970), 2nd ed. pp. 237ff.
[2] H. R. Krishnamurthy, H. S. Mani, and H. C. Verma, J. Phys. A **15**, 2131 (1982). See also J. W. Turner, J. Phys. A **17**, 2791 (1984).
[3] G. B. Shaw, J. Phys. A **7**, 1357 (1974).
[4] M. G. Lamé, *Leçons sur la Théorie Mathématique de l'Elasticité des Corps Solides* (Bachelier, Paris, 1852), §57.
[5] W-K. Li, J. Chem. Educ. **61**, 1034 (1984).
[6] J. W. Turner, J. Phys. A **17**, 2791 (1984).
[7] W-K. Li, Am. J. Phys. **50**, 666 (1982).
[8] See, for example, E. D. Bolker, *Elementary Number Theory* (Benjamin, New York, 1970), p. 121.

Vector coherent state representation theory

D. J. Rowe
Department of Physics, University of Toronto, Toronto, Ontario, Canada M5S 1A7
G. Rosensteel
Department of Physics, Tulane University, New Orleans, Louisiana 70118
R. Gilmore
Department of Physics and Atmospheric Science, Drexel University, Philadelphia, Pennsylvania 19014

(Received 1 May 1985; accepted for publication 12 June 1985)

A vector coherent state theory is formulated as a natural extension of standard coherent state theory. It is shown that the Godement representations and the coherent state representations of the Sp(N,R) groups of Rowe and of Deenen and Quesne are special cases of this more general theory.

I. INTRODUCTION

We give the theory of vector coherent states as a natural extension of standard coherent state theory. The extension is related to Mackey's induced representation theory.[1] Both the standard and vector coherent state representations may be regarded as induced from a subgroup K_0 of a group G_0. However, for the standard coherent states, the representations are induced from one-dimensional representations of K_0, whereas in the vector generalization, the induction is from finite-dimensional vector representations of K_0.

Vector coherent state representations were recently introduced for the noncompact Sp(N,R) groups.[2–4] They proved to be extraordinarily powerful for the evaluation of matrix elements of the Sp(N,R) algebra in a U(N) basis. In particular, they provided analytic expressions for the matrix elements whenever the Sp(N,R) $\mapsto$U(N) branching is multiplicity-free and simple recursion relations for the evaluation of matrix elements in general.[5,6]

The Sp(N,R) vector coherent states are holomorphic vector-valued functions defined on the generalized unit disk D_N, i.e., the complex manifold of $N\times N$ symmetric complex matrices w with $I - w^*w$ positive definite. We shall prove that this set of coherent state representations is directly related to the Sp(N,R) discrete series of Godement,[7] which was studied by Rosensteel[8] and Rosensteel and Rowe.[9] The representation spaces of this series were given therein as Hilbert spaces of holomorphic vector-valued functions on the Siegel half-plane S_N, i.e., the complex manifold of symmetric $N\times N$ complex matrices z with Im z positive definite. The connection is provided by the bijective and analytic Cayley transformation[10] from the unit disk onto the upper half-plane $w\rightarrow z = i(I+w)(I-w)^{-1}$. Thus, the Godement representations may be regarded as a special case of vector coherent state theory.

More generally, for the discrete series of a noncompact group G_0, Okamoto[11] has shown that the representations induced from its maximal compact subgroup K_0 decompose naturally into irreducible subspaces of holomorphic sections of vector bundles over G_0/K_0. The Sp(N,R) case fits into the Okamoto scheme when we identify Sp(N,R)/U(N) with either the Siegel half-plane or the unit disk. Moreover, the strategy of attempting to realize the discrete series representations in Hilbert spaces of holomorphic functions was initiated by Bargmann[12] for SU(1,1) and generalized to the Hermitian symmetric case G_0/K_0 by Harish-Chandra[13] and Schmid.[14]

Boson representations of the Lie algebra are given naturally in this setting from the Lie derivatives of the group actions. Thus, the Sp(N,R) boson representations on the Siegel half-plane were determined by Rosensteel[8] and Rosensteel and Rowe.[9] In spaces of analytic functions defined on the generalized unit disk, Deenen and Quesne[3] and Rowe[4] computed the boson representations using coherent state ideas.

The plan of this paper is to first present the general theory of vector-valued coherent state representations, then apply the construction to the Sp(N,R) case, thereby recovering the results of Rowe,[4] and finally relating this construction to the realizations on spaces of functions defined on the unit disk and half-plane.[8,9]

II. VECTOR-VALUED COHERENT STATE REPRESENTATIONS

Let G_0 be a semisimple Lie group with Lie algebra $\mathbf{g}_0$ and suppose G_0 has a faithful finite-dimensional representation. Let $\mathbf{g}$ be the complex extension of $\mathbf{g}_0$ and let G be the corresponding Lie group. Let K_0 be a compact semisimple Lie subgroup of G_0 with Lie algebra $\mathbf{k}_0$. Let $\mathbf{k}$ be the complex extension of $\mathbf{k}_0$ and K the corresponding Lie group. We suppose that K_0 contains a Cartan subgroup of G_0. Then, $\mathbf{g}$ may be decomposed as the direct sum

$$\mathbf{g} = \mathbf{n}_+ + \mathbf{k} + \mathbf{n}_-, \tag{1}$$

where $\mathbf{n}_+$ and $\mathbf{n}_-$ are, respectively, spaces of positive and negative roots. Let $P\subset G$ be the parabolic subgroup with Lie algebra

$$\mathbf{p} = \mathbf{n}_+ + \mathbf{k}. \tag{2}$$

Let U be a unitary irreducible lowest weight (often called highest weight) representation of G_0 acting on a Hilbert space H with lowest weight state $|0\rangle$. We require that $|0\rangle$ also be a lowest weight vector for K_0. Then, let u be the irreducible representation of K_0 acting on the subspace $H_u \subset H$ containing $|0\rangle$. Suppose that T is an extension of U to $\mathbf{g}$ and ρ is the extension of u to K (see Ref. 15).

We canonically identify G_0/K_0 with an open submani-

 0022-2488/85/112787-05$02.50

fold of G/P (Harish-Chandra,[13] Schmid,[14] Bott,[16] Griffiths and Schmid[17]). Let $\{E_i^-\}$ denote a basis for $\mathbf{n}_-$. An arbitrary vector in $\mathbf{n}_-$ can then be expanded as $z\cdot E^- \equiv \Sigma_i z_i E_i^-$, where (z_i) are complex numbers. By regarding $z\cdot E^-$ as a representative of a left coset

$$P\exp(z\cdot E^-)\in G/P, \tag{3}$$

the complex numbers $z=(z_i)$ become complex coordinates for G/P. If the unitary irreducible representation (UIR) u of K_0 is of dimension 1, so that

$$u(x)|0\rangle = e^{i\lambda(x)}|0\rangle,\quad x\in K_0,$$

and if $T(X)|0\rangle$ is not in H_u for any nonzero $X\in\mathbf{n}_+$, then K_0 is the little group of U at $|0\rangle$ and we have the situation for standard coherent state theory. For any normalized state $|\psi\rangle\in H$, we then have the unnormalized coherent state wave function

$$\psi(z) = \langle 0|T(e^{z\cdot E^-})|\psi\rangle. \tag{4}$$

With this normalization, $\psi(z)$ is evidently a holomorphic function on G_0/K_0.

More generally, if $\dim H_u > 1$ and $T(X)|\alpha\rangle$ is not in $\mathbf{H}_u$ for any $|\alpha\rangle\in H_u$ and any nonzero $X\in n_+$, we can regard K_0 as the little group that leaves H_u invariant. If $(|\alpha\rangle)$ is an orthonormal basis for H_u we define a vector coherent state wave function $\psi(z)$, for any $|\psi\rangle\in H$, by

$$\psi(z) = \sum_\alpha |\alpha\rangle\,\langle\alpha|T(e^{z\cdot E^-})|\psi\rangle. \tag{5}$$

This is evidently a vector-valued holomorphic function $G/K\rightarrow H$.

The group action is given by

$$\Gamma(g)\psi(z) = \sum_\alpha |\alpha\rangle\,\langle\alpha|T(e^{z\cdot E^-})U(g)|\psi\rangle,\quad g\epsilon G. \tag{6}$$

III. THE IDENTITY RESOLUTION AND THE INNER PRODUCT

For square-integrable irreducible representations, it is well known that

$$I = \int_{G_0} U(g^{-1})|0\rangle\,\langle 0|U(g)dv(g), \tag{7}$$

where $dv(g)$ is the G_0-invariant volume element [suitably normalized for each U], $|0\rangle$ is the lowest weight state, and I is the identity on H (see Ref. 18).

The coset in G/P defined by z was given by Eq. (3). To identify the corresponding coset in $G_0/K_0\sim G/P$ we need to specify the diffeomorphism. This can be done by choosing coset representatives

$$G/P\rightarrow G_0,\quad P\exp(z\cdot E^-)\mapsto\kappa(z) \tag{8}$$

having the property

$$\kappa(z) = e^{\omega(z)\cdot E^+}x(z)e^{z\cdot E^-}, \tag{9}$$

where $\omega(z)\cdot E^+\in\mathbf{n}_+$ and $x(z)\in K$. We then have

$$G/P\rightarrow G_0/K_0,\quad P\exp(z\cdot E^-)\mapsto K_0\kappa(z). \tag{10}$$

An arbitrary group element $g\in K_0\kappa(z)$ can now be expressed $g=k\kappa(z)$ for some $k\epsilon K_0$. Hence

$$I = \int_{G_0} T(e^{z\cdot E^-})^\dagger\rho(x(z))^\dagger u(k^{-1})|0\rangle\,\langle 0| \times u(k)\,\rho(x(z))T(e^{z\cdot E^-})dv(g). \tag{11}$$

It now follows from a theorem of Helgason[19] (Theorem 1.7) that

$$I = \int_{G_0/K_0} T(e^{z\cdot E^-})^\dagger\rho(x(z))^\dagger \times\left\{\int_{K_0} u(k^{-1})|0\rangle\,\langle 0|u(k)dv(k)\right\} \times\rho(x(z))T(e^{z\cdot E^-})d\mu(z), \tag{12}$$

where $dv(k)$ is the invariant measure on K_0 and $d\mu(z)$ is the G_0-invariant measure on G_0/K_0. But the quantity in parenthesis is just the identity on H_u, which also can be expressed as

$$\int_{K_0} u(k^{-1})|0\rangle\,\langle 0|u(k)dv(k) = \sum_\alpha|\alpha\rangle\,\langle\alpha|. \tag{13}$$

Therefore, we have the resolution of the identity

$$I = \sum_\alpha\int_{G_0/K_0} T(e^{z\cdot E^-})^\dagger\rho(x(z))^\dagger|\alpha\rangle\,\langle\alpha| \times\rho(x(z))T(e^{z\cdot E^-})d\mu(z). \tag{14}$$

The inner product for coherent state wave functions is now given immediately by

$$\langle\Psi|\Phi\rangle = \int_{G_0/K_0}(\rho(x(z))\Psi(z),\rho(x(z))\Phi(z))d\mu(z), \tag{15}$$

where $(\ ,\ )$ is the Hilbert space inner product for H_u.

IV. APPLICATION TO Sp(N,R)⊃U(N)

It is convenient to regard $G_0=\mathrm{Sp}(N,R)$ as the subgroup of $\mathrm{Sp}(N,C)$ given by the isomorphism

$$\mathrm{Sp}(N,R)\sim\mathrm{Sp}(N,C)\cap\mathrm{U}(N,N). \tag{16}$$

Thus an element $g\in\mathrm{Sp}(N,R)$ is a $2N\times 2N$ matrix of the form

$$g=\begin{pmatrix}\alpha & \beta\\ \beta^* & \alpha^*\end{pmatrix} \tag{17}$$

with

$$\alpha\alpha^\dagger-\beta\beta^\dagger=I,\quad \alpha\tilde{\beta}=\beta\tilde{\alpha}, \tag{18}$$

where I is now the $N\times N$ identity matrix and $\tilde{\beta}$ is β transposed.

The group $K_0=\mathrm{U}(N)$ of $N\times N$ unitary matrices $\{\alpha;\alpha\alpha^\dagger=I\}$ is embedded in $\mathrm{Sp}(N,R)$ as the subgroup of $2N\times 2N$ matrices of the form

$$\begin{pmatrix}\alpha & 0\\ 0 & \alpha^*\end{pmatrix}\in\mathrm{Sp}(N,R). \tag{19}$$

A basis for the Lie algebra $\mathbf{k}$ of K is given by the matrices

$$C_{ij}=\begin{pmatrix}E_{ij} & 0\\ 0 & -E_{ij}\end{pmatrix},\quad i,j=1,\ldots,N, \tag{20}$$

where E_{ij} is the $N\times N$ matrix with elements

$$(E_{ij})_{lk}=\delta_{il}\delta_{jk}. \tag{21}$$

Bases for $\mathbf{n}_+$ and $\mathbf{n}_-$ are given, respectively, by

$$A_{ij} = A_{ji} = \begin{pmatrix} 0 & -E_{ij} - E_{ji} \\ 0 & 0 \end{pmatrix}, \tag{22}$$

$$B_{ij} = B_{ji} = \begin{pmatrix} 0 & 0 \\ E_{ij} + E_{ji} & 0 \end{pmatrix}. \tag{23}$$

A realization T of $\mathbf{g}$ that restricts to a unitary realization U of $\mathbf{g}_0 = \mathrm{sp}(N,R)$ is given in terms of the familiar n-particle harmonic oscillator raising and lowering (Weyl) operators by

$$\begin{aligned} T(A_{ij}) &= \sum_{\alpha=1}^{n} b^{\dagger}_{\alpha i} b^{\dagger}_{\alpha j}, \\ T(B_{ij}) &= \sum_{\alpha=1}^{n} b_{\alpha i} b_{\alpha j}, \\ T(C_{ij}) &= \frac{1}{2} \sum_{\alpha=1}^{n} (b^{\dagger}_{\alpha i} b_{\alpha j} + b_{\alpha j} b^{\dagger}_{\alpha i}), \end{aligned} \tag{24}$$

where

$$\begin{aligned} [b_{\alpha i}, b^{\dagger}_{\beta j}] &= \delta_{\alpha\beta}\delta_{ij}, \\ [b_{\alpha i}, b_{\beta j}] &= [b^{\dagger}_{\alpha i}, b^{\dagger}_{\beta j}] = 0. \end{aligned} \tag{25}$$

This realization exponentiates to a unitary realization of the Sp(N,R) group for n even (and, in general, of the twofold metaplectic covering group).

A UIR $\sigma = (\sigma_1, \sigma_2, ..., \sigma_N)$ under this action is defined by a lowest (often called highest) weight state $|\sigma\rangle$ that satisfies

$$\begin{aligned} &T(B_{ij})|\sigma\rangle = 0, \quad i,j = 1,...,N, \\ &T(C_{ij})|\sigma\rangle = 0, \quad i<j, \\ &T(C_{ii})|\sigma\rangle = \sigma_i|\sigma\rangle, \quad i = 1,...,N. \end{aligned} \tag{26}$$

Evidently $|\sigma\rangle$ is a lowest weight state for a UIR u of the subgroup $K_0 = \mathrm{U}(N)$.

It is convenient to define

$$z\cdot E^- = \frac{1}{2} \sum_{ij} z_{ij} B_{ij}, \tag{27}$$

where $z = (z_{ij})$ is a complex symmetric matrix. The Sp(N,R) vector-valued coherent states are then obtained from the general definition

$$\psi(z) = \sum_{\alpha} |\alpha\rangle\, \langle\alpha|\, \exp\left[\frac{1}{2} \sum_{ij} z_{ij} T(B_{ij})\right] |\psi\rangle. \tag{28}$$

To obtain the group action, Eq. (6), observe that any Sp(N,C) matrix can be factored as

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} I & bd^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} \tilde{d}^{-1} & 0 \\ 0 & d \end{pmatrix} \begin{pmatrix} I & 0 \\ d^{-1}c & I \end{pmatrix}. \tag{29}$$

Then, for $g \in \mathrm{Sp}(N,R)$ of the form (17),

$$e^{z\cdot E^-} g = \begin{pmatrix} I & 0 \\ z & I \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \beta^* & \alpha^* \end{pmatrix} \tag{30}$$

can be rearranged

$$\begin{aligned} e^{z\cdot E^-} g = &\begin{pmatrix} I & \beta(z\beta + \alpha^*)^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} (\tilde{\beta}z + \alpha^{\dagger})^{-1} & 0 \\ 0 & z\beta + \alpha^* \end{pmatrix} \\ &\times \begin{pmatrix} I & 0 \\ (z\beta + \alpha^*)^{-1}(z\alpha + \beta^*) & I \end{pmatrix}. \end{aligned} \tag{31}$$

One easily ascertains, using Eq. (18), that $(z\beta + \alpha^*)^{-1} \times (z\alpha + \beta^*)$ is symmetric. Hence we obtain the coherent state action

$$\Gamma(g)\Psi(z) = \rho((\tilde{\beta}z + \alpha^{\dagger})^{-1})\Psi((\tilde{\alpha}z + \beta^{\dagger})(\tilde{\beta}z + \alpha^{\dagger})^{-1}). \tag{32}$$

Note that, on restriction to U(N),

$$\Gamma(\alpha)\Psi(z) = u(\alpha)\Psi(\tilde{\alpha}z\alpha), \quad \alpha \in \mathrm{U}(N). \tag{33}$$

The coherent state realization of the Sp(N,R) algebra was derived in Refs. 2–6. If, for convenience of notation, we define

$$\mathbb{C}_{ij} = \rho(C_{ij}), \tag{34}$$

the coherent state realization is given in general by

$$\begin{aligned} \Gamma(A_{ij}) &= (\mathbb{C}z)_{ij} + (\mathbb{C}z)_{ji} + (z\nabla z)_{ij} - (N+1)z_{ij}, \\ \Gamma(B_{ij}) &= \nabla_{ij}, \\ \Gamma(C_{ij}) &= \mathbb{C}_{ij} + (z\nabla)_{ij}, \end{aligned} \tag{35}$$

where

$$\nabla_{ij} = (1 + \delta_{ij}) \frac{\partial}{\partial z_{ij}}. \tag{36}$$

The practical utility of this realization of the Sp(N,R) algebra for the evaluation of matrix elements lies in its simplicity, which becomes apparent when $\Gamma(A_{ij})$ is expressed in the form

$$\Gamma(A_{ij}) = [\Lambda, z_{ij}], \tag{37}$$

where Λ is the U(N) invariant operator

$$\Lambda = \tfrac{1}{2}\mathrm{Tr}[(\mathbb{C} + z\nabla)(\mathbb{C} + z\nabla)] - \tfrac{1}{4}\mathrm{Tr}(z\nabla z\nabla) - \tfrac{1}{4}(N+1)\mathrm{Tr}(z\nabla), \tag{38}$$

which is diagonal in the natural $\mathrm{Sp}(N,R) \supset \mathrm{U}(N)$ basis. The discovery of Eq. (37) proved to be the vital step which facilitated the use of vector coherent state theory to evaluate matrix elements of the Sp(N,R) algebras for arbitrary discrete series representations.[5–6] Similar expressions hold for other Lie algebras.[20,21]

Finally, to obtain the identity resolution and the inner product, we choose the coset representatives of Eq. (9) by

$$\begin{aligned} \kappa(z) &= \begin{pmatrix} \alpha & \alpha z^* \\ \alpha^* z & \alpha^* \end{pmatrix} \\ &= \begin{pmatrix} I & \alpha z^* \alpha^{-1*} \\ 0 & I \end{pmatrix} \begin{pmatrix} \alpha^{-1\dagger} & 0 \\ 0 & \alpha^* \end{pmatrix} \begin{pmatrix} I & 0 \\ z & I \end{pmatrix}, \end{aligned} \tag{39}$$

with

$$\alpha = \alpha^{\dagger} = (I - z^* z)^{-1/2}. \tag{40}$$

This choice defines the domain of the complex coordinates (z_{ij}) by

$$\det(I - z^* z) > 0 \tag{41}$$

to be the multidimensional unit disk. The identity resolution is now given, from Eq. (14), by

$$\begin{aligned} I = \sum_{\alpha} \int_{\mathrm{Sp}(N,R)/\mathrm{U}(N)} &T(e^{z\cdot E^-})^{\dagger} \rho((I - z^* z)^{1/2})^{\dagger} |\alpha\rangle\, \langle\alpha| \\ &\times \rho((I - z^* z)^{1/2}) T(e^{z\cdot E^-}) d\mu(z) \end{aligned} \tag{42}$$

and the inner product by

$$\langle\Psi|\Phi\rangle = \int_{\mathrm{Sp}(N,R)/\mathrm{U}(N)} (\rho((I-z^*z)^{1/2})\Psi(z), \rho((I-z^*z)^{1/2})\Phi(z)) d\mu(z). \tag{43}$$

To obtain the Sp(N,R) invariant measure $d\mu(z)$ on Sp(N,R)/U(N) it is sufficient to consider the special representations ($\sigma_1 = \cdots = \sigma_N = \sigma$) for which the corresponding representations ($\sigma,\sigma,\ldots,\sigma$) of U($N$) are one dimensional. Then,

$$\rho(I-z^*z)^{1/2}|0\rangle = |0\rangle \det(I-z^*z)^{\sigma/2}. \tag{44}$$

The coherent state wave functions reduce to

$$\Psi(z) = |0\rangle\ \langle 0|T(e^{z\cdot E^-})|\Psi\rangle, \tag{45}$$

with inner product

$$\langle\Psi|\Phi\rangle = \int_{\mathrm{Sp}(N,R)/\mathrm{U}(N)} \langle 0|\Psi(z)\rangle^* \langle 0|\Phi(z)\rangle \det(I-z^*z)^\sigma\, d\mu(z). \tag{46}$$

One way to determine $d\mu(z)$ is to require that the group action should be unitary. This is most easily done infinitesimally by requiring that the Lie algebra has the desired Hermitian adjoint properties, eg., $\Gamma(A_{ij})^\dagger = \Gamma(B_{ij})$, etc. One obtains

$$d\mu(z) = k \det(I-z^*z)^{-(N+1)} \prod_{i>j} dz_{ij}\, dz_{ij}^*, \tag{47}$$

where k is a constant.[4] This simple technique for determining $d\mu(z)$ is consistent with other methods (cf. Sec. V and Ref. 22).

V. REALIZATIONS ON THE SIEGEL HALF-PLANE

We now consider explicitly the relationship between the vector coherent state representations of Sp(N,R) and the discrete series studied previously by Godement[7] and Rosensteel[8] and Rosensteel and Rowe.[9] A unified treatment of both realizations is achieved by recognizing that Sp(N,R)/U(N) is identified in the former case with the generalized unit disk D_N and in the latter case with the upper half-plane S_N.

Set

$$\Lambda \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} I & I \\ iI & -iI \end{pmatrix} \in \mathrm{GL}(2N,C). \tag{48}$$

Then, Sp(N,R) is embedded in U(N,N) via the map

$$\mathrm{Sp}(N,R) \to \mathrm{U}(N,N), \quad M \to \Lambda^\dagger M \Lambda. \tag{49}$$

GL($2N,C$) acts on the $N\times N$ complex matrices $M_N(C)$ as generalized linear fractional transformations according to the rule

$$g\cdot z \equiv (\alpha z+\beta)(\gamma z+\delta)^{-1}, \tag{50}$$

where

$$g = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{GL}(2N,C)$$

and $z\in M_N(C)$. Note that $g_1\cdot(g_2\cdot z) = (g_1 g_2)\cdot z$.

The upper half-plane S_N and the unit disk D_N are given by

$$S_N \equiv \{z\in M_N(C) | \tilde{z} = z,\ \mathrm{Im}\, z > 0\}, \quad D_N \equiv \{w\in M_N(C) | \tilde{w} = w,\ \det(I-w^*w) > 0\}. \tag{51}$$

They are in 1–1 correspondence by the analytic Cayley transformation[10]

$$D_N \to S_N, \quad w \to z = \Lambda\cdot w = i(I+w)(I-w)^{-1}, \tag{52}$$

and its inverse

$$S_N \to D_N, \quad z \to w = \Lambda^\dagger\cdot z = (z-i)(z+i)^{-1}. \tag{53}$$

Moreover, the action of Sp(N,R) on S_N and the action of Sp(N,R)$\supset$U(N,N) on D_N commute:

$$\begin{tikzcd} S_N \arrow[r, "z\to M\cdot z"] \arrow[d, "\Lambda^\dagger"'] & S_N \arrow[d, "\Lambda^\dagger"] \\ D_N & D_N \arrow[l, "w\to\Lambda^\dagger M\Lambda\cdot w"] \end{tikzcd}, \tag{54}$$

where $M\in$Sp(N,R) and $\Lambda^\dagger M\Lambda\in$Sp($N,R$)$\subset$U($N,N$).

Now let ρ be a finite-dimensional irreducible representation of GL(N,C) and let H_ρ denote its carrier space. Define a right (i.e., anti-) representation R_ρ of GL($2N,C$) on the space of holomorphic functions f from $M_N(C)$ into H_ρ:

$$(R_\rho(g)f)(z) \equiv \rho(\gamma z+\delta)^{-1} f(g\cdot z), \tag{55}$$

for

$$g = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{GL}(2N,C)$$

and $z\in M_N(C)$. Note that, for historical reasons, we use here the right representation, for which $R_\rho(g_2)R_\rho(g_1) = R_\rho(g_1 g_2)$. To obtain the standard (left) representation of Eq. (32) one has simply to replace g by $\tilde{g}$; i.e., put $\Gamma(g) = R_\rho(\tilde{g})$.

The irreducible discrete series representations of Sp(N,R) and Sp(N,R)$\subset$U(N,N) are realized on the Hilbert spaces $H_\rho(S_N)$ and $H_\rho(D_N)$, respectively

$$H_\rho(S_N) \equiv \left\{ f: S_N \to H_\rho \,|\, \text{(i)}\, f \text{ is holomorphic, (ii)} \int_{S_N} d\Omega(z) \|\rho(y)^{1/2} f(z)\|^2_{H_\rho} < \infty,\quad y = \mathrm{Im}\, z \right\}, \tag{56}$$

$$H_\rho(D_N) \equiv \left\{ \psi: D_N \to H_\rho \,|\, \text{(i)}\, \psi \text{ is holomorphic, (ii)} \int_{D_N} d\Omega(w) \|\rho(I-w^*w)^{1/2}\psi(w)\|^2_{H_\rho} < \infty \right\}. \tag{57}$$

Here $d\Omega(z)$ and $d\Omega(w)$ are the Sp(N,R) invariant measures on S_N and D_N, respectively,

$$S_N : d\Omega(z) \equiv (\det y)^{-N-1} \prod_{p<q} dx_{pq}\, dy_{pq},$$

$$z_{pq} = x_{pq} + iy_{pq} \in S_N, \tag{58}$$

$$D_N : d\Omega(w) \equiv \det(I - w^* w)^{-N-1} \prod_{p<q} dx_{pq}\, dy_{pq},$$

$$w_{pq} = x_{pq} + iy_{pq} \in D_N. \tag{59}$$

The factors $(\det y)^{-N-1}$ and $\det(I-w^*w)^{-N-1}$, which give the invariant measures in terms of the Euclidean measures, are known as the Bergman kernels.[22]

The irreducible representations of $\mathrm{Sp}(N,R)$ on $H_\rho(S_N)$ (denoted by Π_ρ) and $\mathrm{Sp}(N,R) \subset \mathrm{U}(N,N)$ on $H_\rho(D_N)$ (denoted by Θ_ρ) are just the relevant restrictions of R_ρ,

$$\begin{aligned} &\Pi_\rho(M) f \equiv R_\rho(M) f, \quad M \in \mathrm{Sp}(N,R), \quad f \in H_\rho(S_N), \\ &\Theta_\rho(\Lambda^\dagger M \Lambda)\psi \equiv R_\rho(\Lambda^\dagger M \Lambda)\psi, \\ &\Lambda^\dagger M \Lambda \in \mathrm{Sp}(N,R) \subset \mathrm{U}(N,N), \quad \psi \in H_\rho(D_N). \end{aligned} \tag{60}$$

These two irreducible representations Π_ρ and Θ_ρ are unitarily equivalent. The intertwining is given by the isometry U_ρ,

$$\begin{aligned} &U_\rho : H_\rho(S_N) \mapsto H_\rho(D_N), \\ &U_\rho \equiv R_\rho(\Lambda). \end{aligned} \tag{61}$$

Clearly, U_ρ intertwines Π_ρ and Θ_ρ since all these operators are defined in terms of the representation R_ρ. The inverse transformation is evidently $U_\rho(\Lambda^\dagger)$. Moreover, U_ρ is an isometry:

$$\begin{aligned} &\| U_\rho f \|^2_{H_\rho(D_N)} \\ &\equiv \int_{D_N} d\Omega(w) \| \rho(I - w^* w)^{1/2} (U_\rho f)(w) \|^2_{H_\rho} \\ &= \int d\Omega(w) \| \rho(I - w^* w)^{1/2} \rho\left(\frac{1}{i\sqrt{2}}(I-w)\right)^{-1} \\ &\quad \times f(\Lambda \cdot w) \|^2_{H_\rho}. \end{aligned} \tag{62}$$

Let $z = \Lambda \cdot w \in S_N$. Then, we have the identity

$$\begin{aligned} &[(I - w^* w)^{1/2}(I-w)^{-1}]^\dagger [(I - w^* w)^{1/2}(I-w)^{-1}] \\ &= (I - w^*)^{-1}(I - w^* w)(I-w)^{-1} \\ &= \mathrm{Im}\, z. \\ &= y. \end{aligned} \tag{63}$$

Hence,

$$\begin{aligned} \| U_\rho f \|^2_{H_\rho(D_N)} &= \int_{S_N} d\Omega(z) \| \rho(y)^{1/2} f(z) \|^2_{H_\rho} \\ &= \| f \|^2_{H_\rho(S_N)}, \end{aligned} \tag{64}$$

where the measure has been normalized appropriately.

ACKNOWLEDGMENT

The authors are pleased to acknowledge constructive comments from J. Repka.

[1] G. W. Mackey, *Induced Representations of Groups and Quantum Mechanics* (Benjamin, New York, 1968).
[2] D. J. Rowe, "Coherent states, contractions and classical limits of the noncompact symplectic groups," in *Proceedings of the XIII International Colloquium on Group Theoretical Methods in Physics*, College Park, Maryland, edited by W. W. Zachary (World Scientific, Singapore, 1984).
[3] J. Deenen and C. Quesne, J. Math. Phys. **25**, 2354 (1984).
[4] D. J. Rowe, J. Math. Phys. **25**, 2662 (1984).
[5] D. J. Rowe, G. Rosensteel, and R. Carr, J. Phys. A: Math. Gen. **17**, L399 (1984).
[6] D. J. Rowe, B. G. Wybourne, and P. H. Butler, J. Phys. A: Math. Gen. **18**, 939 (1985).
[7] R. Godement, Seminaire Cartan, Ecole Normale Superieure, Paris, 1958.
[8] G. Rosensteel, Ph.D. thesis, University of Toronto, 1975.
[9] G. Rosensteel and D. J. Rowe, Int. J. Theor. Phys. **1**, 63 (1977).
[10] C. L. Siegel, Am. J. Math. **65**, 1 (1943).
[11] K. Okamoto, Osaka J. Math. **4**, 85 (1967).
[12] V. Bargmann, Ann. Math. **48**, 569 (1947).
[13] Harish-Chandra, Am. J. Math. **78**, 1 (1956).
[14] W. Schmid, Ann. Math. **93**, 1 (1971).
[15] R. T. Moore, Bull. Am. Math. Soc. **71**, 903 (1965).
[16] R. Bott, Ann. Math. **66**, 203 (1957).
[17] P. Griffiths and W. Schmid, Acta. Math. **123**, 253 (1969).
[18] J. Dixmier, *Les C*-Algebres et Leurs Representations* (Gauthier–Villars, Paris, 1964).
[19] S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962).
[20] K. T. Hecht and J. P. Elliot, Nucl. Phys. A **438**, 29 (1985).
[21] R. LeBlanc and D. J. Rowe, "Models for SU(3) in terms of SO(n,2) and SO*(n) algebras," University of Toronto preprint, 1985.
[22] L. K. Hua, *Harmonic Analysis of Functions of Several Complex Variables in the Classical Domains*" (Am. Math. Soc. Translation, Math. Monographs, Providence, RI, 1963).

On the spectrum of a two-level system

Marek Kuś

Department of Physics, University of Arkansas, Fayetteville, Arkansas 72701 and Institute of Theoretical Physics, Warsaw University, Hoza 69, 00-681 Warsaw, Poland[a)]

(Received 4 October 1984; accepted for publication 26 April 1985)

Doubly degenerate energy levels of the two level atom interacting with a single mode of the electromagnetic field are exactly calculated. The dependence of the number of such levels on the values of the level separation energy and a coupling constant is determined. Some general conclusions about the spectrum are drawn.

I. INTRODUCTION

The simplest model used in quantum optics to describe interaction of light and matter is that of a two-level atom coupled to a single quantized mode of the electromagnetic field, for which the Hamiltonian reads

$$H = \tfrac{1}{2}\omega_0\sigma_3 + \omega a^+a + \lambda\sigma_1(a^+ + a). \tag{1.1}$$

Here, the Pauli matrices σ_1, σ_2, σ_3 with the commutation relations $[\sigma_i, \sigma_j] = 2i\sigma_{ijk}\sigma_k$ describe the two atomic levels separated by the energy ω_0, a^+ and a are the Bose operators of the quantized electromagnetic mode with frequency ω, and λ is the atom-field coupling constant proportional to the dipole moment of the transition.

Despite its simplicity, the exact solutions of the two-level model are not known. Instead, many approximate as well as numerical methods for finding energy levels were developed. For example, widely studied approximation (so-called rotating-wave approximation) consists of disregarding in (1.1) the terms σ^+a^+ and σ^-a [where $\sigma^\pm = \frac{1}{2}(\sigma_1 \pm i\sigma_2)$]. This leads to the exactly solvable case known as the Jaynes–Cummings model.[1] Such an approximation can be well justified in a nearly resonant case ($\omega_0 \simeq \omega$) and indeed was successfully adapted in analysis of many fundamental optical processes (see, for example, Ref. 2).

Recently the properties of the Hamiltonian (1.1) attracted increasing interest, since it has been shown that its classical counterpart exhibits a chaotic dynamical behavior.[3,4] The natural question arises whether (1.1) can be treated as a simple example describing the phenomenon of "quantum chaos."[5]

It is believed that such a phenomenon is connected with certain properties of the energy spectrum like irregularities in the level spacings, sensitive dependence on variation of the parameters, etc.,[6] though it should be stressed that there do not exist either definition or criteria of quantum chaos which are widely accepted.

The aim of this paper is to present and investigate some special solutions to the model described by the Hamiltonian (1.1). As a result it will be possible to find all values of the parameters ω_0, ω, and λ for which doubly degenerate energy levels exist. This can give a deeper insight into the structure of all energy levels as well as the dependence of this structure on the involved parameters ω_0, ω, λ. The presented analysis seems to help in the further investigation of possible "quantum chaos" of the model (1.1).

The possibility of finding special, exact solutions of the problem was recently pointed out by Reik, Nusser, and Amarante Ribeiro.[7] The authors made use of the Bargmann representation[8] of boson operators a^+ and a (see next section) and expanded the stationary Schrödinger wave functions in Neumann series. They observed that the series terminates, giving the exact solution of the Schrödinger equation with energy $E = n - \lambda^2$ ($n = 1, 2, \ldots$), every time when a certain condition connecting parameters ω_0, ω, and λ is satisfied. Because all such conditions have the form

$$W_n(\lambda^2,(\omega_0/\omega)^2) = 0, \tag{1.2}$$

where W_n is an n-degree polynomial, it is not obvious that the physical values of parameters [i.e., $\lambda^2 > 0$, $(\omega_0/\omega)^2 > 0$] fulfilling (1.2) can be found. For $n = 1$ and 2, Eqs. (1.2) are simple and can be easily analyzed. For the next several values of n, the existence of the above-described exact solutions was confirmed by the authors of Ref. 7 by the very elegant numerical method which they used to find all low-lying energy levels of (1.1).

In the following part of the present paper I shall investigate carefully the existence of the above-described solutions for an arbitrary n as well as the dependence of the number of these solutions on the parameters ω_0, ω, and λ.

The occurrence of the solutions with the energies $E_n = n - \lambda^2$ can also be deduced from the formula obtained by Schweber,[9] who also used the Bargmann representation and was able to produce a transcendental equation for energy levels. (See Ref. 10 for the generalization to other quantum-optical models).

II. BARGMANN REPRESENTATION, ANALYTICITY OF SOLUTIONS

To simplify further calculations it is convenient to perform the unitary transformation which puts the interaction term $\lambda\sigma_1(a^+ + a)$ into a diagonal form $\lambda\sigma_3(a^+ + a)$ and (without losing generality) change the time scale to obtain $\omega = 1$. After these operations the Hamiltonian reads

$$H(\lambda,\mu) = \mu\sigma_1 + a^+a + \lambda\sigma_3(a^+ + a), \tag{2.1}$$

where $\mu = \frac{1}{2}\omega_0$ and H is labeled by the actual values of the parameters λ and μ.

In what follows we shall extensively use the Bargmann representation of the boson operators.[8] Here a^+ and a are realized as the multiplication and differentiation over the complex variable z ($a^+ \to z$, $a \to d/dz$) and act in the Hilbert

[a)] Permanent address.

 0022-2488/85/112792-04$02.50

space of entire functions of the order less than 2, equipped with a scalar product

$$(f|g) = \frac{1}{\pi}\int \overline{f(z)}g(z)e^{-|z|^2}\,d(\mathrm{Re}\,z)d(\mathrm{Im}\,z). \tag{2.2}$$

The stationary Schrödinger equations for the two-component eigenfunction

$$\psi = \begin{bmatrix}\psi_1(z)\\ \psi_2(z)\end{bmatrix} \tag{2.3}$$

have in this representation the following form:

$$(z+\lambda)\frac{d}{dz}\psi_1 = (E-\lambda z)\psi_1 - \mu\psi_2, \tag{2.4a}$$

$$(z-\lambda)\frac{d}{dz}\psi_2 = (E+\lambda z)\psi_2 - \mu\psi_1, \tag{2.4b}$$

where E is an energy eigenvalue.

Because we are looking for the analytic solutions of (2.4) it is worthwhile to perform the standard Frobenius analysis[11] to investigate analytic properties of ψ_1 and ψ_2 in the singular points $z = \pm\lambda$. Inserting

$$\psi_{1,2} = (z-\lambda)^s \sum_{n=0}^{\infty} c_n^{1,2}(z-\lambda)^n$$

into Eqs. (2.4) we obtain the "indicial equation"

$$s(E+\lambda^2-s) = 0. \tag{2.5}$$

A similar analysis in the point $z = -\lambda$ leads to the identical equation. Because one solution of (2.5) is $s=0$ there always exists one solution $\psi = \begin{bmatrix}\psi_1\\ \psi_2\end{bmatrix}$ analytic in the vicinity of $z = \pm\lambda$. On the other hand, a second linearly independent analytic solution can occur (in consequence the energy level can be degenerate) only when $E = n - \lambda^2$, where n is a non-negative integer.

III. EXACT SOLUTIONS

Let us start with considering several simple properties of the Hamiltonian (2.1) and Eqs. (2.4)

(a) If

$$\psi = \begin{bmatrix}\psi_1(z)\\ \psi_2(z)\end{bmatrix}$$

is a solution of (2.4) then

$$\psi' = \begin{bmatrix}\psi_2(-z)\\ \psi_1(-z)\end{bmatrix}$$

is also a solution. This is an immediate consequence of the fact that there exists a constant of motion $P = \exp(i\pi(a^+a + \frac{1}{2}(\sigma_1+1)))$ and can be proved by inspection. Consequently, eigenfunctions can be classified with respect to their parity, i.e., according to the sign in the equation

$$\begin{bmatrix}\psi_1(z)\\ \psi_2(z)\end{bmatrix} = \pm\begin{bmatrix}\psi_2(-z)\\ \psi_1(-z)\end{bmatrix}.$$

(b) $H(\lambda,0)$ has a doubly degenerate spectrum with energy levels $E_n = n - \lambda^2$, $n = 0,1,2,\ldots$. The same is true for $H(-\lambda,0)$. Indeed, Eqs. (2.4) for $\mu = 0$ reduce to

$$(z+\lambda)\frac{d}{dz}\psi_1 = (E-\lambda z)\psi_1,$$

$$(z+\lambda)\frac{d}{dz}\psi_2 = (E+\lambda z)\psi_2,$$

and have the solutions

$$\psi_1(z) = e^{-\lambda z}(z+\lambda)^{E+\lambda^2},$$
$$\psi_2(z) = e^{\lambda z}(z-\lambda)^{E+\lambda^2}. \tag{3.1}$$

Because solutions should be analytic, $E+\lambda^2$ must be equal to a non-negative integer, i.e., $E = n-\lambda^2$, $n = 0,1,2,\ldots$.

From (3.1) one can construct two linearly independent solutions of opposite parity

$$\Phi_n^+ = \begin{bmatrix}(z+\lambda)^n e^{-\lambda z}\\ (-1)^n(z-\lambda)^n e^{\lambda z}\end{bmatrix},$$
$$\Phi_n^- = \begin{bmatrix}(z+\lambda)^n e^{-\lambda z}\\ (-1)^{n+1}(z-\lambda)^n e^{\lambda z}\end{bmatrix}. \tag{3.2}$$

Similarly,

$$H(-\lambda,0)\widetilde{\Phi}_n^\pm = (n-\lambda)^2\widetilde{\Phi}_n^\pm,$$

where

$$\widetilde{\Phi}_n^+ = \begin{bmatrix}(-1)^n(z-\lambda)^n e^{\lambda z}\\ (z+\lambda)^n e^{-\lambda z}\end{bmatrix},$$
$$\widetilde{\Phi}_n^- = \begin{bmatrix}(-1)^{n+1}(z-\lambda)^n e^{\lambda z}\\ (z+\lambda)^n e^{-\lambda z}\end{bmatrix} \tag{3.3}$$

are the eigenvectors of positive and negative parity, respectively.

(c) Simple calculations show that

$$\sigma_1\Phi_n^+ = \widetilde{\Phi}_n^\pm, \quad \sigma_1\widetilde{\Phi}_n^+ = \Phi_n^+, \tag{3.4}$$

$$H(\lambda,0)\widetilde{\Phi}_n^+ = (n+3\lambda^2)\widetilde{\Phi}_n^+ - 2n\lambda\widetilde{\Phi}_{n-1}^+ - 2\lambda\widetilde{\Phi}_{n+1}^+, \tag{3.5}$$

and quite similar relations hold for Φ_n^- and $\widetilde{\Phi}_n^-$.

Let us now construct the following finite linear combination of the vectors Φ_n^+ and $\widetilde{\Phi}_n^+$:

$$\psi_n^+ = (2\lambda)^n\Phi_n^+ + \mu\sum_{l=1}^{n}\frac{(2\lambda)^{n-l}}{l!} \times P_{l-1}^{n}(\mu\Phi_{n-l}^+ + l\Phi_{n-l}^+), \tag{3.6}$$

where the $P_k^{(n)}$ are defined by the recursion relations

$$P_0^{(n)} = 1, \quad P_1^{(n)} = (2\lambda)^2 + \mu^2 - 1,$$
$$P_k^{(n)} = (k(2\lambda)^2 + \mu^2 - k^2)P_{k-1}^{(n)} - k(k-1)(n-k+1)(2\lambda)^2 P_{k-2}^{(n)}. \tag{3.7}$$

Using identities (3.4) and (3.5) one can establish that

$$H(\lambda,\mu)\psi_n^+ = (H(\lambda,0) + \mu\sigma_1)\psi_n^+ = (n-\lambda^2)\psi_n^+, \quad n = 1,2,\ldots, \tag{3.8}$$

provided that

$$(n(2\lambda)^2 + \mu^2 - n^2)P_{n-1}^{(n)} - n(n-1)(2\lambda)^2 P_{n-2}^{(n)} = 0, \tag{3.9}$$

which can be written in the form

$$P_n^{(n)}((2\lambda)^2,\mu^2) = 0. \tag{3.10}$$

The analogous formula can be constructed for the states of negative parity

$$\psi_n^- = (2\lambda)^n \Phi_n^- + \mu \sum_{l=1}^{n} \frac{(2\lambda)^{n-l}}{l!} P_{l-1}^{(n)} (\mu \Phi_{n-l}^- + l \widetilde{\Phi}_{n-l}^-), \tag{3.11}$$

$$H(\lambda,\mu)\psi_n^- = (n - \lambda^2)\psi_n^- . \tag{3.12}$$

Thus we are able to identify two linearly independent eigenfunctions of $H(\lambda,\mu)$ corresponding to the energy $E_n = n - \lambda^2$. According to the previous analysis only these eigenvalues can be degenerate.

The possibility of constructions (3.6) and (3.11) is limited only by the existence of physical ($\lambda^2 > 0, \mu^2 > 0$) values which satisfy (3.10).

IV. EXISTENCE OF EXACT SOLUTIONS

From the recurrence relations (3.7) one can easily deduce that $P_n(\lambda,\mu)$ is the polynomial of n degree in the variables $(2\lambda)^2$ and μ^2. After establishing the value of μ we ask whether solutions of Eq. (3.10) with positive $(2\lambda)^2$ exist. To simplify the notation let us fix the value of n and denote $(2\lambda)^2$ by x. The formulas (3.7) now have the form

$$\begin{aligned} &Q_0(x) = 1, \quad Q_1(x) = (x - A_1), \\ &Q_k(x) = (x - A_k)Q_{k-1}(x) - B_k x Q_{k-2}(x), \end{aligned} \tag{4.1}$$

where

$$Q_k(x) = (1/k!)P_k^{(n)}(x,\mu^2), \tag{4.2}$$

$$A_k = (k^2 - \mu^2)/k, \quad B_k = (n - k + 1), \tag{4.3}$$

and Q_k depends on μ^2 parametrically.

The following three theorems establish the existence of positive roots of $Q_k(x)$.

Theorem 1: For $0 < \mu < 1$, $Q_k(x)$ has exactly k different, positive roots $a_1^{(k)},\ldots, a_k^{(k)}$; moreover,

$$0 < a_1^{(k)} < a_1^{(k-1)} < a_2^{(k)} < a_2^{(k-1)} < \cdots < a_{k-1}^{(k-1)} < a_k^{(k)}, \tag{4.4}$$

where $a_1^{(k-1)} \cdots a_{k-1}^{(k-1)}$ denote the roots of Q_{k-1}.

Proof: From (4.3) we have $A_k > 0$ for $0 < \mu < 1$ and always $B_k > 0$. From (4.1), $a_1^{(l)} = A_1 > 0$ and $q_2(x) = (x - A_2)(x - a_1^{(l)}) - B_2 x$; thus

$$\begin{aligned} &Q_2(0) = A_2 a_1^{(l)} > 0, \ Q_2(a_1^{(l)}) = -B_2 a_1^{(l)} < 0, \\ &\qquad \operatorname{sgn} Q_2(\infty) = 1, \end{aligned} \tag{4.5}$$

where $\operatorname{sgn} a = -1, 0, +1$ for $a < 0, a = 0, a > 0$, respectively. Relations (4.4) prove that $Q_2(x) = (x - a_2^{(1)})(x - a_2^{(2)})$ and $0 < a_2^{(1)} < a_1^{(1)} < a_2^{(2)}$.

Let us assume that the theorem is valid for $l < k$, i.e.,

$$\begin{aligned} Q_{k-1} &= (x - a_1^{(k-1)} \cdots (x - a_{k-1}^{(k-1)}), \\ Q_{k-2} &= (x - a_1^{(k-2)}) \cdots (x - a_{k-2}^{(k-2)}), \end{aligned}$$

and

$$0 < a_1^{(k-1)} < a_1^{(k-2)} < a_2^{(k-1)} < a_2^{(k-2)} < \cdots < a_{k-2}^{(k-2)} < a_{k-1}^{(k-1)}. \tag{4.6}$$

Then from (4.1) we have

$$\begin{aligned} Q_k &= (x - A_k)(x - a_1^{(k-1)}) \cdots (x - a_{k-1}^{(k-1)}) \\ &\quad - B_k x (x - a_1^{(k-2)}) \cdots (x - a_{k-2}^{(k-2)}). \end{aligned} \tag{4.7}$$

Thus,

$$\operatorname{sgn} Q_k(0) = \operatorname{sgn}(-1)^k A_k a_1^{(k-1)} \cdots a_{k-1}^{(k-1)} = (-1)^k,$$

$$\begin{aligned} &\operatorname{sgn} Q_k(a_i^{(k-1)}) \\ &\quad = -\operatorname{sgn}(B_k a_i^{(k-1)}(a_i^{(k-1)} - a_1^{(k-2)}) \cdots (a_i^{(k-1)} - a_{i-1}^{(k-2)}) \\ &\qquad \times a_i^{(k-1)} - a_i^{(k-2)}) \cdots (a_i^{(k-1)} - a_{k-2}^{(k-2)})) \\ &\quad = (-1)^{k-i}, \end{aligned}$$

$$\operatorname{sgn} Q_k(\infty) = 1.$$

It means that $Q_k(x)$ changes sign on each of the intervals $(0, a_1^{(k-1)}), (a_i^{(k-1)}, a_{i+1}^{(k-1)})$, $i = 1,\ldots, k-2$ and $(a_{k-1}^{(k-1)}, \infty)$, which proves that

$$Q_k(x) = (x - a_1^{(k)}) \cdots (x - a_k^{(k)}),$$

and $a_1^{(k)},\ldots, a_k^{(k)}$ fulfill the inequalities (4.4).

Thus we proved the theorem by induction.

In the exactly analogous way one can prove the following theorem.

Theorem 2: For $1 < \mu < 2$, $Q_k(x)$ has exactly $k - 1$ different roots, i.e., $Q_k(x) = (x - \alpha_k)(x - a_1^{(k)}) \cdots (x - a_{k-1}^{(k)})$, where

$$\alpha_k < 0 < a_1^{(k)} < a_1^{(k-1)} < a_2^{(k)} < a_2^{(k-1)} < \cdots < a_{k-2}^{(k-1)} < a_{k-1}^{(k)},$$

and $a_1^{(k-1)} \cdots a_{k-2}^{(k-1)}$ denote the positive roots of $Q_{k-1}(x)$. To perform the proof by mathematical induction it is sufficient to observe that $A_1 < 0$ and $A_k > 0$, $k = 2,3,\ldots,n$, $\operatorname{sgn} Q_k(-\infty) = (-1)^k$, and $\operatorname{sgn} Q_k(0) = \operatorname{sgn}(-1)^k A_1 \cdots A_k = (-1)^{k+1}$, and repeat the scheme of the preceding proof.

The above theorems can be generalized to the following.

Theorem 3: For $k < \mu < k+1$, $Q_{k+m}(x)$, $m = 1,2,3,\ldots, n-k$ has at least m positive roots.

To prove this statement we observe the following.

(1) $Q_l(0) = (-1)^l A_1 A_2 \cdots A_l$,

thus

$$\operatorname{sgn} Q_l(0) = \operatorname{sgn}(-1)^l A_1 \cdots A_l = (-1)^{2l} = 1, \quad 0 < l < k,$$

$$\begin{aligned} \operatorname{sgn} Q_{k+m}(0) &= \operatorname{sgn}(-1)^{k+m} A_1 \cdots A_k A_{k+1} \cdots A_{k+m} \\ &= (-1)^{k+m}(-1)^k = (-1)^m, \end{aligned}$$

$$m = 1,2,\ldots, n-k.$$

(2) If $a > 0$ and $Q_i(a) = 0$, then $Q_{i+1}(a)$ and $Q_{i+1}(a)$ are of opposite sign. This follows immediately from the basic recurrence relations (4.1).

(3) If $Q_i(a) = 0$ for $a \neq 0$ then $Q_{i-1}(a) \neq 0$. Indeed, assuming $Q_i(a) = 0$ and $Q_{i-1}(a) = 0$, we obtain

$$Q_{i-2}(a) = (l/B_i a)(Q_i(a) - (a - A_i)Q_{i-1}(a)) = 0,$$

and consequently, by the continuation of this procedure,

$$Q_{i-3}(a) = Q_{i-4}(a) = \cdots Q_0(a) = 0,$$

which contradicts the definition of $Q_0(x)$.

The proof of Theorem 2 is now analogous to a proof of the Sturm theorem.[12]

Let us denote by $V(x)$ the number of changes in sign in the sequence

$$Q_{k+m}(x), Q_{k+m-1}(x), \ldots, Q_0(x),$$

for fixed $x \geqslant 0$.

When x is varied from 0 to ∞, $V(x)$ can change its value only when x passes through a root of the functions Q_l,

$l = 0,1, \ldots, k+m$. As a consequence of (2) and (3) above, $V(x)$ cannot change when x passes a root of $Q_l, l = 1,2, \ldots, k+m-1$. When x passes through a root of Q_{k+m}, $V(x)$ has a change of ± 1. On the other hand,

$$Q_{k+m}(0), Q_{k+m-1}(0), \ldots, Q_{k+1}(0), Q_k(0), \ldots, Q_0(0)$$

have signs

$$(-1)^m, (-1)^{m-1}, \ldots, -1, 1, \ldots, 1,$$

which means that $V(0) = m$, while $\operatorname{sgn} Q_l(\infty) = 1$, i.e., $V(\infty) = 0$. Thus, when x passes from 0 to ∞, $V(x)$ changes from m to 0, which indicates the existence of at least m positive roots of $Q_{k+m}(x)$.

From the definition (4.2) it is obvious that $Q_n(x)$ has exactly the same roots as $P_n^{(n)}(x,\mu^2)$, thus we have established that for $0 < \mu < 1$ there exist μ different values of μ for which eigenfunctions (3.6) and (3.11) can be constructed. For $1 < \mu < 2$ there exist $n-1$ such values (and generally for $k < \mu < k+1$ at least $n-k$ of them).

V. CONCLUDING REMARKS

The situation is particularly simple when $0 < \mu < 1$ (which includes the most interesting resonant case $\mu = \frac{1}{2}$) and from the existence of the solutions (3.6) we can draw some conclusions about the whole energy spectrum. If, for an established value of μ, we plot the energy versus the squared coupling constant the picture will have the following characteristic features.

(1) The neighboring energy levels will cross on the parallel straight lines $E = n - \lambda^2$.

(2) For each n there will be n crossings.

(3) There will be no crossings not lying on one of these lines.

(4) For $\lambda \to \infty$ two neighboring levels will tend to the line $E = n - \lambda^2$ corresponding to the doubly degenerate level of unperturbed Hamiltonian $H(\lambda,0)$(here μ is treated as a perturbation parameter).

Moreover, the changes of parameter μ (until $0 < \mu < 1$) will not disturb the above picture (only the positions of the crossings can change). This observation seems to have a deeper meaning for the directions of investigation of quantum chaos in the model: it suggests that the spectrum can be rather insensitive on the changes of the perturbation parameter μ. Although a careful numerical analysis is undoubtedly needed it seems that "quantum chaotic" behavior cannot be observed in the resonance.

ACKNOWLEDGMENTS

The author would like to acknowledge Professor P. W. Milonni for his invitation and warm hospitality in the University of Arkansas.

This research was supported by National Science Foundation (NSF) Grant PHY-8308048, NSF EPSCOR Grant No. ISP-8011447, and the State of Arkansas.

[1]E. T. Jaynes and F. W. Cummings, Proc. IEEE **51**, 126 (1963).
[2]L. Allen and J. H. Eberly, *Optical Resonance and Two Level Atoms* (Wiley, New York, 1975).
[3]P. I. Beloborov, G. M. Zaslavskii, and G. Kh. Tartakowski, Zh. Eksp. Teor. Fiz. **71**, 1799 (1976).
[4]P. W. Milonni, J. R. Ackerhalt, and H. W. Galbraith, Phys. Rev. Lett. **50**, 966 (1983).
[5]R. Graham and M. Höhnerbach, Phys. Lett. A **101**, 61 (1984).
[6]L. C. Percival, J. Phys. B **6**, L229 (1973).
[7]H. G. Reik, H. Nusser, and L. A. Amarante Ribeiro, J. Phys. A **15**, 3431 (1982).
[8]V. Bargmann, Commun. Pure Appl. Math. **14**, 187 (1961).
[9]S. Schweber, Ann. Phys. (NY) **41**, 205 (1967).
[10]M. Kuś and M. Lewenstein, submitted to J. Phys. B.
[11]E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956).
[12]J. V. Uspensky, *Theory of Equations* (McGraw–Hill, New York, 1948), p. 142.

Spectral sum rule for time delay in $\mathbb{R}^2$

T. A. Osborn and K. B. Sinha[a]
Department of Physics, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada
D. Bollé[b] and C. Danneels
Instituut voor Theoretische Fysica, Universiteit Leuven, B-3030 Leuven, Belgium

(Received 3 May 1985; accepted for publication 17 June 1985)

A local spectral sum rule for nonrelativistic scattering in two dimensions is derived for the potential class $v \in L^{4/3}(\mathbb{R}^2)$. The sum rule relates the integral over all scattering energies of the trace of the time-delay operator for a finite region $\Sigma \subset \mathbb{R}^2$ to the contributions in Σ of the pure point and singularly continuous spectra.

I. INTRODUCTION

Spectral sum rules involving the time delay for a region Σ of finite volume and the bound-state density for the same region were derived in Ref. 1 in the context of classical scattering. Here we rigorously derive the quantum mechanical counterpart of these local sum rules in two Euclidean dimensions (see Theorem 4).

We consider the quantum mechanics of a single spinless particle in two dimensions. The state space is a Hilbert space $\mathscr{H} \equiv L^2(\mathbb{R}^2)$, in which K_0 denotes the self-adjoint extension of $-\Delta$ describing the free Hamiltonian of the particle (with $\hbar^2 = 2m = 1$). We shall assume that the potential v, describing the interaction, is a measurable function in $L^{4/3}(\mathbb{R}^2)$.

The total Hamiltonian $H = K_0 + v$ will be defined by the quadratic form method[2] and we write $\mathscr{H}_{ac}(H)$ and $\mathscr{H}_s(H)$ for the absolutely continuous and singular spectral subspaces respectively for the self-adjoint operator H. Also R_z, $\rho(H)$, and E [$R_z^0, \rho(K_0)$, and E^0, respectively] will denote the resolvent, the resolvent set, and the spectral measure, respectively, for H (for K_0). The symbols $\mathscr{B}$, $\mathscr{B}_0$, $\mathscr{B}_2$, and $\mathscr{B}_1$ denote the linear spaces of all bounded, compact, Hilbert–Schmidt, and trace-class operators in $\mathscr{H}$ with $\|\cdot\|$, $\|\cdot\|_2$, and $\|\cdot\|_1$ denoting the operator, Hilbert–Schmidt, and trace norms, respectively. We also set $\mathscr{B}_4 = \{A \in \mathscr{B}_0 | A^* A \in \mathscr{B}_2\}$. Then one has $\mathscr{B}_1 \subseteq \mathscr{B}_2 \subseteq \mathscr{B}_4 \subseteq \mathscr{B}_0 \subseteq \mathscr{B}$. We shall use the factorization scheme $u(x) = |v(x)|^{1/2}$, $w(x) = \operatorname{sgn} v(x) u(x)$, so that $u, w \in L^{8/3}(\mathbb{R}^2)$. The first theorem collects the results relating to the definition of H.

Theorem 1: Let $v \in L^{4/3}(\mathbb{R}^2)$.

(a) For every $\chi^2 > 0$, $u(K_0 + \chi^2)^{-1/2}$ and $w(K_0 + \chi^2)^{-1/2}$ belong to $\mathscr{B}_4$.

(b) The total Hamiltonian $H = K_0 + v$, defined as a quadratic form on $D(K_0^{1/2})$, the domain of $K_0^{1/2}$, can be extended as the quadratic form of a self-adjoint operator, also denoted by H, which is bounded below. Also, $D(|H|^{1/2}) = D(K_0^{1/2})$.

(c) For every $z \in \mathbb{C} - \{0\}$, the integral kernel $A(z)(x,y) \equiv u(x) R_z^0(x,y) w(y)$ defines a $\mathscr{B}_2$ operator, also denoted $A(z)$, which is $\mathscr{B}_2$ holomorphic in the open upper- and lower-half planes separately.

(d) $\|A(z)\|_2 \to 0$ as $|z| \to \infty$, and $A(z)$ has boundary values in $\mathscr{B}_2$ norm as $z \to \lambda \pm i0$, uniformly for λ in every closed subset of $\mathbb{R} - \{0\}$.

(e) For $z \in \rho(H) \cap \rho(K_0)$, $[1 + A(z)]^{-1} \in \mathscr{B}$ and one has the second resolvent equation

$$R_z - R_z^0 = -R_z^0 w [1 + A(z)]^{-1} u R_z^0. \tag{1}$$

Furthermore, the function $z \to [1 + A(z)]^{-1}$ is $\mathscr{B}$ holomorphic in the open upper- and lower-half planes.

Since many of the calculations are standard we only sketch the proof.

Proof: The Green's function for the free Hamiltonian is $R_z^0(x,y) = (i/4) H_0^{(1)}(\sqrt{z}|x-y|)$, where $H_0^{(1)}$ is the Hankel function of the first kind, and where we have chosen the branch of the square root so that $\operatorname{Im}\sqrt{z} > 0$. Using the bound

$$|H_0^{(1)}(\alpha)| < c_0 |\alpha|^{-1/2} e^{-\operatorname{Im}\alpha}, \tag{2}$$

for all $\alpha \in \mathbb{C} - \{0\}$ with $\operatorname{Im}\alpha \geqslant 0$ (see Ref. 3, pp. 962 and 963), we have that for $z \in \mathbb{C} - \{0\}$,

$$\|A(z)\|_2 = \frac{1}{16} \int\int dx\, dy |u(x)|^2 |H_0^{(1)}(\sqrt{z}|x-y|)|^2 |w(y)|^2 \leqslant \frac{c_0^2}{16|z|^{1/2}} \int\int dx\, dy \frac{|v(x)|\,|v(y)|}{|x-y|} \leqslant \frac{c\|v\|_{4/3}^2}{|z|^{1/2}}, \tag{3}$$

by an application of the Sobolev inequality[4] in $\mathbb{R}^2$. This proves (a) and parts of (c) and (d). The $\mathscr{B}_2$ holomorphy of $A(z)$ follows by writing

$$A(z) = u(K_0 + \chi^2)^{-1/2} [I + (z + \chi^2) R_z^0] \times (w(K_0 + \chi^2)^{-1/2})^*,$$

and observing that while the middle factor is clearly $\mathscr{B}$ holomorphic, each of the other two are $\mathscr{B}_4$.

Part (b) follows from (a) on using standard results on quadratic forms.[2,5,6] The existence of boundary values uniformly in λ is the consequence of an application of the dominated convergence theorem and the estimate (3). The resolvent equation (1) can be established as in Refs. 2 or 7. □

Scattering theory for such a system can be developed along standard lines and the next theorem summarizes the results.

Theorem 2: Let $v \in L^{4/3}(\mathbb{R}^2)$. Set $\mathscr{E} = \{0\} \cup \{\lambda \in \mathbb{R} \setminus \{0\} | I + A(\lambda + i0)$ or $I + A(\lambda - i0)$ is not 1–1$\}$.

(a) $\mathscr{E}$ is a closed and bounded set of Lebesgue measure 0.

(b) u is K_0 bounded and w is H bounded. Furthermore, uE_Δ^0 and wE_Δ are K_0 and H smooth, respectively (see Refs. 5

[a] On leave from the Indian Statistical Institute, New Delhi, 110016, India.
[b] Onderzoeksleider N. F. W. O., Belgium.

and 8 for the definition of smoothness), where Δ is any half-open interval in $R - \mathscr{E}$.

(c) The wave operators

$$\Omega_{\pm} \equiv \operatorname*{s-lim}_{t \to \pm \infty} e^{+iHt} e^{-iK_0 t},$$

and

$$\Omega'_{\pm} E_{R-\mathscr{E}} \equiv \operatorname*{s-lim}_{t \to \pm \infty} e^{iK_0 t} e^{-iHt} E_{R-\mathscr{E}}$$

exist. The scattering system defined by the pair (H, K_0) is asymptotically complete, i.e.,

$$\text{Range } \Omega_+ = \text{Range } \Omega_- = \mathscr{H}_{\text{ac}}(H) = E_{R-\mathscr{E}} \mathscr{H},$$

and $\mathscr{H}_s(H) \subseteq E_{\mathscr{E}} \mathscr{H}$.

Sketch of the proof: As in Ref. 5, p. 364, the observation that $\|A(\lambda \pm i0)\|_2 \to 0$ as $|\lambda| \to \infty$ and an application of the analytic Fredholm theorem gives us (a). For (b), we use the resolvent equation (1) and note that $\|[I + A(\lambda \pm i\eta)]^{-1}\|$ is bounded, uniformly for $\lambda \in \Delta$ and $0 < \eta < 1$. The part (c) then follows by an application of Kato–Lavine theory (Proposition 9.16 in Ref. 5). □

II. TIME DELAY AND A TRACE THEOREM

Following the reasoning in Ref. 9 we see that the expression

$$\tau(f,\Sigma) \equiv \int_{-\infty}^{\infty} (f, e^{iK_0 t} [\Omega_+^* P_\Sigma \Omega_+ - P_\Sigma] e^{-iK_0 t} f) dt$$

formally describes the time delay in the state $f \in \mathscr{H}$ for the region $\Sigma \subseteq \mathbf{R}^2$, where we have written P_Σ for the orthogonal projection defined by multiplication with the characteristic function χ_Σ.

Let $\mathscr{H}_0 \equiv L^2(T)$, with $(\cdot,\cdot)_0$ denoting the inner product and where T is the unit circle embedded in $\mathbf{R}^2$, and let $\mathscr{U}_0$: $L^2(\mathbf{R}^2) \mapsto L^2([0,\infty), \mathscr{H}_0)$ be the spectral transformation (see Ref. 5 for details) for the free Hamiltonian K_0 so that $(\mathscr{U}_0 K_0 f)_\lambda = \lambda (\mathscr{U}_0 f)_\lambda$ for a.a. $\lambda \in [0,\infty)$ and for $f \in D(K_0)$. Then one has the following theorem describing the properties of τ (see Theorem 2 in Ref. 9).

Theorem 3: Let K_0 and H be as described in Theorem 1, and let Σ be a measurable subset of $\mathbf{R}^2$ with finite Lebesgue measure, i.e., $|\Sigma| < \infty$. Then we have the following.

(a) $P_\Sigma R_z^0$ and $P_\Sigma R_z$ are both $\mathscr{B}_2$ operators for every $z \in \rho(H) \cap \rho(K_0)$.

(b) Set $\mathscr{D}_0 \equiv \{ f \in \mathscr{H} | \lambda \to \|(\mathscr{U}_0 f)_\lambda\|_0$ is a bounded function of bounded support in $[0,\infty)\}$. Then $\mathscr{D}_0$ is dense in $\mathscr{H}$ and there exists a unique measurable family $Q(\lambda,\Sigma)$ of trace-class operators in $\mathscr{H}_0$, interpreted as the energy-shell time-delay operator, such that

$$\tau(f,\Sigma) = \int_0^\infty ((\mathscr{U}_0 f)_\lambda, Q(\lambda,\Sigma)(\mathscr{U}_0 f)_\lambda)_0 \, d\lambda,$$

for every $f \in \mathscr{D}_0$.

(c) Denoting $q(\lambda,\Sigma) = \text{tr}_0 \, Q(\lambda,\Sigma)$, the trace of $Q(\lambda,\Sigma)$ in $\mathscr{H}_0$, one has furthermore that

$$\int_0^\infty \frac{|q(\lambda,\Sigma)|}{\lambda^2 + 1} d\lambda < \infty, \tag{4}$$

$$\int_0^\infty \frac{q(\lambda,\Sigma)}{|\lambda - z|^2} d\lambda = 2\pi \, \text{tr} \, R_z^{0*} [\Omega_+^* P_\Sigma \Omega_+ - P_\Sigma] R_z^0, \tag{5}$$

and

$$\frac{1}{2\pi} \int_0^\infty q(\lambda,\Sigma) \text{Im} \frac{1}{\lambda - z} d\lambda = \text{tr} \, P_\Sigma \, \text{Im} \, (R_z E_{\text{ac}} - R_z^0) P_\Sigma, \tag{6}$$

for every $z \in \rho(H) \cap \rho(K_0)$.

The function $q(\lambda,\Sigma)$ is interpreted as the average time-delay function of energy λ for the region Σ.

Proof: Since $|\Sigma| < \infty$, it is easy to see that $P_\Sigma R_z^0 \in \mathscr{B}_2$. This combined with (1) proves that $P_\Sigma R_z \in \mathscr{B}_2$. Part (b) is proved as in Ref. 9 by using the intertwining relation and noting that

$$R_z^{0*} [\Omega_+^* P_\Sigma \Omega_+ - P_\Sigma] R_z^0 = \Omega_+^* R_z^* P_\Sigma R_z \Omega_+ - R_z^{0*} P_\Sigma R_z^0$$

is a $\mathscr{B}_1$ operator. Equations (4) and (5) are consequences of this as in Ref. 9.

Using the cyclicity of the trace, the asymptotic completeness of Ω_+, and the resolvent equation, we write

$$\begin{aligned} \text{tr} \, R_z^{0*} & [\Omega_+^* P_\Sigma \Omega_+ - P_\Sigma] R_z^0 \\ &= \text{tr} \, R_z^* P_\Sigma R_z \Omega_+ \Omega_+^* - \text{tr} \, R_z^{0*} P_\Sigma R_z^0 \\ &= \text{tr} \, P_\Sigma (R_z E_{\text{ac}} R_{\bar z} - R_z^0 R_{\bar z}^0) P_\Sigma \\ &= (z - \bar z)^{-1} \, \text{tr} \, P_\Sigma \{ (R_z - R_{\bar z}) E_{\text{ac}} - (R_z^0 - R_{\bar z}^0) \} P_\Sigma, \end{aligned}$$

which leads to (6). □

III. SUM RULE

A spectral sum rule for the time delay $q(\cdot,\Sigma)$ is derived in this section. It is convenient to introduce a standard notation[6] for the Fourier transform that maps $L^q(\mathbf{R}^n)$ $(1 < q < 2)$ into its conjugate space $L^p(\mathbf{R}^n)$ $(p^{-1} + q^{-1} = 1)$. The Fourier image of an element $f \in L^q(\mathbf{R}^n)$ will be denoted by $\tilde f \in L^p(\mathbf{R}^n)$. With this notation our main result may be stated as follows.

Theorem 4 (Spectral Sum Rule): (i) Suppose $v \in L^{4/3}(\mathbf{R}^2)$ and let Σ be a measurable subset of $\mathbf{R}^2$ with finite Lebesgue measure, i.e., $|\Sigma| < \infty$.

(ii) Assume, furthermore, that $\tilde v \tilde\chi_\Sigma \in L^1(\mathbf{R}^2)$. Then the function $q(\cdot,\Sigma)$: $[0,\infty) \mapsto R$ has a finite improper integral

$$\int_0^\infty q(\lambda,\Sigma) \, d\lambda \equiv \lim_{n \to \infty} \int_0^n q(\lambda,\Sigma) \, d\lambda,$$

which satisfies

$$\int_0^\infty q(\lambda,\Sigma) \, d\lambda = -2\pi \, \text{tr} \, P_\Sigma E_s P_\Sigma - \frac{1}{2} \int_\Sigma v(x) \, dx. \tag{7}$$

Theorem 4 is demonstrated by breaking the proof into three propositions. The basic idea is to apply Cauchy's integral theorem to the holomorphic function $z \to \text{tr} \, P_\Sigma (R_z - R_z^0) P_\Sigma + \text{tr} \, P_\Sigma R_z^0 v R_z^0 P_\Sigma$ on a suitable contour in $\rho(H) \cap \rho(K_0)$. Proposition 5 determines the real axis contribution of $\text{tr} \, P_\Sigma (R_z - R_z^0) P_\Sigma$. The second factor proportional to v is the Born term and its real axis contribution is found in Proposition 6. Finally the large radius contribution of both terms to the Cauchy integral is described in Proposition 9. In Propositions 5, 6, and 9 the set Σ is defined to be a measurable subset of $\mathbf{R}^2$.

Before proceeding to these propositions it is helpful to identify the region in z (the complex energy plane) where Born dominance prevails. Let Π be the canonically cut plane composed of the complex plane with the non-negative reals removed. Theorem 1 (d) shows that $A(z), z\in\Pi$, has $\mathscr{B}_2$-norm continuous extensions to the real axis from either above or below. For positive reals these two extensions are different. Take Π_c to be the closure of the canonically cut plane which maintains the distinction between the two possible boundary values along the positive real axis. The large z bound for $\|A(z)\|_2$ allows the following definition of $\Lambda_\theta < \infty$.

Definition: For each $\theta\in(0,1)$, let Λ_θ be the infinum of the set

$$\{\Lambda\in R^+ \mid \|A(z)\|_2\leqslant\theta<1,\quad \forall z\in\Pi_c \text{ with } |z|>\Lambda\}.$$

In the Born dominant region of Π_c, i.e., $|z|>\Lambda_\theta$, it is evident that $[1+A(z)]^{-1}$ is a bounded operator on $\mathscr{H}$ and has norm bound $\|[1+A(z)]^{-1}\|\leqslant(1-\theta)^{-1}$. Thus for each $\theta\in(0,1)$, $\mathscr{E}$ is contained in $[-\Lambda_\theta, \Lambda_\theta]$. Our first proposition describes the behavior of $\operatorname{tr} P_\Sigma \operatorname{Im}[R_{\lambda+i0} - R^0_{\lambda+i0}] P_\Sigma$ on the finite intervals of the real axis that contain $\mathscr{E}$.

Proposition 5: Suppose $v\in L^{4/3}(\mathbb{R}^2)$ and $|\Sigma|<\infty$. For every finite interval $(a,b)\supset[-\Lambda_\theta,\Lambda_\theta]\supset\mathscr{E}$, $1>\theta>0$,

$$\lim_{\delta\to0^+}\int_a^b d\lambda \operatorname{tr} P_\Sigma \operatorname{Im}[R_{\lambda+i\delta} - R^0_{\lambda+i\delta}] P_\Sigma$$
$$=\frac{1}{2}\int_0^b q(\lambda,\Sigma)\,d\lambda + \pi \operatorname{tr} P_\Sigma E_s P_\Sigma. \tag{8}$$

Proof: Take $\delta>0$. Theorem 2 (a) and the resolvent equation (1) for R_z implies that $P_\Sigma \operatorname{Im} R_{\lambda+i\delta} P_\Sigma\in\mathscr{B}_1$. The spectral decomposition of $\mathscr{H}=\mathscr{H}_{ac}\oplus\mathscr{H}_s$ (with the associated orthogonal projectors E_{ac} and E_s) leads to

$$\operatorname{tr} P_\Sigma \operatorname{Im} R_{\lambda+i\delta} P_\Sigma$$
$$=\operatorname{tr} P_\Sigma \operatorname{Im} R_{\lambda+i\delta} E_{ac} P_\Sigma + \operatorname{tr} P_\Sigma \operatorname{Im} R_{\lambda+i\delta} E_s P_\Sigma.$$

Thus (for $\delta>0$) the left-hand-side integral in (8) is the sum $I_{ac}+I_s$, where

$$I_{ac}(\delta)=\int_a^b d\lambda \operatorname{tr} P_\Sigma \operatorname{Im}[R_{\lambda+i\delta}E_{ac} - R^0_{\lambda+i\delta}] P_\Sigma,$$

$$I_s(\delta)=\int_a^b d\lambda \operatorname{tr} P_\Sigma \operatorname{Im} R_{\lambda+i\delta} E_s P_\Sigma.$$

First, consider the $\delta\to0^+$ limit of $I_{ac}(\delta)$. Theorem 3, Eq. (6), gives us the representation

$$I_{ac}(\delta)=\frac{1}{2\pi}\int_a^b d\lambda\int_0^\infty d\mu \frac{\delta}{(\mu-\lambda)^2+\delta^2}q(\mu,\Sigma). \tag{9}$$

The elementary $d\lambda$ integral can be written in either of two equivalent forms:

$$\int_a^b d\lambda\frac{\delta}{(\mu-\lambda)^2+\delta^2}=\tan^{-1}\frac{b-\mu}{\delta}+\tan^{-1}\frac{\mu-a}{\delta}$$
$$=\tan^{-1}\frac{\delta(b-a)}{(\mu-\mu_+)(\mu-\mu_-)}. \tag{10}$$

If $2\delta<b-a$ the roots $\mu_\pm$ are real, given by $2\mu_\pm=b+a\pm[(b-a)^2-4\delta^2]^{1/2}$, and always fall inside (a,b). Specifically, $\mu_+=b-\epsilon_+$ and $\mu_-=a+\epsilon_-$, where $\epsilon_\pm\to0^+$ as $\delta\to0$. The inequality $|\tan^{-1}\mu|\leqslant|\mu|$ and estimate (4) suffices to establish that the double integral in (9) is absolutely convergent. Fubini's theorem allows a change of integration order whereby (9) becomes

$$I_{ac}(\delta)=\frac{1}{2\pi}\int_0^\infty d\mu\, q(\mu,\Sigma)$$
$$\times\left[\tan^{-1}\frac{b-\mu}{\delta}+\tan^{-1}\frac{\mu-a}{\delta}\right]. \tag{11}$$

Treating the $\mu>2b$ and the $\mu<2b$ contributions to integral (11) separately leads to the construction of a δ-independent $L^1(d\mu)$ majorant. For $0<\delta<1$ and $\mu>2b$ a majorizing function is $|q(\mu,\Sigma)|(b-a)[(\mu-b)(\mu-a)]^{-1}$, whereas for $0\leqslant\mu\leqslant2b$ the bounding function is $\pi|q(\mu,\Sigma)|$. Theorem 3, estimate (4), confirms that this majorant is $L^1(d\mu)$. Dominated convergence now applies to (11) yielding

$$\lim_{\delta\to0^+} I_{ac}(\delta)=\frac{1}{2}\int_0^b d\mu\, q(\mu,\Sigma). \tag{12}$$

It remains to investigate the limit of $I_s(\delta)$. A useful intermediate result is the following. Suppose $\{C_n\}$ is a sequence of operators in $\mathscr{B}$ converging strongly to C. If $A,B\in\mathscr{B}_2$, then

$$\lim_{n\to\infty}\operatorname{tr} AC_nB=\operatorname{tr} ACB. \tag{13}$$

(See Ref. 5, Lemma 8.23.)

Recall $\sigma_s(H)\subseteq\mathscr{E}\subset[a,b]$. Since $E_{[a,b]}P_\Sigma\in\mathscr{B}_2$, it follows that $E_sP_\Sigma\in\mathscr{B}_2$. The function $[a,b]\ni\lambda\to\operatorname{Im} R_{\lambda+i\delta}\in\mathscr{B}$ is $\mathscr{B}$-norm continuous (for $\delta>0$) and has a $\mathscr{B}$-valued strong Riemann integral on $[a,b]$. Likewise the map $[a,b]\ni\lambda\to P_\Sigma E_s \operatorname{Im} R_{\lambda+i\delta} E_s P_\Sigma\in\mathscr{B}_1$ is $\mathscr{B}_1$-norm continuous and so $\lambda\to\operatorname{tr} P_\Sigma E_s \operatorname{Im} R_{\lambda+i\delta} E_s P_\Sigma$ has an ordinary Riemann integral on $[a,b]$. By the definition of these two integrals, the linearity of the trace, and (13) it follows that

$$I_s(\delta)=\operatorname{tr} P_\Sigma E_s\left[\int_a^b d\lambda \operatorname{Im} R_{\lambda+i\delta}\right]E_s P_\Sigma. \tag{14}$$

Neither a nor b are eigenvalues of H. The strong Riemann integral of $\operatorname{Im} R_{\lambda+i\delta}$ gives the standard result (Ref. 5, p. 360)

$$\operatorname*{s-lim}_{\delta\to0^+}\int_a^b d\lambda \operatorname{Im} R_{\lambda+i\delta}=\pi E_{[a,b]}. \tag{15}$$

A second application of (13) together with (15) controls the $\delta\to0^+$ limit of (14),

$$\lim_{\delta\to0^+} I_s(\delta)=\pi\operatorname{tr} P_\Sigma E_s E_{[a,b]} E_s P_\Sigma=\pi\operatorname{tr} P_\Sigma E_s P_\Sigma. \quad\square$$

Proposition 6: Let $v\in L^{4/3}(\mathbb{R}^2)$ and suppose that (i) $|\Sigma|<\infty$, and (ii) $\tilde{v}\tilde{\chi}_\Sigma\in L^1(\mathbb{R}^2)$. Then for $a<-\Lambda_\theta$ with $\theta\in(0,1)$,

$$\lim_{b\to\infty}\lim_{\delta\to0^+} I(b,\delta)$$
$$\equiv\lim_{b\to\infty}\lim_{\delta\to0^+}\int_a^b d\lambda \operatorname{tr} P_\Sigma \operatorname{Im}(R^0_{\lambda+i\delta}vR^0_{\lambda+i\delta})P_\Sigma$$
$$=-\frac{1}{4}\int_\Sigma v(x)dx.$$

Proof: Note that as in the proof of Theorem 1, Eq. (3), we have $\|P_\Sigma R^0_{\lambda\pm i\delta}u\|_2\leqslant c|\lambda|^{-1/4}$ (c independent of λ and δ),

$$2i\,\mathrm{tr}\,P_\Sigma\,\mathrm{Im}\,(R^0_{\lambda+i\delta}\,vR^0_{\lambda+i\delta})\,P_\Sigma$$
$$\to\mathrm{tr}\,(P_\Sigma\,R^0_{\lambda+i0}\,vR^0_{\lambda+i0}\,P_\Sigma-P_\Sigma\,R^0_{\lambda-i0}\,vR^0_{\lambda-i0}\,P_\Sigma),$$

for every $\lambda\neq0$ as $\delta\to0^+$, and that $P_\Sigma\,R^0_{\lambda+i0}\,vR^0_{\lambda+i0}\,P_\Sigma=P_\Sigma\,R^0_{\lambda-i0}\,vR^0_{\lambda-i0}\,P_\Sigma$, for $\lambda<0$. Thus by an application of the dominated convergence theorem,

$$\begin{aligned}2i\,I(b,0^+)&\equiv2i\lim_{\delta\to0^+}I(b,\delta)\\&=\int_0^b d\lambda\,\mathrm{tr}\,(P_\Sigma\,R^0_{\lambda+i0}\,vR^0_{\lambda+i0}\,P_\Sigma\\&\quad-P_\Sigma\,R^0_{\lambda-i0}\,vR^0_{\lambda-i0}\,P_\Sigma).\end{aligned}\tag{16}$$

Upon writing

$$\begin{aligned}&P_\Sigma\,R^0_{\lambda+i0}\,vR^0_{\lambda+i0}\,P_\Sigma-P_\Sigma\,R^0_{\lambda-i0}\,vR^0_{\lambda-i0}\,P_\Sigma\\&\quad=P_\Sigma\,R^0_{\lambda+i0}\,v(R^0_{\lambda+i0}-R^0_{\lambda-i0})\,P_\Sigma\\&\qquad+P_\Sigma\,(R^0_{\lambda+i0}-R^0_{\lambda-i0})vR^0_{\lambda-i0}\,P_\Sigma,\end{aligned}$$

and observing that the trace of the product of two $\mathscr{B}_2$ operators can be evaluated as the iterated integral of the associated L^2 kernels (see Ref. 10, p. 524), one has that the integrand in (16) is

$$\int dx\int dy\,\chi_\Sigma(x)[R^0_{\lambda+i0}+R^0_{\lambda-i0}](x,y)$$
$$\times v(y)[R^0_{\lambda+i0}-R^0_{\lambda-i0}](x,y)\chi_\Sigma(x),$$

where we have also used the fact that $R^0_{\lambda\pm i0}(x,y)=R^0_{\lambda\pm i0}(y,x)$.

Note that $R^0_{\lambda+i0}(x,y)=(i/4)\,H_0^{(1)}(\sqrt{\lambda}\,|x-y|)$ for $\lambda>0$ and then the choice of the branch of $\sqrt{z}$ leads to $R^0_{\lambda-i0}(x,y)=-(i/4)\,H_0^{(2)}(\sqrt{\lambda}\,|x-y|)$, so that

$$[R^0_{\lambda+i0}+R^0_{\lambda-i0}](x,y)=-\tfrac{1}{2}N_0(\sqrt{\lambda}\,|x-y|)$$

and

$$[R^0_{\lambda+i0}-R^0_{\lambda-i0}](x,y)=\frac{i}{2}J_0(\sqrt{\lambda}\,|x-y|),$$

where J_0 and N_0 are the Bessel and Neumann functions of order 0. Thus

$$2i\,I(b,0^+)=-\frac{i}{4}\int_0^b d\lambda\int dx\,\chi_\Sigma(x)\int dy\,J_0\,(\sqrt{\lambda}\,|x-y|)$$
$$\times N_0\,(\sqrt{\lambda}\,|x-y|)v(y).$$

Denoting $s_\lambda(x)\equiv s_\lambda(|x|)=-(i/4)\,J_0(\sqrt{\lambda}\,|x|)\,N_0(\sqrt{\lambda}\,|x|)$ for $\lambda>0$, we can rewrite this as

$$\begin{aligned}2i\,I(b,0^+)&=\int_0^b d\lambda\int dx\,\chi_\Sigma(x)\int dy\,s_\lambda(x-y)v(y)\\&=\int_0^b d\lambda\int dx\,s_\lambda(x)(\chi_\Sigma*v)(x),\end{aligned}\tag{17}$$

where we have written $(\chi_\Sigma*v)(x)=\int\chi_\Sigma(x+y)v(y)\,dy$ and also noted that the above integral converges absolutely by the estimate $(|s_\lambda(x)|<c|x|^{-1})$, and by an application of the Sobolev inequality so that Fubini's theorem can be used.

From Ref. 3, p. 673, formula (6), we note that the improper Riemann Fourier transform of s_λ exists, i.e.,

$$\frac{1}{2\pi}\int_{|x|<n}e^{-ik\cdot x}s_\lambda(x)\,dx$$
$$=-\frac{i}{4}\int_0^n J_0(|k|r)\,J_0(\sqrt{\lambda}\,r)\,N_0(\sqrt{\lambda}\,r)r\,dr$$

converges pointwise for $0<|k|\neq2\sqrt{\lambda}$ as $n\to\infty$ to a function $\tilde{s}_\lambda$ with

$$\tilde{s}_\lambda(k)=\frac{i}{4}\begin{cases}0,&\text{if }0<|\mathrm{k}|<2\sqrt{\lambda},\\(2/\pi)|k|^{-1}(k^2-4\lambda)^{-1/2},&\text{if }2\sqrt{\lambda}<|k|<\infty.\end{cases}\tag{18}$$

It is clear that $\tilde{s}_\lambda\in L^p(\mathbb{R}^2)$ with $1<p<2$ and thus by the Hausdorff–Young theorem (Ref. 6, p. 11) its inverse Fourier transform $\mathscr{F}^{-1}\,\tilde{s}_\lambda\in L^q(\mathbb{R}^2)$ with $p^{-1}+q^{-1}=1$ and furthermore $\mathscr{F}^{-1}$ is a continuous linear map from $L^p(\mathbb{R}^2)$ into $L^q(\mathbb{R}^2)$. Also since convergence in the L^q norm implies convergence pointwise almost everywhere for a subsequence (Ref. 4, p. 18) we conclude that the improper Riemann inverse Fourier transform of $\tilde{s}_\lambda$,

$$\lim_{N\to\infty}\frac{1}{2\pi}\int_{|k|<N}e^{ik\cdot x}\,\tilde{s}_\lambda(k)dk,$$

if it exists, equals $(\mathscr{F}^{-1}\,\tilde{s}_\lambda)(x)$ a.e. That this improper Riemann integral exists and is equal to $s_\lambda(x)$ is the formula (6) of Ref. 3, p. 682. Therefore, by Lemma 8, Eq. (17) reduces to

$$\begin{aligned}2i\,I(b,0^+)&=\int_0^b d\lambda\int\tilde{s}_\lambda(k)\widetilde{\chi_\Sigma*v}(k)dk\\&=2\pi\int_0^b d\lambda\int\tilde{s}_\lambda(k)\,\tilde{\chi}_\Sigma(k)\tilde{v}(-k)dk,\end{aligned}\tag{19}$$

where we have observed that since $\chi_\Sigma\in L^1(\mathbb{R}^2)$, $(\widetilde{\chi_\Sigma*v})(k)=2\pi\,\tilde{\chi}_\Sigma(k)\tilde{v}(-k)$.

An elementary integration shows that

$$\tilde{S}_b(k)\equiv2\pi\int_0^b\tilde{s}_\lambda(k)d\lambda$$
$$=\frac{i}{2}\begin{cases}1,&\text{if }0<|k|\leqslant2\sqrt{b},\\1-(1-4b/k^2)^{1/2},&\text{if }2\sqrt{b}<|k|.\end{cases}$$

Since $|\tilde{s}_\lambda(k)|=-i\,\tilde{s}_\lambda(k)$, it follows that

$$\int_0^b|\tilde{s}_\lambda(k)|d\lambda=-i\int_0^b\tilde{s}_\lambda(k)dk=-\frac{i}{2\pi}\tilde{S}_b(k)\leqslant(4\pi)^{-1}$$

for all $|k|>0$, and recalling the hypothesis $\tilde{\chi}_\Sigma\tilde{v}\in L^1$, we can apply Fubini's theorem to (19) and obtain

$$2i\,I(b,0^+)=\int\tilde{S}_b(k)\tilde{\chi}_\Sigma(k)\tilde{v}(-k)dk.\tag{20}$$

Note that $\tilde{S}_b$ converges to $i/2$ pointwise for all $|k|>0$ as $b\to\infty$ and that $|\tilde{S}_b(k)|\leqslant\frac{1}{2}$. Therefore, we apply dominated convergence to (20) to arrive at

$$\lim_{b\to\infty}I(b,0^+)=\frac{1}{4}\int\tilde{\chi}_\Sigma(k)\tilde{v}(-k)\,dk.\tag{21}$$

Finally an application of Lemma 7 to (21) gives the required result. □

Lemma 7: Let $\psi\in L^r(\mathbb{R}^n)$ for some $r\in[1,2]$ and $f\in L^2\cap L^t(\mathbb{R}^n)$, where $r^{-1}+t^{-1}=1$. Assume furthermore that $\tilde{\psi}\tilde{f}\in L^1(\mathbb{R}^n)$. Then

$$\int\overline{\psi(x)}\,f(x)dx=\int\overline{\tilde{\psi}(k)}\tilde{f}(k)dk.\tag{22}$$

Proof: See the Appendix. □

Lemma 8: Assume $v \in L^{4/3}(\mathbb{R}^2)$ and $|\Sigma| < \infty$. Let s_λ and $\tilde{s}_\lambda$ be as defined in Proposition 6. Then

$$\int s_\lambda(x)\chi_\Sigma * v(x)dx = \int \tilde{s}_\lambda(k)\widetilde{\chi_\Sigma * v}(k)dk. \tag{23}$$

Proof: Set $\psi \equiv \tilde{s}_\lambda$ and $f \equiv \widetilde{\chi_\Sigma * v}$ and utilize Lemma 7 with $\mathbb{R}^n = \mathbb{R}^2$ and $r = \frac{4}{3}$. As noted in Proposition 6, $\tilde{s}_\lambda \in L^{4/3}$. The function f is the Fourier transform of a convolution and is proportional to the product $\tilde{\chi}_\Sigma(k)\tilde{v}(-k)$. Because $\tilde{\chi}_\Sigma \in L^2 \cap L^\infty$ and $\tilde{v} \in L^4$ we have from Hölder's inequality that $f \in L^2 \cap L^4$. It remains only to verify that the requirement $\tilde{\psi}\tilde{f} \in L^1$ is met. Both $\widetilde{\chi_\Sigma * v}$ and $\chi_\Sigma * v$ are in L^2 and thus a.e. $\tilde{f}(x) = \chi_\Sigma * v(-x)$. Since $\chi_\Sigma \in L^1$ and $v \in L^{4/3}$ it follows that $\tilde{f} \in L^{4/3}$. Finally, $s_\lambda \in L^4$, so Hölder's inequality implies $\tilde{\psi}\tilde{f} \in L^1$.

Observe that $\tilde{\psi}(x) = s_\lambda(-x)$ and that both $s_\lambda(x)$ and $\tilde{s}_\lambda(k)$ have purely imaginary values. Thereby, it is seen that (23), with $\psi = \tilde{s}_\lambda$ and $f = \widetilde{\chi_\Sigma * v}$, is equivalent to the identity (22). □

For $b > \Lambda_\theta$, define a large radius integration contour in Π by $C_\delta(b) = \{z \in \Pi \mid |z| = \sqrt{b^2 + \delta^2}$ and $|\mathrm{Im}\, z| \geqslant \delta$ if $\mathrm{Re}\, z > 0\}$. The contour integral over $C_\delta(b)$ will be taken in the conventional right-hand sense.

Proposition 9: Suppose $v \in L^{4/3}(\mathbb{R}^2)$ and $|\Sigma| < \infty$. Then

$$\lim_{b \to \infty} \lim_{\delta \to 0^+} \int_{C_\delta(b)} \mathrm{tr}\, P_\Sigma [R_z - R_z^0 + R_z^0 v R_z^0] P_\Sigma \, dz = 0. \tag{24}$$

Proof: The identities (valid for $z \in \Pi_c$, $|z| > \Lambda_\theta$, $1 > \theta > 0$) $[1 + A(z)]^{-1} = 1 - A(z) + A(z)^2 - A(z)^3[1 + A(z)]^{-1}$ and $R_z^0 v R_z^0 = (R_z^0 w)(u R_z^0)$, when combined with (1), give

$$P_\Sigma [R_z - R_z^0 + R_z^0 v R_z^0] P_\Sigma = \sum_{i=1}^{3} K_i(z),$$

where

$$K_3(z) = P_\Sigma R_z^0 w A(z)^3 [1 + A(z)]^{-1} u R_z^0 P_\Sigma,$$

$$K_i(z) = (-1)^{i+1} P_\Sigma R_z^0 w [A(z)]^i u R_z^0 P_\Sigma, \quad i = 1,2.$$

Consider the K_3 contribution first. If we take the polar representation of $z = |z| \exp(i\psi)$, $\psi \in [0, 2\pi]$, then $z \in C_\delta(b)$ requires $\phi \leqslant \psi \leqslant 2\pi - \phi$, where $\tan \phi = \delta / b$. Bound estimate (3) is of the form $\|A(z)\|_2 = O(|z|^{-1/4})$. Since $\chi_\Sigma \in L^{4/3}(\mathbb{R}^2)$, a similar Sobolev estimate shows that both $\|P_\Sigma R_z^0 w\|_2$ and $\|u R_z^0 P_\Sigma\|_2$ decay like $O(|z|^{-1/4})$ for large $|z|$. After noting that $\|[1 + A(z)]^{-1}\| \leqslant (1 - \theta)^{-1}$, one finds

$$\left| \int_{C_b(\delta)} \mathrm{tr}\, K_3(z) dz \right| \leqslant \frac{c^5}{(b^2 + \delta^2)^{1/8}} \left\{ \frac{2(\pi - \phi)}{1 - \theta} \|\chi_\Sigma\|_{4/3} \|v\|_{4/3}^4 \right\}, \tag{25}$$

where c is the constant arising in the Sobolev estimate. The right side of (25) vanishes in the double limit $\delta \to 0^+$, $b \to \infty$.

The analysis of the contribution of both K_1 and K_2 to the limit in (24) is similar, so we shall restrict the discussion to the K_1 term. The operator K_1 is the product of three $\mathscr{B}$ operators, so $\mathrm{tr}\, K_1$ may be calculated as the triple iterated integral of the kernels associated with these Hilbert–Schmidt operators (Ref. 10, p. 524). Upon using estimate (2) for $R_z^0(x,y)$ and setting $\Gamma = (b^2 + \delta^2)^{1/2}$ we have, for $z \in C_\delta(b)$,

$$|\mathrm{tr}\, K_1(z)| < \frac{c}{\Gamma^{3/4}} \int dx\, \chi_\Sigma(x) \int \int dy_1\, dy_2 \times \frac{|v(y_1) v(y_2)| e^{-r\,\mathrm{Im}\sqrt{z}}}{|x - y_1|^{1/2} |y_1 - y_2|^{1/2} |y_2 - x|^{1/2}},$$

where $r = |x - y_1| + |y_1 - y_2| + |y_2 - x|$. Doing the $|dz|$ integral along contour $C_\delta(b)$ gives the bound

$$\int_{C_b(\delta)} |\mathrm{tr}\, K_1(z)|\, |dz| < \frac{2\pi c}{\Gamma^{1/4}} \int \int \int dx\, dy_1\, dy_2 \times \frac{\chi_\Sigma(x) |v(y_1)|\, |v(y_2)|}{r |x - y_1|^{1/2} |y_1 - y_2|^{1/2} |y_2 - x|^{1/2}}, \tag{26}$$

where the fact that the integrand is non-negative has been used to justify changing the order of integration. Clearly if the triple integral in (26) is finite, then the $b \to \infty$, $\delta \to 0^+$ limit of the K_1 term in (24) vanishes. The finiteness of this triple integral follows by the inequality $r \geqslant |x - y_1|^{1/4} |y_1 - y_2|^{1/2} |y_2 - x|^{1/4}$ together with Schwartz inequality to bound the dx integral and the Sobolev inequality to estimate the $dy_1\, dy_2$ integral. □

Proof of Theorem 4: For $z \in \rho(H) \cap \rho(K_0)$ define $\Phi(z) \in \mathscr{B}$ by

$$\Phi(z) \equiv P_\Sigma [R_z - R_z^0 + R_z^0 v R_z^0] P_\Sigma.$$

From Theorem 1, Eq. (1) it is seen that Φ may also be represented as

$$\Phi(z) = (P_\Sigma R_z^0 w) A(z) [1 + A(z)]^{-1} (u R_z^0 P_\Sigma). \tag{27}$$

The outer two factors on the right side of (27) are $\mathscr{B}_2$ holomorphic in $\rho(K_0)$ while the inner factors are norm holomorphic in $\rho(H)$. It follows that $z \to \Phi(z)$ is trace-norm holomorphic on the domain $\rho(H) \cap \rho(K_0)$.

Select a and b such that $(a,b) \supset [-\Lambda_\theta, \Lambda_\theta]$ for some $0 < \theta < 1$. For fixed a, b, and $\delta > 0$ choose a closed contour in the canonical cut plane Π to be $C_T \equiv C_\delta(a) + C_\pm(b,\delta) + C_\delta(b)$, where $C_\delta(b)$ has been given above and

$$C_\pm(b,\delta) = \{z \in \Pi \mid z = \lambda \pm i\delta, \quad \lambda \in [a,b]\},$$

$$C_\delta(a) = \{z \in \Pi \mid z = a + i\eta, \quad \eta \in [-\delta, \delta]\}.$$

Define a holomorphic function on $\rho(H) \cap \rho(K_0) \subset \mathbb{C}$ by setting $h(z) = \mathrm{tr}\, \Phi(z)$. Cauchy's integral theorem asserts that the C_T contour integral of $h(z)$ vanishes. Specifically, for each $\delta > 0$,

$$i \int_{-\delta}^{\delta} h(a + i\eta) d\eta + \int_{C_\delta(b)} h(z) dz + 2i \int_a^b \mathrm{Im}\, h(\lambda + i\delta) d\lambda = 0. \tag{28}$$

Consider that $\delta \to 0^+$ limit of the first integral in (28). Use (27) to rewrite the argument of $\mathrm{tr}\, \Phi$. After applying the Sobolev inequality to estimate the $\|\cdot\|_2$ norm of $P_\Sigma R_z^0 w$, $A(z)$, and $u R_z^0 P_\Sigma$, and using $\|[1 + A(z)]^{-1}\| \leqslant (1 - \theta)^{-1}$, it follows (if $a < -\Lambda_\theta$) that $|h(a+i\eta)|$ is uniformly bounded in η. Thus the integral $\int_{-\delta}^{\delta} h(a+i\eta)d\eta$ vanishes as $\delta \to 0^+$.

Now take the $\delta \to 0^+$, $b \to \infty$ limit of identity (28). The limiting value of the middle term is determined by Proposition 9 to be zero, leaving us with

$$\lim_{b\to\infty} \lim_{\delta\to 0^+} \int_a^b \operatorname{Im} h(\lambda + i\delta)d\lambda = 0. \tag{29}$$

Inserting the results of Propositions 5 and 6 into (29) yields

$$\lim_{b\to\infty} \frac{1}{2}\int_0^b q(\lambda,\Sigma)d\lambda + \pi \operatorname{tr} P_\Sigma\, E_s\, P_\Sigma + \frac{1}{4}\int_\Sigma v(x)\,dx = 0.$$

Here, the second and third factors are both finite. This requires that

$$\lim_{b\to\infty} \int_0^b q(\lambda,\Sigma)d\lambda$$

be finite; i.e., the improper integral of $\lambda \to q(\lambda,\Sigma)$ satisfies (7). □

IV. DISCUSSION

We conclude by making a number of remarks concerning the spectral sum rule.

(1) Consider the behavior of hypothesis (ii) in Theorem 4. The condition (ii) acts as a joint constraint on v and Σ. Given a fixed set Σ, (ii) restricts the choice of v; or given a fixed $v \in L^{4/3}$, (ii) defines an admissible class of sets $\Sigma \subset \mathbb{R}^2$. Two examples illustrate how (ii) works. For every $v \in L^{4/3}$, one can find a Σ such that (ii) is valid. Suppose Σ is a rectangle. Then $\tilde{\chi}_{\rm rect} \in L^{4/3}$, and furtheremore, since $\tilde{v} \in L^4$, Hölder's inequality implies $\tilde{v}\tilde{\chi}_{\rm rect} \in L^1$. On the other hand, hypothesis (ii) need not be fulfilled by all pairs (v,Σ) allowed by (i). Let Σ be a disk. Then $\tilde{\chi}_{\rm disk} \in L^{4/3+} \cap L^\infty$. In this case, if the potential class is restricted to $v \in L^{4/3} \cap L^{4/3+}$, then (ii) will be satisfied for the disk. Finally, we observe that if the potential class is further narrowed to $v \in L^{4/3} \cap L^2$, then (ii) is obeyed for all Σ with $|\Sigma| < \infty$.

(2) It is often desirable to separate the contributions of the point spectrum and the singularly continuous spectrum. Suppose $\{\psi_i\}$ is the family of independent $L^2(\mathbb{R}^2)$ eigenfunctions of H having eigenvalues λ_i and normalization $\|\psi_i\| = 1$. These eigenvalues always lie within the interval $[-\Lambda_\theta, \Lambda_\theta]$ and may assume negative, zero, or positive values. The family $\{\psi_i\}$ may be empty, finite, or infinite. (In particular, the assumption $v \in L^{4/3}$ is not known to rule out an infinite number of positive eigenvalues.) The spectral subspace decomposition $E_s = E_{\rm pp} + E_{\rm sc}$ implies

$$\operatorname{tr} P_\Sigma\, E_s\, P_\Sigma = \sum_i \int_\Sigma |\psi_i(x)|^2 dx + \operatorname{tr} P_\Sigma\, E_{\rm sc}\, P_\Sigma.$$

(3) Various sufficient conditions on v are known to ensure the absence of the singular continuous spectrum and of the positive point spectrum of H. We quote only two representative results.

Theorem: Let $(1+|x|)^\nu v(x) \in L^{4/3}(\mathbb{R}^2) + L^\infty(\mathbb{R}^2)$, $\nu > 1$. Then $\mathscr{H}_{\rm sc}(H) = \{0\}$. Furthermore, there are a finite number of positive eigenvalues of H with finite multiplicity in every compact subset of $(0,\infty)$.

This result follows from both time-dependent Enss–Mourre theory[11] as well as from time-independent theory.[12]

Theorem: Let $(1+|x|)v(x) \in L^2(\mathbb{R}^2)$. Then H has no positive eigenvalues.

This is a specialized version of the more general results obtained by Froese *et al.*[13]

(4) For any $|\Sigma| < \infty$, Remark (1) shows the spectral sum rule identity (7) is valid for all $v \in L^{4/3} \cap L^2$. The potential class $L^{4/3} \cap L^2$ does not prohibit the appearance of zero energy resonances (see Refs. 14 and 15). For example, if one varies v in $L^{4/3} \cap L^2$ by changing the coupling constant it is possible to introduce zero energy resonances in the scattering system. However, the local spectral sum rule (7) [takes the same form (7) for all $v \in L^{4/3} \cap L^2$, and so] is structurally insensitive to the presence or absence of a zero energy resonance.

(5) Global sum rules (Levinson's theorem) obtain if $\Sigma = \mathbb{R}^2$. A result of the literature that is closely related to the spectral sum rule in Theorem 4 is the $\mathbb{R}^2$–Levinson theorem derived by Cheney.[16] Let $S(k)$: $L^2(T) \mapsto L^2(T)$ denote the energy-shell S-matrix operator, where $|k| = \sqrt{\lambda} \geqslant 0$. Then for a potential class that prohibits (1) the singularly continuous spectrum, (2) non-negative eigenvalues, and (3) zero-energy resonances, it is found that[16]

$$i[\log\det S(0) - \log\det S(\infty)] = -2\pi N - \frac{1}{2}\int_{\mathbb{R}^2} v(x)dx,$$

where N is the number of negative energy bound states.

For scattering in $\mathbb{R}^3$ the effect of zero-energy resonances on the form of Levinson's theorem has been discussed several times.[17,18] In a notation analogous to the above, Newton[17] finds

$$\delta(0) - \lim_{k\to\infty}\left[\delta(k) + \frac{k}{4\pi}\int_{\mathbb{R}^3} v(x)dx\right] = \pi\left(N + \frac{1}{2}q\right),$$

where $\delta(k)$ is an appropriately chosen phase parametrization for the S matrix, $\ln(\det S(k)) = 2i\,\delta(k)$. The factor $q = 0$, if there are no zero-energy resonances, and $q = 1$, otherwise.

Here N is the number of zero-energy and negative-energy eigenfunctions. It is this type of zero-energy resonance modification of Levinson's global sum rule that does not occur in the local sum rule of Theorem 4.

ACKNOWLEDGMENTS

Work on this project has been carried out at the Universities of Bielefeld, Leuven, Maryland, and Manitoba. The authors would like to thank all these institutions for their support and hospitality, which has made this collaboration possible. This work has been supported by grants from the Department of Energy (U. S.) and the Natural Sciences and Engineering Research Council (Canada).

APPENDIX: PROOF OF LEMMA 7

Set $\phi(x) = (2\pi)^{-n/2}\exp(-x^2/2)$ and for every $\epsilon > 0$, $\phi_\epsilon(x) = \epsilon^{-n}\phi(x/\epsilon)$ so that $\int \phi_\epsilon(x)\,dx = 1$ and $\tilde{\phi}_\epsilon(k) = (2\pi)^{-n/2}\exp(-k^2\epsilon^2/2)$. Define

$$\psi_\epsilon(x) = \int \psi(x+y)\,\phi_\epsilon(y)\,dy = \int \psi(x+\epsilon y)\,\phi(y)\,dy. \tag{30}$$

Note that $\psi_\epsilon \in L^r(\mathbb{R}^n)$ and $\|\psi_\epsilon\|_r \leq \|\psi\|_r$. Since the map $\lambda \to |\lambda|^r$ is convex on R^+ for $r \geq 1$ and since $\int \phi(y)dy = 1$, we have, by Jensen's inequality[19] and (30),

$$\|\psi - \psi_\epsilon\|_r^r = \int |\psi(x) - \psi_\epsilon(x)|^r \, dx$$

$$\leq \int dx \int |\psi(x + \epsilon y) - \psi(x)|^r \phi(y) \, dy$$

$$= \int \|(T_{\epsilon y} - I)\psi\|_r^r \, \phi(y) dy,$$

where $(T_y \psi)(x) = \psi(x + y)$.

Now $T_{\epsilon y} \psi \to \psi$ in L^r norm as $\epsilon \to 0^+$ for every y fixed. Furthermore T_y is an isometry. Therefore, by dominated convergence one has that $\|\psi - \psi_\epsilon\|_r \to 0$ as $\epsilon \to 0^+$. Since $\psi_\epsilon \in L^2(\mathbb{R}^n)$ by Young's theorem (Ref. 6, p. 28), we have by Plancherel's theorem that

$$\int \overline{\psi_\epsilon(x)} f(x) dx = \int \overline{\tilde{\psi}_\epsilon(k)} \tilde{f}(k) dk. \tag{31}$$

The left-hand side of (31) converges to $\int \overline{\psi(x)} f(x) dx$ since $\|\psi_\epsilon - \psi\|_r \to 0$ as $\epsilon \to 0^+$ and since $f \in L^2(\mathbb{R}^n)$. On the other hand, $\tilde{\psi}_\epsilon(k) = (2\pi)^{n/2} \tilde{\psi}(k) \tilde{\phi}_\epsilon(-k) = \tilde{\psi}(k) e^{-k^2\epsilon^2/2} \to \tilde{\psi}(k)$ pointwise and $|\tilde{\psi}_\epsilon(k)| \leq |\tilde{\psi}(k)|$ so that an application of the dominated convergence theorem to the right side of (31) along with the hypothesis $\tilde{\psi} \tilde{f} \in L^1(\mathbb{R}^n)$ leads to the original result. □

[1]D. Bollé and T. A. Osborn, J. Math. Phys. **22**, 883 (1981).
[2]B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms* (Princeton U. P., Princeton, NJ, 1971).
[3]I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integral Series and Products* (Academic, New York, 1965).
[4]M. Reed and B. Simon, *Methods of Modern Mathematical Physics I: Functional Analysis* (Academic, New York, 1972).
[5]W. O. Amrein, J. M. Jauch, and K. B. Sinha, *Scattering Theory in Quantum Mechanics* (Benjamin, Reading, MA, 1977).
[6]M. Reed and B. Simon, *Methods of Modern Mathematical Physics II: Fourier Analysis, Self-Adjointness* (Academic, New York, 1975).
[7]T. Kato, Math. Ann. **162**, 258 (1966).
[8]M. Reed and B. Simon, *Methods of Modern Mathematical Physics IV: Analysis of Operators* (Academic, New York, 1978).
[9]J. M. Jauch, K. B. Sinha, and B. Misra, Helv. Phys. Acta **45**, 398 (1972).
[10]T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1980), 2nd ed.
[11]P. Perry, *Scattering Theory by the Enss Method, Mathematical Reports* (Harwood Academic, London, 1983), Vol. 1, Part 1.
[12]S. Agmon, Ann. Scoula Normale Sup. Pisa, Ser. IV **2**, 151 (1975).
[13]R. Froese, I. Herbst, M. Hoffmann-Ostenhof, and T. Hoffmann-Ostenhof, J. Anal. Math. **41**, 272 (1982).
[14]A. Jensen and T. Kato, Duke Math. J. **46**, 583 (1979).
[15]M. Murata, J. Funct. Anal. **49**, 10 (1982).
[16]M. Cheney, J. Math. Phys. **25**, 1449 (1984).
[17]R. Newton, J. Math. Phys. **18**, 1348 (1977).
[18]D. Bollé and S. F. J. Wilk, J. Math. Phys. **24**, 1555 (1983).
[19]H. L. Royden, *Real Analysis* (MacMillan, London, 1968), 2nd ed.

Three-dimensional inverse scattering: Plasma and variable velocity wave equations

James H. Rose
Ames Laboratory—United States Department of Energy,[a] *Iowa State University, Ames, Iowa 50011*

Margaret Cheney
Department of Mathematics, Duke University, Durham, North Carolina 27706

Brian DeFacio[b]
Department of Physical Mathematics, University of Science and Technology at Languedoc, Place Eugene Batillion, Montpelier Cedex 34060, France

(Received 14 March 1985; accepted for publication 12 June 1985)

Exact equations governing three-dimensional time-domain inverse scattering are derived for the plasma wave equation and the variable velocity classical wave equation. This derivation was announced for the plasma wave equation in a short note by the authors. That work was motivated by Newton's three-dimensional generalization of Marchenko's equation. This paper gives the details of the new derivation and extends it to the classical wave equation. For the time domain derivation in this paper, the scattering region is assumed to have compact support and smoothly joins the surrounding three-dimensional infinite medium. The derivation contains the following ingredients: (1) a representation of the solution at a point in terms of its values on a large sphere, (2) the far-field form of the Green's function, (3) causality, and (4) information carried in the wave front of the solution. The derivation of the classical wave inverse scattering equation requires that the velocity in the scattering region be less than that of the surrounding medium. This condition is natural, for example, in the scalar wave model of electromagnetic scattering from dielectric nonconducting bodies in free space. Finally, an experiment to verify the inverse scattering equations is proposed.

I. INTRODUCTION

In a recent note[1] the authors proposed a simple physically motivated derivation of an exact Marchenko-like inverse scattering equation for the three-dimensional linear plasma wave equation[2–6]

$$(\Delta - \partial_t^2 - V(\mathbf{x}))U(t,\hat{e},\mathbf{x}) = 0. \tag{1}$$

Here Δ is the Laplacian and ∂_t^2 is the second derivative with respect to time. The potential $V(\mathbf{x})$ is a scalar function. For simplicity we make the following assumptions about V: (1) $-\Delta + V$ supports no bound states, and (2) V is infinitely differentiable and has compact support (i.e., $V \in C_0^\infty$). Assumption (2) can certainly be relaxed. Also, $U(t,\hat{e},\mathbf{x})$ is the wave field, $\hat{e}$ denotes the direction of incidence, t the time, and $\mathbf{x}$ a point in R^3. Note that potentials and field quantities are denoted by captial letters when discussing Eq. (1).

The derivation just noted will be generalized to provide a set of exact inverse scattering equations for the variable velocity scalar wave equation[7–9] (from now on called the classical wave equation):

$$(\Delta - [1/c^2(\mathbf{x})]\partial_t^2)\ u(t,\hat{e},\mathbf{x}) = 0. \tag{2}$$

Here $c(\mathbf{x})$ is a positive real valued function which denotes the velocity at $\mathbf{x}$ and $u(t,\hat{e},\mathbf{x})$ is the corresponding wave field. We assume that $c(\mathbf{x})$ is asymptotically constant; in this sense the scatterer is situated in an otherwise isotropic and homogeneous host medium. We assume, moreover, that the velocity $c(\mathbf{x})$ is everywhere less than or equal to that of the host medium. [This assumption is physically natural if Eq. (2) is used, for example, to model electromagnetic scattering from a dielectric body in free space.] In addition, we assume that $c(\mathbf{x})$ is in C_0^∞ and differs from the host velocity only in a region of compact support. Lowercase letters will be used to denote the velocity and field quantities for the classical wave equation. The wave equations used imply that for the underlying physical problems (1) attenuation is negligible and (2) the system is passive; i.e., there is no spontaneous introduction of energy into the system.

The historical development of our method proceeds as follows. In 1950, Marchenko[10,11] introduced a method for solving the inverse scattering problem for the time-independent Schrödinger equation (single particle scattering from a scalar potential) assuming a spherically symmetric potential. Recently, Newton,[12–17] in an important development, generalized Marchenko's approach to the three-dimensional Schrödinger equation. The authors then showed that, for positive $V(\mathbf{x})$, the same approach can be used to solve the three-dimensional inverse problem for the plasma wave equation. The exploration of this result led to (1) the physically motivated derivation to be discussed in this paper and (2) its generalization to the classical wave equation.

The derivations to be carried out in this paper have the following structure. One starts with the scattered wave field measured on a large sphere S at infinity. If the exact far-field Green's function, including the effect of the potential, is known, a representation theorem can be used to determine the wave field at all points interior to the sphere S. That is,

[a] Ames Laboratory is operated for the U.S. Department of Energy by Iowa State University under Contract No. W-7405-ENG-82.

[b] Permanent address: Department of Physics and Astronomy, University of Missouri, Columbia, Missouri 65211.

the wave field is determined by the scattering data and the far-field Green's function. Since, as we discuss, the far-field Green's function is proportional to the wave field itself, we obtain an integral equation for the wave field in terms of the scattering data. Under the conditions noted on $V(\mathbf{x})$ and $c(\mathbf{x})$ this integral equation reduces to the three-dimensional analog of Marchenko's approach.

Causality is an essential element of exact inverse scattering methods. In frequency-domain derivations, causality shows up via repeated use of the analytic features of the wave functions. By contrast, our time-domain approach includes causality in a transparent way in terms of the wave fronts (characteristic surfaces). We feel this adds considerable intuitive clarity to our approach.

The structure of this paper is as follows. Section II is devoted to notation and preliminaries. Section III is used to review and discuss the time-domain representation theorem. The far-field form of Green's function and its physical interpretation is given in the fourth section. Section V contains the derivation of the integral equation for the wave field mentioned above. In Section VI the extraction of the potential from the wavefield is described. At this step, the treatment of the two wave equations differs. In particular, the role of curved wave fronts and caustics for the classical wave equation is discussed. The last section notes that the two wave equations can be physically realized. In particular, the plasma wave equation governs acoustic scattering if the velocity of the fluid is the same at all points in space (the density and compressibility may, however, vary). The paper is concluded with two appendices. These appendices give a careful frequency-domain treatment of certain aspects of the derivation. These results are used to support the more heuristic time-domain discussion given in the main text.

II. PRELIMINARIES AND NOTATION

The scattering geometry is shown in Fig. 1 for both wave equations. To make our two equations appear as similar as possible, we assume that $c(\mathbf{x})$ in Eq. (2) is asymptotically 1. The scattering region is defined to be that portion Ω of R^3 in which $V(\mathbf{x})$ differs from zero and $c(\mathbf{x})$ from 1. We chose the origin of coordinates to be in Ω. A scattering experiment may be qualitatively described as follows. At very early time the wave field is described by an incident plane wave delta

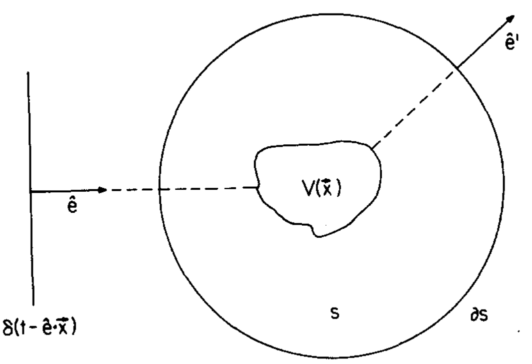


FIG. 1. The scattering geometry is shown. Here $\hat{e}$ and $\hat{e}'$ are the directions of incidence and scattering, respectively. The data is measured on the large spherical surface labeled ∂S.

function $\delta(t-\hat{e}\cdot\mathbf{x})$. This stipulation takes the place of both initial conditions so that for large negative times the time derivative of the wave field is specified by $\partial_t\delta(t-\hat{e}\cdot\mathbf{x})$. This field propagates freely until it intersects the scattering region Ω. The wave field then interacts with the potential in a complicated way. The wave field is finally measured on a large spherical surface ∂S, centered about the origin. Generally the scattering data will be described in the far-field limit and the radius of S will be taken as arbitrarily large. For very early times the wave field is zero in the scattering region since the incident wave has not yet arrived. It will be assumed that the wave field is zero in the scattering region Ω for large positive time. That is, for times sufficiently far in the future all of the energy has propagated out of Ω. (See Appendix A.)

A. Plasma wave scattering formalism

Scattering theory can be written using integral wave equations defined in terms of the Green's function. In particular, the plasma wave equation may be written in two alternative forms

$$U^{\pm}(t,\hat{e},\mathbf{x}) = U_0^{\pm}(t,\hat{e},\mathbf{x}) + \int d^3\mathbf{x}'\,dt' \times G_0^{\pm}(t-t',\mathbf{x},\mathbf{x}')V(\mathbf{x}')U^{\pm}(t',\hat{e},\mathbf{x}') \tag{3a}$$

or

$$U^{\pm}(t,\hat{e},\mathbf{x}) = U_0^{\pm}(t,\hat{e},\mathbf{x}) + \int d^3\mathbf{x}'\,dt' \times G^{\pm}(t-t',\mathbf{x},\mathbf{x}')V(\mathbf{x}')U_0^{+}(t',\hat{e},\mathbf{x}'). \tag{3b}$$

Here $U_0^{\pm}$ denotes the incident field and for this paper we uniformly require that $U_0^{\pm}(t,\hat{e},\mathbf{x}) = \delta(t-\hat{e}\cdot\mathbf{x})$. The Green's functions appearing in Eq. (3) satisfy

$$(\Delta - \partial_t^2)G_0^{\pm}(t-t',\mathbf{x},\mathbf{x}') = \delta^3(\mathbf{x}-\mathbf{x}')\delta(t-t'), \tag{4a}$$

$$(\Delta - \partial_t^2 - V(\mathbf{x}))G^{\pm}(t-t',\mathbf{x},\mathbf{x}') = \delta^3(\mathbf{x}-\mathbf{x}')\delta(t-t'). \tag{4b}$$

Explicitly

$$G_0^{\pm}(t,\mathbf{x},\mathbf{x}') = -(4\pi|\mathbf{x}-\mathbf{x}'|)^{-1}\delta(|\mathbf{x}-\mathbf{x}'| \mp t), \tag{5}$$

where the $+$ and $-$ signs correspond to radiation and incoming boundary conditions, respectively. We note that the free-space Green's functions $G_0^{\pm}$ depend only on $|\mathbf{x}-\mathbf{x}'|$ rather than on $\mathbf{x}$ and $\mathbf{x}'$ independently. The functions $G^{\pm}$ and $G_0^{\pm}$ are related via the equation

$$G^{\pm}(t-t',\mathbf{x},\mathbf{x}') = G_0^{\pm}(t-t',\mathbf{x},\mathbf{x}') + \int d^3\mathbf{x}'' \int dt'' \times G^{\pm}(t-t'',\mathbf{x},\mathbf{x}'')V(\mathbf{x}'')G_0^{\pm}(t''-t',\mathbf{x}'',\mathbf{x}'). \tag{6}$$

Finally the solutions U^+ and U^- are related via

$$U^-(t,\hat{e},\mathbf{x}) = U^+(-t,-\hat{e},\mathbf{x}), \tag{7}$$

as follows from Eqs. (3a) and (5).

Causality plays a crucial role in inverse scattering theory. The wave front (a characteristic surface) occurs at $t=\hat{e}\cdot\mathbf{x}$. For $t<\hat{e}\cdot\mathbf{x}$ the wave field $U^+(t,\hat{e},\mathbf{x})$ is identically zero. Similarly, from Eq. (7), $U^-(t,\hat{e},\mathbf{x})$ is identically zero for

$t > \hat{e}\cdot\mathbf{x}$. The Green's functions also satisfy causality and $G_0^+(t-t',\mathbf{x},\mathbf{x}')$ and $G^+(t-t',\mathbf{x},\mathbf{x}')$ are zero if $t-t' < |\mathbf{x}-\mathbf{x}'|$. Since $|\mathbf{x}-\mathbf{x}'|$ is always positive it follows that G_0^+ and G^+ are zero if $t-t'<0$. This ensures that the response at $(\mathbf{x},t)$ cannot occur before the cause at $(\mathbf{x}',t')$. Thus G_0^+ and G^+ propagate events from the past to the future. On the other hand, G_0^- and G^- are zero if $t-t' > |\mathbf{x}-\mathbf{x}'|$ or $t-t'>0$ and propagate events from the future to the past.

The scattering data is derived from the far-field asymptotic value of the wave field and is conveniently represented by the impulse response function

$$R(\hat{e},\hat{e}',\tau) = \lim_{\substack{t,|\mathbf{x}|\to\infty \\ \tau=t-|\mathbf{x}|}} |\mathbf{x}|\{U(t,\hat{e},\mathbf{x}) - \delta(t-\hat{e}\cdot\mathbf{x})\}. \tag{8a}$$

In other words, for large $|\mathbf{x}|$,

$$U^{\pm}(t,\hat{e},\mathbf{x}) = \delta(t-\hat{e}\cdot\mathbf{x}) + \frac{R(\pm\hat{e},\hat{e}',\pm t - |\mathbf{x}|)}{|\mathbf{x}|} + O\left(\frac{1}{|\mathbf{x}|^2}\right). \tag{8b}$$

Here, $\hat{e}' = \mathbf{x}/|\mathbf{x}|$ is the direction of scattering, and $\hat{e}$ denotes the direction of incidence. Equation (8a) may be evaluated using (3a):

$$R(\hat{e},\hat{e}',\tau) = -(4\pi)^{-1}\int d^3\mathbf{x}\, U^+(\tau+\hat{e}'\cdot\mathbf{x},\hat{e},\mathbf{x})V(\mathbf{x}). \tag{9}$$

It was noted earlier that in the absence of bound states, $U^+(t,\hat{e},\mathbf{x})$ vanishes in the scattering region as $t\to\pm\infty$. Consequently, we may infer from Eq. (9) that $R(\hat{e},\hat{e}',\tau)$ vanishes as $\tau\to\pm\infty$ under the same conditions.

The inverse problem is to determine the potential $V(\mathbf{x})$ given the impulse response function $R(\hat{e},\hat{e}',\tau)$.

B. Variable velocity scalar wave equation

The classical wave equation may be written more conveniently as

$$[\Delta - \partial_t^2 - v(\mathbf{x})\partial_t^2]u(t,\hat{e},\mathbf{x}) = 0. \tag{10}$$

Here

$$v(\mathbf{x}) = [1/c^2(\mathbf{x})] - 1. \tag{11}$$

We remind the reader that for the classical wave equation we use lowercase characters to denote the potential and field variables. For the scattering experiment described, Eq. (10) can be rewritten as

$$u^{\pm}(t,\hat{e},\mathbf{x}) = u_0^{\pm}(t,\hat{e},\mathbf{x}) + \int d^3\mathbf{x}'\,dt' \times g_0^{\pm}(t-t',\mathbf{x},\mathbf{x}')v(\mathbf{x}')\partial_{t'}^2 u^{\pm}(t,\hat{e},\mathbf{x}') \tag{12a}$$

or

$$u^{\pm}(t,\hat{e},\mathbf{x}) = u_0^{\pm}(t,\hat{e},\mathbf{x}) + \int d^3\mathbf{x}'\,dt' \times g^{\pm}(t-t',\mathbf{x},\mathbf{x}')v(\mathbf{x}')\partial_{t'}^2 u_0^{\pm}(t',\hat{e},\mathbf{x}'). \tag{12b}$$

Here $u_0^{\pm}(t,\hat{e},\mathbf{x})$ is given by $\delta(t-\hat{e}\cdot\mathbf{x})$. The Green's functions appearing in Eq. (12) are defined by

$$(\Delta - \partial_t^2)g_0^{\pm}(t-t',\mathbf{x},\mathbf{x}') = \delta^3(\mathbf{x}-\mathbf{x}')\delta(t-t'), \tag{13a}$$

and

$$(\Delta - \partial_t^2 - v(\mathbf{x})\partial_t^2)g^{\pm}(t-t',\mathbf{x},\mathbf{x}') = \delta^3(\mathbf{x}-\mathbf{x}')\delta(t-t'). \tag{13b}$$

Equation (13a) is identical to (4a) and explicitly

$$g_0^{\pm}(\mathbf{x},\mathbf{x}',t-t') = -(4\pi|\mathbf{x}-\mathbf{x}'|)^{-1}\delta(t-t'\mp|\mathbf{x}-\mathbf{x}'|). \tag{14}$$

Again, $+$ and $-$ refer, respectively, to radiation and incoming boundary conditions. Further, $g^{\pm}$ and $g_0^{\pm}$ are related via

$$g^{\pm}(t-t',\mathbf{x},\mathbf{x}') = g_0^{\pm}(t-t',\mathbf{x},\mathbf{x}') + \int d^3\mathbf{x}''\,dt'' \times g^{\pm}(t-t'',\mathbf{x},\mathbf{x}'')v(\mathbf{x}'')\partial_{t''}^2 \times g_0^{\pm}(t''-t',\mathbf{x}'',\mathbf{x}). \tag{15}$$

Finally, u^+ and u^- are related via

$$u^-(t,\hat{e},\mathbf{x}) = u^+(-t,-\hat{e},\mathbf{x}). \tag{16}$$

Equation (16) follows immediately from Eqs. (12b) and (14).

Causality also plays a crucial role for the variable velocity scalar wave equation. By hypothesis we have chosen the velocity everywhere to be less than or equal to 1. Consequently, as with the plasma wave equation, $u^+(t,\hat{e},\mathbf{x}) = 0$ if $t<\hat{e}\cdot\mathbf{x}$. Further $g^+(t-t',\mathbf{x},\mathbf{x}')$ and $g_0^+(t-t',\mathbf{x},\mathbf{x}')$ are zero for $(t-t') < |\mathbf{x}-\mathbf{x}'|$ and $g^-(t-t',\mathbf{x},\mathbf{x}')$ and $g_0^-(t-t',\mathbf{x},\mathbf{x}')$ are zero for $(t-t') > |\mathbf{x}-\mathbf{x}'|$.

The impulse response for the classical wave equation is defined by

$$r(\hat{e},\hat{e}',\tau) = \lim_{\substack{t,|\mathbf{x}|\to\infty \\ \tau=t-|\mathbf{x}|}} |\mathbf{x}|(u(t,\hat{e},\mathbf{x}) - \delta(t-\hat{e}\cdot\mathbf{x})). \tag{17a}$$

In other words, for large $|\mathbf{x}|$,

$$u^{\pm}(t,\hat{e},\mathbf{x}) = \delta(t-\hat{e}\cdot\mathbf{x}) + \frac{r(\pm\hat{e},\hat{x},\pm t-|\mathbf{x}|)}{|\mathbf{x}|} + O\left(\frac{1}{|\mathbf{x}|^2}\right). \tag{17b}$$

Large $|\mathbf{x}|$ analysis of (12a) yields

$$r(\hat{e},\hat{e}',\tau) = -(4\pi)^{-1}\int d^3\mathbf{x}\,\partial_\tau^2 u^+(\tau+\hat{e}'\cdot\mathbf{x},\hat{e},\mathbf{x})v(\mathbf{x}), \tag{18}$$

and as above $r(\hat{e},\hat{e}',\tau) = 0$ as $\tau\to\pm\infty$.

The inverse problem is to recover the velocity, $c(\mathbf{x})$, given the impulse response function.

III. TIME DOMAIN REPRESENTATION THEOREM

The representation theorem states that if the Green's function ($g^{\pm}$ or $G^{\pm}$) is known and if the wave field and its normal derivative are specified for all time on a closed, simply connected smooth surface $\partial S'$, then the field can be found at all interior points $(\mathbf{x},t)$. Thus the representation theorem gives a formal solution of the boundary value problem provided the Green's function is known. Below we will review the representation theorem for the two wave equations we are discussing. The approach in the text is designed for maximum physical clarity. A careful treatment of the representation theorem is presented in Appendix A. This derivation is given since we have not been able to find a reference for the results in the form needed below.

A. Plasma wave equation

The derivation of the time domain representation theorem is presented for completeness. First multiply the plasma wave equation [Eq. (1)] by $G^{\pm}(t-t',\mathbf{x},\mathbf{x}')$ to obtain

$$G^{\pm}(t-t',\mathbf{x},\mathbf{x}')(\Delta' - \partial_{t'}^2 - V(\mathbf{x}'))U^+(t',\hat{e},\mathbf{x}') = 0. \tag{19}$$

Here Δ' is the Laplacian for $\mathbf{x}'$ coordinates. Then multiply Eq. (4b) (which defines $G^{\pm}$) by $U^+(t',\hat{e},\mathbf{x}')$ to obtain

$$\begin{aligned} U^+(t',\hat{e},\mathbf{x}')(\Delta' - \partial_{t'}^2 - V(\mathbf{x}'))G^{\pm}(t-t',\mathbf{x},\mathbf{x}') \\ = \delta(t-t')\delta^3(\mathbf{x}-\mathbf{x}')U^+(t',\hat{e},\mathbf{x}'). \end{aligned} \tag{20}$$

Subtracting (19) from (20)

$$\begin{aligned} &U^+(t',\hat{e},\mathbf{x}')\delta(t-t')\delta^3(\mathbf{x}-\mathbf{x}') \\ &= [U^+(t',\hat{e},\mathbf{x}')\Delta' G^{\pm}(t-t',\mathbf{x},\mathbf{x}') \\ &\quad - G^{\pm}(t-t',\mathbf{x},\mathbf{x}')\Delta' U^+(t',\hat{e},\mathbf{x}')] \\ &\quad - [U^+(t',\hat{e},\mathbf{x}')\partial_{t'}^2 G^{\pm}(t-t',\mathbf{x},\mathbf{x}') \\ &\quad - G^{\pm}(t-t',\mathbf{x},\mathbf{x}')\partial_{t'}^2 U^+(t',\hat{e},\mathbf{x}')]. \end{aligned} \tag{21}$$

Let S' be the volume contained within the surface $\partial S'$. Upon integrating (21) over S' and $-\infty < t' < \infty$ one finds

$$\begin{aligned} &U^+(t,e,\mathbf{x}) \\ &= \int_{S'} d^3x' \int_{-\infty}^{\infty} dt' \, \nabla' \cdot (U^+\nabla' G^{\pm} - G^{\pm}\nabla' U^+) \\ &\quad - \int_{S'} d^3x' \int_{-\infty}^{\infty} dt' \frac{\partial}{\partial t'} \\ &\quad \times (U^+ \partial_{t'} G^{\pm} - G^{\pm} \partial_{t'} U^+). \end{aligned} \tag{22}$$

The second term on the right-hand side vanishes upon integration since $U^+(t',\hat{e},\mathbf{x}') = 0$ as $t' \to \pm\infty$. Using Green's theorem the first term on the right-hand side leads to

$$\begin{aligned} &U^+(t,\hat{e},\mathbf{x}) \\ &= \int_{\partial S'} dS' \int_{-\infty}^{\infty} dt' \Big[U^+(t',\hat{e},\mathbf{x}') \frac{\partial G^{\pm}}{\partial n}(t-t',\mathbf{x},\mathbf{x}') \\ &\quad - G^{\pm}(t-t',\mathbf{x},\mathbf{x}') \frac{\partial U^+}{\partial n}(t',\hat{e},\mathbf{x}') \Big], \end{aligned} \tag{23}$$

for $\mathbf{x} \in S'$ and $U^+(t,\hat{e},\mathbf{x}) = 0$ otherwise. Equation (23) is the general form of the representation theorem. For the inverse scattering problem we will let S' be a ball with arbitrarily large radius centered about the origin. The integration over the boundary $\partial S'$ then involves only the far-field form of Green's function and the scattering data $R(t',\hat{e},\mathbf{x}')$, which suffice to determine $U^{\pm}(t',\hat{e},\mathbf{x}')$ as $|x'| \to \infty$. Consequently the wave field for a point $\mathbf{x}$ interior to $\partial S'$ can be determined from the scattering data once the far-field Green's function is known.

The representation theorem (23) is actually two equations since it holds if either G^+ or G^- is used. The derivation sketched above also holds if U^- is substituted everywhere for U^+. One finds

$$\begin{aligned} &U^-(t,\hat{e},\mathbf{x}) \\ &= \int_{\partial S'} dS' \int_{-\infty}^{\infty} dt' \Big[U^-(t',\hat{e},\mathbf{x}') \frac{\partial}{\partial n} G^+(t-t',\mathbf{x},\mathbf{x}') \\ &\quad - G^+(t-t',\mathbf{x},\mathbf{x}') \frac{\partial U^-}{\partial n}(t',\hat{e},\mathbf{x}') \Big], \end{aligned} \tag{24}$$

and

$$\begin{aligned} &U^-(t,\hat{e},\mathbf{x}) \\ &= \int_{\partial S'} dS' \int_{-\infty}^{\infty} dt' \Big[U^-(t',\hat{e},\mathbf{x}') \frac{\partial}{\partial n} G^-(t-t',\mathbf{x},\mathbf{x}') \\ &\quad - G^-(t-t',\mathbf{x},\mathbf{x}') \frac{\partial U^-}{\partial n}(t',\hat{e},\mathbf{x}') \Big]. \end{aligned} \tag{25}$$

It follows from our discussion of the causal properties of G^+ that the representation theorem in the form

$$U^+ = -\int dS' \, dt' \Big[G^+ \frac{\partial U^+}{\partial n} - U^+ \frac{\partial G^+}{\partial n} \Big] \tag{26}$$

takes the initial data (incident field) and propagates it to the future. On the other hand, the form of the representation theorem

$$U^+ = -\int dS' \, dt' \Big[G^- \frac{\partial U^+}{\partial n} - U^+ \frac{\partial G^-}{\partial n} \Big] \tag{27}$$

propagates the final data (scattered field) into the past. Since we know the scattered data at late time and wish to reconstruct the field at earlier times, Eq. (27) is central to our development of inverse scattering.

B. Classical wave equation

The derivation of the representation theorem for the classical wave equation is almost identical to that for the plasma wave equation. The major difference can be seen by looking at the analog of Eq. (22), which is

$$\begin{aligned} u^+(t,\hat{e},\mathbf{x}) &= \int_{S'} d^3\mathbf{x}' \int_{-\infty}^{\infty} dt' \, \nabla' \cdot (u^+\nabla' g^{\pm} - g^{\pm}\nabla' u^+) \\ &\quad - \int_{S'} d^3\mathbf{x}' \int_{-\infty}^{\infty} dt'(1 + v(\mathbf{x}')) \\ &\quad \times \frac{\partial}{\partial t'}(u^+ \partial_{t'} g^{\pm} - g^{\pm} \partial_{t'} u^+). \end{aligned} \tag{28}$$

In the second term on the right-hand side of (28) the potential occurs. However, the entire term vanishes upon integration by parts since $u^+(t,\hat{e},\mathbf{x}) = 0$ for fixed $\mathbf{x}$ as $t \to \pm\infty$. Consequently the representation theorem for the classical wave equation is exactly of the same form as for the plasma wave equation

$$\begin{aligned} u^{\pm}(t,\hat{e},\mathbf{x}) &= \int_{\partial S'} dS' \int_{-\infty}^{\infty} dt' \Big\{ u^{\pm}(t',\hat{e},\mathbf{x}') \frac{\partial g^{\pm}}{\partial n}(t-t',\mathbf{x},\mathbf{x}') \\ &\quad - g^{\pm}(t-t',\mathbf{x},\mathbf{x}') \frac{\partial u^{\pm}}{\partial n}(t',\hat{e},\mathbf{x}') \Big\}. \end{aligned} \tag{29}$$

The interpretation of Eq. (29) in terms of the causal properties of $u^{\pm}$ and $g^{\pm}$ remains the same as previously.

IV. FAR-FIELD FORM OF GREEN'S FUNCTION

The Green's function simplifies in the far-field limit.

A. Plasma wave equation

The far-field form of the free-space Green's function is obtained by letting $|\mathbf{x}'|$ become arbitrarily large in Eq. (5) and is

$$G_0^{\pm}(t-t',\mathbf{x},\mathbf{x}') = -\frac{1}{4\pi|\mathbf{x}'|}\delta(|\mathbf{x}'|-\hat{x}'\cdot\mathbf{x}\mp(t-t'))+O\left(\frac{1}{|\mathbf{x}'|^2}\right). \tag{30}$$

Substitution of $G_0^{\pm}$ into the equation $G^{\pm} = G_0^{\pm} + G_0^{\pm} VG^{\pm}$ for $|\mathbf{x}'|$ arbitrarily large yields

$$\begin{aligned} G^{\pm}(t-t',\mathbf{x},\mathbf{x}') &= -\frac{1}{4\pi|\mathbf{x}'|}\Big[\delta(|\mathbf{x}'|-\hat{x}'\cdot\mathbf{x}\mp(t-t')) \\ &\quad + \int d^3\mathbf{x}''\,dt''\,G^{\pm}(t-t'',\mathbf{x},\mathbf{x}'')V(\mathbf{x}'') \\ &\quad \times\delta(|\mathbf{x}'|-\hat{x}'\cdot\mathbf{x}''\mp(t''-t'))\Big]+O\left(\frac{1}{|\mathbf{x}'|^2}\right). \end{aligned} \tag{31}$$

The term inside the bracket is the right-hand side of Eq. (36) with $u_0=\delta(t-t'\mp(|\mathbf{x}'|+\hat{x}'\cdot\mathbf{x}))$. Consequently, Eq. (31) for $|\mathbf{x}'|$ arbitrarily large becomes

$$G^{\pm}(t-t',\mathbf{x},\mathbf{x}') = -\frac{1}{4\pi|\mathbf{x}'|}U^{\pm}(t-t'\mp|\mathbf{x}'|,\hat{x}',\mathbf{x})+O\left(\frac{1}{|\mathbf{x}'|^2}\right). \tag{32}$$

Equation (32) is derived in Appendix B where the order of the remainder term is established.

B. Classical wave equation

The derivation is again precisely the same as above. Equation (30) is replaced by

$$g_0^{\pm}(t-t',\mathbf{x},\mathbf{x}') = -\frac{1}{4\pi|\mathbf{x}'|}\delta(|\mathbf{x}'|-\hat{x}'\cdot\mathbf{x}\mp(t-t')) \tag{33}$$

as $|\mathbf{x}'|$ becomes arbitrarily large. Substitution of (33) in the equation $g^{\pm}=g_0^{\pm}+g_0^{\pm}v\,\partial_t^2 g_0^{\pm}$ yields

$$\begin{aligned} g^{\pm}(t-t',\mathbf{x},\mathbf{x}') &= -\frac{1}{4\pi|\mathbf{x}'|}\Big[\delta(|\mathbf{x}'|-\hat{x}'\cdot\mathbf{x}\mp(t-t')) \\ &\quad + \int d^3\mathbf{x}''\,dt''\,g^{\pm}(t-t'',\mathbf{x},\mathbf{x}'')v(\mathbf{x}'') \\ &\quad \times\partial_{t''}^2\delta(|\mathbf{x}|-\hat{x}'\cdot\mathbf{x}''\mp(t''-t'))\Big]. \end{aligned} \tag{34}$$

Equation (34) simplifies to yield the large $|\mathbf{x}'|$ expansion

$$g^{\pm}(t-t',\mathbf{x},\mathbf{x}') = -\frac{1}{4\pi|\mathbf{x}'|}u^{\pm}(t-t'\mp|\mathbf{x}'|,\hat{x}',\mathbf{x}) + O\left(\frac{1}{|\mathbf{x}'|^2}\right). \tag{35}$$

C. Heuristic arguments

An essential feature of the above results for $G^{\pm}$ and $g^{\pm}$ is that in the far field the Green's function becomes proportional to the wave field. Below we present a heuristic argument, which is designed to explain the physical origin of this proportionality. A careful derivation is given in Appendix B. This result could also be obtained by applying a stationary phase argument to the bilinear expansion of the Green's function in terms of the eigenfunctions.

The far-field Green's functions just derived can be justified on physical grounds. The Green's function $g^+(\mathbf{x},\mathbf{x}',t-t')$ gives the wave field at $(\mathbf{x},t)$ due to a delta function excitation at $(\mathbf{x}',t')$. The response to such an excitation is a spherically spreading wave

$$-\frac{1}{4\pi}\frac{\delta(t-t'-|\mathbf{x}-\mathbf{x}'|)}{|\mathbf{x}-\mathbf{x}'|}. \tag{36}$$

In the absence of the potential the response at $(\mathbf{x},t)$ as $|\mathbf{x}-\mathbf{x}'|\to\infty$ is

$$-(4\pi|\mathbf{x}'|)^{-1}\delta(t-t'-|\mathbf{x}'|+\hat{x}'\cdot\mathbf{x}). \tag{37}$$

That is, the wave field is essentially a plane wave propagating in the direction $-\hat{x}'$ if the potential is absent. If a potential is present the wave field is given by u^+, the field induced at $(\mathbf{x},t)$ by the incident plane wave. The Green's function thus is argued to be

$$g^+(t-t',\mathbf{x},\mathbf{x}') = -(4\pi|\mathbf{x}'|)^{-1}u^+(t-t'-|\mathbf{x}'|,\hat{x}',\mathbf{x}) + O(1/|\mathbf{x}'|^2). \tag{38}$$

This is, of course, in agreement with Eqs. (32) and (35). The argument for the plasma wave equation is the same.

V. INVERSE SCATTERING EQUATIONS

In this section we derive an integral equation that relates the wave field to the scattering data. The scattering data are now assumed to be measured on the boundary $\partial S'$ of a large ball S' centered about the origin of coordinates. Roughly, the integral equation is obtained by using the far-field asymptotics to evaluate the representation theorem on the surface $\partial S'$. The result yields the wave field interior to S'. The integral equation comes about since the far-field form of Green's function is proportional to the wave field [Eqs. (32) and (35)].

A. Plasma wave equation

The integrations in Eqs. (25) and (27) (the representation theorem for U^+ and U^-) are taken over the surface $\partial S'$ and the equations are then subtracted. The result is

$$\begin{aligned} &U^+(t,\hat{e},\mathbf{x})-U^-(t,\hat{e},\mathbf{x}) \\ &= \int_{\partial S'} dS'\int_{-\infty}^{\infty}dt'\Big\{(U^+(t',\hat{e},\mathbf{x}')-U^-(t',\hat{e},\mathbf{x}')) \\ &\quad \times\frac{\partial G^-}{\partial n}(t-t',\mathbf{x},\mathbf{x}')-G^-(t-t',\mathbf{x},\mathbf{x}') \\ &\quad \times\left(\frac{\partial U^+}{\partial n}(t',\hat{e},\mathbf{x}')-\frac{\partial U^-}{\partial n}(t',\hat{e},\mathbf{x}')\right)\Big\}. \end{aligned} \tag{39}$$

In (39), we use (8b) and (32); this gives

$$\begin{aligned} &U^+(t,\hat{e},\mathbf{x})-U^-(t,\hat{e},\mathbf{x}) \\ &= -\frac{1}{4\pi}\int_{S^2}d^2\hat{x}'\int_{-\infty}^{\infty}dt'[R(\hat{e},\hat{x}',t'-|\mathbf{x}'|) \\ &\quad -R(-\hat{e},\hat{x}',-t'-|\mathbf{x}'|)] \\ &\quad \times\frac{\partial}{\partial|\mathbf{x}'|}U^-(t-t'+|\mathbf{x}'|,\hat{x}',\mathbf{x}) \end{aligned}$$

$$+\frac{1}{4\pi}\int_{S^2} d^2\hat{x}'\int_{-\infty}^{\infty} dt'\, U^-(t-t'-|\mathbf{x}'|,\hat{x}',\mathbf{x})$$
$$\times\frac{\partial}{\partial|\mathbf{x}'|}\,[R(\hat{e},\hat{x}',t'-|\mathbf{x}'|)$$
$$-R(-\hat{e},\hat{x}',-t'-|\mathbf{x}'|)]. \tag{40}$$

Here S^2 denotes the unit sphere. In (40), we change the $|\mathbf{x}'|$ derivatives to t' derivatives (putting in the necessary minus signs) and integrate by parts in the first t' integral. The terms involving $\partial_{t'}R(-\hat{e},\hat{x}',-t'-|\mathbf{x}'|)$ cancel, leaving

$$U^+(t,\hat{e},\mathbf{x})-U^-(t,\hat{e},\mathbf{x})$$
$$=-\frac{1}{2\pi}\int_{S^2} d^2\hat{x}'\int_{-\infty}^{\infty} dt'\, U^-(t-t'+|\mathbf{x}'|,\hat{x}',\mathbf{x})$$
$$\times\partial_{t'}R(\hat{e},\hat{x}',t'-|\mathbf{x}'|). \tag{41}$$

In (41), we make the change of variables $\tau=t'-|\mathbf{x}'|$ and use (7). We obtain

$$U^+(t,\hat{e},\mathbf{x})-U^-(t,\hat{e},\mathbf{x})$$
$$=-\frac{1}{2\pi}\int_{S^2} d^2\hat{x}'\int_{-\infty}^{\infty} d\tau\, U^+(\tau-t,-\hat{x}',\mathbf{x})$$
$$\times\partial_\tau R(\hat{e},\hat{x}',\tau). \tag{42}$$

For $t>\hat{e}\cdot\mathbf{x}$ causality insures $U^-(t,\hat{e},\mathbf{x})=0$.

If we write

$$U^+(t,\hat{e},\mathbf{x})=\delta(t-\hat{e}\cdot\mathbf{x})+U^{+\mathrm{sc}}(t,\hat{e},\mathbf{x}), \tag{43}$$

then for $t>\hat{e}\cdot\mathbf{x}$

$$U^{+\mathrm{sc}}(t,\hat{e},\mathbf{x})$$
$$=-\frac{1}{2\pi}\int_{S^2} d^2\hat{x}'\,\frac{d}{dt}R(\hat{e},\hat{x}',t-\hat{x}'\cdot\mathbf{x})$$
$$-\frac{1}{2\pi}\int_{S^2} d^2\hat{x}'\int_{-\infty}^{\infty} d\tau\,\frac{d}{dt}R(\hat{e},\hat{x}',\tau+t)$$
$$\times U^{+\mathrm{sc}}(\tau,-\hat{x}',\mathbf{x}). \tag{44}$$

This is precisely the Newton–Marchenko equation for the plasma wave equation. It is a Fredholm II equation with eigenvalues in the closed interval $[-1,1]$ (see Refs. 14 and 15). For sufficiently weak potentials, Eq. (44) can be solved by iteration for U^+ if the scattering data are given. As we will see in Sec. VI, the potential can be recovered from the wave field U^+.

B. Classical wave equation

The same procedure works for the classical wave equation. We subtract the equations (29) involving g^- to obtain a representation for u^+-u^-. We then use the asymptotics (35) and (17b) as before to obtain

$$u^+(t,\hat{e},\mathbf{x})$$
$$=u^-(t,\hat{e},\mathbf{x})-\frac{1}{2\pi}\int_{S^2} d^2\hat{x}'\,\frac{d}{dt}r(\hat{e},\hat{x}',t-\hat{x}'\cdot\mathbf{x})$$
$$-\frac{1}{2\pi}\int_{S^2} d^2\hat{x}'\int_{-\infty}^{\infty} d\tau\,\frac{d}{dt}r(\hat{e},\hat{x}',\tau+t)$$
$$\times u^{+\mathrm{sc}}(\tau,-\hat{x}',\mathbf{x}). \tag{45}$$

Since we require that $c(\mathbf{x})\leqslant 1$ everywhere, it follows that $u^-(t,\hat{e},\mathbf{x})=0$ if $t>\hat{e}\cdot\mathbf{x}$ (see Ref. 18). Consequently the following integral equation is obtained for $t>\hat{e}\cdot\mathbf{x}$

$$u^{+\mathrm{sc}}(t,\hat{e},\mathbf{x})=-\frac{1}{2\pi}\int d^2\hat{x}'\,\frac{d}{dt}r(\hat{e},\hat{x}',t-\hat{x}'\cdot\mathbf{x})$$
$$-\frac{1}{2\pi}\int_{S^2} d^2\hat{x}'\int_{-\infty}^{\infty} d\tau\,\frac{d}{dt}r(\hat{e},\hat{x}',\tau+t)$$
$$\times u^{+\mathrm{sc}}(\tau,-\hat{x}',\mathbf{x}). \tag{46}$$

The reader is reminded that Eq. (46) is quite different from Eq. (44). Indeed little is known about (46) at present. This is due to the fact that the potential term for the classical wave equation has the form $V(\mathbf{x})\partial_t^2 u$ in contrast to the plasma wave equation whose potential term is Vu. Physically we note that this corresponds to the fact that PWE solutions propagate in straight lines, whereas in general the characteristics are curved for the classical wave equation. If Eq. (46) has a unique solution it determines the wave field from the scattered data. The extraction of the potential from this wave field solution is presented in the next section.

VI. RECOVERY OF THE VELOCITY/POTENTIAL

The integral equation that allows the reconstruction of the wave field from the impulse response function is formally identical for both wave equations. There are significant differences and also similarities in the recovery of the potential from the reconstructed wave field. The wave fronts (characteristic surfaces) play an essential role. However, the way the velocity or the potential is extracted differs.

A. Plasma wave equation

For this equation the characteristic surface occurs at $t=\hat{e}\cdot\mathbf{x}$. Further, the wave field near the characteristic surface may be expanded as

$$U^+(t,\hat{e},\mathbf{x})=\delta(t-\hat{e}\cdot\mathbf{x})+B(\hat{e},\mathbf{x})\Theta(t-\hat{e}\cdot\mathbf{x})+\dots . \tag{47}$$

Here $\Theta(y)$ is the Heaviside step function

$$\Theta(y)=1,\quad y\geqslant 0,$$
$$\Theta(y)=0,\quad y<0.$$

As discussed by Morawetz[5] and others,

$$V(\mathbf{x})=-2\hat{e}\cdot\nabla B(\hat{e},\mathbf{x}). \tag{48}$$

That is, the potential is determined if we can reconstruct the wave field near the characteristic surface.

B. Classical wave equation

In this section the recovery of the velocity function, $c(\mathbf{x})$, from the reconstructed wave field $u^+(t,\hat{e},\mathbf{x})$ is discussed. The situation is complicated by the possible presence of caustics (e.g., a focal point within the scattering region). If we initiate the scattering from an incident delta pulse we find at a later time that the solution near the characteristic surface is

$$u^+(t,\hat{e},\mathbf{x})=A(\hat{e},\mathbf{x})\delta(t-s(\hat{e},\mathbf{x}))+\cdots . \tag{49}$$

Here A and s are functions determined by the equations of geometrical optics. In particular, the eikonal equation is

$$|\nabla s(\hat{e},\mathbf{x})|^2=c^{-2}(x) \tag{50}$$

and the first transport equation[19] is

$$2\nabla A\cdot\nabla s+(\Delta s)A=0. \tag{51}$$

Since the wave field u is supposedly known [as a result of solving Eq. (46)], the position of the δ function and therefore $s(\hat{e},\mathbf{x})$ are also known. Equation (50) can then be used to recover $c(\mathbf{x})$.

If caustics occur in the reconstructed wave field the method just described may not be sufficient to determine the velocity everywhere. Clearly the velocity at those points that are not crossed by a characteristic surface cannot be inferred in this way. In this case the velocity can be reconstructed using the wave equation directly, but difficulties are expected to arise due to the lack of high-frequency data in these regions.

VII. PHYSICAL CONSIDERATIONS

The plasma and classical wave equations have been used in other contexts to model many problems in acoustics, electromagnetics, and elastodynamics. Consequently, it is expected that the exact inverse scattering equations developed in this paper, if tractable, will have widespread application. In particular, they provide a common framework in which various aproximate methods can be compared and evaluated.

The simple scalar nature of the wave equations leads one to wonder if there is a physical situation governed essentially exactly by these equations. Such a physical situation could be used for a precise experimental test of the inverse scattering equations.

In this section, we show that both the classical and plasma wave equations may be realized by the propagation of pressure waves in a quiescent fluid. The classical wave equation is realized when the fluid is chosen to have the same density everywhere [however, the compressibility, $\kappa(\mathbf{x})$, may vary in the scattering region]. As will be discussed, the plasma wave equation is realized if the velocity is constant throughout the fluid; however, $\kappa(\mathbf{x})$ and the density $\rho(\mathbf{x})$ vary inversely to each other in the scattering region.

Pressure propagation in an isotropic quiescent fluid is governed by the equation

$$[1/\rho(\mathbf{x})]\Delta p - [1/\rho^2(\mathbf{x})]\nabla\rho\cdot\nabla p - \kappa(\mathbf{x})\partial_t^2 p = 0, \quad (52)$$

where $p(t,\mathbf{x})$ is the pressure at $(\mathbf{x},t)$.

If the density is assumed to be constant, Eq. (52) becomes

$$\Delta p - [1/c^2(\mathbf{x})]\partial_t^2 p = 0. \quad (53)$$

Here we have used $c^2(\mathbf{x}) = 1/(\kappa(\mathbf{x})\rho(\mathbf{x}))$. Equation (53) is just the classical wave equation. Consequently, the construction of experiments (sound scattering in fluid tanks) to test the classical wave inverse scattering equation appear feasible.

The plasma wave equation may also be realized by sound scattering in a fluid tank. First make the transformation $p(t,\mathbf{x}) = \sqrt{\rho(x)}\,\psi(t,\mathbf{x})$ (see Ref. 20). One obtains

$$\Delta\psi - v(\mathbf{x})\psi - [1/c^2(x)]\partial_t^2\psi = 0. \quad (54)$$

Here

$$v(\mathbf{x}) = \tfrac{3}{4}(|\nabla\rho|^2/\rho^2 - \tfrac{2}{3}(\Delta\rho/\rho)). \quad (55)$$

If the velocity $c(\mathbf{x})$ is chosen to be constant everywhere, Eq. (54) reduces to the plasma wave equation.

An experiment to test the plasma wave inverse method might proceed as follows. First, fluids with a common acoustic velocity but differing densities are mixed to obtain the potential given in Eq. (55). Then sound waves are scattered from the potential and the pressure is measured in the far field. Far from the scattering region the density is a constant ρ_0 and the scattered wave field ψ^{sc} is related to the scattered pressure by

$$\psi^{sc} = p^{sc}\sqrt{\rho_0}. \quad (56)$$

This provides the input data for the inversion procedure, Eq. (45).

The plasma wave equation can be related to the Schrödinger equation by a Fourier transform. Consequently, the pressure waves scattered by the potential in Eq. (54) can in principle be related to the quantum mechanical scattering amplitudes for the same potential.

ACKNOWLEDGMENTS

We would like to thank Professor P. Perry, Professor P. C. Sabatier, and Professor C. Hammer for useful discussions. This work was begun and Ref. 1 was written while two of the authors JHR and BDF were guest scientists at the Institute for Theoretical Physics, Chalmers Technical University, and while MC was a research associate at Stanford University.

This work was supported in part by the Office of Naval Research under Contract Nos. N000-14-83-K0038 and N000-14-85-K0224. The work at Stanford of one author (MC) was also supported in part by the Air Force Office of Scientific Research, the National Science Foundation, The Office of Naval Research, and the Army Research Office.

APPENDIX A: DERIVATION OF THE REPRESENTATION THEOREM

This appendix contains a careful derivation of the representation theorem. The derivations are essentially the same for the plasma wave equation and for the variable velocity wave equation; we derive the representation theorem for the former and then explain the modifications necessary for the latter.

1. Plasma wave equation

The derivation will be carried out in the frequency domain; in other words, it will be carried out explicitly for the Schrödinger equation

$$[\Delta + k^2 + V(\mathbf{x})]\psi(k,\mathbf{x}) = 0. \quad (A1)$$

When there are no bound or half-bound states,[21] Eqs. (A1) and (1) are related[22] by the Fourier transform

$$U(t,\mathbf{x}) = F\psi(k,\mathbf{x}) = (2\pi)^{-1}\int_{-\infty}^{\infty} \exp(-ikt)\psi(k,\mathbf{x})dk. \quad (A2)$$

When V is positive and in C_0^∞, there are no bound states. Frequency domain notation is defined in Ref. 22.

In our proof of the representation theorem, we will need information about the smoothness of the Green's function $G(k,\mathbf{x},\mathbf{y})$, which is the kernel of the resolvent $(\Delta - V + k^2)^{-1}$. We will obtain this information from the theory of elliptic partial differential equations.[23] To do this, we will compare G to G_0, which is the kernel of $(\Delta + k^2)^{-1}$. In what follows, we shall suppress the dependence of G and G_0 on k.

The kernels G and G_0 are related by the equation

$$G(\mathbf{x},\mathbf{y}) - G_0(\mathbf{x},\mathbf{y}) = \int G_0(\mathbf{x},\mathbf{z})V(\mathbf{z})G(\mathbf{z},\mathbf{y})d^3\mathbf{z}. \qquad \text{(A3)}$$

We multiply (A3) by $|V(\mathbf{x})|^{1/2}$ and define

$$Kf(\mathbf{x}) \equiv \int |V(\mathbf{x})|^{1/2}G_0(\mathbf{x},\mathbf{y})V(\mathbf{y})|V(\mathbf{y})|^{-1/2}f(\mathbf{y})d^3\mathbf{y}.$$

With this notation, (A3) can be written

$$|V|^{1/2}G_0 = (I-K)(|V|^{1/2}G). \qquad \text{(A4)}$$

Newton[24] has shown that under our asumptions on V the operator $I-K$ (which depends on k) is invertible on $L^2(R^3)$ for all k. We use this fact in proving the following proposition.

Proposition: Let $V(\mathbf{x})$ be a bounded L^1 function, and assume that $I-K$ is invertible. Then for each y, $G(\cdot,\mathbf{y}) - G_0(\cdot,\mathbf{y})$ is bounded.

Proof: We write (A4) as

$$|V|^{1/2}G = (I-K)^{-1}(|V|^{1/2}G_0). \qquad \text{(A5)}$$

It is easily seen that for fixed $\mathbf{y}$, $|V|^{1/2}G_0(\cdot,\mathbf{y})$ is in L^2. Since $(I-K)$ is invertible, $|V|^{1/2}G$ is also in L^2. We apply the Schwarz inequality to (A3), obtaining

$$|G-G_0| = \left|\int GVG_0\right| \leqslant \|G|V|^{1/2}\|_2\,\|G_0|V|^{1/2}\|_2 < \infty.$$

Q.E.D.

In what follows we denote by $\Omega_\mathbf{y}$ any bounded, open set whose closure does not contain the point $\mathbf{y}$.

Next we will show that $G-G_0$ is actually Hölder continuous; we recall that $G-G_0$ is said to be Hölder continuous at x_0 with exponent α if

$$[G-G_0]_{\alpha;\mathbf{x}_0} = \sup_{\mathbf{x}\in\Omega_\mathbf{y}} \frac{|(G-G_0)(\mathbf{x}) - (G-G_0)(\mathbf{x}_0)|}{|\mathbf{x}-\mathbf{x}_0|^\alpha}$$

is finite.

Lemma: Let $V(\mathbf{x})$ be a bounded L^1 function, and assume that $I-K$ is invertible. Then $G(\cdot,y)$ is Hölder continuous in $\Omega_\mathbf{y}$ with exponent $\alpha \leqslant 1$.

Proof: We shall show that G is Hölder continuous with exponent α by showing that the same is true for $G-G_0$.

We show below that $G_0(\mathbf{x},\mathbf{y}) = -\exp(ik|\mathbf{x}-\mathbf{y}|)/(4\pi|\mathbf{x}-\mathbf{y}|)$ is Hölder continuous in $\Omega_\mathbf{y}$.

We must show that

$$[G-G_0]_{\alpha;\mathbf{x}_0} = \sup_{\mathbf{x}\in\Omega_\mathbf{y}} \frac{|(G-G_0)(\mathbf{x}) - (G-G_0)(\mathbf{x}_0)|}{|\mathbf{x}-\mathbf{x}_0|^\alpha} \qquad \text{(A6)}$$

is finite. Clearly, by the above proposition, the right side of (A6) is bounded when $\mathbf{x}-\mathbf{x}_0$ is bounded away from zero. We therefore restrict our attention to a ball $B_\epsilon(\mathbf{x}_0)$ of radius ϵ about $\mathbf{x}_0$. We use (A3) to estimate the right side of (A6):

$$\sup_{\mathbf{x}\in\Omega_\mathbf{y}\cap B_\epsilon(\mathbf{x}_0)} \frac{|(G-G_0)(\mathbf{x}) - (G-G_0)(\mathbf{x}_0)|}{|\mathbf{x}-\mathbf{x}_0|^\alpha}$$
$$< \sup_{\mathbf{x}\in\Omega_\mathbf{y}\cap B_\epsilon(\mathbf{x}_0)} \int \frac{|\mathbf{x}_0-\mathbf{z}|e^{ik|\mathbf{x}-\mathbf{z}|} - |\mathbf{x}-\mathbf{z}|e^{ik|\mathbf{x}_0-\mathbf{z}|}|}{|\mathbf{x}-\mathbf{z}|\,|\mathbf{x}_0-\mathbf{z}|\,|\mathbf{x}-\mathbf{x}_0|^\alpha}$$
$$\times|V(\mathbf{z})G(\mathbf{z},\mathbf{y})|d^3\mathbf{z}. \qquad \text{(A7)}$$

Next, we need the estimates

$$|\mathbf{x}_0-\mathbf{z}|e^{ik|\mathbf{x}-\mathbf{z}|} = |\mathbf{x}-\mathbf{z}|e^{ik|\mathbf{x}-\mathbf{z}|} + O(|\mathbf{x}-\mathbf{x}_0|)$$

and

$$\exp(ik|\mathbf{x}_0-\mathbf{z}|) = \exp(ik|\mathbf{x}-\mathbf{z}|) + k\,O(|\mathbf{x}-\mathbf{x}_0|).$$

These we use in the right side of (A7), which is then bounded by

$$\text{const} \sup_{\Omega_\mathbf{y}\cap B_\epsilon(\mathbf{x}_0)} \int \frac{1+k|\mathbf{x}-\mathbf{z}|}{|\mathbf{x}-\mathbf{z}|\,|\mathbf{x}_0-\mathbf{z}|}$$
$$\times V(\mathbf{z})G(\mathbf{z},\mathbf{y})|\mathbf{x}-\mathbf{x}_0|^{1-\alpha}d^3\mathbf{z}.$$

For large values of $|\mathbf{x}-\mathbf{z}|$, the fraction $(1+k|\mathbf{x}-\mathbf{z}|)|\mathbf{x}-\mathbf{z}|^{-1}|\mathbf{x}_0-\mathbf{z}|^{-1}$ is bounded; the only difficulty is for small $|\mathbf{x}-\mathbf{z}|$. For small $|\mathbf{x}-\mathbf{z}|$, $k|\mathbf{x}-\mathbf{z}|$ is negligible compared with 1. We therefore consider the expression

$$\int \frac{|V(\mathbf{z})G(\mathbf{z},\mathbf{y})|}{|\mathbf{x}-\mathbf{z}|\,|\mathbf{x}_0-\mathbf{z}|}d^3\mathbf{z}. \qquad \text{(A8)}$$

We split the integral (A8) into two pieces, an integral over $\Omega_\mathbf{y}$ and an integral over its complement. In the former, we note that V and G are bounded; this gives us

$$\int_{\Omega_\mathbf{y}} \frac{|V(\mathbf{z})G(\mathbf{z},\mathbf{y})|}{|\mathbf{x}-\mathbf{z}|\,|\mathbf{x}_0-\mathbf{z}|}d^3\mathbf{z} \leqslant c\int_{\Omega_\mathbf{y}} \frac{d^3\mathbf{z}}{|\mathbf{x}-\mathbf{z}|\,|\mathbf{x}_0-\mathbf{z}|} < \infty.$$

In the integral over the complement of $\Omega_\mathbf{y}$, we note that $|\mathbf{x}-\mathbf{z}|$ and $|\mathbf{x}_0-\mathbf{z}|$ are both bounded away from zero (because $\mathbf{x}$ and $\mathbf{x}_0$ are in $\Omega_\mathbf{y}$). We apply the Schwarz inequality, obtaining

$$\int_{R^3-\Omega_\mathbf{y}} \frac{|V(\mathbf{z})G(\mathbf{z},\mathbf{y})|}{|\mathbf{x}-\mathbf{z}|\,|\mathbf{x}_0-\mathbf{z}|}d^3\mathbf{z}$$
$$< \text{const}\int_{R^3-\Omega_\mathbf{y}} |V(\mathbf{z})|^{1/2}|V(\mathbf{z})|^{1/2}|G(\mathbf{z},\mathbf{y})|d^3\mathbf{z}$$
$$< \text{const}\,\|V\|_1^{1/2}\|G|V|^{1/2}\|_2 < \infty.$$

Q.E.D.

Theorem: Let $V(\mathbf{x})$ be a bounded L^1 function that is Hölder continuous with exponent $\alpha \leqslant 1$. Assume that $(I-K)$ is invertible. Let $\Lambda_\mathbf{y}$ be an open, bounded, and connected set with $\partial\Lambda_\mathbf{y} \subset \Omega_\mathbf{y}$. Then $G(\cdot,\mathbf{y})$ is in $C^2(\Lambda_\mathbf{y})$.

Proof: In $\Omega_\mathbf{y}$, G satisfies $\Delta G = k^2G - VG$. Let $\phi_\mathbf{y}(\cdot) = G(\cdot,\mathbf{y})|_{\partial\Lambda_\mathbf{y}}$. Then $\phi_\mathbf{y}(\cdot)$ is continuous, and G is a solution of the Dirichlet problem

$$\Delta G = f \quad \text{in } \Lambda_\mathbf{y}$$

and

$$G = \phi_\mathbf{y} \quad \text{on } \partial\Lambda_\mathbf{y},$$

where $f = k^2G - VG$. Since f is bounded and Hölder continuous (with exponent $\alpha \leqslant 1$) in $\Lambda_\mathbf{y}$ this Dirichlet problem has a unique solution given by the Newtonian potential of f (see Ref. 23, p. 55). This Newtonian potential is in $C^2(\Lambda_\mathbf{y})$. This shows that G is in $C^2(\Lambda_\mathbf{y})$.

Q.E.D.

We have completed our investigation of the smoothness of G; we now turn to the representation theorem. This theorem concerns behavior of the Green's function G defined by (A3) and behavior of the wave function ψ defined by the Lippmann–Schwinger equation

$$\psi(k,\hat{e},\mathbf{x}) = \exp(ik\hat{e}\cdot\mathbf{x})$$
$$+ \int G_0(k,\mathbf{x},\mathbf{y})V(\mathbf{y})\psi(k,\hat{e},\mathbf{y})d^3\mathbf{y}. \qquad \text{(A9)}$$

Theorem: Let V satisfy the hypotheses of the above theorem (V positive and C_0^∞ certainly suffices). Then the following equation holds:

$$\int_{S_R(0)} \left[G(k,\mathbf{x},\mathbf{y}) \frac{\partial}{\partial n} \psi(k,\hat{e},\mathbf{x}) - \psi(k,\hat{e},\mathbf{x}) \frac{\partial}{\partial n} G(k,\mathbf{x},\mathbf{y}) \right] dS_x = -\psi(k,\hat{e},\mathbf{y}),$$

where $S_R(0)$ denotes the sphere of radius R centered at zero, n denotes the unit outward normal, and dS_x is the surface area element on $S_R(0)$.

Proof: We begin with the following equations for ψ and G:

$$[\Delta + k^2 - V(\mathbf{x})]\psi(k,\hat{e},\mathbf{x}) = 0, \tag{A10}$$

$$[\Delta + k^2 - V(\mathbf{x})]G(k,\mathbf{x},\mathbf{y}) = \delta(\mathbf{x}-\mathbf{y}). \tag{A11}$$

We multiply (A10) by G and (A11) by ψ and subtract, obtaining

$$\int_{\Omega_{y,\epsilon,R}} [G(k,\mathbf{x},\mathbf{y})\Delta\psi(k,\hat{e},\mathbf{x}) - \psi(k,\hat{e},\mathbf{x})\Delta G(k,\mathbf{x},\mathbf{y})]d^3\mathbf{x} = \int_{\Omega_{y,\epsilon,R}} \psi(k,\hat{e},\mathbf{x})\delta(\mathbf{x}-\mathbf{y})d^3\mathbf{x}, \tag{A12}$$

where we have written $\Omega_{y,\epsilon,R} = B_R(0) - B_\epsilon(\mathbf{y})$. (See Fig. 2.)

The right side of (A12) is zero because $\mathbf{y} \notin \Omega_{y,\epsilon,R}$. Next we apply Green's theorem to the left side. This is legitimate for the following reason. Green's theorem requires continuity of $\nabla \cdot (G\nabla\psi - \psi\nabla G)$ and $G\nabla\psi - \psi\nabla G$. We have shown that G is in $C^2(\Omega_{y,\epsilon,R})$. We need to know that ψ is in $C^2(\Omega_{y,\epsilon,R})$ also. This, however, follows from the same sort of proof as was used to show that G is C^2.

Equation (A12) then becomes

$$\int_{S_R(0)} \left[G(k,\mathbf{x},\mathbf{y}) \frac{\partial}{\partial n} \psi(k,\hat{e},\mathbf{x}) - \psi(k,\hat{e},\mathbf{x}) \frac{\partial}{\partial n} G(k,\mathbf{x},\mathbf{y}) \right] dS_x + \int_{S_\epsilon(\mathbf{y})} \left[G(k,\mathbf{x},\mathbf{y}) \frac{\partial}{\partial \nu} \psi(k,\hat{e},\mathbf{x}) - \psi(k,\hat{e},\mathbf{x}) \frac{\partial}{\partial \nu} G(k,\mathbf{x},\mathbf{y}) \right] dS_x = 0, \tag{A13}$$

where we have used the notation $S_\epsilon(\mathbf{y})$ for the sphere of radius ϵ centered at the point y, and n and ν are normals as shown in Fig. 2.

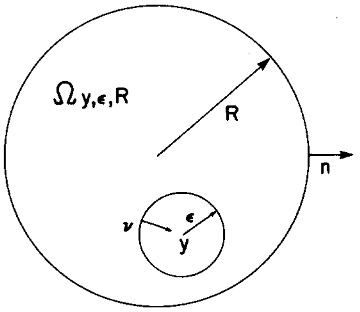


FIG. 2. The geometry appropriate for deriving the representation theorem is shown. The wave field and its derivative are assumed to be known on the large sphere. The value of the wave field is found at the point y.

We want to compute the second integral of (A13). We know that on $\Omega_{y,\epsilon,R}$

$$G(k,\mathbf{x},\mathbf{y}) = G_0(k,\mathbf{x},\mathbf{y}) + f(k,\mathbf{x},\mathbf{y}),$$

where for fixed k and $\mathbf{y}$, f is in $C^2(\Omega_{y,\epsilon,R})$. We can then write the $S_\epsilon(\mathbf{y})$ integral of (A13) as

$$\int_{S_\epsilon(\mathbf{y})} G_0(k,\mathbf{x},\mathbf{y}) \frac{\partial}{\partial \nu} \psi(k,\hat{e},\mathbf{x}) dS_x + \int_{S_\epsilon(\mathbf{y})} \left[f(k,\mathbf{x},\mathbf{y}) \frac{\partial}{\partial \nu} \psi(k,\hat{e},\mathbf{x}) - \psi(k,\hat{e},\mathbf{x}) \frac{\partial}{\partial \nu} f(k,\mathbf{x},\mathbf{y}) ds_x - \int_{S_\epsilon(\mathbf{y})} \psi(k,\hat{e},\mathbf{x}) \frac{\partial}{\partial \nu} G_0(k,\mathbf{x},\mathbf{y}) \right] dS_x. \tag{A14}$$

The integrand of the second integral in (A14) is continuous; this term will therefore disappear in the $\epsilon \to 0$ limit.

In the first term we make the change of variables $\epsilon\hat{\phi} = \mathbf{x} - \mathbf{y}$, with $|\hat{\phi}| = 1$. This first term of (A14) is then

$$-\int_{S_1(0)} \exp[ik\epsilon](4\pi\epsilon)^{-1} \frac{\partial}{\partial \nu} \psi(k,\hat{e},\mathbf{y} + \epsilon\hat{\phi})\epsilon^2\, d^2\hat{\phi}. \tag{A15}$$

Clearly this term also vanishes in the $\epsilon \to 0$ limit. The last term of (A14) is

$$-\int_{S_\epsilon(\mathbf{y})} \psi(k,\hat{e},\mathbf{x}) \left[-\frac{ike^{ik|\mathbf{x}-\mathbf{y}|}}{4\pi|\mathbf{x}-\mathbf{y}|} + \frac{e^{ik|\mathbf{x}-\mathbf{y}|}}{4\pi|\mathbf{x}-\mathbf{y}|^2} \right] \hat{\nu} \cdot \frac{\mathbf{x}-\mathbf{y}}{|\mathbf{x}-\mathbf{y}|} dS_x. \tag{A16}$$

The first term of (A16) disappears in the $\epsilon \to 0$ limit for the same reason that (A15) does. The second term, however, is

$$\int_{S_1(0)} \psi(k,\hat{e},\mathbf{y} + \epsilon\hat{\phi}) \frac{e^{ik\epsilon}}{4\pi\epsilon^2} \epsilon^2\, d^2\hat{\phi},$$

which approaches $\psi(k,\hat{e},\mathbf{y})$ as $\epsilon \to 0$.

We have thus computed the second term of (A13) in the $\epsilon \to 0$ limit. Equation (A13) can therefore be written

$$\int_{S_R(0)} \left[G(k,\mathbf{x},\mathbf{y}) \frac{\partial}{\partial n} \psi(k,\hat{e},\mathbf{x}) - \psi(k,\hat{e},\mathbf{x}) \frac{\partial}{\partial n} G(k,\mathbf{x},\mathbf{y}) \right] dS_x = -\psi(k,\hat{e},\mathbf{y}).$$

Q.E.D.

2. Classical wave equation

Next we consider the modifications necessary to carry out the above arguments for the variable velocity wave equation (2).

We begin with the frequency-domain version of Eq. (10), namely

$$[\Delta + k^2 + k^2V(\mathbf{x})]\psi(k,\mathbf{x}) = 0. \tag{A17}$$

The existence of the distributional Fourier transform needed to relate (A17) to the wave equation (2) or (10) can be inferred from local existence (see below) together with the high-frequency estimates of Ref. 25. These estimates are valid given the hypotheses on the velocity stated in the Introduction, together with the additional condition that

$c^{-2} + \mathbf{x} \cdot \nabla c^{-2}$ be strictly positive. Necessary conditions for existence of the Fourier transform are unknown. The Lippmann–Schwinger equation corresponding to (A17) is

$$\psi^{\pm}(k,\hat{e},\mathbf{x}) = \exp(ik\hat{e}\cdot\mathbf{x}) + \frac{1}{4\pi}\int \frac{\exp(\pm ik|\mathbf{x}-\mathbf{y}|)}{|\mathbf{x}-\mathbf{y}|} \times k^2 V(\mathbf{y})\psi^{\pm}(k,\hat{e},\mathbf{y})d^3\mathbf{y}. \tag{A18}$$

We multiply (A18) by $|V(\mathbf{x})|^{1/2}$, obtaining

$$\xi^{\pm}(k,\hat{e},\mathbf{x}) = \xi^{0}(k,\hat{e},\mathbf{x}) + \int K^{\pm}(k,\mathbf{x},\mathbf{y})\xi^{\pm}(k,\hat{e},\mathbf{y})d^3\mathbf{y}, \tag{A19}$$

where

$$\xi^{\pm}(k,\hat{e},\mathbf{x}) = |V(\mathbf{x})|^{1/2}\psi^{\pm}(k,\hat{e},\mathbf{x}),$$

$$\xi^{0}(k,\hat{e},\mathbf{x}) = |V(\mathbf{x})|^{1/2}\exp(ik\hat{e}\cdot\mathbf{x}),$$

$$K^{\pm}(k,\mathbf{x},\mathbf{y}) = \frac{k^2}{4\pi}\frac{|V(\mathbf{x})|^{1/2}\exp(\pm ik|\mathbf{x}-\mathbf{y}|)}{|\mathbf{x}-\mathbf{y}|\,|V(\mathbf{y})|^{1/2}}V(\mathbf{y}).$$

Agmon's proof[26] shows that the operator $I-K(k)$ is invertible for all real nonzero k. (Potentials that depend on k are not mentioned in Ref. 26, but the relevant arguments are exactly the same for k^2V as for V.) Moreover, since $K(0)=0$, $I-K(k)$ is invertible for all real k.

The derivation of the representation theorem for (A17) is exactly the same as for (A1); V is merely replaced by k^2V.

APPENDIX B: ASYMPTOTICS

In this appendix, we compute the large-x asymptotic forms of ψ, $(\partial/\partial|\mathbf{x}|)\psi$, G, and $(\partial/\partial|\mathbf{x}|)G$. We do this explicitly for the Schrödinger equation (which is the Fourier transform of the plasma wave equation). For an alternative treatment see Ref. 27. The arguments are exactly the same for the Fourier-transformed wave equation, except that V is replaced by k^2V.

We shall use the following large-$|\mathbf{x}|$ expansions, which are valid for $|\mathbf{x}|>|\mathbf{y}|$:

$$|\mathbf{x}-\mathbf{y}| = |\mathbf{x}| - \mathbf{x}\cdot\mathbf{y} + O(|\mathbf{x}|^{-1}), \tag{B1}$$

$$|\mathbf{x}-\mathbf{y}|^{-1} = |\mathbf{x}|^{-1}(1+\hat{x}\cdot\mathbf{y}|\mathbf{x}|^{-1} + O(|\mathbf{x}|^{-2})), \tag{B2}$$

$$\frac{e^{ik|\mathbf{x}-\mathbf{y}|}}{|\mathbf{x}-\mathbf{y}|} = \frac{e^{ik(|\mathbf{x}|-\hat{x}\cdot\mathbf{y})}}{|\mathbf{x}|} + O(|\mathbf{x}|^{-2}), \tag{B3}$$

$$\frac{\partial}{\partial|\mathbf{x}|}\frac{e^{ik|\mathbf{x}-\mathbf{y}|}}{|\mathbf{x}-\mathbf{y}|} = ik\frac{e^{ik(|\mathbf{x}|-\hat{x}\cdot\mathbf{y})}}{|\mathbf{x}|} + O(|\mathbf{x}|^{-2}). \tag{B4}$$

Equation (B1) can be obtained as follows. We write

$$|\mathbf{x}-\mathbf{y}| = |\mathbf{x}|(1-2s\hat{x}\cdot\mathbf{y} + s^2|\mathbf{y}|^2)^{1/2},$$

where $s=|\mathbf{x}|^{-1}$. The radical can be expanded in a Taylor series about $s=0$. The remainder, when written in Lagrange form, is easily seen to be $O(|\mathbf{x}|^{-2})$ provided $|\mathbf{x}|>|\mathbf{y}|$. Equation (B2) is obtained in a similar manner. Equations (B3) and (B4) can be deduced from (B1) and (B2).

Proposition: Let $V(\mathbf{x})$ be a bounded function of compact support. Then

$$\psi(k,\mathbf{e},\mathbf{x}) = \exp(ik\hat{e}\cdot\mathbf{x}) + A(k,\hat{x},\hat{e})|\mathbf{x}|^{-1}\exp(ik|\mathbf{x}|) + O(|\mathbf{x}|^{-2}). \tag{B5}$$

Proof: We use the Lippmann–Schwinger equation together with (B3).

Proposition: Let $V(\mathbf{x})$ be a bounded function of compact support. Then

$$\frac{\partial}{\partial|\mathbf{x}|}\psi(k,\hat{e},\mathbf{x}) = ik\hat{e}\cdot\hat{x}\exp(ik\hat{e}\cdot\mathbf{x}) + ikA(k,\hat{x},\hat{e})|\mathbf{x}|^{-1}\exp(ik|\mathbf{x}|) + O(|\mathbf{x}|^{-2}). \tag{B6}$$

Proof: We differentiate the Lippmann–Schwinger equation and use (B4).

Proposition: Let V be a bounded function of compact support. Then for $|\mathbf{y}|>|\mathbf{x}|$,

$$G^{-}(\mathbf{x},\mathbf{y}) = -(4\pi|\mathbf{y}|)^{-1}e^{-ik|\mathbf{y}|}\psi^{-}(k,\hat{y},\mathbf{x}) + O(|\mathbf{y}|^{-2}).$$

Proof: In the equation

$$G^{-}(\mathbf{x},\mathbf{y}) = G_0^{-}(\mathbf{x},\mathbf{y}) + \int G^{-}(\mathbf{x},\mathbf{z})V(\mathbf{z})G_0^{-}(\mathbf{z},\mathbf{y})d^3\mathbf{z}, \tag{B7}$$

we use expansion (B3) with the variables switched, obtaining

$$G^{-}(\mathbf{x},\mathbf{y}) = \frac{-e^{-ik(|\mathbf{y}|-\hat{y}\cdot\mathbf{x})}}{4\pi|\mathbf{y}|} + O(|\mathbf{y}|^{-2}) - \frac{1}{4\pi}\int G^{-}(\mathbf{x},\mathbf{z})V(\mathbf{z})\frac{e^{-ik(|\mathbf{y}|-\hat{y}\cdot\mathbf{z})}}{|\mathbf{y}|}d^3\mathbf{z} - \frac{1}{4\pi}\int G^{-}(\mathbf{x},\mathbf{z})V(\mathbf{z})\left[\frac{e^{-ik|\mathbf{y}-\mathbf{z}|}}{|\mathbf{y}-\mathbf{z}|} - \frac{e^{-ik(|\mathbf{y}|-\mathbf{y}\cdot\mathbf{z})}}{|\mathbf{y}|}\right]d^3\mathbf{z}. \tag{B8}$$

Next we recall the following representation of the wave function ψ^{-}:

$$\psi^{-}(k,\hat{y},\mathbf{x}) = e^{ik\hat{y}\cdot\mathbf{x}} + \int G^{-}(\mathbf{x},\mathbf{z})V(\mathbf{z})e^{ik\hat{y}\cdot\mathbf{z}}d^3\mathbf{z}.$$

This can be obtained from combining (B7) with the Lippmann–Schwinger equation. Equation (B8) is therefore

$$G^{-}(\mathbf{x},\mathbf{y}) = -\frac{e^{-ik|\mathbf{y}|}}{4\pi|\mathbf{y}|}\psi^{-}(k,\hat{y},\mathbf{x}) + O(|\mathbf{y}|^{-2}) - \frac{1}{4\pi}\int G^{-}(\mathbf{x},\mathbf{z})V(\mathbf{z})\left[\frac{e^{-ik|\mathbf{y}-\mathbf{z}|}}{|\mathbf{y}-\mathbf{z}|} - \frac{e^{-ik(|\mathbf{y}|-\hat{y}\cdot\mathbf{z})}}{|\mathbf{y}|}\right]d^3\mathbf{z}. \tag{B9}$$

We apply the Schwarz inequality to the last term of (B9), considering the integrand to be the product $G|V|^{1/2}|V|^{1/2}[\cdots]$. The expression in brackets is $O(|\mathbf{y}|^{-2})$ for $\mathbf{y}$ outside the support of V. This last term of (B9) is therefore $O(|\mathbf{y}|^{-2})$. Q.E.D.

Proposition: Let V be a bounded function of compact support. Then for $|\mathbf{y}|>|\mathbf{x}|$

$$\frac{\partial}{\partial|\mathbf{y}|}G^{-}(\mathbf{x},\mathbf{y}) = \frac{ike^{-ik|\mathbf{y}|}}{4\pi|\mathbf{y}|}\psi^{-}(k,\hat{y},\mathbf{x}) + O(|\mathbf{y}|^{-2}). \tag{B10}$$

Proof: We differentiate Eq. (B7) and use (B4); the reasoning is similar to the proof of the previous proposition.

[1]J. H. Rose, B. DeFacio, and M. Cheney, J. Opt. Soc. Am. A (in press) and Chalmers Technical University, Institute of Theoretical Physics Report #84-31.

[2]K. Budden, *Radio Waves in the Ionosphere* (Cambridge U. P., London, 1961).
[3]A. K. Jordan and S. Ahn, Proc. IEEE **126**, 945 (1979).
[4]G. N. Balanis, J. Math. Phys. **13**, 1001 (1972).
[5]C. S. Morawetz, Comput. Math. Appl. **7**, 319 (1981).
[6]B. DeFacio and J. H. Rose, Phys. Rev. A **31**, 897 (1985).
[7]F. G. Friedlander, *Sound Pulses* (Cambridge U. P., Cambridge, 1958).
[8]C. H. Wilcox, "Scattering theory for d'Alembert's equation in exterior domains" in *Lecture Notes in Mathematics*, Vol. 442 (Springer, New York, 1975).
[9]G. A. Hagedorn, Commun. Pure Appl. Math. **37**, 91 (1984).
[10]V. A. Marchenko, Dokl. Akad. Nauk. SSSR **72**, 457 (1950).
[11]V. A. Marchenko, Trudy Moskov. Mat. Obsc. **1**, 327 (1952).
[12]R. G. Newton, Phys. Rev. Lett. **43**, 541 (1979).
[13]R. G. Newton, J. Math. Phys. **21**, 1698 (1980).
[14]R. G. Newton, J. Math. Phys. **22**, 631 (1981).
[15]R. G. Newton, J. Math. Phys. **22**, 2191 (1981).
[16]R. G. Newton, J. Math. Phys. **23**, 594 (1982).
[17]R. G. Newton, in *Inverse Scattering: Theory and Applications*, edited by J. B. Bednar, R. Redner, E. Robinson, and A. Weglein (SIAM, Philadelphia, 1983). This reference presents an optimal introduction to the ideas in Refs. 12–16. Few of the proofs of the original papers are repeated but the notation is improved, minor errors are corrected, and the path is more straightforward.
[18]M. E. Taylor, *Pseudo differential Operators* (Princeton U. P., Princeton, NJ, 1981).
[19]M. Kline and I. Kay, *Electromagnetic Theory and Geometrical Optics* (Wiley, New York, 1965).
[20]S. Coen, M. Cheney, and A. Weglein, J. Math. Phys. **25**, 1857 (1984).
[21]R. G. Newton, Phys. Rev. A **31**, 3305 (1985).
[22]J. H. Rose, M. Cheney, and B. DeFacio, J. Math. Phys. **25**, 2995 (1984).
[23]D. Gilbarg and N. S. Trudinger, *Elliptic Partial Differential Equations of Second Order* (Springer, New York, 1977).
[24]R. G. Newton, J. Math. Phys. **18**, 1348 (1977).
[25]C. O. Bloom and N. D. Kazarinoff, *Short Wave Radiation Problems in Inhomogeneous Media: Asymptotic Solutions* (Springer, New York, 1976).
[26]S. Agmon, Ann. Scuola Norm. Sup. Pisa Cl. Sci. II **2**, 151 (1975); M. Reed and B. Simon, *Analysis of Operators* (Academic, New York, 1980); L. Hörmander, *Linear Partial Differential Operators* (Springer, New York, 1983), Vol. 2.
[27]B. S. Buslaev, in *Spectral Theory and Wave Processes*, Vol. 1 of the series *Topics in Mathematical Physics*, edited by M. Sh. Birman (Consultants Bureau, New York, 1967), p. 69.

Generating infinitesimal transformations with second-order-infinitesimal accuracy for proving covariance of commutation relations under finite transformations

Frederik J. Belinfante[a)]
Department of Physics, Purdue University, Lafayette, Indiana 47907

(Received 8 February 1984; accepted for publication 13 June 1985)

As dynamical quantization of Einstein's gravitational theory meets unsolved problems, it is worth considering the alternative method of quantization suggested by Fermi's quantization of special-relativistic electrodynamics, which for that theory has been the starting point of most modern applications of quantum electrodynamics. This method avoids first-class constraints by an alteration of the Lagrangian. In physical formulas, this introduces unwanted terms, that are at the end equated weakly to zero by auxiliary conditions. By absence of first-class constraints, this theory could be quantized canonically, if it would not contain fermion fields. As shown by Dirac, in the presence of fermion fields the canonical commutation relations of the altered theory have to be replaced by modified commutation relations. The ultimate purpose of this and following papers is to prove the covariance of Dirac's modified commutation relations, first under infinitesimal transformations, and thence under finite transformations. As usual, this requires the proof of existence of a conserved and invariant generator for the transformations admitted, here coordinate transformations and local Lorentz transformations of the tetrad field. For infinitesimal transformations, the conventional method of deriving from the Lagrangian density $\mathscr{L}$ a generator T_1 linear in the infinitesimal parameters that determine the transformations is used. However, for guaranteeing integrability of the procedure for generating transformations of the field variables, from infinitesimal transformations to finite transformations, it is necessary to show the existence of a more accurate generator T, no longer linear in the parameters, which will by $e^{iT}Fe^{-iT} - F$ generate the transformations $\bar{\delta}F$ of the field variables F with second-order accuracy. While it is left to a following paper to discuss the exact form of Dirac's modified commutation relations and to prove the conservation and covariance of the generator T and to prove that the conventional first-order generator T_1 will, this time *by Dirac's modified commutation relations*, generate the first-order substantial variations $\bar{\delta}_1 F$ of the field variables, in the present paper formulas are derived for expressing the second-order-infinitesimal terms in the transformation formulas by means of the first-order terms, and a method is derived for obtaining from T_1 the more accurate generator T. Finally, it is verified that the transformations $F(x) \mapsto F'(x') = F(x) + \bar{\delta}F$ generated by the generators T do satisfy the transformation group property with second-order accuracy.

I. INTRODUCTION

A. Covariance of quantization

Proofs of covariance of canonical quantization in quantum field theory date back to the papers of Heisenberg and Pauli.[1,2] Since their second paper,[2] most proofs introduce generators T, depending upon the parameters $(\xi^\mu, \epsilon_{1(\alpha\kappa\beta)})$ of the coordinate transformations $(x'^\mu = x^\mu + \xi^\mu(x))$ and local Lorentz transformations of tetrads that may be allowed. These T are shown to generate the transformations $F(P) \mapsto F'(P')$ of the canonical field variables F at points P, according to

$$F'(P') = e^{iT(\xi,\epsilon)} F(P) e^{-iT(\xi,\epsilon)}, \quad \text{for} \quad x'_{P'} = x_P . \tag{1}$$

Here, T is expressed entirely in terms of the canonical field variables on a hypersurface $x^0 = \text{const}$, on which the commutation relations are given in a form

$$[F(P);G(Q)] = iC(P)\delta_3(\mathbf{x}_P - \mathbf{x}_Q), \quad \text{for} \quad x_P^0 = x_Q^0 , \tag{2}$$

and P in (1) is to lie on that same hypersurface. However, T is then shown to be conserved for all values of x^0 and to be invariant under the transformations allowed. Then, it follows automatically from (2) by (1) that also, after the transformation,

$$[F'(P');G'(Q')] = iC'(P')\delta_3(\mathbf{x}'_P - \mathbf{x}'_Q), \quad \text{for} \quad x'^0_{P'} = x'^0_{Q'} . \tag{2'}$$

Originally, this method was applied just to Lorentz transformations in flat space-time, with the tetrad transformations coupled to the coordinate transformations by $h^{\mu}_{(\alpha)} = \delta^{\mu}_{\alpha}$ in all Lorentz frames. In the presence of fermion fields, the $[F;G]$ should be "graded" commutators,[3,4] i.e., anticommutators between fermion fields and fermion fields. Then, also distinction has to be made between differentiations from the left and from the right.[4] The method of Heisenberg and Pauli also was generalized to the case when there are *derived variables*,[4] as there are, for instance, in theories using a first-order Lagrangian for boson fields.

[a)] Present address: P. O. Box 901, Gresham, Oregon 97030.

 0022-2488/85/112814-13$02.50

B. Altered theory

As canonical quantization is self-contradictory in the presence of first-class constraints, and the latter are unavoidable in theories suffering from general gauge invariance or general coordinate invariance,[5,6] Fermi[7] "altered" quantum electrodynamics by adding terms quadratic in $S \equiv \nabla_\mu A^\mu$ to the Lagrangian. This breaks the general gauge invariance of Maxwell's theory, and restricts gauge invariance by the Lorentz condition $S = 0$. As, however, S in the altered theory is related to the canonical conjugate to A_0, the "auxiliary condition" $S = 0$ then had to be a "weak" equation, rather than a "strong" (q-number) relation. Various interpretations of such weak equations have been proposed in the literature, like (1) $\int \Psi^\dagger S \Psi = 0$, (2) $S\Psi = 0$,[8,9] (3) $S^{(+)}\Psi = 0$,[10] or (4) $\langle a|F|b\rangle = \int_{S=0} \Psi_a^\dagger F \Psi_b$.[11] Interpretation (3) requires splitting up fields F into their creation parts $F^{(-)}$ and annihilation parts $F^{(+)}$, which can be done Lorentz covariantly only in an interaction picture,[12] in which, however, S or $S^{(+)}$ is no longer zero.[13] Moreover, for being able to define a "bare vacuum" state at least Lorentz covariantly, one must assume an indefinite metric in the (generalization of a) Hilbert space of quantum states allowed by the commutation relations of the altered theory.[10] If space-time is curved, fields in an interaction picture are often regarded as fields in a flat space-time, upon which (part of) curved space-time is mapped.[14,15] These complications are avoided by interpretation (4).

In general-relativistic gravitational theory, the alteration of the theory includes addition of terms quadratic and bilinear in the $S^\mu \equiv (g^{\mu\nu}\sqrt{-g})_{,\nu}$, and the De Donder conditions[16] $S^\mu = 0$ are therefore among the auxiliary conditions.[17] In the presence of fermion fields, one will want to use the tetrad field rather than the metric field as the Lagrangian field varaibles, and the existence of tetrad constraints then requires further alteration of the Lagrangian, and additional auxiliary conditions.[17] Presently, we will not discuss all these details, or the interpretation of the auxiliary conditions for the altered theory.

C. Why no dynamical quantization?

It would, of course, be beautiful, if we could quantize the fields dynamically. Dynamical quantization means canonical quantization of only the dynamical variables of the theory (the variables initially specifiable as Cauchy data), with all other field variables treated as *nonlocally derived* variables. Such treatment requires assumption of appropriate strong (q-number) supplementary conditions, like div $\mathbf{A} = 0$ is assumed in Maxwell's theory as a strong equation in addition to the strong equations $p^{A_0} = 0$ and the time derivatives div $\mathbf{A}_{,0} = 0$ and div $\mathbf{E} = 4\pi\rho$, if we want to quantize flat-space quantum electrodynamics dynamically, by quantizing only the transverse fields, and solving for A_0 as the instantaneous Coulomb potential from ρ. The development of a theory of dynamical quantization of Einstein's gravitational theory was much boosted by the work of Dirac[18] and of Arnowitt, Deser, and Misner,[19] but bogged down by our lack of knowledge of general solutions of nonlinear differential equations for what should be the nonlocally derived variables. Therefore, we will here not discuss dynamical quantization, either.

Sticking to the quantization of the altered theory may have one advantage. In dynamical quantization, the strong supplementary conditions are sometimes interpreted as q-number "coordinate conditions," and this might endanger the c-number character of the coordinates used as parameters labeling points in space-time. As the auxiliary conditions of the altered theory are weak equations, this danger there does not exist. In the interpretations (2) and (3), auxiliary conditions merely tell us, in the oversized "Hilbert" space of the altered theory, what kinds of "unphysical" gravitons would be absent or would occur virtually only in what combinations. Thus, these conditions would then determine in this oversized "Hilbert" space a physical subspace, in whatever c-number coordinate system we have chosen to do our work. Interpretation (4) removes the objection of nonnormalizability of Ψ, which interpretation (2) may produce.[9]

D. Second-class constraints and canonization of fermion fields

Canonical quantization of field theory is hampered not only by first-class constraints. When fermion fields occur in a general-relativistic theory, there also occur second-class constraints of a type that cannot be treated by Belinfante's method of derived variables.[4] These second-class constraints contradict the commutation relations of direct canonical quantization. Dirac resolved this problem by replacing customary canonical quantization (which equates the commutator of two field variables to $i\hbar\times$ their Poisson bracket), by a similar procedure, in which the Poisson brackets are replaced by Dirac's modified Poisson brackets.[20] The physical meaning of this change in quantization procedure was explained by DeWitt and DeWitt.[21] For interpreting canonically quantized fermion fields, like in the old theory of "second quantization,"[22,23] in terms of creation and annihilation of particles satisfying Fermi statistics,[24] it is necessary to use fermion fields $\tilde\psi$ and $\tilde\psi^\dagger$ such that $i\hbar\tilde\psi^\dagger$ is the canonical conjugate to $\tilde\psi$. However, $i\hbar\psi^\dagger$ in the general-relativistic theory of fermions is not the canonical conjugate of ψ. Therefore, the DeWitts proposed a transformation $\tilde\psi = \Theta\psi$ from the ψ field occurring in the Lagrangian, with its simple transformation properties as an undor,[25] to the canonized[17] $\tilde\psi$ field, with its simple interpretation in terms of annihilation and antiparticle-creation operators.[21] The complicated explicit form of the transformation Θ was later given by Belinfante *et al.*[17] After canonically quantizing the canonized fields, one then may transform back to the Lagrangian fields, and one finds Dirac's modified commutation relations. Explicit formulas will be given in a following paper.[26]

E. Proof of covariance

Since the transformation properties of $\tilde\psi$ are awful, it is easier to establish the covariance of Dirac's quantization procedure in terms of the ψ field. The generalization of the method of Eq. (1) to the case of Dirac's modified quantization has not been published before. *This generalization is the main purpose of this sequence of papers.* For finding out how far this generalization may go, we start by admitting *general* coordinate transformations, and will place restrictions upon them only later, when we must. As general coordinate transformations do not form a Lie group, we will here avoid use of the theory of Lie groups.[27] Later we will want the Lagrangian to be invariant under our coordinate transformations. In an altered theory, this will place restrictions upon the allowable coordinate group, which then will become a Lie group. Yet, this need not necessarily preclude general covariance of our quantization procedure,[26] even when, contrary to Fock, we would admit nonaffine coordinate transformations.[28]

We need auxiliary conditions for the physical interpretation of our altered theory. In this sequence of papers, we will not further specify these, or discuss their various possible interpretations.

F. Infinitesimal generator

Later in the present paper, we will show how Eq. (1) may be derived from its first-order-infinitesimal form,

$$\bar{\delta} F \equiv F'(P') - F(P) \underset{2}{=} \bar{\delta}_1 F = [iT_1(\xi,\epsilon); F(P)]\,, \qquad (3)$$

where ξ and ϵ are infinitesimal, and $\underset{n}{=}$ means that here nth-order infinitesimal terms are neglected. Here, T_1 is linear in the parameters ξ^{μ}, $\epsilon_{1(\alpha)(\beta)}$, and $\xi^{\mu}_{,\nu}$ that describe the transformation from coordinate system and tetrad field Σ to Σ', and $x'_{P'} = x_P$. For T_1, we will use the conventional expression in terms of the Lagrangian density $\mathscr{L}$, even though in the commutator in (3) we will use the less usual commutation relations of Dirac. Below, we will discuss how T is obtained from T_1. For the proof of (3) we refer to the following paper.[26]

G. Integrability to finite transformations

Inclusion of fermion fields and use of Dirac's modified commutation relations are not the only way in which this sequence of papers goes beyond publications of the past. Many authors seem to believe that it is clear without proof that from (3) it will follow that there exists a T such that (1) will follow automatically. However, for integrability of (3) to (1) it is necessary that T for a finite transformation from Σ to Σ', as constructed by applying (3) infinitely many times for infinitesimal intermediate steps of transformation, will by the right-hand member of (3) yield the same result (independent of the path of transformation between Σ and Σ'), like the left-hand member will automatically on account of the fact that the transformations of F form a representation of the transformation group. However, if we subdivide the transformation path from Σ to Σ' into N ($\to \infty$) infinitesimal steps, and when we change one step at a time by an infinitesimal amount ($\propto 1/N$) until we reach a final path of transformation finitely different from an original one, the number of intermediate paths of transformation will be of the order of N^2, with each intermediate path differing from the preceding one by an infinitesimal change of path of a transformation to which (3) would be applicable. If, however, in each transformation the right-hand member of (3) is allowed to make a second-order-infinitesimal error, after N^2 infinitesimal changes of the path the result of the overall transformation from Σ to Σ' may have changed by a finite amount, and there is no guarantee that a finite change of path of a finite transformation would result in the same result given by the same T in the right-hand member of (1). Therefore, there is no trivial guarantee of integrability of (3).

For ensuring integrability of an infinitesimal verification of (1), we have to replace (3) by an infinitesimal application of (1) verified with second-order accuracy, that is, we should first verify

$$\bar{\delta} F \underset{3}{=} \bar{\delta}_{12} F \underset{3}{=} e^{iT_{12}(\xi,\epsilon)} F e^{-iT_{12}(\xi,\epsilon)} - F\,. \qquad (4)$$

Here, T_{12} and $\bar{\delta}_{12} F$ will contain terms of both first and second order in the parameters ξ and ϵ of the transformation, and, for infinitesimal transformations, $T_{12} \underset{3}{=} T$. In this paper, we will derive an expression for T_{12} which satisfies (4), expressing it in terms of T_1.

H. Products of transformations

As the transformations of the field variables should form a representation of the transformation group, the result of two successive transformations of F should be the same as the result of one single transformation of F under the coordinate and tetrad transformation that is the resultant of the original two consecutive transformations. When we label these transformations by (ξ,ϵ) or by (η, v), we mean by that a coordinate transformation ξ or η, *preceded* by a Lorentz transformation ϵ or v of the tetrad field. The result of successive infinitesimal transformations

$$\Sigma \xrightarrow{(\eta,\, v)} \Sigma'' \xrightarrow{(\xi,\epsilon)} \Sigma''''$$

(compare Fig. 1) will at a point P be

$$\begin{aligned} x''''^{\mu}_{P} &= x''^{\mu}_{P} + \xi^{\mu}(x''_P) \\ &\underset{3}{=} x^{\mu}_{P} + \eta^{\mu}(x_P) + \xi^{\mu}(x_P) + \xi^{\mu}_{,\lambda}(x_P)\eta^{\lambda}(x_P) \\ &\underset{3}{=} x^{\mu}_{P} + \rho^{\mu}(x_P)\,, \end{aligned}$$

with $\rho^{\mu} = \xi^{\mu} + \eta^{\mu} + \xi^{\mu}_{,\lambda}\,\eta^{\lambda}$, while the cumulative tetrad transformation ν will be given by $\nu = \epsilon + v + \epsilon_{,\lambda}\eta^{\lambda}$. The integrability of (3) will now be guaranteed by

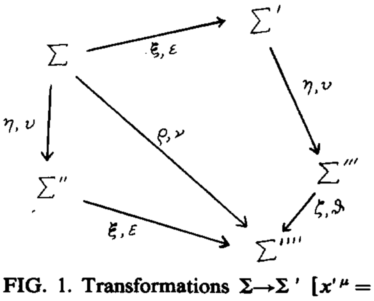


FIG. 1. Transformations $\Sigma \to \Sigma'$ $[x'^{\mu} = x^{\mu} + \xi^{\mu}(x)]$, $\Sigma'' \to \Sigma''''$ $[x''''^{\mu} = x''^{\mu} + \xi^{\mu}(x'')]$, and so on, each preceded by a tetrad rotation (ξ by ϵ, η by v, ρ by ν, ζ by ϑ). The difference between Σ''' and Σ'''' shows that transformations are not commutative. The transformation ζ preceded by ϑ is the commutator of transformations ξ with ϵ and η with v.

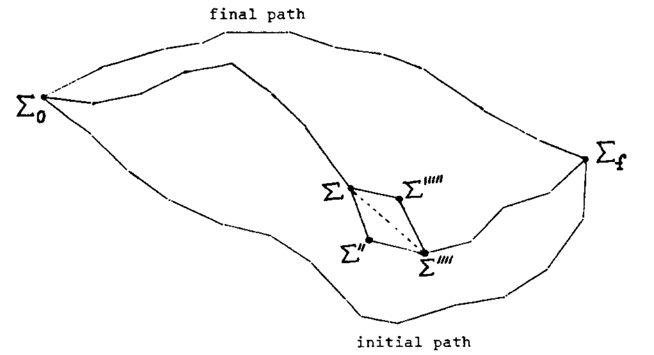


FIG. 2. Each change of one infinitesimal pair of steps in an intermediate path of transformation from Σ_0 to Σ_f, using two applications of (5) for calculating its effect upon a field variable F, will move the intermediate path of transformation in Σ space across an $\mathcal{O}_2$ area ($\propto N^{-2}$, infinitesimal of second order). A finite change of a finite path of transformation requires a move across a finite area in Σ space. This may be achieved by of the order of N^2 infinitesimal changes of path. If the transformation result calculated by an infinitesimal generator is to be independent of the path of transformation, there should be no second-order infinitesimal errors in the calculation of each infinitesimal step.

$$F''''(x'''') \underset{3}{=} e^{iT_{12}(\xi,\epsilon)} e^{iT_{12}(\eta,\nu)} F(x) e^{-iT_{12}(\eta,\nu)} e^{-iT_{12}(\xi,\epsilon)}$$

$$\underset{3}{=} e^{iT_{12}(\rho,\nu)} F(x) e^{-iT_{12}(\rho,\nu)}, \tag{5}$$

because, in the limit $N \to \infty$, the third-order-infinitesimal uncertainties in a finite transformation $\Sigma_0 \to \Sigma_f$, caused by changing each pair of infinitesimal steps $\Sigma \to \Sigma'' \to \Sigma''''$ in it, by means of (5) via $\Sigma \to \Sigma''''$ into a different infinitesimal path $\Sigma \to \Sigma''''' \to \Sigma''''$ (see Fig. 2), will accumulate after N^2 of such infinitesimal alterations of the path of transformation from Σ_0 to Σ_f, to a resultant change of the effect of this path of integrations upon fields F, by an amount of the order of $N^2 \times \mathcal{O}_3 = \mathcal{O}_1 \propto N^{-1}$ = first-order infinitesimal = vanishing. For the T_{12} derived below, the validity of (5) may be verified directly.

I. Summary of sections

We will now briefly summarize the contents of the following sections.

1. General tensor transformations

After an explanation of our terminology in Secs. II–V, we will in Secs. VI–X discuss the local transformation $\delta F(P) = F'(P) - F(P)$ for field variables $F = \varphi_i$ that are "tensors" in a generalized sense (which includes tensor densities). As in this part of the paper we do not use the metric, results here obtained would also be valid, if space-time were not Riemannian. In particular, we derive relations between infinitesimal transformations and their commutators, and we express the second-order terms in infinitesimal tensor transformations in terms of the first-order terms. These results may be derived from the fact that tensor transformations form a representation of the coordinate group, but these results may also be verified individually, and then by inversion of the reasoning they may be used for proving that tensor transformations form a representation (Sec. X).

2. Lorentz transformations of tetrad fields and undor transformations

As our work was prompted by Dirac's modification of the commutation relations, which was required by the presence of fermion fields, some of the fields will be undors[25] (four-component "Dirac spinors"), which, like ordinary (two-component) spinors,[29] are defined relative to local Lorentz frames given in curved space-time by local tetrads. Undors therefore transform under local rotations and Lorentz transformations of the tetrads. (These tetrad transformations could, but need not, be coupled to coordinate transformations.) Therefore, in Secs. XI–XIV we generalize the results of Secs. VI–X to include local tetrad transformations, and we give explicit formulas for infinitesimal undor transformations including second-order infinitesimals. As any tetrad field determines a metric field, here space-time is assumed to be Riemannian.

3. Obtaining the second-order generator T_{12} for the second-order substantial variations $\delta_{12} F$ of the field variables, from the first-order generator T_1

After discussion of the field momenta in Sec. XV, we introduce a simplified notation for dealing with transformations of fields F that may be either the original field variables[30] φ_i or their canonical conjugates $\not{p}^i$. We then derive in Sec. XVI second-order formulas also for the substantial variations $\bar{\delta} F$, and show how $\bar{\delta}_{12} F$ can be calculated starting from the formula for the first-order $\bar{\delta}_1 F$.

In Sec. XVII we mention (but do not yet use) the special form of the first-order-infinitesimal generator T_1, of which in the next paper[26] we will prove the conservation and invariance and the property Eq. (3). In the present paper we explain how from T_1 we obtain T_{12}, and why this satisfies (1), if we neglect third-order infinitesimals. In Sec. XVIII we verify Eq. (5) directly. In Sec. XIX we warn against using (1) or (3) outside their limited domain of validity, and in Sec. XX we briefly summarize our results.

II. THE COORDINATE PATCH

In the following we will consider an N-dimensional coordinate patch, in which a metric may be given or not given. We admit here "general" coordinate transformations inside this patch, $x'^{\mu} = f^{\mu}(x)$, where (x) stands for $(x^1, x^2, \ldots, x^N)$, and where the f^{μ} are continuous functions that have a sufficient number of partial derivatives, while the x^{ν} should be determined uniquely by the x'^{μ}.

Points in the coordinate patch may be labeled by their coordinates in any of the coordinate systems obtainable from each other by the coordinate transformations. So, two points P' in a coordinate system (frame) Σ' and P'' in a frame Σ'' are called identical, if in some arbitrary frame Σ or Σ''' we have $x_{P'} = x_{P''}$ or $x'''_{P'} = x'''_{P''}$. We then interpret coordinate transformations as equations valid at any fixed point P, that is, as $x'^{\mu}_P = f^{\mu}(x_P)$.

III. THE GROUP OF COORDINATE TRANSFORMATIONS

We define the product h of two coordinate transformations f and g by

$$h^{\mu}(x) = g^{\mu}(f(x)), \tag{6}$$

which results by inserting $x_P'^{\lambda} = f^{\lambda}(x_P)$ in $x_P''^{\mu} = g^{\mu}(x_P')$, and identifying $h^{\mu}(x_P)$ with $x_P''^{\mu}$ as a function of the x_P. The coordinate transformations then automatically form a group. We call f and g here the first and the second step in the product transformation h. As generally $g(f(x)) \neq f(g(x))$, the coordinate transformation group is not Abelian.

IV. SCALAR FIELDS

We call $S(x)$ a scalar field, if after a transformation f from Σ to Σ' we have $S'(P) = S(P)$ so that $S'(x_P') = S(x_P)$ when $x_P'^{\mu} = f^{\mu}(x_P)$. The local variation of a field F is defined by

$$\delta F(P) \equiv F'(P) - F(P); \tag{7a}$$

for a scalar field it vanishes ($\delta S = 0$), while the substantial variation

$$\bar{\delta} F(x) \equiv F'(x_{P'}') - F(x_P), \quad \text{with} \quad x_{P'}' = x_P = x \tag{7b}$$

in general does not vanish for a scalar field.

V. TENSOR FIELDS

Our definition here of a tensor field will be more general (will include more quantities) than the usual definition. What we here will call a tensor field, will be a set of $\mathscr{N}$ fields (say, $q_{\rm i}$ with ${\rm i} = 1,2,\ldots,\mathscr{N}$), which, under any coordinate transformation with descriptors $\xi^{\mu}(x)$, transform among each other homogeneous-linearly according to

$$q_{\rm i}'(P) = S_{\rm i}^{\ \rm j} q_{\rm j}(P), \tag{8}$$

where we assume automatic summation over indices that occur both up and down in one term, and where the coefficients $S_{\rm i}^{\ \rm j}$ are functions of the derivatives $\xi^{\mu}_{\ ,\nu}$ of the descriptors of the coordinate transformation. With the descriptors given by

$$\xi^{\mu}(x_P) \equiv \delta x_P^{\mu} \equiv x_P'^{\mu} - x_P^{\mu} = f^{\mu}(x_P) - x_P^{\mu}, \tag{9}$$

tensors are (at any fixed point P) invariant under a shift of origin given by constant descriptors. In $\xi^{\mu}_{\ ,\nu} \equiv \partial \xi^{\mu}/\partial x^{\nu}$, the derivatives are taken with respect to the coordinates before the transformation.

Not only scalars and tensors in the usual sense are tensors in the sense of this paper. We also include here tensor densities, obtained from an ordinary tensor by multiplying all of its components by some positive or negative power of the determinant of some (other) tensor of the second rank. For tensor densities, the $S_{\rm i}^{\ \rm j}$ may no longer be polynomials in the $\xi^{\mu}_{\ ,\nu}$, but, for small $\xi^{\mu}_{\ ,\nu}$, we still may expand the $S_{\rm i}^{\ \rm j}$ as a series in products of powers of the various $\xi^{\mu}_{\ ,\nu}$.

Furthermore, the $q_{\rm i}$ may include components of *several* tensors. For instance, for an electromagnetic field, the $q_{\rm i}$ might be the $4+6+1+6=17$ components of A_{μ}, $F_{\mu\nu}$, $\mathfrak{S} = \partial_{\nu}(g^{\mu\nu} A_{\mu}\sqrt{-g})$, and $\mathfrak{F}^{\mu\nu} \equiv g^{\mu\alpha} g^{\nu\beta} F_{\alpha\beta}\sqrt{-g}$.

Not all fields are tensor fields. Fermions are described by four-component undor fields[25] $\psi_{\rm A}$ and $\bar{\psi}^{\rm B}$, which are built up out of various kinds of two-component spinor fields.[29] Their physical meaning is defined relative to local Lorentz frames described by tetrad fields $h^{\ \mu}_{(\alpha)}$. The latter then describe the gravitational field, as the metric can be expressed in terms of the tetrad field. Instead of the tetrad field, one may well prefer to use its inverse, the field $h^{\ (\alpha)}_{\mu}$. The tetrads and their inverse do not merely (through the index μ) transform as tensors, but also [through the index (α)] undergo Lorentz transformations, as do the undors. When these Lorentz transformations are infinitesimal, they may be described by antisymmetric descriptors $\epsilon_{(\alpha)(\beta)} = -\epsilon_{(\beta)(\alpha)}$. They depend, however, directly upon these descriptors and not upon the derivatives of these descriptors. Through Sec. X, we will not discuss these Lorentz transformations of tetrads and of undors, and therefore do not deal with fermion fields.

VI. EXPANSION OF THE TENSOR TRANSFORMATIONS

The $S_{\rm i}^{\ \rm j}$ in (8), through the $\xi^{\mu}_{\ ,\nu}$, depend upon the old as well as the new coordinates. Unless otherwise specified, we treat them as functions of the old coordinates.

When we expand the $S_{\rm i}^{\ \rm j}$ in terms of the $\xi^{\mu}_{\ ,\nu}$, the coefficients are constants. For instance, a mixed tensor density $\mathfrak{t}_{\mu}^{\ \nu} = t_{\mu}^{\ \nu}\sqrt{-g}$ transforms according to

$$\mathfrak{t}_{\rho'}^{\ \kappa'} = \frac{\partial x'^{\kappa}}{\partial x^{\lambda}} \frac{\partial x^{\sigma}}{\partial x'^{\rho}} \det\left\{\frac{\partial x^{\alpha}}{\partial x'^{\beta}}\right\} \mathfrak{t}_{\sigma}^{\ \lambda}, \tag{10}$$

with

$$\frac{\partial x'^{\kappa}}{\partial x^{\lambda}} = \delta^{\kappa}_{\lambda} + \xi^{\kappa}_{\ ,\lambda},$$
$$\frac{\partial x^{\sigma}}{\partial x'^{\rho}} = \delta^{\sigma}_{\rho} - \xi^{\sigma}_{\ ,\rho} + \xi^{\sigma}_{\ ,\epsilon}\xi^{\epsilon}_{\ ,\rho} + \cdots, \tag{11a}$$

so that

$$\det\left\{\frac{\partial x^{\alpha}}{\partial x'^{\beta}}\right\} = 1 - \xi^{\alpha}_{\ ,\alpha} + \tfrac{1}{2}(\xi^{\alpha}_{\ ,\alpha}\xi^{\beta}_{\ ,\beta} + \xi^{\alpha}_{\ ,\beta}\xi^{\beta}_{\ ,\alpha}) + \cdots. \tag{11b}$$

Therefore, when in general we write

$$S_{\rm i}^{\ \rm j} \equiv \delta_{\rm i}^{\rm j} + \delta S_{\rm i}^{\ \rm j}, \quad \delta S_{\rm i}^{\ \rm j} = \delta_1 S_{\rm i}^{\ \rm j} + \delta_2 S_{\rm i}^{\ \rm j} + \cdots, \tag{12a}$$

$$\delta_1 S_{\rm i}^{\ \rm j} \equiv s_{\rm i}^{\ \rm j}{}_{\mu}^{\ \nu}\xi^{\mu}_{\ ,\nu},$$
$$\delta_2 S_{\rm i}^{\ \rm j} \equiv s_{\rm i}^{\ \rm j}{}_{\mu}^{\ \nu}{}_{\alpha}^{\ \beta}\xi^{\mu}_{\ ,\nu}\xi^{\alpha}_{\ ,\beta}, \tag{12b}$$
$$\text{with } s_{\rm i}^{\ \rm j}{}_{\mu}^{\ \nu}{}_{\alpha}^{\ \beta} = s_{\rm i}^{\ \rm j}{}_{\alpha}^{\ \beta}{}_{\mu}^{\ \nu},$$

we would for $\mathfrak{t}_{\rho}^{\ \kappa}$ find, from (11a) and (11b),

$$s_{\rm i}^{\ \rm j}{}_{\mu}^{\ \nu} \equiv s^{\kappa\sigma}_{\rho\lambda\mu}{}^{\nu} = \delta^{\kappa}_{\mu}\delta^{\nu}_{\lambda}\delta^{\sigma}_{\rho} - \delta^{\kappa}_{\lambda}\delta^{\sigma}_{\mu}\delta^{\nu}_{\rho} - \delta^{\kappa}_{\lambda}\delta^{\sigma}_{\rho}\delta^{\nu}_{\mu}, \tag{10'}$$

$$s_{\rm i}^{\ \rm j}{}_{\mu}^{\ \nu}{}_{\alpha}^{\ \beta} \equiv s^{\kappa\sigma}_{\rho\lambda\mu}{}^{\nu}{}_{\alpha}{}^{\beta}$$
$$= \mathrm{Sym}\{-\delta^{\kappa}_{\mu}\delta^{\nu}_{\lambda}\delta^{\sigma}_{\alpha}\delta^{\beta}_{\rho} - \delta^{\kappa}_{\mu}\delta^{\nu}_{\lambda}\delta^{\sigma}_{\rho}\delta^{\beta}_{\alpha} + \delta^{\kappa}_{\lambda}\delta^{\sigma}_{\mu}\delta^{\nu}_{\rho}\delta^{\beta}_{\alpha}$$
$$+ \delta^{\kappa}_{\lambda}\delta^{\sigma}_{\mu}\delta^{\nu}_{\alpha}\delta^{\beta}_{\rho} + \tfrac{1}{2}\delta^{\kappa}_{\lambda}\delta^{\sigma}_{\rho}\delta^{\nu}_{\alpha}\delta^{\beta}_{\mu} + \tfrac{1}{2}\delta^{\kappa}_{\lambda}\delta^{\sigma}_{\rho}\delta^{\nu}_{\mu}\delta^{\beta}_{\alpha}\}, \tag{10''}$$

$$\mathrm{Sym}\{A_{\mu}^{\ \nu}{}_{\alpha}^{\ \beta}\} = \tfrac{1}{2}\{A_{\mu}^{\ \nu}{}_{\alpha}^{\ \beta} + A_{\alpha}^{\ \beta}{}_{\mu}^{\ \nu}\}. \tag{13}$$

The coefficients $s_{\rm i}^{\ \rm j}{}_{\ldots}^{\ \cdots}$ in the expansions (12b), therefore, are independent of the coordinates and of the coordinate system used. Below, we will derive a few important relations between these coefficients $s_{\rm i}^{\ \rm j}{}_{\mu}^{\ \nu}$ and $s_{\rm i}^{\ \rm j}{}_{\mu}^{\ \nu}{}_{\alpha}^{\ \beta}$ for arbitrary "tensors."

VII. TENSOR REPRESENTATIONS OF THE COORDINATE TRANSFORMATION GROUP

Let a tensor field with components $q_{\rm i}$ be given originally in some fixed coordinate system Σ, of which we may also use the coordinates x as the labels of points P while we discuss transformations between various coordinate systems Σ, Σ', Σ'', etc. Let $q'_{\rm i}$, $q''_{\rm i}$, ... be the components of the transformed tensor in the coordinate systems Σ', Σ'', ... We may treat them all as functions of the points P in space, writing $q'_{\rm i}(P)$, $q''_{\rm i}(P)$, etc. (This could be regarded as considering them functions of the coordinates x_P in Σ, even when we take the components of $q'_{\rm i}$, $q''_{\rm i}$, ... along the coordinate axes of Σ', Σ'', ...) As an alternative, we may consider them functions $q'_{\rm i}(x')$, $q''_{\rm i}(x'')$, ... of the coordinates in the frames in which we take the components of the tensor.

In the former case, we call $q''_{\rm i}(P) - q'_{\rm i}(P)$ the local variation at P, of the tensor components, under the (infinitesimal) transformation from Σ' to Σ''. In the latter case, we call $q''_{\rm i}(P'') - q'_{\rm i}(P')$ with $x''_{P''} = x'_{P'}$ the substantial variation of the tensor at given values of the coordinates, under this same transformation. For local variations we use the symbol δ, and for substantial variations the symbol $\bar{\delta}$. So, if $\Delta x \equiv x_{P''} - x_{P'}$, and $\Delta q''_{\rm i} \equiv q''_{\rm i}(P'') - q''_{\rm i}(P')$, then $\bar{\delta} q_{\rm i} = \delta q_{\rm i}(P') + \Delta q''_{\rm i}$ for this transformation. Or, also, $\bar{\delta} q_{\rm i} = \delta q_{\rm i}(P'') + \Delta q'_{\rm i}$.

Irrespective of whether the result of the transformation is given as $q''_{\rm i}(P)$ at given x_P, or as $q''_{\rm i}(P'')$ at given $x''_{P''}$, the result for $q''_{\rm i}$ will be uniquely determined by the choice of the coordinate system (here Σ''), of the tensor component (here labeled by i), and of the location (value of coordinates x or x''), and will not depend on whether we obtain it by transformation directly from the $q_{\rm i}(P)$ field originally given in Σ, or indirectly by several steps of transformation (like $\Sigma \to \Sigma' \to \Sigma''$). We call this the *representation property* of tensor transformations. We will use it below for deriving relations between the constant coefficients $s_{\rm i}{}^{\rm j}{}_\mu{}^\nu$, etc.

The use of the word "representation" here is justified by the fact that, for each given type of tensor field with given constant coefficients $s_{\rm i}{}^{\rm j}{}_\mu{}^\nu$, etc., for any given coordinate transformation $x'^\mu = f^\mu(x)$, the transformation matrix $S_{\rm i}{}^{\rm j}$ is given uniquely by (12a) with (12b) with $\xi^\mu_{,\nu} = \partial f^\mu / \partial x^\nu - \delta^\mu_\nu$, while the matrix for the tensor transformation that belongs to the product (6) of two coordinate transformations f and g is the matrix product of the tensor transformation matrices for the individual coordinate transformations f and g. So, for the product (6) of coordinate transformations, the matrices $S^{(f)}{}_{\rm i}{}^{\rm j}$, $S^{(g)}{}_{\rm i}{}^{\rm j}$, and $S^{(h)}{}_{\rm i}{}^{\rm j}$ in $q'_{\rm i}(P) = S^{(f)}{}_{\rm i}{}^{\rm j} q_{\rm j}(P)$ and in $q''_{\rm i}(P) = S^{(g)}{}_{\rm i}{}^{\rm j} q'_{\rm j}(P) = S^{(h)}{}_{\rm i}{}^{\rm j} q_{\rm j}(P)$ will satisfy the group property

$$S^{(h)}{}_{\rm i}{}^{\rm k} = S^{(g)}{}_{\rm i}{}^{\rm j} S^{(f)}{}_{\rm j}{}^{\rm k} . \tag{14}$$

For any tensor field originally given in Σ, this guarantees that in any different coordinate system Σ'' the tensor components $q''_{\rm i}$ will have the same value irrespective of whether they are from Σ obtained directly, or via Σ', or via any other detour. This is what we called above the representation property of tensor transformations.

VIII. THE COMMUTATOR OF TRANSFORMATIONS

From an initial coordinate system Σ we may transform to Σ' using a descriptor field $\xi^\mu(x)$, or to Σ'' using a descriptor field $\eta^\mu(x)$. From Σ', let us transform to Σ''' using the descriptor field $\eta^\mu(x')$, and transform from Σ'' to Σ'''' using $\xi^\mu(x'')$. (See Fig. 1.) Then, in general, Σ''' and Σ'''' will differ, because, for infinitesimal transformations, at any fixed point P,

$$x'''^\mu - x^\mu = \xi^\mu(x) + \eta^\mu(x + \delta x) = \xi^\mu + \eta^\mu + \eta^\mu_{,\nu} \xi^\nu , \tag{15a}$$

$$x''''^\mu - x^\mu = \eta^\mu + \xi^\mu + \xi^\mu_{,\nu} \eta^\nu \quad (= \rho^\mu; \text{ see Fig. 1}). \tag{15b}$$

We may now transform from Σ''' to Σ'''', by using a descriptor field

$$\zeta^\mu(x''') = x''''^\mu - x'''^\mu = \xi^\mu_{,\nu} \eta^\nu - \eta^\mu_{,\nu} \xi^\nu . \tag{16a}$$

As this is $\mathscr{O}_2$ already, neglecting third-order infinitesimals we may write

$$\zeta^\mu(x) = \zeta^\mu(x''') . \tag{16b}$$

We may call this transformation the commutator of the transformations with descriptors ξ^μ and η^μ.

Because of the representation property of tensor transformations, $q''''_{\rm i}(P)$ in Σ'''' will be the same, irrespective of whether we obtain it from $q_{\rm i}(P)$ in Σ by two steps, $\Sigma \to \Sigma'' \to \Sigma''''$, or by three steps, $\Sigma \to \Sigma' \to \Sigma''' \to \Sigma''''$. Therefore, up to second order, we find for $q''''_{\rm i}(P) - q_{\rm i}(P)$,

$$\begin{aligned} &\eta^\mu_{,\nu}(x) s_{\rm i}{}^{\rm k}{}_\mu{}^\nu q_{\rm k} + \eta^\mu_{,\nu} \eta^\alpha_{,\beta} s_{\rm i}{}^{\rm k}{}_\mu{}^\nu{}_\alpha{}^\beta q_{\rm k} \\ &\quad + \xi^\mu_{,\nu}(x + \eta) s_{\rm i}{}^{\rm j}{}_\mu{}^\nu \{ q_{\rm j} + \eta^\alpha_{,\beta} s_{\rm j}{}^{\rm k}{}_\alpha{}^\beta q_{\rm k} \} \\ &\quad + \xi^\mu_{,\nu} \xi^\alpha_{,\beta} s_{\rm i}{}^{\rm k}{}_\mu{}^\nu{}_\alpha{}^\beta q_{\rm k} \\ &\quad = \{\text{same with } \xi \text{ and } \eta \text{ interchanged}\} + \zeta^\mu_{,\nu} s_{\rm i}{}^{\rm k}{}_\mu{}^\nu q_{\rm k} . \end{aligned} \tag{17}$$

By $\xi^\mu_{,\nu}(x + \eta) = \xi^\mu_{,\nu} + \xi^\mu_{,\nu\lambda} \eta^\lambda$, this gives

$$\begin{aligned} &\xi^\mu_{,\nu} \eta^\alpha_{,\beta} (s_{\rm i}{}^{\rm j}{}_\mu{}^\nu s_{\rm j}{}^{\rm k}{}_\alpha{}^\beta - s_{\rm i}{}^{\rm j}{}_\alpha{}^\beta s_{\rm j}{}^{k}{}_\mu{}^\nu) q_{\rm k} \\ &\quad + (\xi^\mu_{,\nu\lambda} \eta^\lambda - \eta^\mu_{,\nu\lambda} \xi^\lambda - \zeta^\mu_{,\nu}) s_{\rm i}{}^{\rm k}{}_\mu{}^\nu q_{\rm k} = 0 , \end{aligned} \tag{18a}$$

or, by (16a),

$$\begin{aligned} &\xi^\mu_{,\nu} \eta^\alpha_{,\beta} \{ s_{\rm i}{}^{\rm j}{}_\mu{}^\nu s_{\rm j}{}^{\rm k}{}_\alpha{}^\beta - s_{\rm i}{}^{\rm j}{}_\alpha{}^\beta s_{\rm j}{}^{k}{}_\mu{}^\nu \\ &\quad + \delta^\beta_\mu s_{\rm i}{}^{\rm k}{}_\alpha{}^\nu - \delta^\nu_\alpha s_{\rm i}{}^{\rm k}{}_\mu{}^\beta \} q_{\rm k} = 0 . \end{aligned} \tag{18b}$$

As this is to be valid for all $q_{\rm k}$ and for any fields ξ^μ and η^μ, it follows that

$$s_{\rm i}{}^{\rm j}{}_\mu{}^\nu s_{\rm j}{}^{\rm k}{}_\alpha{}^\beta - s_{\rm i}{}^{\rm j}{}_\alpha{}^\beta s_{\rm j}{}^{\rm k}{}_\mu{}^\nu = \delta^\nu_\alpha s_{\rm i}{}^{\rm k}{}_\mu{}^\beta - \delta^\beta_\mu s_{\rm i}{}^{\rm k}{}_\alpha{}^\nu . \tag{19}$$

IX. SECOND-ORDER COEFFICIENTS IN TERMS OF FIRST-ORDER COEFFICIENTS

Consider again the transformation $\Sigma \to \Sigma'' \to \Sigma''''$, the first step with descriptor η^μ, the second step with descriptor $\xi^\mu + \xi^\mu_{,\lambda} \eta^\lambda$, so that the total transformation has descriptor $x''''^\mu - x^\mu = \xi^\mu + \eta^\mu + \xi^\mu{}_{,\lambda} \eta^\lambda$. Therefore, for the transformation of $q_{\rm i}$ in one step from Σ to Σ'''' we find, to second order,

$$\begin{aligned} \delta q_{\rm i} = \{ &(\xi^\mu_{,\nu} + \eta^\mu_{,\nu} + \xi^\mu_{,\lambda\nu} \eta^\lambda + \xi^\mu_{,\lambda} \eta^\lambda_{,\nu}) s_{\rm i}{}^{\rm k}{}_\mu{}^\nu \\ &+ (\xi^\mu_{,\nu} + \eta^\mu_{,\nu})(\xi^\alpha_{,\beta} + \eta^\alpha_{,\beta}) s_{\rm i}{}^{\rm k}{}_\mu{}^\nu{}_\alpha{}^\beta \} q_{\rm k} . \end{aligned} \tag{20}$$

For this transformation in two steps, the left-hand member of (18) gave

$$\delta q_{\mathrm{i}} = \{(\xi^{\mu}_{,\nu} + \eta^{\mu}_{,\nu} + \xi^{\mu}_{,\nu\lambda}\,\eta^{\lambda})s_{\mathrm{i}}{}^{\mathrm{k}}{}_{\mu}{}^{\nu} + \xi^{\mu}_{,\nu}\eta^{\alpha}_{,\beta}\,s_{\mathrm{i}}{}^{\mathrm{j}}{}_{\mu}{}^{\nu}s_{\mathrm{j}}{}^{\mathrm{k}}{}_{\alpha}{}^{\beta}$$
$$+ (\xi^{\mu}_{,\nu}\,\xi^{\alpha}_{,\beta} + \eta^{\mu}_{,\nu}\eta^{\alpha}_{,\beta})s_{\mathrm{i}}{}^{\mathrm{k}}{}_{\mu}{}^{\nu}{}_{\alpha}{}^{\beta}\}q_{\mathrm{k}}\,. \qquad (21)$$

By the representation property of tensor transformations, the two expressions (20) and (21) must be equal. Therefore,

$$\xi^{\mu}_{,\nu}\eta^{\alpha}_{,\beta}\delta^{\nu}_{\alpha}s_{\mathrm{i}}{}^{\mathrm{k}}{}_{\mu}{}^{\beta}q_{\mathrm{k}} + \xi^{\mu}_{,\nu}\eta^{\alpha}_{,\beta}(s_{\mathrm{i}}{}^{\mathrm{k}}{}_{\mu}{}^{\nu}{}_{\alpha}{}^{\beta} + s_{\mathrm{i}}{}^{k}{}_{\alpha}{}^{\beta}{}_{\mu}{}^{\nu})q_{\mathrm{k}}\,.$$
$$= \xi^{\mu}_{,\nu}\eta^{\alpha}_{,\beta}s_{\mathrm{i}}{}^{\mathrm{j}}{}_{\mu}{}^{\nu}s_{\mathrm{j}}{}^{\mathrm{k}}{}_{\alpha}{}^{\beta}q_{\mathrm{k}}\,. \qquad (22)$$

Thence,

$$\delta^{\nu}_{\alpha}s_{\mathrm{i}}{}^{\mathrm{k}}{}_{\mu}{}^{\beta} + s_{\mathrm{i}}{}^{\mathrm{k}}{}_{\mu}{}^{\nu}{}_{\alpha}{}^{\beta} + s_{\mathrm{i}}{}^{\mathrm{k}}{}_{\alpha}{}^{\beta}{}_{\mu}{}^{\nu} = s_{\mathrm{i}}{}^{\mathrm{j}}{}_{\mu}{}^{\nu}s_{\mathrm{j}}{}^{\mathrm{k}}{}_{\alpha}{}^{\beta}\,. \qquad (23)$$

The part of this *antisymmetric* under an interchange of ${}_{\mu}{}^{\nu}$ with ${}_{\alpha}{}^{\beta}$ confirms our previous result, Eq. (19). The part *symmetric* under ${}_{\mu}{}^{\nu} \rightleftarrows {}_{\alpha}{}^{\beta}$ yields the new relation

$$s_{\mathrm{i}}{}^{\mathrm{k}}{}_{\mu}{}^{\nu}{}_{\alpha}{}^{\beta} + s_{\mathrm{i}}{}^{\mathrm{k}}{}_{\alpha}{}^{\beta}{}_{\mu}{}^{\nu} = \tfrac{1}{2}\{s_{\mathrm{i}}{}^{\mathrm{j}}{}_{\mu}{}^{\nu}s_{\mathrm{j}}{}^{\mathrm{k}}{}_{\alpha}{}^{\beta} + s_{\mathrm{i}}{}^{\mathrm{j}}{}_{\alpha}{}^{\beta}s_{\mathrm{j}}{}^{\mathrm{k}}{}_{\mu}{}^{\nu}$$
$$- \delta^{\nu}_{\alpha}s_{\mathrm{i}}{}^{\mathrm{k}}{}_{\mu}{}^{\beta} - \delta^{\beta}_{\mu}s_{\mathrm{i}}{}^{\mathrm{k}}{}_{\alpha}{}^{\nu}\}\,. \qquad (24)$$

This equation shows us that the second-order terms $\delta_2 q_{\mathrm{i}}$ in δq_{i} are not independent of the first-order terms $\delta_1 q_{\mathrm{i}}$, and that their coefficients $s_{\mathrm{i}}{}^{\mathrm{j}}{}_{\mu}{}^{\nu}{}_{\alpha}{}^{\beta} = s_{\mathrm{i}}{}^{\mathrm{j}}{}_{\alpha}{}^{\beta}{}_{\mu}{}^{\nu}$ can be expressed by (24) in terms of the first-order coefficients $s_{\mathrm{i}}{}^{\mathrm{j}}{}_{\mu}{}^{\nu}$. Thus, (8) with (12a) with (12b) becomes

$$\delta q_{\mathrm{i}} = \delta_1 q_{\mathrm{i}} + \delta_2 q_{\mathrm{i}} + \cdots, \quad \delta_1 q_{\mathrm{i}} = \xi^{\mu}_{,\nu}s_{\mathrm{i}}{}^{\mathrm{j}}{}_{\mu}{}^{\nu}q_{\mathrm{j}}\,,$$
$$\delta_2 q_{\mathrm{i}} = \tfrac{1}{2}\xi^{\mu}_{,\nu}\xi^{\rho}_{,\sigma}s_{\mathrm{i}}{}^{\mathrm{j}}{}_{\mu}{}^{\nu}s_{\mathrm{j}}{}^{\mathrm{k}}{}_{\rho}{}^{\sigma}q_{\mathrm{k}} - \tfrac{1}{2}\xi^{\mu}_{,\lambda}\xi^{\lambda}_{,\nu}s_{\mathrm{i}}{}^{\mathrm{j}}{}_{\mu}{}^{\nu}q_{\mathrm{j}}\,. \qquad (25)$$

X. EXAMPLES OF THE RELATIONS (19) AND (24)

We will verify here the validity of Eqs. (19) and (24) for the case of a mixed tensor density field $q_{\mathrm{i}} \equiv \mathfrak{t}^{\kappa}_{\rho} \equiv t^{\kappa}_{\rho}\sqrt{-g}$ considered already in Eqs. (10), (10′), and (10″). We will write $\mathfrak{t}^{\lambda}_{\sigma}$ for q_{j}, and $\mathfrak{t}^{\pi}_{\tau}$ for q_{k}. Calculating each side of Eq. (19) individually, we find that *both* sides are in this case equal to

$$\delta^{\nu}_{\alpha}(\delta^{\kappa}_{\mu}\delta^{\beta}_{\pi}\delta^{\tau}_{\rho} - \delta^{\tau}_{\mu}\delta^{\beta}_{\rho}\delta^{\kappa}_{\pi}) - \delta^{\beta}_{\mu}(\delta^{\kappa}_{\alpha}\delta^{\nu}_{\pi}\delta^{\tau}_{\rho} - \delta^{\tau}_{\alpha}\delta^{\nu}_{\rho}\delta^{\kappa}_{\pi})\,, \qquad (19')$$

so that (19) is satisfied.

Similarly, Eq. (24) is for this case verified, as both sides now are found to be equal to

$$\delta^{\beta}_{\rho}\delta^{\kappa}_{\pi}(\delta^{\nu}_{\alpha}\delta^{\tau}_{\mu} + \delta^{\nu}_{\mu}\delta^{\tau}_{\alpha}) + \delta^{\nu}_{\rho}\delta^{\kappa}_{\pi}(\delta^{\beta}_{\mu}\delta^{\tau}_{\alpha} + \delta^{\beta}_{\alpha}\delta^{\tau}_{\mu})$$
$$+ \delta^{\kappa}_{\pi}\delta^{\tau}_{\rho}(\delta^{\beta}_{\alpha}\delta^{\nu}_{\mu} + \delta^{\beta}_{\mu}\delta^{\nu}_{\alpha}) - \delta^{\kappa}_{\mu}\delta^{\nu}_{\pi}(\delta^{\beta}_{\rho}\delta^{\tau}_{\alpha} + \delta^{\beta}_{\alpha}\delta^{\tau}_{\rho})$$
$$- \delta^{\kappa}_{\alpha}\delta^{\beta}_{\pi}(\delta^{\nu}_{\rho}\delta^{\tau}_{\mu} + \delta^{\nu}_{\mu}\delta^{\tau}_{\rho})\,. \qquad (24')$$

In this way, the relations (19) and (24) could be shown to hold for any other chosen type of tensor field. Therefore, the reasoning of Secs. VIII and IX could be reversed, deriving the representation property of tensor transformations from the validity of the relations (19) and (24), instead of deriving these relations from the representation property.

XI. SECOND-ORDER LOCAL LORENTZ TRANSFORMATIONS OF TETRADS AND OF LOCAL TENSORS

In Eq. (20), the distinction between first-order and second-order-infinitesimal terms was well determined, with vanishing second-order uncertainty in the first-order term $\delta_1 q_{\mathrm{i}}$, because, even for finite transformations, there is no uncertainty in ξ^{μ}, as we defined it rigorously as $x'^{\mu} - x^{\mu}$, so that $\delta_1 x^{\mu} = \xi^{\mu} = \delta x^{\mu}$ with $\delta_2 x^{\mu} = 0$.

For achieving a similar result for the transformations of local tensors[31] by infinitesimal local rotations or Lorentz transformations of the tetrad field with respect to which the local tensor components are taken, we must have $\mathscr{O}_2$ precision also in the definition of the descriptors for the latter infinitesimal transformations.

We therefore start here by defining descriptors $\epsilon^{(\alpha)}_{(\beta)}$ which for local tensor transformations shall have exactly the same meaning as the $\xi^{\alpha}_{,\beta}$ would have for the corresponding world tensor transformations.[31] That is, in absence of coordinate transformations, the inverse tetrads are under their local rotations and Lorentz transformations to transform *rigorously* according to $h'^{(\alpha)}_{\mu} = h^{(\alpha)}_{\mu} + \epsilon^{(\alpha)}{}_{(\beta)}h^{(\beta)}_{\mu}$, or, omitting the (here irrelevant) index μ, we define $\epsilon^{(\alpha)}{}_{(\beta)}$ by postulating that a contravariant local[31] four-vector $h^{(\alpha)}$ shall rigorously transform according to

$$h'^{(\alpha)} = h^{(\alpha)} + \epsilon^{(\alpha)}{}_{(\beta)}h^{(\beta)}\,, \qquad (26a)$$

even for finite transformations. [Compare Eq. (26a) with the transformation $h'^{\alpha} = (\delta^{\alpha}_{\beta} + \xi^{\alpha}_{,\beta})h^{\beta}$ of a world[31] four-vector h^{α} (like dx^{α}), and see how $\xi^{\alpha}_{,\beta}$ here is replaced in (26a) by $\epsilon^{(\alpha)}{}_{(\beta)}$.] The (possibly finite) $\epsilon_{(\alpha)(\beta)} \equiv \mathring{g}_{(\alpha)(\beta)}\epsilon^{(\gamma)}{}_{(\beta)}$ is now no longer antisymmetric. [See Eq. (28a) below.]

In general, for local tensors, in first-order approximation, we may write infinitesimal transformations in the form

$$\delta_1 \mathrm{q}_{\mathrm{i}} = \epsilon^{(\alpha)}_1{}_{(\beta)}\sigma_{\mathrm{i}}{}^{\mathrm{j}}{}_{(\alpha)}{}^{(\beta)}\mathrm{q}_{\mathrm{j}}\,, \qquad (26b)$$

where $\epsilon^{(\alpha)}_1{}_{(\beta)}$ is the $\mathscr{O}_1$ part of $\epsilon^{(\alpha)}{}_{(\beta)}$, and will rigorously be defined below. Equation (26b) may be compared with $\delta_1 q_{\mathrm{i}} = \xi^{\mu}_{,\nu}s_{\mathrm{i}}{}^{\mathrm{j}}{}_{\mu}{}^{\nu}q_{\mathrm{j}}$ for world tensors, except that by definition we had $\xi^{\mu}_{1,\nu} = \xi^{\mu}_{,\nu}$ and $\xi^{\mu}_{2,\nu} = 0$, while below we will have $\epsilon^{(\alpha)}_2{}_{(\beta)} \neq 0$. For contravariant local four-vectors $\mathrm{q}_{\mathrm{i}} = h^{(\mu)}$, comparison of (26b) with the $\mathscr{O}_1$ part of (26a) gives

$$\sigma_{\mathrm{i}}{}^{\mathrm{j}}{}_{(\alpha)}{}^{(\beta)} \equiv \sigma^{(\mu)}{}_{(\nu)(\alpha)}{}^{(\beta)} = \delta^{(\mu)}_{(\alpha)}\delta^{(\beta)}_{(\nu)}\,. \qquad (26c)$$

For a covariant local four-vector $\mathrm{q}_{\mathrm{i}} = h_{(\mu)}$, we would have

$$\sigma_{\mathrm{i}}{}^{\mathrm{j}}{}_{(\alpha)}{}^{(\beta)} \equiv \sigma_{(\mu)}{}^{(\nu)}{}_{(\alpha)}{}^{(\beta)} = -\delta^{(\nu)}_{(\alpha)}\delta^{(\beta)}_{(\mu)}\,. \qquad (26d)$$

When local (tetrad) and world (coordinate) transformations are applied simultaneously, the components $h^{\mu}_{(\alpha)}$ of a tetrad field would transform according to

$$h'^{\mu}_{(\alpha)} = (\delta^{\mu}_{\nu} + \xi^{\mu}_{,\nu})\{\delta^{(\beta)}_{(\alpha)} - \epsilon^{(\beta)}{}_{(\alpha)} + \epsilon^{(\beta)}{}_{(\gamma)}\epsilon^{(\gamma)}{}_{(\alpha)}$$
$$- \epsilon^{(\beta)}{}_{(\delta)}\epsilon^{(\delta)}{}_{(\gamma)}\epsilon^{(\gamma)}{}_{(\alpha)} + \cdots\}h^{\nu}_{(\beta)}\,. \qquad (26e)$$

We will now define $\epsilon_{1(\alpha)(\beta)} = \mathring{g}_{(\alpha)(\gamma)}\epsilon^{(\gamma)}_1{}_{(\beta)}$ rigorously (also for finite local tensor transformations) as the antisymmetric part of $\epsilon_{(\alpha)(\beta)}$:

$$\epsilon_{1(\alpha)(\beta)} \equiv \tfrac{1}{2}(\epsilon_{(\alpha)(\beta)} - \epsilon_{(\beta)(\alpha)})\,, \qquad (27a)$$

and define $\epsilon_{2(\alpha)(\beta)}$ as its symmetric part:

$$\epsilon_{2(\alpha)(\beta)} \equiv \tfrac{1}{2}(\epsilon_{(\alpha)(\beta)} + \epsilon_{(\beta)(\alpha)})\,, \qquad (27b)$$

so that

$$\epsilon_{(\alpha)(\beta)} = \epsilon_{1(\alpha)(\beta)} + \epsilon_{2(\alpha)(\beta)}\,. \qquad (27c)$$

The invariance of the inverse Minkowski metric $\mathring{g}^{(\alpha)(\beta)}$ under local Lorentz transformations now requires

$$\epsilon^{(\alpha)(\beta)} + \epsilon^{(\beta)(\alpha)} + \epsilon^{(\alpha)(\delta)}\epsilon^{(\beta)}{}_{(\delta)} = 0\,, \qquad (28a)$$

that is,

$$\epsilon^{(\alpha)}_2{}_{(\beta)} = \epsilon_{2(\beta)}{}^{\alpha} = -\tfrac{1}{2}\epsilon^{(\alpha)(\lambda)}\epsilon_{(\beta)(\lambda)}\,. \qquad (28b)$$

It follows that $\epsilon_2^{(\alpha)}{}_{(\beta)}$ is small of second order, so that we may rigorously identify the $\mathscr{O}_1$ part of $\epsilon^{(\alpha)}{}_{(\beta)}$ with its antisymmetric part $\epsilon_1^{(\alpha)}{}_{(\beta)} \equiv \mathring{g}^{(\alpha)(\gamma)}\epsilon_{1(\gamma)(\beta)}$. Inserting now (27c) on the right in (28b), we can solve in successive approximations for $\epsilon_2^{(\alpha)}{}_{(\beta)}$ in terms of $\epsilon_1^{(\alpha)}{}_{(\beta)}$. As shown in the Appendix, this yields

$$\epsilon_2^{(\alpha)}{}_{(\beta)} = \tfrac{1}{2}(\epsilon_1{}^2)^{(\alpha)}{}_{(\beta)} - \tfrac{1}{8}(\epsilon_1{}^4)^{(\alpha)}{}_{(\beta)} + \cdots \tag{28c}$$

$$= \sum_{n=1}^{\infty} \binom{\frac{1}{2}}{n} (\epsilon_1{}^{2n})^{(\alpha)}{}_{(\beta)}, \tag{28d}$$

where $(\epsilon_1{}^N)^{(\alpha)}{}_{(\beta)}$ is the chain matrix product

$$\epsilon_1^{(\alpha)}{}_{(\gamma)}\epsilon_1^{(\gamma)}{}_{(\delta)}\epsilon_1^{(\delta)}{}_{(\epsilon)} \cdots \epsilon_1^{(\kappa)}{}_{(\lambda)}\epsilon_1^{(\lambda)}{}_{(\beta)}$$

of N factors $\epsilon_1^{(\eta)}{}_{(\theta)}$. Therefore,

$$\epsilon^{(\alpha)}{}_{(\beta)} = \epsilon_1^{(\alpha)}{}_{(\beta)} + \tfrac{1}{2}\epsilon_1^{(\alpha)}{}_{(\gamma)}\epsilon_1^{(\gamma)}{}_{(\beta)} + \mathscr{O}_4 . \tag{28e}$$

This shows that we do not need all 16 components of $\epsilon^{(\alpha)}{}_{(\beta)}$ as descriptors: As expected, six descriptors suffice for describing all Lorentz transformations, and we see from (28d) that we can use for these descriptors the antisymmetric $\epsilon_{1(\alpha)(\beta)}$.

XII. COMPARISON OF LORENTZ TRANSFORMATIONS OF WORLD TENSORS AND OF LOCAL TENSORS

As, up to $\mathscr{O}_2$,

$$\epsilon_1^{(\alpha)}{}_{(\beta)} = \epsilon^{(\alpha)}{}_{(\beta)} - \tfrac{1}{2}\epsilon^{(\alpha)}{}_{(\gamma)}\epsilon^{(\gamma)}{}_{(\beta)}, \tag{29a}$$

the descriptors $\epsilon_1^{(\alpha)}{}_{(\beta)}$ for local Lorentz transformations are equivalent to the combinations

$$\xi^{\mu}_{,\nu} - \tfrac{1}{2}\xi^{\mu}_{,\lambda}\xi^{\lambda}_{,\nu} \tag{29b}$$

of descriptors for coordinate transformations. When we insert (12) with (24) in Eq. (8), this combination (29b) becomes the coefficient of $s_i{}^j{}_\mu{}^\nu q_j$. Because in (26b) we placed the entire combination (29a) in $\delta_1 q_i$, we find in $\delta_2 q_i$ no term that would for local transformations correspond to the term with $-\tfrac{1}{2}\xi^{\mu}_{,\lambda}\xi^{\lambda}_{,\nu}$ in $\delta_2 q_i$, because that term is already part of the term with $\epsilon_1^{(\alpha)}{}_{(\beta)}$ in $\delta_1 q_i$.

XIII. COMBINATION OF WORLD TENSOR AND LOCAL TENSOR TRANSFORMATIONS

We shall now generalize the discussions of Secs. VIII–X to fields q_i that have local tensor indices (α), (β), ... as well as world tensor indices μ, ν, ..., like, for instance, tetrads. Like in (12), we first write $\delta q_i = \delta_1 q_i + \delta_2 q_i + \mathscr{O}_3$ with

$$\delta_1 q_i = \{\xi^{\mu}_{,\nu} s_i{}^j{}_\mu{}^\nu + \epsilon_1^{(\alpha)}{}_{(\beta)}\sigma_i{}^j{}_{(\alpha)}{}^{(\beta)}\} q_j \equiv (\delta_1 \hat{S}_i{}^j) q_j , \tag{30a}$$

$$\delta_2 q_i = \{\xi^{\mu}_{,\nu}\xi^{\rho}_{,\sigma} s_i{}^k{}_\mu{}^\nu{}_\rho{}^\sigma + \xi^{\mu}_{,\nu}\epsilon_1^{(\alpha)}{}_{(\beta)} S_i{}^k{}_\mu{}^\nu{}_{(\alpha)}{}^{(\beta)} + \epsilon_1^{(\alpha)}{}_{(\beta)}\epsilon_1^{(\gamma)}{}_{(\delta)}\sigma_i{}^k{}_{(\alpha)}{}^{(\beta)}{}_{(\gamma)}{}^{(\delta)}\} q_k , \tag{30b}$$

and then follow the reasoning of Secs. VIII and IX. By $\Sigma \to \Sigma'$ we this time understand a coordinate transformation with simultaneous[32] independent local rotations or Lorentz transformations of the tetrads. For $\Sigma \to \Sigma'$ (and $\Sigma'' \to \Sigma''''$) and for $\Sigma \to \Sigma''$ (and $\Sigma' \to \Sigma'''$) we used in Sec. VIII transformations with descriptors ξ^μ and η^μ. This time we use ξ^μ together with $\epsilon_1^{(\alpha)}{}_{(\beta)}$, and η^μ together with $v_1^{(\alpha)}{}_{(\beta)}$. In the transformation of a local four-vector $h^{(\alpha)}$ from Σ' to Σ''', we now use, according to (26a) and (28e),

$$h'''^{(\alpha)} = h'^{(\alpha)} + v_1'^{(\alpha)}{}_{(\beta)} h'^{(\beta)} + \tfrac{1}{2} v_1'^{(\alpha)}{}_{(\epsilon)} v_1'^{(\epsilon)}{}_{(\beta)} h'^{(\beta)}, \tag{31a}$$

with

$$v_1'^{(\alpha)}{}_{(\beta)} = v_1^{(\alpha)}{}_{(\beta)} + v^{(\alpha)}_{1(\beta),\lambda}\xi^\lambda . \tag{31b}$$

We then find

$$h''''^{(\alpha)} - h'''^{(\alpha)} = \vartheta_1^{(\alpha)}{}_{(\beta)} h^{(\beta)}, \tag{32a}$$

with

$$\vartheta_1^{(\alpha)}{}_{(\beta)} = \epsilon_1^{(\alpha)}{}_{(\epsilon)} v_1^{(\epsilon)}{}_{(\beta)} - v_1^{(\alpha)}{}_{(\epsilon)}\epsilon_1^{(\epsilon)}{}_{(\beta)} + \epsilon_1^{(\alpha)}{}_{(\beta),\lambda}\eta^\lambda - v^{(\alpha)}{}_{(\beta),\lambda}\xi^\lambda , \tag{32b}$$

which, together with (16a), shows that $\Sigma''' \to \Sigma''''$ is given by descriptors ζ^μ and $\vartheta_1^{(\alpha)}{}_{(\beta)}$.

We now compare the transformation $\Sigma \to \Sigma'' \to \Sigma''''$ of q_i again first with $\Sigma \to \Sigma' \to \Sigma''' \to \Sigma''''$ as in Sec. VIII, and then with $\Sigma \to \Sigma''''$ as in Sec. IX. The result for $q_i \to q_i''''$ should be always the same. The first comparison gives not only Eq. (19), but additionally gives the relations

$$s_i{}^j{}_\mu{}^\nu[\sigma_j{}^{k(\alpha)(\beta)} - \sigma_j{}^{k(\beta)(\alpha)}] = [\sigma_i{}^{j(\alpha)(\beta)} - \sigma_i{}^{j(\beta)(\alpha)}] s_j{}^k{}_\mu{}^\nu \tag{33}$$

and

$$\epsilon_1^{(\alpha)}{}_{(\beta)} v_1^{(\gamma)}{}_{(\delta)} [\sigma_i{}^j{}_{(\alpha)}{}^{(\beta)}\sigma_j{}^k{}_{(\gamma)}{}^{(\delta)} - \sigma_i{}^j{}_{(\gamma)}{}^{(\delta)}\sigma_j{}^k{}_{(\alpha)}{}^{(\beta)}] = \epsilon_1^{(\alpha)}{}_{(\beta)} v_1^{(\gamma)}{}_{(\delta)} [\sigma_i{}^k{}_{(\alpha)}{}^{(\delta)}\delta^{(\beta)}_{(\gamma)} - \sigma_i{}^k{}_{(\gamma)}{}^{(\beta)}\delta^{(\delta)}_{(\alpha)}] . \tag{34}$$

The second comparison requires knowledge of the descriptor $v_1^{(\alpha)}{}_{(\beta)} = v^{(\alpha)}{}_{(\beta)} - \tfrac{1}{2}v^{(\alpha)}{}_{(\epsilon)}v^{(\epsilon)}{}_{(\beta)}$ for the tetrad transformation directly from Σ to Σ''''. Since, by (32a), (32b), (31a), (31b), and (26a),

$$h''''^{(\alpha)} - h^{(\alpha)} = \{\epsilon^{(\alpha)}{}_{(\beta)} + v^{(\alpha)}{}_{(\beta)} + \epsilon^{(\alpha)}{}_{(\epsilon)} v^{(\epsilon)}{}_{(\beta)} + \epsilon^{(\alpha)}{}_{(\beta),\lambda}\eta^\lambda\} h^{(\beta)} = v^{(\alpha)}{}_{(\beta)} h^{(\beta)}, \tag{35}$$

we find, by (28e),

$$v_1^{(\alpha)}{}_{(\beta)} = \epsilon_1^{(\alpha)}{}_{(\beta)} + v_1^{(\alpha)}{}_{(\beta)} + \tfrac{1}{2}\epsilon_1^{(\alpha)}{}_{(\epsilon)} v_1^{(\epsilon)}{}_{(\beta)} - \tfrac{1}{2} v_1^{(\alpha)}{}_{(\epsilon)}\epsilon_1^{(\epsilon)}{}_{(\beta)} + \epsilon_1^{(\alpha)}{}_{(\beta),\lambda}\eta^\lambda , \tag{36}$$

and $q_i'''' - q_i$ is given by the sum of (30a) and (30b) with ξ^μ replaced by $\xi^\mu + \eta^\mu + \xi^{\mu}_{,\lambda}\eta^\lambda$, and with $\epsilon_1^{(\alpha)}{}_{(\beta)}$ replaced by $v_1^{(\alpha)}{}_{(\beta)}$.

The second comparison now gives not only Eq. (24), but also

$$S_i{}^k{}_\mu{}^{\nu(\alpha)(\beta)} - S_i{}^k{}_\mu{}^{\nu(\beta)(\alpha)} = s_i{}^j{}_\mu{}^\nu[\sigma_j{}^{k(\alpha)(\beta)} - \sigma_j{}^{k(\beta)(\alpha)}] \tag{37}$$

and

$$\epsilon_1^{(\alpha)}{}_{(\beta)} v_1^{(\gamma)}{}_{(\delta)} \{\sigma_i{}^k{}_{(\alpha)}{}^{(\beta)}{}_{(\gamma)}{}^{\delta} + \sigma_i{}^k{}_{(\gamma)}{}^{(\delta)}{}_{(\alpha)}{}^{(\beta)} - \sigma_i{}^j{}_{(\alpha)}{}^{(\beta)}\sigma_j{}^k{}_{(\gamma)}{}^{(\delta)} + \tfrac{1}{2}(\sigma_i{}^k{}_{(\alpha)}{}^{(\delta)}\delta^{(\beta)}_{(\gamma)} - \sigma_i{}^k{}_{(\gamma)}{}^{(\beta)}\delta^{(\delta)}_{(\alpha)})\} = 0 . \tag{38}$$

By (34), this gives

$$\epsilon_1^{(\alpha)}{}_{(\beta)} v_1^{(\gamma)}{}_{(\delta)}\sigma_i{}^k{}_{(\alpha)}{}^{(\beta)}{}_{(\gamma)}{}^{(\delta)} = \tfrac{1}{4}\epsilon_1^{(\alpha)}{}_{(\beta)} v_1^{(\gamma)}{}_{(\delta)} [\sigma_i{}^j{}_{(\alpha)}{}^{(\beta)}\sigma_j{}^k{}_{(\gamma)}{}^{(\delta)} + \sigma_i{}^j{}_{(\gamma)}{}^{(\delta)}\sigma_j{}^k{}_{(\alpha)}{}^{(\beta)}] . \tag{39}$$

Now using Eqs. (24), (37) with (33), and (39) with $v_1^{(\alpha)}{}_{(\beta)} = \epsilon_1^{(\alpha)}{}_{(\beta)}$, we find from (30b)

$$\delta_2 q_i = \tfrac{1}{2}(\delta_1 \hat{S}_i{}^j)(\delta_1 \hat{S}_j{}^k) q_k - \tfrac{1}{2}\xi^{\mu}_{,\lambda}\xi^{\lambda}_{,\nu} s_i{}^k{}_\mu{}^\nu q_k , \tag{30b$'$}$$

with $\delta_1 \hat{S}_i{}^j$ from (30a). Thence,

$$\delta q_i = [(\xi^{\mu}_{,\nu} - \tfrac{1}{2}\xi^{\mu}_{,\lambda}\xi^{\lambda}_{,\nu}) s_i{}^k{}_\mu{}^\nu + \epsilon_1^{(\alpha)}{}_{(\beta)}\sigma_i{}^k{}_{(\alpha)}{}^{(\beta)} + \tfrac{1}{2}(\delta_1 \hat{S}_i{}^j)(\delta_1 \hat{S}_j{}^k)] q_k + \mathscr{O}_3 . \tag{30b$''$}$$

Examples: As a first example, we may apply (30b″) to $q_i \equiv h^{(\mu)}$, with $\sigma_i{}^j{}_{(\alpha)}{}^{(\beta)}$ given by (26c) and with $s_i{}^j{}_\mu{}^\nu = 0$, so that $\delta_1 \hat{S}_i{}^k = \delta \hat{S}^{(\mu)}{}_{(\nu)} = \epsilon_1^{(\mu)}{}_{(\nu)}$. By (28e), then, Eq. (30b″) gives (26a). As a second example, for $q_i = h^{\mu}_{(\alpha)}$, with $\delta_1 \hat{S}_i{}^k = \delta_1 \hat{S}^{\mu(\beta)}_{(\alpha)\nu} = \xi^{\mu}_{,\nu} \delta^{(\beta)}_{(\alpha)} - \delta^{\mu}_{\nu} \epsilon_1^{(\beta)}{}_{(\alpha)}$ and $s_i{}^k{}_\rho{}^\sigma = s^{\mu(\beta)\sigma}_{(\alpha)\nu\rho} = \delta^{\mu}_{\rho} \delta^{\sigma}_{\nu} \delta^{(\beta)}_{(\alpha)}$, Eq. (30b″) gives (26e) up to (and including) the $\mathscr{O}_2$ terms.

XIV. UNDOR TRANSFORMATIONS

In Secs. V–X, we wrote q_i for "tensor" fields (in a generalized sense). In Secs. XI and XII, we wrote φ_i for boson fields, and we wrote q_i for local tensors (which transform under changes of the tetrad field). In Sec. XIII, we considered both coordinate and tetrad field transformations, so that the field variables here could be world tensors as well as local tensors, or combinations of both, like tetrads, or also fermion fields. Therefore, we changed our notation in Sec. XIII from φ_i to q_i.

The q_i therefore may also include undors[3] ψ and $\bar{\psi}$. As is well known, Dirac wave functions of fermions transform locally under Lorentz transformations of the local tetrad, infinitesimally to first order, according to

$$\delta_1 \psi = \tfrac{1}{8} \epsilon^{(\alpha)}{}_{(\beta)} (\gamma_{(\alpha)} \gamma^{(\beta)} - \gamma^{(\beta)} \gamma_{(\alpha)}) \psi = \tfrac{1}{4} \epsilon_{1(\alpha)(\beta)} \gamma^{(\alpha)} \gamma^{(\beta)} \psi , \tag{40}$$

where the $\gamma^{(\alpha)}$ are the Dirac matrices of flat-space theory, related among each other and with the γ^{μ} of the curved-space Dirac equation by

$$\gamma^{(\alpha)} \gamma^{(\beta)} + \gamma^{(\beta)} \gamma^{(\alpha)} = 2 \mathring{g}^{(\alpha)(\beta)} , \tag{41a}$$

$$\gamma^{\mu} = h^{\mu}_{(\alpha)} \gamma^{(\alpha)} , \tag{41b}$$

$$\gamma^{\mu} \gamma^{\nu} + \gamma^{\nu} \gamma^{\mu} = 2 g^{\mu\nu} , \tag{41c}$$

and related to the β in $\bar{\psi} = \psi^{\dagger} \beta$ by

$$\gamma^{(\alpha)\dagger} \beta = -\beta \gamma^{(\alpha)} , \tag{41d}$$

$$\gamma^{\mu\dagger} \beta = -\beta \gamma^{\mu} . \tag{41e}$$

Rewriting (40) as $\delta_1 \psi = \Omega \psi$, we obtain

$$\delta_1 \bar{\psi} = (\delta_1 \psi)^{\dagger} \beta = \bar{\psi} \beta \Omega^{\dagger} \beta = \tfrac{1}{4} \bar{\psi} \epsilon_{1(\alpha)(\beta)} \gamma^{(\alpha)} \gamma^{(\beta)} = -\bar{\psi} \Omega . \tag{40'}$$

Comparison of (40) and (40′) with (26b) shows that, for $q_i = \psi_A$ with $q_j = \psi_B$,

$$\sigma_i{}^j{}_{(\alpha)}{}^{(\beta)} = \sigma_A{}^B{}_{(\alpha)}{}^{(\beta)} = \tfrac{1}{8} [\gamma_{(\alpha)} \gamma^{(\beta)} - \gamma^{(\beta)} \gamma_{(\alpha)}]_A{}^B , \tag{42a}$$

while, for $q_i = \bar{\psi}^A$ with $q_j = \bar{\psi}^B$,

$$\sigma_i{}^j{}_{(\alpha)}{}^{(\beta)} = \sigma^A{}_{B(\alpha)}{}^{(\beta)} = -\tfrac{1}{8} [\gamma_{(\alpha)} \gamma^{(\beta)} - \gamma^{(\beta)} \gamma_{(\alpha)}]_B{}^A . \tag{42b}$$

Note that, by the antisymmetry of the $\sigma_i{}^{j(\alpha)(\beta)}$ between their last two indices, the middle member of Eq. (40) contains only the antisymmetric part $\epsilon_{1(\alpha)(\beta)}$ of the $\epsilon_{(\alpha)(\beta)}$, in agreement with what we postulated in Eq. (26b).

We may now obtain $\delta_2 \psi$ and $\delta_2 \bar{\psi}$ from (40) and (40′) either by Eq. (30b′) or by noting that $\bar{\psi}\psi$ is scalar, so that

$$0 = \delta_2(\bar{\psi}\psi) = (\delta_2 \bar{\psi})\psi + (\delta_1 \bar{\psi})(\delta_1 \psi) + \bar{\psi}(\delta_2 \psi)$$

gives

$$(\delta_2 \bar{\psi})\psi + \bar{\psi}(\delta_2 \psi) = \bar{\psi} \Omega \Omega \psi . \tag{43}$$

Either method gives

$$\delta_2 \psi = \tfrac{1}{2} \Omega \Omega \psi = \tfrac{1}{2} \delta_1 \delta_1 \psi$$
$$= \tfrac{1}{32} \epsilon^{(\alpha)}_{1(\beta)} \epsilon^{(\gamma)}_{1(\delta)} \gamma_{(\alpha)} \gamma^{(\beta)} \gamma_{(\gamma)} \gamma^{(\delta)} \psi , \tag{40''}$$

$$\delta_2 \bar{\psi} = \tfrac{1}{2} \bar{\psi} \beta \Omega^{\dagger} \Omega^{\dagger} \beta$$
$$= \tfrac{1}{32} \epsilon^{(\alpha)}_{1(\beta)} \epsilon^{(\gamma)}_{1(\delta)} \bar{\psi} \gamma^{(\delta)} \gamma_{(\gamma)} \gamma^{(\beta)} \gamma_{(\alpha)}$$
$$= \tfrac{1}{2} \bar{\psi} \Omega \Omega . \tag{40'''}$$

By Eqs. (40), (40′), and (28e) it then follows with the help of (41a)–(41e) that, to second order,

$$\delta(\bar{\psi} \gamma^{(\mu)} \psi) = \epsilon^{(\mu)}{}_{(\nu)} (\bar{\psi} \gamma^{(\nu)} \psi) , \tag{44}$$

in agreement with (26a).

As seen from the above, Eqs. (30a), (30b), (30b′), and (30b″) are now valid for any q, be it a world tensor, a local tensor, or an undor.

XV. TRANSFORMATIONS OF FIELD MOMENTA

If we want to prove the covariance of the commutation relations assumed to be valid in coordinate system and tetrad field Σ, under transformation from Σ to Σ', we first must know how the field momenta p^i will transform. This is one of the reasons why we must make an assumption about the transformation properties of the Lagrangian function $L = \mathscr{L}/\sqrt{-g}$, and for simplicity we will assume it to be scalar under the transformations considered. This is not as much of a restriction as it may seem at first sight, as the space-time integral of $\mathscr{L}$ cannot be (but for boundary terms) invariant under general coordinate transformations anyhow, if the alteration of $\mathscr{L}$ is to have removed all first-class constraints. The altered L usually proposed are scalar under affine coordinate transformations and under local or sometimes only global Lorentz transformations; whatever its invariance properties are will then restrict the transformations which we may consider in the following.

With $\mathscr{L}$ now a scalar density, and with $\mathfrak{P}^{i\mu} \equiv \partial \mathscr{L} / \partial \varphi_{i,\mu}$, the sum (over i) $\mathfrak{P}^{i\mu} \varphi_i$ is a four-vector density. This determines the $\delta \mathfrak{P}^{i\mu}$, and therefore the $\delta p^i \equiv \delta \mathfrak{P}^{i0}/c$. In particular, splitting up all local variations into parts $\mathscr{O}_1$ and $\mathscr{O}_2$ in the descriptors ξ^{μ} and $\epsilon_{1(\alpha)(\beta)}$ and using the notation $\delta_1 \hat{S}_i{}^j$ of Eq. (30a), we obtain

$$\delta_1 \mathfrak{P}^{k\mu} = \xi^{\mu}_{,\lambda} \mathfrak{P}^{k\lambda} - \mathfrak{P}^{k\mu} \xi^{\beta}_{,\beta} - \mathfrak{P}^{i\mu} \delta_1 \hat{S}_i{}^k , \tag{45a}$$

$$\delta_2 \mathfrak{P}^{k\mu} = \tfrac{1}{2} (\xi^{\alpha}_{,\alpha} \xi^{\beta}_{,\beta} + \xi^{\alpha}_{,\beta} \xi^{\beta}_{,\alpha}) \mathfrak{P}^{k\mu} - \xi^{\beta}_{,\beta} \xi^{\mu}_{,\lambda} \mathfrak{P}^{k\lambda}$$
$$+ (\xi^{\beta}_{,\beta} \mathfrak{P}^{i\mu} - \xi^{\mu}_{,\lambda} \mathfrak{P}^{i\lambda}) \delta_1 \hat{S}_i{}^k + \tfrac{1}{2} \mathfrak{P}^{i\mu} (\xi^{\alpha}_{,\lambda} \xi^{\lambda}_{,\beta} s_i{}^k{}_\alpha{}^\beta)$$
$$+ \tfrac{1}{2} \mathfrak{P}^{i\mu} (\delta_1 \hat{S}_i{}^j)(\delta_1 \hat{S}_j{}^k) , \tag{45b}$$

so that

$$\delta_1 p^k = \xi^{0}_{,n} \mathfrak{P}^{kn}/c - \xi^{n}_{,n} p^k - p^i \delta_1 \hat{S}_i{}^k , \tag{46a}$$

$$\delta_2 p^k = (\xi^{0}_{,n} \xi^{n}_{,0} + \tfrac{1}{2} \xi^{m}_{,n} \xi^{n}_{,m} + \tfrac{1}{2} \xi^{m}_{,m} \xi^{n}_{,n}) p^k - \xi^{\beta}_{,\beta} \xi^{0}_{,n} \mathfrak{P}^{kn}/c$$
$$+ (\xi^{n}_{,n} p^i - \xi^{0}_{,n} \mathfrak{P}^{in}/c) \delta_1 \hat{S}_i{}^k + \tfrac{1}{2} p^i \xi^{\alpha}_{,\lambda} \xi^{\lambda}_{,\beta} s_i{}^k{}_\alpha{}^\beta$$
$$+ \tfrac{1}{2} p^i (\delta_1 \hat{S}_i{}^j)(\delta_1 \hat{S}_j{}^k). \tag{46b}$$

If δ denotes the transformation of field variables like p^k, $\mathfrak{P}^{kn}$, or q_i, but not any transformation of $\xi^{\mu}_{,\nu}$ or $\delta_1 \hat{S}_i{}^j$, we see from (46a) with (45a) that

$$\tfrac{1}{2}\delta_1\delta_1 p^k = \tfrac{1}{2}\xi^0_{,n}\delta_1\mathfrak{P}^{kn}/c - \tfrac{1}{2}\xi^n_{,n}\delta p^k - \tfrac{1}{2}\delta p^i\delta_1\hat{S}_i{}^k$$
$$= (\tfrac{1}{2}\xi^0_{,\lambda}\xi^\lambda_{,n} - \xi^\lambda_{,\lambda}\xi^0_{,n})\mathfrak{P}^{kn}/c$$
$$+ \tfrac{1}{2}(\xi^m_{,m}\xi^n_{,n} + \xi^n_{,0}\xi^0_{,n})p^k$$
$$+ (\xi^n_{,n}p^i - \xi^0_{,n}\mathfrak{P}^{in}/c)\delta_1\hat{S}_i{}^k + \tfrac{1}{2}p^i(\delta_1\hat{S}_i{}^j)(\delta_1\hat{S}_j{}^k), \quad (47)$$

so that

$$\delta_2 p^k = \tfrac{1}{2}\delta_1\delta_1 p^k - \tfrac{1}{2}\delta_1^\times p^k, \quad (48a)$$

where $\delta_1^\times p^k$ is obtained from the terms in $\delta_1 p^k$ that contain $\xi^\mu_{,\nu}$, by replacing this $\xi^\mu_{,\nu}$ by $\xi^\mu_{,\lambda}\xi^\lambda_{,\nu}$. Note that, similarly, (30b') may be written as

$$\delta_2 q_i = \tfrac{1}{2}\delta_1\delta_1 q_i - \tfrac{1}{2}\delta_1^\times q_i, \quad (48b)$$

so that, if F is any q_i or p^k, in general we may write

$$\delta F = \delta_1 F + \delta_2 F = \delta_1 F + \tfrac{1}{2}\delta_1\delta_1 F - \tfrac{1}{2}\delta_1^\times F. \quad (48c)$$

Like $\delta_1 q_i$, also $\delta_1 p^k$, and in general $\delta_1 F$ is linear in the coefficients $\xi^\mu_{,\nu}$ and/or $\epsilon_{1(\alpha)(\beta)}$. We will in the following understand $\delta_{1,\lambda} F$ as a differentiation of these coefficients in $\delta_1 F$ with respect to x^λ, while in $\delta_1 F_{,\lambda}$ or $\delta_1(\xi^\lambda F_{,\lambda})$ the δ_1 again describes the transformation of F only, so that, for instance, $\delta_1(q_{i,\lambda})$ means $(\delta_1\hat{S}_i{}^j)q_{j,\lambda}$ with $\delta_1\hat{S}_i{}^j$ unchanged from Eq. (30a).

XVI. SUBSTANTIAL VARIATION OF FIELDS UP TO SECOND ORDER

For discussing Eq. (3) and its integrability to (1) for finite transformations, we also need expressions with second-order accuracy for the substantial variation $\bar{\delta}F(P)$ from $F(P)$ to $F'(P')$, where $x'^\mu(P') = x^\mu(P)$. Using the notation $\Delta x^\mu \equiv x^\mu(P') - x^\mu(P)$, we find from

$$x'^\mu_{P'} = [x^\mu_P + \Delta x^\mu] + \xi^\mu(P')$$
$$= x^\mu_P + \Delta x^\mu + \xi^\mu + \xi^\mu_{,\nu}\Delta x^\nu = x^\mu_P$$

that, up to (and including) $\mathcal{O}_2$,

$$\Delta x^\mu = -\xi^\mu + \xi^\mu_{,\lambda}\xi^\lambda. \quad (49)$$

With F again any q_i or p^k, and with the notation $\Delta F(P) = F(P') - F(P)$, we then have

$$\bar{\delta}F(P) = F'(P') - F(P) = \delta F(P') + \Delta F(P)$$
$$= \delta_1(F + F_{,\lambda}\Delta x^\lambda) + (\delta_{1,\lambda} F)\Delta x^\lambda + \delta_2 F + F_{,\mu}\Delta x^\mu$$
$$+ \tfrac{1}{2}F_{,\mu\nu}\Delta x^\mu\Delta x^\nu. \quad (50)$$

Inserting (44) in (45), we find

$$\bar{\delta}F = \bar{\delta}_1 F + \bar{\delta}_2 F, \quad (51a)$$

$$\bar{\delta}_1 F = \delta_1 F - \xi^\mu F_{,\mu}, \quad (51b)$$

$$\bar{\delta}_2 F = \delta_2 F - \xi^\lambda\delta_1 F_{,\lambda} - \xi^\lambda\delta_{1,\lambda} F$$
$$+ \xi^\lambda\xi^\mu_{,\lambda}F_{,\mu} + \tfrac{1}{2}\xi^\mu\xi^\nu F_{,\mu\nu}. \quad (51c)$$

Now, let δ_{12} stand for δ_1 with ξ^μ and $\epsilon_{1(\alpha)(\beta)}$ replaced by[33]

$$\varphi^\mu \equiv \xi^\mu - \tfrac{1}{2}\xi^\mu_{,\lambda}\xi^\lambda \quad (52a)$$

and

$$\omega_{1(\alpha)(\beta)} \equiv \epsilon_{1(\alpha)(\beta)} - \tfrac{1}{2}\epsilon_{1(\alpha)(\beta),\lambda}\xi^\lambda, \quad (52b)$$

so that $\xi^\mu_{,\nu}$ is replaced by

$$\varphi^\mu_{,\nu} = \xi^\mu_{,\nu} - \tfrac{1}{2}\xi^\mu_{,\lambda}\xi^\lambda_{,\nu} - \tfrac{1}{2}\xi^\mu_{,\nu\lambda}\xi^\lambda. \quad (52c)$$

Then,

$$\delta_{12} F = \delta_1 F - \tfrac{1}{2}\xi^\lambda\delta_{1,\lambda} F - \tfrac{1}{2}\delta_1^\times F. \quad (53)$$

Therefore, Eqs. (51a)–(51c) with (48c) give

$$\bar{\delta}F = \delta_{12} F - \varphi^\mu F_{,\mu} + \tfrac{1}{2}\delta_1\delta_1 F - \tfrac{1}{2}\xi^\lambda\delta_{1,\lambda} F$$
$$- \xi^\lambda\delta_1 F_{,\lambda} + \tfrac{1}{2}\xi^\lambda\xi^\mu_{,\lambda}F_{,\mu} + \tfrac{1}{2}\xi^\mu\xi^\nu F_{,\mu\nu}$$
$$= \bar{\delta}_{12} F + \tfrac{1}{2}(\delta_1 - \xi^\lambda\partial_\lambda)(\delta_1 - \xi^\mu\partial_\mu)F, \quad (54a)$$

where $\bar{\delta}_{12} F$ is obtained from $\bar{\delta}_1 F$ by the same substitutions (52a)–(52c) as by which $\delta_{12} F$ was obtained from $\delta_1 F$. Equation (51b) allows us to introduce the abbreviation

$$\bar{\delta}_1 = \delta_1 - \xi^\mu\partial_\mu, \quad (54b)$$

so that

$$\bar{\delta}F = \bar{\delta}_{12} F + \tfrac{1}{2}\bar{\delta}_1\bar{\delta}_1 F, \quad (54c)$$

neglecting third-order infinitesimals. This tells us how $\bar{\delta}F$ may be obtained from $\bar{\delta}_1 F$ and $\bar{\delta}_1(\bar{\delta}_1 F)$.

XVII. THE TRANSFORMATION GENERATOR UP TO SECOND ORDER

For proving the covariance of the assumed commutation relations (which will be Dirac's *modified* commutation relations[17,20,26]) we will have to show the existence of a conserved and invariant generator T with the property (1). For this purpose, we start by defining on a hypersurface $x^0 = \text{const}$ in Σ the familiar first-order generator[2,4]

$$T_1(\xi,\epsilon) = \hbar^{-1}\int_{t=\text{const}} d^3x\{p^i[\delta_1 q_i - \xi^n q_{i,n}] - \xi^0\mathscr{H}/c\}, \quad (55a)$$

where $\mathscr{H} = cp^i q_{i,0} - \mathscr{L}$ is here used as Hamiltonian density. Equation (55a) is in terms of canonical variables, what in covariant notation would be

$$T_1 = (\hbar c)^{-1}\int_0 d\sigma_\nu\left\{\frac{\partial\mathscr{L}}{\partial q_{i,\nu}}\bar{\delta}_1 q_i + \mathscr{L}\xi^\nu\right\}. \quad (55b)$$

This first-order-infinitesimal expression (linear in the descriptors ξ and ϵ of the transformations) is identical with the generator used in papers by others who are discussing different quantization procedures.[34]

From (55a) we will now construct, by the substitutions (52a)–(52c), the new generator

$$T \equiv T_{12}(\xi,\epsilon) = T_1(\varphi,\omega)$$
$$= \hbar^{-1}\int_{t=\text{const}} d^3x\{p^i\bar{\delta}_{12}q_i - \varphi^0\mathscr{H}/c\}. \quad (56)$$

In our next paper[26] we will then first prove that (but for irrelevant boundary terms that commute with all field variables F at finite positions) the q number (56) is conserved and is invariant, provided that under the transformations $\Sigma\to\Sigma'$ the descriptors transform as their indices suggest, and that we confine transformations to those, under which $\mathscr{L}$ transforms like a scalar density. Next, by the assumed (modified) commutation relations,[26] we will there[26] calculate the commutators of $T_1(\xi,\epsilon)$ with the q_i and the p^k, and with second-order precision[36] we will find[26]

$$[iT_1(\xi,\epsilon);F]=\bar{\delta}_1 F, \tag{57a}$$

that is Eq. (3). As here both sides are linear in the descriptors ξ^{μ} and $\epsilon_{1(\alpha)(\beta)}$, it follows at once that the substitutions (52a)–(52c) give

$$[iT;F]=[iT_{12}(\xi,\epsilon);F]=[iT_1(\varphi,\omega);F]=\bar{\delta}_{12}F. \tag{57b}$$

Thence, always neglecting $\mathscr{O}_3$,

$$e^{iT}Fe^{-iT}=F+[iT;F]+\tfrac{1}{2}[iT;(IT;F)]$$
$$=F+\bar{\delta}_{12}F+\tfrac{1}{2}[iT;\bar{\delta}F]. \tag{58}$$

By $\bar{\delta}F=F'(P')-F(P)$ and $T'=T$, we obtain with $\mathscr{O}_2$ accuracy

$$[iT;\bar{\delta}F]=[iT';F'(P')]-[iT;F]$$
$$=\bar{\delta}_1F'(P')-\bar{\delta}_1F(P)=\bar{\delta}_1\bar{\delta}_1F, \tag{59}$$

and therefore

$$e^{iT}Fe^{-iT}=F+\bar{\delta}_{12}F+\tfrac{1}{2}\bar{\delta}_1\bar{\delta}_1F=F(P)+\bar{\delta}F=F'(P'), \tag{60}$$

according to (54c). This shows that, once we will have proved that $T_1(\xi,\epsilon)$ is with second-order precision[36] the first-order generator of Eq. (3), the T defined by (56) as $T_1(\varphi,\omega)$ will automatically, up to second order, be the generator T of Eq. (1) for the infinitesimal substantial variation of field variables, and, as explained before, will by successive application in infinitesimal steps be integrable to a similar generation of finite transformations. We see that for infinitesimal transformations the second-order generator is given by the first-order generator calculated with second-order precision at the point halfway between P and P'.[33]

XVIII. VERIFICATION OF INTEGRABILITY

We will now directly verify the integrability condition (5), that the result $F''''(P'''')$ of two successive infinitesimal transformations $\Sigma\to\Sigma''$ [with $v_{1(\alpha)(\beta)}$ and with $x''^{\mu}=x^{\mu}+\eta^{\mu}(x)$, see Fig. 1] and $\Sigma''\to\Sigma''''$ [with $\epsilon_{1(\alpha)(\beta)}(x'')$ and with $x''''^{\mu}=x''^{\mu}+\xi^{\mu}(x'')$], obtained by

$$F''''(P'''')=e^{iT_{12}(\xi,\epsilon)}e^{iT_{12}(\eta,v)}F(P)e^{-iT_{12}(\eta,v)}e^{-iT_{12}(\xi,\epsilon)}, \tag{61a}$$

is the same as the result that is obtained by a direct transformation

$$e^{iT_{12}(\rho,\nu)}F(P)e^{-iT_{12}(\rho,\nu)}, \tag{61b}$$

where ρ^{μ} and $\nu_{1(\alpha)(\beta)}$ are given by Eqs. (15b) and (36).

For expressing the various T_{12} here in terms of the T_1, we use the abbreviations (52a) and (52b) and

$$\chi^{\mu}=\eta^{\mu}-\tfrac{1}{2}\eta^{\mu}_{,\lambda}\eta^{\lambda}, \tag{62a}$$

$$\gamma^{(\alpha)}_{1\ (\beta)}=v^{(\alpha)}_{1\ (\beta)}-\tfrac{1}{2}v^{(\alpha)}_{1\ (\beta),\lambda}\eta^{\lambda}, \tag{62b}$$

$$\psi^{\mu}=\rho^{\mu}-\tfrac{1}{2}\rho^{\mu}_{,\lambda}\rho^{\lambda}, \tag{62c}$$

$$\theta^{(\alpha)}_{1\ (\beta)}=\nu^{(\alpha)}_{1\ (\beta)}-\tfrac{1}{2}\nu^{(\alpha)}_{1\ (\beta),\lambda}\rho^{\lambda}, \tag{62d}$$

so that, by (15b) with (16a) and by (36) with (32b),

$$\psi^{\mu}=\varphi^{\mu}+\chi^{\mu}+\tfrac{1}{2}\zeta^{\mu}, \tag{63a}$$

$$\theta^{(\alpha)}_{1\ (\beta)}=\omega^{(\alpha)}_{1\ (\beta)}+\gamma^{(\alpha)}_{1\ (\beta)}+\tfrac{1}{2}\vartheta^{(\alpha)}_{1\ (\beta)}. \tag{63b}$$

Expanding all exponentials and neglecting $\mathscr{O}_3$, we find

$$e^{ia}e^{ib}=\exp\{i(a+b)+\tfrac{1}{2}[ia;ib]\}. \tag{64}$$

Therefore, by the linearity of T_1,

$$e^{iT_{12}(\xi,\epsilon)}e^{iT_{12}(\eta,v)}\equiv e^{iT_1(\varphi,\omega)}e^{iT_1(\chi,\gamma)}$$
$$=\exp\{iT_1(\varphi+\chi,\ \omega+\gamma)$$
$$+\tfrac{1}{2}[iT_1(\varphi,\omega);iT_1(\chi,\gamma)]\} \tag{65a}$$

are the left-hand factors in (61a) for the two-step transformation, while the left-hand factor in (61b) for the one-step transformation is, according to (63a) and (63b),

$$e^{iT_{12}(\rho,\nu)}\equiv e^{iT_1(\psi,\theta)}$$
$$=\exp\{iT_1(\varphi+\chi,\ \omega+\gamma)+\tfrac{1}{2}iT_1(\zeta,\vartheta)\}. \tag{65b}$$

Equality of (61a) and (61b) then requires that in $e^{iT}Fe^{-iT}$ it should make no difference whether iT contains the term $\tfrac{1}{2}[iT_1(\varphi,\omega);iT_1(\chi,\gamma)]$ of (65a), or whether it contains the term $\tfrac{1}{2}iT_1(\zeta,\vartheta)$ of (65b). Yet, these terms need not be equal. It suffices that they have equal commutators with the field variables F to which Eqs. (1), (61a), and (61b) are applied; i.e.,

$$[[iT_1(\varphi,\omega);iT_1(\chi,\gamma)];F]$$
$$\text{should equal } [iT_1(\zeta,\vartheta);F]. \tag{66}$$

We calculate these two expressions by using several times Eq. (57b) with Eq. (60). As in (66) we neglect $\mathscr{O}_3$, the term with $\bar{\delta}_1\bar{\delta}_1F$ in (60) will not contribute. Thus, the first one of the two expressions (66) is equal to

$$[iT_1(\varphi,\omega);[iT_1(\chi,\gamma);F]]-[iT_1(\chi,\gamma);[iT_1(\varphi,\omega);F]]$$
$$=[iT_1(\varphi,\omega);\bar{\delta}_{(\eta,v)}F]-[iT_1(\chi,\gamma);\bar{\delta}_{(\xi,\epsilon)}F]$$
$$=\{\bar{\delta}_{(\xi,\epsilon)}(F''(P'')-F(P))\}$$
$$-\{\bar{\delta}_{(\eta,v)}(F'(P')-F(P))\}$$
(see Fig. 1)
$$=\{(F''''-F'')-(F'-F)\}$$
$$-\{(F'''-F')-(F''-F)\}$$
$$=F''''(P'''')-F'''(P''')\quad(\text{neglecting } \mathscr{O}_3). \tag{67a}$$

On the second and third lines from the bottom of (67a), we wrote F for $F(P)$, F' for $F'(P')$, and so on. Similarly, as ζ and ϑ are $\mathscr{O}_2$, the second expression (66) differs only by $\mathscr{O}_3$ terms from

$$[iT_{12}(\zeta,\vartheta);F]=[iT_{12}(\zeta,\vartheta);F''']$$
$$=\bar{\delta}_{(\zeta,\vartheta)}F'''=F''''(P'''')-F'''(P'''). \tag{67b}$$

Comparison with (67a) then shows that, in second order, (61a) and (61b) are equal.

XIX. WARNING AGAINST AN OBJECTION

The reader should be warned that the result (57a) (of which the proof was postponed to our next paper[26]), and therefore also our conclusion (57b), is valid only if F is one of the field variables. It is not generally valid for a product of a q-number F with a c-number field f, because there is no guarantee that $f(x)$ would satisfy $\bar{\delta}f(x)=0$, while a c-number $f(x)$ would commute with iT.

As an example, consider again the formula $x''^{\mu}=x^{\mu}+\eta^{\mu}$ for the transformation $\Sigma\to\Sigma''$. As η^{μ} and $\Delta_{(\eta)}x^{\mu}(PP'')=-\eta^{\mu}+\mathscr{O}_2$ both are four-vectors, it is seen that, under a transformation with descriptor ξ^{μ} (from Σ to Σ' and from Σ'' to Σ''''), we have[37]

$$\bar{\delta}_{(\xi)}\,\eta^{\mu} = \zeta^{\mu}\,, \tag{68a}$$

$$\bar{\delta}_{(\xi)}\,\Delta x^{\mu} = -\,\zeta^{\mu}\,. \tag{68b}$$

Yet, both η^μ and Δx^μ are c-numbers, and therefore commute with iT_1. Therefore, with f a c-number field and neglecting $\mathscr{O}_2$, we should expect

$$[iT_1; fF] = f[iT_1; F] = f\bar{\delta}_1 F \neq \bar{\delta}_1(fF)\,. \tag{69}$$

In particular, in the first expression (66), we therefore should not make the error of equating $[iT_1(\varphi,\omega); iT_1(\chi,\gamma)]$ to $\delta_{1(\varphi,\omega)} \times iT_1(\chi,\gamma)$, because also $T_1(\chi,\gamma)$ contains c-number fields χ^μ and $\gamma_{1(\alpha)(\beta)}$.

XX. SUMMARY

After an introduction explaining the purpose of this paper and a paper to follow,[26] we showed in the present paper that, for infinitesimal coordinate transformations possibly preceded[32] by infinitesimal Lorentz transformations of the tetrad field, the second-order-infinitesimal contributions to local and to substantial transformations of tensors and undors[25] can be expressed in terms of the first-order-infinitesimal transformations. [See Eqs. (48a)–(48c) and (54a)–(54c).] When the Lagrangian is altered so as to allow canonical quantization, for fermions it is not allowed to quantize the undor fields ψ and $\psi^\dagger$ themselves canonically, nor the tetrad field with its ordinary canonical conjugates. These fields will satisfy Dirac's modified commutation relations, as we canonically quantize the canonized fermion fields $\tilde{\psi}$ and $\tilde{\psi}^\dagger$, and the tetrad field with its canonized canonical conjugates derived from the canonized Lagrangian.[17,20,21] In the next paper,[26] we will give explicit formulas for all of this, and we will prove that the expression T_1 [(55a) or (55b)] is *by this modified quantization* the first-order generator of infinitesimal coordinate and tetrad transformations, provided these transformations leave the Lagrangian function invariant. We also prove there the conservation and invariance of T_{12}, obtained from T_1 by the substitutions (52a)–(52c) for the descriptors of the transformations. In the present paper, we show that this T_{12} then with second-order-infinitesimal accuracy generates by Eq. (1) the substantial infinitesimal transformations of the field variables, and we explain why this guarantees the existence of a generator T also for finite transformations. This, in turn, guarantees the covariance of Dirac's modified commutation relations under the transformations here allowed. Because of the alteration of the Lagrangian, these are affine transformations only. Yet, it may be argued[26] that this does not necessarily preclude general coordinate covariance of this (modified) canonical quantization procedure.

APPENDIX: SOLVING EQ. (28b) FOR $\epsilon_2^{(\alpha)}{}_{(\beta)}$ IN TERMS OF $\epsilon_1^{(\alpha)}{}_{(\beta)}$

For simplicity we write in this Appendix α, β, ..., for (α), (β), ..., so that by $g_{\alpha\beta}$ we will mean the Minkowski metric.

For a Lorentz transformation between Lorentz frames, let $\epsilon^\alpha{}_\beta = \partial x'^\alpha/\partial x^\beta - \delta^\alpha_\beta$. This transformation leaves $g^{\alpha\beta}$ unchanged, so

$$g^{\alpha\beta} = (\delta^\alpha_\gamma + \epsilon^\alpha{}_\gamma)(\delta^\beta_\delta + \epsilon^\beta{}_\delta)\, g^{\gamma\delta}$$
$$= g^{\alpha\beta} + \epsilon^{\alpha\beta} + \epsilon^{\beta\alpha} + \epsilon^{\alpha\delta}\epsilon^\beta{}_\delta\,.$$

Let $\epsilon_1^{\alpha\beta} = \frac{1}{2}(\epsilon^{\alpha\beta} - \epsilon^{\beta\alpha})$ and $\epsilon_2^{\alpha\beta} = \frac{1}{2}(\epsilon^{\alpha\beta} + \epsilon^{\beta\alpha})$, so that $\epsilon_2^{\alpha\beta} = -\frac{1}{2}\epsilon^{\alpha\delta}\epsilon^\beta{}_\delta = -\frac{1}{2}\epsilon^\alpha{}_\delta(\epsilon_1^{\beta\delta} + \epsilon_2^{\beta\delta})$, and therefore

$$\epsilon_2{}^\alpha{}_\beta = \tfrac{1}{2}(\epsilon_1{}^\alpha{}_\delta + \epsilon_2{}^\alpha{}_\delta)(\epsilon_1{}^\delta{}_\beta - \epsilon_2{}^\delta{}_\beta)\,. \tag{A1}$$

We assume the ϵ_1 to be "small of first order." Then we see from (A1) that the matrix ϵ_2 is small of second order.

We expand the matrix ϵ_2 as a sum of chain products of matrices ϵ_1, by iterating the substitution of (A1) into its own right-hand member. In each substitution, the leading term in ϵ_2 is $\frac{1}{2}\epsilon_1\epsilon_1$, which does not require further substitution. A term obtained after n substitutions will be a polynomial at least of order $(n+2)$, so that, for obtaining an expansion up to the Nth order in ϵ_1, we need no more than $(N-2)$ substitutions for obtaining any of the terms. Up to Nth order, the resulting terms then are products of factors ϵ_1 only. Therefore, in the resulting terms up to this order, the order of sequence of the factors does no longer matter, and the contributions from the terms $-\frac{1}{2}\epsilon_1\epsilon_2$ and $+\frac{1}{2}\epsilon_2\epsilon_1$ in (A1) will cancel out. Therefore, we obtain the same expansion up to Nth order, if we replace (A1) by

$$\epsilon_2 = \tfrac{1}{2}\epsilon_1\epsilon_1 - \tfrac{1}{2}\epsilon_2\epsilon_2\,. \tag{A2}$$

[For instance, after zero substitutions, up to second order in ϵ_1, we obtain just the first term in (A1), and the omission of the third-order terms is irrelevant in the expansion up to second order here considered.]

Compare (A2) with the equation $y = \frac{1}{2}(x^2 - y^2)$, which is solved for y by $y = (1+x^2)^{1/2} - 1 = \Sigma_{n=1}^{\infty}\binom{1/2}{n}x^{2n}$. Similarly, (A2) may be verified to have the solution already mentioned in Eq. (28d).

[1] W. Heisenberg and W. Pauli, Z. Phys. **56**, 1 (1929) [correction: L. Rosenfeld, Z. Phys. **63**, 574 (1930)].

[2] W. Heisenberg and W. Pauli, Z. Phys. **59**, 168 (1929).

[3] L. Corwin, Y. Ne'eman, and S. Sternberg, Rev. Mod. Phys. **47**, 573 (1975), definitions on p. 574.

[4] F. J. Belinfante, Physica **7**, 765 (1940).

[5] L. Rosenfeld, Ann. Phys. Leipzig **5**, 113 (1930); Ann. Inst. H. Poincaré **2**, 25 (1932).

[6] P. G. Bergmann, Phys. Rev. **75**, 680 (1949); J. L. Anderson and P. G. Bergmann, Phys. Rev. **83**, 1018 (1951).

[7] E. Fermi, Rev. Mod. Phys. **4**, 87 (1932).

[8] J. Schwinger, Phys. Rev. **74**, 1439 (1948), Eq. (2.12).

[9] F. J. Belinfante, Phys. Rev. **76**, 226 (1949).

[10] S. N. Gupta, Proc. Phys. Soc. London Sect. A **63**, 681 (1950); A **66**, 129 (1953).

[11] F. J. Belinfante, Physica **12**, 17 (1946), p. 27 (top half) and pp. 31–32.

[12] J. Schwinger, Phys. Rev. **75**, 651 (1949).

[13] Reference 8, Eq. (2.32).

[14] J. C. Garrison, Phys. Rev. **129**, 1424 (1963).

[15] M. C. V. Bergère, "The interaction picture for gravity interacting with fermions", Ph. D. thesis, Purdue University, 1970 (University Microfilms No. 70-18, 595, Ann Arbor, 1970).

[16] Th. De Donder, *La Gravifique Einsteinienne* (Gauthier-Villars, Paris, 1921), p. 40. See also Th. De Donder, *Mathematical Theory of Relativity* (M. I. T., Cambridge, MA, 1927), pp. 50–51, Eqs. (31) and (32).

[17] F. J. Belinfante, D. I. Caplan, and W. L. Kennedy, Rev. Mod. Phys. **29**, 518 (1957).

[18] P. A. M. Dirac, Phys. Rev. **114**, 924 (1959).

[19] R. Arnowitt, S. Deser, and C. W. Misner, Phys. Rev. **117**, 1595 (1960); **118**, 1100 (1960).

[20] P. A. M. Dirac, Can. J. Math. **2**, 129 (1950).

[21] B. S. DeWitt and C. M. DeWitt, Phys. Rev. **87**, 116 (1952).

[22] V. Fock, Z. Phys. **75**, 622 (1932); **76**, 852 (1932).

[23] See pp. 328–334 of H. A. Kramers, *Quantum Mechanics* (Interscience, New York, 1957).

[24]P. Jordan and E. Wigner, Z. Phys. **47**, 631 (1928).

[25]F. J. Belinfante, Physica **6**, 849 (1939).

[26]F. J. Belinfante, J. Math. Phys. **26**, 2827 (1985).

[27]See, for instance, p. 10 of J. L. Anderson, *Principles of Relativity Physics* (Academic, New York, 1967).

[28]F. J. Belinfante, J. Math. Phys. **26**, 2836 (1985).

[29]B. L. van der Waerden, Göttinger Nachr. **1929**, 100. See also B. L. van der Waerden, *Die Gruppentheoretische Methode in der Quanten Mechanik* (Springer, Berlin, 1932), Sec. 20.

[30]These may include gravitational field variables, like the metric field components when fermions are absent, or tetrad field components when fermions are present.

[31]F. J. Belinfante, Physica **7**, 305 (1940).

[32]For simplicity we assume that in "simultaneous" coordinate and tetrad transformations the tetrad transformation precedes the coordinate transformation. [The last line in Eq. (32b) shows that the order of sequence makes a second-order difference.]

[33]Note that $\varphi^{\mu}(P) = \frac{1}{2}\{\xi^{\mu}(P) + \xi^{\mu}(P')\}$ and $\omega_{1(\alpha)(\beta)}(P) = \frac{1}{2}\{\epsilon_{1(\alpha)(\beta)}(P) + \epsilon_{1(\alpha)(\beta)}(P')\}$ are the descriptors of coordinate transformation and of tetrad Lorentz transformation at $x_P + \frac{1}{2}\Delta x_P$, halfway between the points P and P' in $\bar{\delta}F(P) = F'(P') - F(P)$.

[34]Compare, for instance, $\int C^{\rho}\, d\sigma_{\rho}$ with C^{ρ} from Eq. (3.4) of P. G. Bergmann and A. Komar, "The phase-space formulation of general relativity and approaches toward its canonical quantization,"[35] where Q^{ρ} from their Eq. (3.2) is easily seen to be $= \mathscr{L}\Delta x^{\rho} = -\mathscr{L}\xi^{\rho}$ in their first-order-infinitesimal approximation.

[35]Pages 227–254 of Vol. 1 of *General Relativity and Gravitation One Hundred Years After the Birth of A. Einstein*, edited by A. Held (Plenum, New York, 1980).

[36]That is, without neglecting any second-order infinitesimals.

[37]From $\delta_{(\xi)}\eta^{\mu} = \xi^{\mu}_{,\nu}\eta^{\nu}$ it follows that $\bar{\delta}_{(\xi)}\eta^{\mu} = \delta_{(\xi)}\eta^{\mu} - \eta^{\mu}_{,\nu}\xi^{\nu} = \zeta^{\mu}$ by (16a). Also, $\bar{\delta}_{(\xi)}\Delta x^{\mu} \equiv \bar{\delta}_{(\xi)}\{x^{\mu}(P'') - x^{\mu}(P)\} = \{x'^{\mu}(P'''') - x'^{\mu}(P')\} - \{x^{\mu}(P'') - x^{\mu}(P)\}$, where, in terms of coordinates and descriptors in Σ at P, $x'^{\mu}(P'''') = x^{\mu} - \eta^{\mu} - \xi^{\mu}_{,\lambda}\eta^{\lambda} + \eta^{\mu}_{,\lambda}\xi^{\lambda} + \eta^{\mu}_{,\lambda}\eta^{\lambda}$ and $x^{\mu}(P'') = x^{\mu} - \eta^{\mu} + \eta^{\mu}_{,\lambda}\eta^{\lambda}$ and $x'^{\mu}(P') = x^{\mu}(P)$, so that (68b) follows with the help of (16a).

On the covariance of Dirac's modified canonical commutation relations for fermions interacting with gravity and bosons

Frederik J. Belinfante[a)]
Department of Physics, Purdue University, Lafayette, Indiana 47907

(Received 8 February 1984; accepted for publication 13 June 1985)

For fermions in curved space-time, where $i\hbar\psi^\dagger$ is no longer the canonical conjugate of ψ, canonical quantization according to Dirac requires use of "modified" graded commutation relations, in which the canonical conjugates to the tetrad field components are no longer commutative with each other or with the fermion field. As shown earlier by the DeWitts, these modified commutation relations may be understood as a canonical quantization of new field variables that can conventionally be interpreted in terms of creation and annihilation operators. Because of the horrible transformation properties of these new variables, covariance of this quantization of fermion fields and of the gravitational field with which the fermions interact is best proved in terms of the old variables. Like canonical quantization, also this modified quantization requires alteration of the theory destroying general invariance of the Lagrangian. As discussed in a preceding paper, covariance of this modified quantization under a group of finite transformations will follow, under the conditions that (1) it can be proved for a generator T_1, linear in infinitesimal descriptors of the transformations, that by these modified commutation relations it will generate the substantial variations of the canonical field variables, while (2) the second-order generator T_{12}, constructed from T_1 as previously discussed, will be conserved and invariant. It is here proved that these two conditions are satisfied, provided that the transformations leave the Lagrangian function L invariant. This restriction limits the proof of covariance of quantization to affine transformations only. It is discussed why this yet leaves the possibility of general covariance of the "physical part" of this quantization procedure.

I. INTRODUCTION

In its simplest form, canonical quantization uses the "graded" commutation relations[1]

$$\left.\begin{aligned}[\mathrm{q}_i(x);\mathrm{q}_j(y)] &= [\mathrm{p}^i(x);\mathrm{p}^j(y)] = 0,\\ [\mathrm{q}_i(x);\mathrm{p}^j(y)] &= i\hbar\delta_i^j\delta_3(\mathbf{x}-\mathbf{y}),\end{aligned}\right\}\quad \text{for } x^0=y^0. \tag{1}$$

Here, $c\mathrm{p}^i = \partial\mathscr{L}/\partial \mathrm{q}_{i,0}$. The [;] denote "graded" commutators. By graded we here mean that between two fermion field components they are *anti*commutators.[2]

Canonical quantization, however, seldom is as simple as (1). We will here first discuss some deviations from (1).

II. DERIVED VARIABLES

There often exist "field identities" (field equations not containing any time derivatives), which allow us to express certain field variables Q^d (called derived variables) as functions of other field variables and their spatial derivatives, while the $Q^d_{,0} \equiv \partial_0 Q^d$ do not occur in the Lagrangian density $\mathscr{L}$. In particular, these Q^d arise when a boson theory is derived from a first-order Lagrangian (linear in the $\mathrm{q}_{,\mu}$). Examples are $\mathbf{B}$ (= curl $\mathbf{A}$) in quantum electrodynamics, and, in neutral vector meson theory, Φ ($=\kappa^{-2}[g\psi^\dagger\psi - \mathrm{div}\,\mathbf{E}]$). In such cases, Eqs. (1) do not apply directly to the Q^d, but apply only to the canonical field coordinates q_c (of which the time derivatives occur in $\mathscr{L}$) and their canonically conjugated field momenta p^c. The commutation relations for the Q^d then follow by the field identities from those for the q_c and p^c. The covariance of this method of quantization under Poincaré transformations has been proved.[3]

III. MODIFIED COMMUTATION RELATIONS

The above field identities are examples of what Dirac[4] has called "second-class constraints." Another type of second-class constraints occurs in fermion theory in curved space, where in the Lagrangian density $\mathscr{L}$ the special-relativistic term $i\hbar c\,\psi^\dagger\psi_{,0}$ is replaced by[5,6]

$$-\tfrac{1}{2}\hbar c\not{h}\,h^{\,0}_{(\alpha)}\{\bar\psi\gamma^{(\alpha)}\psi_{,0} - \bar\psi_{,0}\gamma^{(\alpha)}\psi\}, \tag{2}$$

where the $h^{\,\mu}_{(\alpha)}$ are tetrad components and $\not{h}\equiv\sqrt{-g}$ = the determinant of the $h_\nu^{(\beta)}$, while the notation of the Dirac matrices is as in Sec. XIV of a preceding paper.[7] Then, the canonical conjugate to ψ is no longer $i\hbar\psi^\dagger$. Instead, both ψ and $\bar\psi$ according to (2) would be canonical field coordinates q_c (as their time derivatives appear in $\mathscr{L}$), and the corresponding $\mathrm{p}^c \equiv \partial^R\mathscr{L}/c\,\partial \mathrm{q}_{c,0}$ then are functions of the q_c.[8] In particular, now[8]

$$\begin{aligned}\mathrm{p}^\psi &= -\tfrac{1}{2}\hbar\not{h}\bar\psi\gamma^0,\\ \mathrm{p}^{\bar\psi} &\equiv \frac{\partial^R\mathscr{L}}{c\,\partial\bar\psi_{,0}} = -\frac{\partial^L\mathscr{L}}{c\,\partial\bar\psi_{,0}} = -\tfrac{1}{2}\hbar\not{h}\gamma^0\psi.\end{aligned} \tag{3}$$

To this case, Belinfante's method of quantizing derived variables[3] is not applicable, but one can use here Dirac's method of quantization,[4] equating the graded commutators of field variables to $i\hbar\times$ the graded "modified Poisson brackets" introduced by Dirac. The resulting commutation relations then are no longer all of the form (1). For instance, Dirac's modified commutation relations provide the anticommutation relations

$$\begin{aligned}[\psi_A(x);\bar\psi^B(y)] &= -(i\kappa/\not{h})(\gamma^0)_A{}^B\delta_3(\mathbf{x}-\mathbf{y}),\\ [\psi_A;\psi_B] &= [\bar\psi^A;\bar\psi^B] = 0;\ \text{all for } x^0=y^0,\end{aligned} \tag{4}$$

a) Present address: P. O. Box 901, Gresham, Oregon 97030.

where

$$k \equiv 1/g^{00}, \quad \text{so}, \quad (\gamma^0)^{-1} = k\gamma^0. \tag{5}$$

By (3), this gives $[\psi(x);p^{\psi}(y)]$ at $x^0 = y^0$ only half of the conventional value (1). A more spectacular difference from (1) is that the field momenta canonically conjugate to the tetrad components now no longer commute with each other or with the ψ or $\bar\psi$ fields. (See below.)

IV. CANONIZATION OF FIELD VARIABLES

Dirac's modified commutation relations may also be derived by the method of "canonization."[5] It amounts to performing a transformation proposed by DeWitt and DeWitt,[9]

$$\tilde\psi = \psi\Theta, \quad \Theta = \Theta^\dagger, \quad \tilde\psi^\dagger = \bar\psi\beta\Theta, \tag{6}$$

with

$$\begin{aligned}\Theta^2 &= i g\beta\gamma^0 = i g h^{\,0}_{(\alpha)}\,\beta\gamma^{(\alpha)},\\ \Theta^{-2} &= -i g^{-1}(\gamma^0)^{-1}\beta = (k/g)\gamma^0\gamma^{(0)},\end{aligned} \tag{7}$$

so that, according to Belinfante, Caplan, and Kennedy,[5]

$$\Theta = \frac{g^{1/2}\{1-(-k)^{1/2}\gamma^{(0)}\gamma^0\}}{(-k)^{1/4}2^{1/2}[1+(-k)^{1/2}h^{\,0}_{(0)}]^{1/2}}, \tag{8a}$$

$$\Theta^{-1} = \frac{(-k)^{1/4}\{1-(-k)^{1/2}\gamma^0\gamma^{(0)}\}}{2^{1/2}g^{1/2}[1+(-k)^{1/2}h^{\,0}_{(0)}]^{1/2}}. \tag{8b}$$

The beauty of the transformation (6) is that it gives (2) the form

$$\tfrac{1}{2}i\hbar c\{\psi^\dagger\Theta^2\psi_{,0} - \psi^\dagger_{,0}\Theta^2\psi\}. \tag{9}$$

In the Lagrangian $\Lambda = \int \mathscr{L}\, d^4x/c$ we now perform an integration by parts with respect to time, as we change $\mathscr{L}$ into $\tilde{\mathscr{L}}$ by adding the amount

$$\tilde{\mathscr{L}} - \mathscr{L} \equiv \tfrac{1}{2}i\hbar c(\psi^\dagger\Theta^2\psi)_{,0} = \tfrac{1}{2}i\hbar c(\tilde\psi^\dagger\tilde\psi)_{,0}, \tag{10}$$

which combines with (9) to

$$\begin{aligned}&\tfrac{1}{2}i\hbar c\{\tilde\psi^\dagger\tilde\psi_{,0} + \tfrac{1}{2}\psi^\dagger(\Theta_{,0}\Theta - \Theta\Theta_{,0})\psi\}\\ &\quad= \tfrac{1}{2}i\hbar c\{\tilde\psi^\dagger\tilde\psi_{,0} + \tfrac{1}{2}\tilde\psi^\dagger(\Theta^{-1}\Theta_{,0} - \Theta_{,0}\Theta^{-1})\tilde\psi\}.\end{aligned} \tag{11}$$

After this change of fermion field variables and of Lagrangian density, the canonical conjugate to $\tilde\psi$ is now $i\hbar\tilde\psi^\dagger$, and therefore $\tilde\psi$ can be interpreted in terms of electron annihilation and positon creation operators in the way ψ was interpreted in flat-space theory, provided we quantize $\tilde\psi$ in the conventional way by (1). However, since Θ by (8) depends upon the tetrad field components q_i both directly and through the metric, the new canonical field momenta $\tilde p^i = (\partial\tilde{\mathscr{L}}/c\,\partial q_{i,0})_{\tilde\psi}$ now differ from the old $p^i = (\partial\mathscr{L}/c\,\partial q_{i,0})_\psi$ by

$$\tilde p^i - p^i = \frac{\partial^R\mathscr{L}}{c\,\partial\psi_{,0}}\left(\frac{\partial\psi_{,0}}{\partial q_{i,0}}\right)_{\tilde\psi} + \left(\frac{\partial\psi^\dagger_{,0}}{\partial q_{i,0}}\right)_{\tilde\psi^\dagger}\frac{\partial^L\mathscr{L}}{c\,\partial\psi^\dagger_{,0}}, \tag{12}$$

so, by (9),

$$\begin{aligned}\tilde p^i - p^i &= \tfrac{1}{2}i\hbar\left\{\psi^\dagger\Theta^2\frac{\partial(\Theta^{-1})}{\partial q_i}\tilde\psi - \tilde\psi^\dagger\frac{\partial(\Theta^{-1})}{\partial q_i}\Theta^2\psi\right\}\\ &= \tfrac{1}{2}i\hbar\tilde\psi^\dagger\left\{\Theta^{-1}\frac{\partial\Theta}{\partial q_i} - \frac{\partial\Theta}{\partial q_i}\Theta^{-1}\right\}\tilde\psi\\ &= \tfrac{1}{2}i\hbar\psi^\dagger\left\{\frac{\partial\Theta}{\partial q_i}\Theta - \Theta\frac{\partial\Theta}{\partial q_i}\right\}\psi.\end{aligned} \tag{13}$$

Note that from (10) there is no contribution $\{\partial(\tilde{\mathscr{L}} - \mathscr{L})/\partial q_{i,0}\}_{\tilde\psi}$ to (12).

If we start from Dirac's modified quantization, it is found[9] that, after the canonization, the conventional graded commutation relations (1) will be valid for $\tilde\psi$ with canonical conjugate $i\hbar\tilde\psi^\dagger$, and for q_i with $\tilde p^i$. It is simpler to invert the reasoning, and say that ordinary canonical quantization (1) is allowed after canonization of the fields. From this, we then can derive Dirac's modified graded commutation relations for the original ψ, $\bar\psi$, q_i, and p^i. We thus first find

$$[p^i;\Theta^{-1}] = \frac{\hbar}{i}\frac{\partial\Theta^{-1}}{\partial q_i}\delta_3 = i\hbar\Theta^{-1}\frac{\partial\Theta}{\partial q_i}\Theta^{-1}\delta_3, \tag{14}$$

where we omitted the arguments x and y (with $x^0 = y^0$), and where δ_3 is short for $\delta_3(\mathbf{x}-\mathbf{y})$. Thence, in similar notation,

$$\begin{aligned}[p^i(x);\psi(y)] &= [\tilde p^i;\Theta^{-1}\tilde\psi] + [\Theta^{-1}\tilde\psi;(\tilde p^i - p^i)]\\ &= i\hbar\left\{\Theta^{-1}\frac{\partial\Theta}{\partial q_i}\psi + \tfrac{1}{2}\Theta^{-1}\left[\Theta^{-1}\frac{\partial\Theta}{\partial q_i} - \frac{\partial\Theta}{\partial q_i}\Theta^{-1}\right]\Theta\psi\right\}\delta_3\\ &= \tfrac{1}{2}i\hbar\left\{\Theta^{-2}\frac{\partial(\Theta^2)}{\partial q_i}\right\}\psi\delta_3\\ &= \tfrac{1}{2}i\hbar\frac{k\gamma^0}{g}\frac{\partial(g\gamma^0)}{\partial q_i}\psi\delta_3(\mathbf{x}-\mathbf{y}), \text{ for } x^0 = y^0,\end{aligned} \tag{15a}$$

on account of (7). Similarly, for $x^0 = y^0$,

$$\begin{aligned}[p^i(x);\bar\psi(y)] &= \tfrac{1}{2}i\hbar\bar\psi\frac{\partial(g\gamma^0)}{\partial q_i}\frac{k\gamma^0}{g}\delta_3(\mathbf{x}-\mathbf{y})\\ &= \tfrac{1}{2}i\hbar\psi^\dagger\frac{\partial(\Theta^2)}{\partial q_i}\Theta^{-2}\beta\delta_3(\mathbf{x}-\mathbf{y})\end{aligned} \tag{15b}$$

and

$$\begin{aligned}&[p^i(x);p^j(y)]\\ &\quad= \tfrac{1}{4}\hbar^2\psi^\dagger\left[\frac{\partial(\Theta^2)}{\partial q_i}\Theta^{-2}\frac{\partial(\Theta^2)}{\partial q_j} - (i\rightleftarrows j)\right]\psi\delta_3(\mathbf{x}-\mathbf{y})\\ &\quad= \tfrac{1}{4}i\hbar^2\bar\psi\left[\frac{\partial(g\gamma^0)}{\partial q_i}\frac{k\gamma^0}{g}\frac{\partial(g\gamma^0)}{\partial q_j} - (i\rightleftarrows j)\right]\psi\delta_3(\mathbf{x}-\mathbf{y}).\end{aligned} \tag{16}$$

Also (4) follows similarly by (7) from

$$[\psi(x);\bar\psi(y)] = \Theta^{-2}\beta\delta_3(\mathbf{x}-\mathbf{y}), \quad \text{for } x^0 = y^0. \tag{17}$$

For boson fields q_i, including tetrad fields, we find, unchanged

$$[q_i(x);p^j(y)] = [q_i(x);\tilde p^j(y)] = i\hbar\delta^j_i\delta_3(\mathbf{x}-\mathbf{y}), \quad \text{for } x^0 = y^0. \tag{18}$$

As gravitational field components q_i we may use in the above the tetrad field $h^{\,\mu}_{(\alpha)}$, or its inverse matrix $h^{\,(\alpha)}_{\mu}$. The forms taken by Eqs. (15a), (15b), and (16) for the former choice have been published before.[5] For the latter choice, (16) becomes

$$[p^{\,\mu}_{(\alpha)}(x);p^{\,\nu}_{(\beta)}(y)] = -2i\hbar^2 g k h^{\,0}_{(\alpha)}h^{\,0}_{(\beta)}\bar\psi\Sigma^{\mu0\nu}\psi\delta_3(\mathbf{x}-\mathbf{y}), \tag{19}$$

where we introduced the abbreviation

$$\Sigma^{\alpha\epsilon\beta} \equiv \tfrac{1}{8}(\gamma^{\alpha}\gamma^{\epsilon}\gamma^{\beta} - \gamma^{\beta}\gamma^{\epsilon}\gamma^{\alpha}). \tag{20}$$

The proof of covariance of these commutation relations will start in Sec. VII. For this proof, it is simpler to use the original field variables ψ, $\overline{\psi}$, φ_i, with $\not{p}^i$ and the above, more complicated, modified commutation relations, than to use the canonized field variables $\tilde{\psi}$, $\tilde{\psi}^\dagger$, and φ_i with $\tilde{\not{p}}^i$ and the simpler commutation relations (1), because under coordinate transformations the canonized field variables transform in a horrible nonlinear way. See Eqs. (6) and (13) with (8a) and (8b).

V. THE ALTERED THEORY OF FIELDS

The first-class[4] constraints, which the general coordinate invariance of the Lagrangian causes[10–12] in Einstein's gravitational theory, give rise to well-known complications with the quantization, comparable to (but worse than) complications met in the quantization of Maxwell's theory because of its general gauge invariance. In the latter case, the theory can be quantized "dynamically"[7] in the transverse gauge,[13] but attempts at dynamical quantization of the theory of gravity[14] interacting with matter including fermion fields meets with unsolved problems. Therefore, we consider here, as previously,[7] only "Fermi-type" quantization of the so-called "altered" theory of fields.

This alternative method of quantization was first applied in 1929–1930 by Fermi[15] to flat-space quantum electrodynamics. Though originally expressed in terms of Fourier components of the fields, it can also be expressed in terms of the fields as functions of space-time coordinates.[16] This method removes the cause of first-class constraints by altering the Lagrangian, by adding terms to it that have only *restricted* gauge invariance. In quantum electrodynamics, the altered Lagrangian is invariant only under those gauge transformations that keep $S \equiv \text{div}\,\mathbf{A} + \partial\Phi/c\,\partial t$ invariant. The Lagrangian of the altered theory of gravitation is invariant under *affine* coordinate transformations only, so as to make

$$S^{\mu} \equiv \mathfrak{g}^{\mu\nu}{}_{,\nu} \equiv \partial_{\nu}(g^{\mu\nu}\sqrt{-g}) \tag{21a}$$

transform as a four-vector. The altered or "muddified"[17] Lagrangian functions L and densities $\mathscr{L}$ then differ from the original ones by "mud" terms quadratic or bilinear in the quantities S and/or S^{μ} or the like. This would change expressions for physical quantities by terms at least linear in S and the like, and thus we obtain a theory which describes nature only in those states, in which the quantities S and S^{μ} and the like have the value zero, even though as q numbers they cannot vanish, on account of the commutation relations which we assume.

The altered field equations, which follow from the altered Lagrangian, now allow us to solve for all field velocities $\mathrm{q}_{i,0}$ in terms of the (altered) field momenta p^{j}. As first-class constraints now are absent, we can now quantize canonically by (1), provided that in the presence of fermion fields we first canonize, which is equivalent to using Dirac's modified commutation relations (15)–(18) for the fields originally occurring in the altered Lagrangian.

The resulting theory then allows for more different quantum states than can occur in nature, so it gives us too large a "Hilbert space." (The quotation marks here refer to the lack of positive definiteness of the metric of this function space, as discussed by Gupta.[18]) This oversized space contains states that describe the presence of free "unphysical" photons and gravitons, created and annihilated by "nondynamical" field components.[7] In the "physical part" of this oversized Hilbert space, only photons created by *transverse* fields can occur as free particles, and the nonphysical photons can occur only virtually in certain combinations that describe the effects of the Coulomb interaction between charged particles. Similarly for gravitons, where combinations of virtual nonphysical gravitons should be responsible for gravity as Newton knew it. The physical part of Hilbert space is selected from the oversized space by the auxiliary conditions

$$S=0, \quad S^{\mu}=0, \tag{21b}$$

that is, by the quantized version of the Lorentz condition and of the De Donder condition.[19] This Hilbert subspace (21b) should be equivalent to the Hilbert space that would be defined by successful dynamical quantization.

VI. TETRADS OR METRIC?

The Dirac matrices γ^{μ} in the general-relativistic theory of fermions are expressible in terms of the constant numerical Dirac matrices $\gamma^{(\alpha)}$ of the Lorentz-covariant flat-space theory and the tetrad field $h^{\ \mu}_{(\alpha)}$ by

$$\gamma^{\mu}(x) = \gamma^{(\alpha)} h^{\ \mu}_{(\alpha)}(x). \tag{22a}$$

The inverse tetrad field $h^{\ (\beta)}_{\nu}$, given by $h^{\ \mu}_{(\alpha)} h^{\ (\alpha)}_{\nu} = \delta^{\mu}_{\nu}$, is related to the metric $g_{\mu\nu}$ by

$$g_{\mu\nu} = \mathring{g}_{(\alpha)(\beta)} h^{\ (\alpha)}_{\mu} h^{\ (\beta)}_{\nu}. \tag{22b}$$

Quantization of the gravitational field now may be accomplished by quantization of the $h^{\ \mu}_{(\alpha)}$ or of the $h^{\ (\alpha)}_{\mu}$. Most of the following will be written in a form independent of this choice.

The main reason for describing the gravitational field by the tetrads instead of by the metric is that, for invariance of fermion theory under local rotations or Lorentz transformations of the tetrads, Fock[20] found it necessary to include, in the expression for the covariant derivative $\nabla_{\mu}\psi$ of the Dirac wave function, terms containing derivatives of the tetrad field, but not expressible in terms of derivatives of the metric field.[21] So, we need both the tetrad field and the metric field, but they are interdependent, and it is easier to express the metric field as a function of the tetrad field than to express the tetrad field in terms of the metric field.

It is true that Pauli has proposed expressing a particular tetrad field in terms of a given metric field by an infinite expansion.[22] Here, we will not use such a complicated way of expressing a particular choice for one variable in terms of other variables.

The tetrad field has 16 components, while the metric field has only ten. The additional six components of the tetrad field correspond to the arbitrary choice of the orientation of the tetrad in each point of space-time. The invariance of fermion theory under changes of this choice, postulated by

Fock, introduces six additional first-class constraints. For dealing with this, the altered theory would require additional terms in the Lagrangian, bilinear or quadratic in the six components of an antisymmetric $S^{(\alpha)(\beta)}$ that is not completely independent of the orientation of the tetrads, and in addition to (21b) there would be the auxiliary condition

$$S^{(\alpha)(\beta)} = 0. \tag{21b$'$}$$

An alternative treatment, by Pellegrini and Plebanski,[23,24] assumes that the tetrad field may be arbitrarily chosen at one time only, and from there on will satisfy field equations of motion of its own, derived from a different addition to the Lagrangian. This ascribes a physical meaning to the tetrad field. As there is now no longer a requirement of invariance under changes of the tetrad field, the Fock terms

$$\tfrac{1}{2}\hbar c[(\nabla_\mu\bar{\psi} - \bar{\psi}_{,\mu})\gamma^\mu\psi - \bar{\psi}\gamma^\mu(\nabla_\mu\psi - \psi_{,\mu})]\sqrt{-g} \tag{23}$$

then may be omitted from $\mathscr{L}$, and the matter Lagrangian then would feature ordinary instead of covariant derivatives not only for boson fields, but also for fermion fields.

Here, we will not specify which one of these two treatments of the tetrad field is used.

VII. COVARIANCE OF COMMUTATION RELATIONS

Under a coordinate transformation $x \to x'$, let field components $F_i(x)$ (which may be q's or p's or their spatial derivatives) be transformed into primed field components $F'_{i'}(x')$. To each point P with coordinates x_P in the unprimed frame Σ, let there belong a point P' in Σ' with coordinates $x'_{P'}$ satisfying $x'_{P'} = x_P$. If we now can prove that, for all field variables F_i at points P on a given spacelike hypersurface σ ($t = \text{const}$ in Σ),

$$F'_{i'}(x'_{P'}) = UF_i(x_P)U^{-1}, \quad \text{for } x'_{P'} = x_P \text{ with } t'_{P'} = t_P = t_\sigma, \tag{24}$$

with the same U for all fields and for all points P on σ, then, for P and Q on σ, and with $f_{jk}(x)$ a polynomial in the q's and p's and their spatial derivatives at x, it will follow by (24) from the assumed validity of the commutation (or anticommutation) relations

$$[F_j(x_P); F_k(x_Q)] = if_{jk}(x_P)\delta_3(\mathbf{x}_P - \mathbf{x}_Q) \tag{25}$$

on σ, that for the corresponding points P' and Q' on σ' ($t' = t_\sigma$ in Σ') the commutation relations

$$[F'_{j'}(x'_{P'}); F'_{k'}(x'_{Q'})] = if'_{j'k'}(x'_{P'})\delta_3(\mathbf{x}'_{P'} - \mathbf{x}'_{Q'}) \tag{25$'$}$$

will be valid.[7]

Therefore, proof of (24) suffices for proving the covariance (or, more precisely, the form invariance) of the commutation relations under the transformations for which (24) is proved.

If this transformation can be of the form $\mathbf{x}' = \mathbf{x}$ with $t' = t + (t_\sigma - t_{\sigma'})$, the fields $F'_{i'}(x'_{P'})$ will be simply the fields $F_i(x_{P'})$ at points P' shifted from P in the time direction, from σ to σ', without change of spatial coordinates. Therefore, we will then have proved not only the covariance of the commutation relations, but also that they remain valid at times different from the time at the hypersurface σ, on which they were originally assumed.

For infinitesimal transformations, U will be close to the unit operator, and it is customary to write

$$U^{\pm 1} = e^{\pm iT} = 1 \pm iT - \tfrac{1}{2}T^2 \mp \cdots. \tag{26}$$

Again, T may be expanded as $T = T_1 + T_2 + \cdots$ in terms of first order and higher orders in the infinitesimal descriptor fields $\xi^\mu = x'^\mu - x^\mu$ that describe the infinitesimal coordinate transformation.

In the preceding[7] paper (I), we discussed whether the conclusion (24), when proved for infinitesimal transformations, will be integrable to finite transformations. If a finite transformation is split up into a succession of N infinitesimal steps, with correspondingly U split up into factors $U_N U_{N-1} \cdots U_3 U_2 U_1$, the question is whether the product U will be independent of the way in which the finite transformation was split up into infinitesimal steps. In our preceding paper (I), we explained how this uniqueness of U is guaranteed by the known fact that the transformations $F_i(x) \to F'_{i'}(x')$ form a *representation* of the group of coordinate transformations $x \to x'$ that we allow,[25] provided that for each infinitesimal step in the transformation we can prove (24) with second-order precision.[7] [See Fig. 2 of (I).] We then showed that for this it suffices to find a first-order generator T_1 with the property

$$\bar{\delta}_1 F = [iT_1; F], \tag{27}$$

where $\bar{\delta}_1 F$ was the part of $\bar{\delta} F$ linear in the infinitesimal descriptors ξ^μ and $\epsilon_{1(\alpha)(\beta)}$. Here, the latter were given with $\mathscr{O}_2$ accuracy as $\epsilon_{1(\alpha)(\beta)} = \epsilon_{(\alpha)(\beta)} - \tfrac{1}{2}\epsilon_{(\alpha)(\gamma)}\epsilon^{(\gamma)}{}_{(\beta)}$. Thence T_{12}, defined as T_1 with ξ^μ and $\epsilon_{1(\alpha)(\beta)}$ replaced by the φ^μ and $\omega_{1(\alpha)(\beta)}$ of[26] Eqs. (I: 52a) and (I: 52b), could with $\mathscr{O}_2$ accuracy serve as the generator T, for which $U = e^{iT}$ would satisfy Eq. (24) above. This then would ensure covariance of the commutation relations used for calculating the right-hand members of Eqs. (27), provided that T_{12} would also be conserved and invariant under the transformations described by the descriptors used in T_1. In Sec. IX, we will prove the latter property of T_{12}, and in Sec. XII we will prove Eq. (27) for all canonical field variables F.

VIII. CHOICE OF THE GENERATOR

By (I: 51b) with (I: 30a), the first-order substantial variations of the field coordinates q_i are

$$\bar{\delta}_1 q_i = \delta_1 q_i - \xi^\mu q_{i,\mu} \equiv (\delta_1 \hat{S}_i{}^j) q_j - \xi^\mu q_{i,\mu}, \tag{28}$$

with

$$\delta_1 \hat{S}_i{}^j \equiv \xi^\mu{}_{,\nu} s_i{}^j{}_\mu{}^\nu + \epsilon_1^{(\alpha)}{}_{(\beta)} \sigma_i{}^j{}_{(\alpha)}{}^{(\beta)}. \tag{29}$$

Here, we use the first-order descriptors ξ^μ and $\epsilon_1^{(\alpha)}{}_{(\beta)}$, which in (I) were defined with second-order precision for making the second-order terms unambiguous.[7]

As (28) according to (I: 3) should be obtainable by

$$\bar{\delta}_1 q_i = [iT_1; q_i], \tag{30}$$

and this should also be true in the absence of fermions, when the modified commutation relations simply are the conventional ones, our first guess at an appropriate T_1 would be

$$T_1 = \hbar^{-1} \int d^3\mathbf{x} \Big\{ p^i [\delta_1 q_i - q_{i,n}\xi^n] - \frac{\xi^0 \mathscr{H}}{c} \Big\}, \tag{31}$$

where

$$\mathscr{H} = c p^i q_{i,0} - \mathscr{L} \tag{32}$$

is the Hamiltonian density. In (31), the p^i for $q_i = \psi$ or $\bar{\psi}$ are given by (3), and we interpret in (31) $p^{\bar{\psi}} \delta_1 \bar{\psi}$ as[6]

$$-(\delta_1 \bar{\psi}) p^{\bar{\psi}} = \tfrac{1}{2}\hbar (\delta_1 \bar{\psi}) \not{\gamma}^0 \psi, \tag{33}$$

with $\delta_1 \bar{\psi}$ from (I: 40'), while in (32) we interpret $c p^{\bar{\psi}} \bar{\psi}_{,0}$ as[6]

$$-c \bar{\psi}_{,0} p^{\bar{\psi}} = \tfrac{1}{2}\hbar c \bar{\psi}_{,0} \not{\gamma}^0 \psi. \tag{34}$$

Below we will verify that the choice (31) for T_1 will provide validity of (30) also when the initial commutation relations are Dirac's modified ones. Moreover we will verify that (27) now also holds for $F = p^i$. But, first we will prove that T_{12} is conserved and invariant.

In the following we will write T for T_{12}, and we will write δ_T for δ_{12}, that is, for the first-order-infinitesimal variations δ_1 of various quantities, with the ξ^μ and $\epsilon_{1(\alpha)(\beta)}$ in δ_1 replaced according to (I: 52a) and (I: 52b) by

$$\varphi^\mu = \xi^\mu - \tfrac{1}{2}\xi^\mu_{,\lambda} \xi^\lambda, \tag{35}$$

$$\omega_{1(\alpha)(\beta)} = \epsilon_{1(\alpha)(\beta)} - \tfrac{1}{2}\epsilon_{1(\alpha)(\beta),\lambda} \xi^\lambda. \tag{36}$$

IX. CONSERVATION OF THE GENERATOR

For proving the conservation of T, it is preferable to write T in more covariant notation as a function of the q_i and $q_{i,\mu}$, rather than as a function of the q_i, the $q_{i,n}$, and the p^i. So, we write

$$T = (\hbar c)^{-1} \int_\sigma d\sigma_\nu \left\{ \frac{\partial^R \mathscr{L}}{\partial q_{i,\nu}} \bar{\delta}_T q_i + \mathscr{L} \varphi^\nu \right\}, \tag{37}$$

where the integral is over the spacelike hypersurface σ ($x^0 = \text{const}$), so that $d\sigma_\nu = \delta^0_\nu\, d^3\mathbf{x}$. In (37), again, for $q_i = \bar{\psi}$, we interpret $(\partial^R \mathscr{L}/\partial \bar{\psi}) f(\bar{\psi})$ for $\mathscr{L} = \bar{\psi} g(\psi) + \cdots$ as $-f(\bar{\psi})(\partial^R \mathscr{L}/\partial \bar{\psi}) = +f(\bar{\psi}) g(\psi)$,[3] or rather the Wick-ordered product of f and g.[6]

The proof of conservation of T is similar to the usual proof of conservation of energy and momentum in field theory: We prove the validity of a continuity equation. For (37), we will prove

$$\partial_\nu \left\{ \frac{\partial^R \mathscr{L}}{\partial q_{i,\nu}} \bar{\delta}_T q_i + \mathscr{L} \varphi^\nu \right\} = 0, \tag{38}$$

and, using Gauss's theorem, we conclude that the difference between T (37) calculated on two different spacelike hypersurfaces σ_1 and σ_2 is an integral of the form of (37), over a world tube at $r \to \infty$ between σ_1 and σ_2:

$$T(\sigma_2) - T(\sigma_1) = -(\hbar c)^{-1} \lim_{r\to\infty} \oiint r^2 \sin\theta\, d\theta\, d\phi \times \int_{t_1(r,\theta,\phi)}^{t_2(r,\theta,\phi)} c\, dt \{\ \}, \tag{39}$$

where $\{\ \}$ is the same as in the integrand of (37).

Since we use T only for calculating its commutators with fields $F(P)$ at finite points P on σ, and the points over which we integrate in (39) lie at a spacelike distance from these P, the integral (39) commutes with $F(P)$. We therefore may as well neglect terms like (39) in T, and treat T as if it were conserved from σ_1 to σ_2.

For proving (38), we first use the field equations

$$\partial_\nu \left(\frac{\partial \mathscr{L}}{\partial q_{i,\nu}} \right) = \frac{\partial \mathscr{L}}{\partial q_i}, \tag{40}$$

so that the first term on the left in (38) becomes

$$\frac{\partial \mathscr{L}}{\partial q_i} \bar{\delta}_T q_i + \frac{\partial \mathscr{L}}{\partial q_{i,\nu}} \bar{\delta}_T (q_{i,\nu}) = \bar{\delta}_T \mathscr{L} \tag{41}$$

on account of $\bar{\delta}(q_{i,\nu}) = (\bar{\delta} q_i)_{,\nu}$. In (41), $\bar{\delta}_T \mathscr{L}$ stands for the terms *linear* in φ^μ and $\omega^{(\alpha)}_{1(\beta)}$ in the substantial variation of $\mathscr{L}$ when the descriptors are φ^μ and $\omega^{(\alpha)}_{1(\beta)}$ rather than ξ^μ and $\epsilon^{(\alpha)}_{1\ (\beta)}$.

We also assume that $\mathscr{L}$ is a scalar density under the transformations allowed. This gives, with $\mathscr{O}_2$ precision[27] for the $\mathscr{O}_1$ terms of the local variation of $\mathscr{L}$ with our changed descriptors,

$$\delta_T \mathscr{L} = -\mathscr{L} \varphi^\mu_{,\mu} \tag{42}$$

and

$$\bar{\delta}_T \mathscr{L} = \delta_T \mathscr{L} - \mathscr{L}_{,\mu} \varphi^\mu = -(\mathscr{L} \varphi^\mu)_{,\mu}. \tag{43}$$

Therefore, the first term (41) on the left in (38) cancels the second term on the left in (38). This completes the proof of the validity of (38), and therefore proves the conservation of T, but for the irrelevant boundary term (39).

We now do a comparison with the case of Lorentz transformations in flat space-time. The proof of covariance of the commutation relations which we are giving here shows great similarity to the proof of this covariance under Lorentz transformations given in flat space.[1,3] The generator used there is equivalent for that special case to our present expression (31), but, with the transformation coefficients there constants, we have there the special conditions $\xi^\mu_{,\nu\lambda} = 0$ and $\epsilon^{(\alpha)}_{1\ (\beta),\lambda} = 0$, which make T there expressible in terms of the total energy, momentum, and angular momentum, thus guaranteeing in that case the conservation of T. For proving the conservation of our present T, however, we did not explicitly need conditions of constancy of $\xi^\mu_{,\nu}$ and of $\epsilon^{(\alpha)}_{1\ (\beta)}$.

Nevertheless, these conditions are *implicitly* imposed by our postulate of invariance of $L = \mathscr{L}/\sqrt{-g}$, as we use the altered Lagrangian.

X. THE LAGRANGIAN AND HAMILTONIAN DENSITIES

Using the commutation relations (15)–(19), we next should calculate the commutators $[iT; F(P)]$. For this, we need $\mathscr{H}$ (contained in T). For calculating this, we write $\mathscr{L}$ in the general form

$$\mathscr{L} = \tfrac{1}{2}\mathscr{D}^{i\mu j\nu} \varphi_{i,\mu} \varphi_{j,\nu} - \mathscr{U} + \hbar c \{ \tfrac{1}{2}\bar{\psi}_{,\mu} c^\mu \psi - \tfrac{1}{2}\bar{\psi} c^\mu \psi_{,\mu} - m \bar{\psi}\psi + \bar{\psi} \omega^{i\mu} \psi \varphi_{i,\mu} + \bar{\psi} \not{f} \psi \}. \tag{44}$$

Here, the φ_i are the components of the boson fields, including the tetrad fields. Our Lagrangian density $\mathscr{L}$ is second order in the derivatives of these φ_i, with coefficients $\mathscr{D}^{i\mu j\nu} = \mathscr{D}^{j\nu i\mu}$ that are functions of the φ only. Also the $\omega^{i\mu}$ and the c^μ and $\not{f} - m$ are functions of the φ (in particular of the tetrads), and contain also the constant flat-space Dirac matrices. In particular,

$$c^\mu = \not{h}^\mu_{(\alpha)} \gamma^{(\alpha)}, \quad m = \not{\kappa} = \not{m}c/\hbar. \tag{45}$$

The term with m is the fermion mass term; f stands for the interaction between fermion and boson fields; the terms with $\omega^{\mathrm{i}\mu}$ are the Fock terms,

$$\bar{\psi}\omega^{\mathrm{i}\mu}\psi q_{\mathrm{i},\mu} = f\bar{\psi}\Sigma_{(\alpha)\lambda}{}^{\mu}\psi h^{(\alpha)\lambda}{}_{,\mu} = f\bar{\psi}\Sigma^{\mu\lambda(\alpha)}\psi h_{(\alpha)\lambda,\mu}, \tag{46}$$

with

$$\Sigma_{(\alpha)\lambda}{}^{\mu} = \tfrac{1}{8}(\gamma_{(\alpha)}\gamma_{\lambda}\gamma^{\mu} - \gamma^{\mu}\gamma_{\lambda}\gamma_{(\alpha)}) \tag{47}$$

according to (20), unless these terms (46) are omitted, as in the theory of Pellegrini and Plebański.[23,24] The nonlinearity of the theory resides in the dependence of the $\mathscr{D}^{\mathrm{i}\mu\mathrm{j}\nu}$, etc., upon the q_{k}. The canonical conjugates p^{i} to the q_{i} are now given by

$$cp^{\mathrm{i}} = \mathscr{D}^{\mathrm{i}0\mathrm{j}\nu}q_{\mathrm{j},\nu} + \hbar c\bar{\psi}\omega^{\mathrm{i}0}\psi. \tag{48}$$

In the altered theory, we can solve from (48) for all of the boson field velocities by

$$q_{\mathrm{i},0} = K_{\mathrm{ij}}\pi^{\mathrm{j}} \tag{49a}$$

with

$$\pi^{\mathrm{j}} \equiv cp^{\mathrm{j}} - \mathscr{D}^{\mathrm{j}0\mathrm{k}n}q_{\mathrm{k},n} - \hbar c\bar{\psi}\omega^{\mathrm{j}0}\psi, \tag{49b}$$

where the K_{ij} satisfy

$$\mathscr{D}^{\mathrm{h}0\mathrm{i}0}K_{\mathrm{ij}} = \delta^{\mathrm{h}}_{\mathrm{j}}, \quad K_{\mathrm{ih}}\mathscr{D}^{\mathrm{h}0\mathrm{j}0} = \delta^{\mathrm{j}}_{\mathrm{i}}, \quad K_{\mathrm{ij}} = K_{\mathrm{ji}}, \tag{50}$$

and like the $\mathscr{D}^{\mathrm{i}\mu\mathrm{j}\nu}$ are functions of the q.

We now define

$$\mathscr{H} = cp^{\mathrm{i}}q_{\mathrm{i},0} + \tfrac{1}{2}\hbar c(\bar{\psi}_{,0}c^{0}\psi - \bar{\psi}c^{0}\psi_{,0}) - \mathscr{L}, \tag{51}$$

and eliminate from this the $q_{\mathrm{i},0}$ by (49) and the $\psi_{,0}$ and $\bar{\psi}_{,0}$ by the Dirac equations, which here take the form

$$\mathrm{m}\psi + c^{\mu}\psi_{,\mu} + \tfrac{1}{2}c^{\mu}{}_{,\mu}\psi - \omega^{\mathrm{i}\mu}\psi q_{\mathrm{i},\mu} - f\psi = 0, \tag{52a}$$

$$-\bar{\psi}\mathrm{m} + \bar{\psi}_{,\mu}c^{\mu} + \tfrac{1}{2}\bar{\psi}c^{\mu}{}_{,\mu} + \bar{\psi}\omega^{\mathrm{i}\mu}q_{\mathrm{i},\mu} + \bar{\psi}f = 0. \tag{52b}$$

This gives

$$\mathscr{H} \equiv \tfrac{1}{2}K_{\mathrm{ij}}\pi^{\mathrm{i}}\pi^{\mathrm{j}} - \tfrac{1}{2}\mathscr{D}^{\mathrm{i}m\mathrm{j}n}q_{\mathrm{i},m}q_{\mathrm{j},n} + \mathscr{U} + \hbar c\{\mathrm{m}\bar{\psi}\psi + \tfrac{1}{2}\bar{\psi}c^{n}\psi_{,n} - \tfrac{1}{2}\bar{\psi}_{,n}c^{n}\psi - \bar{\psi}\omega^{\mathrm{i}n}\psi q_{\mathrm{i},n} - \bar{\psi}f\psi\}. \tag{53}$$

XI. GENERALIZED HEISENBERG EQUATIONS OF MOTION

In the following, let ξ be any c-number function of the coordinates. Using our modified (anti)commutation relations (15)–(19), we now calculate the commutators $[F; \int\mathscr{H}\xi\, d^{3}\mathbf{x}]$ for $F = q_{\mathrm{i}}, p^{\mathrm{i}}, \psi_{\mathrm{A}}$, or $\bar{\psi}^{\mathrm{A}}$. Derivatives of ξ will appear in the results, by integration by parts of spatial derivatives of Dirac delta functions. Because we did not pay proper attention to the ordering of factors in $\mathscr{L}$, we cannot do so either in $\mathscr{H}$ or in its commutators.[6] As some factors may have commutators containing factors $\delta_{3}(0) = \infty$, we simply must hope for the best that there exists an ordering for which such infinite terms do not occur in the results. With this understanding, we find, by (49a) and (49b)

$$\left[q_{\mathrm{i}}; \int\mathscr{H}\xi\, d^{3}\mathbf{x}\right] = i\hbar c q_{\mathrm{i},0}\xi, \tag{54a}$$

and, by (51a) and (51b),

$$\left[\psi; \int\mathscr{H}\xi\, d^{3}\mathbf{x}\right] = i\hbar c\{\psi_{,0}\xi - \tfrac{1}{2}k\gamma^{0}\gamma^{n}\psi\xi_{,n}\}, \tag{54b}$$

$$\left[\bar{\psi}; \int\mathscr{H}\xi\, d^{3}\mathbf{x}\right] = i\hbar c\{\bar{\psi}_{,0}\xi - \tfrac{1}{2}k\bar{\psi}\gamma^{n}\gamma^{0}\xi_{,n}\}, \tag{54c}$$

where the last term in (54b) arises from the commutator $[\psi; \hbar c\int - \tfrac{1}{2}(\bar{\psi}c^{n}\psi)_{,n}\xi\, d^{3}\mathbf{x}]$, which arises when we write $-\tfrac{1}{2}\bar{\psi}_{,n}c^{n}\psi$ as $-\tfrac{1}{2}(\bar{\psi}c^{n}\psi)_{,n} + \tfrac{1}{2}\bar{\psi}(c^{n}\psi)_{,n}$. A bit more complicated is the derivation of

$$\left[cp^{\mathrm{i}}; \int\mathscr{H}\xi\, d^{3}\mathbf{x}\right]$$

$$= i\hbar c\Big\{cp^{\mathrm{i}}_{,0}\xi - \xi_{,n}\Big[\mathscr{D}^{\mathrm{j}\nu\mathrm{i}n}q_{\mathrm{j},\nu} + \hbar c\bar{\psi}\omega^{\mathrm{i}n}\psi + \tfrac{1}{4}\hbar ck\bar{\psi}\Big(\gamma^{n}\gamma^{0}\frac{\partial c^{0}}{\partial q_{\mathrm{i}}} - \frac{\partial c^{0}}{\partial q_{\mathrm{i}}}\gamma^{0}\gamma^{n}\Big)\psi\Big]\Big\}, \tag{54d}$$

where we must use Eqs. (15a), (15b), and (16). Here, $cp^{\mathrm{i}}_{,0}$ on the right is the time derivative of p^{i} given by Eq. (48). Equality of the two sides of Eq. (54d) follows from the field equations obtained by varying the $q_{\mathrm{j}}(x)$ in this Lagrangian. For functions $\mathscr{F}(q)$, frequent use is also made, in the derivation, of $\mathscr{F}_{,\mu} = (\partial\mathscr{F}/\partial q_{\mathrm{i}})q_{\mathrm{i},\mu}$, $\partial c^{\mu}/\partial q_{\mathrm{i}} = \gamma^{(\alpha)}\partial(fh^{\mu}_{(\alpha)})/\partial q_{\mathrm{i}}$, $k\gamma^{0}\gamma^{0} = 1$, and the anticommutation relations of the Dirac matrices, like $\gamma^{(\alpha)}\gamma^{0} + \gamma^{0}\gamma^{(\alpha)} = 2h^{(\alpha)0}$.

For $\xi = \mathrm{const}$, Eqs. (54a)–(54d) become the Heisenberg equations of motion.

XII. COMMUTATORS WITH T_1

Using Eqs. (3), (I: 40), (I: 40′), and (20), we write out Eq. (31) as

$$T_{1} = \int d^{3}\mathbf{x}\{\hbar^{-1}p^{\mathrm{i}}(\delta_{1}q_{\mathrm{i}} - q_{\mathrm{i},n}\xi^{n}) + \epsilon_{1(\alpha)(\beta)}f\bar{\psi}\Sigma^{(\alpha)(\beta)}\psi + \tfrac{1}{2}f(\bar{\psi}\gamma^{0}\psi_{,n} - \bar{\psi}_{,n}\gamma^{0}\psi)\xi^{n} - (\hbar c)^{-1}\mathscr{H}\xi^{0}\}. \tag{55}$$

Using the modified commutation relations and the above results (54a)–(54d), we find by rather straightforward calculation

$$[iT_{1}; F] = \bar{\delta}_{1}F, \tag{56}$$

for $F = q_{\mathrm{i}}, p^{\mathrm{i}}, \psi$, or $\bar{\psi}$. The calculations for $\mathscr{F} = p^{\mathrm{i}}$ are again the most complicated ones, with many terms that ultimately cancel each other.

Using (56), we can now calculate $[iT_{1}; q_{\mathrm{i},0}]$ by expressing $q_{\mathrm{i},0}$ first by (49a) with (49b) in terms of the q's and the p's. It is then found to be equal to $(\bar{\delta}_{1}q_{\mathrm{i}})_{,0}$.

It is easier to use the fact that T_{1} (like T) is conserved but for irrelevant boundary terms at spatial infinity, as seen from Sec. X with φ^{μ} replaced by ξ^{μ} and with $\bar{\delta}_{T}$ replaced by $\bar{\delta}_{1}$. Therefore,

$$\bar{\delta}_{1}(\mathrm{q}_{\mathrm{i},\mu}) = (\bar{\delta}_{1}\mathrm{q}_{\mathrm{i}})_{,\mu} = [iT_{1};\mathrm{q}_{\mathrm{i}}]_{,\mu} = [iT_{1};\mathrm{q}_{\mathrm{i},\mu}], \tag{57}$$

also for $\mu = 0$. From (57) together with (56) for $F = q$, then,

$$\bar{\delta}_{1}F(\mathrm{q},\mathrm{q}_{,\mu}) = [iT_{1}; F(\mathrm{q},\mathrm{q}_{,\mu})]. \tag{58}$$

Therefore, we may avoid the lengthy direct verification of (56) for $F = p^{\mathrm{i}}$, as it will follow much more easily from (58) by (48).

We thus have verified Eq. (27) for all canonical field variables F, which completes the parts of the proof of covariance of Dirac's modified commutation relations, also for finite coordinate transformations and tetrad rotations, that in the preceding paper (I) were delegated to the present paper.

XIII. LIMITATIONS OF COVARIANCE OF OUR QUANTIZATION PROCEDURE

So far, we have proved covariance of our quantization procedure only under those transformations that leave our altered Lagrangian function L invariant. In practice, this means limitation to affine transformations.

It has often been suggested to confine coordinate systems to those that at space-like infinity asymptotically are Lorentz frames. With this restriction, we have shown above merely covariance of the modified commutation relations for the altered theory under Poincaré transformations only.

Fock has suggested that Poincaré transformations anyhow would be the only coordinate transformations allowed in a theory that impose the De Donder condition and that allows asymptotically only Lorentz frames, provided that one assumes absence of incoming radiation.[28] His justification of this claim is more a plausibility argument than a rigorous proof. In a following paper,[29] we will further discuss whether Fock can be right. The answer to this question, however, may well be irrelevant, if it would be true, as suggested in Sec. XV below, that the *physical part* of the above quantization procedure would be covariant under rather *general* coordinate transformations.

XIV. ANOTHER PROOF THAT THE COMMUTATION RELATIONS ARE INVARIANT UNDER AFFINE TRANSFORMATIONS ONLY

Consider again the definition (37) of T, for the special case that $\omega_1^{(\alpha)}{}_{(\beta)} = 0$ and $\varphi^\mu = \text{const}$, so that $\delta \mathrm{q_i} = 0$. Then, T takes the form

$$T = -\hbar^{-1}\mathscr{P}_\mu \varphi^\mu, \tag{59a}$$

with

$$c\mathscr{P}_\mu = -\int_\sigma \mathrm{t}_\mu{}^\nu \, d\sigma_\nu, \tag{59b}$$

where

$$\mathrm{t}_\mu{}^\nu \equiv \mathscr{L}\delta_\mu^\nu - \frac{\partial^R \mathscr{L}}{\partial \mathrm{q}_{\mathrm{i},\nu}} \mathrm{q}_{\mathrm{i},\mu}. \tag{59c}$$

Also in (38) now φ^μ can be factorized out, and we find from the arbitrariness of the φ^μ the four continuity equations[30]

$$\partial_\nu \mathrm{t}_\mu{}^\nu = 0. \tag{60}$$

Using Gauss's theorem, Einstein[31] has shown that, on account of (60), $\mathscr{P}_\mu$ defined by (59b) transforms as a covariant "free four-vector," that is, as a covariant four-vector at spatial infinity, under transformations that at spatial infinity are affine, but that are *not* affine at finite spatial distances.

In coordinate systems Σ or Σ', in which we assume our modified commutation relations to be valid, we still can derive by direct calculation the validity of the relations (56), which in our present case simplify to

$$-i\hbar^{-1}[\mathscr{P}_\mu \xi^\mu; F] = -\xi^\mu F_{,\mu}, \tag{61}$$

where the ξ^μ are c numbers independent of x. In particular, let us take F at some point P, and for ξ^μ let us choose the infinitesimal four-vector dx^μ pointing from P to Q. Thus we obtain

$$\begin{aligned}[F; c\mathscr{P}_\mu]\,dx^\mu &= i\hbar c F_{,\mu}\,dx^\mu \\ &= i\hbar c[F(Q) - F(P)] + \mathscr{O}_2. \end{aligned} \tag{62}$$

Under our assumption of validity of the commutation relations in Σ as well as in Σ', therefore, (62) should be valid in both Σ and Σ'.

In particular, let us take for $\Sigma \rightarrow \Sigma'$ a transformation that does not change the coordinates at all at spatial infinity, though it does change the coordinates and dx^μ at P and Q. According to Einstein,[31] then, this transformation does not change the $\mathscr{P}_\mu$. If, for simplicity, we take F to be a scalar field, this transformation would change nothing in the right-hand member of (62), while in the left-hand member of (62) it would merely change the dx^μ. This would create a contradiction, so that our original assumption must have been wrong, that we could perform a transformation $\Sigma \rightarrow \Sigma'$ that was not affine, and yet the commutation relations would hold in both Σ and Σ'. We conclude that only under affine transformations the commutation relations remain valid.

We may formulate this argument slightly differently as follows. If both members of (62) are to transform in the same way (for avoiding contradictions of the kind we found above), then $\mathscr{P}_\mu \, dx^\mu$ in the left-hand member of (62) should be a scalar q number. Therefore, the transformation of dx^μ should be contragredient to the transformation of $\mathscr{P}_\mu$. As the latter according to Einstein transforms as a covariant four-vector at infinity, it follows that our dx^μ (which was defined at P) should transform as a contravariant four-vector at infinity. That is, the transformation of a contravariant four-vector should be the same at P as at infinity. This is so only if the coordinate transformation is affine. So, if the commutation relations hold in both Σ and Σ', the transformation from Σ to Σ' must be affine.

XV. POSSIBLE GENERAL COVARIANCE OF THE PHYSICAL PART OF THE QUANTIZED ALTERED THEORY

In dynamic quantization of the original unaltered theory,[14] in linearized approximation,[32] not all ten polarization modes of the metrical field are quantized. The three longitudinal modes[33] $g^{\parallel}_{mn} = \Delta^{-1}\{g_{im,in} + g_{in,im} - \Delta^{-1} g_{ij,ijmn}\}$ are assumed to be zero[34]; the transverse trace component $g_{\mathrm{T}} = g_{ii} - \Delta^{-1} g_{ij,ij}$ becomes the nonlocally derived variable $-2\kappa\Delta^{-1}\mathscr{T}^{00}$; the four timelike components $g_{\mu 0}$ become the nonlocally derived variables $g_{n0} = 2\kappa\Delta^{-1}\{\mathscr{T}^{n0} - \frac{3}{4}\Delta^{-1}\mathscr{T}^{0i}{}_{,in}\}$ and[32] $g_{00} = -1 - \kappa\Delta^{-1}\{\mathscr{T}^{\mu\mu} - \Delta^{-1}\mathscr{T}^{ij}{}_{,ij}\}$. Thus, only the two traceless transverse modes of polarization of $g_{\mu\nu}$ are quantized. Since in the altered theory we quantize all ten components of $g_{\mu\nu}$, or even all 16 components of $h^{\mu}_{(\alpha)}$ or $h^{(\alpha)}_{\mu}$; in the absence of auxiliary conditions this theory allows far more quantum states than are physically possible. That is, "altered Hilbert space" has far too many dimensions; physically possible quantum states form only a small subspace of it.

Now consider the set Ω_1 of coordinate systems obtainable from a given frame of reference Σ_1 by affine transformations. Such a set is called an orbit under the affine group in the space of coordinate systems, or, briefly, an *affine orbit*. Every coordinate system belongs to one affine orbit. Coordinate transformations inside one orbit are affine; nonaffine coordinate transformations lead from one orbit to a different one. Similarly, "Poincaré orbits" would be sets of coordinate systems obtainable from each other by Poincaré transformations.

Now, the Lagrangian function of an altered theory is invariant only in the affine orbit in which it was defined. When we use a Lagrangian of the same form in a different orbit Ω_2, it determines a theory which does *not* follow by coordinate transformation from the altered theory in the original orbit Ω_1. Instead, it is a different theory. *Each affine orbit has its own altered theory.*

Therefore, we should not expect to obtain the commutation relations of the altered theory in the second orbit, by mere coordinate transformation, from the commutation relations of the different altered theory in the first orbit.

As there are infinitely many affine orbits, there are infinitely many altered theories. Whichever affine orbit we select as our starting point, our theory will be covariantly quantized inside that orbit. Each of these theories, by its quantization alone, determines an oversized space of quantum states, of which some are physical states and some are unphysical states. The physical states are then selected by the auxiliary conditions applied in the orbit in which the theory was defined. These auxiliary conditions reduce the altered field equations and observables, in the Hilbert subspace determined by these conditions, to the physically correct equations and quantities of the generally covariant theory, provided that our method of quantization is correct. If also dynamical quantization would be possible and would be a correct method of quantization, the auxiliary conditions should reduce each altered Hilbert space to the Hilbert space defined by dynamical quantization. If dynamical quantization would be a correct procedure, the commutation relations obtained by it should also be covariant under general coordinate transformations, though this would be equally difficult to prove explicitly,[35] as the covariance of the "physical part" of the altered commutation relations.

The equivalence of the physical part of the quantized altered theory to the dynamically quantized original covariant theory is suggested by two facts: (1) in the altered theory, the auxiliary conditions reduce the altered Lagrangian, and therefore also the altered field equations, to the Lagrangian and the field equations for the field variables of the covariant unaltered theory, to which we apply dynamical quantization; and (2) in dynamical quantization we assume for the dynamical field variables (like the transverse fields in Maxwell's theory) the same commutation relations as which would apply to these variables in the altered theory.

Dynamical quantization of the general-relativistic theory and modified canonical quantization of the altered theory of gravity are both generalizations of what is done in quantizing special-relativistic field theory. If we would know for sure that these generalizations are correct methods of quantization, there would be no problem. Compare the birth of special relativity theory. The "microscopic" electromagnetic field equations and equations of motion for charged particles were guessed by Lorentz on the basis of experimental results combined with a number of hypotheses (like the Lorentz contraction of electrons, etc.). Taking these equations for granted, one can prove explicitly their covariance under Lorentz transformations. This proof was given by Lorentz,[36] with some corrections made by Poincaré.[37] Then, Einstein[38] much simplified matters by *postulating* covariance of electromagnetic theory, and thence deriving the equations for moving charges. Similarly, if we knew for sure that our quantization procedures were correct, we could postulate their general covariance as an *application of the general relativity principle*, and from this postulate we could possibly derive some relations between the different quantized altered theories, or between different dynamical quantizations of the covariant theory, in different affine orbits, or in different Poincaré orbits, if we assume space to be asymptotically flat and if we confine ourselves to coordinate systems that asymptotically are Lorentz frames.

However, quantum field theory is an unfinished theory, as shown by the need for swindles in renormalization theory, such as treating logarithmically divergent integrals as small quantities. This decreases a person's confidence in extrapolations of the methods of special-relativistic field theory to the general-relativistic case, and one would wish to have an explicit mathematical proof of general covariance of the quantized theory left after use of the auxiliary conditions, so as to give us more confidence that we are on the right track with our attempts at quantization.

So far, however, no explicit proof is available, so that all we can presently do is trust that the covariance is there, by lack of evidence to the contrary, and by a combination of hope for the correctness of our quantization procedure (within the limits of correctness of present-day quantum theory), with our belief in the general relativity principle.

[1]W. Heisenberg and W. Pauli, Z. Phys. **56**, 1 (1929) [correction: L. Rosenfeld, Z. Phys. **63**, 574 (1930)]; W. Heisenberg and W. Pauli, Z. Phys. **59**, 168 (1929).

[2]L. Corwin, Y. Ne'eman, and S. Sternberg, Rev. Mod. Phys. **47**, 573 (1975); see definitions on p. 574.

[3]F. J. Belinfante, Physica **7**, 765 (1940).

[4]P. A. M. Dirac, Can. J. Math. **2**, 129 (1950).

[5]F. J. Belinfante, D. I. Caplan, and W. L. Kennedy, Rev. Mod. Phys. **29**, 518 (1957).

[6]In the terms of the Lagrangian, products are usually interpreted as Wick products (normal-ordered products). In the present paper, however, we disregard questions about ordering of factors.

[7]F. J. Belinfante, J. Math. Phys. **26**, 2814 (1985), referred to in the following as (I).

[8]Here, ∂^L/∂ and ∂^R/∂ denote differentiations from the left and from the right. These were earlier called *ante*derivatives and *post* derivatives.[3]

[9]B. S. DeWitt and C. M. DeWitt, Phys. Rev. **87**, 116 (1952).

[10]L. Rosenfeld, Ann. Phys. Leipzig **5**, 113 (1930), Ann. Inst. H. Poincaré **2**, 25 (1932).

[11]P. G. Bergmann, Phys. Rev. **75**, 680 (1949); J. L. Anderson and P. G. Bergmann, Phys. Rev. **83**, 1018 (1951).

[12]F. A. E. Pirani, A. Schild, and R. Skinner, Phys. Rev. **87**, 452 (1952).

[13]F. J. Belinfante and J. S. Lomont, Phys. Rev. **84**, 541 (1951).

[14]R. Arnowitt, S. Deser, and C. W. Misner, Phys. Rev. **117**, 1595 (1960); **118**, 1100 (1960).

[15]E. Fermi, Rev. Mod. Phys. **4**, 87 (1932).

[16]F. J. Belinfante, Physica **12**, 1, 17 (1946).

[17]The name "altered" was an invention of the Editor of the Rev. Mod. Phys.,[5] who did not consider Belinfante's original name "muddified" sufficiently dignified.

[18]S. N. Gupta, Proc. Phys. Soc. London Sect. A **63**, 681 (1950); A **66**, 129 (1953).

[19]Th. De Donder, *La Gravifique Einsteinienne* (Gauthier-Villars, Paris, 1921), p. 67–68.

[20]V. Fock, Z. Phys. **57**, 261 (1929).

[21]F. J. Belinfante, Physica **7**, 305 (1940).

[22]See Footnote 7 of Ref. 9.

[23]C. Pellegrini and J. Plebanski, K. Dan. Vidensk. Selsk. Mat. Fys. Skr. **2**, No. 4 (1963).

[24]C. Møller, K. Dan. Vidensk. Selsk. Mat. Fys. Medd. **34**, No. 3 (1964).

[25]For this group we will choose the group of transformations that leave the Lagrangian function L invariant.

[26]By (I: 52a) we mean Eq. (52a) of paper (I).[7]

[27]Remember that, in an expansion in powers of the infinitesimal descriptors and their derivatives, $\delta_T \mathscr{L}$ represents, in the local variation of $\mathscr{L}$, the terms *linear* in the changed descriptors φ^{μ} and $\omega_1^{(\alpha)}{}_{(\beta)}$ and their derivatives, like $\delta_1 \mathscr{L}$ did for the original descriptors. Therefore, in $\delta_T \mathscr{L}$ the second-order terms $\frac{1}{2}\mathscr{L}(\varphi^{\mu}{}_{,\lambda}\varphi^{\lambda}{}_{,\mu} + \varphi^{\mu}{}_{,\mu}\varphi^{\lambda}{}_{,\lambda})$ from $L\delta\sqrt{-g}$ are absent.

[28]V. Fock (1955), translation *The Theory of Space Time and Gravitation*, (Pergamon, Mac Millan, New York, 1964), 2nd revised ed., pp. 369–375.

[29]F. J. Belinfante, J. Math. Phys. **26**, 2836 (1985).

[30]The validity of these continuity equations (60) now follows directly from the field equations (40), *without* use of the transformation properties of $\mathscr{L}$.

[31]A. Einstein, Sitzungsber. Preuss. Akad. Wiss. Phys. Math. Kl. **1918**, 448; see also W. Pauli, "Relativitätstheorie" in *Encyklopädie der Mathematischen Wissenschaften* (Teubner, Leipzig, 1921), Article V 19, pp. 610–613; translation *Theory of Relativity* (Pergamon, London, 1958; Dover reprint, New York, 1981), pp. 60–62.

[32]In the linearized approximation, the contributions of $\mathscr{T}^{ll}$ or of $\Delta^{-1}\mathscr{T}^{lj}{}_{,lj}$ to g_{00} cannot be trusted, because this approximation neglects the difference between $\mathscr{T}^{\mu\nu}{}_{,\nu}$ and $\nabla_\nu \mathscr{T}^{\mu\nu}$.

[33]Here, Δ^{-1} is the inverse of the Laplacian, so, $\Delta^{-1}f(x,t) = -\int d^3\mathbf{x}' f(\mathbf{x}',t)/4\pi r$.

[34]Their canonical conjugates are nonlocally derived variables, according to $p^{rs} = -(2c\Delta)^{-1}\{\mathscr{T}^{0r}{}_{,s} + \mathscr{T}^{0s}{}_{,r} - \Delta^{-1}\mathscr{T}^{0l}{}_{,lrs}\}$.

[35]We proved the covariance of dynamical quantization of special-relativistic quantum electrodynamics under Lorentz transformations only,[13] by a method which presently cannot be generalized to a proof of general covariance of nonlinear gravitational theory.

[36]H. A. Lorentz, Proc. Acad. Amsterdam **6**, 809 (1904), or *Collected Papers* (Nijhoff, The Hague, 1937), Vol. 5, pp. 172–197.

[37]H. Poincaré, C. R. Acad. Sci. Paris **140**, 1504 (1905); Rend. Circ. Mat. Palermo **21**, 129 (1906).

[38]A. Einstein, Ann. Phys. **17**, 891 (1905).

The De Donder condition and the Poincaré group in quantized general relativity

Frederik J. Belinfante[a)]
Department of Physics, Purdue University, Lafayette, Indiana 47907

(Received 8 February 1984; accepted for publication 13 June 1985)

The De Donder coordinate condition for an unquantized metrical field is discussed first. While the coordinate transformations allowed by this condition do not form a group in the ordinary sense, it is possible in an infinite number of ways to make these transformations elements of a group of *metric-dependent* coordinate transformations. In the unquantized theory, the question what coordinate transformations the De Donder condition will allow remains not generally answered, but it is not of primary importance, because the imposition of the De Donder condition in the unquantized theory is at best a question of convenience. The question is more important in the quantized "altered theory" of gravity, in which the De Donder condition serves for selecting the physical part of a theory that without this condition would also describe unphysical states. There are as many "altered" quantum theories of gravity, as there are affine orbits in the space of coordinate systems. Each orbit has its own altered theory, and in each separate altered theory with its own De Donder condition the only coordinate transformations that make sense are the affine ones, or, with proper boundary conditions imposed, Poincaré transformations. The coordinate transformation between two frames in which the De Donder condition is valid, either weakly or strongly, must be an affine (or Poincaré) transformation, if the coordinates are required to be c numbers.

I. INTRODUCTION

In his book *The Theory of Space Time and Gravitation*, Fock[1] had claimed that harmonic coordinate transformations would automatically be Poincaré transformations. His "proof" of this claim is not rigorous, and therefore we should call it a *conjecture*. The validity of his claim has been attacked by Bergmann, who has claimed that Fock's harmonic coordinate transformations would not even form a group.[2] Before we discuss these contradictory claims, first we provide some definitions.

II. DE DONDER TRANSFORMATIONS

A De Donder frame is a coordinate system, in which the metric satisfies the De Donder condition[3]

$$g^{\mu\nu}{}_{,\nu} = 0. \tag{1}$$

A coordinate transformation $\Sigma \rightarrow \Sigma'$ applied to an unprimed De Donder frame is called a De Donder transformation, if the result Σ' is again a De Donder frame. It is easily verified that such a De Donder transformation $x^{\lambda'} = f^{\lambda'}(x)$ must satisfy the four equations

$$g^{\mu\nu} \frac{\partial^2 f^{\lambda'}}{\partial x^\mu \, \partial x^\nu} = 0. \tag{2}$$

Therefore, if the same transformation $f^{\lambda'}$ is applied to a different De Donder frame Σ'', and the result is $x^{\lambda'''} = f^{\lambda'}(x'')$, by $g^{\mu''\nu''} \neq g^{\mu\nu}$ we cannot expect $g^{\mu''\nu''}(x'')\partial^2 f^{\lambda'}(x'')/\partial x^{\mu''} \, \partial x^{\nu''}$ to vanish, and therefore a transformation which is a De Donder transformation when applied to one De Donder frame is not a De Donder transformation when applied to a different De Donder frame. Therefore, De Donder transformations do not form a group, as, without knowledge of the frame of reference upon which a transformation is applied, we cannot even tell whether or not a transformation $f^{\lambda'}$ would be a De Donder transformation.

Instead of a *group*, what the De Donder transformations do form might be called a *network*, of connections between all pairs of De Donder frames. Therefore, the De Donder network does not form a group of coordinate transformations if we regard coordinate transformations simply as given one-to-one mappings between old and new coordinates, independent of the metric. As the answer to the question what is a De Donder transformation depends not merely upon the coordinates, but also upon the metrical field, the group property is lost for transformations that do *not* depend upon the metrical field.

If, by hook or by crook, we still want to consider De Donder transformations as forming a group, this can be done, in infinitely many ways, provided that we interpret a coordinate transformation as a mapping that *depends upon the metrical field originally present*. For an example, see the Appendix.

The question whether or not the De Donder network may be regarded as a group, however, is of rather limited interest.

III. BOUNDARY CONDITIONS AND HARMONIC TRANSFORMATIONS

In a space-time that at spatial infinity is asymptotically flat, we may want to confine coordinate systems to those that at spatial infinity asymptotically become Lorentz frames. With this boundary condition, affine transformations automatically become Poincaré transformations, while arbitrary coordinate transformations by this boundary condition become merely asymptotically Poincaré transformations.

[a)] Present address: P. O. Box 901, Gresham, Oregon 97030.

In such a space-time, harmonic frames according to Fock[1] are De Donder frames that satisfy not only this boundary condition, but that also satisfy an additional boundary condition, which Fock calls "absence of incoming radiation." Harmonic coordinate transformations now are defined as coordinate transformations between harmonic frames, like De Donder transformations were between De Donder frames. By the boundary condition, therefore, harmonic transformations are asymptotically Poincaré transformations. According to Fock, by Eq. (2) and the additional condition of absence of incoming radiation, they would be Poincaré transformations rigorously throughout space-time.

It should here be emphasized that Fock makes this claim for *classical* (unquantized) gravitational theory. We will find the use of the De Donder condition (1) more interesting in quantized gravitational theory than in the classical theory. We will discuss the two cases separately. We first will finish the discussion of the classical case, and will start the discussion of the quantized case in Sec. VI.

IV. CLASSICAL ADVANTAGES OF AUXILIARY CONDITIONS

The De Donder condition in gravitational theory may be compared to the Lorentz condition in electrodynamics. We know that in Lorentz-covariant electrodynamics in flat space-time, irrespective of the choice of gauge, we may define advanced and retarded field strengths **E** and **B**. We also may derive field strengths from electrodynamic potentials **A** and Φ. If we postulate the Lorentz condition $\partial_\nu A^\nu = 0$, the electrodynamic potentials become components of a four-vector, and the advanced and retarded field strengths then will be derived from advanced and retarded potentials. These are advantages of the Lorentz gauge. (If we would use the transverse gauge, the electrostatic potential would become the instantaneous Coulomb potential, and the vector potential would be the transverse part of the advanced or retarded potential.)

The advantages of auxiliary conditions are partially lost in the nonlinear general-relativistic theory. The De Donder condition keeps some advantages, as discussed by Papapetrou[4] and by Gupta.[5] It simplifies the classical gravitational Lagrangian. However, nonlinearity of the theory remains, and severely restricts the advantages of this condition.

Also in static problems, the advantage of the De Donder condition is limited. The De Donder coordinates or harmonic coordinates which describe the Schwarzschild field are Cartesian coordinates that are obtainable from spherical coordinates r, θ, ϕ, where r is related to the Schwarzschild radial coordinate $\mathscr{r}$ by $\mathrm{r} = \mathscr{r} - \mathscr{m}$ (with $\mathscr{m} = GM/c^2$).[6] The harmonic spherically symmetric static coordinates and their uniqueness have been investigated for more general cases, in the presence of electric charge or of spherically symmetric extended mass or energy density distributions.[7]

If use of these coordinate systems that satisfy the De Donder condition for any particular purpose may have any advantage, glory be, but in that case what is the purpose of transforming to a different and probably less advantageous coordinate system that would also satisfy the De Donder condition?

In short, the study of the network of De Donder transformations or of harmonic coordinate transformations in the classical theory of gravity does not seem to be of very great importance.

V. THE QUESTION OF UNIQUENESS, MODULO POINCARÉ TRANSFORMATIONS, OF HARMONIC FRAMES

The question whether Fock's conjecture was right or wrong is not solved by Bergmann's claim that harmonic transformations (regarded merely as coordinate substitutions independent of the metric[8]) would not form a group.[2] In fact, if Fock were right, Bergmann's claim obviously would be wrong (as in that case harmonic coordinate transformations would form the Poincaré group).[9] Bergmann's claim, on the other hand, would be justified, if Fock's claim were wrong. Therefore, all that can be said so far is that either Fock or Bergmann is right, and then the other one is wrong. This does not establish *who* is right. Though investigations of special static spherically symmetric cases seem to support Fock's point of view,[7] there is no convincing evidence either way in the general case. But, as remarked already above, in the unquantized theory the answer to this mathematical question is for the physicist merely a matter of curiosity, and not of earthshaking importance.

VI. THE DE DONDER CONDITION IN THE QUANTUM THEORY OF GRAVITY

Contrary to the above, the assumption of the De Donder condition becomes more than a matter of convenience in the altered quantum theory of gravity, discussed in our preceding paper.[10] There, we had to impose this condition upon the states physically realizable, for separating valid predictions of the theory from invalid ones. In this quantum theory, however, the question what transformations are allowable has a very definite answer.

In the altered theory, terms quadratic and bilinear in $S^\mu \equiv \partial_\nu \mathfrak{g}^{\mu\nu}$ are added to the Lagrangian. This creates a theory that is covariant under affine coordinate transformations only. We have seen[10] that then also the canonical commutation relations (in their "modified" form, as appropriate for a theory describing in curved space-time fermions as well as bosons) will be covariant under affine transformations only. These affine transformations become Poincaré transformations, when we impose the boundary conditions mentioned above in Sec. III.

Therefore, a first reason for confining oneself to affine transformations (or to Poincaré transformations), in a theory which requires the vanishing of S^μ for obtaining physical results, is that under other coordinate transformations the theory would not be covariant. By different transformations one would end up in a different affine orbit or Poincaré orbit, in which a different altered theory should be used with its own auxiliary conditions, not following by coordinate transformation from the auxiliary conditions in the original altered theory. As discussed in our preceding paper,[10] the existence of an altered theory for every affine orbit might guarantee general covariance of the physical part of all these altered theories taken together.

The conclusion that the imposing of the De Donder condition in a given field theory must be confined to one affine orbit, may also be obtained by centering our attention upon the auxiliary conditions $S^{\mu} = 0$ themselves. Suppose Σ and Σ' are two different coordinate systems, in which we impose the De Donder condition as a "weak" condition that places limitations upon the quantum-mechanical states that should be regarded as physically possible.[11] Then, also Eq. (2) should be valid as a weak equation. Equation (2), however, differs from the De Donder condition (1) in that it contains the entire metric $g^{\mu\nu}$, and not merely its derivatives. If the metric is decomposed into its dynamical parts, its derived parts, and its nonphysical parts,[12] like the electrodynamic potentials in the Lorentz gauge can be decomposed into their transverse dynamic parts, their derived parts (the instantaneous Coulomb potential), and their nonphysical parts (the longitudinal vector potential), Eq. (2) will contain also the dynamical part of the metric, and therefore, as an auxiliary condition, Eq. (2) would impose limitations upon the dynamical part of the field as well as upon its other parts.[13] This is not acceptable, and therefore the dynamical parts of the metric should appear in (2) with zero coefficients. This means that we must have

$$\partial_\mu \partial_\nu x^{\lambda'} + \partial_\nu \partial_\mu x^{\lambda'} = 0, \tag{3}$$

if the auxiliary conditions are to be valid in Σ' as well as in Σ. The only transformations $\Sigma \to \Sigma'$ that satisfy (3) are the affine transformations

$$x^{\lambda'} = a_\mu^{\lambda'} x^\mu + b^{\lambda'} \tag{4}$$

($a_\mu^{\lambda'}$ and $b^{\lambda'}$ constant), which, by the boundary conditions of Sec. III, become Poincaré transformations. This again shows why we should confine the validity of the De Donder condition as a weak auxiliary condition upon the physical states in a given altered theory, to the one affine orbit in the space of coordinate systems, in which this altered theory was formulated,[10] or, with the above boundary conditions, to one Poincaré orbit.

Note that we never said here anything about lack of incoming radiation. We do not need here that restriction.

VII. USE OF c-NUMBER COORDINATES

In the conclusion (3) from (2) with quantized metric, a tacit assumption was made. We assumed here that the only q numbers entering (2) were the $g^{\mu\nu}$ appearing as a factor multiplying the second derivatives of the new coordinates, so that the latter, as coefficients of those q numbers, had to vanish, for avoiding an unacceptable auxiliary condition involving the dynamical gravitational field.

In other words, we assumed that the new coordinates had to be c numbers, if the old coordinates were already c numbers. This condition is automatically fulfilled, if we postulate that, as coordinates, we will use c numbers only.

In quantum theory, postulating that coordinates should be c numbers excludes coordinates that in the usual terminology of quantum theory are called "observables" (quantities described by Hermitian operators that do not commute with all other q numbers). For instance, the definition "z = distance from the floor" (an observable) does not provide a rigorous definition of z, because quantum mechanically there is for the position of the top molecules of the floor at best some probability distribution, but no mathematical certainty. Therefore, for labeling points rigorously, we need c numbers. That is, we want coordinates of points to have unambiguous values, so that they can be introduced only by postulate, in contrast to distances between points, which are observables that depend through the metric upon the presence of nearby matter, for which there are only probability distributions. Therefore, the metric components must be q numbers.

If we made distances q numbers by making coordinates q numbers, while the components of the metric then might be c numbers, the metric field and other fields would be functions of q numbers instead of functions of c numbers. Also integrals over space and over space-time would be integrals over q numbers. This would be unwieldy, and would much complicate an understanding of these fields and integrals. This is an additional reason for wanting to avoid q number coordinates. It is one of the advantages of the Fermi-type[14] quantization of the altered theory of gravitation that it allows us to avoid q number coordinates by interpreting auxiliary conditions like the De Donder condition as a method of selecting physical Hilbert space inside an oversized space of quantum states, rather than as a coordinate condition that would make the coordinates depend upon the metrical field.

APPENDIX: EXAMPLE OF HOW A DE DONDER NETWORK MAY BE REGARDED AS A GROUP OF METRIC-DEPENDENT TRANSFORMATIONS

Because $g^{00} \neq 0$, it is possible to solve from Eqs. (2) for the $\partial_0 \partial_0 f^{\lambda'}$ as functions of the metrical field and the $\partial_n(\partial_0 f^{\lambda'})$ and $\partial_m \partial_n f^{\lambda'}$. Therefore, if in Σ on the hypersurface $\sigma(x^0 = \text{const})$ we know the initial values of the eight fields $f^{\lambda'}$ and $\partial_0 f^{\lambda'}$ as functions of the spatial coordinates x^n, we could in a given (unquantized) metric field (and in the absence of singularities) solve by integration for the fields $f^{\lambda'}(x)$ and $\partial_0 f^{\lambda'}(x)$ at different values of x^0.

It is preferable to use here $\xi^\lambda = f^{\lambda'}(x) - x^\lambda$ instead of $f^{\lambda'}$ itself, so that the identity transformation is given by $\xi^\lambda = 0$ and $\partial_0 \xi^\lambda = 0$ on σ. A De Donder transformation starting from a De Donder frame Σ therefore is determined by the choice of ξ^λ and $\partial_0 \xi^\lambda$ at $x^0 = 0$, as functions of the x^n. Though the explicit form of $f^{\lambda'}(x) = x^\lambda + \xi^\lambda(x)$ for $x^0 \neq 0$ will depend upon the metric field present, we will in the following call different fields $\xi^\lambda(x)$ in different metric fields "the same coordinate transformation," as long as their initial ξ^λ and $\partial_0 \xi^\lambda$ on σ are the same.

We may easily generalize this. We pick any arbitrary De Donder frame as a preferred unprimed coordinate system Σ. When the actual coordinate system is a De Donder frame Σ', we still will write x^μ and σ for the coordinates in the preferred frame Σ and the hypersurface $x^0 = 0$ in it. A De Donder transformation $\Sigma' \to \Sigma''$ will satisfy $g^{\mu'\nu'}(x')\partial_{\mu'} \partial_{\nu'} \xi^\lambda = 0$ now with $\xi^\lambda = x''^\lambda - x'^\lambda$. For a given metric field, this transformation will still be given unambiguously, if on σ we know the eight components of ξ^λ and $\partial\xi^\lambda/\partial x^0$ as functions of the preferred spatial coordinates x^n on σ. (Note that we used here ∂_0 and not $\partial_{0'}$.)

Therefore, we will regard the eight fields ξ^λ and $\partial_0\xi^\lambda$ on σ as the parameters determining a De Donder transformation $\Sigma' \rightarrow \Sigma''$ throughout space-time in whatever metric field happens to be present. For the unit element in the De Donder group, these eight parameters will be zero. For the inverse transformation, they will have the opposite sign. For successive De Donder transformations, these parameters on the one preferred hypersurface σ in Σ will be additive. With the De Donder transformations labeled by these eight parameters given as functions of the spatial coordinates in Σ, all four group properties are satisfied, and the De Donder transformations form a group. If we had chosen a different preferred frame Σ and thus a different hypersurface σ, we would have obtained a different group. There are, therefore, as many De Donder groups of this kind, as there are choices of preferred De Donder frames Σ.

[1]V. Fock (1955), translation *The Theory of Space Time and Gravitation* (Pergamon, Mac Millan, New York, 1964), 2nd ed., pp. 369–375.

[2]P. G. Bergmann, Phys. Rev. **124**, 274 (1961), Sec. III.

[3]Th. De Donder, *La Gravifique Einsteinienne* (Gauthier-Villars, Paris, 1921), p. 40, where our Eq. (1) is written in the form $g^{\sigma\tau}(g_{\sigma\tau,\alpha} - 2g_{\alpha\sigma,\tau}) = 0$. See also Th. De Donder, *Mathematical Theory of Relativity* (M.I.T., Cambridge, MA, 1927), pp. 50–51, Eqs. (31) and (32).

[4]A. Papapetrou, Proc. R. Ir. Acad. Sect.A **52**, 11 (1948).

[5]S. N. Gupta, Proc. Phys. Soc. London Sect. A **65**, 161 (1952).

[6]F. J. Belinfante, Phys. Rev. **98**, 793 (1955).

[7]F. J. Belinfante and J. C. Garrison, Phys. Rev. **125**, 1124 (1962).

[8]That is, not interpreted as in the Appendix as metric-dependent mappings.

[9]If Fock is right, Bergmann's argument fails, because harmonic transformations then satisfy Eqs. (4) and (3), so that (2) will be satisfied *irrespective of the metric*.

[10]F. J. Belinfante, J. Math. Phys. **26**, 2827 (1985).

[11]If we would impose the De Donder condition as a strong (q-number) condition in coordinate systems Σ and Σ', so that Eq. (2) would be a strong equation, it is easily seen that the only c-number solutions of (2) for $x^{\lambda'}(x)$ would be the affine solutions (4), so that the postulate of c-number coordinates would guarantee that $\Sigma \rightarrow \Sigma'$ would be an affine transformation.

[12]R. Arnowitt, S. Deser, and C. W. Misner, Phys. Rev. **117**, 1595 (1960); **118**, 1100 (1960).

[13]The dynamical part of the metric entered Eq. (2), because under nonaffine transformations the nonphysical part of the graviton field mixes with the dynamical part.

[14]E. Fermi, Rev. Mod. Phys. **4**, 87 (1932); F. J. Belinfante, Physica **12**, 1, 17 (1946).

Equivalence of a magnetohydrodynamic fluid and a viscous fluid with heat flux

S. R. Maiti and S. N. Das
Theoretical Physics Centre, Department of Physics, Midnapore College, Midnapore 721 101, India

(Received 22 May 1984; accepted for publication 19 April 1985)

In this paper the equivalence between a perfect fluid with electromagnetic field and a viscous fluid with heat flux is considered. Heat flux, shear tensor, and other fluid parameters are readily calculated out of the magnetohydrodynamic fluid parameters.

I. INTRODUCTION

In an illuminating paper Tupper[1] has discussed in depth the equivalence of a viscous fluid and an electromagnetic field under general relativity. First he has shown the electrovac case to be equivalent to a viscous fluid provided that some restriction on geometry is assumed (viz., the eigenvalues of the shear tensor are indistinct—see also Raychaudhuri and Saha[2]). Next he introduces a heat conduction term along with the viscous fluid and has shown that such a heat conduction term is not admissible unless the electromagnetic field is null. Again he introduces a perfect fluid along with an electromagnetic field and obtains some equivalent conditions with a viscous fluid, and in such cases also some restrictions on geometry need be assumed.[1,3] Thus one should have no other alternative but to introduce a heat flux along with the viscous fluid to make it equivalent to an electromagnetic fluid in general. In the present paper it is shown that such an introduction of a heat flux is indeed possible. The heat flux and shear tensor terms are readily obtainable in terms of electromagnetic field components and no further restriction on geometry need be assumed.

II. RAINICH CONDITIONS AND EQUIVALENCE OF TWO FLUIDS

We consider a situation where the electromagnetic field is associated with a perfect fluid. Thus Einstein's equation is

$$G_{\mu\nu} = -K_{\mu\nu}, \tag{2.1a}$$

with the energy momentum tensor $K_{\mu\nu}$ given as

$$K_{\mu\nu} = (\bar{p}+\bar{\rho})v_\mu v_\nu - \bar{p}g_{\mu\nu} + E_{\mu\nu}, \tag{2.1b}$$

where v^μ is the magnetohydrodynamic fluid flow vector (i.e., $v^\alpha v_\alpha = 1$) and $\bar{p}, \bar{\rho}$ are the pressure and density of the fluid. Here $E_{\mu\nu}$ is the electromagnetic energy momentum field tensor such that (cf. Lichnerowicz[4])

$$E_{\mu\nu} = (\tfrac{1}{2}g_{\mu\nu} - v_\mu v_\nu)(E_\alpha E^\alpha + B^\alpha B_\alpha) - (E_\mu E_\nu + B_\mu B_\nu) - (S_\mu v_\nu + S_\nu v_\mu), \tag{2.1c}$$

where E^α, B^α, and S^α are the electric field, the magnetic field, and the Poynting vector, respectively. Retaining the space-time to be the same (i.e., with the same $g_{\alpha\beta}$) one can replace (2.1) by a viscous fluid with a heat flux given as

$$G_{\mu\nu} = -H_{\mu\nu}, \tag{2.2a}$$

where the energy momentum tensor of the fluid is

$$H_{\mu\nu} = (p+\rho)u_\mu u_\nu - pg_{\mu\nu} + 2\eta\sigma_{\mu\nu} + q_\mu u_\nu + q_\nu u_\mu. \tag{2.2b}$$

Here u^μ is the fluid flow vector such that $u^\alpha u_\alpha = 1$ and p, ρ are the pressure and density of the fluid, η is the coefficient of viscosity of the fluid, $\sigma_{\alpha\beta}$ is the shear tensor derived from u^μ, and q^μ is the heat flux vector such that $q_\alpha u^\alpha = 0$ and $q^2 = -q^\alpha q_\alpha$.

The equivalence of (2.1) and (2.2) demands

$$E_{\mu\nu} = (p+\rho)u_\mu u_\nu - (\bar{p}+\bar{\rho})v_\mu v_\nu - (p-\bar{p})g_{\mu\nu} + 2\eta\sigma_{\mu\nu} + q_\mu u_\nu + q_\nu u_\mu. \tag{2.3}$$

As is evident from (2.1c), $E^\alpha{}_\alpha = 0$, one should have from (2.3)

$$(\rho - \bar{\rho}) = 3(p - \bar{p}) \tag{2.4}$$

to be satisfied. Now writing

$$M = p+\rho \quad \text{and} \quad N = \bar{p}+\bar{\rho}, \tag{2.5}$$

Eq. (2.3) can be rewritten as

$$E_{\mu\nu} = Mu_\mu u_\nu - Nv_\mu v_\nu - \tfrac{1}{4}(M-N)g_{\mu\nu} + 2\eta\sigma_{\mu\nu} + q_\mu u_\nu + q_\nu u_\mu. \tag{2.6}$$

Now we introduce a unit spacelike vector n^α, orthogonal to u^α such that

$$v^\alpha = \mu u^\alpha + \lambda n^\alpha, \tag{2.7a}$$

with

$$\mu^2 - \lambda^2 = 1 \tag{2.7b}$$

and $n^\alpha n_\alpha = -1$, $u^\alpha n_\alpha = 0$, $v^\alpha u_\alpha = \mu$, and $v^\alpha n_\alpha = -\lambda$. Also we define

$$\psi_{\alpha\beta} = 2\eta\sigma_{\alpha\beta} - \lambda^2 N n_\alpha n_\beta, \tag{2.8a}$$

$$\psi = \psi^\alpha{}_\alpha = \lambda^2 N, \tag{2.8b}$$

$$\bar{q}^\alpha = q^\alpha - \mu\lambda N n^\alpha, \tag{2.8c}$$

$$\bar{q}^2 = -\bar{q}^\alpha\bar{q}_\alpha = q^2 - 2\mu\beta q N + \mu^2\lambda^2 N^2, \tag{2.8d}$$

where $q_\alpha v^\alpha = \lambda q_\alpha n^\alpha = -\beta q$. Here $\psi_{\alpha\beta}u^\beta = 0$ and $\bar{q}_\alpha u^\alpha = 0$. With the help of (2.8) one can write (2.6) as

$$E_{\alpha\beta} = (M - N\mu^2)u_\alpha u_\beta - \tfrac{1}{4}(M-N)g_{\alpha\beta} + \psi_{\alpha\beta} + \bar{q}_\alpha u_\beta + \bar{q}_\beta u_\alpha. \tag{2.9}$$

As is evident from the property of the electromagnetic field, one can expect that (2.9) should also satisfy the Rainich condition

$$E_{\alpha\mu}E^{\mu\beta} = \delta^\beta{}_\alpha \tfrac{1}{4}E_{\mu\nu}E^{\mu\nu}. \tag{2.10}$$

Using Eq. (2.9)

$$E_{\alpha\beta}E^{\alpha\beta} = (M - N\mu^2)^2 - \tfrac{1}{2}(M-N)(M-N\mu^2) + \tfrac{1}{4}(M-N)^2 - \tfrac{1}{2}(M-N)\psi + \psi_{\alpha\beta}\psi^{\alpha\beta} - 2\bar{q}^2, \tag{2.11}$$

and (2.10) can be calculated from (2.9) as

$$\begin{aligned}&[(M-N\mu^2)^2 - \tfrac{1}{2}(M-N)(M-N\mu^2) - \bar{q}^2]u_\alpha u_\beta \\ &+ [\tfrac{1}{2}(M+N) - N\mu^2](\bar{q}_\alpha u_\beta + \bar{q}_\beta u_\alpha) \\ &- \tfrac{1}{4}g_{\alpha\beta}[(M-N\mu^2)^2 - \tfrac{1}{2}(M-N)(M-N\mu^2) \\ &- 2\bar{q}^2 - \tfrac{1}{2}(M-N)\psi + \psi_{\mu\nu}\psi^{\mu\nu}] \\ &- \tfrac{1}{2}(M-N)\psi_{\alpha\beta} + \psi_{\alpha\mu}\psi^{\mu}{}_{\beta} + \psi_{\alpha\mu}\bar{q}^{\mu}u_\beta \\ &+ \psi_{\beta\mu}\bar{q}^{\mu}u_\alpha + \bar{q}_\alpha\bar{q}_\beta = 0.\end{aligned} \tag{2.12}$$

Contracting by $u^\alpha u^\beta$ we have

$$\psi_{\alpha\beta}\psi^{\alpha\beta} - \tfrac{1}{2}(M-N)\psi = 3(M-N\mu^2)[\tfrac{1}{2}(M+N) - N\mu^2] - 2\bar{q}^2. \tag{2.13}$$

Substituting in (2.12) we have

$$\begin{aligned}&- h_{\alpha\beta}\{(M-N\mu^2)[\tfrac{1}{2}(M+N) - N\mu^2] - \bar{q}^2\} \\ &+ [\tfrac{1}{2}(M+N) - N\mu^2](\bar{q}_\alpha u_\beta + \bar{q}_\beta u_\alpha) \\ &- \tfrac{1}{2}(M-N)\psi_{\alpha\beta} + \psi_{\alpha\mu}\psi^{\mu}{}_{\beta} + \psi_{\alpha\mu}\bar{q}^{\mu}u_\beta \\ &+ \psi_{\beta\mu}\bar{q}^{\mu}u_\alpha + \bar{q}_\alpha\bar{q}_\beta = 0.\end{aligned} \tag{2.14}$$

Contracting further by u^β we have

$$[\tfrac{1}{2}(M+N) - N\mu^2]\bar{q}_\alpha + \psi_{\alpha\beta}\bar{q}^\beta = 0. \tag{2.15}$$

Thus $\bar{q}_\alpha$ is an eigenvector of $\psi_{\alpha\beta}$. Substituting (2.15) in (2.14) we have

$$\begin{aligned}&- h_{\alpha\beta}\{(M-N\mu^2)[\tfrac{1}{2}(M+N) - N\mu^2] - \bar{q}^2\} \\ &- \tfrac{1}{2}(M-N)\psi_{\alpha\beta} + \psi_{\alpha\mu}\psi^{\mu}{}_{\beta} + \bar{q}_\alpha\bar{q}_\beta = 0.\end{aligned} \tag{2.16}$$

Now contracting (2.9) by u^β and $\bar{q}^\beta$ separately we have

$$E_{\alpha\beta}u^\beta = mu_\alpha + \bar{q}_\alpha, \tag{2.17a}$$

$$E_{\alpha\beta}\bar{q}^\beta = -m\bar{q}_\alpha - \bar{q}^2 u_\alpha, \tag{2.17b}$$

where

$$m = \tfrac{1}{4}(3M+N) - N\mu^2. \tag{2.17c}$$

Thus neither u^α nor $\bar{q}^\alpha$ is an eigenvector of $E_{\alpha\beta}$, but there are two eigenvectors lying in the plane containing u^α and $\bar{q}^\alpha$. By considering eigenvectors of the form $u^\mu + \alpha\bar{q}^\mu$ one can show that [from (2.17)]

$$\alpha = m \pm (m^2 - \bar{q}^2)^{1/2}, \tag{2.18}$$

and the corresponding eigenvalues are

$$m - \bar{q}^2[m \pm (m^2 - \bar{q}^2)^{1/2}]. \tag{2.19}$$

Now from (2.11) and (2.13)

$$E_{\alpha\beta}E^{\alpha\beta} = 4(m^2 - \bar{q}^2). \tag{2.20}$$

One can easily see from the Rainich condition (2.10) and (2.20) that the eigenvalues of $E_{\alpha\beta}$ for a non-null field are degenerate and are given by

$$\pm(\tfrac{1}{4}E_{\alpha\beta}E^{\alpha\beta})^{1/2} = \pm(m^2 - \bar{q}^2)^{1/2}. \tag{2.21}$$

Eigenvalues (2.19) and (2.21) are consistent when

$$\bar{q} = 0, \tag{2.22}$$

i.e., $\bar{q}_\alpha = 0$. Further, (2.9) should satisfy the reality conditions

$$E_{\alpha\beta}u^\alpha u^\beta = m \geqslant 0, \tag{2.23a}$$

$$E_{\alpha\beta}v^\alpha v^\beta = M\mu^2 - \tfrac{1}{4}(M+3N) - 2\mu\beta q + 2\eta\sigma_{\alpha\beta}v^\alpha v^\beta \geqslant 0. \tag{2.23b}$$

Thus equivalence of two types of fields is possible under the conditions (2.4), (2.13), (2.15), (2.16), (2.22), and (2.23) for non-null fields (2.1).

III. NON-NULL FIELD ($\bar{q} = 0$)

Since $\bar{q}^\alpha$ is a spacelike vector $\bar{q} = 0$ implies $\bar{q}_\alpha = 0$. Hence from (2.8c)

$$q_\alpha = \mu\lambda N n_\alpha = \mu N(v_\alpha - \mu u_\alpha), \tag{3.1a}$$

$$q^2 = -q^\alpha q_\alpha = \mu^2\lambda^2 N^2. \tag{3.1b}$$

Thus from a magnetohydrodynamic fluid, one can construct the equivalent viscous fluid with heat flux in the following way. First, resolve v^α into two mutually perpendicular components; the timelike direction is along the fluid flow line (of the viscous fluid with heat flux), the spacelike direction is the heat flux direction, and the magnitude of heat flux is given by (3.1b). The heat flux is zero when $N = 0$, as is shown by Tupper in a different manner. Now Eq. (2.9) is written as

$$E_{\alpha\beta} = (M - N\mu^2)u_\alpha u_\beta - \tfrac{1}{4}(M-N)g_{\alpha\beta} + \psi_{\alpha\beta}. \tag{3.2}$$

Equations (2.13) and (2.16) are also modified as

$$\psi^{\alpha\beta}\psi_{\alpha\beta} - \tfrac{1}{2}\psi(M-N) = 3(M-N\mu^2)[\tfrac{1}{2}(M+N) - N\mu^2], \tag{3.3a}$$

$$h_{\alpha\beta}(M-N\mu^2)[\tfrac{1}{2}(M+N) - N\mu^2] + \tfrac{1}{2}(M-N)\psi_{\alpha\beta} - \psi_{\alpha\mu}\psi^{\mu}{}_{\beta} = 0. \tag{3.3b}$$

Also from (2.17a) one can see u^α to be an eigenvector of $E_{\alpha\beta}$ with eigenvalue m. Contracting (3.2) and (2.1c) by v^β and equating we have

$$S_\alpha = -\mu(M-N\mu^2)u_\alpha - [\tfrac{1}{2}\varphi - \tfrac{1}{4}(M-N)]v_\alpha - \psi_{\alpha\beta}v^\beta, \tag{3.4}$$

where $\varphi = E^\alpha E_\alpha + B^\alpha B_\alpha$.

IV. CHOICE OF TETRAD

We choose a suitable tetrad to study the problem in depth. Since u^α, n^α are two orthogonal vectors, we choose another spacelike unit vector ξ^α orthogonal to both u^α and n^α (i.e., $\xi^\alpha\xi_\alpha = -1$, $\xi_\alpha u^\alpha = \xi_\alpha n^\alpha = 0$) such that

$$E_\alpha = x_1 u_\alpha + y_1 n_\alpha + z_1\xi_\alpha,$$

and consider another spacelike unit vector l^α orthogonal to u^α, n^α, and ξ^α. Then

$$B_\alpha = x_2 u_\alpha + y_2 n_\alpha + z_2\xi_\alpha + r_2 l_\alpha.$$

Now $E_\alpha v^\alpha = B_\alpha v^\alpha = 0$, so that $y_1 = \mu x_1/\lambda$, $y_2 = \mu x_2/\lambda$. Therefore,

$$E_\alpha = x_1 u_\alpha + (\mu x_1/\lambda)n_\alpha + z_1\xi_\alpha, \tag{4.1a}$$

$$B_\alpha = x_2 u_\alpha + (\mu x_2/\lambda)n_\alpha + z_2\xi_\alpha + r_2 l_\alpha, \tag{4.1b}$$

and similarly

$$S_\alpha = x_3 u_\alpha + (\mu/\lambda)x_3 n_\alpha + z_3\xi_\alpha + r_3 l_\alpha. \tag{4.1c}$$

Further, $S_\alpha E^\alpha = S_\alpha B^\alpha = 0$. So from (4.1) we have

$$x_1x_3/\lambda^2 + z_1z_3 = 0, \tag{4.2a}$$

$$x_2x_3/\lambda^2 + z_2z_3 + r_2r_3 = 0, \tag{4.2b}$$

and the electromagnetic field energy density is

$$\varphi = ((x_1^2 + x_2^2)/\lambda^2 + z_1^2 + z_2^2 + r_2^2). \tag{4.3}$$

Now one can compare the components of $E_{\alpha\beta}u^\beta$ from (3.2) and (2.1c) and can have the following four equations:

$$(\tfrac{1}{2} - \mu^2)\varphi - 2\mu x_3 - (x_1^2 + x_2^2) = \tfrac{1}{4}(3M + N) - N\mu^2 = m, \tag{4.4a}$$

$$\mu\lambda\varphi + (\mu/\lambda)(x_1^2 + x_2^2) + [(\mu^2 + \lambda^2)/\lambda]x_3 = 0, \tag{4.4b}$$

$$z_1x_1 + z_2x_2 + \mu z_3 = 0, \tag{4.4c}$$

$$r_2x_2 + \mu r_3 = 0. \tag{4.4d}$$

In the same way, if we compute components of $E_{\alpha\beta}\, n^\beta$, $E_{\alpha\beta}\xi^\beta$, and $E_{\alpha\beta}l^\beta$ from the above two equations and compare, we have further the following six independent equations:

$$\psi_{\alpha\beta}n^\alpha n^\beta = -\tfrac{1}{4}(M - N) - \tfrac{1}{2}(1 + 2\lambda^2)\varphi - (\mu^2/\lambda^2)(x_1^2 + x_2^2) - 2\mu x_3, \tag{4.4e}$$

$$\psi_{\alpha\beta}\xi^\alpha\xi^\beta = -\tfrac{1}{4}(M - N) - \tfrac{1}{2}\varphi - (z_1^2 + z_2^2), \tag{4.4f}$$

$$\psi_{\alpha\beta}l^\alpha l^\beta = -\tfrac{1}{4}(M - N) - \tfrac{1}{2}\varphi - r_2^2, \tag{4.4g}$$

$$\psi_{\alpha\beta}\xi^\alpha n^\beta = -(\mu/\lambda)(x_1z_1 + x_2z_2) - \lambda z_3, \tag{4.4h}$$

$$\psi_{\alpha\beta}l^\alpha n^\beta = -(\mu/\lambda)x_2r_2 - \lambda r_3, \tag{4.4i}$$

$$\psi_{\alpha\beta}\xi^\alpha l^\beta = -r_2z_2. \tag{4.4j}$$

Now from (4.4a) and (4.4b) and from (4.4b) and (4.3),

$$m = -\tfrac{1}{2}\varphi - x_3/\mu, \tag{4.5a}$$

$$(\mu^2 + \lambda^2)x_3 = (z_1^2 + z_2^2 + r_2^2)\lambda^2\mu. \tag{4.5b}$$

Also from (4.4b) and (4.4e), (4.4c) and (4.4h), and (4.4d) and (4.4g),

$$\psi_{\alpha\beta}n^\alpha n^\beta = -\tfrac{1}{4}(M - N) + \tfrac{1}{2}\varphi + \mu x_3/\lambda^2, \tag{4.6a}$$

$$\psi_{\alpha\beta}\xi^\alpha \mathrm{n}^\beta = z_3/\lambda, \tag{4.6b}$$

$$\psi_{\alpha\beta}l^\alpha n^\beta = r_3/\lambda. \tag{4.6c}$$

Hence the tensor $\psi_{\alpha\beta}$ may be computed using (4.4e)–(4.4j) and (4.6) as

$$\begin{aligned}\psi_{\alpha\beta} = &-\tfrac{1}{4}(M - N)(n_\alpha n_\beta + \xi_\alpha\xi_\beta + l_\alpha l_\beta) \\ &+ \tfrac{1}{2}\varphi(n_\alpha n_\beta - \xi_\alpha\xi_\beta - l_\alpha l_\beta) \\ &+ \frac{\mu}{\lambda^2}x_3n_\alpha n_\beta + \frac{z_3}{\lambda}(n_\alpha\xi_\beta + n_\beta\xi_\alpha) \\ &+ \frac{r_3}{\lambda}(n_\alpha l_\beta + n_\beta l_\alpha) - r_2z_2(\xi_\alpha l_\beta + \xi_\beta l_\alpha) \\ &- (z_1^2 + z_2^2)\xi_\alpha\xi_\beta - r_2^2 l_\alpha l_\beta.\end{aligned} \tag{4.7}$$

Here the metric tensor is written in forms of the tetrad as

$$g_{\alpha\beta} = u_\alpha u_\beta - n_\alpha n_\beta - \xi_\alpha\xi_\beta - l_\alpha l_\beta, \tag{4.8a}$$

and the expansion is written with respect to u^α as (in terms of the tetrad)

$$\theta = u^\alpha{}_{;\alpha} = \theta_1{}^1 + \theta_2{}^2 + \theta_3{}^3, \tag{4.8b}$$

where $\theta_{(ab)} = u_{\alpha;\beta}\, e^\alpha{}_{((a)}\, e^\beta{}_{(b))}$. The tetrad components of the shear tensor are

$$\sigma_{ab} = \theta_{(ab)} - \tfrac{1}{3}\theta\eta_{ab}. \tag{4.8c}$$

Here a,b are tetrad indices.

Further, the Poynting vector is defined as

$$S_\alpha = \eta_{\alpha\beta\lambda\mu}v^\beta E^\lambda B^\mu, \tag{4.9}$$

and consequently the various components can be computed by using (4.1a) and (4.1b) as

$$x_3 = \lambda z_1r_2, \tag{4.10a}$$

$$z_3 = -(x_1/\lambda)r_2, \tag{4.10b}$$

$$r_3 = (x_1z_2 - x_2z_1)/\lambda. \tag{4.10c}$$

Substituting these in (4.2) we have

$$z_1x_1r_2 = 0, \tag{4.11a}$$

$$z_2x_1r_2 = 0. \tag{4.11b}$$

Thus three situations may arise (i) $z_1 = z_2 = 0$, (ii) $x_1 = 0$, and (iii) $r_2 = 0$.

A. Case 1: $z_1 = z_2 = 0$

In this case E^α is lying in the u^α-n^α plane. From (4.10) $x_3 = r_3 = 0$. Also from (4.5b) one can see that $r_2 = 0$ and hence from (4.10b) $z_3 = 0$ also. Thus the Poynting vector cannot exist in this case. Hence from (4.9) either the electric field or magnetic field is zero, or they are parallel. In any case, diagonal components of $\psi_{\alpha\beta}$ will exist and can be written from (4.7) as

$$\begin{aligned}\psi_{\alpha\beta} = &-\tfrac{1}{4}(M - N)(n_\alpha n_\beta + \xi_\alpha\xi_\beta + l_\alpha l_\beta) \\ &+ \tfrac{1}{2}\varphi(n_\alpha n_\beta - \xi_\alpha\xi_\beta - l_\alpha l_\beta).\end{aligned} \tag{4.12}$$

Here $\psi_{\alpha\beta}$ or the shear tensor is diagonal and two eigenvalues are the same, corresponding to eigenvectors ξ^α and l^α. Here ξ^α and l^α are to some extent arbitrary, only they are orthogonal to u^α and n^α and orthogonal to each other. Thus M can be taken from (4.5a) and (2.17c) as

$$M = -\tfrac{2}{3}\varphi + \tfrac{4}{3}(\mu^2 - \tfrac{1}{4})N. \tag{4.13a}$$

In this case tetrad components of the expansion $\theta_2{}^2 = \theta_3{}^3$. The value of η can be determined from (4.8c), (4.12), and (4.13a) as

$$2\eta = -(\tfrac{2}{3}\varphi - \tfrac{2}{3}\lambda^2N)/(\theta_1{}^1 - \tfrac{1}{3}\theta). \tag{4.13b}$$

Thus u^α is such that $\theta_1{}^1 > \tfrac{1}{3}\theta$ for $-\tfrac{2}{3}\varphi > \tfrac{2}{3}\lambda^2N$ or $\theta_1{}^1 < \tfrac{1}{3}\theta$ for $-\tfrac{2}{3}\varphi < \tfrac{2}{3}\lambda^2N$.

B. Case 2: $x_1 = 0$

Here the electric field is along ξ^α and from (4.10b), $z_3 = 0$, x_3 is given by (4.10a), and r_3 is, from (4.10c),

$$r_3 = -x_2z_1/\lambda. \tag{4.14}$$

Again from (4.4c) and (4.4d)

$$z_2x_2 = 0, \tag{4.15a}$$

$$x_2(\lambda r_2 - \mu z_1) = 0. \tag{4.15b}$$

Thus, either (i) $x_2 = 0$ or (ii) $z_2 = 0$ and $r_2 = \mu z_1/\lambda$.

1. Case (2a): $x_2 = 0$

From (4.14), $r_3 = 0$ also. Thus S^α is in the u^α-n^α plane and B^α is in the ξ^α-l^α plane. From (4.3) and (4.4b)

$$\varphi = -(z_1^2 + z_2^2 + r_2^2) \tag{4.16a}$$
$$= [(\mu^2 + \lambda^2)/\mu\lambda] z_1 r_2. \tag{4.16b}$$
Then the tensor $\psi_{\alpha\beta}$ can be written from (4.7) as
$$\begin{aligned}\psi_{\alpha\beta} = &-\tfrac{1}{4}(M-N)(n_\alpha n_\beta + \xi_\alpha\xi_\beta + l_\alpha l_\beta)\\ &+\tfrac{1}{2}\varphi(n_\alpha n_\beta - \xi_\alpha\xi_\beta - l_\alpha l_\beta)\\ &-[\mu^2\varphi/(\mu^2+\lambda^2)]n_\alpha n_\beta - (z_1^2+z_2^2)\xi_\alpha\xi_\beta\\ &- r_2^2 l_\alpha l_\beta - r_2 z_2(l_\alpha\xi_\beta + l_\beta\xi_\alpha),\end{aligned} \tag{4.17}$$
and M is calculated from (4.5a) and (4.16b) as
$$M = \tfrac{4}{3}N(\mu^2 - \tfrac{1}{4}) - \tfrac{2}{3}[\varphi/(\mu^2+\lambda^2)]. \tag{4.18a}$$
The coefficient of viscosity can be determined from (4.8c) and (4.17) as
$$2\eta = \tfrac{1}{2}(M-N)/(\tfrac{1}{3}\theta - \theta_1{}^1). \tag{4.18b}$$
Now $M > N$, so u^α is such that $\tfrac{1}{3}\theta > \theta_1{}^1$.

2. Case (2b): $z_2 = 0$ and $r_2 = \mu z_1/\lambda$

Here r_3 is given by (4.14) and from (4.10a), x_3 is
$$x_3 = \mu z_1^2, \tag{4.19}$$
and from (4.4b) or (4.3), φ is
$$\varphi = -[x_2^2/\lambda^2 + z_1^2(1 + \mu^2/\lambda^2)]. \tag{4.20}$$
The tensor $\psi_{\alpha\beta}$ and M are calculated from (4.7) and (4.5a) as
$$\begin{aligned}\psi_{\alpha\beta} = &-\tfrac{1}{4}(M-N)(n_\alpha n_\beta + \xi_\alpha\xi_\beta + l_\alpha l_\beta)\\ &+\tfrac{1}{2}\varphi(n_\alpha n_\beta - \xi_\alpha\xi_\beta - l_\alpha l_\beta)\\ &+(\mu^2/\lambda^2)z_1^2(n_\alpha n_\beta - l_\alpha l_\beta)\\ &-(x_2 z_1/\lambda^2)(n_\alpha l_\beta + n_\beta l_\alpha) - z_1^2\xi_\alpha\xi_\beta,\end{aligned} \tag{4.21}$$
$$M = \tfrac{4}{3}N(\mu^2 - \tfrac{1}{4}) - \tfrac{2}{3}\varphi - \tfrac{4}{3}z_1^2. \tag{4.22a}$$
The value of η in this case is obtained from (4.8c) and (4.21)
$$2\eta = [\tfrac{1}{2}(M-N) - x_2^2/\lambda^2]/(\tfrac{1}{3}\theta - \theta_1{}^1). \tag{4.22b}$$
Thus u^α is such that $\eta > 0$.

C. Case 3: $r_2 = 0$

In this case one can see from (4.10) that $x_3 = z_3 = 0$ and r_3 is given by (4.10c). Again from (4.5b) we have
$$z_1^2 + z_2^2 = 0, \tag{4.23}$$
i.e., $z_1 = z_2 = 0$. Hence from (4.10c), $r_3 = 0$ also, and the Poynting vector vanishes. So the case is the same as case 1.

Thus in all three cases one can construct a viscous fluid with a heat flux which is equivalent to a magnetohydrodynamic fluid.

V. NULL FIELD

For null electromagnetic fields $E_{\alpha\beta}E^{\alpha\beta} = 0$. So from (2.20) and (2.8d)
$$m^2 = \bar{q}^2 = q^2 - 2\mu N\beta q + \mu^2\lambda^2 N^2, \tag{5.1}$$
and (2.13) and (2.16) can be written as
$$\psi_{\alpha\beta}\psi^{\alpha\beta} - \tfrac{1}{2}(M-N)\psi = m^2 - \tfrac{3}{16}(M-N)^2, \tag{5.2}$$
$$\tfrac{1}{16}(M-N)^2 h_{\alpha\beta} - \tfrac{1}{2}(M-N)\psi_{\alpha\beta} + \psi_{\alpha\mu}\psi^\mu{}_\beta + \bar{q}_\alpha\bar{q}_\beta = 0. \tag{5.3}$$
Equations (4.1)–(4.3) and (4.9)–(4.11) are equally valid here. Further, for the null field $E_\alpha B^\alpha = 0$ and $E_\alpha E^\alpha = B^\alpha B_\alpha = \tfrac{1}{2}\varphi$, hence from (4.1)
$$x_1 x_2/\lambda^2 + z_1 z_2 = 0, \tag{5.4}$$
$$x_1^2/\lambda^2 + z_1^2 = x_2^2/\lambda^2 + z_2^2 + r_2^2 = -\tfrac{1}{2}\varphi. \tag{5.5}$$
Now we write q_α in component form,
$$q_\alpha = (\beta q/\lambda)n_\alpha + z_4\xi_\alpha + r_4 l_\alpha. \tag{5.6}$$
Comparing $E_{\alpha\beta}u^\beta$, $E_{\alpha\beta}n^\beta$, $E_{\alpha\beta}\xi^\beta$, and $E_{\alpha\beta}l^\beta$ from (2.1c) and (2.6) in a similar way (as is done in Sec. IV) we can get ten independent equations. Out of these the first four are
$$x_2 r_2 + \mu r_3 = -r_4, \tag{5.7a}$$
$$x_1 z_1 + x_2 z_2 + \mu z_3 = -z_4, \tag{5.7b}$$
$$\mu\lambda N - \frac{\beta q}{\lambda} = \mu\lambda\varphi + \frac{\mu^2+\lambda^2}{\lambda}x_3 + \frac{\mu}{\lambda}(x_1^2 + x_2^2), \tag{5.7c}$$
$$(\tfrac{1}{2} - \mu^2)\varphi - 2\mu x_3 - (x_1^2 + x_2^2) = \tfrac{1}{4}(3M + N) - N\mu^2 = m, \tag{5.7d}$$
and from the next six equations one can construct the shear tensor as
$$\begin{aligned}2\eta\sigma_{\alpha\beta} = &-\tfrac{1}{4}(M-N)(n_\alpha n_\beta + \xi_\alpha\xi_\beta + l_\alpha l_\beta)\\ &+\tfrac{1}{2}\varphi(n_\alpha n_\beta - \xi_\alpha\xi_\beta - l_\alpha l_\beta)\\ &+[\lambda^2 N - 2\mu x_3 + \mu^2(z_1^2 + z_2^2 + r_2^2)]n_\alpha n_\beta\\ &-(z_1^2+z_2^2)\xi_\alpha\xi_\beta - r_2^2 l_\alpha l_\beta\\ &-[(\mu/\lambda)(x_1 z_1 + x_2 z_2) + \lambda z_3](n_\alpha\xi_\beta\\ &+ n_\beta\xi_\alpha) - ((\mu/\lambda)x_2 r_2 + \lambda r_3)(l_\alpha n_\beta + l_\beta n_\alpha)\\ &- r_2 z_2(l_\alpha\xi_\beta + l_\beta\xi_\alpha).\end{aligned} \tag{5.8}$$
As (4.9)–(4.11) are equally valid, there will arise the same three situations.

A. Case 1: $z_1 = z_2 = 0$

From (5.4) and (5.5) one should have $x_2 = 0$ and
$$-\tfrac{1}{2}\varphi = x_1^2/\lambda^2 = r_2^2. \tag{5.9}$$
Again from (4.10), $x_3 = r_3 = 0$ and z_3 is given by (4.10b). Now from (5.7a), $r_4 = 0$ and from (5.7b)
$$z_4 = (\mu/\lambda)x_1 r_2. \tag{5.10}$$
Also, from (5.7c) and (5.9)
$$\beta q = \mu\lambda^2(N - \tfrac{1}{2}\varphi). \tag{5.11}$$
Thus q_α can be written as [using (5.6) and (5.9)–(5.11)]
$$q_\alpha = \mu\lambda(N - \tfrac{1}{2}\varphi)n_\alpha \mp \tfrac{1}{2}\mu\varphi\xi_\alpha, \tag{5.12}$$
where
$$q^2 = \mu^2\lambda^2 N^2 - \mu^2\lambda^2 N\varphi + \tfrac{1}{4}\mu^4\varphi^2. \tag{5.13}$$
Again from (5.1) and (5.11)
$$m^2 = q^2 - \mu^2\lambda^2 N^2 + \mu^2\lambda^2\varphi N. \tag{5.14}$$
Thus the heat flux cannot be zero in this case. Substituting q^2 from (5.13) we have
$$m^2 = \tfrac{1}{4}\mu^4\varphi^2. \tag{5.15}$$
Again from (5.7d)
$$m = -\tfrac{1}{2}\mu^2\varphi, \tag{5.16}$$

or

$$M = \tfrac{4}{3}N(\mu^2 - \tfrac{1}{4}) - \tfrac{2}{3}\mu^2\varphi. \tag{5.16a}$$

Thus Eqs. (5.15) and (5.16) are consistent. Also, the shear tensor is obtained from (5.8) and (5.9) as

$$\begin{aligned}2\eta\sigma_{\alpha\beta} = & -\tfrac{1}{4}(M-N)(n_\alpha n_\beta + \xi_\alpha\xi_\beta + l_\alpha l_\beta) \\ & + \tfrac{1}{2}\varphi(n_\alpha n_\beta - \xi_\alpha\xi_\beta) + (\lambda^2N - \tfrac{1}{2}\mu^2\varphi)n_\alpha n_\beta \\ & \mp \tfrac{1}{2}\varphi\lambda(n_\alpha\xi_\beta + n_\beta\xi_\alpha).\end{aligned} \tag{5.16b}$$

The value of η can be evaluated from (4.8c) and (5.16b) as

$$\begin{aligned}2\eta &= \frac{-\tfrac{1}{4}(M-N) - \tfrac{1}{2}\varphi}{\theta_1{}^1 - \tfrac{1}{3}\theta} = \frac{-\tfrac{1}{4}(M-N) - \tfrac{1}{2}\varphi}{\tfrac{1}{3}\theta - \theta_2{}^2} \\ &= \frac{-\tfrac{1}{4}(M-N)}{\tfrac{1}{3}\theta - \theta_3{}^3}.\end{aligned} \tag{5.16c}$$

Since $M > N$, u^α is such that $\theta_3{}^3 > \tfrac{1}{3}\theta$.

B. Case 2: $x_1 = 0$

Here from (5.4) and (5.5), $z_2 = 0$ (since $z_1 \neq 0$, whence φ will be zero) and

$$-\tfrac{1}{2}\varphi = z_1{}^2 = x_2{}^2/\lambda^2 + r_2{}^2. \tag{5.17}$$

Again from (4.10), $z_3 = 0$, x_3 is given by (4.10a), and

$$r_3 = -z_1x_2/\lambda. \tag{5.18}$$

Thus from (5.7a) and (5.7b)

$$r_4 = -x_2r_2 + (\mu/\lambda)z_1x_2, \tag{5.19}$$

and $z_4 = 0$. Also from (5.7c)

$$\beta q = \mu\lambda^2(N + z_1{}^2 + r_2{}^2) - (\mu^2 + \lambda^2)\lambda z_1r_2. \tag{5.20}$$

Hence q_α can be written as

$$\begin{aligned}q_\alpha = & [\mu\lambda(N + z_1{}^2 + r_2{}^2) - (\mu^2 + \lambda^2)z_1r_2]n_\alpha \\ & + [(\mu/\lambda)z_1x_2 - x_2r_2]l_\alpha.\end{aligned} \tag{5.21}$$

Again substituting (5.20) in (5.1)

$$\begin{aligned}m^2 = & q^2 - 2\mu^2\lambda^2N(z_1{}^2 + r_2{}^2) \\ & + 2\mu\lambda Nz_1r_2(\mu^2 + \lambda^2) - \mu^2\lambda^2N^2.\end{aligned} \tag{5.22}$$

Further, from (5.7d)

$$m = -\tfrac{1}{2}\mu^2\varphi + \lambda^2r_2{}^2 - 2\mu\lambda z_1r_2 = (\mu z_1 - \lambda r_2)^2, \tag{5.23a}$$

or

$$M = \tfrac{4}{3}N(\mu^2 - \tfrac{1}{4}) + \tfrac{4}{3}(\mu z_1 - \lambda r_2)^2. \tag{5.23b}$$

Substituting m from (5.22) one can calculate q^2. Here, q^2 so calculated is consistent with that obtained from (5.21). The shear tensor is also obtained from (5.8) as

$$\begin{aligned}2\eta\sigma_{\alpha\beta} = & -\tfrac{1}{4}(M-N)(n_\alpha n_\beta + \xi_\alpha\xi_\beta + l_\alpha l_\beta) \\ & + \tfrac{1}{2}\varphi(n_\alpha n_\beta - l_\alpha l_\beta) + [\lambda^2N - 2\mu\lambda z_1r_2 \\ & + \mu^2(z_1{}^2 + r_2{}^2)]n_\alpha n_\beta - r_2{}^2l_\alpha l_\beta \\ & - ((\mu/\lambda)x_2r_2 - z_1x_2)(l_\alpha n_\beta + l_\beta n_\alpha).\end{aligned} \tag{5.24a}$$

The coefficient of viscosity is calculated from (4.8c) and (5.24a) as

$$2\eta = \frac{\tfrac{1}{4}(M-N) + \tfrac{1}{2}\varphi + r_2{}^2}{\tfrac{1}{3}\theta - \theta_1{}^1} = \frac{-\tfrac{1}{4}(M-N)}{\tfrac{1}{3}\theta - \theta_2{}^2}. \tag{5.24b}$$

Since $M > N$, u^α is such that $\theta_2{}^2 > \tfrac{1}{3}\theta$.

C. Case 3: $r_2 = 0$

Here, from (4.10), $x_3 = z_3 = 0$, but r_3 is given by (4.10c). The Poynting vector is directed along l^α. The conditions (5.4) and (5.5) give

$$x_1x_2/\lambda^2 + z_1z_2 = 0, \tag{5.25a}$$

$$-\tfrac{1}{2}\varphi = x_1{}^2/\lambda^2 + z_1{}^2 = x_2{}^2/\lambda^2 + z_2{}^2. \tag{5.25b}$$

From (5.7a)–(5.7c)

$$r_4 = -(\mu/\lambda)(x_1z_2 - x_2z_1), \tag{5.26a}$$

$$z_4 = -(x_1z_1 + x_2z_2), \tag{5.26b}$$

$$\beta q = \mu\lambda^2(N + z_1{}^2 + z_2{}^2). \tag{5.27}$$

Thus one can have q_α to be

$$\begin{aligned}q_\alpha = & \mu\lambda(N + z_1{}^2 + z_2{}^2)n_\alpha - (x_1z_1 + x_2z_2)\xi_\alpha \\ & - (\mu/\lambda)(x_1z_2 - x_2z_1)l_\alpha,\end{aligned} \tag{5.28}$$

and from (5.7d)

$$m = -\tfrac{1}{2}\varphi + \lambda^2(z_1{}^2 + z_2{}^2), \tag{5.29a}$$

or

$$M = \tfrac{4}{3}N(\mu^2 - \tfrac{1}{4}) - \tfrac{2}{3}\varphi + \tfrac{4}{3}\lambda^2(z_1{}^2 + z_2{}^2). \tag{5.29b}$$

Again from (5.1) and (5.27)

$$m^2 = q^2 - \mu^2\lambda^2N^2 - 2\mu^2\lambda^2N(z_1{}^2 + z_2{}^2). \tag{5.30}$$

Thus q^2 cannot be zero in this case. If values of q^2 and m^2 are calculated from (5.28) and (5.29) and are substituted, one can have, using (5.25),

$$z_1{}^2 + z_2{}^2 = -\tfrac{1}{2}\varphi. \tag{5.31}$$

So from (5.25) one should have either

$$z_1 = x_2/\lambda \quad \text{and} \quad z_2 = -x_1/\lambda \tag{5.32a}$$

or

$$z_1 = -x_2/\lambda \quad \text{and} \quad z_2 = x_1/\lambda. \tag{5.32b}$$

In the two cases from (5.28)

$$q_\alpha = \mu\lambda(N - \tfrac{1}{2}\varphi)n_\alpha \mp \tfrac{1}{2}\mu\varphi l_\alpha, \tag{5.33}$$

and from (5.29)

$$m = -\tfrac{1}{2}\mu^2\varphi, \tag{5.34a}$$

or

$$M = \tfrac{4}{3}N(\mu^2 - \tfrac{1}{4}) - \tfrac{2}{3}\mu^2\varphi. \tag{5.34b}$$

The shear tensor is from (5.8)

$$\begin{aligned}2\eta\sigma_{\alpha\beta} = & -\tfrac{1}{4}(M-N)(n_\alpha n_\beta + \xi_\alpha\xi_\beta + l_\alpha l_\beta) \\ & + \tfrac{1}{2}\varphi(n_\alpha n_\beta - l_\alpha l_\beta) + (\lambda^2N - \tfrac{1}{2}\mu^2\varphi)n_\alpha n_\beta \\ & \mp \tfrac{1}{2}\lambda\varphi(l_\alpha n_\beta + l_\beta n_\alpha).\end{aligned} \tag{5.35a}$$

The coefficient of viscosity is determined as

$$2\eta = \frac{\tfrac{1}{4}(M-N) + \tfrac{1}{2}\varphi}{\tfrac{1}{3}\theta - \theta_1{}^1} = \frac{-\tfrac{1}{4}(M-N)}{\tfrac{1}{3}\theta - \theta_2{}^2}. \tag{5.35b}$$

So here also u^α is such that $\theta_2{}^2 > \tfrac{1}{3}\theta$.

Thus in all the three cases for a null field q^α cannot be zero. From the magnetohydrodynamic field, an equivalent viscous fluid with a heat flux is obtainable.

VI. CASE $v^{\alpha} = u^{\alpha}$

In this case $\mu = 1$ and $\lambda = 0$. So from (2.6)

$$E_{\alpha\beta} = (M-N)u_\alpha u_\beta - \tfrac{1}{4}(M-N)g_{\alpha\beta} + 2\eta\sigma_{\alpha\beta} + q_\alpha u_\beta + q_\beta u_\alpha. \tag{6.1}$$

Thus all the equations in Sec. II are also valid with $\mu = 1$ and $\lambda = 0$. Also $\psi_{\alpha\beta} = 2\eta\sigma_{\alpha\beta}$, $\psi = 0$, and $\bar{q}_\alpha = q_\alpha$. Proceeding in a similar way as in Sec. II, we have for a non-null field, instead of (2.22), $q = 0$. That is, a heat flux is not admissible. For a null field $E_{\alpha\beta}E^{\alpha\beta} = 0$ and (2.20) and (2.17c), we have

$$m^2 = q^2 = \tfrac{9}{16}(M-N)^2. \tag{6.2}$$

Again comparing $E_{\alpha\beta}u^\beta$ from (6.1) and (2.1c) we have

$$q_\alpha = -S_\alpha, \tag{6.3a}$$

$$(M-N) = -\tfrac{2}{3}\varphi. \tag{6.3b}$$

Also from (6.2) and (6.3a)

$$q^2 = \tfrac{1}{4}\varphi^2 = S^2, \tag{6.4}$$

and the shear tensor obtained from (6.1) and (2.1c) using (6.3) is

$$2\eta\sigma_{\alpha\beta} = \tfrac{1}{3}\varphi h_{\alpha\beta} - (E_\alpha E_\beta + B_\alpha B_\beta). \tag{6.5a}$$

Further, for null fields $E_\alpha B^\alpha = 0$ and $E_\alpha E^\alpha = B^\alpha B_\alpha = \tfrac{1}{2}\varphi$, so from (6.4a)

$$8\eta^2\sigma^2 = \tfrac{1}{6}\varphi^2. \tag{6.5b}$$

Also, the same relation is obtained from (2.13) and (2.17c). Equation (6.4) ensures that the null field must admit heat flux. If one chooses unit vectors along E^α, B^α, and S^α, then along with the fluid flow vector, they form an orthonormal set of tetrad. In this tetrad system one can see from (6.5a) that the shear tensor has eigenvalues $(0, -(1/12\eta)\varphi, -(1/12\eta)\varphi, +(1/6\eta)\varphi)$.

Using the above tetrad system the coefficient of viscosity can be calculated as

$$2\eta = \frac{-\varphi/6}{\theta_1{}^1 - \tfrac{1}{3}\theta} = \frac{-\varphi/6}{\theta_2{}^2 - \tfrac{1}{3}\theta} = \frac{\varphi/3}{\theta_3{}^3 - \tfrac{1}{3}\theta}. \tag{6.6}$$

Here $\theta_1{}^1 = \theta_2{}^2$. Also $\theta_1{}^1 > \tfrac{1}{3}\theta$ and $\theta_3{}^3 < \tfrac{1}{3}\theta$ for η to be positive.

Thus the equivalent conditions for the two fluids with parallel flow lines are that the electromagnetic field must be null and condition (6.4) must also be satisfied along with $\theta_1{}^1 > \tfrac{1}{3}\theta$ and $\theta_1{}^1 = \theta_2{}^2$.

VII. CONCLUDING REMARKS

It has been shown that the equivalence of two types of fluids (2.1b) and (2.2b) is possible. If the flow lines of two fluids are not parallel, then for non-null electromagnetic fields, the equivalence is possible only when v^α can be resolved into a timelike component along the flow lines of the viscous fluid and a spacelike component in the direction of the heat flux vector; and for null fields, the heat flux must be nonzero. If the two fluid flows are parallel, then the heat flux is nonzero when the electromagnetic field is null along with $S^2 = \tfrac{1}{4}\varphi^2$, and the shear tensor must have degenerate eigenvalues. In other words, one can construct a viscous fluid with heat flux, which is equivalent to a magnetohydrodynamic fluid (2.1b), but u^α must be so chosen that $\eta > 0$. The construction with a parallel fluid flow vector is not always possible. This is possible only when the electromagnetic field is null with $S^2 = \tfrac{1}{4}\varphi^2$ and expansion components along E^α and B^α are equal and are greater than the average expansion. In the Appendix a few examples of the construction of a viscous fluid (2.2b) are attempted, using the solutions of Dunn and Tupper[5] for a magnetohydrodynamic fluid.

ACKNOWLEDGMENT

The authors wish to thank an anonymous referee for some valuable suggestions.

APPENDIX: SOME EXAMPLES

To find an equivalent viscous fluid (2.2b), as an example, we consider the type I cosmological model[5] with metric

$$ds^2 = dt^2 - t^{2a}dx^2 - t^{2b}(dy^2 + dz^2), \tag{A1}$$

which satisfies (2.1) for a non-null electromagnetic field with $v^\alpha = (1,0,0,0)$ provided $b > a$, $a - 2b + 1 < 0$, and $b < \tfrac{1}{2}$. From their paper one can calculate

$$N = \bar{p} + \bar{\rho} = 2b(1 + a - b)t^{-2} \tag{A2}$$

and the electric and magnetic fields are of the form

$$E_\alpha = E_1\delta_\alpha{}^1, \quad B_\alpha = B_1\delta_\alpha{}^1. \tag{A3}$$

Construction of a viscous fluid of the form (2.2b) is possible. Consider

$$u^\alpha \equiv [t^a, (1 - t^{-2a})^{1/2}, 0, 0]. \tag{A4a}$$

Then $\mu = u^0$ and $\lambda^2 = (u^0)^2 - 1$. Hence n^α is computed from (2.7a)

$$n^\alpha \equiv (-\lambda, -u^0u^1/\lambda, 0, 0). \tag{A4b}$$

Now $x_1 = E_1u^1$ and $x^2 = E_2u^2$, hence from (A3), (A4), and (4.1) one can have $z_1 = z_2 = r_2 = 0$. So the situation is under case 1 in Sec. IV. Again $\theta_1{}^1$ and θ are readily calculated as

$$\theta_1{}^1 = 2at^{a-1}, \tag{A5a}$$

$$\theta = 2(a + b)t^{a-1}. \tag{A5b}$$

So

$$\theta_1{}^1 - \tfrac{1}{3}\theta = -\tfrac{2}{3}(a - 2b)t^{a-1}. \tag{A5c}$$

Thus from (4.13) one can show that

$$2\eta = (2b - a)^{-1}[(b - a)(1 - a)t^{-(1+a)} + 2b(a - b + 1)t^{a-1}] \tag{A6}$$

and is positive. Heat flux and shear tensor are easily calculated from (3.1a) and (4.7).

The same metric (A1) also satisfies (2.1) for a null electromagnetic field but in this case $a + b = 1$ and $\tfrac{1}{2} < a < 1$ with

$$v^\alpha = (\sqrt{8a})^{-1}[(2a + 1), (2a - 1)t^{-a}, 0, 0], \tag{A7}$$

and the electric and magnetic fields are of the form

$$E_\alpha = (0, 0, E_2, E_3), \quad B_\alpha = (0, 0, B_2, B_3). \tag{A8}$$

If we choose u^α to be identical to v^α, η becomes negative and such a choice is not possible.[1] Again, if we consider

$$u^\alpha = [\mu(2a+1)/\sqrt{8a},\ \mu(2a-1)t^{-a}/\sqrt{8a},\ \lambda t^{-b},\ 0] \tag{A9}$$

(e.g., $\mu = \sqrt{2}$ and $\lambda = 1$), η is still negative. The viscous fluid interpretation is not possible in this case.

Thus one may conclude that a metric representing two different types of magnetohydrodynamic fluid may not always have a viscous fluid interpretation in both the cases. The same metric can have two different reinterpretations, one of which is physically acceptable while the other is not.

[1]B. O. J. Tupper, J. Math. Phys. **22**, 2666 (1981).

[2]A. K. Raychaudhuri and S. K. Saha, J. Math. Phys. **22**, 2237 (1981).

[3]A. K. Raychaudhuri and S. K. Saha, J. Math. Phys. **23**, 2554 (1982).

[4]A. Lichnerowicz, *Relativistic Hydrodynamics and Magnetohydrodynamics* (Benjamin, New York, 1967).

[5]K. A. Dunn and B. O. J. Tupper, Astrophys. J. **235**, 307 (1980).

Confined gravitational fields produced by anisotropic fluids

L. Herrera[a),b)]
Departamento de Física, Facultad de Ciencias, Universidad Central de Venezuela, Caracas 1051, Venezuela
J. Ponce de León
Departamento de Física, Facultad de Ciencias, Universidad Central de Venezuela, Caracas 1051, Venezuela and Centro de Física, I.V.I.C., Apartado 1827, Caracas 1010-A, Venezuela

(Received 21 February 1985; accepted for publication 25 April 1985)

A family of solutions of the Einstein equations for a spherically symmetric distribution of anisotropic matter is presented, which can be matched with the flat (Minkowskian) space-time on the boundary of the matter, although the energy density and stresses are nonvanishing within the sphere.

I. THE FIELD EQUATIONS

Let us consider a nonstatic distribution of matter represented by an anisotropic fluid and which is spherically symmetric.

In comoving coordinates the line element may be written as[1]

$$ds^2 = e^{\nu}\, dt^2 - e^{\lambda}\, dr^2 - e^{\mu} d\Omega^2, \tag{1}$$

with

$$d\Omega^2 = d\theta^2 + \sin^2\theta\, d\phi^2, \quad x^{0,1,2,3} \equiv t,r,\theta,\phi, \tag{2}$$

where λ, ν, and μ are functions of r and t. For the energy momentum tensor we have the usual expression

$$T^{\mu}_{\nu} = (\rho + p_\perp) U^{\mu} U_{\nu} - p_\perp \delta^{\mu}_{\nu} + (p_r - p_\perp) X^{\mu} X_{\nu}, \tag{3}$$

with ρ, p_r, and U^{μ} denoting the energy density, the pressure in the direction of X_{μ}, and the four-velocity of the fluid, respectively, and X_{μ} and $p_\perp$ denoting a unit spacelike vector (in the radial direction) orthogonal to U^{μ} and the pressure on the two-space orthogonal to X_{μ}. Also, since we are in a comoving frame,

$$U^{\mu} = \delta^{\mu}_{0}\, e^{-\nu/2}. \tag{4}$$

Next, we shall assume that the space-time admits a one-parameter group of conformal motions, i.e.,

$$\underset{\xi}{L}\, g_{\alpha\beta} = \psi g_{\alpha\beta}, \tag{5}$$

where the left-hand side is the Lie derivative of the metric tensor and ψ is an arbitrary function of the coordinates.

We shall further restrict the vector field ξ^{α}, by demanding

$$\xi^{\alpha} U_{\alpha} = 0. \tag{6}$$

Then it can be shown (for details, see Refs. 2 and 3) that the metric functions ν, λ, and μ become

$$e^{-\nu/2} = e^{-\mu/2} = e^{f(t)/2}[h_1(r) + h_2(t)], \tag{7}$$

$$e^{-\lambda/2} = h_1(r) + h_2(t), \tag{8}$$

where h_1, h_2, and f are three unknown functions of their arguments.

The function ψ will be

$$\psi = A h_1'(r) e^{\lambda/2}, \tag{9}$$

where A is a constant and a prime denotes differentiation with respect to r.

Thus, the line element (1) reduces to

$$ds^2 = R^2(r,t)[dt^2 - e^{f(t)} dr^2 - d\Omega^2], \tag{10}$$

with

$$R(r,t) \equiv e^{-f(t)/2}/[h_1(r) + h_2(t)].$$

The Einstein field equations corresponding to this line element are

$$\begin{aligned}-8\pi T^1_1 = 8\pi p_r = [&3h_1'^2(r) - 3\dot h_2^2(t) e^{f(t)}] \\ &+ e^{-\lambda/2} e^{f(t)}[2\ddot h_2(t) - \dot h_2(t)\dot f(t)] \\ &+ e^{-\lambda} e^{f(t)}[\ddot f(t) - \dot f^2(t)/4 - 1],\end{aligned} \tag{11}$$

$$\begin{aligned}-8\pi T^2_2 = 8\pi p_\perp = [&3h_1'^2(r) - 3\dot h_2^2(t) e^{f(t)}] \\ &- 2e^{-\lambda/2}[h_1''(r) - \ddot h_2(t) e^{f(t)}] + \tfrac{1}{2}\ddot f(t) e^{-\lambda} e^{f(t)},\end{aligned} \tag{12}$$

$$\begin{aligned}8\pi T^0_0 = 8\pi\rho = -&[3h_1'^2(r) - 3\dot h_2^2(t) e^{f(t)}] \\ &+ 2e^{-\lambda/2}[h_1''(r) + \dot h_2(t)\dot f(t) e^{f(t)}] \\ &+ e^{-\lambda} e^{f(t)}[\dot f^2/4 + 1]\end{aligned} \tag{13}$$

(dots denote differentiation with respect to t).

II. THE JUNCTION CONDITIONS AND THE EQUATION OF THE BOUNDARY SURFACE

Next, we shall match the line element (10) with the flat space-time on the boundary of the matter for any possible choice of the functions $h_1(r)$, $h_2(t)$, and $f(t)$. We recall that two regions of the space-time are said to match across a separating hypersurface (say S) if the first and second fundamental form are continuous across S (Darmois conditions). Now, the line element outside the sphere will be given, in coordinates T, R, θ, ϕ, by

$$ds_E^2 = dT^2 - dR^2 - R^2\, d\Omega^2, \tag{14}$$

where the subscript E stands for exterior, and

---

a) This work was completed when one of the authors (L.H.) was on leave from Universidad Central de Venezuela at the Mathematics Department of the University of California, Berkeley.

b) Postal address: Apartado Postal 80793, Caracas 1080-A, Venezuela.

 0022-2488/85/112847-03$02.50

$$d\Omega^2 = d\theta^2 + \sin^2\theta\, d\phi^2.$$

In these coordinates the equation of the boundary takes the form

$$R_b = R_b(T), \tag{15}$$

where b stands for boundary.

Then the induced metric on the boundary surface (from the outside) is

$$(ds^2)_b^+ = \left[1 - \left(\frac{dR_b}{dT}\right)^2\right] dT^2 - R_b^2\, d\Omega^2 \tag{16}$$

and the corresponding line element on the boundary from the inside reads

$$(ds^2)_b^- = \{e^{-f(t)}/[h_1(r_0) + h_2(t)]^2\}[dt^2 - d\Omega^2], \tag{17}$$

where we have used (10) and the fact that the equation of the boundary in the comoving coordinates (t, r, θ, ϕ) reads

$$r = r_0 = \text{const.}$$

Then, demanding the first fundamental form to be continuous across the boundary, we get at once

$$R_b(T) = e^{-f(t)/2}/[h_1(r_0) + h_2(t)] \tag{18}$$

and

$$dT\left[1 - \left(\frac{dR_b}{dT}\right)^2\right]^{1/2} = R_b\, dt. \tag{19}$$

Next, it can be shown that the continuity of the second fundamental form across the boundary surface is equivalent to the continuity of the mass function and the radial pressure (across the same boundary surface).[4-7] Since we are matching the line element (10) with the Minkowskian metric, then

$$M\,(\text{total mass}) \equiv m(r_0,t) = 0, \tag{20}$$

$$P_r(r_0,t) = 0, \tag{21}$$

where $m(r,t)$ is the mass function introduced by Misner and Sharp[4]

$$2m(r,t) \equiv e^{\mu/2}\left[1 + e^{-\nu}\left(\frac{\partial e^{\mu/2}}{\partial t}\right)^2 - e^{-\lambda}\left(\frac{\partial e^{\mu/2}}{\partial r}\right)^2\right]. \tag{22}$$

Using (7), (11), and (22), Eqs. (20) and (21) become

$$2\ddot{R}_b R_b - \dot{R}_b^2 = 3h_1'^2(r_0) R_b^4 - R_b^2, \tag{23}$$

$$\dot{R}_b^2 = -R_b^2 + h_1'^2(r_0) R_b^4. \tag{24}$$

Since the first integral of (23) is given by (24), we only have to integrate this last equation. We obtain

$$R_b = 1/h_1'(r_0)\cos(t - t_0), \tag{25}$$

with $t_0 = \text{const.}$

III. THE MODELS

For sake of simplicity we shall restrict further our solutions with the choice $h_2(t) = 0$. With this condition, and using (18), we obtain, from (11)–(13),

$$8\pi p_r = 3h_1'^2(r) - 3\omega^2 h_1^2(r), \tag{26}$$

$$8\pi p_\perp = 3h_1'^2(r) - 2h_1(r)h_1''(r) - \omega^2 h_1^2(r), \tag{27}$$

$$8\pi\rho = -3h_1'^2(r) + 2h_1(r)h_1''(r) + \omega^2 h_1^2(r), \tag{28}$$

with

$$\omega^2 \equiv h_1'^2(r_0)/h_1^2(r_0).$$

Next, to display an explicit solution we still have to specify the function $h_1(r)$ (which is equivalent to the specification of the function ψ). In this paper we shall guess the function $h_1(r)$ from the condition of the positiveness of ρ.

A sufficient condition to meet this last requirement is

$$-3h_1'^2(r) + 2h_1(r)h_1''(r) = 0, \tag{29}$$

from which

$$h_1(r) = 1/(Cr + B)^2, \tag{30}$$

With C and B constants. Then Eqs. (26)–(28) become

$$8\pi p_r = \frac{12C^2}{(Cr + B)^4 (Cr_0 + B)^2}\left[\frac{(Cr_0 + B)^2}{(Cr + B)^2} - 1\right], \tag{31}$$

$$8\pi p_\perp = \frac{-4C^2}{(Cr_0 + B)^2 (Cr + B)^4}, \tag{32}$$

$$8\pi\rho = \frac{4C^2}{(Cr_0 + B)^2 (Cr + B)^4}. \tag{33}$$

For the line element we get

$$ds^2 = \frac{(Cr + B)^4}{\omega^2\cos^2(t - t_0)}[dt^2 - \omega^2\cos^2(t - t_0)dr^2 - d\Omega^2]. \tag{34}$$

In order to ensure that the energy density is larger or (at least) equal to the radial pressure, we may choose $Cr_0 = 1$ and $B = 3 + 2\sqrt{3}$. Then it is easily seen that

$$\rho \geqslant p_r,$$

where the equality holds for $r = 0$. It should be noted, however, that for the election of the constants C and B above, $R(r,t)$ does not satisfy the regularity condition

$$R(0,t) = 0.$$

In fact, we have excluded the center of symmetry $R = 0$, and R varies in the interval $[3R_b/4, R_b]$. In order to overcome this inconvenience we may choose $B = 0$. Then the regularity condition is satisfied and the energy density will be larger than or equal to the radial pressure in the region $r_0\sqrt{3}/2 \leqslant r \leqslant r_0$ (we will call this region I). We can now match our solution in region I with any other conformally symmetric solution in the region $0 \leqslant r \leqslant r_0\sqrt{3}/2$ (we will call this region II). It is important to remark that in matching the two regions (I and II) we do not require the mass function to vanish at the inner boundary $r = r_0\sqrt{3}/2$ (of course it should be continuous across that surface).

We would like to finish with the following remarks.

(a) Configurations of the kind we have just discussed have been suggested by Zel'dovich and Novikov some years ago.[8] Also, a solution with a vanishing gravitational mass has been found, in a completely different context, by Gross and Perry.[9]

(b) It should be understood that the configuration above would represent the late stages of a self-similar evolution scenario, when the sphere has radiated away all its gravitational mass. Since we are considering adiabatic evolution, it is obvious that our solutions cannot describe the system during the radiation process.

(c) It is not difficult to prove that all the configurations above [with $h_2(t) = 0$] do not admit a one-parameter group of homothetic motions unless $\rho = p_r = p_\perp = 0$. In other words the function ψ is not a constant except for the trivial solution.

ACKNOWLEDGMENTS

We would like to thank A. Taub for helpful comments and suggestions.

This work was partially supported by C.O.N.I.C.I.T. and the National Science Foundation.

[1]L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Addison–Wesley, Reading MA, 1951), p. 311.
[2]L. Herrera and J. Ponce de León, J. Math. Phys. **26**, 778 (1985).
[3]L. Herrera and J. Ponce de León, J. Math. Phys. **26**, 2018 (1985).
[4]C. W. Misner and D. M. Sharp, Phys. Rev. B **136**, 571 (1964).
[5]E. H. Robson, Ann. Inst. H. Poincaré, **16**, 41 (1972).
[6]M. E. Cahill and A. H. Taub, Commun. Math. Phys. **21**, 1 (1971).
[7]M. E. Cahill and G. C. McVittie, J. Math. Phys. **11**, 1392 (1970).
[8]Ya. B. Zel'dovich and I. D. Novikov, *Relativistic Astrophysics* (Chicago U. P., Chicago, 1971).
[9]D. J. Gross and M. J. Perry, Nucl. Phys. B **226**, 29 (1983).

Statistical mechanics in a conformally static setting. II

Henry E. Kandrup
Center for Relativity and Center for Theoretical Physics, University of Texas, Austin, Texas 78712 and Center for Theoretical Physics, University of Maryland, College Park, Maryland 20742 and Institut d'Astrophysique, 75014 Paris, France

(Received 19 December 1984; accepted for publication 24 April 1985)

This paper examines the statistical mechanics of a collection of N identical, classical point masses, which interact relativistically via a simple scalar field in a conformally static background geometry. The model system considered here should be representative of any system in which the particle and field equations are both linear. Attention focuses first upon the formulation of exact equations for the evolution of appropriately defined reduced distribution functions and the interpretation of these relations. In particular, a projection operator formalism is used to derive exact coupled equations for the evolution of the irreducible one-particle and one-oscillator distributions, which contain no explicit reference to more complicated particle–oscillator correlation functions. Attention focuses also upon the issue of how the analysis would be further complicated by allowing for nonlinear effects, e.g., in the particle equations of motion. The subtleties that arise in this case serve to indicate the limitations of the kinetic theory of self-gravitating systems developed by Israel and Kandrup, which entails the consideration of particle and field equations linearized about some (possibly highly nontrivial!) background solution.

I. INTRODUCTION

The first paper in this series,[1] hereafter referred to as Paper I, began an investigation of the statistical mechanics of a collection of N identical particles interacting relativistically via a simple scalar field or a linearized gravitational interaction in a fixed, conformally static, background space-time. One principle objective was the formulation of a "complete" many-particle and oscillator description, in which the system is characterized by a distribution function μ that involves the degrees of freedom of both the particles and the fields and satisfies an appropriate Liouville, or conservation, equation. The other principle objective was to understand how, in a simple approximation, this complete description implies a simpler mean field theory. Unlike much of the earlier work in a similar vein,[2–7] the analysis in Paper I broke manifest covariance by implementing the "natural" $3 + 1$ splitting into space plus time suggested by the conformal time translation symmetry. This procedure, albeit displeasing aesthetically, is certainly legitimate mathematically, and, significantly, eliminates many potential ambiguities and questions which might otherwise arise.[8,9]

The objective of this second paper is to investigate in greater detail various properties of reduced distribution functions derived from μ. Principally for the sake of computational simplicity, attention here will focus exclusively upon the linearized scalar interaction and its nonlinear generalization considered by Hakim[2,3] and Kandrup.[1,5] Aside from issues of gauge, the only additional complications that arise in the gravitational case involve the more intricate couplings buried in the field equations for a second rank tensor field.

The questions to be addressed include the following.

(1) In what ways, conceptually and otherwise, will the sort of description presented here differ from the more conventional description of a Newtonian or special relativistic system?

(2) How might one proceed to formulate useful exact equations for the evolution of reduced distributions, and how, physically, are these equations to be interpreted?

(3) How would the situation be further complicated by allowing for equations of motion that involve the fundamental fields in a nonlinear way?

A satisfactory answer to this third question is of particular importance for the understanding of self-gravitating systems. In this case, the "true" particle equations of motion, as well as the field equations, are nonlinear, although it may well be legitimate in some approximation to linearize about some "background" solution.[4] This linearization is in fact crucial in the simplest attempts to transcend a naive mean field description for self-gravitating systems. Moreover, as will be evidenced below in Sec. IV, an allowance for nonlinear equations already leads to complications at the level of the mean field theory (which, however, are customarily ignored).[10,11]

The program of this paper is as follows. Section II recalls and extends the statistical mechanical description formulated in Paper I. The "complete" description is used as a starting point for the formulation of a relativistic analog of the BBGKY hierarchy of equations, and this in turn is used to extract the simple mean field description. Section III uses a projection operator formalism to demonstrate explicitly how one may derive exact coupled equations for the evolution of reduced one-particle and one-oscillator distributions, which contain no explicit reference to higher-order correlation functions. The general approach smacks of the formalism of Balescu, Prigogine, and their co-workers, but differs in its allowance for a dynamical, time-dependent background geometry. Section IV focuses upon the physical implications of these exact equations, and then indicates how the situation would be complicated by allowing for nonlinear equations of motion. The discussion there suggests a important subtlety (and possible inconsistency!) inherent in any attempt to describe too carefully the effects of interparticle correlations without allowing correctly for the effects of nonlinearities. Section V summarizes the principal results and suggests future avenues of research.

It should be emphasized that much of the formalism developed here resembles quite closely more conventional techniques from statistical mechanics. Indeed, to the extent that the effects of a nontrivial background space-time may be ignored, and that one may proceed as if the space-time were really flat, it is reasonable to exploit the more powerful, but unfortunately less general, formalism of Prigogine and Balescu. Given, however, the potential implications in astrophysics and cosmology, it is important to formulate the basic issues from the perspective of, and in the language of, general relativity.

Finally, it should be noted that, without exception, the notation here parallels that of Paper I.

II. BASIC EQUATIONS FOR REDUCED DISTRIBUTIONS

Given that the background space-time is conformally static, the line element may be written in the form

$$ds^2 = g_{\mu\nu}\, dx^\mu\, dx^\nu = \Omega^2(\eta)(\gamma_{\eta\eta}\, d\eta^2 + \gamma_{ab}\, dx^a\, dx^b), \tag{2.1}$$

where η is the conformal time, and $\gamma_{\eta\eta}$ and γ_{ab} are functions only of the spatial x^c. In terms of $\gamma_{\mu\nu}$, the linearized equations of motion for the ith particle are

$$\frac{dx_i^a}{d\eta} = \frac{d\tau_i}{d\eta}\frac{l}{\Omega^2 m}\gamma^{ab}(i)p_b^i$$

and

$$\frac{dp_a^i}{d\eta} = \frac{d\tau_i}{d\eta}\left[\frac{-1}{2m\Omega^2}p_\mu^i\, p_\nu^i\, \partial_a^i\gamma^{\mu\nu}(i) - \lambda\Delta_a^{\ \mu}(i)\partial_\mu^i\Phi(i)\right]. \tag{2.2}$$

In these expressions, λ is a coupling constant, m is the particle mass, τ_i is the proper time of the ith particle, p_μ^i is the physical four-momentum, the η component of which is to be viewed as a function of the spatial components p_a^i and the space-time coordinates x_i^a and η, and $\Delta_\mu^{\ \nu}(i) \equiv \delta_\mu^\nu + p_\mu^i\, p_i^\nu/(\Omega^2 m^2)$ is the spatial projection tensor.

Similarly, if one allows for the possibility of a mass κ and a coupling with the scalar curvature, the field equations for $\Phi(x^a,\eta)$ assume the covariant form[1]

$$\nabla_\mu\nabla^\mu\Phi + \mathrm{n}\mathrm{R}[g]\Phi - \kappa^2\Phi = 4\pi\lambda\,\rho/\mathrm{m}. \tag{2.3}$$

Here

$$\rho \equiv m\sum_i \frac{d\tau_i}{d\eta}(-g)^{1/2}\delta^{(3)}[x^a - x_i^a(\eta)]$$
$$= m\Omega^4\sum_i \frac{d\tau_i}{d\eta}(-\gamma)^{1/2}\delta^{(3)}[x^a - x_i^a(\eta)] \tag{2.4}$$

is the particle density, ∇_μ and $R[g]$ denote, respectively, the covariant derivative operator and scalar curvature associated with $g_{\mu\nu}$, and n is a numerical constant that characterizes the coupling with R.

If one expresses $\nabla_\mu\nabla^\mu$ in terms of Ω^2 and the conformal $\gamma_{\mu\nu}$, and views $R[g]$ as a function of $\Omega(\eta)$ and $R[\gamma]$, the scalar curvature associated with $\gamma_{\mu\nu}$, the field equations may be seen to involve the second-order time-independent differential operator

$$\Delta \equiv (-\gamma^{\eta\eta})^{-1}\{(-\gamma)^{-1/2}\partial_a(-\gamma)^{1/2}\gamma^{ab}\partial_b + nR[\gamma]\}, \tag{2.5}$$

a natural generalization of the flat space Laplacian. It is therefore convenient to expand Φ in terms of a complete set of orthogonal [with respect to the inner product (2.22) of Paper I] eigenvectors $\{\psi_A\}$ of Δ, chosen to satisfy

$$\Delta\psi_A + \omega_A^2\psi_A = 0. \tag{2.6}$$

Without loss of generality, these ψ_A's may be chosen to satisfy the normalization

$$\int d^3x(-\gamma)^{1/2}(-\gamma^{\eta\eta})\psi_A(x^a)\psi_B(x^a) = 4\pi\delta_{AB}, \tag{2.7}$$

so that an assumption of completeness implies that

$$\sum_A \psi_A(x^a)\psi_A(y^a) = 4\pi(-\gamma^{\eta\eta})^{-1}(-\gamma)^{-1/2}\delta^{(3)}(x^a - y^a). \tag{2.8}$$

Thus, if one supposes that

$$\Phi(x_i^a,\eta) = \Omega^{-1}\chi(x_i^a,\eta) = \Omega^{-1}\sum_{A=1}^{\infty} a_A(\eta)\psi_A(x_i^a), \tag{2.9}$$

one is led immediately to an infinite set of oscillator equations of the form

$$\frac{dq_A}{d\eta} = -p_A$$

and

$$\frac{dp_A}{d\eta} = \left[\omega_A^2 + \frac{\kappa^2\Omega^2}{(-\gamma^{\eta\eta})} - (1+6n)\frac{\Omega''}{\Omega}\right]q_A + \frac{\lambda}{\Omega}\sum_i \frac{d\tau_i}{d\eta}\psi_A(i), \tag{2.10}$$

where a prime denotes differentiation with respect to η.

The "complete" distribution function $\mu(x_1^a, p_a^1, \ldots; q_1, p_1, \ldots; \eta)$ is to be defined as a probability density for finding the system with coordinates and momenta in the neighborhood of the stated values at time η. Then, as discussed in Paper I, the evolution of μ will be given by the Liouville equation

$$\frac{\partial\mu}{\partial\eta} + \sum_i \frac{\partial}{\partial x_i^a}\left(\frac{dx_i^a}{d\eta}\mu\right) + \sum_i \frac{\partial}{\partial p_a^i}\left(\frac{dp_a^i}{d\eta}\mu\right) + \sum_A \frac{\partial}{\partial q_A}\left(\frac{dq_A}{d\eta}\mu\right) + \sum_A \frac{\partial}{\partial p_A}\left(\frac{dp_A}{d\eta}\mu\right) = 0, \tag{2.11}$$

where, e.g., $dx_i^a/d\eta$ and $dp_A/d\eta$ are given by Eqs. (2.2) and (2.10).

Given this μ, one is in a position to define various reduced distribution functions. Thus, for example, the irreducible one-particle and one-oscillator distributions take the forms

$$f(i) = \int \prod_B dB \prod_{j\neq i} dj\,\mu \tag{2.12}$$

and

$$g(A) = \int \prod_{B\neq A} dB \prod_j dj\,\mu, \tag{2.13}$$

where

$$dj \equiv d^3x_j\, d^3p^j \quad \text{and} \quad dA \equiv dq_A\, dp_A \tag{2.14}$$

denote the appropriate reduced phase space volume elements. Similarly, one may define the particle-oscillator joint distribution

$$h(i,A) = \int \prod_{B \neq A} dB \prod_{j \neq i} dj\, \mu. \tag{2.15}$$

More complicated particle, field, and particle–field distributions may be defined in the obvious way.

It is straightforward to derive equations for the evolution of $f(i)$ and $g(A)$ in terms of $h(i,A)$. Thus, for example, by integrating over the degrees of freedom of all the oscillators and $N-1$ of the particles, one is led to an equation of the form

$$\frac{\partial f(i)}{\partial \eta} + \frac{\partial}{\partial x^a}\left(\frac{d\tau}{d\eta}\frac{1}{\Omega^2 m}p^a f\right) - \frac{\partial}{\partial p_a}\left(\frac{d\tau}{d\eta}\frac{1}{2m\Omega^2}p_\mu p_\nu\, \partial_a \gamma^{\mu\nu} f\right) - \frac{\partial}{\partial p_a}\left(\lambda \Delta^{\mu}_{a}\, \partial_\mu \Omega^{-1} \sum_A \int dA\, q_A \psi_A h(i,A)\right) = 0. \tag{2.16}$$

And, analogously, one finds that

$$\frac{\partial g(A)}{\partial \eta} - p_A \frac{\partial g}{\partial q_A} + \left[\omega_A^2 + \frac{\kappa^2 \Omega^2}{(-\gamma^{\eta\eta})} - (1+6n)\frac{\Omega''}{\Omega}\right] \times q_A \frac{\partial g}{\partial p_A} + \frac{\lambda}{\Omega}\sum_i \int di \frac{d\tau_i}{d\eta}\psi_A(i)\frac{\partial h(i,A)}{\partial p_A} = 0. \tag{2.17}$$

Similarly, it is easy enough to formulate an equation for the evolution of $h(i,A)$ in terms of $I(i,A,B)$, the reduced distribution for a single particle and a pair of oscillators, and $J(i,j,A)$, the distribution for a pair of particles and a single oscillator.

Equations (2.16) and (2.17) constitute the first two equations in a relativistic analog of the ordinary BBGKY hierarchy of coupled equations. The crucial difference between Eq. (2.16) and its Newtonian analog should, however, be emphasized. In a relativistic theory, the particle interactions are mediated via fields, and, as such, the evolution of $f(i)$ is affected by particle–oscillator correlations buried in $h(i,A)$. In a Newtonian theory, one envisions instead a direct particle–particle interaction, so that $h(i,A)$ will be replaced by a two-particle $H(i,j)$.

In any case, given Eqs. (2.16) and (2.17), it is easy to obtain a relativistic analog of the ordinary Vlasov equation or self-consistent field approximation (SCFA). Indeed, this sort of mean field theory follows immediately in the limit that correlations between particles and fields may be ignored, so that

$$h(i,A) \simeq f(i) g(A). \tag{2.18}$$

In this approximation, one concludes that

$$\frac{\partial f(i)}{\partial \eta} + \frac{\partial}{\partial x^a}\left(\frac{d\tau}{d\eta}\frac{1}{\Omega^2 m}p^a f\right) - \frac{\partial}{\partial p_a}\left(\frac{d\tau}{d\eta}\frac{1}{2m\Omega^2}p_\mu p_\nu\, \partial_a \gamma^{\mu\nu} f\right) - \frac{\partial}{\partial p_a}\left(\lambda \Delta^{\mu}_{a}\, \partial_\mu \Omega^{-1} \sum_A \langle q_A \psi_A \rangle f\right) = 0, \tag{2.19}$$

where

$$\langle q_A \psi_A(i)\rangle \equiv \int dA\, g(A) q_A \psi_A(i) \tag{2.20}$$

denotes an average value defined with respect to the one-oscillator $g(A)$. Similarly, one finds that

$$\frac{\partial g(A)}{\partial \eta} - p_A \frac{\partial g}{\partial q_A} + \left[\omega_A^2 + \frac{\kappa^2 \Omega^2}{(-\gamma^{\eta\eta})} - (1+6n)\frac{\Omega''}{\Omega}\right] \times q_A \frac{\partial g}{\partial p_A} + \frac{\lambda}{\Omega}\sum_i \left\{\frac{d\tau_i}{d\eta}\psi_A(i)\right\}\frac{\partial g}{\partial p_A} = 0, \tag{2.21}$$

where

$$\left\{\frac{d\tau_i}{d\eta}\psi_A(i)\right\} \equiv \int di\, f(i)\frac{d\tau_i}{d\eta}\psi_A(\mathrm{i}) \tag{2.22}$$

denotes an average value defined with respect to $f(i)$. Quite generally, given any function ξ of the particle and field variables, one may define the average values

$$\langle \xi \rangle \equiv \int \prod_B dB\, g(B)\, \xi(1,\ldots,N{:}1,2,\ldots) \tag{2.23}$$

and

$$\{\xi\} \equiv \int \prod_j dj\, f(j)\, \xi(1,\ldots,N;1,2,\ldots). \tag{2.24}$$

The connection between Eqs. (2.19) and (2.21) and the ordinary SCFA is easy to establish. Thus, if one introduces the "average" field

$$\langle \Phi \rangle \equiv \Omega^{-1} \sum_A \int dA\, g(A) q_A \psi_A, \tag{2.25}$$

and supposes that μ is symmetric under particle interchange, it follows from Eq. (2.21) that

$$\Omega^{-1}\partial_\eta^2\, \Omega \langle \Phi \rangle - \Delta \langle \Phi \rangle + \frac{\kappa^2 \Omega^2}{(-\gamma^{\eta\eta})}\langle \Phi \rangle - (1+6n)\frac{\Omega''}{\Omega}\langle \Phi \rangle = \frac{4\pi\lambda N}{(-g^{\eta\eta})}\int \frac{d^3p}{(-g)^{1/2}}\frac{d\tau}{d\eta} f, \tag{2.26}$$

or, equivalently, that

$$\nabla_\mu \nabla^\mu \langle \Phi \rangle + nR\,[g]\langle \Phi \rangle - \kappa^2 \langle \Phi \rangle = 4\pi\lambda\, \bar{\rho}/m, \tag{2.27}$$

where

$$\bar{\rho}(i) \equiv Nm \int \frac{d^3p^i}{(-g)^{1/2}}\frac{d\tau_i}{d\eta} f(i) \tag{2.28}$$

is the covariant mean density associated with $f(i)$. In terms of $\langle \Phi \rangle$, Eq. (2.19) takes the form

$$\frac{\partial f}{\partial \eta} + \frac{\partial}{\partial x^a}\left(\frac{d\tau}{d\eta}\frac{1}{\Omega^2 m}p^a f\right) - \frac{\partial}{\partial p_a}\left(\frac{d\tau}{d\eta}\frac{1}{2m\Omega^2}p_\mu p_\nu\, \partial_a \gamma^{\mu\nu} f\right) - \frac{\partial}{\partial p_a}(\lambda \Delta^{\mu}_{a}\, \partial_\mu \langle \Phi \rangle f) = 0. \tag{2.29}$$

Equations (2.27) and (2.29) are precisely what is usually meant by a relativistic mean field theory.[10]

III. A PROJECTION OPERATOR APPROACH

In the limit that the particle–oscillator distribution $h(i,A)$ may be approximated as an uncorrelated product $f(i)g(A)$, the exact equations (2.16) and (2.17) reduce to the coupled system (2.19) and (2.21) involving only $f(i)$ and $g(A)$. The problem, however, is that $h(i,A)$ does not factorize *exactly* in this simple way: particle–oscillator correlations will, of course, exist, and they will exert a nontrivial influence upon the evolution of f and g. What one might, therefore, like to do, and what can in fact be done, is express the "correlated component" of h, namely $\bar{h}(i,A) \equiv h(i,A) - f(i)g(A)$, as a functional of f and g, enabling one thereby to obtain exact equations involving only f and g.

The physical picture that one envisions is simple enough. In a sense that is to be made precise, the pieces of the total distribution function μ that do and do not involve correlations are mutually "orthogonal." And, as such, it is reasonable to think of the uncorrelated component as serving as a source for the correlated component, and the correlated component as serving as a source for the piece not involving correlations. More precisely, one expects to be able to decompose μ into a sum of two orthogonal pieces, a "relevant" μ_R and an "irrelevant" μ_I, which satisfy a pair of coupled equations; and, given this coupled system, one should be able to obtain a single equation for μ_R containing no explicit reference to μ_I except, possibly, through an initial condition.

This physical picture may be realized mathematically by means of a projection operator formalism. The first object is to write μ in the form

$$\mu = \mu_R + \mu_I \equiv \prod_i f(i) \prod_A g(A) + \mu_I, \tag{3.1}$$

i.e., as a sum of the relevant component μ_R, constructed as a product of irreducible contributions, and an irrelevant contribution $\mu_I \equiv \mu - \mu_R$, which contains all the information about correlations. The idea then is to construct a projection operator P, which, when acting upon μ, yields the desired μ_R, i.e., an operator P which satisfies the requirements

$$P\mu = \mu_R \quad \text{and} \quad (1-P)\mu = \mu_I \tag{3.2}$$

and

$$P(\eta)P(\eta') = P(\eta) \quad (\eta \geqslant \eta'). \tag{3.3}$$

Such a P will render precise the picture that one envisions. An important additional demand is that P be "compatible" with the notion of time translation, i.e., that the commutator of P (which may well involve η!) and ∂_η annihilate μ:

$$[P(\eta), \partial_\eta]\mu = 0. \tag{3.4}$$

Within this general framework, there are still a number of different ways in which one might proceed. In the context of Newtonian statistical mechanics, the most elegant approach would entail an application of techniques developed by Balescu, Prigogine, and their co-workers.[12,13] As discussed, e.g., in the classic paper by Balescu and Wallenborn,[12] this procedure enables one to extract an exact Markovian equation for the evolution of a "kinetic component" of the total distribution function; and, as is evidenced by the work of Balescu and Paiva-Veretennicoff,[14] that program may be adapted readily to such special relativistic problems as the consideration of an electrostatic plasma in Minkowski space.

The principal restriction inherent in this approach is that it relies upon the existence of a time-independent Hamiltonian. The necessity for a Hamiltonian in and of itself is perhaps not that much of a problem, but the demand that it be time independent, so that the equations of motion for the particles and field involve no time-dependent forces, is much more serious. For an arbitrary dynamical background, this is simply impossible! For the special case of a static space-time, the only explicit time dependence is to be found in the functions $\Phi(x^a,\eta)$ and $\rho(x^a,\eta)$, and even this dependence is eliminated by the introduction of the field oscillators. If, however, the space-time is only conformally static, the equations of motion will involve $\Omega(\eta)$ and, for a general space-time, things only get worse.

For this reason, this paper will adopt a simpler and less elegant, but more general, approach along the lines suggested by Willis and Picard,[15] which has already found fruitful applications in the study of Newtonian self-gravitating systems.[16–18]

What is needed is an explicit representation of P. Were one dealing with a collection of N particles and $\mathscr{N}$ oscillators, N and $\mathscr{N}$ both finite, this would be straightforward. Given, however, that there are infinitely many oscillators, one must be somewhat careful. The potential complications may be avoided best by ordering the eigenvectors ψ_A so that P may be defined as a limit with $\mathscr{N} \to \infty$:

$$\begin{aligned} P \equiv & \sum_{i=1}^{N} \prod_{j\neq i} f(j) \int dj \prod_{\forall B} g(B) \int dB \\ & + \lim_{\mathscr{N}\to\infty} \sum_{A=1}^{\mathscr{N}} \prod_{B\neq A} g(B) \int dB \prod_{\forall j} f(j) \int dj \\ & - \lim_{\mathscr{N}\to\infty} (N+\mathscr{N}-1) \prod_{\forall j} f(j) \int dj \prod_{\forall B} g(B) \int dB. \end{aligned} \tag{3.5}$$

This P, defined as an operator acting upon any function ξ, is itself constructed from f and g, and, therefore, is clearly time-dependent. It is, however, easy to see that Eq. (3.4) will hold and, moreover, that $P\mu = \mu_R$. That P is idempotent is less obvious, but a proof may be constructed along the lines used for the Newtonian analog.[15,18]

Given this P, the game is in fact quite simple. Start with the Liouville equation (2.11), viewed as an operator equation

$$\partial_\eta \mu(\eta) = -iL(\eta)\mu(\eta), \tag{3.6}$$

where $L(\eta)$, like $P(\eta)$, is a linear operator. The idea then is to act on Eq. (3.6) with P and $(1-P)$ to conclude that

$$\partial_\eta \mu_R + iPL\mu_R = -iPL\mu_I \tag{3.7}$$

and

$$\partial_\eta \mu_I + i(1-P)L\mu_I = -i(1-P)L\mu_R. \tag{3.8}$$

These relations demonstrate explicitly that μ_I and μ_R can, in fact, be interpreted, respectively, as sources for μ_R and μ_I. It is then easy to write down a formal solution to Eq. (3.8), yielding μ_I as a functional of μ_R, and to substitute that formal solution back into Eq. (3.7). Thus, in terms of the initial condition $\mu_I(\eta=0)$, one sees that

$$\partial_\eta \mu_R + iPL\mu_R = -iPL(\eta)\mathscr{G}(\eta,0)\mu_I(\eta=0) - \int_0^\eta d\eta' P(\eta)L(\eta)\mathscr{G}(\eta,\eta-\eta') \times[1-P(\eta-\eta')] \times L(\eta-\eta')\mu_R(\eta-\eta'), \tag{3.9}$$

where

$$\mathscr{G}(\eta_2,\eta_1) = T\exp\left\{-i\int_{\eta_1}^{\eta_2}[1-P(\eta)]L(\eta)\right\}, \tag{3.10}$$

and T is a (conformal) time-ordering operator. In the limit that $\mu_I(\eta=0)=0$, i.e., in the absence of initial correlations, one has an exact, closed equation for the evolution of μ_R. When $\mu_I(\eta=0)\neq 0$, one may speak of a propagation of nontrivial initial conditions.

At this stage it is convenient to introduce a bit of notation. Let

$$v_i^a \equiv \frac{d\tau_i}{d\eta}\frac{1}{\Omega^2 m}\gamma^{ab}(i)p_b^i \tag{3.11}$$

denote the ordinary three-velocity of the ith particle. Similarly, introduce the quantities

$$F_a(i) \equiv \frac{d\tau_i}{d\eta}\left[\frac{-1}{2m\Omega^2}p_\mu^i p_\nu^i \partial_a^i \gamma^{\mu\nu}(i) - \lambda\,\Delta_a^\mu(i)\,\partial_\mu^i\,\Omega^{-1}\sum_A q_A\psi_A(i)\right] \tag{3.12}$$

and

$$F(A) \equiv \left[\omega_A^2 + \frac{\kappa^2\Omega^2}{(-\gamma^{\eta\eta})} - (1+6n)\frac{\Omega''}{\Omega}\right]q_A + \frac{\lambda}{\Omega}\sum_i \frac{d\tau_i}{d\eta}\psi_A(i), \tag{3.13}$$

which may be identified, respectively, as the total forces acting upon the ith particle and the A th oscillator. In terms of these definitions, it is easy to verify that

$$iPL\mu_R = \sum_i \frac{\partial}{\partial x_i^a}(v_i^a\mu_R) + \sum_i \frac{\partial}{\partial p_a^i}(\langle F_a(i)\rangle\mu_R) - \sum_A \frac{\partial}{\partial q_A}(p_A\mu_R) + \sum_A \frac{\partial}{\partial p_A}(\{F(A)\}\mu_R), \tag{3.14}$$

where, recall, $\langle\ \rangle$ and $\{\ \}$ denote averages defined with respect to the g's and f's, respectively. Thus, setting $\mu_I(\eta=0)=0$, one concludes that

$$\int\prod_{\forall B} dB\int\prod_{j\neq i} dj[\partial_\eta\mu_R + iPL\mu_R] = \partial_\eta f(i) + \frac{\partial}{\partial x_i^a}[v_i^a f(i)] + \frac{\partial}{\partial p_a^i}[\langle F_a(i)\rangle f(i)] = -\int\prod_{\forall B} dB\int\prod_{j\neq i} dj\, iP(\eta)L(\eta)\int_0^\eta d\eta' \times\mathscr{G}(\eta,\eta-\eta')i[1-P(\eta-\eta')] \times L(\eta-\eta')\mu_R(\eta-\eta'), \tag{3.15}$$

and, similarly, that

$$\partial_\eta g(A) - p_A\frac{\partial g(A)}{\partial q_A} + \frac{\partial}{\partial p_a}[\{F(A)\}g(A)] = -\int\prod_{B\neq A} dB\int\prod_{\forall j} dj\, iP(\eta)L(\eta)\int_0^\eta d\eta' \times\mathscr{G}(\eta,\eta-\eta')i[1-P(\eta-\eta')] \times L(\eta-\eta')\mu_R(\eta-\eta'). \tag{3.16}$$

The left-hand side of these equations, which involve $iPL\mu_R$, would, if equated to zero, yield precisely the SCFA. The right-hand sides, which involve $iPL\mu_I$, reflect the effects of particle–oscillator correlations.

The right-hand sides can in fact be recast in a simpler, and more suggestive, form. Note first of all that, for any function ξ,

$$\int\prod_{B\neq A} dB\int\prod_{\forall j} dj(1-P)\xi = 0 = \int\prod_{\forall B} dB\int\prod_{j\neq i} dj(1-P)\xi. \tag{3.17}$$

This implies that one may ignore the first $P(\eta)$ in Eqs. (3.15) and (3.16). Similarly, it follows from Eq. (3.1) that

$$\int\prod_{B\neq A} dB\int\prod_{\forall j} dj\,\mu_I = 0 = \int\prod_{\forall B} dB\int\prod_{j\neq i} dj\,\mu_I. \tag{3.18}$$

These relations imply that one need only consider those contributions to $iP(\eta)L(\eta)$ that involve explicitly the particle–oscillator couplings. Thus, for example,

$$\int\prod_{\forall B} dB\int\prod_{j\neq i} dj\, iPL\mu_I = \int\prod_{\forall B} dB\int\prod_{j\neq i} dj\, iL\mu_I = \int\prod_{\forall B} dB\int\prod_{j\neq i} dj\left[\frac{\partial}{\partial x_i^a}v_i^a + \frac{\partial}{\partial p_a^i}F_a(i)\right]\mu_I = \int\prod_{\forall B} dB\int\prod_{j\neq i} dj\frac{\partial}{\partial p_a^i}\sum_A F_a(A\to i)\mu_I, \tag{3.19}$$

where

$$F_a(A\to i) = -\lambda\frac{d\tau_i}{d\eta}\Delta_a^\mu(i)\,\partial_\mu^i\,\Omega^{-1}q_A\psi_A(i) \tag{3.20}$$

is the force exerted upon particle i by oscillator A. And, similarly, one sees that

$$\int\prod_{B\neq A} dB\int\prod_{\forall j} dj\, iPL\mu_I = \int\prod_{B\neq A} dB\int\prod_{\forall j} dj\frac{\partial}{\partial p_a}\sum_i F(i\to A)\mu_I, \tag{3.21}$$

where

$$F(i\to A) = \frac{\lambda}{\Omega}\frac{d\tau_i}{d\eta}\psi_A(i) \tag{3.22}$$

is the force that particle i exerts upon oscillator A. In terms of these pairwise forces, one finds that

$$i(1-P)L\mu_R = \sum_i \sum_A \left[\frac{\partial}{\partial p_a^i}\mathscr{F}_a(A\to i) + \frac{\partial}{\partial p_A}\mathscr{F}(i\to A)\right]\mu_R, \qquad (3.23)$$

where

$$\mathscr{F}_a(A\to i) = F_a(A\to i) - \langle F_a(A\to i)\rangle \qquad (3.24)$$

and

$$\mathscr{F}(i\to A) = F(i\to A) - \{F(i\to A)\} \qquad (3.25)$$

represent suitably defined "fluctuating forces." Finally, one may observe that, in Eqs. (3.19) and (3.21), the true $F_a(A\to i)$ and $F(i\to a)$ may be replaced by $\mathscr{F}_a(A\to i)$ and $\mathscr{F}(i\to A)$, yielding thereby a somewhat more symmetric form. Thus, for example, it follows from Eq. (3.18) that the quantity proportional to $\langle F_a(A\to i)\rangle$, which one wishes to append, will vanish identically since the average value is a function only of x_i^a and p_a^i.

By implementing these definitions and simplifications, one is led to the desired coupled equations, written in the forms

$$\begin{aligned}\partial_\eta f(i) &+ \frac{\partial}{\partial x_i^a}[v_i^a f(i)] + \frac{\partial}{\partial p_a^i}[\langle F_a(i)\rangle f(i)] \\ &= \int \prod_{\forall C} dC \int \prod_{k\neq i} dk \int_0^\eta d\eta' \\ &\times \sum_A \frac{\partial}{\partial p_a^i}\mathscr{F}_a(A\to i;\eta)\mathscr{G}(\eta,\eta-\eta') \\ &\times \sum_{\forall B}\sum_{\forall j}\left[\frac{\partial}{\partial p_b^j}\mathscr{F}_b(B\to j;\eta-\eta') + \frac{\partial}{\partial p_B}\mathscr{F}(j\to B;\eta-\eta')\right]\mu_R(\eta-\eta') \end{aligned} \qquad (3.26)$$

and

$$\begin{aligned}\partial_\eta g(A) &- \frac{\partial}{\partial q_A}[p_A g(A)] + \frac{\partial}{\partial p_A}[\{F(A)\}g(A)] \\ &= \int \prod_{C\neq A} dC \int \prod_{\forall k} dk \int_0^\eta d\eta' \\ &\times \sum_i \frac{\partial}{\partial p_A}\mathscr{F}(i\to A;\eta)\mathscr{G}(\eta,\eta-\eta') \\ &\times \sum_{\forall B}\sum_{\forall j}\left[\frac{\partial}{\partial p_b^j}\mathscr{F}_b(B\to j;\eta-\eta') + \frac{\partial}{\partial p_B}\mathscr{F}(j\to B;\eta-\eta')\right]\mu_R(\eta-\eta'). \end{aligned} \qquad (3.27)$$

IV. PHYSICAL INTERPRETATION AND THE EFFECT OF NONLINEARITIES

As emphasized already, the exact equations (3.26) and (3.27) embody two sorts of influences, namely (i) "average" effects, involving $\langle F_a(A\to i)\rangle$ and $\{F(i\to A)\}$, which derive from the quantity $iPL\mu_R$ and, as such, do not involve explicitly the effects of correlations; and (ii) the effects of "deviations from average conditions," involving the fluctuating forces $\mathscr{F}_a(A\to i)$ and $\mathscr{F}(i\to A)$, which derive from $iPL\mu_I$ and, as such, reflect explicitly the effects of correlations. In particular, in the limit that $iPL\mu_I$ may be neglected altogether, one recovers the SCFA of Sec. II.

All this is very nice. One potential subtlety in the interpretation of these equations should, however, be discussed. Specifically, in that one is not assuming that all particle–oscillator correlations may be neglected, it is no longer obvious that the average $\langle \Phi \rangle$ entering into the definition of $\langle F_a(i)\rangle$ still satisfies the average field equation (2.27). One might anticipate that $\langle \Phi \rangle$ would depend not only upon $f(i)$, but upon some higher-order correlation functions. It is therefore important to verify explicitly that

$$\langle F_a(A\to i)\rangle = -\lambda \frac{d\tau_i}{d\eta}\Delta_a^\mu(i)\,\partial_\mu^i\,\Omega^{-1}\langle q_A\psi_A(i)\rangle, \qquad (4.1)$$

where $\langle q_A\psi_A\rangle$ is determined by the SCFA equations.

The proof of this assertion is not difficult. Thus, if one exploits the exact relation (2.17) to express the time evolution of $g(A)$ in terms of $h(i,A)$, one sees immediately that

$$\begin{aligned}\partial_\eta^2\langle q_A\psi_A\rangle &= \partial_\eta \int dA\, q_A\psi_A\partial_\eta g(A) \\ &= \partial_\eta \int dA\, q_A\psi_A\left[p_A\frac{\partial g(A)}{\partial q_A} + \cdots\right] \\ &= -\partial_\eta\langle P_A\psi_A\rangle. \end{aligned} \qquad (4.2)$$

Here the third equality follows from an integration by parts and the observation that terms involving $\partial g(A)/\partial p_A$ do not contribute. A second application of Eq. (2.17) then implies that

$$\begin{aligned}&\partial_\eta^2\langle q_A\psi_A(i)\rangle \\ &= \int dA\, p_A\psi_A(i)\left[\omega_A^2 + \frac{\kappa^2\Omega^2}{(-\gamma^{\eta\eta})} - (1+6n)\frac{\Omega''}{\Omega}\right]q_A\frac{\partial g(A)}{\partial p_A} \\ &\quad + \int dA\, p_A\psi_A(i)\frac{\lambda}{\Omega}\sum_k\int dk\frac{d\tau_k}{d\eta}\psi_A(k)\frac{\partial h(k,A)}{\partial p_A} \\ &= -\left[\omega_A^2 + \frac{\kappa^2\Omega^2}{(-\gamma^{\eta\eta})} - (1+6n)\frac{\Omega''}{\Omega}\right]\langle q_A\psi_A(i)\rangle \\ &\quad - \int dA\,\psi_A(i)\frac{\lambda}{\Omega}\sum_k\int dk\frac{d\tau_k}{d\eta}\psi_A(k)h(k,A). \end{aligned} \qquad (4.3)$$

One knows, however, that

$$\int dA\, h(k,A) = f(k), \qquad (4.4)$$

so that the field equation for $\langle q_A\psi_A\rangle$ takes the form

$$\begin{aligned}&\partial_\eta^2\langle q_A\psi_A(i)\rangle + \left[\omega_A^2 + \frac{\kappa^2\Omega^2}{(-\gamma^{\eta\eta})} - (1+6n)\frac{\Omega''}{\Omega}\right]\langle q_A\psi_A(i)\rangle \\ &= -\frac{\lambda}{\Omega}\sum_k\int dk\frac{d\tau_k}{d\eta}\psi_A(i)\psi_A(k)f(k). \end{aligned} \qquad (4.5)$$

If one sums this relation over all A and recalls the completeness relation (2.8), one recovers the SCFA [Eq. (2.26)].

It should, however, be stressed that although the equations for $\langle q_A\psi_A\rangle$ or $\langle \Phi \rangle$ do not involve the effects of correlations, the equations for a more complicated object like

$\langle q_A^2\, \psi_A^2 \rangle$ or $\langle \Phi^2 \rangle$ most definitely will! Indeed, an analysis paralleling Eqs. (4.2)–(4.5) implies that

$$\partial_\eta^2 \langle q_A^2\, \psi_A^2\,(i)\rangle - 2\langle p_A^2\, \psi_A^2\,(i)\rangle + 2\left[\omega_A^2 + \frac{\kappa^2\Omega^2}{(-\gamma^{\eta\eta})} - (1+6n)\frac{\Omega''}{\Omega}\right]\langle q_A^2\, \psi_A^2\,(i)\rangle = -2\int dA\, q_A \psi_A{}^2(i)\frac{\lambda}{\Omega}\sum_k \int dk\,\frac{d\tau_k}{d\eta}\,\psi_A(k)h(k,A), \tag{4.6}$$

a relation which involves $h(i,A)$ irreducibly.

One can actually choose to view the correlations buried in $h(i,A)$ as providing a source for the difference $\langle q_A^2\, \psi_A^2\rangle - \langle q_A \psi_A\rangle^2$. One knows, of course, that

$$\partial_\eta^2[\langle q_A\psi_A\rangle^2] = 2\langle q_A\psi_A\rangle\partial_\eta^2\langle q_A\psi_A\rangle + 2[\partial_\eta\langle q_A\psi_A\rangle]^2, \tag{4.7}$$

and, therefore, it is easy to see that

$$\begin{aligned}\partial_\eta^2 \langle q_A\psi_A(i)\rangle^2 &- 2\langle p_A\psi_A(i)\rangle^2 + 2\left[\omega_A^2 + \frac{\kappa^2\Omega^2}{(-\gamma^{\eta\eta})} - (1+6n)\frac{\Omega''}{\Omega}\right]\langle q_A\psi_A(i)\rangle^2\\ &= -2\langle q_A\psi_A(i)\rangle\int dA\,\psi_A(i)\frac{\lambda}{\Omega}\times\sum_k\int dk\,\frac{d\tau_k}{d\eta}\,\psi_A(k)h(k,A)\\ &= -2\langle q_A\psi_A(i)\rangle\frac{\lambda}{\Omega}\sum_k\int dk\,\frac{d\tau_k}{d\eta}\,\psi_A(i)\psi_A(k)f(k).\end{aligned} \tag{4.8}$$

Let

$$\tilde h(k,A) \equiv h(k,A) - f(k)g(A) \tag{4.9}$$

again denote the particle–oscillator correlation function. Then, in terms of the quantities

$$\delta_A^2 \equiv \langle q_A^2\, \psi_A^2\rangle - \langle q_A\psi_A\rangle^2 \tag{4.10}$$

and

$$D_A^2 \equiv \langle p_A^2\, \psi_A^2\rangle - \langle p_A\psi_A\rangle^2, \tag{4.11}$$

Eqs. (4.6) and (4.8) imply that

$$\left\{\partial_\eta^2 + 2\left[\omega_A^2 + \frac{\kappa^2\Omega^2}{(-\gamma^{\eta\eta})} - (1+6n)\frac{\Omega''}{\Omega}\right]\right\}\delta_A^2(i) - 2D_A^2(i) = -2\langle q_A\psi_A(i)\rangle\frac{\lambda}{\Omega}\sum_k\int dk\,\frac{d\tau_k}{d\eta}\,\psi_A(k)\tilde h(k,A). \tag{4.12}$$

In the SCFA, $\tilde h$ is assumed to vanish, so that the right-hand side of Eq. (4.11) disappears. If, however, one allows correctly for a nontrivial $\tilde h(i,A)$, this is no longer true: in this case, it would be inconsistent to demand that δ_A^2 and D_A^2 simultaneously vanish identically.

These considerations suggest the importance of understanding how the basic formalism and interpretation would be altered if one were to allow for nonlinear equations of motion. As a concrete example, consider a collection of particles that interact via fields that still satisfy the linear equations (2.3), but in which the particle equations of motion involve a nonlinear coupling with Φ of the form considered by Hakim[2,3] or Kandrup[5]:

$$\frac{dx_i^a}{d\eta} = \frac{d\tau_i}{d\eta}\frac{1}{\Omega^2 m}\gamma^{ab}(i)p_b^i$$

and

$$\frac{dp_a^i}{d\eta} = \frac{d\tau_i}{d\eta}\left[\frac{-1}{2m\Omega^2}p_\mu^i\, p_\nu^i\,\partial_a^i\,\gamma^{\mu\nu}(i) - m\Delta^\mu_a(i)\,\partial_\mu^i\log\left(1+\frac{\lambda\Phi(i)}{m}\right)\right]. \tag{4.13}$$

In the limit that $|\lambda\Phi/m| \ll 1$, this reduces to Eq. (2.2), which only involves Φ linearly. In general, however, all powers of Φ will be involved.

Given that the field equations remain unchanged, one can still expand Φ in terms of the oscillators of Eq. (2.9) to obtain Eqs. (2.10). And, given the particle and oscillator equations, one can again formulate the "complete" Liouville equation and obtain expressions for the evolution of the reduced $f(i)$ and $g(A)$ in terms of higher-order correlation functions. Because the field equations are still given by Eq. (2.3), the equation for $\partial_\eta g(A)$ remains unchanged. The nonlinear couplings in Eq. (4.12) will, however, have a profound influence upon the equation for $\partial_\eta f(i)$. Thus, one finds explicitly that, in this case,

$$\begin{aligned}\frac{\partial f(i)}{\partial\eta} &+ \frac{\partial}{\partial x_i^a}\left[\frac{d\tau_i}{d\eta}\frac{1}{\Omega^2 m}p_i^a f(i)\right] - \frac{\partial}{\partial p_a^i}\left[\frac{d\tau_i}{d\eta}\frac{1}{2m\Omega^2}p_\mu^i\, p_\nu^i\,\partial_a^i\,\gamma^{\mu\nu}(i)f(i)\right]\\ &- \frac{\partial}{\partial p_a^i}\left[m\frac{d\tau_i}{d\eta}\Delta^\mu_a(i)\int\prod_{\forall B}dB\,\partial_\mu^i \times\log\left(1+\frac{\lambda}{m}\Omega^{-1}\sum_A q_A\psi_A(i)\right)\right]\mathscr{F}(i;1,2,\ldots) = 0,\end{aligned} \tag{4.14}$$

where, now,

$$\mathscr{F}(i;1,2,\ldots) \equiv \int\prod_{j\neq i}dj\,\mu \tag{4.15}$$

is the reduced distribution appropriate for a single particle and *all* of the oscillators. The equations of motion no longer involve a simple pairwise particle–oscillator interaction, and, therefore, Eq. (4.14) involves the full $\mathscr{F}(i;1,2,\ldots)$, rather than $h(i,A)$.

In the limit that $|\lambda\Phi/m| \ll 1$, one may approximate that $\log(1+u) \simeq u$ and recover thereby Eq. (2.17). To next higher order, one has that $\log(1+u) \simeq u - \frac{1}{2}u^2$, so that one acquires an additional quadratic contribution of the form

$$\begin{aligned}-\frac{\partial}{\partial p_a^i}&\frac{\lambda^2}{m}\frac{d\tau_i}{d\eta}\Delta^\mu_a(i)\int\prod_{\forall C}dC\,\partial_\mu^i\Omega^{-2}\times\sum_A q_A\psi_A(i)\sum_B q_B\psi_B(i)\mathscr{F}(i;1,2,\ldots)\\ &= -\frac{\partial}{\partial p_a^i}\frac{\lambda^2}{m}\frac{d\tau_i}{d\eta}\Delta^\mu_a(i)\times\left[\sum_A\int dA\,\partial_\mu^i\Omega^{-2}q_A^2\,\psi_A^2(i)h(i,A) + \sum_{B\neq A}\int dA\int dB\,\partial_\mu^i\,\Omega^{-2}q_A q_B\psi_A(i)\psi_B(i)I(i;A,B)\right],\end{aligned} \tag{4.16}$$

where, explicitly

$$I(i;A,B) \equiv \int \prod_{j \neq i} dj \prod_{C \neq A,B} dC\, \mu. \tag{4.17}$$

At this stage, it is again natural to implement a relativistic SCFA by supposing that

$$h(i,A) \simeq f(i) g(A), \quad I(i;A,B) \simeq f(i) g(A) g(B). \tag{4.18}$$

In this approximation, Eq. (4.18) reduces to

$$-\frac{\partial}{\partial p_a^i} \frac{\lambda^2}{m} \frac{d\tau_i}{d\eta} \Delta_a^\mu(i)\, \partial_\mu^i \Big[\sum_A \Omega^{-2} \langle q_A^2\, \psi_A^2(i) \rangle + \sum_{B \neq A} \Omega^{-2} \langle q_A \psi_A(i) \rangle \langle q_B \psi_B(i) \rangle \Big] f(i)$$
$$= -\frac{\partial}{\partial p_a^i} \frac{\lambda^2}{m} \frac{d\tau_i}{d\eta} \Delta_a^\mu(i)\, \partial_\mu^i \langle \Phi^2(i) \rangle f(i). \tag{4.19}$$

Equation (4.20) does not yet coincide completely with the sort of equation that one typically encounters when describing a self-gravitating system. The additional nontrivial assumption that one requires is that $\langle q_A^2\, \psi_A^2 \rangle \simeq \langle q_A \psi_A \rangle^2$, so that $\langle \Phi^2 \rangle = \langle \Phi \rangle^2$. As is evident from Eq. (4.11), the difference between these two quantities is related to the correlation function $\tilde{h}(i,A)$ and, therefore, this approximation ought not to be that unreasonable in the limit that

$$\mu \simeq \mu_R = \prod_i f(i) \prod_A g(A). \tag{4.20}$$

One sees quite generally that, if Eq. (4.20) is valid,

$$\frac{\partial f(i)}{\partial \eta} + \frac{\partial}{\partial x_i^a} \left[\frac{d\tau_i}{d\eta} \frac{1}{\Omega^2 m} p_i^a f(i) \right] - \frac{\partial}{\partial p_a^i} \left[\frac{d\tau_i}{d\eta} \frac{1}{2m\Omega^2} p_\mu^i\, p_\nu^i\, \partial_a^i\, \gamma^{\mu\nu}(i) f(i) \right]$$
$$- \frac{\partial}{\partial p_a^i} \left[m \frac{d\tau_i}{d\eta} \Delta_A^\mu(i) \times \partial_\mu^i \left\langle \log\left(1 + \frac{\lambda \Phi(i)}{m}\right) \right\rangle f(i) \right] = 0, \tag{4.21}$$

and, in the limit that $\langle q_A^n\, \psi_A^n \rangle \simeq \langle q_A \psi_A \rangle^n$, this takes the form

$$\frac{\partial f(i)}{\partial \eta} + \frac{\partial}{\partial x_i^a} \left[\frac{d\tau_i}{d\eta} \frac{1}{\Omega^2 m} p_i^a f(i) \right] - \frac{\partial}{\partial p_a^i} \left[\frac{d\tau_i}{d\eta} \frac{1}{2m\Omega^2} p_\mu^i\, p_\nu^i\, \partial_a^i\, \gamma^{\mu\nu}(i) f(i) \right]$$
$$- \frac{\partial}{\partial p_a^i} \left[m \frac{d\tau_i}{d\eta} \Delta_a^\mu(i)\, \partial_\mu^i \log\left(1 + \frac{\lambda \langle \Phi(i) \rangle}{m}\right) \right] f(i) = 0, \tag{4.22}$$

precisely the type of expression that one would ordinarily consider in a simple mean field theory.

The consideration of nonlinear field equations also introduces the possibility of a logical inconsistency. Specifically, a "natural" thing to do in many cases, e.g., for a self-gravitating system, would be to linearize the basic equations at the very outset and then proceed as if the "fundamental" theory really were linear. Thus, for example, if one were working with a flat, static background, one could realize the effects of the "fluctuating forces" in a perturbation series which really amounts to an expansion in powers of the coupling constant λ (see Refs. 13 and 14): the lowest-order contributions will be $O(\lambda^2)$ and higher-order contributions $O(\lambda^n)$ could, at least in principle, be evaluated systematically. The obvious problem, however, is that the terms that one is ignoring because of the linearization may also be realized in powers of λ, and that, in particular, one will be neglecting contributions to the mean field $O(\lambda^2)$!

On a purely formal level, one cannot evaluate the correlational effects associated with fluctuating forces even to lowest order without also introducing nonlinear mean field effects!

The ultimate justification for so doing would seem to be based not simply upon naive power counting, but upon the notion that the mean field and correlational effects are somehow decoupled, the former representing large-scale "global" phenomena, the latter representing "microscopic" or "localized" phenomena. Thus, for example, the kinetic theory of self-gravitating systems developed by Israel and Kandrup[4–7] corresponds to a situation in which one considers the dominant mean field contributions, which, in this language, would be $O(\lambda)$, and the dominant correlational effects, which would be $O(\lambda^2)$, but ignores higher-order contributions or couplings between these effects. Any attempt to improve upon that admittedly naive approach will involve important questions of principle!

V. DISCUSSION

The principal objective of this paper was to indicate how, starting from a "complete" statistical description of a collection of particles interacting via a simple scalar field in a fixed, conformally static space-time, one could derive useful information about the evolution of reduced one-particle and one-oscillator distribution functions. The analysis is of intrinsic interest in its own right, and, moreover, should indicate how one might hope to describe the evolution of realistic self-gravitating systems, such as a relativistic cosmology or a cluster of stars.

In Sec. II, attention focused primarily upon the problem of trying to understand in the clearest possible way the physical content of the sort of SCFA encountered, for example, in conventional descriptions of self-gravitating systems in general relativity.[10,11] Section III then demonstrated explicitly how, at least in principle, one can transcend that sort of mean field theory to derive exact coupled equations for the evolution of reduced distributions that contain no explicit reference to higher-order correlation functions.

The analysis in these sections was predicated entirely upon the fact that the particle and field equations are both linear. This is certainly true for the case of the electromagnetic interaction and, as discussed by Israel and Kandrup,[4] such an assumption should also be legitimate in some approximation for the description of self-gravitating systems, provided that "deviations" from some "average" conditions are not too large. Nevertheless, it is clear that linearized interactions are not the whole story, and, for this reason, much of Sec. IV was devoted to the question of how an allowance for nonlinear effects might alter the basic physics. Thus, for example, it was emphasized that the usual sorts of perturbation expansions, in terms of powers of a coupling constant λ, may require a reassessment. And, moreover, it was indicated that the standard sort of mean field theory, as applied to nonlinear interactions, amounts, in the context of the SCFA, to the assumptions (i) that particle-field correlations may be neglected entirely and (ii) that expectation values of products may be equated with products of expectation values, so that, e.g., $\langle q_A^2\ \psi_A^2\rangle \simeq \langle q_A \psi_A\rangle^2$.

The types of problems that remain should also be emphasized. The most important of these is probably the formulation of a tractable approximate expression for the "collision operator" that described the effects of the fluctuating forces, or, better yet, the development of a usable systematic perturbation expansion in terms of which to describe these effects. For the case of a homogeneous configuration in Minkowski space, this should not be difficult. The Balescu–Prigogine formalism, as applied to an electromagnetic plasma,[13] clearly does the trick. It is, however, evident already at the level of the Newtonian theory that an allowance for spatial inhomogeneities[19] or a nontrivial background spacetime, e.g., a $k = 0$ Friedmann cosmology,[20] can lead to qualitatively different results! In particular, the absence of a static background almost certainly precludes the possibility of an "equilibrium."

Another important issue concerns the possibility of obtaining useful information about other, more complicated, objects of interest, such as the two-particle correlation function. The point is that, when considering such astrophysical phenomena as the clustering of galaxies or the evolution of "clumps" in a cluster of stars, one wishes to know not only the one-particle $f(i)$ but, in addition, such quantities as the two-particle distribution (say) $H(i,j)$. It is, for example, this object that yields the observed spatial galaxy covariance function measured by Peebles[21] and his co-workers. The formulation of exact equations for these sorts of objects, independent of higher-order correlations, is, in a Newtonian framework, straightforward albeit messy.[18] The relativistic generalization introduces no new issues of principle aside from those considered in this paper.

That relativistic effects will in fact prove of crucial importance for the understanding of galaxy clustering, or even the evolution of dense stellar systems,[22,23] may well be unlikely (although these effects ought to be connected with the formation of massive black holes at the center of a galaxy!). The desire to understand how to pose the problem correctly in a relativistic framework is, however, a natural one.

Finally, there remains the problem of trying to understand how one could deal with nonlinear interactions in a completely satisfactory way: only by solving this problem might one claim to "understand" self-gravitating systems. It is, of course, useful to have a simple linearized theory, and, for most practical purposes, such a theory may be enough. Ultimately, however, one needs to ascertain precisely what the "completely correct" description entails, so that one may appreciate fully what is, and is not, being ignored!

ACKNOWLEDGMENTS

It is a pleasure to acknowledge useful discussions with Alan Grecos, Remi Hakim, and, especially, George Severne. It is also a pleasure to acknowledge encouragement provided by S. C. Hill and T. L. Thompson.

Partial financial support was provided by the National Science Foundation and by the Center for Theoretical Physics at the University of Maryland.

[1]H. E. Kandrup, J. Math. Phys. **25**, 3286 (1984).
[2]R. Hakim, J. Math. Phys. **8**, 1315 (1967).
[3]R. Hakim, J. Math. Phys. **8**, 1379 (1967).
[4]W. Israel and H. E. Kandrup, Ann. Phys. (NY) **152**, 30 (1984).
[5]H. E. Kandrup, Ann. Phys. (NY) **153**, 44 (1984).
[6]H. E. Kandrup, Physica A **126**, 461 (1984).
[7]H. E. Kandrup, Astrophys. J. **282**, 361 (1984).
[8]These issues are discussed from the viewpoint of a relativist by J. Ehlers, in *General Relativity and Gravity*, edited by G. Shaviv and J. Rosen (Wiley, New York, 1975).
[9]A more conventional statistical mechanical viewpoint may be found in R. Balescu and T. Kotera, Physica **33**, 558 (1967).
[10]W. Israel, in *General Relativity: Papers in Honour of J. L. Synge*, edited by L. O'Raifertaigh (Clarendon, Oxford, 1972).
[11]J. Ehlers, in *General Relativity and Gravitation: Enrico Fermi Course XLVII* (Academic, New York, 1971).
[12]R. Balescu and J. Wallenborn, Physica **54**, 477 (1971).
[13]R. Balescu, *Equilibrium and Non-Equilibrium Statistical Mechanics* (Wiley, New York, 1975).
[14]R. Balescu and I. Paiva-Veretennicoff, Bull. Cl. Sci., Acad. Roy. Belg. **57**, 457 (1971).
[15]C. Willis and R. Picard, Phys. Rev. A **9**, 1343 (1974).
[16]H. E. Kandrup, Astrophys. J. **244**, 316 (1981).
[17]H. E. Kandrup, Astrophys. J. **259**, 1 (1984).
[18]H. E. Kandrup and S. Hill Kandrup, Astrophys. J. **277**, 1 (1984).
[19]H. D. Kandrup, Astrophys. Space Sci. (in press).
[20]H. E. Kandrup, Phys. Rev. D **30**, 2067.
[21]See, e.g., P. J. E. Peebles, *The Large-Scale Structure of the Universe* (Princeton U.P., Princeton, 1980) for an exhaustive list of references.
[22]E. D. Fackerell, J. R. Ipser, and K. S. Thorne, Comm. Astrophys. Space Phys. **1**, 134 (1969).
[23]H. E. Kandrup, in *Proceedings of I. A. U. Symposium* 113: *Dynamics of Star Clusters*, edited by P. Hut (Reidel, Dordrecht, in press).

Gravitational radiation from dust

R. A. Isaacson
Physics Division, National Science Foundation, Washington, D.C. 20550
J. S. Welling and J. Winicour
Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

(Received 22 April 1985; accepted for publication 18 June 1985)

A dust cloud is examined within the framework of the general relativistic characteristic initial value problem. Unique gravitational initial data are obtained by requiring that the space-time be quasi-Newtonian. Explicit calculations of metric and matter fields are presented, which include all post-Newtonian corrections necessary to discuss the major physical properties of null infinity. These results establish a curved space version of the Einstein quadrupole formula, in the form "news function equals third time derivative of transverse quadrupole moment," for this system. However, these results imply that some weakened notion of asymptotic flatness is necessary for the description of quasi-Newtonian systems.

I. INTRODUCTION

In this paper, we describe the details of a new approach to the calculation of gravitational radiation from a quasi-Newtonian source. We work within the framework of the fully relativistic initial value problem on a null cone and take the Newtonian limit to obtain a hierarchy of corrections to purely Newtonian behavior. The novel features of our computational scheme include the use of the true curved-space null cones for the propagation of radiation; a unified treatment of the space-time region extending from the material sources to null infinity; a well-defined prescription designed to eliminate unphysical incoming gravitational waves; and a geometrical treatment of null infinity for the evaluation of the Bondi news function. These techniques have two key advantages. First, they lead to a unique prescription, within general relativity, for producing physically reasonable initial gravitational data corresponding to a Newtonian system. Second, they give a means of calculating post-Newtonian effects, including corrections to the equations of motion and extensions to the Einstein quadrupole formula.

At this stage, it is premature to attempt a comparison and critical review of how our new scheme relates to the large literature of alternative approaches to this subject. Rather, the purpose of this article is to present an exposition of our results in the context of a specific Newtonian model. We also take this opportunity to present a number of new and useful results, extending previous work[1,2] to higher order where the first outgoing radiation terms appear.

The use of a simple physical model illustrates the computational possibilities for our formalism. The higher-order radiation calculations lead to quite long expressions. These, computed using MACSYMA, are presented in the appendices. The body of this paper is used to discuss both general formalism and computational strategy, as well as specific tricks used for the model calculation. Section II reviews Newtonian dynamics, procedures for evaluating time derivatives of Newtonian quantities, and the treatment of matter discontinuities. Section III presents the general rules governing quasi-Newtonian dynamics, including formulas determining the metric up to third-order corrections. It also discusses post-Newtonian equations of motion. Section IV applies these general techniques to our model system. Section V deals with asymptotic issues such as the comparison of the gravitational radiation with the Einstein quadrupole formula.[3] Section VI summarizes our results and discusses both some unresolved issues remaining in the analytic approach and the prospects for numerical computation. The notation and conventions used throughout the paper are presented in Appendix A. Other appendices contain explicit expressions for lengthy results discussed in the corresponding main sections.

II. THE NEWTONIAN CALCULATION

We review some Newtonian computational techniques for self-gravitating fluids. These are of interest, not only as a foundation for understanding relativistic corrections, but also for the evaluation of many terms occurring in the quasi-Newtonian dynamics.

Our main interest in this paper is not purely formal techniques but their application to a simple physical model. The model consists of an initially homogeneous ball of dust, centered at the origin. The initial density of the dust has the discontinuous form

$$\rho = \begin{cases} k\,, & r<R\,,\\ 0\,, & r>R\,, \end{cases} \tag{2.1}$$

while its initial Newtonian velocity is anisotropic

$$v_1 = vr^3 Y_2\,. \tag{2.2}$$

The evolution of this model follows from the coupled gravitational and hydrodynamical equations of motion, conveniently expressed in Cartesian coordinates. In order to agree with the signature convention for the full space-time in the following sections, we choose the signature for the flat Newtonian three-geometry to be $(-,-,-)$. Then the Poisson–Euler equations are

$$\nabla^2\Phi \equiv -\Phi^{;k}{}_{;k} = 4\pi\rho\,, \tag{2.3}$$

$$\rho_{,0} + (\rho v^i)_{;i} = 0\,, \tag{2.4}$$

$$\rho v_{i,0} = -\rho v^k v_{i;k} + p_{;i} + \rho\Phi_{;i}\,. \tag{2.5}$$

 0022-2488/85/112859-10$02.50

The fluid has a matter stress tensor

$$T_{ij} = \rho v_i v_j - p g_{ij} \tag{2.6}$$

and a gravitational stress tensor

$$t_{ij} = (4\pi)^{-1}[\Phi_{;i}\Phi_{;j} - \tfrac{1}{2}\Phi^{;k}\Phi_{;k}g_{ij}], \tag{2.7}$$

satisfying

$$t_i{}^k{}_{;k} = -\rho\Phi_{;i}. \tag{2.8}$$

Consequently, the momentum flux is given by

$$(\rho v_i)_{,0} = -\Pi_i{}^k{}_{;k}, \tag{2.9}$$

where

$$\Pi_{ij} = T_{ij} + t_{ij}. \tag{2.10}$$

The inertia tensor of the fluid is

$$I_{ij} = \int \rho x_i x_j \, d^3V, \tag{2.11}$$

with trace

$$I = g^{ij} I_{ij} \tag{2.12}$$

and tracefree part

$$Q_{ij} = I_{ij} - \frac{1}{3} I g_{ij} = \int \rho\left(x_i x_j - \frac{1}{3}\delta_{ij} r^2\right) d^3V. \tag{2.13}$$

To calculate time derivatives of this tensor at a given instant, we use the dynamical equations (2.3)–(2.5), (2.9), and (2.10). By integrating over all space and using the divergence theorem to eliminate surface terms, we obtain for matter with compact support,

$$\begin{aligned} I_{ij,0} &= \int \rho_{,0} x_i x_j \, d^3V = -\int (\rho v^k)_{;k} x_i x_j \, d^3V \\ &= 2\int \rho v_{(i} v_{j)} \, d^3V. \end{aligned} \tag{2.14}$$

Similarly,

$$I_{ij,00} = 2\int \Pi_{ij} \, d^3V, \tag{2.15}$$

$$\begin{aligned} I_{ij,000} = \int \{ &-g_{ij}[2p_{,0} + \rho(\Phi_{,0} + \Phi_{;k}v^k)] \\ &+ 4[p_{;(i} + \rho\Phi_{;(i}]v_{j)} - \Phi_{,0}\Phi_{;ij}/\pi\} d^3V. \end{aligned} \tag{2.16}$$

For the initial dust ball described by (2.1) and (2.2) we may easily solve (2.3), choosing the zero of potential at the origin. For results, see Appendix B.

To further develop this model, it is natural to work in spherical coordinates. In addition, to establish correspondence with the quasi-Newtonian formalism described in Sec. IV, we choose a reference frame with freely falling origin. Equations (2.3)–(2.5) then take the form[4]

$$r^{-1}(r\Phi^*)_{,11} + r^{-2}\Phi^{*:A}{}_{:A} = 4\pi\rho, \tag{2.17}$$

$$\rho_{,0} - (r^2\rho v_1)_{,1}/r^2 - (\rho v_B)^{:B}/r^2 = 0, \tag{2.18}$$

$$\begin{aligned} (\rho v_A)_{,0} &- (r^2\rho v_A v_1)_{,1}/r^2 - (\rho v_A v_B)^{:B}/r^2 \\ &- p_{,A} - \rho\Phi^*{}_{,A} = 0, \end{aligned} \tag{2.19}$$

$$\begin{aligned} (\rho v_1)_{,0} &- (r^2\rho v_1 v_1)_{,1}/r^2 - (\rho v_1 v_A)^{:A}/r^2 + \rho v_A v^A/r^3 \\ &- p_{,1} - \rho\Phi^*{}_{,1} = 0. \end{aligned} \tag{2.20}$$

Here v_1 and v_A are the covariant velocity components and Φ^* the Newtonian potential in this frame. As boundary conditions, we require that Φ^* and its gradient vanish at the origin, in accord with the free-fall behavior of the origin world line.

To determine $I_{ij,000}$ for our spherical dust model from (2.16) we must still compute $\Phi^*{}_{,0}$. From the time derivative of (2.17) and (2.18), we obtain

$$(r\Phi^*{}_{,0})_{,11} = 4\pi r\rho_{,0} = 4\pi(r^2\rho v_1)_{,1}/r. \tag{2.21}$$

At this point, we must proceed carefully since ρ has a step discontinuity at $r = R$ so that the right side of (2.21) acts as a δ-function source. To compute the magnitude of the resulting jump in the radial derivative of $\Phi^*{}_{,0}$ we integrate (2.21) across the discontinuity

$$\begin{aligned} \Delta(\Phi^*{}_{,01}) &\equiv R^{-1}\int_{R-}^{R+} (r\Phi^*{}_{,0})_{,11} \, dr \\ &= 4\pi R^{-1}\int_{R-}^{R+} (r^2\rho v_1)_{,1} r^{-1} \, dr \\ &= [-4\pi R^{-2}(r^2\rho v_1)]_{R-} = -4\pi k v R^3 Y_2. \end{aligned} \tag{2.22}$$

Despite the density discontinity, $\Phi^*{}_{,0}$ itself is continuous at $r = R$. Now, it is straightforward to determine $\Phi^*{}_{,0}$ from

$$(r\Phi^*{}_{,0})_{,11} = \begin{cases} 20\pi k v r^3 Y_2, & r < R, \\ 0, & r > R, \end{cases} \tag{2.23}$$

together with the conditions at $r = R$. For results see Appendix B.

In a similar way, we may calculate $\Phi^*{}_{,00}$. Here, the Poisson equation has source terms involving both δ functions and their first derivatives. The source for $\Phi^*{}_{,000}$ is even more singular, containing second derivatives of δ functions. We avoid these pathologies by reformulating the calculation in terms of integrals of these quantities. A useful trick is to switch from Poisson equations of the form

$$\nabla^2 f = S, \tag{2.24}$$

with singular sources S, by introducing a new variable τ defined by[4]

$$\tau \equiv \int \frac{f}{r}, \tag{2.25}$$

which satisfies the smoothed equation

$$r^2\nabla^2\tau = \int rS \tag{2.26}$$

by virtue of the operator identity $[r^2\nabla^2, \int r^{-1}] = 0$. Jumps in τ and its derivatives at $r = R$ may be calculated by integrating the Poisson equation to get

$$\Delta(\tau_{,1}) = R^{-2}\Delta\left(\int\int rS\right), \tag{2.27}$$

$$\Delta(\tau) = R^2\Delta\left(\int\int\int rS\right). \tag{2.28}$$

This procedure is used to compute $\int \Phi^*{}_{,000}/r$. For results, see Appendix B.

Now that we can compute time derivatives of fluid quantities and the Newtonian potential we may return to the evaluation of $I^{ij}{}_{,000}$. For our dust model it is now straightforward to show that (2.16) reduces to

$$I_{ij,000} = \tfrac{128}{75}\pi^2 k^2 v R^7(\delta_{i3}\delta_{j3} - \tfrac{1}{3}\delta_{ij}). \tag{2.29}$$

This has vanishing trace, so $Q_{ij,000} = I_{ij,000}$.

III. THE QUASI-NEWTONIAN FORMALISM

For a given Newtonian space-time, such as the dust model in the previous section, there exists a procedure[1,2] for constructing a λ-dependent sequence of quasi-Newtonian general relativistic space-times. These space-times are described on a common manifold in such a way that they share a family of null cones emanating from a geodesic world line. In a null coordinate system based upon these null cones, with $x^\alpha = (x^0, x^1, x^A) = (u, r, \theta, \phi)$,

$$ds^2 = [e^{2\lambda^2\beta}(1+\lambda^2 W/r) - \lambda^4 r^2 h^{AB}U_A U_B]du^2 + 2\lambda e^{2\lambda^2\beta}\,du\,dr + 2\lambda^3 r^2 U_A\,du\,dx^A - \lambda^2 r^2 h_{AB}\,dx^A\,dx^B, \tag{3.1}$$

where $h^{AB}h_{BC} = \delta^A{}_B$, $\det(h_{AB}) = \sin^2\theta$, and $h_{AB} = q_{AB} + \lambda^2\gamma_{AB}$. The factors of λ ensure that (3.1) induces a Newton–Cartan geometry in the limit $\lambda = 0$. In this version of Newtonian theory, the absolute time slices are null hypersurfaces. Smoothness at the origin of the null coordinates must be interpreted in terms of the local Fermi coordinates $t = u + \lambda r$, $x = r\sin\theta\cos\phi$, $y = r\sin\theta\sin\phi$, $z = r\cos\theta$.

The matter source consists of the λ-dependent ideal fluid energy momentum tensor

$$T_{\mu\nu} = (\rho + \lambda^2 p)w_\mu w_\mu - \lambda^2 p g_{\mu\nu},$$

where the four-velocity has the form $w_\mu = t_{,\mu} + \lambda^2 v_\mu$. For $\lambda = 0$, the contravariant components satisfy $w^\alpha = (1, v^i)$, with $v^i = (v^1, v^A)$ the polar coordinates of the fluid velocity in the background Newtonian theory.

Einstein's equation $G_{\mu\nu} = -8\pi T_{\mu\nu}$ decomposes into hypersurface equations which determine β, U_A, and W in terms of the gravitational null data $c_{AB} \equiv \gamma_{AB,1}$ and the initial matter data ρ, p, and v_i; a gravitational evolution equation which determines the time derivative of c_{AB}; and the matter evolution equation $T^{\mu\nu}{}_{;\mu} = 0$ which determines the time derivative of the matter data.

These equations are easiest to examine, in terms of a λ expansion, by introducing the spin-weight zero potentials[4] Z and α satisfying $U_A q^A = \eth Z/\sqrt{2}$ and $c_{AB}q^A q^B = \eth^2\alpha$. Here c_{AB} is the "shear tensor" of the null cones and α the "shear potential." The hypersurface equations then take the form

$$-4r\beta_{,1} = J_\beta, \tag{3.2}$$

$$(r^4 Z_{,1})_{,1} = 2r^4(\beta/r^2)_{,1} - (2 + \eth\bar\eth)r^2\alpha + J_Z, \tag{3.3}$$

$$W_{,1} = \frac{1}{4}\bar\eth^2\eth^2\int(\alpha + \bar\alpha) + (2 - \eth\bar\eth)\beta + (1/4r^2)[r^4\eth\bar\eth(Z + \bar Z)]_{,1} + J_W, \tag{3.4}$$

where the J's are hypersurface quantities. More specifically, the λ-expansion coefficients $J_\beta{}^{(n)}$, $J_Z{}^{(n)}$, and $J_W{}^{(n)}$ involve the matter fields ρ, v_α, and p up to order n and the hypersurface gravitational fields β, Z, W, and α up to order $n-2$. The gravitational evolution equation takes the form

$$[r^2(\alpha - Z)]_{,1} = -2\beta + J_\alpha, \tag{3.5}$$

where

$$J_\alpha = 2\lambda r\int\frac{(r^2\alpha_{,0})_{,1}}{r} + K_\alpha, \tag{3.6}$$

and $K_\alpha{}^{(n)}$ involves matter fields up to order n and hypersurface gravitational fields up to order $n-2$.

Equations (3.2)–(3.5) can be combined to yield the Poisson equation

$$r^2\nabla^2(r^2\alpha)_{,1} = r^2 S \equiv J_\beta + J_{Z,1} + (r^2 J_{\alpha,1})_{,1} - 2J_\alpha, \tag{3.7}$$

where

$$\nabla^2 = r^{-2}\frac{\partial}{\partial r}r^2\frac{\partial}{\partial r} + r^{-2}\eth\bar\eth$$

is the Laplacian for Euclidean three-space. The boundary conditions for solving this Poisson equation are that $(r^2\alpha)_{,1}$ and its gradient both vanish at the origin, in accord with the condition that the vertices of the null cones trace out a timelike geodesic. In the Newtonian limit, this leads to a freely falling reference frame with potential Φ^*, as introduced in Sec. II. The method of calculating the various J's has been discussed and their forms given up through $O(\lambda^2)$ terms. Here we will need the J's up through $O(\lambda^3)$. Results to this order are given in Appendix C.

In order for a general relativistic system to have a Newtonian system with potential Φ^* as its limit it is necessary that

$$\eth^2(r\alpha^{(0)})_{,1} = -2\,\eth^2\Phi^*.$$

Here, for simplicity in removing the $\eth$ operators, we set

$$(r^2\alpha^{(0)})_{,1} = -2\mathscr{P}\Phi^*, \tag{3.8}$$

where $\mathscr{P}$ projects out the $l = 0$ and $l = 1$ parts. This removes the monopole and dipole terms in $\alpha^{(0)}$ which play the role of gauge terms and do not affect any quantities of physical interest.

The initial gravitational data α, at $u = u_0$, for the corresponding quasi-Newtonian relativistic system are determined by requiring that (3.8) hold for $u > u_0$, at least in the formal sense of matching time derivatives. The details of this procedure and the salient features of the resulting space-time have been worked out quite generally up through the determination of $\alpha^{(2)}$.[2] Here we need $\alpha^{(3)}$ for the Newtonian model presented in Sec. II. Proceeding at first in general, the derivation begins with the specification of the initial Newtonian matter data and the initial determination of Φ^*. These quantities are then evolved according to Euler's equations. For our particular model in Sec. II, this was carried out up to the determination of $\Phi^*{}_{,000}$. Next, the initial value of $\alpha^{(0)}$ is found from (3.8) and all other initial zeroth-order quantities obtained from the hypersurface equations (3.2)–(3.4). The initial time derivatives of $\alpha^{(0)}$ are then obtained from those of Φ^*, also via (3.6). The time derivatives of the hypersurface equations then give the time derivatives of all other zeroth-order quantities.

Given this start at the $\alpha^{(0)}$ level, the iteration scheme proceeds by determining the initial value of $\alpha^{(1)}$ using the Poisson equation (3.5). Note that $\alpha^{(0)}{}_{,0}$ appears in the source for $\alpha^{(1)}$ but it has already been determined. The time derivatives of $\alpha^{(1)}$ are found from the time derivatives of the Poisson equation. A new feature arises here—the source contains $\rho_{,0}$. Inititially, ρ is chosen to equal the density of the background Newtonian system and has no λ dependence. However, the time dependence of ρ, as determined by the matter evolution equation $T^{\mu\nu}{}_{;\nu}$, will in general lead to λ-dependent time derivatives of ρ even at $u = u_0$. The same considerations apply to the velocity and pressure. The post-Newtonian corrections to the Euler equation, obtained this way, are given in Appendix C, at least to the order required in this paper.

The continuation of this iteration scheme becomes more burdensome at each order because the source for $\alpha^{(n)}$ involves the u derivative of $\alpha^{(n-1)}$, which in turn involves the second u derivative of $\alpha^{(n-2)}$ continuing down to the nth u derivative of the Newtonian potential. Thus to procede from the $(n-1)$ level to the n level requires the solution of n additional Poisson equations and the assembly of n complicated source terms, in which nth-order post-Newtonian corrections appear. The details of this procedure will be explicitly given in the next section when we calculate the initial quasi-Newtonian data, up through $\alpha^{(3)}$, for the Newtonian background presented in Sec. II.

IV. THE QUASI-NEWTONIAN CALCULATION

The leading term in the news function describing gravitational radiation appears at order λ^3, for our quasi-Newtonian model. This is in accord with the Einstein quadrupole radiation formula which involves three time derivatives, each one carrying with it a factor of λ. Thus it will be necessary to calculate α to $O(\lambda^3)$. The time derivative of α also enters the news function but, because of the factor of λ provided by time differentiation, $\alpha_{,0}$ is needed only to $O(\lambda^2)$. As explained in Sec. III, the derivation of these terms requires $\alpha_{,00}$ to $O(\lambda)$, which in turn requires $\alpha_{,000}$ to Newtonian order.

The initial spherical symmetry of our dust model substantially simplifies the initial values of the J's which form the source S for the Poisson equation (3.7). Explicit formulas for the $S^{(n)}$, up to third order, appear in Appendix D, as well as formulas for the u derivatives of these source terms. As explained, in Sec. III, the latter are required to determine the u derivatives of the $\alpha^{(n)}$'s which appear in $S^{(n+1)}$. In obtaining these formulas, we have inverted the $\eth$ operators which occur in (C4) and (C6). In so doing, we have set to zero the arbitrary monopole and dipole gauge terms in the resulting α's.

When carrying out explicit calculations for this model, great care must be taken in handling boundary conditions at the outer boundary of the dust. This requires keeping a "zoology" of the boundary behavior of the various functions involved in order to calculate the boundary contributions to the radial integrations. The introduction of τ, as described in Sec. II, reduces the degree of discontinuity encountered in the Poisson equations. This is necessary to avoid delta function terms at λ^3 order. In terms of τ, Eq. (3.7) takes the form

$$r^2\nabla^2\tau^{(n)} = \int rs^{(n)}, \tag{4.1}$$

where

$$\tau^{(n)} = \int \frac{(r^2\alpha^{(n)})_{,1}}{r}. \tag{4.2}$$

For uniformity of notation we extend these definitions to the Newtonian variables, defining $\tau^{(0)}$ to be $\int \Phi^*/r$, so that $S^{(0)} = 4\pi\rho^{(0)}$ (where $\rho^{(0)}$ is the Newtonian density described in Sec. II). When these methods are used, the entire calculation can be carried out using functions no more discontinuous than a step function. The calculation proceeds in analogy with Eqs. (2.24)–(2.28).

To begin the calculation of α for the current model, Eq. (3.8) is first used to produce $\alpha^{(0)}$. Because of the initial spherical symmetry of the matter distribution, this consists only of a monopole part. As explained above, such terms will be dropped.

It is helpful to carry out the remaining calculation entirely in terms of $\tau^{(n)}$. With $\tau^{(n)}$ in hand, $\alpha^{(n)}$ can be recovered via Eq. (4.2) (using integration by parts to avoid difficulty from boundary terms). Figures 1–3 illustrate the flow of the calculation at these orders. Equation numbers given in the diagrams refer to the equations used to move from one node to the next. It is useful to consider these diagrams as collections of cells, each of which consists of the construction of an $S^{(n)}$ and the solution of a Poisson equation for a new $\tau^{(n)}$. Thus each cell represents one iteration of the quasi-Newtonian scheme. In each case the new source is constructed via Eq. (3.7) (or some u derivative thereof), and each Poisson equation is of the form (4.1). Convenient expressions for the needed source terms can be found in Appendix D. The nth-order source terms contain lower-order hypersurface terms which can be calculated using the hypersurface equations (3.2)–(3.4).

Consider first Fig. 1, and the calculation of $\tau^{(1)}$. The first needed term is $S^{(0)}{}_{,0} \equiv 4\pi\rho^{(0)}{}_{,0}$, derived from the (Newtonian) hydrodynamical equation (2.18). This is used as a source to produce $\tau^{(0)}{}_{,0}$ via the Poisson equation (4.1). The result is essentially $\Phi^*{}_{,0}$, given by Eq. (B2). From $\tau^{(0)}{}_{,0}$ and miscellaneous known matter terms, Eq. (D1) yields $\tau^{(1)}$.

Figure 2 illustrates how the calculation of $\tau^{(2)}$ first involves the time derivative of the $\tau^{(1)}$ calculation to obtain $\tau^{(1)}{}_{,0}$. In a manner exactly analogous to the previous example, $\tau^{(1)}{}_{,0}$ is constructed from the Newtonian quantity $S^{(0)}{}_{,00}$ and terms derivable from previous results (via the hypersurface equations). Given $\tau^{(1)}{}_{,0}$ and various known matter terms, (D2) yields $S^{(2)}$. Then $\tau^{(2)}$ is provided by the inversion of the Poisson equation (4.1).

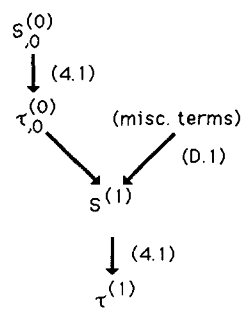


FIG. 1. Diagram of the calculation of $\tau^{(1)}$.

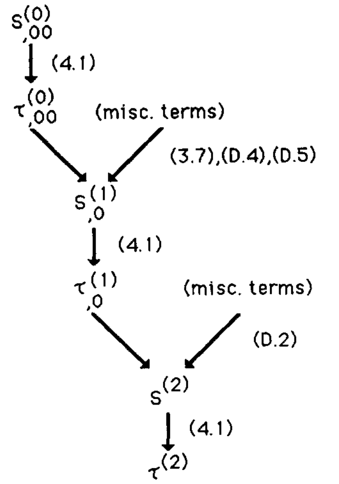


FIG. 2. Diagram of the calculation of $\tau^{(2)}$.

Finally, Fig. 3 shows that the calculation of $\tau^{(3)}$ first involves the calculation of $\tau^{(2)}{}_{,0}$. Equation (D3) is used to construct $S^{(3)}$ from $\tau^{(2)}{}_{,0}$ and lower-order hypersurface terms. At this level, note that relativistic corrections to the Newtonian equations of motion are required in the construction of $S^{(1)}{}_{,00}$ and $S^{(2)}{}_{,0}$. The necessary ingredients are supplied in (D4)–(D11), in Appendix D.

V. THE FLUX AND OTHER ASYMPTOTIC ISSUES

To discuss the asymptotic properties of the quasi-Newtonian system, it is necessary to introduce large r expansions of the metric variables. For simplicity we do this in the axisymmetric Bondi formalism[5] appropriate for the dust model described earlier. The traditional Bondi metric for the λ-dependent system is then

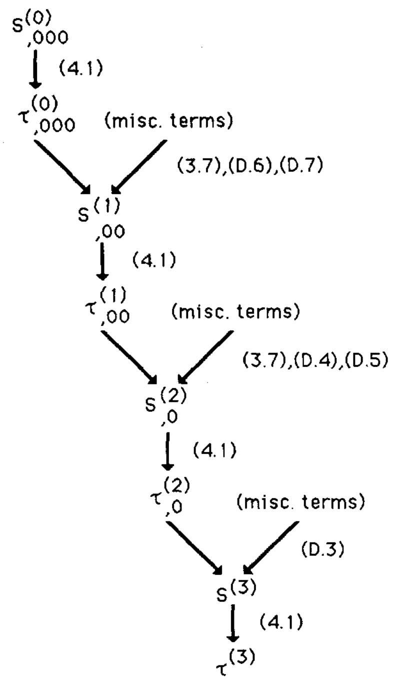


FIG. 3. Diagram of the calculation of $\tau^{(3)}$.

$$ds^2 = (Vr^{-1}e^{2\lambda^2\beta} - \lambda^4 U^2 r^2 e^{2\lambda^2\gamma})du^2 + 2\lambda e^{2\lambda^2\beta}\,du\,dr + 2\lambda^3 U r^2 e^{2\lambda^2\gamma}\,du\,d\theta - \lambda^2 r^2(e^{2\lambda^2\gamma}\,d\theta^2 + e^{-2\lambda^2\gamma}\sin^2\theta\,d\phi^2)\,, \tag{5.1}$$

where

$$h_{AB}\,dx^A\,dx^B = e^{2\lambda^2\gamma}\,d\theta^2 + e^{-2\lambda^2\gamma}\sin^2\theta\,d\phi^2\,, \tag{5.2}$$

$$U_A = Ue^{2\lambda^2\gamma}\delta^2_A\,, \tag{5.3}$$

and

$$\lambda^2 W = V - r\,, \tag{5.4}$$

so that the shear of a null hypersurface of constant u is proportional to $\gamma_{,1}$. The asymptotic behavior of γ is $\gamma = K + c/r + O(r^{-2})$. The hypersurface equations then give the asymptotic forms of the other metric variables. The leading order (constant) terms in the radial expansions of β and U shall be called H and L, respectively.

Consider a conformal factor ω and a new angular coordinate θ_B, such that

$$\omega^2[e^{2\lambda^2K}\,d\theta^2 + e^{-2\lambda^2K}\sin^2\theta\,d\phi^2] = d\theta_B^2 + \sin^2\theta_B\,d\phi^2\,. \tag{5.5}$$

This transformation produces a standard Bondi frame in the large r limit, rather than the nonstandard frame imposed on the original coordinate system by the requirement of smoothness at the origin. Details are given in Ref. 6. The following formulas fix ω and θ:

$$d\theta_B/\sin\theta_B = e^{2\lambda^2K}\,d\theta/\sin\theta\,, \tag{5.6}$$

$$\omega = e^{\lambda^2K}\sin\theta_B/\sin\theta\,. \tag{5.7}$$

In terms of the above quantities, Ref. 6 gives a formula for the leading-order Bondi news function. With the inclusion of the proper λ factors, this formula becomes

$$N = c_{,0}/\lambda^2 + c_{,2}L + (cL_{,2} + cL\cot\theta)/2 + e^{-2\lambda^2K}\omega\sin\theta\,[(e^{2\lambda^2H}\omega)_{,2}(\omega^2\sin\theta)^{-1}]_{,2}/(2\lambda^5)\,. \tag{5.8}$$

In standard Bondi coordinates[5] all the terms on the right-hand side of (5.8) would vanish except the $c_{,0}$ term.

With these results in hand, the news function for the dust model of the previous sections can be evaluated by a straightforward but lengthy calculation. The value of ω, an intermediate result, can be found in Appendix D. The final result for the leading λ dependence is

$$N = \tfrac{64}{75}\pi^2k^2vR^7\sin^2\theta + O(\lambda)\,. \tag{5.9}$$

Note that this result involves the complete cancellation of individual terms in (5.8) which are $O(\lambda^{-2})$ and $O(\lambda^{-1})$. Even in the leading order, $N^{(0)}$ is only about one-tenth the size of its constituent terms. Observe that $N^{(0)}$ has a pure spin-2 quadrupole angular dependence. Also, it has quadratic k dependence, indicative of the matter–matter interactions necessary for a dust system to radiate.

The Einstein quadrupole formula, derived via linearized theory, may be put in the form $N_L = Q_{,000}$, where N_L is the linearized "news" and Q is the transverse Newtonian quadrupole moment $q^iq^jQ_{ij}$. Here q^i is the Cartesian coordinate version of the dyad vector q^A. The term $Q_{,000}$ can be calculated using Eq. (2.29), a purely Newtonian result. For this particular model at $u = u_0$, we find $N^{(0)} = Q_{,000}$. This extends the validity of the Einstein quadrupole formula and suggests its more general validity for quasi-Newtonian systems.

One outstanding asymptotic feature of our solution is the appearance of a $\ln r/r^3$ term in $\alpha^{(3)}$. This behavior is not in accord with the conventional description of asymptotic flatness for which the peeling property of the Weyl tensor implies $\Psi_i = O(r^{-5+i})$, in terms of the Newman–Penrose components of the Weyl tensor. For the dust model of this paper, all the Ψ's are in agreement with this peeling property, at $u = u_0$, except for Ψ_0, which is given by

$$\Psi_0 = -\tfrac{128}{35}\pi^2 k^2 v R^{10} \lambda^3 \sin^2\theta\,[\ln(r/R)/r^5] + O(\lambda/r^5). \tag{5.10}$$

This departure from conventional asymptotic behavior is disconcerting but it should be taken seriously as an indication that the description of radiation from isolated physical systems possibly requires a broader description of asymptotic flatness. In fact, there exists a class of logarithmically asymptotically flat (LAF) space-times which include our dust model as a special case.[7] These LAF space-times are asymptotic solutions of Einstein's equations whose Weyl tensor (in terms of a Penrose compactification[8,9] with conformal factor Ω) vanishes at null infinity as $O(\Omega \ln \Omega)$, rather than $O(\Omega)$ as would be required by the peeling property. In all other regards, LAF space-times have essentially conventional asymptotic features.

In our model, logarithmic asymptotic flatness results from the requirement of a Newtonian limit for times $u > u_0$. Is it possible to restore conventional asymptotic flatness by requiring only a certain degree of tangency in time between the Newtonian system and the $\lambda = 0$ limit of the general relativistic system? By requiring that Eq. (3.8), and its first and second u derivatives, hold *only* at $u = u_0$, the initial gravitational data would be freed up at orders $n \geqslant 3$. This offers a possibility of removing the logarithmic term in $\alpha^{(3)}$. Of course, for this procedure to have any physical justification there should be no concomitant change in the news function, so that the quadrupole radiation formula remains intact. In this regard, note that $\alpha^{(3)}$ determines $\alpha^{(2)}{}_{,0}$ in a nonlocal way (through the gravitational evolution equation) and that the asymptotic parts of both $\alpha^{(3)}$ and $\alpha^{(2)}{}_{,0}$ enter in the news function, so that there are severe global constraints on any acceptable modification of $\alpha^{(3)}$. Nevertheless, it turns out that the source modification[4]

$$\delta S^{(3)} = (20/3r)(r\Phi^* \mathscr{P} \Phi^*{}_{,0})_{,1} \tag{5.11}$$

in the Poisson equation (3.7), leads to a $\delta\alpha^{(3)}$, which satisfies these constraints and which cancels the logarithmic term in $\alpha^{(3)}$. Thus it is possible to restore the peeling property at the expense of a Newtonian limit which is only tangential in time.

There is then a choice of strategies: (i) a strict Newtonian limit may be imposed at the expense of a weakened asymptotic flatness or (ii) conventional asymptotic flatness may be imposed at the expense of a weakened Newtonian limit. From a formal standpoint, the present evidence favors alternative (i) since it fits into a canonical LAF formalism whereas there exists some arbitrariness in the choice of modification $\delta S^{(3)}$ in the case of our dust model and some uncertainty as to how this modification might generalize. A consideration of the Newman–Penrose quantities[10] offers a possible physical explanation for the existence of logarithmic asymptotic behavior.[7] These quantities are well defined and are constants of the motion in the conventional asymptotically flat case. For a system with spatial reflection symmetry, they reduce to the product of the mass and quadrupole moment in the static case, so that it would seem plausible that they also equal this product in the slow-motion approximation implied by the Newtonian limit. But this product is not a Newtonian constant of the motion. The logarithmic behavior resolves this paradox by providing a setting in which the Newman–Penrose quantities cannot be defined. Our dust model lends credence to this explanation since the asymptotic part of $\delta S^{(3)}$ which is associated with the logarithmic behavior is proportional to the time derivative of the product of Newtonian mass and Newtonian quadrupole moment.

VI. SUMMARY

We have demonstrated that the quasi-Newtonian formalism can be implemented to obtain a completely analytic treatment of the initial gravitational radiation from a simple model. As might have been expected from a first attempt at this approach, the calculations were immense. We hope the success of this model will lead to the development of more powerful techniques. In particular, calculational techniques of a more general nature would be desirable, especially to obtain a generalization of the quadrupole radiation formula established in Sec. V.

Our results imply that some weakened form of asymptotic flatness is necessary for the description of quasi-Newtonian systems. One possibility is the LAF version,[7] but we regard the present status of this issue as tentative. Versions of the more general type investigated by Couch and Torrence[11] and by Goldberg and Novak[12,13] may turn out to be more appropriate.

Our formalism supplies a method for prescribing quasi-Newtonian gravitational initial data for numerical evolution of the characteristic initial value problem.[14] However, present attempts to solve the Einstein equation numerically with initial data given by this formalism have met with limited success in calculating the flux. For our dust model, this is because there are individual terms of order $1/\lambda$ and $1/\lambda^2$ in the right-hand side of Eq. (5.8) for the news function. As mentioned earlier, these terms exactly cancel analytically to produce a news function whose leading term $N^{(0)}$ is $O(\lambda^{(0)})$. Numerically, however, the code must be extremely accurate to resolve $N^{(0)}$. In a model without initial spherical symmetry, there are also $O(1/\lambda^3)$ terms which must combine to cancel. In this sense $N^{(0)}$ is produced by terms of order λ^3. Thus to numerically calculate $N^{(0)}$ in the general (nonspherical) case one must resolve the $O(\lambda^{(0)})$ terms to one part in λ^{-3}. If one wishes to operate in the Newtonian regime, where λ is of order 10^{-2}, this implies an accuracy of one part in 10^6 in the numerical solution. A code in which these cancellations are included analytically might be possible, but no such method has yet been developed.

Our approach provides immediate access to other matters of physical importance which have never been explored. The Bondi mass M is a prime example. One would expect that the leading terms in a λ expansion would satisfy $M = m + \lambda^2 B + O(\lambda^3)$, where m is the Newtonian mass and B the Newtonian binding energy. But this elementary demand on the reasonableness of the Bondi mass has never been established for a radiating system. If it is in fact satisfied, what are the post-Newtonian corrections? Similar considerations apply to the reasonableness of angular momentum expressions in general relativity. The prospect of having a firm grasp on the relationship between matter sources and asymptotic behavior at null infinity is very exciting.

ACKNOWLEDGMENT

J. S. Welling and J. Winicour were supported in part by National Science Foundation Grant No. PHY-8403708.

APPENDIX A: CONVENTIONS

Our conventions are adopted to agree, as closely as possible, with those of Refs. 5 and 6. We use signature $+\,-\,-\,-$; units for which $G = c = 1$; Greek letters ranging over 0–3 for space–time indices; lowercase Latin letters ranging over 1–3 for spatial indices; capital Latin letters ranging over 2 to 3 for indices on topologically spherical two-spaces; a semicolon to denote space-time covariant differentiation; a colon to represent covariant differentiation with respect to the unit sphere metric q_{AB}; a comma for partial differentiation; a unit sphere dyad for which $q_{AB} = 2q_{(A}\bar{q}_{B)}$; and Eisenhart's curvature conventions,[15] for which $v_{\mu;\alpha\beta} - v_{\mu;\beta\alpha} = v_\nu R^{\nu}{}_{\mu\alpha\beta}$, $R_{\mu\nu} = R^{\alpha}{}_{\mu\nu\alpha}$, $R = R^{\alpha}{}_{\alpha}$ and the intrinsic scalar curvature of the unit sphere equals -2. The numerical conventions for the unit sphere spin-weight ladder operator $\eth$ are fixed by the examples $v_{A:B}q^A\bar{q}^B = \bar{\eth}(v_A q^A)/\sqrt{2}$, $f^{:A}{}_A = \eth\,\bar{\eth} f$, and $(\bar{\eth}\,\eth - \eth\,\bar{\eth})\eta = 2s\eta$, for a spin-weight s quantity η. We write $f = \Sigma f^{(n)}\lambda^n$ for the expansions of λ-dependent fields. We use the shorthand notation

$$\int f = \int_0^r f(s)ds\,.$$

We define angular functions proportional to the Y_{lm} by

$$Y_2 = 3\cos^2\theta - 1\,,\quad Y_4 = 35\cos^4\theta - 30\cos^2\theta + 3\,,$$
$$Y_6 = 231\cos^6\theta - 315\cos^4\theta + 105\cos^2\theta - 5\,.$$

We denote by $\mathscr{P}$ the operator which projects out $l = 0$ and $l = 1$ harmonics, e.g., $\mathscr{P}(A + B\cos\theta + C\cos^2\theta) = C(\cos^2\theta - \frac{1}{3})$.

APPENDIX B: NEWTONIAN QUANTITIES

Using the methods described in Sec. II, the Newtonian gravitational potential and its time derivatives can be calculated for the spherical dust model described by Eq. (2.1) and (2.2). The results, to the level needed in the calculations that follow, are

$$\Phi^* = \begin{cases} \frac{2}{3}\pi k r^2\,, & r<R\,,\\ 2\pi k R^2 - \frac{4}{3}\pi k R^3/r\,, & r>R\,, \end{cases} \tag{B1}$$

$$\Phi^*{}_{,0} = \begin{cases} \frac{10}{7}\pi k v[r^4 - \frac{21}{25}R^2 r^2]Y_2\,, & r<R\,,\\ \frac{8}{35}\pi k v[R^7/r^3]Y_2\,, & r>R\,, \end{cases} \tag{B2}$$

$$\Phi^*{}_{,00} = \begin{cases} \pi k v^2[(\frac{144}{55}r^6 - \frac{12}{7}R^2 r^4)Y_4 + (\frac{32}{9}r^6 - \frac{48}{35}R^4 r^2)Y_2 + \frac{64}{15}r^6] + \frac{8}{3}\pi^2 k^2 r^2\,, & r<R\,,\\ \pi k v^2[-\frac{48}{385}(R^{11}/r^5)Y_4 - \frac{32}{315}(R^9/r^3)Y_2 + \frac{16}{15}R^6] + \frac{8}{3}\pi^2 k^2 R^2\,, & r>R\,, \end{cases} \tag{B3}$$

$$\int\frac{\Phi^*{}_{,000}}{r} = \begin{cases} \pi k v^3[(\frac{27}{14}r^8 - \frac{216}{143}R^2 r^6)Y_6 + (\frac{243}{91}r^8 - \frac{18}{11}R^4 r^4)Y_4 + (\frac{90}{7}r^8 - \frac{48}{7}R^6 r^2)Y_2] & \\ \quad + \pi^2 k^2 v(\frac{200}{21}r^4 - \frac{2008}{175}R^2 r^2)Y_2\,, & r<R\,,\\ \pi k v^3[(-\frac{288}{7007}(R^{15}/r^7) - \frac{9}{1078}R^8)Y_6 + (-\frac{1152}{25025}(R^{13}/r^5) - \frac{81}{1925}R^8)Y_4 - \frac{6}{7}R^8 Y_2] & \\ \quad + \pi^2 k^2 v[\frac{64}{225}(R^7/r^3) - \frac{704}{315}R^4]Y_2\,, & r>R\,. \end{cases} \tag{B4}$$

APPENDIX C: QUASI-NEWTONIAN FORMULAS

To calculate the J's introduced in Sec. III, Einstein's equation is expanded in terms of β, Z, W, α, ρ, v_α, and p. In this process, it is convenient to express the contravariant two-metric h^{AB} in terms of a dyad $h^{AB} = 2m^{(A}\bar{m}^{B)}$ with the expansion

$$m^A = (1 + \lambda^4 Q)q^A + \lambda^2 P\bar{q}^A\,, \tag{C1}$$

in terms of the auxiliary variables P and Q. The phase freedom in m^A is fixed here by the requirements

$$[m^A]_{r=0} = q^A \quad\text{and}\quad m^A{}_{,1}\bar{m}_A = 0\,.$$

To the order required in this paper, only $P^{(0)}$ appears in the J's and may be reexpressed in terms of $\alpha^{(0)}$ by

$$2P^{(0)} = -\eth^2\int\alpha^{(0)}\,. \tag{C2}$$

Straightforward calculation then leads to[4]

$$J_\beta = -8\pi r^2(\rho + \lambda^2 p)(1 + \lambda v_1)^2 - \tfrac{1}{2}\lambda^2 r^2(\eth^2)\bar{\eth}^2\bar{\alpha} + O(\lambda^4)\,, \tag{C3}$$

$$\begin{aligned}\eth J_Z = {}& 16\pi\lambda\sqrt{2}r^2(\rho + \lambda^2 p)(1 + \lambda v_1)q^A v_A + 2\lambda^2(r^4\beta\,\eth Z_{,1})_{,1} + \lambda^2[r^4(\eth^2\alpha)\bar{\eth}\bar{Z}]_{,1}\\ &+ r^2\lambda^2[P\,\eth\,\bar{\eth}^2\bar{\alpha} - \bar{P}\,\eth^3\alpha - 2(\eth\bar{P})\eth^2\alpha] + O(\lambda^4)\,,\end{aligned} \tag{C4}$$

$$J_W = -4\pi[\rho r^2 + \lambda^2\rho(2r^2\beta + q^{AB}v_A v_B) - \lambda^2 p r^2] - (3\lambda^2/2)\eth\,\bar\eth(P\bar P) + (\lambda^2/2)\eth(P\,\bar\eth\bar P) + (\lambda^2/2)\bar\eth(\bar P\,\eth P)$$
$$- \lambda^2\beta(\bar\eth^2 P + \eth^2\bar P) + 2\lambda^2\beta(1 - \eth\,\bar\eth)\beta - \lambda^2(\eth\beta)\bar\eth\beta - \lambda^2\,\bar\eth(P\,\bar\eth\beta) - \lambda^2\,\eth(\bar P\,\eth\beta)$$
$$+ (\lambda^2/2r^2)[r^4\,\bar\eth(P\,\bar\eth\bar Z) + r^4\,\eth(\bar P\,\eth Z)]_{,1} - (\lambda^2 r^4/4)(\bar\eth\bar Z_{,1})\eth Z_{,1} + O(\lambda^4), \tag{C5}$$

$$\eth^2 J_\alpha = 2\lambda r\,\eth^2\int r^{-1}(r^2\alpha_{,0})_{,1} - 16\pi\lambda^2\rho(v_A q^A)^2 - \lambda^2(rW\,\eth^2\alpha)_{,1} - 4\lambda^2\beta\,\eth^2\beta - 2\lambda^2(\eth\beta)^2$$
$$- 4\lambda^2 P\,\eth\,\bar\eth\beta + 2\lambda^2[(\eth\beta)\bar\eth P - (\bar\eth\beta)\eth P] + \lambda^2 P\,\eth\,\bar\eth[r^2(Z + \bar Z)]_{,1} - (\lambda^2 r^4/2)(\eth Z_{,1})^2$$
$$+ \lambda^2 r^2(\bar\eth\,\eth^2\alpha)\eth Z + (\lambda^2 r^2/2)(\eth^2\alpha)\eth\,\bar\eth(Z - \bar Z) + \lambda^2[(\eth P)(r^2\,\bar\eth\bar Z)_{,1} - (\bar\eth P)(r^2\,\eth Z)_{,1}] + O(\lambda^4). \tag{C6}$$

It is also convenient to have the following combination which appears in S:

$$\eth(J_\beta + J_{Z,1}) = -8\pi r^2\,\eth[(\rho + \lambda^2 p)(1 + \lambda v_1)^2] + 16\pi\lambda\sqrt{2}[r^2(\rho + \lambda^2 p)(1 + \lambda v_1)q^A v_A]_{,1} + \lambda^2\{r^4[2\beta\,\eth Z_{,1} + (\eth^2\alpha)\bar\eth\bar Z]\}_{,11}$$
$$+ \lambda^2[P\,\eth\,\bar\eth^2 - \bar P\,\eth^3 - 2(\eth\bar P)\eth^2](r^2\alpha)_{,1} + O(\lambda^4). \tag{C7}$$

The post-Newtonian equations of motion are obtained from a λ expansion of the matter evolution equation. In this paper, we need only the equations for dust ($p = 0$). A straightforward calculation to the required order leads to

$$O(\lambda^3) = [r^2\rho(1 + \lambda v_1)]_{,0} + \lambda^2(r^2\rho)_{,0}v_0 + \lambda\{r^2\rho[v_0 - (W/r)(1 + \lambda v_1) + \lambda U^A v_A]\}_{,1} - [r^2\rho(1 + \lambda^2 v_0)v_1]_{,1}$$
$$+ \lambda[r^2\rho U^B(1 + \lambda v_1)]_{:B} - [\rho e^{2\lambda^2\beta}(1 + \lambda^2 v_0)h^{AB}v_A]_{:B} + \lambda^2\rho r^2 v_{1,0}v_1 + \lambda^2\rho v_{A,0}q^{AB}v_B, \tag{C8}$$

$$O(\lambda^3) = [r^2\rho(1 + \lambda v_1)^2]_{,0} + (1 + \lambda v_1)\{\rho r^2[-v_1 - (\lambda W/r)(1 + \lambda v_1)]\}_{,1} + \lambda^2(\rho r^2 U^A)_{,1}v_A$$
$$+ (1 + \lambda v_1)\lambda v_0(\rho r^2)_{,1} + \{\rho r^2(1 + \lambda v_1)[U^B\lambda(1 + \lambda v_1) - r^{-2}e^{2\lambda^2\beta}h^{BC}v_C]\}_{:B} + \lambda\rho v_{B,1}q^{BC}v_C, \tag{C9}$$

$$O(\lambda^2) = [r^2\rho(1 + \lambda v_1)v_A]_{,0} + [\rho r^2 v_A(-r^{-1}\lambda W + \lambda v_0 - v_1)]_{,1} + [\rho r^2 v_A(\lambda U^B - r^{-2}q^{BC}v_C)]_{:B}$$
$$- \rho r^2 v_{1:A}(-\lambda r^{-1}W + \lambda v_0 - v_1) - \rho r^2 v_{C:A}(\lambda U^C - r^{-2}q^{BC}v_B) - \rho r^2(1 + \lambda v_1)v_{0:A}. \tag{C10}$$

Here v_0 can be eliminated using

$$v_0 = (1 - \lambda v_1)[(W/2r) + \beta + (1/2)v_1^2 + (1/2r^2)q^{AB}v_A v_B] + (\lambda v_1 W/r) - \lambda U^A v_A + O(\lambda^2), \tag{C11}$$

which follows from the normalization condition $w^\alpha w_\alpha = -1$.

APPENDIX D: EXPLICIT FORMULAS FOR THE MODEL

For $u = u_0$, the source terms are given by

$$r^2 S^{(1)} = -16\pi r^2\rho v_1 - (4/r)(r^3\Phi^*_{,0})_{,1}, \tag{D1}$$

$$r^2 S^{(2)} = -8\pi r^2\rho v_1^2 + (2/r)[r^3(r^2\alpha^{(1)}_{,0})_{,1}]_{,1}, \tag{D2}$$

$$r^2 S^{(3)} = (2/r)[r^3(r^2\alpha^{(2)}_{,0})_{,1}]_{,1} + 2[r^4\beta_{,0}Z^{(1)}_{,1}]_{,11}$$
$$+ \left[2 - \frac{\partial}{\partial r}r^2\frac{\partial}{\partial r}\right][(rW^{(0)}\alpha^{(1)})_{,1} + 4\beta^{(0)}\beta^{(1)}]. \tag{D3}$$

The u derivatives of these source terms, at $u = u_0$, are obtained by differentiating the general formulas given in Appendix C, and then specializing to the initial conditions of our model. In this way, they can be obtained from a λ expansion of the following equations for the u derivatives of J_α and for the combination $J_\beta + J_{Z,1}$:

$$J_{\alpha,0} + O(\lambda^3) = 2\lambda r\int\frac{(r^2\alpha)_{,100}}{r} - \lambda^2(rW\alpha_{,0})_{,1} - 4\lambda^2\beta\beta_{,0}, \tag{D4}$$

$$J_{\beta,0} + J_{Z,10} + O(\lambda^3)$$
$$= -8\pi r^2[\rho(1 + \lambda v_1)^2]_{,0} + 16\pi\sqrt{2}\lambda\,\eth^{-1}[r^2\rho(1$$
$$+ \lambda v_1)q^A v_A]_{,01} + 2\lambda^2[r^4\beta Z_{,01}]_{,11}, \tag{D5}$$

$$J_{\alpha,00} + O(\lambda^2) = -4\lambda r\int\frac{\Phi^*_{,000}}{r}, \tag{D6}$$

$$J_{\beta,00} + J_{Z,100} + O(\lambda^2) = -8\pi r^2[\rho(1 + \lambda v_1)^2]_{,00}$$
$$+ 16\pi\sqrt{2}\lambda\,\eth^{-1}[r^2\rho q^A v_A]_{,100}. \tag{D7}$$

The initial u derivatives of the matter variables appearing in these equations are given by

$$[r^2\rho(1 + \lambda v_1)^2]_{,0} + O(\lambda^3)$$
$$= (\rho r^2 v_1)_{,1} + (\lambda/2)\rho_{,1}rW + \lambda\rho rW_{,1} + (\lambda/2)(\rho r^2 v_1^2)_{,1}$$
$$- \lambda\beta(\rho r^2)_{,1} - \lambda\rho r^2\,\eth\,\bar\eth Z + \lambda^2\rho_{,1}rv_1 W$$
$$+ 2\lambda^2\rho r v_1 W_{,1} + \lambda^2\rho r W v_{1,1}, \tag{D8}$$

$$[r^2\rho(1 + \lambda v_1)q^A v_A]_{,0} + O(\lambda^2)$$
$$= (1/\sqrt{2})\rho r^2\,\eth[(W/2r) + \beta], \tag{D9}$$

$$[r^2(\rho(1 + \lambda v_1)^2]_{,00} + O(\lambda^2)$$
$$= [\rho r^2(W/2r + \beta)_{,1} + (\rho r^2 v_1^2)_{,1}]_{,1}$$
$$+ \rho\,\eth\,\bar\eth(W/2r + \beta) + \lambda[2\rho r^2 v_1\Phi^* - 4\rho r^2 v_1\beta$$
$$- \rho r^2 v_1^3]_{,11} + \lambda(\rho r^2 v_1)_{,1}\Phi^*_{,1} - 2\lambda(\rho r^2\beta_{,0})_{,1}$$
$$+ \lambda(\rho r^2\Phi^*_{,0})_{,1} + \lambda\rho r^2\Phi^*_{,01}$$
$$- \lambda\rho\,\eth\,\bar\eth\left[r^2\alpha_{,0} + 2\int\beta_{,0}\right], \tag{D10}$$

$$[r^2\rho q^A v_A]_{,00} + O(\lambda) = (1/\sqrt{2})\rho r^2\,\eth\Phi^*_{,0}, \tag{D11}$$

which follow from the specialization of the post-Newtonian equations (C8)–(C10) to our model. Note that the ð's appearing in (D9) and (D11) cancel the inverse δ's in (D5) and (D7), respectively.

Using these sources and the formalism described in Secs. III and IV, we can calculate α to $O(\lambda^3)$. These results, plus many needed hypersurface quantities, are given below.

For clarity we introduce $X \equiv r/R$, and denote the value of a quantity f in the interior or exterior of the dust sphere by $f_<$ or $f_>$, respectively. As discussed earlier, monopole terms which are physically irrelevant have been dropped.

The values of the terms of α can be shown to be

$$\alpha^{(0)} = 0\,, \tag{D12}$$

$$\begin{aligned}
\alpha_<{}^{(1)} &= \pi k v R^4 \left[\tfrac{-7}{18} X^4 + X^2 - \tfrac{16}{45} X\right] Y_2\,,\\
\alpha_>{}^{(1)} &= \pi k v R^4 \left[\tfrac{5}{9} X^{-2} - \tfrac{3}{10} X^{-4}\right] Y_2\,,
\end{aligned} \tag{D13}$$

$$\begin{aligned}
\alpha_<{}^{(2)} &= \pi k v^2 R^7 \{ [\tfrac{-6998}{45045} X^7 + \tfrac{192}{385} X^5 - \tfrac{16}{45} X^4 + \tfrac{26}{1225} X^3] Y_4 + [\tfrac{-496}{3465} X^7 + \tfrac{192}{245} X^3 - \tfrac{16}{35} X^2 - \tfrac{16}{105} X] Y_2 \}\,,\\
\alpha_>{}^{(2)} &= \pi k v^2 R^7 \{ [\tfrac{22}{1575} X^{-2} + \tfrac{-24}{2695} X^{-4} + \tfrac{4}{63} X^{-5} - \tfrac{298}{5005} X^{-6}] Y_4 + [\tfrac{64}{315} X^{-2} + \tfrac{-464}{2695} X^{-4}] Y_2 \}\,,
\end{aligned} \tag{D14}$$

$$\begin{aligned}
\alpha_<{}^{(3)} &= \pi^2 k^2 v R^6 Y_2 [\tfrac{-2524}{2625} X^6 + \tfrac{2386}{315} X^4 - \tfrac{312}{25} X^3 + \tfrac{237}{35} X^2 - \tfrac{6136}{7875} X]\\
&\quad + \pi k v^3 R^{10} \{ [\tfrac{-26}{255} X^{10} + \tfrac{324}{455} X^8 - \tfrac{1152}{1001} X^7 + \tfrac{7452}{11011} X^6 - \tfrac{576}{4235} X^5] Y_6\\
&\quad + [\tfrac{-438}{8575} X^{10} + \tfrac{504}{605} X^6 - \tfrac{4608}{3025} X^5 + \tfrac{32}{35} X^4 - \tfrac{3264}{18865} X^3] Y_4 + [\tfrac{-440}{3087} X^{10} + \tfrac{80}{21} X^4 - \tfrac{2304}{343} X^3 + \tfrac{24}{7} X^2 - \tfrac{128}{315} X] Y_2 \}\,,\\
\alpha_>{}^{(3)} &= \pi^2 k^2 v R^6 Y_2 [\tfrac{64}{105} \log(X)/X^4 - \tfrac{2056}{1575} X^{-2} + \tfrac{21767}{7875} X^{-4} - \tfrac{4}{3} X^{-5}]\\
&\quad + \pi k v^3 R^{10} \{ [\tfrac{256}{55055} X^{-5} - \tfrac{270}{11011} X^{-6} + \tfrac{288}{7735} X^{-7} - \tfrac{4}{231} X^{-8}] Y_6\\
&\quad + [\tfrac{-288}{18865} X^{-4} + \tfrac{832}{21175} X^{-5} - \tfrac{102}{4235} X^{-6}] Y_4 + [\tfrac{-48}{1715} X^{-4}] Y_2 \}\,.
\end{aligned} \tag{D15}$$

These quantities are related via the hypersurface equations to the remaining metric variables. To the order needed for the calculation described in Sec. IV, these are

$$\beta_<{}^{(0)} = \pi k r^2\,, \quad \beta_>{}^{(0)} = \pi k R^2\,, \tag{D16}$$

$$\beta_<{}^{(1)} = \pi k v [\tfrac{4}{3} r^5] Y_2\,, \quad \beta_>{}^{(1)} = \pi k v [\tfrac{4}{3} R^5] Y_2 \tag{D17}$$

$$Z^{(0)} = 0\,, \tag{D18}$$

$$\begin{aligned}
Z_<{}^{(1)} &= \pi k v R^4 [\tfrac{73}{630} X^4 + \tfrac{2}{3} X^2 - \tfrac{16}{45} X] Y_2\,,\\
Z_>{}^{(1)} &= \pi k v R^4 [\tfrac{-1}{3} + \tfrac{8}{5} X^{-1} - \tfrac{10}{9} X^{-2} + \tfrac{32}{105} X^{-3} - \tfrac{3}{10} X^{-4}] Y_2\,,
\end{aligned} \tag{D19}$$

$$W_<{}^{(0)} = \tfrac{-2}{3} \pi k r^3\,, \quad W_>{}^{(0)} = \pi k R^3 [2X - \tfrac{8}{3}]\,, \tag{D20}$$

$$\begin{aligned}
W_<{}^{(1)} &= \pi k v R^6 [\tfrac{47}{105} X^6 - \tfrac{4}{3} X^4 + \tfrac{16}{15} X^3] Y_2\,,\\
\mathrm{W}_>{}^{(1)} &= \pi k v R^6 [2X^2 - \tfrac{8}{5} X + \tfrac{32}{35} X^{-1} - \tfrac{3}{5} X^{-2}] Y_2\,.
\end{aligned} \tag{D21}$$

To the order required for the calculation of Sec. IV, the u derivatives of τ, β, and Z are

$$\begin{aligned}
\tau_<{}^{(1)}{}_{,0} &= \pi k v^2 R^7 \{ [\tfrac{-272}{539} X^7 + \tfrac{24}{25} X^5 - \tfrac{8}{21} X^4] Y_4 + [\tfrac{-704}{1225} X^7 + \tfrac{32}{21} X^3 - \tfrac{96}{175} X^2] Y_2 \}\,,\\
\tau_>{}^{(1)}{}_{,0} &= \pi k v^2 R^7 \{ [\tfrac{-12}{385} X^{-4} + \tfrac{8}{105} X^{-5} + \tfrac{36}{1225}] Y_4 + [\tfrac{32}{175} X^{-3} + \tfrac{32}{147}] Y_2 \}\,,
\end{aligned} \tag{D22}$$

$$\begin{aligned}
\tau_<{}^{(1)}{}_{,00} &= \pi k v^3 R^9 \{ [\tfrac{-4}{3} X^9 + \tfrac{23328}{7007} X^7 - \tfrac{288}{143} X^6] Y_6 + [\tfrac{-3504}{3184} X^9 + \tfrac{1008}{275} X^5\\
&\quad - \tfrac{144}{55} X^4] Y_4 + [\tfrac{-1760}{441} X^9 + \tfrac{320}{21} X^3 - \tfrac{576}{49} X^2] Y_2 \} + \pi^2 k^2 v [- \tfrac{52}{35} X^5 + \tfrac{1472}{63} X^3 - \tfrac{1654}{75} X^2] Y_2\,,\\
\tau_>{}^{(1)}{}_{,00} &= \pi k v^3 R^9 \{ [\tfrac{-8}{1617} + \tfrac{64}{1001} X^{-6} - \tfrac{540}{7007} X^{-7}] Y_6 + [\tfrac{-48}{1925} + \tfrac{288}{5005} X^{-4} - \tfrac{1152}{13475} X^{-5}] Y_4 + [\tfrac{-32}{63}] Y_2 \}\\
&\quad + \pi^2 k^2 v R^5 [\tfrac{-346}{315} + \tfrac{208}{225} X^{-3}] Y_2\,,
\end{aligned} \tag{D23}$$

$$\begin{aligned}
\tau_<{}^{(2)}{}_{,0} &= \pi k v^3 R^{10} \{ [\tfrac{-36}{85} X^{10} + \tfrac{1944}{1001} X^8 - \tfrac{15552}{7007} X^7 + \tfrac{1104}{1573} X^6] Y_6\\
&\quad + [\tfrac{-21024}{79625} X^{10} + \tfrac{1344}{605} X^6 - \tfrac{4032}{1375} X^5 + \tfrac{48}{49} X^4] Y_4 + [\tfrac{-528}{637} X^{10} + \tfrac{480}{49} X^4 - \tfrac{640}{49} X^3 + \tfrac{144}{35} X^2] Y_2 \}\\
&\quad + \pi^2 k^2 v R^6 [\tfrac{-4372}{2835} X^6 + \tfrac{3988}{245} X^4 - \tfrac{3116}{135} X^3 + \tfrac{1422}{175} X^2] Y_2\,,\\
\tau_>{}^{(2)}{}_{,0} &= \pi k v^3 R^{10} \{ [\tfrac{1152}{55055} X^{-5} - \tfrac{60}{1001} X^{-6} + \tfrac{432}{10829} X^{-7}] Y_6\\
&\quad + [\tfrac{384}{35035} X^{-3} - \tfrac{144}{2695} X^{-4} + \tfrac{4992}{105875} X^{-5}] Y_4 + [\tfrac{64}{3185} X^{-3}] Y_2 \} + \pi^2 k^2 v R^6 [\tfrac{646}{95} - \tfrac{165296}{99225} X^{-3} + \tfrac{16}{21} X^{-4}] Y_2\,,
\end{aligned} \tag{D24}$$

$$\beta_<{}^{(0)}{}_{,0} = \tfrac{5}{2} \pi k v r^4 Y_2\,, \quad \beta_>{}^{(0)}{}_{,0} = \tfrac{1}{2} \pi k v R^4 Y_2\,, \tag{D25}$$

$$Z_<{}^{(0)}{}_{,0} = \pi k v R^3 [\tfrac{3}{7} X^3 + \tfrac{4}{3} X] Y_2\,, \quad Z_>{}^{(0)}{}_{,0} = \pi k v R^3 [X^{-1} + \tfrac{8}{35} X^{-4}] Y_2\,. \tag{D26}$$

The conformal factor ω, used in the construction of the "news" in Sec. V, has the following value:

$$\omega = 1 - \tfrac{8}{5}\pi k v R^{5}\lambda^{3} Y_2 - \tfrac{1}{70}\pi k v^2 R^{8}\lambda^{4}[9Y_4 + 20Y_2] + \tfrac{5783}{3675}\pi^2 k^{2} v R^{7}\lambda^{5} Y_2 + \cdots . \tag{D27}$$

[1] J. Winicour, J. Math. Phys. **24**, 1193 (1983).

[2] J. Winicour, J. Math. Phys. **25**, 2506 (1984).

[3] R. A. Isaacson, J. S. Welling, and J. Winicour, Phys. Rev. Lett. **53**, 1870 (1984).

[4] See Appendix A for notation and conventions.

[5] H. Bondi, M. G. J. Van der Burg, and A. W. K. Metzner, Proc. R. Soc. London Sec. A **269**, 21 (1962).

[6] R. A. Isaacson, J. S. Welling, and J. Winicour, J. Math. Phys. **24**, 1824 (1983).

[7] J. Winicour, Found. Phys. **15**, 605 (1985).

[8] R. Penrose, Phys. Rev. Lett. **10**, 66 (1963).

[9] R. Geroch, in *Asymptotic Structures of Space-Time*, edited by F. P. Esposito and L. Witten (Plenum, New York, 1977), p. 1.

[10] E. T. Newman and R. Penrose, Proc. R. Soc. London Sec. A **305**, 175 (1968).

[11] W. E. Couch and R. J. Torrence, J. Math. Phys. **13**, 69 (1972).

[12] S. Novak and J. N. Goldberg, Gen. Relativ. Gravit. **13**, 79 (1981).

[13] S. Novak and J. N. Goldberg, Gen. Relativ. Gravit. **14**, 655 (1982).

[14] R. A. Isaacson, J. S. Welling, and J. Winicour, in *Problems of Collapse and Numerical Relativity*, edited by D. Bancel and M. Signore (Reidel, Dordrecht, 1984), p. 239.

[15] L. P. Eisenhart, *Riemannian Geometry* (Princeton U. P., Princeton, NJ, 1925).

Collision-free gases in spatially homogeneous space-times

R. Maartens and S. D. Maharaj
Department of Applied Mathematics, University of the Witwatersrand, 1 Jan Smuts Avenue, Johannesburg 2001, South Africa

(Received 19 December 1984; accepted for publication 16 May 1985)

The kinematical and dynamical properties of one-component collision-free gases in spatially homogeneous, locally rotationally symmetric (LRS) space-times are analyzed. Following Ray and Zimmerman [Nuovo Cimento B **42**, 183 (1977)], it is assumed that the distribution function f of the gas inherits the symmetry of space-time, in order to construct solutions of Liouville's equation. The redundancy of their further assumption that f be based on Killing vector constants of the motion is shown. The Ray and Zimmerman results for Kantowski–Sachs space-time are extended to all spatially homogeneous LRS space-times. It is shown that in all these space-times the kinematic average four-velocity u^i can be tilted relative to the homogeneous hypersurfaces. This differs from the perfect fluid case, in which only one space-time admits tilted u^i, as shown by King and Ellis [Commun. Math. Phys. **31**, 209 (1973)]. As a consequence, it is shown that all space-times admit nonzero acceleration and heat flow, while a subclass admits nonzero vorticity. The stress π_{ij} is proportional to the shear σ_{ij} by virtue of the invariance of the distribution function. The evolution of tilt and the existence of perfect fluid solutions are also discussed.

I. INTRODUCTION

General relativistic kinetic theory provides a self-consistent approach to the study of matter, which incorporates the particle structure and dispenses with the additional phenomenological equations required in the usual fluid models of matter. Its study can therefore help to deepen an understanding of matter in general relativity, and to clarify the basis of the idealized fluid picture. Further, the theory allows for a unified treatment (at a classical level) of massive particles and massless particles representing radiation. A self-gravitating gas of galactic particles provides a cosmological model that is in many ways more fundamental and realistic than the usual fluid models. For a gas of massless particles moving in a background geometry, we obtain a model of radiation in cosmology.

The difficulty of translating these advantages of the theory into useful results is the extreme complexity of the equations. In the general case of a multicomponent charged self-gravitating gas with collisions, the Einstein–Maxwell–Boltzmann system of equations governs the gas behavior and the space-time geometry. At this general level, the theory is mainly concerned with constructing collision integrals and deriving the EMB system and its general properties (see, for example, Refs. 1–4 for further discussion and references). To make further progress towards an understanding of the matter and geometry, it is necessary to impose simplifying assumptions. These assumptions are broadly of three types.

(1) Assumptions about the nature and behavior of the gas. For example, under certain conditions on the collision integral, approximation theories may be developed, leading to foundations for relativistic thermodynamics.[4–6]

(2) Assumptions on the space-time geometry, considered either as background (test gas), or as generated by the gas (self-gravitating gas). An example of the former type is the analysis of photons or neutrinos in a symmetric cosmological background.[7] Self-gravitating gases have been investigated in static spherically symmetric space-times,[8,9] Robertson–Walker space-times,[10–13] and Bianchi space-times.[4,9,12,14–16]

(3) Assumptions which impose symmetries on the distribution function f. For example, isotropy of f in momentum space leads in the uncharged collision-free case to a Robertson–Walker geometry (apart from special cases),[17] and to weaker, but significant, restrictions on the geometry in the case with collisions.[18] In addition, a more general class of "matter symmetries" admitted by f (uncharged collision-free case) has been considered.[19,20]

In this paper our principal concern is to establish relationships between the space-time geometry and the kinematics and dynamics of the gas, in a cosmological context. We provide a development and extension of the work by Ray and Zimmerman.[12] We follow them in making the following assumptions, listed according to the types described above.

(I) The gas is a one-component collision-free gas without charge.

The physical situations we have in mind include galactic particles (considered as identical for simplicity—the extension to a mass spectrum is not difficult[10]), and photons propagating through the universe without interacting with the matter. The collision-free assumption means that the distribution function is a constant of the motion, so that we can appeal to geometric methods in trying to construct reasonable solutions to Liouville's equation. A generalization to a multicomponent collision-free gas is relatively straightforward, since the various components are necessarily noninteracting.

(II) The space-time geometry is spatially homogeneous and LRS.

These space-times admit a G_r of motions, $r = 4$ or 6, acting transitively on spacelike hypersurfaces. The case $r = 6$ gives the Robertson–Walker geometry, which is the standard cosmological geometry.[21] Uncharged collision-free gases in these space-times have received considerable attention, and we will therefore confine attention to the $r = 4$ geometries. These have been studied extensively as cosmolo-

 0022-2488/85/112869-12$02.50

gical models with fluid and electromagnetic source terms.[22–24] For collision-free gases, Ray and Zimmerman have studied a self-gravitating gas in one of the seven such geometries.[9,12] This paper aims to give a systematic analysis for all the $r = 4$ geometries.

(III) The distribution function is invariant under the G_4 of motions.

In general the distribution function f will *not* inherit the space-time symmetries (cf. the somewhat analogous situation regarding the electromagnetic field tensor[25]). Even when the gas itself generates the space-time geometry, it is only an *average* of f over momentum space at each point that has the same symmetry as the metric. (Ellis *et al.*[13] provide an example of an anisotropic f which generates a Robertson–Walker geometry.) However, for a gas under assumptions (I) and (II), the further invariance assumption (III) is a natural and important starting point.

Ray and Zimmerman make the following further assumption.

(IV) The distribution function is a function of Killing vector constants of the motion.

These are linear first integrals generated by the inner product of a Killing vector with the geodesic four-momentum.[14]

We will show in Sec. IV that (IV) in fact follows from (III) via Liouville's equation. First we give a brief summary of the relevant kinetic theory and invariance conditions in Sec. II, followed by a review of the LRS spatially homogeneous metrics and their G_4's of motions in Sec. III. In Sec. IV we derive a unified expression for the invariant distribution functions in all the space-times. This is used to solve Liouville's equation. Using these solutions, we calculate in Sec. V the kinematic quantities of the test gas. In Sec. VI we consider the neutral self-gravitating gas; in particular, we show that the average four-velocity u^i can be tilted in all the space-times. The stress tensor π_{ij} is shown to be proportional to the shear tensor σ_{ij}. Such a relationship emerges as a linear term in approximation theory of nonequilibrium gases.[6] Our result holds exactly for a collision-free gas. Further dynamical properties are discussed in Sec. VI. In Sec. VII we give some concluding remarks.

Notation: We follow the notation and conventions as used in standard references.[1,21] In particular: commas denote partial derivatives; semicolons denote covariant derivatives; $\mathscr{L}_X$ is the Lie derivative along the vector field X; $a, b, c,\ldots$ are indices in the orthonormal tetrad basis; $i, j, k,\ldots$ are indices in the coordinate basis; $\partial_i = \partial/\partial x^i$; and $I, J, K,\ldots$ are indices in the Lie algebra of Killing vectors. Coordinates $x^i = (t, x^\alpha) = (t, x, y, z)$, $\alpha = 1,2,3$ are chosen so that ∂_t is timelike future-directed, and ∂_α are spacelike. Individual components of tensors are labeled using $(0,1,2,3)$ for tetrad components and (t, x, y, z) for coordinate components.

II. COLLISION-FREE GAS

In this section we review briefly the relativistic kinetic theory of an uncharged collision-free one-component gas,[1] and discuss the invariance of the distribution function f under Killing symmetries.[12,19]

Let f be a one-particle distribution function for a Gibbs ensemble of systems of particles of rest mass m (>0). Then $f(x, p)$ determines the number of particles at event x with a four-momentum p. The space-time geometry at x is determined by the metric g. Here, f is a non-negative smooth function on $P(m)$, the phase space for rest mass m. At each x, the momentum space P_x is the region in the tangent space consisting of future-directed nonspacelike tangent vectors. The mass shell at x is the hypersurface $P_x(m) = \{ p \in P_x \mid g_{ij} p^i p^j = -m^2 \}$, in particular, $P_x(0)$ is the future light cone at x. The mass shells are the fibers of phase space: $P(m) = \cup P_x(m)$. Here, $P(m)$ is a hypersurface in the phase space for all rest masses $P = \cup P_x$, which is a region of the tangent bundle. Local coordinates x^i on space-time induce local coordinates (x^i, p^j) on P, where $p = p^i \partial_i$. Then $P(m)$ is given locally by

$$g_{ij}(x) p^i p^j = -m^2. \tag{2.1}$$

Choosing p^α as coordinates on the mass shell, we have local coordinates (x^i, p^α) on $P(m)$, and then p^t is determined on $P(m)$ by (2.1).

A. Liouville's equation

The possible particle motions are given by (2.1) and

$$p^i = \frac{dx^i}{dv}, \quad \frac{dp^i}{dv} = -\Gamma^i{}_{jk} p^j p^k, \tag{2.2}$$

since free particles not subject to collisions follow geodesics. The $\Gamma^i{}_{jk}$ are the connection coefficients and v is an affine parameter: $v = (\text{proper time})/m$ for $m > 0$. The family of intersecting geodesics given by (2.2) is naturally lifted $(x^i(v) \to (x^i(v), dx^i/dv))$ into a nonintersecting congruence of phase orbits in phase space. The tangent vector field to the phase orbits is $(dx^i/dv)\partial/\partial x^i + (dp^i/dv)\partial/\partial p^i$, and by (2.2) this gives the Liouville vector field

$$L = p^i \left(\frac{\partial}{\partial x^i} - \Gamma^j{}_{ik} p^k \frac{\partial}{\partial p^j} \right), \tag{2.3}$$

a directional derivative along the phase flow. Since $Lm = 0$, L is tangent to $P(m)$ and so the restriction of L to $P(m)$ is just[5]

$$L = p^i \left(\frac{\partial}{\partial x^i} - \Gamma^\alpha{}_{ij} p^j \frac{\partial}{\partial p^\alpha} \right). \tag{2.3'}$$

Since the gas is collision-free, f is constant along the phase flow, giving Liouville's equation

$$Lf = 0. \tag{2.4}$$

B. Kinematics and dynamics

f is assumed to vanish sufficiently rapidly at infinity on the mass shell, so that its moments over the mass shell are bounded. The first moment

$$n^i = \int p^i f \pi_m \tag{2.5}$$

is the particle four-current density, and the second moment

$$T^{ij} = \int p^i p^j f \pi_m \tag{2.6}$$

is the energy–momentum tensor. In (2.5) and (2.6), π_m is the covariant volume element on $P_x(m)$, given in any basis by

$$\pi_m = (-\det g)^{1/2}\, dp^{123}/(-p_0), \tag{2.7}$$

and the integrals are taken over the whole mass shell $P_x(m)$. Here, n defines the kinematic average four-velocity u of the gas by

$$n^i = Nu^i, \quad u^i u_i = -1, \tag{2.8}$$

where N is the number density. The kinematic quantities associated with the average behavior of the gas are given by the u congruence

$$u_{i;j} = \sigma_{ij} + \tfrac{1}{3}\theta h_{ij} + \omega_{ij} - \dot{u}_i u_j; \tag{2.9}$$

σ_{ij} is the shear ($\sigma_{[ij]} = \sigma_{ij} u^j = \sigma^i{}_i = 0$), $\theta = u^i{}_{;i}$ the expansion, ω_{ij} the vorticity ($\omega_{(ij)} = \omega_{ij} u^j = 0$), and $\dot{u}_i = u_{i;j} u^j$ the acceleration. Here, $h_{ij} = g_{ij} + u_i u_j$ is the projection tensor into the rest space of u. The average dynamic quantities of the gas are given by T,

$$T^{ij} = \mu u^i u^j + p h^{ij} + q^i u^j + u^i q^j + \pi^{ij}, \tag{2.10}$$

where μ is the energy density, p the isotropic pressure, q^i the heat flow ($q_i u^i = 0$), and π^{ij} the stress (anisotropic pressure) ($\pi_{ij} u^j = 0 = \pi^i{}_i$) (all quantities are measured by a u observer). For the average behavior of the gas to be that of a perfect fluid, the condition is[4]

$$q^i = 0 = \pi^{ij} \tag{2.11}$$

(as measured by the kinematic average velocity u). From (2.4)–(2.6) follow the conservation equations

$$n^i{}_{;i} = 0 = T^{ij}{}_{;j}. \tag{2.12}$$

For a test gas, the kinetic energy–momentum tensor (2.6) makes a negligible contribution to the total energy–momentum tensor, so that the space-time geometry is determined as a background geometry independently of the distribution function, which is restricted only by Liouville's equation. For a self-gravitating gas, the kinetic energy–momentum tensor is the source of the gravitational field. In this case the self-consistent Einstein–Liouville system of equations holds:

$$G^{ij} = T^{ij}, \quad Lf = 0, \tag{2.13}$$

where G is the Einstein tensor. [Consistency follows from (2.12) and the contracted Bianchi identities.]

C. Invariance of the distribution function

If space-time admits a G_r of motions, generated by r Killing vector fields X_I, with structure constants $C^K{}_{IJ}$, then

$$\mathcal{L}_I g_{ij} = 0, \quad [X_I, X_J] = C^K{}_{IJ} X_K, \tag{2.14}$$

where $\mathcal{L}_I \equiv \mathcal{L}_{X_I}$. Does the invariance of the metric lead to any invariance conditions on the distribution function f? Following Ehlers[1] we define the invariance of f under a space-time vector field X by $\widetilde{X} f = 0$, where $\widetilde{X}$ is the complete lift[26] of $X = X^i \partial_i$: $\widetilde{X} = X^i \partial/\partial x^i + X^i{}_{,j}\, p^j\, \partial/\partial p^i$. [This is the natural definition since $\widetilde{X}$ generates the local G_1 $(x,p) \to (\phi_s x, \phi_{s*} p)$, where $x \to \phi_s x$ is the local G_1 generated by X.] Hence the condition for f to be invariant under G_r is

$$\widetilde{X}_I f \equiv X_I{}^i \frac{\partial f}{\partial x^i} + X_I{}^i{}_{,j}\, p^j \frac{\partial f}{\partial p^i} = 0, \quad \text{all } I. \tag{2.15}$$

For a test gas, (2.15) is clearly not in general satisfied. But even for a self-gravitating gas, (2.15) will not hold in general: $\mathcal{L}_I g = 0 \Rightarrow \mathcal{L}_I G = 0$ and so by (2.13), $\mathcal{L}_I T = 0$; however, this imposes by (2.6) invariance only on an *average* of f over the mass shell. Specifically, Berezdivin[19,27] shows that

$$\mathcal{L}_I n^i = \int p^i (\widetilde{X}_I f) \pi_m, \quad \mathcal{L}_I T^{ij} = \int p^i p^j (\widetilde{X}_I f) \pi_m. \tag{2.16}$$

From (2.16) follows the invariance condition on f for a self-gravitating gas in a space-time admitting Killing vectors X_I

$$\int p^i p^j (\widetilde{X}_I f) \pi_m = 0. \tag{2.17}$$

Equation (2.15)$\Rightarrow$(2.17), but not conversely, as shown by Ellis *et al.*,[13] who present in $k = 0$ Robertson–Walker space-time an f that is invariant under the translation subgroup but not under the rotational subgroup of the G_6.

Liouville's equation implies that f is a constant of the motion. Now Killing vectors lead to constants of motion

$$y_I(x,p) = g_{ij}(x) X_I{}^i(x) p^j \Rightarrow L y_I = 0. \tag{2.18}$$

If f is assumed to be a function only of the y_I, and to be invariant under the X_I, then from (2.14), (2.15), and (2.18) follows the result[1]

$$f = F(y_I) \text{ and } \widetilde{X}_I f = 0, \quad \text{for all } I \Rightarrow y_K C^K{}_{IJ} \frac{\partial F}{\partial y_I} = 0. \tag{2.19}$$

What appears not to have been recognized is the simple integrability condition following from (2.19) after differentiation with respect to y_J

$$C^J{}_{IJ} \frac{\partial F}{\partial y_I} = 0. \tag{2.20}$$

III. SPATIALLY HOMOGENEOUS LRS SPACE-TIMES

The space-times considered in this paper admit a G_4 of motions transitive on spacelike hypersurfaces. Hence at each point there is a G_1 of isotropy about a preferred direction. These are the class A space-times in MacCallum's elegant classification.[28] (MacCallum's class A also includes the G_4 transitive on timelike hypersurfaces.) This classification is a considerable improvement on the previous Kruchkovich–Petrov (KP) classification,[29] used by Ray and Zimmerman.[12] The metrics for these space-times may be described in a unified form using canonical coordinates (t,x,y,z)

$$ds^2 = -dt^2 + A(t)^2 [dx + lh(y) dz]^2 + \exp(2jx) B(t)^2 [dy^2 + h'^2 dz^2], \tag{3.1}$$

where $t = \text{const}$ gives the homogeneous hypersurfaces $S_3(t)$; ∂_t is geodesic and normal to $S_3(t)$ and t is proper time; ∂_α are tangent to $S_3(t)$; h is given by

$$h(y) = (y, \cos y, \cosh y), \quad \text{for } k = (0, 1, -1); \tag{3.1$'$}$$

and $l,k,j \in \{0, \pm 1\}$ are parameters distinguishing the different geometries. Seven different group types occur, classified according to their G_3 subgroups, as shown in Table I. The subgroups may act simply transitively (G_3 on S_3) or multiply transitively (G_3 on S_2). Class A1c is the exceptional Kan-

TABLE I. Group classification for spatially homogeneous LRS space-times.

MacCallum type of G_4	KP type of G_4	l	k	j	Bianchi–Behr type of G_3 on S_3	Bianchi–Behr type of G_3 on S_2
A1a	VI_4	0	0	0	I, VII_0	VII_0
A1b	VII	0	-1	0	III	VIII
A1c	VIII	0	1	0	—	IX
A2a	III	1	0	0	II	—
A2b	VIII	1	1	0	IX	—
A2c	VII	1	-1	0	III, VIII	—
A3	V	0	0	1	V, VII_h	VII_0

towski–Sachs space-time[30]—it does not admit a simply transitive G_3. By (3.1), an orthonormal one-form basis is

$$E^a = \{dt, A(dx + lh\,dz),\ \exp(jx)B\,dy,\ \exp(jx)Bh'\,dz\}. \tag{3.2}$$

In this basis, the components of the Einstein tensor constructed from the metric (3.1) give[28]

$$a\neq b \Rightarrow G^{ab}=0, \quad \text{except } G^{01}\neq 0 \text{ in class A3}; \qquad G^{22}=G^{33}. \tag{3.3}$$

The inverse metric of (3.1) is

$$g^{ij} = \mathrm{diag}(-1, A^{-2} + h^2 L, M, L + (1-l)h'^{-2}M) - 2hL\delta^{(i}{}_x\,\delta^{j)}{}_z, \tag{3.4}$$

where $L \equiv l(h'B)^{-2}$, $M \equiv \exp(-2jx)B^{-2}$. The components $\Gamma^\alpha{}_{ij}p^ip^j$ of the connection coefficients in class A space-times will be needed for the Liouville vector (2.3'). These components are given in the Appendix.

MacCallum does not give explicit forms for the basis of Killing vectors. These are given by Petrov[29] (p. 229), except for the A2b and A2c geometries. Petrov's coordinates (x^1,x^2,x^3,x^4) correspond to (x,y,z,t) in (3.1), except in class A2a, where $x^1=y$, $x^2=x$. For classes A2b and A2c, we use the forms given by Kramer *et al.*[31] (p. 127) in complex coordinates (w,t,ζ). The coordinate transformation $t = x+z$, $\zeta = \sqrt{2}\,e^{iz}(\tan y/2,\ \tanh y/2)$ for $k=(1,-1)$, brings their metric (with $l=-k/2$) into the form (3.1) ($l=1, j=0$), after a rescaling of time. Then we can transform the basis of Killing vectors into our canonical coordinates. The results of these transformations are collected in Table II. In the basis of Killing vectors, X_4 generates the local rotational symmetry; X_2, X_3 span the plane of symmetry; and X_1 is along the preferred direction. (X_2 in class A1b disagrees with Petrov's expression[29] for his corresponding X_1 (p. 229). The $C^4{}_{32}$ for class A1c disagrees with the corresponding $C^2{}_{31}$ in Kantowski and Sachs[30] [Eq. (3)].) Table II gives the structure constants, as well as the one-form $C^J{}_{IJ}$, which distinguishes class A3 as exceptional ($C^J{}_{IJ}\neq 0$). This exceptional geometric behavior of A3 [confirmed by (3.3)] will be reflected in exceptional behavior of the distribution function (Sec. IV).

We will now examine the invariant tensors. The metric tensor and all tensors covariantly constructed from it, such as the Einstein tensor, are clearly invariant under the class A G_4 of motions. We now find the most general invariant vector and symmetric spatial (0,2) tensor in class A space-times.

In the non-LRS Bianchi space-times[23] with a maximal G_3 on S_3, the existence of an invariant triad E_α spanning S_3 implies that any vector of the form $F(t)\partial_t + F^\alpha(t)E_\alpha$ is invariant. In the LRS case, however, only one direction at each point in S_3 is invariant, leading to stringent restrictions on invariant vectors and spatial tensors. From Table II we see that for all class A space-times

$$\mathscr{L}_I\,\partial_t = 0 = \mathscr{L}_I\,\partial_x, \quad \text{all } I.$$

Hence ∂_x is the preferred direction in S_3. Since the G_4 is transitive on S_3, we also have $\mathscr{L}_I F = 0$ for all $I \Leftrightarrow F = F(t)$. Then it follows that in all class A space-times

$$\mathscr{L}_I\,v^i = 0, \quad \text{for all } I \Leftrightarrow v = F(t)\partial_t + H(t)\partial_x, \tag{3.5}$$

for some F, H. Any invariant vector lies in the t-x plane at each point, and has constant components over S_3 (in the canonical coordinate basis). Thus there is a one-parameter family of invariant unit vectors. By (3.1) and (3.5) any invariant unit timelike vector u (future directed) satisfies

$$\mathscr{L}_I\,u^i = 0, \quad \text{for all } I, \qquad u^iu_i = -1 \Leftrightarrow u = \cosh\psi\,\partial_t + A^{-1}\sinh\psi\,\partial_x, \tag{3.6}$$

while any invariant spacelike vector c orthogonal to u satisfies

$$\mathscr{L}_I\,c^i = 0, \quad \text{for all } I, \qquad c^ic_i = 1, \quad c^iu_i = 0 \Leftrightarrow c = \sinh\psi\,\partial_t + A^{-1}\cosh\psi\,\partial_x, \tag{3.7}$$

where $\psi = \psi(t)$ is the (hyperbolic) angle of tilt,[32] measuring the deviation of u from the normal ∂_t to S_3. Here, u is normal to $S_3 \Leftrightarrow \psi = 0 \Leftrightarrow c$ is tangent to S_3.

The projection tensor $h_{ij} = g_{ij} + u_i u_j$ into the rest space of u [satisfying (3.6)] is invariant. Hence any invariant symmetric (0,2) tensor in the rest space of u satisfies

$$\mathscr{L}_I\,S_{ij} = 0, \quad \text{for all } I, \qquad S_{[ij]} = 0 = S_{ij}u^j \Leftrightarrow S_{ij} = P(t)c_ic_j + Q(t)h_{ij}, \tag{3.8}$$

for some P, Q. If S is also trace-free ($S^i{}_i = 0$), then

$$S_{ij} = P(t)(c_ic_j - \tfrac{1}{3}h_{ij}). \tag{3.9}$$

The restrictions on invariant tensors given by (3.5)–(3.9) will be important when we consider the average kinematical and dynamical properties of the gas (Secs. V and VI).

IV. INVARIANT DISTRIBUTION FUNCTIONS

We now investigate the consequences of assumptions (I)–(IV) (Sec. I), using the theory and results of Secs. II and III. The analysis applies to both test gases and self-gravitating gases—that is, only Liouville's equation (2.4) is imposed. The further restrictions arising when Einstein's field equations are imposed (self-gravitating gas) are taken up in Sec. VI.

A. Killing vector constants of motion

We begin with the simplest case: all assumptions (I)–(IV) are imposed. The metric is given by (3.1) and Table I, with structure constants and Killing vectors of the G_4 given in Table II. The distribution function automatically satisfies Liouville's equation since it is a function of constants of the

TABLE II. Killing vectors for spatially homogeneous LRS space-times.

MacCallum type of G_4	X_1	X_2	X_3	X_4	Positive structure constants	$C^J{}_{IJ}$
A1a	∂_x	∂_y	∂_z	$-z\,\partial_y + y\,\partial_z$	$C^3_{24}=1$ $C^2_{43}=1$	0
A1b	∂_x	$-\cos z\,\partial_y + (\coth y \sin z - 1)\partial_z$	$\cos z\,\partial_y - (\coth y \sin z + 1)\partial_z$	$\sin z\,\partial_y + \coth y \cos z\,\partial_z$	$C^2_{24}=1$ $C^3_{43}=1$ $C^4_{23}=2$	0
A1c	∂_x	$\sin z\,\partial_y + \cot y \cos z\,\partial_z$	$\cos z\,\partial_y - \cot y \sin z\,\partial_z$	∂_z	$C^4_{23}=1$ $C^2_{34}=1$ $C^3_{42}=1$	0
A2a	∂_x	∂_z	$-\partial_y + z\,\partial_x$	$-z\,\partial_y + \frac{1}{2}(z^2 - y^2)\partial_x + y\,\partial_z$	$C^1_{23}=1$ $C^3_{24}=1$ $C^2_{43}=1$	0
A2b	∂_x	$-\cos z \csc y\,\partial_x + \sin z\,\partial_y + \cos z \cot y\,\partial_z$	$\sin z \csc y\,\partial_x + \cos z\,\partial_y - \sin z \cot y\,\partial_z$	$\partial_z - \partial_x$	$C^2_{34}=1$ $C^3_{42}=1$ $C^4_{23}=1$ $C^1_{23}=1$	0
A2c	∂_x	$-\cos z\,\mathrm{csch}\, y\,\partial_x + \sin z\,\partial_y + \cos z \coth y\,\partial_z$	$\sin z\,\mathrm{csch}\, y\,\partial_x + \cos z\,\partial_y - \sin z \coth y\,\partial_z$	$\partial_z - \partial_x$	$C^2_{34}=1$ $C^3_{24}=1$ $C^4_{32}=1$ $C^1_{32}=1$	0
A3	$-\partial_x + y\,\partial_y + z\,\partial_z$	∂_y	∂_z	$-z\,\partial_y + y\,\partial_z$	$C^3_{31}=1$ $C^3_{24}=1$ $C^2_{21}=1$ $C^2_{43}=1$	$-2\delta_I{}^1$

motion [assumption (IV)]. The problem reduces to the solution of the coupled first-order linear system (2.19) for all seven class A geometries, extending the Ray and Zimmerman result[12] for class A1c. The integrability condition (2.20) is trivially satisfied in classes A1 and A2; but in class A3, it imposes the restriction $\partial F/\partial y_1 = 0$. This restriction is severe: the system (2.19) in class A3 is inconsistent unless $F = \text{const}$, and the constant must be zero since f has bounded moments. In classes A1 and A2, the system (2.19) may be solved by the method of characteristics,[33] and we obtain the distribution function in terms of the constants of motion y_I [(2.18)][34]

$$\text{A1:}\quad f = F[y_1,(1+k)(2-k)(y_2^2+y_3^2) + 2k^2y_4^2 + k(1-k)y_2y_3], \tag{4.1}$$

$$\text{A2:}\quad f = F[y_1, y_2^2+y_3^2+ky_4^2+2(1+k-k^2)y_1y_4],$$

$$\text{A3:}\quad f=0,$$

where F is non-negative and suitably bounded on the mass shell. By (2.18), (3.1), and Table II, the solutions (4.1) may be reexpressed in terms of (x^i, p^α)

$$f(x,p) = (1-j)F[A(t)^2(p^x + lh(y)p^z), B(t)^4(p^{y^2} + h'(y)^2p^{z^2})]. \tag{4.2}$$

[For convenience, we use the same symbol F in (4.1) and (4.2).] Equation (4.2) agrees with Ray's corrected version[9] of the Ray–Zimmerman solution[12] for class A1c ($k=1$, $l=0=j$). Using the orthonormal tetrad (3.2), we obtain from (4.2) a unified form for all class A geometries, which shows clearly how f inherits the space-time symmetry.

Theorem 4.1: In the spatially homogeneous LRS (class A) space-times with a neutral one-component collision-free gas, the distribution function invariant under the G_4 of motions and based on Killing vector constants of motion is spatially homogeneous in space-time and LRS in momentum space. Specifically,

$$f(x,p) = (1-j)F[A(t)p^1, B(t)^2(p^{2^2}+p^{3^2})], \tag{4.3}$$

where F is an arbitrary smooth non-negative function suitably bounded on the mass shell.

In Theorem 4.1, A and B are metric scale functions (in the canonical coordinate system) [(3.1)]; $p^a = E^a(p)$ are components of the four-momentum in the canonical orthonormal tetrad (3.2); and j is a group parameter given in Table I. The exceptional geometry of class A3 ($j=1$), previously noted in Sec. III, leads to a trivial distribution function under the assumptions of Theorem 4.1. An invariant f in class A3 space-time cannot be a function only of Killing vector constants of the motion; alternatively, if f is based on Killing vector constants of motion, then it cannot be invariant under the class A3 G_4 (it will in general be non-LRS in momentum space and inhomogeneous in space-time).

B. Invariance without Killing constants

We now drop the assumption (IV), that f be based on Killing vector constants of the motion. Our motivation is firstly to establish the relationship of this assumption to the invariance assumption (III); secondly, to construct nontrivial solutions for class A3. We begin by solving the invariance condition (2.15), a coupled linear system which again yields to solution by characteristics, though with more difficulty.[34] The solutions for all class A space-times may be expressed in the single form

$$f(x,p) = F[t, p^x + lh(y)p^z, \exp(2jx)(p^{y^2} + h'(y)^2p^{z^2})]. \tag{4.4}$$

In particular, we get nontrivial solutions for class A3. Using the tetrad (3.2), (4.4) becomes

$$f(x,p) = F[t, A(t)^{-1}p^1, B(t)^{-2}(p^{2^2}+p^{3^2})], \tag{4.5}$$

which shows that f is spatially homogeneous in space-time and LRS in momentum space—that is, this feature of Theorem 4.1 follows purely from G_4 invariance.[35]

The invariant distribution functions (4.4) and (4.5) do not satisfy Liouville's equation (2.3′) and (2.4). This must now be imposed on (4.4), using the connection coefficients given in the Appendix. Writing F in (4.4) as $F = F(t,u,v)$ [so that $u = A^{-1}p^1$, $v = B^{-2}(p^{2^2}+p^{3^2})$], we find that Liouville's equation reduces to

$$p^t\left[\frac{\partial F}{\partial t} - 2\frac{A'}{A}u\frac{\partial F}{\partial u} - 4\frac{B'}{B}v\frac{\partial F}{\partial v}\right] + jv\left[\left(\frac{B}{A}\right)^2\frac{\partial F}{\partial u} - 2u\frac{\partial F}{\partial v}\right] = 0, \tag{4.6}$$

where, by (2.1), $p^t = (m^2 + A^2u^2 + B^2v)^{1/2}$. In classes A1 and A2 ($j=0$), (4.6) may be integrated

$$\text{A1,A2:}\quad F = F(A^2u, B^4v). \tag{4.7}$$

Equation (4.7) gives exactly the same f as (4.3). So we have shown the redundancy of assumption (IV) in classes A1 and A2.

Theorem 4.2: In class A1 and A2 space-times with a neutral one-component collision-free gas, if the distribution function is invariant under the G_4 of motions and satisfies Liouville's equation, then it is based on Killing vector constants of motion.

(Then the results of Theorem 4.1 hold.) The distribution function must always be based on constants of the motion by Liouville's equation. For a one-component gas, there are at most six functionally independent such constants.[36] In class A space-times, four constants are generated by the Killing vectors. Theorem 4.2 implies that if f depends on any constants not arising from Killing vectors,[37] then f cannot be G_4 invariant.

Class A3 is again exceptional. With $j=1$ in (4.6) we have been unable to find the general solution (the special solutions with $A \propto B$ give the degenerate Robertson–Walker subclass of A3).

Theorem 4.3: In class A3 space-time under the conditions of Theorem 4.2, the distribution function f is spatially homogeneous in space-time and LRS in momentum space. Specifically, f is of the form (4.5) and subject to (4.6).

V. KINEMATICS OF THE GAS

We use the forms derived in Sec. IV for invariant distribution functions in class A space-times to obtain the kinematic quantities associated with the average behavior of the gas (Sec. II). As in Sec. IV, only Liouville's equation is imposed, so that the results hold for both test gases and self-gravitating gases. The further restrictions imposed by Einstein's field equations in the self-gravitating case are investigated in the next section.

By (2.15) and (2.16), the four-current density n is invariant under the G_4 for invariant f, and then (3.5) gives n

$$\widetilde{X}_I f = 0, \quad \text{for all } I \Rightarrow \mathscr{L}_I\, n^i = 0 \Rightarrow n = \alpha(t)\,\partial_t + \beta(t)\,\partial_x, \tag{5.1}$$

for some α, β. Since ∂_t is normal to the homogeneous hypersurfaces $S_3(t)$, we see that n, and hence the kinematic average four-velocity u [(2.8)], will be tilted relative to $S_3(t)$ if $\beta \neq 0$. We now show that β may be nonzero. In the orthonormal tetrad (3.2), we have $g_{ab} = \operatorname{diag}(-1,1,1,1)$, so that volume element (2.7) on the mass shell is

$$\pi_m = dp^{123}/p^0, \tag{5.2}$$

where by (2.1),

$$p^0 = (m^2 + p^{1^2} + p^{2^2} + p^{3^2})^{1/2}, \tag{5.3}$$

and the mass shell is given by (5.3) with $-\infty < p^1, p^2, p^3 < \infty$. The tetrad components of n are given in terms of the coordinate components by

$$n^a = E^a(n) = (n^t, A\,(n^x + lhn^z), \exp(jx)Bn^y, \exp(jx)Bh'n^z). \tag{5.4}$$

By (2.5) and (5.2)

$$n^a = \int p^a \frac{1}{p^0} f(x,p)\,dp^{123}, \tag{5.5}$$

where the integral is over all of $\mathbb{R}^3$. Then $f(x,p)$ in (5.5) is given by (4.3) in classes A1 and A2

$$\text{A1,A2:} \quad f(x,p) = F\,[Ap^1, B^2(p^{2^2} + p^{3^2})], \tag{5.6}$$

and by (4.5) in class A3

$$\text{A3:} \quad f(x,p) = F\,[t, A^{-1}p^1, B^{-2}(p^{2^2} + p^{3^2})]. \tag{5.7}$$

In (5.7), F is subject to (4.6). In all classes, F is even in p^2 and p^3, so that the integrand in (5.5) for $a = 2,3$ is an odd function, integrated from $-\infty$ to $+\infty$, and hence the integral vanishes: $n^2 = 0 = n^3$. By (5.4), this gives $n^y = 0 = n^z$, so that $n^a = (n^t, An^x, 0, 0)$ in all classes. Now F is not necessarily even in p^1, so that (5.5) may give $n^1 = n^1(t) \neq 0$ in all classes. Hence $n^x = n^x(t)$ may be nonzero. Clearly (5.5) $\Rightarrow n^0 = n^0(t) \neq 0$, so that $n^t = n^t(t) \neq 0$. We have confirmed (5.1) and shown that β may be nonzero.

Equation (5.1) implies invariance of the number density N and kinematic average four-velocity u [(2.8)]: $0 = \mathscr{L}_I\, n^i = \mathscr{L}_I (Nu^i) \Rightarrow \mathscr{L}_I\, N = 0$, $\mathscr{L}_I\, u^i = 0$ (using $u_i\, \mathscr{L}_I\, u^i = 0$). Hence $N = N(t)$, and u is given by (3.6); that is,

$$Lf = 0 = \widetilde{X}_I f, \quad \text{all } I \Rightarrow u = \cosh\psi(t)\partial_t + A(t)^{-1}\sinh\psi(t)\partial_x, \quad \psi \neq 0 \text{ in general} \tag{5.8}$$

($\psi = 0 \Leftrightarrow \beta = 0$). Then the invariant unit spatial direction orthogonal to u is given by (3.7)

$$c = \sinh\psi\,\partial_t + A^{-1}\cosh\psi\,\partial_x. \tag{5.9}$$

[By (5.1), (4.5), and the above analysis of the integral (5.5), it is clear that (5.8) holds whether or not Liouville's equation is imposed.]

The invariance of the kinematic average four-velocity follows from the *invariance of the distribution function.*[38] The invariance of u imposes severe restrictions on the kinematic quantities [(2.9)]

$$\mathscr{L}_I\, u^i = 0 = \mathscr{L}_I\, g_{ij} \Rightarrow \mathscr{L}_I\, u_{i;j} = 0 \Rightarrow \mathscr{L}_I\, \theta = 0,$$
$$\mathscr{L}_I\, \dot{u}^i = 0, \quad \mathscr{L}_I\, \sigma_{ij} = 0 = \mathscr{L}_I\, \omega_{ij}.$$

By (3.5)–(3.9) these invariance conditions imply (cf. Ref. 22)

$$\theta = \theta(t), \quad \dot{u}^i = a(t)c^i, \quad \sigma_{ij} = \sqrt{3}\,\sigma(t)(c_i c_j - \tfrac{1}{3}h_{ij}), \quad \omega^i = \omega(t)c^i, \tag{5.10}$$

where a is the magnitude of the acceleration, σ is the magnitude of the shear[21] ($\sigma^2 = \frac{1}{2}\sigma_{ij}\,\sigma^{ij}$, $\sigma_{ij} = 0 \Leftrightarrow \sigma = 0$), $\omega^i = -\frac{1}{2} *(u \wedge du)^i$ is the vorticity vector,[39] and ω is the magnitude of the vorticity ($\omega_{ij} = 0 \Leftrightarrow \omega^i = 0 \Leftrightarrow \omega = 0$).

The kinematic quantities are thus determined by invariance of the distribution function up to four functions of t. Direct calculation gives

$$\begin{aligned}
\theta &= \cosh\psi(\log AB^2\cosh\psi)' + 2j\sinh\psi/A,\\
a &= \sinh\psi(\log A\sinh\psi)',\\
\sigma &= \sqrt{3}(\theta/3 - \cosh\psi(\log B)' - j\sinh\psi/A),\\
\omega &= l(-A\sinh\psi/2B^2).
\end{aligned} \tag{5.11}$$

By (5.11), $\omega = 0 \Leftrightarrow l\psi = 0$; $a = 0 \Leftrightarrow (A\sinh\psi)' = 0$; $\theta = 0 \Leftrightarrow (AB^2\cosh\psi)' + 2jB^2\sinh\psi = 0$; and $\sigma = 0 \Leftrightarrow (A \times \cosh\psi/B)' - j\sinh\psi/B = 0$. The class A space-times degenerate into higher symmetry if $A'B' = 0$, $(A/B)' = 0$, or $(AB^2)' = 0$. We exclude these possibilities so as to confine attention to nondegenerate class A space-times. We can collect the above results in the following.

Theorem 5.1: For a neutral one-component collision-free gas in spatially homogeneous LRS (class A, nondegenerate) space-time, if the distribution function f is invariant under the G_4 of motions, then the number density, the kinematic average four-velocity u^i, and the kinematic quantities are invariant under the G_4 [(5.8)–(5.11)]; u^i may be tilted relative to the homogeneous hypersurfaces. (This also holds when Liouville's equation is dropped.) In all class A space-times, the acceleration vanishes if and only if the tilt is zero or arcsinh $(A_0/A(t))$. The vorticity always vanishes in classes A1 and A3; in class A2 it vanishes only if the tilt vanishes. The expansion and shear cannot both vanish; either vanishes if and only if the tilt is a solution to

$$(AB^s\cosh\psi)' + sjB^s\sinh\psi = 0,$$
$$\text{where } s = 2\,(\theta = 0) \text{ or } s = -1 \quad (\sigma = 0).$$

The results of Theorem 5.1 are consistent with the conservation equations $T^{ij}{}_{;j} = 0$, since these follow directly from Liouville's equation.[1] However, the conditions on the tilt ψ for the vanishing of various kinematic quantities are unlikely to be consistent with Einstein's field equations. Under what geometric conditions will the tilt vanish? A partial answer is given by the following.

Theorem 5.2: If the distribution function of Theorem 5.1 is also invariant under spatial inversion in momentum space, then the kinematic average four-velocity is orthogonal to the homogeneous hypersurfaces.

Such invariance is expressed using the orthonormal tetrad[3] $f(x, p^0, -p^1, -p^2, -p^3) = f(x, p^0, p^1, p^2, p^3)$. By (5.6) and (5.7), this implies f is even in p^1, and so by (5.5) $n^1 = 0 \Rightarrow n^x = 0$ [by (5.4)]$\Rightarrow u = \partial_t$ (cf. Ref. 39, p. 244).

It is interesting to compare our results with those of King and Ellis,[32] who analyze self-gravitating perfect fluids in spatially homogeneous space-time. They find that for an expanding fluid, the acceleration vanishes if and only if the tilt vanishes or $dp/d\mu = 0$; that there are no shear-free tilted perfect fluids; and that in the class A space-times, only A3 admits a tilted perfect fluid, which has zero vorticity. We see, as expected, that the collision-free test gas allows for a wider range of behavior.

Finally, we look at the conservation equation (2.12) for n. By (2.8), this is $N_{,i}\, u^i + N\theta = 0$ (and hence $\theta = 0 \Leftrightarrow N = \text{const}$). Using (5.8) and (5.11) this can be integrated

$$N = N_0 \exp\left(-2j\int dt\,\frac{\tanh\psi}{A}\right)(AB^2\cosh\psi)^{-1}.$$

[Compare Eq. (3.1) in Ref. 32.]

VI. DYNAMICS: THE SELF-GRAVITATING GAS

The distribution functions invariant under the class A G_4 of motions and satisfying Liouville's equation are given by (5.6) and (5.7) [subject to (4.6)]. We now impose the field equations (2.13), so that the gas itself generates the class A geometry. The geometric form (3.3) of the Einstein tensor implies that the energy–momentum tensor satisfies

$$a \neq b \Rightarrow T^{ab} = 0, \quad \text{except } T^{01} \neq 0 \text{ in general in class A3;} \qquad T^{22} = T^{33}. \tag{6.1}$$

By (2.6), using (5.2), we have

$$T^{ab} = \int p^a p^b \frac{1}{p^0} f(x,p)\,dp^{123}, \tag{6.2}$$

where p^0 is given by (5.3). We want to determine whether (6.1) imposes any further restrictions on the forms (5.6) and (5.7) of the distribution function $f(x,p)$. First we note that the symmetry and evenness of f in p^2 and p^3 imply $T^{22} = T^{33}$ and $T^{02} = T^{03} = T^{12} = T^{13} = T^{23} = 0$, since integration in (6.2) is over $-\infty < p^2, p^3 < \infty$. Hence the only possible restrictions can arise from T^{01}. In class A3, $T^{01} \neq 0$ in general, so that clearly no restrictions are placed on the functional form of $f(x,p)$ as given by (5.7). In classes A1 and A2, we have $T^{01} = 0$

$$\int_{\mathbf{R}^3} p^1 F\,[A(t)p^1, B(t)^2(p^{2^2} + p^{3^2})]\,dp^{123} = 0, \tag{6.3}$$

by (6.1), (6.2), and (5.6). Ray asserts[9] that (6.3) implies F must be even in p^1. Then by (5.5) this implies $n^1 = 0$, so that the kinematic average four-velocity u is forced to be orthogonal to the homogeneous hypersurfaces. This is Ray's conclusion in class A1c. However, in our opinion Ray's assertion and his conclusion are false. The function

$$G(Ap^1) = \exp -(Ap^1 - \alpha)^2 - (\alpha/\beta)\exp -(Ap^1 - \beta)^2, \tag{6.4}$$

where α, β are nonzero constants, provides a counterexample. The distribution function

$$F = G(Ap^1) H\,[B^2(p^{2^2} + p^{3^2})] \tag{6.5}$$

is *not* even in p^1 provided $\alpha \pm \beta \neq 0$, but it does satisfy (6.3), since

$$\int_{-\infty}^{\infty} sG(s)ds = 0 \tag{6.6}$$

(see Ref. 40, p. 307). Further, (6.5) is a physically reasonable distribution function if H is non-negative and suitably bounded on the mass shell, and if $\alpha\beta < 0$. Thus, while (6.3) does restrict the functional form of the distribution function f, the restriction is *not* that f must be even in p^1.

Finally, we show that the condition (6.3) does not force u to be orthogonal. To do this, we show that the counterexample (6.5) leads to a nonzero n^1. Using (6.4) and (5.3) in (5.5), we obtain

$$n^1(t) = A(t)^{-1}\int_{\mathbf{R}^2} dp^{23}\, H\,[B(t)^2(p^{2^2} + p^{3^2})] \times I\,[A(t), p^{2^2} + p^{3^2}], \tag{6.7}$$

where

$$I = \int_{-\infty}^{\infty} s(M^2 + s^2)^{-1/2} G(s)ds,$$

with $s \equiv Ap^1$, $M^2 \equiv A^2(m^2 + p^{2^2} + p^{3^2})$. From (6.7), $I \neq 0 \Rightarrow n^1 \neq 0$, and so our task reduces to showing that $I \neq 0$. Now only the odd part of G makes a contribution to I, so that

$$I = 2\int_0^{\infty} s(M^2 + s^2)^{-1/2} G_-(s)ds, \tag{6.8}$$

where

$$G_-(s) \equiv (G(s) - G(-s))/2 = e^{-s^2}[e^{-\alpha^2}\sinh 2\alpha s - (\alpha/\beta)e^{-\beta^2}\sinh 2\beta s]$$

(with $\alpha\beta < 0 \neq \alpha + \beta$). We note that $sG_-(s)$ is bounded, smooth, and changes sign once for $s > 0$. And $(M^2 + s^2)^{-1/2}$ is bounded, smooth, and monotonic for $s \geqslant 0$. Also, (6.6) $\Rightarrow \int_0^\infty sG_-(s)ds = 0$ [$sG_-(s)$ describes equal areas above and below the s axis]. Hence we can apply the second mean value theorem[40] to (6.8) [using $(M^2 + s^2)^{-1/2} \to 0$ as $s \to \infty$], to get $I = 2M^{-1}\int_0^a sG_-(s)ds$ (where $0 < a < \infty$)$\Rightarrow I \neq 0$ by (6.6). [Qualitatively, what happens is that the precise balance implied by (6.6) is broken by the scaling factor $(M^2 + s^2)^{-1/2}$ in (6.8).]

We can summarize the preceding results in the following.

Theorem 6.1: For a neutral one-component collision-free gas in class A space-time, if the distribution function is invariant under the G_4 of motions and satisfies Liouville's equation, then its functional form [(5.6) and (5.7)] is consistent with Einstein's field equations. Further, the kinematic average four-velocity may be tilted in all class A space-times.

We note that the restriction (6.3) on the distribution function in classes A1 and A2, while admitting noneven F, is a severe constraint on F. A more detailed investigation of the field equations may provide a more lucid condition for $T^{01}=0$ than (6.3), and thus help to clarify the circumstances under which tilted four-velocity occurs. It is also interesting to compare our result with that of King and Ellis.[32] The comparison is more meaningful than in Sec. V, since the gas is self-gravitating. For a self-gravitating perfect fluid, only class A3 admits a tilted u. For the self-gravitating collision-free gas, class A3 is also exceptional, in that the restriction $T^{01}=0$ does not hold, allowing for a wider range of behavior of the distribution function. However, the average four-velocity of the gas, unlike the fluid four-velocity, may still be tilted in all spatially homogeneous LRS space-times.

A. Dynamical quantities

By (2.16), invariance of f leads to invariance of the energy–momentum tensor T

$$\widetilde{X}_I f=0 \Rightarrow \mathscr{L}_I T^{ij}=0. \tag{6.9}$$

This is consistent with the field equations

$$\mathscr{L}_I g_{ij}=0, \quad G^{ij}=T^{ij} \Rightarrow \mathscr{L}_I T^{ij}=0,$$

but we note that *invariance of the energy–momentum tensor follows directly from invariance of the distribution function*—without the Einstein (or even the Liouville) equations. Equation (6.9) imposes severe restrictions on the dynamical quantities defined by T [(2.10)], given the invariance of u [(5.8)]

$$\mathscr{L}_I(\mu u^i u^j + p h^{ij} + q^i u^j + q^j u^i + \pi^{ij})=0,$$

$$\mathscr{L}_I u^i = 0 \Rightarrow \mathscr{L}_I \mu = 0 = \mathscr{L}_I p, \quad \mathscr{L}_I q^i=0, \quad \mathscr{L}_I \pi^{ij}=0;$$

and by (3.5)–(3.9), these invariance conditions imply

$$\mu=\mu(t), \quad p=p(t), \quad q^i=Q(t)c^i, \qquad \pi^{ij}=\sqrt{3}\,\pi(t)(c^ic^j-\tfrac{1}{3}h^{ij}), \tag{6.10}$$

where Q is the magnitude of the heat flow, and π is the magnitude of the anisotropic stress ($\pi^2=\frac{1}{2}\pi_{ij}\pi^{ij}$, $\pi^{ij}=0 \Leftrightarrow \pi=0$).

Comparing (5.10) and (6.10) we obtain the transport law

$$\widetilde{X}_I f=0, \quad \text{for all } I \Rightarrow \pi_{ij} = -\lambda(t)\sigma_{ij}, \tag{6.11}$$

which holds for all class A space-times, tilted or orthogonal (we exclude the degenerate case $\sigma=0$). The relation (6.11) between stress and shear, which emerges as a linear term in approximation theory of the Boltzmann equation,[6] holds exactly by virtue of the invariance of the distribution function in spatially homogeneous LRS space-times. (In their analysis of spatially homogeneous space-times with u orthogonal, MacCallum *et al.*[41] point out that the approximation theory leading to $\pi=-\lambda\sigma$ does not hold in the collision-free case, but that it may hold for small shear.)

Theorem 6.2: For a neutral one-component collision-free gas in (nondegenerate) class A space-time, if the distribution function is invariant under the G_4 of motions, then the energy density and isotropic pressure are spatially homogeneous, and the anisotropic stress is proportional to the shear.

Equation (6.10) also shows that the heat flow q is along the preferred direction. To determine the relation between q and $\dot{u}$, and the forms of the four dynamical functions in (6.10), we must use the field equations to obtain the form of T. By (6.1), (5.8), and (5.9),

$$T^{ab}=\operatorname{diag}(T^{00},T^{11},T^{22},T^{22})+2jT^{01}\delta^{(a}{}_0\,\delta^{b)}{}_1, \tag{6.12}$$

$$u^a=(\cosh\psi,\sinh\psi,0,0), \quad c^a=(\sinh\psi,\cosh\psi,0,0).$$

Then direction calculation using (6.12) gives

$$\begin{aligned}
\mu &= T^{00}\cosh^2\psi + T^{11}\sinh^2\psi - jT^{01}\sinh 2\psi,\\
p &= (T^{00}\sinh^2\psi + T^{11}\cosh^2\psi + 2T^{22} - jT^{01}\sinh 2\psi)/3,\\
Q &= -[(T^{00}+T^{11})\sinh 2\psi - 2jT^{01}\cosh 2\psi]/2,\\
\pi &= \sqrt{3}(p-T^{22}).
\end{aligned} \tag{6.13}$$

Thus the function λ in (6.11), which we may identify as the coefficient of viscosity of the gas, is given by

$$\lambda=\sqrt{3}(T^{22}-p)/\sigma.$$

Equations (6.2), (5.3) and the non-negativity of f imply

$$T^{aa}>0 \quad \text{(no sum)}. \tag{6.14}$$

By (6.13), (5.11), and (6.14), the heat flow has the following properties:

$$\begin{aligned}
&\psi\neq 0 \Rightarrow q^a = -A\,\dot{u}^a.\\
&\psi = 0 \Rightarrow q^a = (jT^{01})c^a,\\
&q^a=0 \Leftrightarrow \tanh 2\psi = 2jT^{01}/(T^{00}+T^{11}),
\end{aligned} \tag{6.15}$$

where (for $\psi\neq 0$)

$$A=(T^{00}+T^{11}-2jT^{01}\coth 2\psi)\times[\operatorname{sech}\psi(\log A\sinh\psi)']^{-1}. \tag{6.16}$$

Thus when u is tilted ($\psi\neq 0$), the heat flow obeys the transport law

$$q^i = -\eta h^{ij}(\tau_{,j}+\tau\dot{u}_j), \tag{6.17}$$

with

$$h^{ij}\tau_{,j}=[(A-\eta\tau)/\eta]\dot{u}^i.$$

Here, τ may be identified as the temperature of the gas, and η as the heat conduction coefficient.[2]

In classes A1 and A2, $q^a=0 \Leftrightarrow \psi=0$ by (6.15), and this is consistent with the constraint field equations[21]

$$q^a = -R^{bc}u_b\,h_c{}^a, \tag{6.18}$$

since R^{bc} is diagonal [by (3.3)] and $u_b=-\delta_b{}^0$. In class A3, $\psi=0$ does not force $q^a=0$ unless $T^{01}=0$. Again, this is consistent with (6.18) since R^{bc} is *not* diagonal [(3.3)]. In fact, using the explicit form $h^{ab}(\frac{2}{3}\theta_{;b}-\sigma_{bc;d}h^{cd})-\eta^{abcd}\times u_b(\omega_{c;d}+2\omega_c\dot{u}_d)$ for the rhs of (6.18),[21] and the forms (5.10) and (5.11) for the kinematic quantities, we find that for the orthogonal self-gravitating gas in class A3

$$\text{A3 } (\psi=0): \quad T^{01}=2(B'/B-A'/A)/A \tag{6.19}$$

[compare Eq. (3.16) in Ref. 32] and thus by (5.11)

$$\text{A3 } (\psi=0): \quad T^{01}=0 \Leftrightarrow \sigma_{ab}=0. \tag{6.20}$$

Now $\dot{u}_a=0=\omega_a$ in the orthogonal case, so that $\sigma_{ab}=0$ reduces the class A3 geometry to its Robertson–Walker subclass. If we consider only nondegenerate class A geometry, this means that T^{01} and hence q^a never vanish in the orthogonal class A3 case.

Theorem 6.3: Under the conditions of Theorem 6.2, if the distribution function also satisfies the Einstein field equations, then the heat flow is proportional to the acceleration if u is tilted; if u is orthogonal, the heat flow is zero in classes A1 and A2, but in class A3 it is nonzero and along the preferred direction [(6.15)].

B. Evolution of tilt

We can use the conservation equations to investigate the evolution of tilt. Using the form (2.10) for T^{ab}, the conservation equations

$$u_a T^{ab}{}_{;b} = 0, \quad h_{ab} T^{bc}{}_{;c} = 0 \tag{6.21}$$

become[21]

$$\dot{\mu} + (\mu + p)\theta + \pi_{ab}\sigma^{ab} + q^a{}_{;a} + q^a \dot{u}_a = 0, \tag{6.22}$$

$$(\mu + p)\dot{u}_a + h_a{}^c(p_{;c} + \pi_c{}^b{}_{;b} + \dot{q}_c) + q^b(\omega_{ab} + \sigma_{ab} + \tfrac{4}{3}\theta h_{ab}) = 0. \tag{6.23}$$

King and Ellis[32] show that a self-gravitating perfect fluid in spatially homogeneous space-time is either always tilted or always orthogonal. This also holds for the collision-free gas in class A1 and A2 space-times.

Theorem 6.4: For a gas under the conditions of Theorem 6.3 in class A1 and A2 space-times, the kinematic average four-velocity u is locally either always orthogonal or always tilted.

To prove this, suppose that $\psi = 0$ at some time $t = t_0$. Then using (5.8)–(5.11), (6.10) and (6.13) in (6.23), we obtain

$$\psi'(T^{00} + T^{22}) - j[T^{01}(\log A^2 B^2 T^{01})' + 2(T^{11} - T^{22})/A] = 0, \tag{6.24}$$

where all quantities are evaluated at $t = t_0$. When $j = 0$, (6.14) and (6.24)$\Rightarrow\psi'(t_0) = 0$. Since t_0 is arbitrary, Theorem 6.4 is proved.[42] In class A3 ($j = 1$), (6.24) does not give a clear answer. For $\psi(t_0) = 0$, the constraint field equations (6.18) retain the form (6.19) for $t = t_0$, while the conservation equation (6.22) gives

$$T^{00}(\log AB^2 T^{00})' + T^{11}(\log A)' + 2T^{22}(\log B)' + jT^{01}(\psi' + 2/A) = 0, \tag{6.25}$$

for $t = t_0$. The Raychaudhuri field equation[21]

$$\dot{\theta} - \dot{u}^a{}_{;a} + \tfrac{1}{3}\theta^2 + 2(\sigma^2 - \omega^a\omega_a) + (\mu + 3p)/2 = 0,$$

evaluated at $t = t_0$ [$\psi(t_0) = 0$] gives

$$(AB^2)''/(AB^2) + 2(\log A/B)'^2 + \psi'^2 + 2j\psi' + T^{00} + T^{11} + T^{22} = 0. \tag{6.26}$$

Equations (6.19) and (6.24)–(6.26) do not seem to force or to rule out $\psi'(t_0) = 0$. The remaining six field equations

$$R^{cd} h_{ac} h_{bd} = \pi_{ab} + (\mu - p) h_{ab}/2,$$

evaluated at $t = t_0$ [$\psi(t_0) = 0$], also appear to leave open the nature of $\psi'(t_0)$. Hence we conclude that Theorem 6.4 may not hold in class A3.

C. Perfect fluid

The question of whether the average behavior of the self-gravitating gas can become that of a perfect fluid was considered by Ray.[9] Ray argued that a gas under the conditions of Theorem 6.3 in class A1c space-time, with u orthogonal, could *not* exhibit perfect fluid behavior, i.e., that conditions (2.11) could not be satisfied. This result backs up the conjecture of Ellis *et al.*[4] that a collision-free gas can behave like a perfect fluid only if the shear vanishes. While Ray's result is plausible to us, it is not clear to us whether his proof, which is based on an extension of the argument applied in static spherically symmetric space-time,[8] can be applied to all class A space-times. By a different approach, we can prove the following partial result.

Theorem 6.5: A gas under the conditions of Theorem 6.3 cannot exhibit perfect fluid behavior in classes A1 and A2 with tilted u, or in class A3 with orthogonal u.

For the tilted class A1 and A2 case, the result follows immediately from (6.13): $\psi\neq 0\Rightarrow q^a\neq 0$, which violates the perfect fluid conditions (2.11). This result is in fact just the corollary of the result due to King and Ellis.[32] In the orthogonal class A3 case, the result follows from (6.15) and (6.20): $\psi = 0\Rightarrow T^{01}\neq 0$ (unless $\sigma_{ab} = 0$, which implies Robertson–Walker geometry)$\Rightarrow q^a\neq 0$. In the remaining cases, the problem reduces to the consideration of certain multiple integral identities. For the orthogonal class A1 and A2 case, $q^a = 0$ by (6.13), so that the perfect fluid condition reduces to $\pi_{ab} = 0$, which by (6.10), (6.13) (with $\psi = 0$) and (6.2), (5.6), becomes

$$I(t) \equiv \int_{\mathbb{R}^3} dp^{123}(p^{1^2} - p^{2^2})(m^2 + p^{1^2} + p^{2^2} + p^{3^2})^{-1/2} \times F[A(t)p^1, B(t)^2(p^{2^2} + p^{3^2})] = 0. \tag{6.27}$$

Perfect fluid behavior will be possible if there exists a physically reasonable F such that (6.27) holds for t in an open interval. Since nonzero values of A and B may be arbitrarily assigned on an initial hypersurface $t = t_0$, it should be possible to get $I(t_0) = 0$ for a reasonable choice of F. However, it seems unlikely that $I'(t_0) = 0$, without restricting $A'(t_0)$, $B'(t_0)$. If the conjecture of Ellis *et al.*[4] is correct, $I'(t_0) = 0$ will require $\sigma_{ab}(t_0) = 0$. In the tilted class A3 case,

$$\begin{aligned} q^a = 0 &\Leftrightarrow (T^{00} + T^{11})\sinh 2\psi - 2T^{01}\cosh 2\psi = 0, \\ \pi_{ab} = 0 &\Leftrightarrow T^{00}\sinh^2\psi + T^{11}\cosh^2\psi - T^{22} - T^{01}\sinh 2\psi = 0 \end{aligned} \tag{6.28}$$

follow from (6.13). Equation (6.28) can be written as a pair of integral identities of the form (6.27), but with the appearance of the additional function $\psi(t)$. By similar arguments to the above, it seems unlikely that these identities can be satisfied by reasonable F of form (5.7) without severe restrictions on A, B, and ψ. [Certainly $\psi = 0$, $(\log A/B)' = 0$ will allow the identities to hold, since these conditions correspond to a Robertson–Walker geometry with F isotropic in momentum space.]

In conclusion, it is likely that Theorem 6.5 will hold for all class A cases, although we have been unable to provide a definite answer on the basis of (6.27) and (6.28).

VII. CONCLUSION

We have achieved the aim of obtaining the invariant distribution functions for all collision-free gases in spatially homogeneous LRS space-times, showing how they are spatially homogeneous in space-time and LRS in momentum space (Theorems 4.1–4.3). We showed that these gases can have average velocities tilted relative to the homogeneous hypersurfaces (Theorem 5.1). Both results were shown to hold also for self-gravitating gases (Theorem 6.1). We obtained the kinematic quantities associated with the average behavior of the gas, showing in particular that only class A2 could have nonzero vorticity (when the tilt is nonzero) (Theorem 5.1). We obtained the dynamical quantities of the self-gravitating gas, and showed that the stress is always proportional to the shear, while the heat flow is proportional to the acceleration (nonzero tilt) (Theorems 6.2 and 6.3). We showed that locally tilt cannot be "created or destroyed" in classes A1 and A2 (Theorem 6.4). Finally, we discussed perfect fluid behavior, obtaining a partial result (Theorem 6.5) and a plausibility argument for the impossibility of such behavior.

Although we were able to obtain some results by use of the conservation and certain field equations, our feeling is that a more detailed analysis using tetrad methods[3,23,32] (and therefore reducing the field equations to first order and supplementing them with the Jacobi identities) would be more successful in clarifying the relationships amongst the kinetic and geometric variables. A suitable tetrad $\{E_a\}$ for both tilted and orthogonal cases would be one in which $E_0 = u$, $E_1 = c$ [(5.8) and (5.9)].

What emerges throughout this paper is the power of the assumption that the distribution function f be invariant under the space-time group of motions [(2.15)]. We showed that the further assumption that f be based on Killing vector constants of the motion is redundant in class A1 and A2 space-times, and too restrictive in class A3. We were then able to derive kinematic and dynamical properties of the gas which rest, finally, on the G_4 invariance of f. Now invariance under the full group of motions is the strongest assumption satisfying the condition (2.17), i.e., $\int (\widetilde{X}_I f) p^i p^j \pi_m = 0$ for all I, which is imposed on f by the field equations for the self-gravitating gas. The next step is to consider ways of weakening the assumption of G_4 invariance. If we cannot find the most general restriction on f that satisfies (2.17), we could try assuming invariance of f under a subgroup of the G_4 of motions. (This is essentially the approach followed by Ellis *et al.*[13] in Robertson–Walker space-time.)

In the case where the gas particles each carry a charge e, Ray[9] has extended the Ray–Zimmerman analysis[12] of class A1c. Ray constructs modified Killing vector constants of motion by assuming G_4 invariance of the electromagnetic vector potential A, and then shows that class A1c space-time cannot admit charged particles. Ray's result may be extended[34]; under Ray's assumptions (G_4 invariance of A and f), charged particles are possible only in class A3 space-time. This follows directly from the straightforward proof that a G_4-invariant A [which must be of the form (3.5)] leads in classes A1 and A2 to a space-like four-current $J^i = F^{ij}{}_{;j}$ ($F = -2dA$), which violates the relation[1] $J^i = en^i$. In class A3, J^i is timelike, so that e may be nonzero.

Finally, the methods of this paper can be applied to the more general non-LRS spatially homogeneous Bianchi space-times. For G_3-invariant f based on Killing vector constants of motion, partial results have been obtained by Misner[14] (type I), Matzner[15] (type V), and MacCallum and Ellis[43] (group types with $a^\alpha = 0$ or $n^\alpha{}_\alpha = 0$). These results may be unified and extended along the lines of this paper,[44] although the problem is clearly more difficult.

ACKNOWLEDGMENT

One of us (SDM) thanks the Council of Scientific and Industrial Research (CSIR) for the support of a scholarship during research work on this paper.

APPENDIX: CONNECTION COEFFICIENTS

The components $\Gamma^\alpha{}_{ij} p^i p^j \equiv \Gamma^\alpha$ in class A space-times are

$$\Gamma^\alpha = 2p^t(p^x A'/A, p^y B'/B, p^z B'/B) + \gamma^\alpha,$$

where

$$\gamma^\alpha = khp^z(0, h' p^z, -2p^y/h') \quad \text{in class A1},$$

$$\begin{aligned}\gamma^\alpha = (&2h(A'/A - B'/B) p^t p^z - (h/h')(A/B)^2 p^x p^y \\ &- h'[(A^2/B^2 - k)(h/h')^2 - k/h'^2 \\ &- (1-k^2)] p^y p^z, -h(A/B)^2 p^x p^z \\ &- h'h(A^2/B^2 - k) p^{z^2}, (1/h')(A/B)^2 p^x p^y \\ &+ (h/h')(A^2/B^2 - 2k) p^y p^z) \quad \text{in class A2},\end{aligned}$$

$$\begin{aligned}\gamma^\alpha = (&-(B/A)^2 \exp(2x)(p^{y^2} + p^{z^2}), \\ &2p^x p^y, 2p^x p^z) \quad \text{in class A3}.\end{aligned}$$

From Γ^α one can read off the connection coefficients $\Gamma^\alpha{}_{ij}$. The remaining connection coefficients are

$$\begin{aligned}\Gamma^t{}_{ij} = \operatorname{diag}(&0, AA', \exp(2jx)BB', \\ &\exp(2jx)h'^2 BB' + lh^2 BB') + l(2hBB')\delta_{(i}{}^x \delta_{j)}{}^z.\end{aligned}$$

[1] J. Ehlers, in *Proceedings of the International Summer School of Physics "Enrico Fermi,"* Course 47, edited by R. K. Sachs (Academic, New York, 1971).
[2] J. Ehlers, in *Relativity, Astrophysics and Cosmology*, edited by W. Israel (Reidel, Dordrecht, 1973).
[3] G. F. R. Ellis, D. R. Matravers, and R. Treciokas, Ann. Phys. **150**, 455 (1983).
[4] G. F. R. Ellis, R. Treciokas, and D. R. Matravers, Ann. Phys. **150**, 487 (1983).
[5] R. W. Lindquist, Ann. Phys. **37**, 487 (1966).
[6] W. Israel and J. M. Stewart, in *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980), Vol. 2; J. M. Stewart, *Nonequilibrium Relativistic Kinetic Theory* (Springer, Berlin, 1971).
[7] R. K. Sachs and J. Ehlers, in *Astrophysics and General Relativity, Brandeis Summer Institute in Theoretical Physics*, Vol. 2, edited by M. Chretien, S. Deser, and J. Goldstein (Gordon and Breach, London, 1971).
[8] E. D. Fackerell, Astrophys. J. **153**, 643 (1968).
[9] J. R. Ray, Astrophys. J. **257**, 578 (1982).
[10] R. Hakim, Phys. Rev. **173**, 1235 (1968).
[11] L. Bel, Astrophys. J. **155**, 83 (1969).
[12] J. R. Ray and J. C. Zimmerman, Nuovo Cimento B **42**, 183 (1977).
[13] G. F. R. Ellis, D. R. Matravers, and R. Treciokas, Gen. Relativ. Gravit. **15**, 931 (1983).
[14] C. W. Misner, Astrophys. J. **151**, 431 (1968).
[15] R. A. Matzner, Astrophys. J. **157**, 1085 (1969).
[16] D. R. Matravers, preprint, Applied Mathematics Department, University of Cape Town, South Africa, 1985.
[17] J. Ehlers, P. Geren, and R. K. Sachs, J. Math. Phys. **9**, 1344 (1968).
[18] R. Treciokas and G. F. R. Ellis, Commun. Math. Phys. **23**, 1 (1971).
[19] R. Berezdivin and R. K. Sachs, J. Math. Phys. **14**, 1254 (1973).

[20]D. R. Oliver and W. R. Davis, Ann. Inst. H. Poincaré A **555**, 339 (1979).
[21]G. F. R. Ellis, in *Càrgese Lectures in Physics*, edited by E. Schatzmann (Gordon and Breach, New York, 1973), Vol. 6.
[22]J. M. Stewart and G. F. R. Ellis, J. Math. Phys. **9**, 1072 (1968).
[23]G. F. R. Ellis and M. A. H. MacCallum, Commun. Math. Phys. **12**, 108 (1969).
[24]M. A. H. MacCallum, in *General Relativity: An Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge U. P., London, 1979).
[25]J. Wainwright and P. E. A. Yaremovicz, Gen. Relativ. Gravit. **7**, 345 (1976). For a *source-free* field, the field tensor is determined by the metric up to a duality transformation. See M. Henneaux, J. Math. Phys. **25**, 2276 (1984).
[26]K. Yano and S. Ishihara, *Tangent and Cotangent Bundles* (Dekker, New York, 1973).
[27]R. Berezdivin, Ph.D. thesis, University of California, Berkeley, California, 1972.
[28]M. A. H. MacCallum, in *Essays in General Relativity: a Festschrift for Abraham Taub*, edited by F. J. Tipler (Academic, New York, 1980). An essentially equivalent classification is given in Refs. 22 and 23.
[29]A. Z. Petrov, *Einstein Spaces* (Pergamon, London, 1969).
[30]R. Kantowski and R. K. Sachs, J. Math. Phys. **7**, 443 (1966).
[31]D. Kramer, H. Stephani, M. A. H. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (Cambridge U. P., London, 1980).
[32]A. R. King and G. F. R. Ellis, Commun. Math. Phys. **31**, 209 (1973).
[33]I. N. Sneddon, *Elements of Partial Differential Equations* (McGraw–Hill, New York, 1957).
[34]S. D. Maharaj, M.Sc. dissertation, Applied Mathematics Department, University of the Witwatersrand, South Africa, 1984.
[35]It is clear that any function defined on the phase space and invariant under the complete lift of the class A G_4, must be spatially homogeneous over space-time and LRS in momentum space, i.e., must be of the form $F[t,H(t,p^1),K(t,p^{2^2}+p^{3^2})]$. However, the invariance condition (2.15) must be solved in order to determine the functions H and K.
[36]Excluding explicit dependence on the path parameter. See G. H. Katzin and J. Levine, J. Math. Phys. **22**, 1878 (1981).
[37]P. Sommers, J. Math. Phys. **14**, 787 (1973); G. Caviglia, *ibid.* **24**, 2065 (1983).
[38]For a self-gravitating perfect fluid, u is invariant under a Killing vector X by virtue of the field equations $\mathscr{L}g=0 \Rightarrow \mathscr{L}G=0 \Rightarrow \mathscr{L}T=0 \Rightarrow \mathscr{L}(\mu u_i u_j + p h_{ij}) = 0 \Rightarrow \mathscr{L}u_i = 0$, $\mathscr{L}\mu = 0 = p$, on using $u^i \mathscr{L} u_i = 0$ ($\mathscr{L} \equiv \mathscr{L}_X$). For a test perfect fluid, invariance of u requires special geometric conditions to be imposed, e.g., $X^i u_i = 0 = \omega_{ij}$.
[39]P. R. Ryan and L. C. Shepley, *Homogeneous Relativistic Cosmologies* (Princeton U. P., Princeton, NJ, 1975).
[40]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York, 1980).
[41]M. A. H. MacCallum, J. M. Stewart, and B. G. Schmidt, Commun. Math. Phys. **17**, 343 (1970).
[42]Strictly, we have to assume analyticity of ψ, and use the fact that $(6.24) \Rightarrow d^n \psi(t_0)/dt^n = 0$ for all n.
[43]M. A. H. MacCallum and G. F. R. Ellis, Commun. Math. Phys. **19**, 31 (1970).
[44]R. Maartens and S. D. Maharaj (in preparation).

Spacelike conformal Killing vectors and spacelike congruences

D. P. Mason
Department of Applied Mathematics, University of the Witwatersrand, 1 Jan Smuts Avenue, Johannesburg 2001, South Africa

M. Tsamparlis
Department of Physics, Division of Mechanics, University of Athens, Athens GR 15771 Zografou, Greece

(Received 5 March 1984; accepted for publication 17 May 1985)

Necessary and sufficient conditions are derived for space-time to admit a spacelike conformal motion with symmetry vector parallel to a unit spacelike vector field n^a. These conditions are expressed in terms of the shear and expansion of the spacelike congruence generated by n^a and in terms of the four-velocity of the observer employed at any given point of the congruence. It is shown that either the expansion or the rotation of this spacelike congruence must vanish if $Dn^a/dp = 0$, where p denotes arc length measured along the integral curves of n^a, and also that there exist no proper spacelike homothetic motions with constant expansion. Propagation equations for the projection tensor and the rotation tensor are derived and it is proved that every isometric spacelike congruence is rigid. Fluid space-times are studied in detail. A relation is established between spacelike conformal motions and material curves in the fluid: if a fluid space-time admits a spacelike conformal Killing vector parallel to n^a and $n_a u^a = 0$, where u^a is the fluid four-velocity, then the integral curves of n^a are material curves in an irrotational fluid, while if the fluid vorticity is nonzero, then the integral curves of n^a are material curves if and only if they are vortex lines. An alternative derivation, based on the theory of spacelike congruences, of some of the results of Collins [J. Math. Phys. **25**, 995 (1984)] on conformal Killing vectors parallel to the local vorticity vector in shear-free perfect fluids with zero magnetic Weyl tensor is given. The necessary and sufficient condition for vortex lines to be material lines is derived and the restriction this places on the flow of a thermodynamical perfect fluid is determined. As an application, a pure magnetic field in a rotational fluid is considered and results similar in nature to Ferraro's law of isorotation are obtained. Throughout, corresponding results for a timelike conformal motion and for Newtonian theory are given for comparison.

I. INTRODUCTION

The properties of a *timelike* conformal motion with symmetry vector parallel to a timelike unit vector field v^a can usefully be expressed in terms of the kinematical quantities of the timelike congruence of curves generated by v^a. For instance, the necessary and sufficient conditions for space-time to admit a timelike conformal Killing vector may be expressed in terms of the shear, expansion, and acceleration of this congruence.[1–3] In this paper we adopt a similar approach to the study of *spacelike* conformal Killing vectors admitted by space-time: we relate the properties of a spacelike conformal motion with symmetry vector parallel to a spacelike unit vector field n^a to quantities such as the shear, expansion, and rotation of the spacelike congruence of curves generated by n^a. We find that the four-velocity w^a of the observer employed to determine the deformation of this congruence at any given point plays an important role.

The theory of spacelike congruences in general relativity was first formulated by Greenberg,[4] who also considered applications to the vortex congruence in a rotational fluid. The theory has been developed and further applications have been considered by several authors.[5–15] In Sec. II the basic aspects of the theory of spacelike congruences required in this paper are reviewed briefly. If the four-velocity w^a of an observer is specified at any one point then the four-velocity of the observer employed at any other point along a spacelike congruence is not arbitrary but is determined by a transport law for w^a along the congruence, derived by Greenberg: for instance, in a fluid space-time, if an observer comoving with the fluid can be chosen at any one point, then a comoving observer can be employed at any other point along a spacelike congruence in the fluid if and only if the curves of the congruence are material curves in the fluid.[15,16]

In Sec. III, necessary and sufficient conditions are derived for space-time to admit a spacelike conformal motion with symmetry vector parallel to the unit tangent vector field n^a; these conditions are expressed in terms of the expansion $\mathscr{E}$ and the shear $\mathscr{S}_{ab}$ of the spacelike congruence generated by n^a and in terms of the four-velocity w^a of the observer. Two forms of the basic theorem are given, the first of which is suitable for specialization later to a fluid space-time and in which four necessary and sufficient tensor conditions are derived. The first two of those conditions correspond directly with the two necessary and sufficient tensor conditions for space-time to admit a timelike conformal motion and also with the two necessary and sufficient tensor conditions for a conformal motion in three-dimensional Euclidean space. Vanishing shear is one of these conditions in each case. The remaining two tensor conditions therefore apply specifically to *spacelike* conformal Killing vectors and are purely relativisitic; they can be combined, with the aid of the Greenberg transport law for w^a, into a single condition on the Lie derivative of w^a with respect to n^a, and this leads to the alternative form of the basic theorem. As applications of the theorems, it is proved that there exist no proper spacelike

 0022-2488/85/112881-21$02.50

homothetic motions with constant expansion $\mathscr{E}$ and that if space-time admits a conformal Killing vector parallel to n^a and $Dn^a/dp = 0$ (p denotes arc length measured along the integral curve of n^a), then either the expansion $\mathscr{E}$ or the rotation $\mathscr{R}$ of the spacelike congruence generated by n^a must vanish. Both these results have direct analogs in the theory of timelike conformal Killing vectors[3,17] and in Newtonian theory.

The propagation equation for the projection tensor p_{ab} (defined in Sec. II) and for the rotation tensor $\mathscr{R}_{ab}$ are considered in Sec. IV. A spacelike congruence can be defined as rigid by direct analogy with the condition for a rigid timelike congruence,[6] and we show with the aid of the propagation equation for p_{ab} that every isometric spacelike congruence generated by a spacelike Killing vector is rigid. It is well established that every isometric timelike congruence is rigid.[18] The equation governing the propagation of $\mathscr{R}_{ab}$ along a spacelike congruence generated by a conformal Killing vector was first derived by Prasad and Sinha.[14] It shows that a close relationship exists between the rotation and the expansion of this congruence, which can be exploited by making use of the concept of a flux tube.

The remainder of the paper is concerned with fluid space-times and with spacelike conformal motions parallel to unit vector fields n^a, which satisfy $n_a u^a = 0$, where u^a is the unit four-velocity of the fluid. This latter condition is satisfied, for instance, by the unit vector parallel to the local vorticity vector of the fluid, ω^a, and to the electric and magnetic field four-vectors in an electrically conducting fluid. In Sec. V, the necessary and sufficient conditions derived in Sec III for space-time to admit a spacelike conformal motion are rewritten for a fluid space-time. Prasad and Sinha[14] established a similar theorem but they did not observe that one of the conditions can be expressed in terms of the vorticity vector, which is important in applications. Our results may be used to derive general properties of spacelike conformal motions. For instance, we establish a connection between spacelike conformal motions and material curves in the fluid.[16] That such a relationship exists for a *timelike* conformal motion is well known: if a rotational fluid space-time admits a timelike conformal motion parallel to u^a, then an acceleration potential will exist and therefore the vortex lines will be material lines in the fluid. We prove that if space-time admits a *spacelike* conformal motion parallel to n^a ($n_a u^a = 0$) and if the fluid is irrotational then the integral curves of n^a must be material curves in the fluid, while if the vorticity of the fluid is nonzero then the integral curves of n^a are material curves if and only if they are vortex lines. This result is purely relativistic and demonstrates the important role played by vorticity in spacelike symmetries of fluid space-times. Although spacelike symmetries with symmetry vector parallel to ω^a have been studied and generalizations of the first Helmholtz theorem derived,[19–21] this close relationship between *spacelike* symmetries, vorticity, and material curves appears to be a new result. We illustrate it, for the special case of Killing vectors, by considering the Friedman–Robertson–Walker models ($\omega = 0$) and the Gödel model ($\omega \neq 0$). Our results may also be used to establish the existence of spacelike conformal motions, and we demonstrate this by giving an alternative derivation of some of the results established recently by Collins[22] on conformal Killing vectors parallel to the local vorticity vector in shear-free perfect fluids with zero magnetic Weyl tensor. Our derivation is based on the theory of spacelike congruences and makes use of general expressions for the expansion $\mathscr{E}(\omega)$, the shear $\mathscr{S}_{ab}(\omega)$, and the rotation $\mathscr{R}_{ab}(\omega)$ of a vortex congruence,[15] as well as the propagation equation for the expansion along the congruence. Both $\mathscr{E}(\omega)$ and $\mathscr{S}_{ab}(\omega)$ depend explicitly on the shear of the fluid σ_{ab} and on the magnetic part of the Weyl tensor H_{ab}, which explains the important role played by σ_{ab} and H_{ab} in determining if a fluid space-time admits a conformal Killing vector parallel to ω^a.

In Sec. VI we investigate further properties of material vortex lines. Starting from the vorticity propagation equation,[23,24] the necessary and sufficient condition for vortex lines to be material lines in the fluid is derived and the restriction this places on the flow of a thermodynamical perfect fluid is examined: except for two special cases, vortex lines will be material lines if and only if they lie along the intersection of the surfaces $S =$ const and $T/f =$ const, where S is the specific entropy, T is the temperature, and f is the index of the fluid.[25] Also, in order to illustrate further the results of Sec. V, we consider a pure magnetic field in a rotational fluid; if the local electric field vanishes identically then it follows directly from Maxwell's equations that the magnetic field lines are "frozen-in" to the fluid.[24] Hence if space-time admits a spacelike conformal Killing vector field everywehre tangent to the magnetic field lines, then the vortex lines must coincide with the magnetic field lines. This flow is investigated and quantities conserved along a magnetic field/vortex line are derived. Results similar in nature to Ferraro's law of isorotation[26] in nonrelativistic magnetohydrodynamics are obtained.

Finally concluding remarks are made in Sec. VII. Throughout the paper corresponding results for a timelike conformal motion and also results from Newtonian theory are given for comparison. Properties which apply specifically to spacelike conformal motions, as well as purely relativistic effects, can therefore be isolated. The notation and conventions of Ellis[23,24] are followed throughout.[27]

II. SPACELIKE CONGRUENCES

We review briefly the basic aspects of the theory of spacelike congruences that will be required in this paper.[4,15].

Consider a spacelike congruence, i.e., a family of nonintersecting spacelike curves,

$$x^a = x^a(\eta^\alpha, p)\,, \tag{2.1}$$

where the three parameters η^α ($\alpha = 1,2,3$) specify the particular spacelike curve of the congruence and where p is some parameter along the curve, which we take to be the arc length measured from some arbitrary section of the congruence. The unit tangent vector n^a at a point P on a curve $\mathscr{C}$ of the congruence is defined by

$$n^a = \left(\frac{\partial x^a}{\partial p}\right)_{\eta^\alpha}. \tag{2.2}$$

We have $n_a n^a = +1$ and since n^a forms a vector field,

$n^a n_{a;b} = 0$. We now introduce an observer at P with unit four-velocity w^a such that w^a is orthogonal to $\mathscr{C}$ at P; thus

$$w_a w^a = -1\,, \quad w_a n^a = 0\,. \tag{2.3}$$

To observe the deformation of the curves of the congruence at P, the observer w^a erects a "screen" orthogonal to $\mathscr{C}$ at P so that the curves pass perpendicularly through the screen at P. The screen is the two-surface dual to the surface formed by w_a and n_a. We introduce the projection tensor

$$p_{ab} = g_{ab} + w_a w_b - n_a n_b\,; \tag{2.4}$$

clearly

$$p_{ab} w^b = 0\,, \quad p_{ab} n^b = 0\,. \tag{2.5}$$

The tensor p_{ab} projects vectors onto the screen at P.

The connecting vector δx^a connects points with the same value of the parameter p on neighboring curves of the congruence:

$$\delta x^a = \left(\frac{\partial x^a}{\partial \eta^\alpha}\right)_p \delta\eta^\alpha\,. \tag{2.6}$$

Since the δx^a connect points of equal p, it is easily verified that

$$\underset{n}{\mathscr{L}}\,\delta x^a = 0\,, \tag{2.7}$$

where $\underset{n}{\mathscr{L}}$ stands for the Lie derivative with respect to the vector field n^a. Now, δx^a need not lie on the screen at P, and we therefore construct at P the orthogonal connecting vector

$${}_\perp\delta x^a = p^a_b\,\delta x^b\,, \tag{2.8}$$

which lies on the screen at P; we have $w_{a\perp}\delta x^a = 0$ and $n_{a\perp}\delta x^a = 0$. With the aid of (2.7) we obtain

$$p_{ab}({}_\perp\delta x^b)^* = (p^c_a\, p^d_b\, n_{c;d})_\perp\delta x^b + p_{ac}(\overset{*}{w}{}^c - \mathring{n}^c) w_b \delta x^b\,, \tag{2.9}$$

where we have introduced the notation

$$\overset{*}{A}{}^a = \frac{DA^a}{dp} = A^a{}_{;b} n^b\,, \tag{2.10}$$

$$\mathring{A}{}^a = A^a{}_{;b} w^b\,. \tag{2.11}$$

Now, except at P, the motions of the observers employed along a curve $\mathscr{C}$ of the congruence have still to be specified. We choose observers such that

$$p_{ac}(\overset{*}{w}{}^c - \mathring{n}^c) = 0\,, \tag{2.12}$$

$$(w_a w^a)^* = 0\,, \quad (w_a n^a)^* = 0\,. \tag{2.13}$$

By (2.3) and (2.13), w^a will be a unit vector orthogonal to n^a along $\mathscr{C}$, and by (2.12), Eq. (2.9) reduces to

$$p_{ab}({}_\perp\delta x^b)^* = (p^c_a\, p^d_b\, n_{c;d})_\perp\delta x^b\,; \tag{2.14}$$

the rate of change of separation of two spacelike curves, as seen by observers w^a in their screens, is given by a linear transformation from the orthogonal connecting vector ${}_\perp\delta x^a$, the transformation being determined by the tensor $p^c_a\, p^d_b\, n_{c;d}$. On splitting $p^c_a\, p^d_b\, n_{c;d}$ into its irreducible parts, we have

$$p^c_a\, p^d_b\, n_{c;d} = \mathscr{R}_{ab} + \tfrac{1}{2}\mathscr{E} p_{ab} + \mathscr{S}_{ab}\,, \tag{2.15}$$

where the rotation tensor $\mathscr{R}_{ab}$, the expansion $\mathscr{E}$, and the shear tensor $\mathscr{S}_{ab}$ of the curves of the congruence are defined by

$$\mathscr{R}_{ab} = p^c_a\, p^d_b\, n_{[c;d]}\,, \tag{2.16}$$

$$\mathscr{E} = p^{ab} n_{a;b}\,, \tag{2.17}$$

$$\mathscr{S}_{ab} = p^c_a\, p^d_b\, n_{(c;d)} - \tfrac{1}{2}\mathscr{E} p_{ab}\,. \tag{2.18}$$

Clearly by (2.5),

$$\begin{aligned} &\mathscr{R}_{ab} w^b = 0\,, \quad \mathscr{R}_{ab} n^b = 0\,; \\ &\mathscr{S}_{ab} w^b = 0\,, \quad \mathscr{S}_{ab} n^b = 0\,, \quad \mathscr{S}^a_a = 0\,. \end{aligned} \tag{2.19}$$

The rotation vector is defined by[28]

$$\mathscr{R}^a = \tfrac{1}{2}\eta^{abcd} w_b \mathscr{R}_{cd}\,. \tag{2.20}$$

Since $\mathscr{R}_{cd}$ lies on the screen, it is easily verified that $\mathscr{R}^a$ is parallel to n^a. Equation (2.20) may be inverted to give

$$\mathscr{R}_{ab} = \eta_{abcd}\mathscr{R}^c w^d\,, \tag{2.21}$$

and the magnitude of the rotation $\mathscr{R}$ is defined by

$$\mathscr{R}^2 = \mathscr{R}_a \mathscr{R}^a = \tfrac{1}{2}\mathscr{R}_{ab}\mathscr{R}^{ab}\,. \tag{2.22}$$

From (2.15) the rotation, expansion, and shear of the congruence determine the covariant derivative of n_a through the identity

$$\begin{aligned} n_{a;b} = {} & \mathscr{R}_{ab} + \tfrac{1}{2}\mathscr{E} p_{ab} + \mathscr{S}_{ab} + \overset{*}{n}_a n_b - \mathring{n}_a w_b - w_a(w^t n_{t;b}) \\ & - w_a w_b (w_t \mathring{n}^t) + w_a n_b (w_t \overset{*}{n}{}^t)\,. \end{aligned} \tag{2.23}$$

Conditions (2.12) and (2.13) are equivalent to the single condition

$$\overset{*}{w}{}^a = \mathring{n}^a + (w_b \mathring{n}^b) w^a - (w_b \overset{*}{n}{}^b) n^a\,, \tag{2.24}$$

which is the transport law for w^a derived by Greenberg.[4] Equation (2.24) may be rewritten equivalently as

$$\hat{h}^a_b \mathring{n}^b = \overset{*}{w}{}^a + (w_b \overset{*}{n}{}^b) n^a\,, \tag{2.25}$$

where

$$\hat{h}^a_b = g^a_b + w^a w_b\,. \tag{2.26}$$

It is shown in Appendix A that the unit four-velocity field u^a of a self-gravitating fluid with $n_a u^a = 0$ satisfies Eq. (2.25) if and only if n^a is tangential to a material curve in the fluid.[16] It follows directly that in a fluid space-time, comoving observers $w^a = u^a$ can be employed all along a spacelike congruence generated by n^a with $n_a u^a = 0$ if and only if the curves of the congruence are material curves in the fluid. [If $n_a u^a = 0$, a comoving observer can always be employed at any one point since (2.3) is satisfied by $w^a = u^a$.]

In Newtonian theory, we denote by

$$x^\alpha = x^\alpha(\eta^\mu, p) \quad (\mu = 1,2) \tag{2.27}$$

a congruence of curves in three-dimensional Euclidean space, where p is some parameter along the curves which we take to be the arc length measured from some arbitrary section of the congruence. The projection tensor $p^{\alpha\beta}$ is defined by

$$p^{\alpha\beta} = h^{\alpha\beta} - n^\alpha n^\beta, \tag{2.28}$$

where $h^{\alpha\beta}$ is the flat three-space metric tensor and the unit tangent vector n^α is defined as in (2.2). The connecting vector δx^α and the orthogonal connecting vector ${}_\perp\delta x^\alpha$ are defined as in general relativity. Since δx^α satisfies an equation of the form (2.7), it is easily verified that

$$p_{\alpha\beta}({}_\perp\delta x^\beta)^* = (p^\lambda_\alpha\, p^\tau_\beta n_{\lambda,\tau})_\perp \delta x^\beta, \tag{2.29}$$

which compares with (2.9); a significant difference between Newtonian theory and general relativity is that in Newtonian theory it is not necessary to employ special observers along the congruence. On decomposing $p^\lambda_\alpha\, p^\tau_\beta n_{\lambda,\tau}$ into its irreducible parts, we have

$$p^\lambda_\alpha\, p^\tau_\beta n_{\lambda,\tau} = \mathscr{R}_{\alpha\beta} + \tfrac{1}{2}\mathscr{E} p_{\alpha\beta} + \mathscr{S}_{\alpha\beta}, \tag{2.30}$$

where

$$\mathscr{R}_{\alpha\beta} = p^\lambda_\alpha\, p^\tau_\beta n_{[\lambda,\tau]}, \tag{2.31}$$

$$\mathscr{E} = p^{\lambda\tau} n_{\lambda,\tau}, \tag{2.32}$$

$$\mathscr{S}_{\alpha\beta} = p^\lambda_\alpha\, p^\tau_\beta n_{(\lambda,\tau)} - \tfrac{1}{2}\mathscr{E} p_{\alpha\beta}. \tag{2.33}$$

Clearly,

$$\mathscr{R}_{\alpha\beta} n^\beta = 0; \quad \mathscr{S}_{\alpha\beta} n^\beta = 0, \quad \mathscr{S}^\alpha_\alpha = 0. \tag{2.34}$$

The rotation vector $\mathscr{R}^\alpha$ is defined by[29]

$$\mathscr{R}^\alpha = \tfrac{1}{2}\eta^{\alpha\beta\gamma}\mathscr{R}_{\beta\gamma}, \tag{2.35}$$

and since $\mathscr{R}_{\beta\gamma} n^\gamma = 0$ it is easily verified that $\mathscr{R}^\alpha$ is parallel to n^α; we also have

$$\mathscr{R}_{\alpha\beta} = \eta_{\alpha\beta\gamma}\mathscr{R}^\gamma, \tag{2.36}$$

and we define

$$\mathscr{R}^2 = \mathscr{R}_\alpha \mathscr{R}^\alpha = \tfrac{1}{2}\mathscr{R}_{\alpha\beta}\mathscr{R}^{\alpha\beta}. \tag{2.37}$$

Finally, it follows from (2.30) that $n_{\alpha,\beta}$ may be decomposed according to

$$n_{\alpha,\beta} = \mathscr{R}_{\alpha\beta} + \tfrac{1}{2}\mathscr{E} p_{\alpha\beta} + \mathscr{S}_{\alpha\beta} + \overset{*}{n}_\alpha n_\beta. \tag{2.38}$$

III. SPACELIKE CONFORMAL MOTIONS

Let $\xi^a = \xi n^a$, where $n_a n^a = +1$ and $\xi = (\xi_a \xi^a)^{1/2} > 0$. The infinitesimal point transformation

$$x^a \to x^a + \epsilon \xi^a, \tag{3.1}$$

defines a spacelike conformal motion of space-time, and ξ^a is a spacelike conformal Killing vector of space-time, if and only if

$$\underset{\xi}{\mathscr{L}} g_{ab} = 2\psi g_{ab}, \tag{3.2}$$

where $\underset{\xi}{\mathscr{L}}$ stands for the Lie derivative along ξ^a and $\psi(x^c)$ is some scalar function of position called the conformal factor. Equation (3.2) may be rewritten equivalently as

$$\xi_{(a;b)} = \psi g_{ab}. \tag{3.3}$$

We express the necessary and sufficient conditions for space-time to admit a spacelike conformal motion with symmetry vector $\xi^a = \xi n^a$ in terms of the properties of the spacelike congruence generated by n^a. We first state the corresponding result for space-time to admit a *timelike* conformal motion with symmetry vector $\eta^a = nv^a$, where $v_a v^a = -1$ and $\eta = (-\eta_a \eta^a)^{1/2} > 0$ (see Refs. 1–3).

Theorem 3.1: Space-time admits a timelike conformal motion with symmetry vector $\eta^a = \eta v^a$ ($v_a v^a = -1, \eta > 0$) if and only if

(i) $$\sigma_{ab} = 0, \tag{3.4}$$

(ii) $$\dot v_a = (\log \eta)_{,a} + (\theta/3) v_a, \tag{3.5}$$

where σ_{ab}, θ, and $\dot v_a$, are, respectively, the shear, expansion, and acceleration of the timelike congruence generated by v^a (see Ref. 30). The conformal factor ψ satisfies

$$\psi = \eta\theta/3. \tag{3.6}$$

It follows directly from (3.6) that a timelike conformal Killing vector is a proper conformal Killing vector if and only if $\theta \neq 0$. For a spacelike vector $\xi^a = \xi n^a$, we have the following result.

Theorem 3.2: Space-time admits a spacelike conformal motion with symmetry vector $\xi^a = \xi n^a$ ($n_a n^a = +1, \xi > 0$) if and only if at any given point

(i) $$\mathscr{S}_{ab} = 0, \tag{3.7}$$

(ii) $$\overset{*}{n}_a = -(\log \xi)_{,a} + \tfrac{1}{2}\mathscr{E} n_a, \tag{3.8}$$

(iii) $$w_a \mathring{n}^a = -\tfrac{1}{2}\mathscr{E}, \tag{3.9}$$

(iv) $$p^b_a(\mathring{n}_b + w^t n_{t;b}) = 0, \tag{3.10}$$

where $\mathscr{S}_{ab}$ and $\mathscr{E}$ are, respectively, the shear and expansion of the spacelike congruence generated by n^a and as measured by an observer with four-velocity w^a ($w_a w^a = -1$, $w_a n^a = 0$). The conformal factor ψ satisfies

$$\psi = \xi\mathscr{E}/2. \tag{3.11}$$

Proof: (a) Suppose first that space-time admits a spacelike conformal motion with symmetry vector $\xi^a = \xi n^a$. Then

$$\xi_{(a;b)} = \psi g_{ab} \tag{3.12}$$

for some scalar function $\psi(x^c)$. Contract (3.12) in turn with the tensors $w^a w^b$, $w^a n^b$, $w^a p^{bc}$, $n^a n^b$, $n^a p^{bc}$, and $p^{ac} p^{bd}$; this gives, with the aid of definition (2.18) for $\mathscr{S}_{ab}$,

$$w^a w^b: \quad w_a \mathring{n}^a = -\psi/\xi, \tag{3.13}$$

$$w^a n^b: \quad w^a\left[\overset{*}{n}_a + (\log \xi)_{,a}\right] = 0, \tag{3.14}$$

$$w^a p^{bc}: \quad p^{ab}\left[\mathring{n}_b + w^t n_{t;b}\right] = 0, \tag{3.15}$$

$$n^a n^b: \quad \overset{*}{\xi} = \psi, \tag{3.16}$$

$$n^a p^{bc}: \quad p^{ab}\left[\overset{*}{n}_b + (\log \xi)_{,b}\right] = 0, \tag{3.17}$$

$$p^{ac} p^{bd}: \quad \mathscr{S}_{ab} + \tfrac{1}{2}(\mathscr{E} - 2\psi/\xi)p_{ab} = 0. \tag{3.18}$$

First, take the trace of (3.18). Since $\mathscr{S}^a_a = 0$, we obtain

$$\mathscr{E} = 2\psi/\xi, \tag{3.19}$$

which gives (3.11). With (3.19), (3.18) reduces to condition (3.7), namely $\mathscr{S}_{ab} = 0$.

To derive (3.8), we note from (3.14) and (3.17) that $\overset{*}{n}_a + (\log \xi)_{,a}$ must be parallel to n_a; thus

$$\overset{*}{n}_a + (\log \xi)_{,a} = A n_a \tag{3.20}$$

for some scalar function A. Contracting (3.20) with n^a gives $A = (\log \xi)^*$ and on using (3.16) and (3.19), this may be rewritten as $A = \frac{1}{2}\mathscr{E}$, which establishes (3.8).

Condition (3.9) follows immediately from (3.13) and (3.19), and (3.10) is simply (3.15).

(b) Conversely, suppose that conditions (3.7)–(3.10) are satisfied. Then using expansion (2.23) with $\mathscr{S}_{ab} = 0$ for $n_{a;b}$ we have

$$\xi_{(a;b)} = \xi \Big[\tfrac{1}{2}\mathscr{E} p_{ab} + ((\log \xi)_{,(a} + \overset{*}{n}_{(a})n_{b)} + w_a w_b (w_t \mathring{n}^t) - w_{(a} p^t_{b)} (\mathring{n}_t + w^s n_{s;t}) \Big] . \tag{3.21}$$

Applying (3.8)–(3.10), we find that (3.21) reduces to

$$\xi_{(a;b)} = (\xi \mathscr{E}/2) g_{ab} . \tag{3.22}$$

Thus ξ^a is a conformal Killing vector of space-time with conformal factor $\psi = \xi\mathscr{E}/2$. □

Conditions (3.7)–(3.10) are in a form suitable for application to fluid space-times, which will be considered in Sec. V; in particular (3.9) provides a useful expression for the expansion $\mathscr{E}$. The two conditions (3.9) and (3.10), however, may be combined into one tensor equation, which can be expressed in several forms. For instance it is easily verified that (3.9) and (3.10) are equivalent to the single equation [obtained by expanding p^b_a in (3.10) and using (3.9)]

$$\mathring{n}_a + w^t n_{t;a} - (w^t \overset{*}{n}_t) n_a - \mathscr{E} w_a = 0 . \tag{3.23}$$

With the aid of the Greenberg transport law (2.12), (3.23) can be rewritten more concisely and in a form more consistent with the aim of expressing the conditions in terms of the properties of the associated spacelike congruence. In the derivation of Theorem 3.2, we did not use the Greenberg transport law, which the four-velocities w^a of all observers employed along the congruence must satisfy.

Theorem 3.3: Space-time admits a spacelike conformal motion with symmetry vector $\xi^a = \xi n^a$ ($n_a n^a = +1, \xi > 0$) if and only if

(i) $\mathscr{S}_{ab} = 0$, (3.24)

(ii) $\overset{*}{n}_a = -(\log \xi)_{,a} + \frac{1}{2}\mathscr{E} n_a$, (3.25)

(iii) $\underset{n}{\mathscr{L}} w_a = \frac{1}{2}\mathscr{E} w_a$, (3.26)

where $\mathscr{S}_{ab}$ and $\mathscr{E}$ are, respectively, the shear and expansion of the spacelike congruence generated by n^a and as measured by an observer with four-velocity w^a ($w_a w^a = -1$, $w_a n^a = 0$), which satisfies the Greenberg transport law (2.12). The conformal factor ψ is given by $\psi = \xi\mathscr{E}/2$.

Proof: We show with the aid of the Greenberg transport law (2.12) that (3.26) is equivalent to (3.9) and (3.10).

Suppose first that (3.26) is satisfied. Contracting (3.26) in turn with w^a and p^a_c yields, respectively, (3.9) and

$$p^a_c(\overset{*}{w}_a + w^t n_{t;a}) = 0 . \tag{3.27}$$

On using the Greenberg transport law (2.12), (3.10) follows immediately from (3.27).

Conversely, suppose that (3.9) and (3.10) are satisfied. Applying (2.12), (3.10) may be rewritten as (3.27), and on noting that $p^a_c = g^a_c + w^a w_c - n^a n_c$, (3.27) can be expressed as

$$\underset{n}{\mathscr{L}} w_c = -(w_b \mathring{n}^b) w_c . \tag{3.28}$$

Finally, making use of (3.9) we obtain (3.26). □

It follows from (3.11) that a spacelike conformal Killing vector is a *proper* conformal Killing vector if and only if $\mathscr{E} = 0$. Also from (3.25),

$$\overset{*}{n}_{[a;b]} = 0 \tag{3.29}$$

is a necessary and sufficient condition for $\xi^a = \xi n^a$ to be a spacelike Killing vector.

For comparison we state the corresponding theorem in Newtonian theory.

Theorem 3.4: Three-dimensional Euclidean space admits a conformal motion with symmetry vector $\xi^\alpha = \xi n^\alpha$ [$n_\alpha n^\alpha = +1, \xi = (\xi_\alpha \xi^\alpha)^{1/2} > 0$] if and only if

(i) $\mathscr{S}_{\alpha\beta} = 0$, (3.30)

(ii) $\overset{*}{n}_\alpha = -(\log \xi)_{,\alpha} + \frac{1}{2}\mathscr{E} n_\alpha$, (3.31)

where $\mathscr{S}_{\alpha\beta}$ and $\mathscr{E}$ are, respectively, the shear and expansion of the congruence of curves generated by n^α. The conformal factor ψ satisfies

$$\psi = \xi\mathscr{E}/2 . \tag{3.32}$$

□

The two conditions (3.24) and (3.25) for a spacelike conformal motion in space-time correspond directly with the two conditions (3.4) and (3.5) for a timelike conformal motion and with the two conditions (3.30) and (3.31) for a conformal motion in three-dimensional Euclidean space. Vanishing shear is a necessary and sufficient condition which applies in all three cases. However, condition (3.26) [or (3.9) and (3.10)] has no analog either for a timelike conformal motion or in Newtonian theory. Results derived from (3.26) therefore describe purely relativistic effects that apply specifically to *spacelike* conformal motions. These effects are most apparent in fluid space-times and will be considered in Sec. V.

We conclude this section with some applications of Theorem 3.3, which make use only of properties (3.25) and (3.11) for the conformal factor ψ. The results therefore have direct analogs in the theory of timelike conformal motions and in Newtonian theory.

It can be shown that two different timelike conformal motions cannot have the same streamlines (Williams[17] established this result for proper timelike homothetic motions) and that there do not exist any proper timelike homothetic motions of constant expansion.[17] For spacelike conformal motions we have the following.

Theorem 3.5: (i) Two different spacelike conformal motions cannot both have symmetry vectors parallel to the same spacelike unit vector n^a.

(ii) There do not exist any proper spacelike homothetic motions of constant expansion.

Proof: (i) Consider two spacelike conformal motions with symmetry vectors $\xi^a = \xi n^a$ and $\zeta^a = \zeta n^a$, where $\xi > 0$, $\zeta > 0$, and $n_a n^a = +1$. Then by (3.25),

$$\overset{*}{n}_a = -(\log \xi)_{,a} + \tfrac{1}{2}\mathscr{E} n_a , \tag{3.33}$$

$$\overset{*}{n}_a = -(\log \zeta)_{,a} + \tfrac{1}{2}\mathscr{E} n_a . \tag{3.34}$$

On subtracting (3.34) from (3.33) we obtain

$$(\log(\xi/\zeta))_{,a} = 0 ; \tag{3.35}$$

hence, $\xi = \alpha\zeta$, where α is a constant, which contradicts the assumption that the conformal motions are different.

(ii) Suppose space-time admits a proper spacelike homothetic motion of constant expansion. From (3.11), $\psi = \xi\mathscr{E}/2$, and we have in general

$$2\psi_{,a} = \xi_{,a}\mathscr{E} + \mathscr{E}_{,a}\xi . \tag{3.36}$$

But for a homothetic motion, ψ is constant and if the expansion $\mathscr{E}$ is also constant then (3.36) reduces to $\mathscr{E}_{,a} = 0$. Hence (3.25) becomes

$$\overset{*}{n}_a = \tfrac{1}{2}\mathscr{E} n_a , \tag{3.37}$$

and on contracting (3.37) with n^a we find that $\mathscr{E} = 0$. Thus by (3.11), $\psi = 0$, which contradicts the assumption that the homothetic motion is a proper homothetic motion. □

It follows from (3.25) and (3.36) that the expansion $\mathscr{E}$ of a spacelike homothetic motion must satisfy the differential equation

$$\overset{*}{\mathscr{E}} + \tfrac{1}{2}\mathscr{E}^2 = 0 . \tag{3.38}$$

Theorem 3.5 remains valid in Newtonian theory.

As a second application of Theorem 3.3 we consider the special case in which $\overset{*}{n}^a = 0$. For a timelike conformal motion, the corresponding condition is $\dot{v}^a = 0$ and the congruence consists of timelike geodesic curves. Oliver and Davis[3] have established the following.

Theorem 3.6: If $\eta^a = \eta v^a$ $(v_a v^a = -1)$ is a timelike conformal Killing vector and $\dot{v}^a = 0$ then either (i) $\theta = 0$ and v^a is a timelike Killing vector or (ii) $\theta \neq 0$ and $\omega = 0$, $h^{\ b}_{a}\theta_{,b} = 0$. □

Since the shear of the congruence σ_{ab} vanishes, this result is reminiscent of the class of results for a shear-free perfect fluid with barotropic equation of state in which necessarily $\omega\theta = 0$, established by Ellis[31] for dust ($\dot{u}^a = 0$) and recently extended by Collins,[22] White and Collins,[32] and Collins and White.[33] For a spacelike congruence we have the following.

Theorem 3.7: If $\xi^a = \xi n^a$ $(n_a n^a = +1, \xi > 0)$ is a spacelike conformal Killing vector and $\overset{*}{n}^a = 0$, then either (i) $\mathscr{E} = 0$ and n^a is a spacelike Killing vector or (ii) $\mathscr{E} \neq 0$ and $\mathscr{R} = 0$, $p^b_a \mathscr{E}_{,b} = 0$.

Proof: Since $\overset{*}{n}^a = 0$, (3.25) reduces to

$$\mathscr{E} n_a = 2(\log \xi)_{,a} . \tag{3.39}$$

(i) $\mathscr{E} = 0$. From (3.11) we have $\psi = \xi\mathscr{E}/2 = 0$ and hence ξn^a is a Killing vector. But since $\mathscr{E} = 0$, it follows from (3.39) that $\xi_{,a} = 0$ and therefore that ξ is constant. Thus n_a itself satisfies Killing's equation.

(ii) $\mathscr{E} \neq 0$. Covariantly differentiate (3.39) with respect to x^b and take the skew part; this gives

$$n_{[a}\mathscr{E}_{,b]} + \mathscr{E} n_{[a;b]} = 0 . \tag{3.40}$$

Project on (3.40) with $p^a_c p^b_d$ and use definition (2.16) for $\mathscr{R}_{ab}$. Then

$$\mathscr{E}\mathscr{R}_{cd} = 0 , \tag{3.41}$$

and since $\mathscr{E} \neq 0$ it follows that $\mathscr{R}_{cd} = 0$. Alternatively, by projecting on (3.40) with $n^a p^{bc}$ and using $\overset{*}{n}_b = 0$, we obtain $p^{cb}\mathscr{E}_{,b} = 0$. □

It follows from Theorem 3.7 that if ξn^a is a spacelike conformal Killing vector and $\overset{*}{n}^a = 0$, then either the expansion $\mathscr{E}$ or the rotation $\mathscr{R}$ of the spacelike congruence generated by n^a, vanishes, i.e., $\mathscr{E}\mathscr{R} = 0$. This result remains valid in Newtonian theory since it was established using only (3.25).

Finally we note that in the statements of Theorems 3.2 and 3.3, only one condition , (3.8) [or equivalently (3.25)], contains the factor ξ. Both theorems may be stated in a way independent of ξ by replacing (3.9) and (3.25) with

$$(\overset{*}{n}_{[a} - \tfrac{1}{2}\mathscr{E} n_{[a})_{;b]} = 0 . \tag{3.42}$$

A similar comment applies to the statements of Theorems 3.1 and 3.4.

IV. PROPAGATION EQUATIONS

Further properties of the spacelike congruence generated by a conformal Killing vector can be determined by considering the propagation equations for tensor quantities along a curve of the congruence. In this section we will consider the propagation equation for the projection tensor p_{ab}, which has application to the concept of a rigid spacelike congruence, and the propagation equation for the rotation tensor $\mathscr{R}_{ab}$. The expansion $\mathscr{E}$ enters these equations as a proportionality factor. An application of the propagation equation for the expansion will be made in Sec. V when determining the condition for a spacelike conformal Killing vector to be homothetic in a certain fluid space-time. Since the shear $\mathscr{S}_{ab}$ of a spacelike congruence generated by a conformal Killing vector vanishes along the congruence, the propagation equation for $\mathscr{S}_{ab}$ is not considered.

A. Projection tensor

Before considering the propagation equation for the projection tensor p_{ab} we briefly review the results for a congruence of timelike curves with unit tangent vector v^a $(v_a v^a = -1)$. The projection tensor onto the instantaneous rest-space of v^a is $h_{ab} = g_{ab} + v_a v_b$.

Theorem 4.1: (i) For a timelike congruence of curves with unit tangent vector v^a,

$$\underset{v}{\mathscr{L}}\, h_{ab} = 2(\sigma_{ab} + (\theta/3)\, h_{ab}) \tag{4.1}$$

and

$$\underset{v}{\mathscr{L}}\, h_{ab} = 0\,, \quad \text{iff } \sigma_{ab} = 0 \text{ and } \theta = 0\,. \tag{4.2}$$

(ii) If space-time admits a timelike conformal motion with symmetry vector parallel to v^a then

$$\underset{v}{\mathscr{L}}\, h_{ab} = \tfrac{2}{3}\theta h_{ab}\,. \tag{4.3}$$

□

A timelike congruence may be defined as rigid if the orthogonal distance between any neighboring pair of curves remains constant along the curves.[17,18,34–38] It can be shown that, according to this definition, a timelike congruence is rigid if and only if

$$\underset{v}{\mathscr{L}}\, h_{ab} = 0\,, \tag{4.4}$$

and hence from (4.2), it is rigid if and only if it is shear-free and expansion-free. A congruence will be called isometric if the unit tangent vector to the curves is parallel to a Killing vector. By (3.4) and (3.6), an isometric timelike congruence is shear-free and expansion-free, and hence every isometric timelike congruence is rigid. Although the converse, that a rigid timelike congruence is isometric, is true for a rotating timelike congruence in flat space-time[18,35,36] and in a space-time of constant curvature,[17] it is not true in general.[18,37]

We now establish corresponding results for a spacelike congruence.

Theorem 4.2: (i) For a spacelike congruence of curves with unit tangent vector n^a,

$$\underset{n}{\mathscr{L}}\, p_{ab} = 2(\mathscr{S}_{ab} + \tfrac{1}{2}\mathscr{E}\, p_{ab})\,, \tag{4.5}$$

where $p_{ab} = g_{ab} + w_a w_b - n_a n_b$ $(w_a w^a = -1,\ w_a n^a = 0)$, w^a satisfies the Greenberg transport law, and $\mathscr{S}_{ab}$ and $\mathscr{E}$ are the shear and expansion of the congruence as measured by w^a. Also,

$$\underset{n}{\mathscr{L}}\, p_{ab} = 0\,, \quad \text{iff } \mathscr{S}_{ab} = 0 \text{ and } \mathscr{E} = 0\,. \tag{4.6}$$

(ii) If space-time admits a spacelike conformal motion with symmetry vector parallel to n^a then

$$\underset{n}{\mathscr{L}}\, p_{ab} = \mathscr{E} p_{ab}\,. \tag{4.7}$$

Proof: (i) We have

$$\underset{n}{\mathscr{L}}\, p_{ab} = \overset{*}{p}_{ab} + 2p_{c(a} n^c{}_{;b)}\,. \tag{4.8}$$

On using (2.23) for $n^c{}_{;b}$, (4.8) becomes

$$\underset{n}{\mathscr{L}}\, p_{ab} = 2(\mathscr{S}_{ab} + \tfrac{1}{2}\mathscr{E} p_{ab}) + 2w_{(a}(\overset{*}{w}_{b)} + n_{b)} w_c \overset{*}{n}{}^c - p^c_{b)} \mathring{n}_c)\,, \tag{4.9}$$

and since $w_c \overset{*}{n}{}^c = -\overset{*}{w}_c n^c$ (because $w_c n^c = 0$), (4.9) may be rewritten as

$$\underset{n}{\mathscr{L}}\, p_{ab} = 2(\mathscr{S}_{ab} + \tfrac{1}{2}\mathscr{E} p_{ab}) + 2w_{(a}\, p^c_{b)}(\overset{*}{w}_c - \mathring{n}_c)\,. \tag{4.10}$$

On applying the Greenberg transport law for w^a in the form (2.12), (4.10) reduces to (4.5).

Clearly if $\mathscr{S}_{ab} = 0$ and $\mathscr{E} = 0$, then $\underset{n}{\mathscr{L}}\, p_{ab} = 0$. Conversely, suppose that $\underset{n}{\mathscr{L}}\, p_{ab} = 0$; then

$$\mathscr{S}_{ab} + \tfrac{1}{2}\mathscr{E} p_{ab} = 0\,. \tag{4.11}$$

Since $\mathscr{S}^a_a = 0$ and $p^a_a = 2$, the trace of (4.11) gives $\mathscr{E} = 0$, and therefore from (4.11), $\mathscr{S}_{ab} = 0$.

(ii) If space-time admits a spacelike conformal motion with symmetry vector parallel to n^a, then by (3.24), $\mathscr{S}_{ab} = 0$, and (4.5) reduces to (4.7). □

By direct analogy with the condition for a rigid timelike congruence, Ciubotariu[6] defined a spacelike congruence to be rigid if the orthogonal distance between any neighboring pair of curves remains constant along the curves, and he proved that a spacelike congruence is rigid if and only if

$$\underset{n}{\mathscr{L}}\, p_{ab} = 0\,. \tag{4.12}$$

Ciubotariu also showed that (4.12) implies that $\mathscr{S}_{ab} = 0$ and $\mathscr{E} = 0$. From (4.6), the converse is also true; a spacelike congruence is rigid if and only if it is shear-free and expansion-free. For an isometric spacelike congruence, the shear and expansion both vanish by (3.7) and (3.11), and hence every isometric spacelike congruence is rigid. We do not investigate here the converse of this result.

These results can be illustrated by the vortex congruence in the Gödel universe[6] and this will be considered briefly in Sec. V. In Newtonian theory the projection tensor is $p_{\alpha\beta} = h_{\alpha\beta} - n_\alpha n_\beta$ and Theorem 4.2 remains valid.

B. Rotation tensor

Before considering the propagation equation for the rotation tensor $\mathscr{R}_{ab}$, we briefly review the results for the vorticity tensor ω_{ab} of a congruence of timelike curves.[30] There is a direct analogy between the equations governing ω_{ab} and $\mathscr{R}_{ab}$.

Theorem 4.3: (i) If space-time admits a timelike conformal motion with symmetry vector parallel to v^a $(v_a v^a = -1)$, then

$$h^c_a h^d_b \dot{\omega}_{cd} + (\theta/3)\omega_{ab} = 0\,, \tag{4.13}$$

$$\underset{v}{\mathscr{L}}\, \omega_{ab} = (\theta/3)\omega_{ab}\,, \tag{4.14}$$

$$\underset{v}{\mathscr{L}}\, \omega^a = -\tfrac{4}{3}\theta\omega^a + (\omega^b \dot{v}_b)v^a\,, \tag{4.15}$$

where ω^a is the vorticity vector.[30]

(ii) If, further, τ denotes proper time measured along the world-line of v^a, and δV is the volume of a material element as measured by v^a, then

$$\frac{D\omega}{d\tau} + \frac{\theta}{3}\omega = 0\,, \tag{4.16}$$

$$\frac{D}{d\tau}(\omega^3 \delta V) = 0\,, \tag{4.17}$$

where $\omega^2 = \omega_a \omega^a = \tfrac{1}{2}\omega_{ab}\omega^{ab}$. □

Equations (4.13)–(4.17) are purely kinematical. Since $\sigma_{ab} = 0$ for a timelike conformal motion, it follows that along the world-line of v^a, $\delta V \propto (\delta l)^3$, where δl is a characteristic length associated with the material element, and (4.17) may be rewritten as[39]

$$\frac{D}{d\tau}(\omega\delta l)=0\,. \tag{4.18}$$

We now establish corresponding results for a spacelike congruence.

Theorem 4.4: If space-time admits a spacelike conformal motion with symmetry vector parallel to n^a ($n_a n^a = +1$) then

(i) $p_a^c\, p_b^d \overset{*}{\mathscr{R}}_{cd} + \tfrac{1}{2}\mathscr{E}\mathscr{R}_{ab} = 0\,,$ (4.19)

(ii) $\underset{n}{\mathscr{L}}\mathscr{R}_{ab} = \tfrac{1}{2}\mathscr{E}\mathscr{R}_{ab}\,,$ (4.20)

(iii) $\underset{n}{\mathscr{L}}\mathscr{R}^a = -\tfrac{1}{2}\mathscr{E}\mathscr{R}^a\,.$ (4.21)

Proof: (i) It is shown in Appendix B that, for any spacelike congruence, the equation governing the propagation of $\mathscr{R}_{ab}$ along a curve of the congruence is

$$p_a^c\, p_b^d(\overset{*}{\mathscr{R}}_{cd} - \overset{*}{n}_{[c;d\,]} + 2w^t n_{t;[c}\mathring{n}_{d\,]}) - 2\mathscr{S}^c_{\ [a}\mathscr{R}_{b\,]c} + \mathscr{E}\mathscr{R}_{ab} = 0\,. \tag{4.22}$$

We use Theorem 3.3 to simplify (4.22). With the aid of (3.25) for $\overset{*}{n}_c$ and definition (2.16) for $\mathscr{R}_{ab}$, it is easily verified that

$$p_a^c\, p_b^d \overset{*}{n}_{[c;d\,]} = \tfrac{1}{2}\mathscr{E}\mathscr{R}_{ab}\,. \tag{4.23}$$

Also, using (3.26) it can be shown that

$$p_a^c\, p_b^d w^t n_{t;[c}\mathring{n}_{d\,]} = -p_a^c\, p_b^d \overset{*}{w}_{[c}\mathring{n}_{d\,]}\,, \tag{4.24}$$

and on applying the Greenberg transport law in the form (2.12) (which the four-velocity w^a of all observers employed along the congruence must satisfy), (4.24) becomes

$$p_a^c\, p_b^d w^t n_{t;[c}\mathring{n}_{d\,]} = -p_a^c\, p_b^d \mathring{n}_{[c}\mathring{n}_{d\,]} = 0\,. \tag{4.25}$$

On substituting from (4.23) and (4.25) into (4.22) and noting also that $\mathscr{S}_{ab}=0$ by (3.24), Eq. (4.19) is obtained.

(ii) It is shown in Appendix B that for any spacelike congruence, (4.22) may be rewritten in the form

$$\underset{n}{\mathscr{L}}\mathscr{R}_{ab} = p_a^c\, p_b^d(\overset{*}{n}_{[c;d\,]} - 2w^t n_{t;[c}\mathring{n}_{d\,]})\,. \tag{4.26}$$

Equation (4.20) follows directly from (4.26) using (4.23) and (4.25).

(iii) Since

$$\mathscr{R}^a = \tfrac{1}{2}\eta^{abcd}w_b\mathscr{R}_{cd}\,, \tag{4.27}$$

we have

$$\underset{n}{\mathscr{L}}\mathscr{R}^a = \tfrac{1}{2}\underset{n}{\mathscr{L}}(\eta^{abcd})w_b\mathscr{R}_{cd} + \tfrac{1}{2}\eta^{abcd}(\underset{n}{\mathscr{L}}w_b)\mathscr{R}_{cd} + \tfrac{1}{2}\eta^{abcd}w_b\underset{n}{\mathscr{L}}\mathscr{R}_{cd}\,. \tag{4.28}$$

But,

$$\underset{n}{\mathscr{L}}\eta^{abcd} = -n^t_{\ ;t}\,\eta^{abcd}\,, \tag{4.29}$$

and from (2.17) and (3.9) [or (3.26)] it follows that $n^t_{\ ;t} = \tfrac{3}{2}\mathscr{E}$; thus

$$\underset{n}{\mathscr{L}}\eta^{abcd} = -\tfrac{3}{2}\mathscr{E}\eta^{abcd}\,. \tag{4.30}$$

On evaluating the second and third terms on the right-hand side of (4.28) using (3.26) and (4.20), (4.21) is obtained. □

Equations (4.19) and (4.20) were first derived by Prasad and Sinha[14] and correspond directly to (4.13) and (4.14) for a congruence of timelike curves. Equations (4.15) and (4.21) differ in form because there is no condition corresponding to (3.26) for a timelike congruence. However, we do have

$$\underset{\xi n}{\mathscr{L}}\mathscr{R}_a = 0\,, \quad \underset{\eta v}{\mathscr{L}}\omega_a = 0 \tag{4.31}$$

(and also $\mathscr{L}_v\omega_a = 0$), which is easily verified. Equations (4.19)–(4.21) show that a close relationship exists between the rotation $\mathscr{R}$ and the expansion $\mathscr{E}$ of a spacelike congruence generated by a conformal Killing vector. To develop this relationship further we make use of the concept of a flux tube.[40] We define a flux tube to be the two-dimensional surface swept out by the curves of the spacelike congruence that pass through a given simple closed curve. At any given point along the flux tube, we denote by δA the cross-sectional area of the flux tube as measured by observer w^a at that point; p denotes arc length measured along a curve of the congruence.

Theorem 4.5: If space-time admits a spacelike conformal motion with symmetry vector parallel to n^a ($n_a n^a = +1$) then

(i) $\dfrac{D\mathscr{R}}{dp} + \tfrac{1}{2}\mathscr{E}\mathscr{R} = 0\,,$ (4.32)

(ii) $\dfrac{D}{dp}(\mathscr{R}^2\delta A) = 0\,,$ (4.33)

where $\mathscr{R}$ and $\mathscr{E}$ are the rotation and expansion of the spacelike congruence generated by n^a and δA is the cross-sectional area of a flux tube formed by curves of the congruence.

Proof: (i) Equation (4.32) can be derived by starting from any one of (4.19)–(4.21). For instance, by contracting (4.19) with $\mathscr{R}^{ab}$, noting that $\mathscr{R}_{ab}\mathscr{R}^{ab} = 2\mathscr{R}^2$, and using definition (2.10), (4.32) is immediately obtained.

(ii) Since[15]

$$\mathscr{E} = \frac{1}{\delta A}\frac{D\delta A}{dp}\,, \tag{4.34}$$

Eq. (4.32) may be written as

$$\frac{1}{\mathscr{R}}\frac{D\mathscr{R}}{dp} + \frac{1}{2\delta A}\frac{D\delta A}{dp} = 0\,, \tag{4.35}$$

from which (4.33) is immediately derived. □

Equation (4.33), which was first derived by Prasad and Sinha,[14] yields a counterpart of the first Helmholtz theorem for vortex tubes,[40] and like the first Helmholtz theorem it is purely kinematical. Since $\mathscr{R}$ and δA are measured by an observer w^a satisfying the Greenberg transport law, (4.33) is valid along a curve of the congruence and not simply at one point. Thus $\mathscr{R}^2\,\delta A$ is conserved along a flux tube and can be defined as the strength of the flux tube. We see that the magnitude of the rotation of the curves varies inversely as the square root of the element of area normal to the curves. Since by (3.24), the shear of the congruence vanishes, along a curve $\delta A \propto (\delta l)^2$, where δl is a characteristic length in the "screen" normal to n^a erected by observer w^a. Hence by (4.33),

$$\frac{D}{dp}(\mathscr{R}\delta l) = 0\,, \tag{4.36}$$

which corresponds directly with (4.18).

For the special case of a spacelike motion, $\mathscr{E} = 0$ by (3.11) and therefore δA is conserved along a flux tube. Equations (4.32), (4.33), and (4.36) all reduce to

$$\frac{D\mathscr{R}}{dp} = 0\,. \tag{4.37}$$

The magnitude of the rotation of the curves of an isometric spacelike congruence is therefore conserved along the congruence and hence $\mathscr{R}$ can be taken as a measure of the strength of a flux tube formed by curves of the congruence.

The results obtained here for rotation have a direct analogy with results in Newtonian theory, which can be derived by starting from the propagation equations[15]

$$p^\lambda_\alpha\, p^\tau_\beta(\overset{*}{\mathscr{R}}_{\lambda\tau} - \overset{*}{n}_{[\lambda,\tau]}) - 2\mathscr{S}_{\tau[\alpha}\mathscr{R}_{\beta\,]}{}^\tau + \mathscr{E}\mathscr{R}_{\alpha\beta} = 0\,, \tag{4.38}$$

$$\underset{n}{\mathscr{L}}\mathscr{R}_{\alpha\beta} = p^\lambda_\alpha\, p^\tau_\beta \overset{*}{n}_{[\lambda,\tau]}\,. \tag{4.39}$$

Theorem 4.4 (with Latin indices replaced by Greek indices) and Theorem 4.5 remain valid in three-dimensional Euclidean space.

V. FLUID SPACE-TIMES

Let u^a $(u_a u^a = -1)$ be the four-velocity field of a self-gravitating fluid. The unit vector tangent to several important spacelike vector fields satisfies the property $n_a u^a = 0$. Examples are the unit vector tangent to vortex lines in a rotational fluid, where the local vorticity vector is defined by

$$\omega^a = \tfrac{1}{2}\eta^{abcd}u_b u_{c;d}\,, \tag{5.1}$$

and the unit vector tangent to electric and magnetic field lines in an electrically conducting fluid, where the electric and magnetic field four-vectors are defined, respectively, in terms of the skew electromagnetic field tensor F_{ab} by

$$E^a = F^{ab}u_b\,,\quad H^a = \tfrac{1}{2}\eta^{abcd}u_b F_{cd}\,. \tag{5.2}$$

In the remainder of this paper we will be concerned exclusively with fluid space-times and with spacelike conformal motions with symmetry vector parallel to n^a, where n^a satisfies the condition $n_a u^a = 0$.

When $n_a u^a = 0$, condition (2.3) is satisfied by u^a and an observer $w^a = u^a$ comoving with the fluid can be employed at any one point of the spacelike congruence generated by n^a. [Having chosen a comoving observer at one point, comoving observers can be employed all along a curve of this congruence if and only if the curve is a material curve in the fluid; this is the necessary and sufficient condition for u^a to satisfy the Greenberg transport law for w^a (see Ref. 15).] We rewrite in terms of u^a the necessary and sufficient conditions for a fluid space-time to admit a spacelike conformal motion. The kinematical quantities of the fluid and in particular the fluid vorticity ω^a are thereby introduced through the covariant derivative $u_{a;b}$ of u_a.

An overhead dot denotes covariant differentiation along a fluid particle world-line and $h_{ab} = g_{ab} + u_a u_b$ (see Ref. 30).

Theorem 5.1: If $n_a u^a = 0$, then a fluid space-time admits a spacelike conformal motion with symmetry vector ξn^a $(n_a n^a = +1, \xi > 0)$ if and only if at any given point,

(i) $\mathscr{S}_{ab} = 0\,,$ (5.3)

(ii) $\overset{*}{n}_a = -(\log\xi)_{,a} + \tfrac{1}{2}\mathscr{E}n_a\,,$ (5.4)

(iii) $u_a\dot{n}^a = -\tfrac{1}{2}\mathscr{E}\,,$ (5.5)

(iv) $\omega^a = (\omega_t n^t)n^a + \tfrac{1}{2}\eta^{abcd}(h^t_{\ b}\dot{n}_t - \overset{*}{u}_b)u_c n_d\,,$ (5.6)

where $\mathscr{S}_{ab}$ and $\mathscr{E}$ are, respectively, the shear and expansion of the spacelike congruence generated by n^a and as measured by an observer with four-velocity u^a, and ω^a is the local vorticity vector of the fluid. The conformal factor ψ satisfies

$$\psi = \xi\mathscr{E}/2\,. \tag{5.7}$$

Proof: Consider any given point P on a curve of the spacelike congruence generated by n^a. Since $n_a u^a = 0$, condition (2.3) is satisfied by $w^a = u^a$ and therefore a comoving observer u^a can be employed at P. Having chosen a comoving observer at P, in general it will not be possible to employ a comoving observer at a neighboring point of the curve since w^a must satisfy the Greenberg transport law; we write

$$w^a = u^a + \lambda^a\,,\quad \lambda^a(P) = 0\,, \tag{5.8}$$

where to satisfy (2.3), $\lambda_a\lambda^a + 2u_a\lambda^a = 0$ and $\lambda_a n^a = 0$. In general, at a neighboring point, $\lambda^a \neq 0$ and

$$\left(\frac{D\lambda^a}{dp}\right)_P = \overset{*}{\lambda}{}^a(P) \neq 0\,. \tag{5.9}$$

We can proceed using either the results of Theorem 3.2 or Theorem 3.3. It is of interest to consider both approaches.

(a) Application of Theorem 3.2: Suppose that the conditions (3.7)–(3.10) of Theorem 3.2 are satisfied. The *derivatives* of w^a do not occur in (3.7)–(3.10) and therefore at the given point P, w^a can simply be replaced by u^a in (3.7)–(3.10). Conditions (5.3)–(5.5) are obtained immediately from (3.7)–(3.9), where now $\mathscr{S}_{ab}$ and $\mathscr{E}$ are the shear and expansion of the congruence as measured by observer u^a. With $w^a = u^a$, (3.10) becomes

$$p^b_a(\dot{n}_b + u^c n_{c;b}) = 0\,, \tag{5.10}$$

where $p^b_a = g^b_a + u_a u^b - n_a n^b$. Equation (5.10) may be rewritten as

$$h^{\ b}_a\dot{n}_b + u^c n_{c;a} + (u^c\dot{n}_c)u_a - (u^c\overset{*}{n}_c)n_a = 0\,. \tag{5.11}$$

But since $n_c u^c = 0$ we have

$$u^c n_{c;a} = -2n^c u_{[c;a]} - \overset{*}{u}_a\,, \tag{5.12}$$

and since

$$u_{c;a} = \theta_{ca} + \omega_{ca} - \dot{u}_c u_a\,, \tag{5.13}$$

where $\theta_{ca} = \theta_{(ca)}$ is the (rate of) expansion tensor and $\omega_{ca} = \omega_{[ca]}$ is the vorticity tensor, (5.12) may be rewritten as

$$u^c n_{c;a} = -2n^c\omega_{ca} - (u^c\dot{n}_c)u_a - \overset{*}{u}_a\,. \tag{5.14}$$

On substituting from (5.14) into (5.11), we obtain

$$2n^c\omega_{ca} = h_a^b\dot{n}_b - \overset{*}{u}_a - (u^c\overset{*}{n}_c)n_a \,. \tag{5.15}$$

But[23,24]

$$\omega_{ca} = \eta_{capq}\omega^p u^q \,, \tag{5.16}$$

and therefore by operating on (5.15) with $\frac{1}{2}\eta^{arst}u_s n_t$, (5.6) is obtained.[28]

The steps are reversible, and therefore conversely conditions (5.3)–(5.6) imply (3.7)–(3.10) with the particular choice $w^a = u^a$.

(b) Application of Theorem 3.3: Suppose that the conditions (3.24)–(3.26) of Theorem 3.3 are satisfied. The derivatives of w^a do not occur in (3.24) and (3.25) and therefore (5.3) and (5.4) are obtained immediately from (3.24) and (3.25) by setting $w^a = u^a$. On substituting (5.8) into (3.26) we obtain at P,

$$\overset{*}{u}_a + \overset{*}{\lambda}_a + u_b n^b{}_{;a} = \tfrac{1}{2}\mathscr{E}u_a \,, \tag{5.17}$$

where $\mathscr{E}$ is the expansion of the congruence as measured by u^a. To obtain $\overset{*}{\lambda}_a$ at P, we substitute (5.8) into the Greenberg transport law in the form (2.25):

$$\overset{*}{\lambda}_a(P) = h_a^b\dot{n}_b - \overset{*}{u}_a - (u_b\overset{*}{n}{}^b)n_a \,. \tag{5.18}$$

[From Appendix A, $\overset{*}{\lambda}{}^a(P) = 0$ if and only if n^a is the unit tangent vector to a material curve in the fluid.] Using (5.18), (5.17) at P assumes the form

$$h_a^b\dot{n}_b + u^b n_{b;a} - (u_b\overset{*}{n}{}^b)n_a = \tfrac{1}{2}\mathscr{E}u_a \,. \tag{5.19}$$

On contracting (5.19) in turn with u^a and h_c^a we obtain (5.5) for $\mathscr{E}$ and (5.11). Equation (5.6) then follows from (5.11) as in part (a).

Again, the steps are reversible, and therefore conversely (5.3) and (5.4) imply (3.24) and (3.25) with the choice $w^a = u^a$, and (5.5) and (5.6) together with the Greenberg transport law imply (3.26) evaluated at P. □

Theorem 5.1 can be established without appeal to the Greenberg transport law as demonstrated in proof (a) above, since the statement of Theorem 3.2 does not contain derivatives of w^a. In proof (b), the Greenberg transport law was essentially applied twice, since it was used originally to derive (3.26), and its effect cancels.

The three conditions (5.3)–(5.5) were derived by Prasad and Sinha,[14] but the fourth condition they left as (5.11) instead of expressing it in terms of the fluid vorticity vector as in (5.6). They stated their theorem in terms of the unit tangent vector n^a to a magnetic field line, but the results apply to any spacelike vector field with $n_a u^a = 0$.

Unlike (5.3) and (5.4), which depend only on the properties of the spacelike congruence generated by n^a, (5.5) and (5.6) depend on the kinematical quantities of the timelike congruence generated by u^a, such as its aceleration $\dot{u}^a(u_a\dot{n}^a = -n_a\dot{u}^a)$ and vorticity ω^a. The expansion $\mathscr{E}$ and the acceleration $\dot{u}^a$ of the two congruences are related through (5.5), while (5.6) is expressed in terms of the timelike congruence only.

When the statement of Theorem 5.1 is compared with the statements of Theorems 3.1 and 3.4, it is clear that conditions (5.5) and (5.6) have no analog either for a timelike conformal motion or for a conformal motion in three-dimensional Euclidean space; (5.5) and (5.6) describe properties that apply only to *spacelike* conformal Killing vectors and these properties are purely relativistic. For instance, it follows from (5.5) [and (5.7)] that a spacelike conformal Killing vector parallel to n^a $(n_a u^a = 0)$ is a proper conformal Killing vector if and only if $n_a\dot{u}^a \neq 0$, a result which does not have an immediate analog either for a timelike conformal Killing vector, or in Newtonian theory. Results derived from (5.6) will be considered in Sec. V A below.

Theorem 5.1 may be applied in two ways which we will consider in turn: it can be used in a fluid space-time either to derive general properties of spacelike conformal motions or to establish the existence of spacelike conformal motions. We will consider the former application first and we will establish a connection between spacelike conformal motions of space-time and material curves in the fluid. As an example of the latter application we will then consider spacelike conformal Killing vectors parallel to the fluid vorticity vector ω^a.

A. Material curves

A material curve in a fluid is a curve which consists at all times of the same fluid particles and therefore it moves with the fluid as the fluid evolves. That *timelike* conformal Killing vectors lead to physically significant material curves in fluid space-times is well established: if a rotational fluid space-time admits a timelike conformal motion parallel to u^a, then the vortex lines are material lines in the fluid. To prove this, we consider (3.5) with $v^a = u^a$ and contract it with u^a to give $\theta/3 = (\log\eta)_{,a}u^a$. Hence (3.5) can now be expressed as

$$\dot{u}^a = -(\log(1/\eta))_{,b}\, h^{ba} \,. \tag{5.20}$$

Thus $1/\eta$ is an acceleration potential and it follows, with the aid of the vorticity propagation equation, that the vortex lines are "frozen-in" to the fluid.[24] This result is due to a property of the flow and is not due to a physical property of the fluid. Although the spacelike analog of (3.5) is (5.4), (5.4) is not used to derive the corresponding result for a spacelike conformal motion; this is obtained from condition (5.6)

Theorem 5.2: Suppose a fluid space-time admits a spacelike conformal motion with symmetry vector parallel to n^a, where $n_a n^a = +1$ and $n_a u^a = 0$.

(i) If the fluid is irrotational $(\omega = 0)$, then the integral curves of n^a must be material curves in the fluid.

(ii) If the vorticity of the fluid is nonzero $(\omega \neq 0)$, then the integral curves of n^a are material curves in the fluid if and only if they are vortex lines.

Proof: Since $n_a u^a = 0$ and space-time admits a conformal Killing vector parallel to n^a, (5.6) must be satisfied:

$$\omega^a = (\omega_t n^t)n^a + \tfrac{1}{2}\eta^{abcd}(h_b^t\dot{n}_t - \overset{*}{u}_b)u_c n_d \,. \tag{5.21}$$

(i) $\omega=0$. When $\omega^a=0$, (5.21) reduces to

$$\eta^{abcd}(h^t_{\ b}\dot n_t-\overset{*}{u}_b)u_c n_d=0\,, \tag{5.22}$$

and by operating on (5.22) with $\eta_{afrs}u^r n^s$ we obtain[28]

$$h^f_{\ t}\dot n^t=\overset{*}{u}{}^f-(n_t\overset{*}{u}{}^t)n^f\,. \tag{5.23}$$

But, from Appendix B, (5.23) is the necessary and sufficient condition for n^a to be the unit tangent vector to a material curve in the fluid.

(ii) $\omega\neq 0$. Suppose first that n^a is the unit tangent vector to a material curve in the fluid. Then n^a must satisfy the propagation Eq. (5.23), and on substituting (5.23) into (5.21) we obtain

$$\omega^a=(\omega_t n^t)n^a\,. \tag{5.24}$$

Thus since $\omega\neq 0$, it follows that $n^a=\pm\omega^a/\omega$ and hence n^a is the unit tangent vector to the vortex lines in the fluid.

Conversely, suppose that $n^a=\pm\omega^a/\omega$. Then (5.21) reduces to (5.22) which, as in part (i), implies (5.23); thus the vortex lines are material lines in the fluid. □

Theorem 5.2 is a direct consequence of condition (5.6), which has no analog in Newtonian theory. The results are therefore purely relativistic. They do not depend on the nature of the fluid, but are a property of the flow. Expressed in terms of spacelike congurences, Theorem 5.2 states that if a fluid space-time admits a spacelike conformal motion parallel to n^a, then the congruence generated by n^a must consist of material curves if $\omega=0$ and therefore move with the fluid, while if $\omega\neq 0$ the congruence is a vortex congruence if and only if it consists of material curves.

We illustrate Theorem 5.2 by two well-known universe models, the Friedman–Robertson–Walker (FRW) models ($\omega=0$) and the Gödel model ($\omega\neq0$).[41–45] The spacelike symmetry vectors are Killing vectors.

1. FRW models

The line element is

$$ds^2=-dx^{0^2}+\frac{R^2(x^0)}{(1+\frac14 kr^2)^2}(dx^{1^2}+dx^{2^2}+dx^{3^2})\,,$$
$$k=0,\pm1\,, \tag{5.25}$$

where $r^2=x^{1^2}+x^{2^2}+x^{3^2}$. The nonzero Christoffel symbols are (no summation over repeated indices)

$$\Gamma^0_{\alpha\alpha}=\dot R R/(1+\tfrac14 kr^2)^2\,,\quad \Gamma^\alpha_{\alpha0}=\Gamma^\alpha_{0\alpha}=\dot R/R\,,$$
$$\Gamma^\alpha_{\beta\beta}=kx^\alpha/2(1+\tfrac14 kr^2)\,, \tag{5.26}$$
$$\Gamma^\beta_{\beta\alpha}=\Gamma^\beta_{\alpha\beta}=-kx^\alpha/2(1+\tfrac14 kr^2)\,,$$

where α and β take the values 1,2,3. Also

$$u^a=\delta^a_0\,,\quad \dot u^a=0\,,\quad \sigma_{ab}=0\,,\quad \omega^a=0\,,\quad \theta=\theta(x^0)\,. \tag{5.27}$$

The metric (5.25) admits six spacelike Killing vectors[46]:

$$\underset{(1)}{\xi}{}^a=(1+\tfrac14 kx^{1^2}-\tfrac14 kx^{2^2}-\tfrac14 kx^{3^2})\delta^a_1$$
$$+\tfrac12 kx^1x^2\delta^a_2+\tfrac12 kx^1x^3\delta^a_3\,, \tag{5.28}$$
$$\underset{(4)}{\xi}{}^a=x^2\delta^a_1-x^1\delta^a_2\,; \tag{5.29}$$

$\underset{(2)}{\xi}{}^a$, $\underset{(3)}{\xi}{}^a$ and $\underset{(5)}{\xi}{}^a$, $\underset{(6)}{\xi}{}^a$ are obtained from (5.28) and (5.29), respectively, by cyclically permuting the indices 1,2,3. Since $\underset{(i)}{\xi}{}_0=0$ for $i=1,2,...,6$ we have

$$\underset{(i)}{\xi}{}_a u^a=0\,,\quad i=1,2,...,6\,. \tag{5.30}$$

Hence, by Theorem 5.2 (i), since $\omega=0$ and $\underset{(i)}{\xi}{}^a$ is a spacelike Killing vector satisfying $\underset{(i)}{\xi}{}_a u^a=0$, it follows that the integral curves of $\underset{(i)}{\xi}{}^a$ must be material curves in the fluid, $i=1,2,...,6$.

We now check that this is indeed the case by showing by a direct calculation that

$$h^a_{\ b}\dot n^b-\overset{*}{u}{}^a+(n_b\overset{*}{u}{}^b)n^a=0\,, \tag{5.31}$$

where n^a is the unit spacelike vector parallel to any one of the six Killing vectors ξ^a:

$$n^a=\xi^a/(\xi_b\xi^b)^{1/2}\,. \tag{5.32}$$

We find from (5.25), (5.28), and (5.29) that

$$(\xi_a\xi^a)^{1/2}=R(x^0)f(x^\alpha) \tag{5.33}$$

for some $f(x^\alpha)$, and it is easily verified with the aid of (5.26) that

$$\dot n^a=0\,,\quad \overset{*}{u}{}^a=(\dot R/R)n^a\,. \tag{5.34}$$

It follows immediately from (5.34) that (5.31) is satisfied.

2. Gödel model

The line element is

$$ds^2=-dx^{0^2}-2e^{\alpha x^1}dx^0dx^2+dx^{1^2}-\tfrac12 e^{2\alpha x^1}dx^{2^2}+dx^{3^2}\,, \tag{5.35}$$

where α is a nonzero constant. The nonzero Christoffel symbols are

$$\Gamma^0_{01}=\Gamma^0_{10}=\alpha\,,\quad \Gamma^0_{12}=\Gamma^0_{21}=(\alpha/2)e^{\alpha x^1}\,, \tag{5.36}$$
$$\Gamma^1_{02}=\Gamma^1_{20}=(\alpha/2)e^{\alpha x^1}\,,\quad \Gamma^1_{22}=(\alpha/2)e^{2\alpha x^1}\,, \tag{5.37}$$
$$\Gamma^2_{01}=\Gamma^2_{10}=-\alpha e^{-\alpha x^1}\,, \tag{5.38}$$

and

$$u^a=\delta^a_0\,,\quad \dot u^a=0\,,\quad \sigma_{ab}=0\,,\quad \theta=0\,,$$
$$\omega^a=(\alpha/\sqrt2)\delta^a_3\,,\quad \omega_{a;b}=0\,. \tag{5.39}$$

The metric (5.35) admits the five Killing vectors

$$\underset{(0)}{\xi}{}^a=\delta^a_0\,, \tag{5.40}$$

$$\underset{(1)}{\xi}{}^a=\delta^a_1-\alpha x^2\delta^a_2\,, \tag{5.41}$$

$$\underset{(2)}{\xi}{}^a=\delta^a_2\,, \tag{5.42}$$

$$\underset{(3)}{\xi}{}^a=\delta^a_3\,, \tag{5.43}$$

$$\underset{(4)}{\xi}{}^a=-2e^{-\alpha x^1}\delta^a_0+\alpha x^2\delta^a_1+(e^{-2\alpha x^1}-\tfrac12(\alpha x^2)^2)\delta^a_2\,. \tag{5.44}$$

The $\underset{(0)}{\xi}{}^a$, $\underset{(2)}{\xi}{}^a$, and $\underset{(4)}{\xi}{}^a$ are timelike. The $\underset{(1)}{\xi}{}^a$ is timelike, null, or spacelike depending on the values of α, x^1, and x^2; also

$$\underset{(1)}{\xi}{}_a u^a = \alpha x^2 e^{\alpha x^1} . \tag{5.45}$$

The $\underset{(3)}{\xi}{}^a$ is spacelike and satisfies $\underset{(3)}{\xi}{}_a u^a = 0$.

Let $n^a = \delta^a_3 = \underset{(3)}{\xi}{}^a$. Then the unit vector n^a is a spacelike Killing vector satisfying $n_a u^a = 0$, and we also have by (5.39) that $\omega^a = (\alpha/\sqrt{2})n^a$. From Theorem 5.2 (ii), we have the following.

(a) Since the unit tangent vector n^a to the vortex lines is a spacelike Killing vector, the vortex lines must be material lines in the fluid. Clearly, in the comoving coordinate system used, this is indeed the case [and can be verified by direct calculation by showing that $\dot{n}^a = \overset{*}{u}{}^a = 0$, since $\Gamma^a_{30} = 0$, which implies that (5.31) is satisfied].

(b) By first noting that δ^a_3 is the unit tangent vector to a material curve and also a spacelike Killing vector, we can deduce that the local vorticity vector ω^a must be in the direction δ^a_3. From (5.45), the only other spacelike Killing vector $\underset{(1)}{\xi}{}^a$ (for certain values of α, x^1, and x^2) does not satisfy $\underset{(1)}{\xi}{}_a u^a = 0$ everywhere, as required by Theorem 5.2.

The Gödel model can also be used to illustrate the results of Sec. IV on rigid and isometric spacelike congruences. Since $\omega_{a;b} = 0$ by (5.39), it follows from (2.17) and (2.18) that for the vortex congruence, $\mathscr{E} = 0$ and $\mathscr{S}_{ab} = 0$. [From (2.16), $\mathscr{R}_{ab}$ also vanishes.] Hence by (4.6), $\underset{n}{\mathscr{L}} p_{ab} = 0$ and therefore the vortex congruence in the Gödel universe is rigid, which was verified by Ciubotariu[6] by calculating $\underset{n}{\mathscr{L}} p_{ab}$ directly. Since in the Gödel model, ω^a/ω is a Killing vector, the vortex congruence is also isometric.

B. Conformal Killing vectors parallel to the vorticity vector

Theorem 5.1 can be used alternatively to aid in the construction of spacelike conformal Killing vectors and to test if a given spacelike vector is a conformal Killing vector of space-time. We will illustrate this by considering conformal Killing vectors parallel to the vorticity vector ω^a. The fourth condition (5.6) of Theorem 5.1 depends on ω^a, which clearly plays a special role in the theory of spacelike conformal motions. In order to apply Theorem 5.1 expressions for the expansion and shear of the corresponding spacelike congruence, as measured by a comoving observer, are required. For the vortex congruence, these quantities are known. The expansion $\mathscr{E}(\omega)$, shear $\mathscr{S}_{ab}(\omega)$, and rotation $\mathscr{R}_{ab}(\omega)$ (the rotation will also be required) of a vortex congruence as measured by an observer with four-velocity u^a are given by[15]

$$\mathscr{E}(\omega) = (1/\omega)(2\omega n_c \dot{u}^c - n^b \sigma_{br;s}(\omega^{rs}/\omega) - H_{ab} n^a n^b) , \tag{5.46}$$

$$\mathscr{S}_{ab}(\omega) = (1/\omega)\, p^c_a\, p^d_b (\eta_{(c}{}^{rst}\sigma_{d)r;s} u_t + H_{cd}) + (1/2\omega)(n^t \sigma_{tc;d}(\omega^{cd}/\omega) + H_{cd} n^c n^d)\, p_{ab} , \tag{5.47}$$

$$\mathscr{R}_{ab}(\omega) = (1/\omega)\, p^c_a\, p^d_b\, \eta_{[c}{}^{rst}\sigma_{d]r;s} u_t + (1/2\omega)\, p^c_a\, p^d_b \eta_{cd}{}^{st}(\tfrac{2}{3}\theta_{,s} - q_s)u_t , \tag{5.48}$$

where $n^a = \omega^a/\omega$, q^a in the energy flux relative to u^a, and H_{ab} is the magnetic part of the Weyl tensor with respect to the fluid flow. Equations (5.46)–(5.48) can be derived from first principles by starting from the Ricci identity for u^a, expressing the Riemann curvature tensor in terms of the Weyl tensor and the Ricci tensor, and using the Einstein field equations; it can be verified that this gives[15]

$$\omega_{a;b} = g_{ab}(\omega_c \dot{u}^c) + H_{ab} + \tfrac{1}{2}\eta_{ab}{}^{st}(\tfrac{2}{3}\theta_{,s} - q_s)u_t + \eta_a{}^{rst}\sigma_{br;s} u_t - \omega_a(\theta u_b + 2\dot{u}_b) + u_a(\omega^c u_{c;b}) + u_b(u_a \omega^c{}_{;c} + u_{a;c}\omega^c - \dot{\omega}_a) . \tag{5.49}$$

By using definitions (2.16)–(2.18), where p^{ab} is given by (2.4) with $w^a = u^a$ and $n^a = \omega^a/\omega$, (5.46)–(5.48) may be obtained. Equation (5.49) for $\omega_{a;b}$ will be required later to evaluate $\overset{*}{n}{}^a$ (see Ref. 47).

As an indication of the kind of result that may be obtained, we note that Theorem 3.1 has been used to show that certain fluid space-times admit *timelike* conformal motions.[3] For instance, in a shear-free perfect fluid with barotropic equation of state $p = p(\mu)$, $\mu + p \neq 0$ (p is the isotropic pressure and μ is the total energy density of the fluid as measured by u^a), if $(p - \frac{1}{3}\mu)^{\cdot} = 0$, then $(1/r)u^a$ is a timelike conformal Killing vector, where[48]

$$r = \exp\left(\int_{p_0}^{p} \frac{dp}{\mu + p}\right) . \tag{5.50}$$

Also, in the FRW models, $R(x^0)u^a$ is a timelike conformal Killing vector[3] for any equation of state, where $R(x^0)$ is the "scale factor" defined by (5.25); in the FRW models, when $p = p(\mu)$ and $(p - \frac{1}{3}\mu)^{\cdot} = 0$, it can be verified that $R(x^0) = \alpha/r(x^0)$, where α is a constant. We will see that similar results can be established for spacelike conformal Killing vectors if u^a is replaced by ω^a/ω.

Equations (5.46) and (5.47) show that the expansion $\mathscr{E}(\omega)$ and shear $\mathscr{S}_{ab}(\omega)$ of a vortex congruence depend explicitly on σ_{ab} and H_{ab}. The simplest case to consider therefore is shear-free flow in fluid space-times with vanishing magnetic Weyl tensor. This has been considered recently, for perfect fluid space-times, by Collins.[22] We will give an alternative derivation, based on Theorem 5.1, of some of the results of Collins on conformal Killing vectors parallel to ω^a. These results can be summarized briefly as follows. Consider a shear-free perfect fluid with equation of state $p = p(\mu)$ satisfying $\mu + p \neq 0$ and such that $H_{ab} = 0$. Collins demonstrated that rotating shear-free solutions with $H_{ab} = 0$ exist irrespective of the relative orientation of ω^a and $\dot{u}^a$. (He also proved that for such solutions, since $\omega^a \neq 0$, the rate of expansion θ must vanish.) If ω^a is orthogaonal to $\dot{u}^a$, or if $\dot{u}^a = 0$, then ω^a/ω is a Killing vector of space-time and $p = \mu + 2\Lambda$, where Λ is the cosmological constant, while if $\omega_a \dot{u}^a \neq 0$, then space-time admits a proper conformal Killing vector parallel to ω^a; this conformal Killing vector is homothetic if and only if $p = \mu + 2\Lambda$. (The relation between these results and those of McIntosh[49,50] on homothetic motions was discussed by Collins and will be commented on later in this section.)

In the following theorem we first establish the properties of the vortex congruence needed to apply Theorem 5.1.

Theorem 5.3: Consider a rotating shear-free perfect fluid with $H_{ab}=0$ and equation of state $p=p(\mu)$, $\mu+p\neq 0$. Einstein's field equations are assumed to be satisfied.

(a) Irrespective of the relative orientation of ω^a ands $\dot{u}^a$,

$$\mathscr{E}(\omega)=2n_a\dot{u}^a\,, \tag{5.51}$$

$$\mathscr{S}_{ab}(\omega)=0\,, \tag{5.52}$$

$$\mathscr{R}_{ab}(\omega)=0\,, \tag{5.53}$$

$$\overset{*}{n}{}^a=0\,, \tag{5.54}$$

where $n^a=\omega^a/\omega$ and $\mathscr{E}(\omega)$, $\mathscr{S}_{ab}(\omega)$, and $\mathscr{R}_{ab}(\omega)$ are the expansion, shear, and rotation of the vortex congruence as measured by a comoving observer with four-velocity u^a.

(b) $\omega_a\dot{u}^a\neq 0$.

(i) There exists a proper conformal Killing vector parallel to ω^a irrespective of the relative orientation of ω^a and $\dot{u}^a$ (excluding ω^a orthogonal to $\dot{u}^a$).

(ii) If ω^a and $\dot{u}^a$ are parallel, then $(1/r)(\omega^a/\omega)$ is a proper conformal Killing vector where r is defined by (5.50).

(iii) The conformal Killing vector is homothetic if and only if $p=\mu+2\Lambda$.

(c) $\omega_a\dot{u}^a=0$. If either ω^a is orthogonal to $\dot{u}^a$ or $\dot{u}^a=0$, then ω^a/ω is a Killing vector and $p=\mu+2\Lambda$.

Proof: (a) When $\sigma_{ab}=0$ and $H_{ab}=0$, (5.46) and (5.47) reduce to $\mathscr{E}(\omega)=2n_a\dot{u}^a$ and $\mathscr{S}_{ab}(\omega)=0$, respectively. Now, under the stated assumptions, when $\omega^a\neq 0$, the rate of expansion θ must vanish.[22] Since also $q^a=0$ for a perfect fluid, (5.48) reduces to $\mathscr{R}_{ab}(\omega)=0$. Further, with $n^a=\omega^a/\omega$ we have

$$\overset{*}{n}{}^a=\overset{*}{\omega}{}^a/\omega-(\overset{*}{\omega}/\omega)n^a\,. \tag{5.55}$$

But on setting $H_{ab}=0$, $\sigma_{ab}=0$, $q^a=0$, and $\theta=0$ (see Ref. 22) in (5.49), it follows that

$$\omega_{a;b}=g_{ab}(\omega_c\dot{u}^c)-2\omega_a\dot{u}_b+u_a(\omega^c u_{c;b})$$
$$+u_b(u_a\omega^c{}_{;c}+u_{a;c}\omega^c-\dot{\omega}_a)\,, \tag{5.56}$$

and hence that

$$\overset{*}{\omega}{}^a=-\omega(n_b\dot{u}^b)n^a\,, \tag{5.57}$$

$$\overset{*}{\omega}=-\omega(n_b\dot{u}^b)\,. \tag{5.58}$$

On substituting (5.57) and (5.58) into (5.55), we obtain $\overset{*}{n}{}^a=0$.

Unlike (5.51) and (5.52), (5.53) and (5.54) depend on Einstein's field equations.[47]

(b) $\omega_a\dot{u}^a\neq 0$. (i) We check that the necessary and sufficient conditions, (5.3) to (5.6), of Theorem 5.1 are satisfied. Throughout $n^a=\omega^a/\omega$.

Since by (5.51) and (5.52), $\mathscr{E}=2n_a\dot{u}^a$ and $\mathscr{S}_{ab}=0$, (5.5) and (5.3) are satisfied. Further, for a perfect fluid with equation of state $p=p(\mu)$ satisfying $\mu+p\neq 0$, it follows from the momentum conservation equation that r defined by (5.50) is an acceleration potential[24]:

$$\dot{u}_a=-h_a^{\ b}(\log r)_{,b}\,, \tag{5.59}$$

and hence that the vorticity propagation equation[23,24]

$$h^{\ a}_{b}\dot{\omega}^b=u^a{}_{;b}\omega^b-\theta\omega^a+\tfrac{1}{2}\eta^{abcd}u_b\dot{u}_{c;d} \tag{5.60}$$

may be rewritten as

$$h^{\ a}_{b}\dot{n}^b=\overset{*}{u}{}^a+(u_b\overset{*}{n}{}^b)n^a\,. \tag{5.61}$$

Condition (5.6) with $n^a=\omega^a/\omega$ is therefore satisfied. [From (5.61), the vortex lines are material lines and all observers employed along the congruence will be comoving if a comoving observer is employed at some initial point.]

It remains to consider condition (5.4), which, since $\overset{*}{n}{}^a=0$, reduces to proving that there always exists a solution ξ to the equation

$$(\log\xi)_{,a}=\tfrac{1}{2}\mathscr{E}n_a\,. \tag{5.62}$$

To verify this, we define $A_a=\frac{1}{2}\mathscr{E}n_a$ and we show that $A_{[a;b]}=0$; we have

$$A_{[a;b]}=\tfrac{1}{2}\mathscr{E}n_{[a;b]}+\tfrac{1}{2}n_{[a}\mathscr{E}_{,b]}\,. \tag{5.63}$$

Consider first $n_{a;b}$ evaluated for a comoving observer $w^a=u^a$. Since by (5.52) and (5.53), $\mathscr{S}_{ab}=0=\mathscr{R}_{ab}$ for a comoving observer, and $\overset{*}{n}{}^a=0$ by (5.54), (2.23) reduces to

$$n_{a;b}=\tfrac{1}{2}\mathscr{E}p_{ab}-\dot{n}_a u_b-u_a(u^t n_{t;b})-u_a u_b(u_t\dot{n}^t)\,. \tag{5.64}$$

Now,

$$\overset{*}{u}_a=u_{a;b}n^b=(\sigma_{ab}+(\theta/3)h_{ab}+\omega_{ab}-\dot{u}_a u_b)(\omega^b/\omega)=0\,, \tag{5.65}$$

and since also $\overset{*}{n}{}^a=0$ and $u_a\dot{n}^a=-\frac{1}{2}\mathscr{E}$, the propagation equation (5.61) reduces to

$$\dot{n}^a=\tfrac{1}{2}\mathscr{E}u^a\,. \tag{5.66}$$

Further, $u^t n_{t;b}=-n^t u_{t;b}=(n^t\dot{u}_t)u_b$ and therefore (5.64) becomes

$$n_{a;b}=\tfrac{1}{2}\mathscr{E}(g_{ab}-n_a n_b)\,; \tag{5.67}$$

hence $n_{[a;b]}=0$. To evaluate (5.63), $\mathscr{E}_{,a}$ is also required. Consider the Ricci identity for n^a,

$$n_{a;bc}-n_{a;cb}=R_{tabc}n^t\,, \tag{5.68}$$

and raise index a and contract with c; this gives[51]

$$n^c{}_{;bc}-n^c{}_{;cb}=R_{tb}n^t\,. \tag{5.69}$$

But it follows directly from (2.17) with $w^a=u^a$ and (5.51) that

$$n^c{}_{;c}=\tfrac{3}{2}\mathscr{E}\,, \tag{5.70}$$

and using (5.67) for $n^c{}_{;b}$ we also have, with the aid of (5.54) and (5.70),

$$n^c{}_{;bc}=\tfrac{1}{2}\mathscr{E}_{,b}-\tfrac{1}{2}(\overset{*}{\mathscr{E}}+\tfrac{3}{2}\mathscr{E}^2)n_b\,. \tag{5.71}$$

Equation (5.69) therefore becomes

$$\mathscr{E}_{,b}+\tfrac{1}{2}(\overset{*}{\mathscr{E}}+\tfrac{3}{2}\mathscr{E}^2)n_b=-R_{tb}n^t\,. \tag{5.72}$$

Also, applying Einstein's field equations[23,24]

$$R_{ab} = T_{ab} - \tfrac{1}{2}Tg_{ab} + \Lambda g_{ab}, \quad (5.73)$$

where for a perfect fluid

$$T_{ab} = \mu u_a u_b + ph_{ab}, \quad (5.74)$$

we find that

$$R_{tb}n^t = \tfrac{1}{2}(\mu - p + 2\Lambda)n_b. \quad (5.75)$$

On substituting from (5.75) into (5.72) we obtain

$$\mathscr{E}_{,b} = -\tfrac{1}{2}(\overset{*}{\mathscr{E}} + \tfrac{3}{2}\mathscr{E}^2)n_b + \tfrac{1}{2}(p - \mu - 2\Lambda)n_b. \quad (5.76)$$

Thus $n_{[a}\mathscr{E}_{,b]} = 0$ and since also $n_{[a;b]} = 0$, it follows from (5.63) that $A_{[a;b]} = 0$. Hence there must exist a function ϕ such that $A_a = \phi_{,a} = (\log \xi)_{,a}$, where $\xi = e^\phi$. Since $A_a = \frac{1}{2}\mathscr{E} n_a$ a solution to Eq. (5.62) for ξ therefore always exists, and condition (5.4) is satisfied.

The conformal Killing vector is a proper conformal Killing vector since by (5.7), $\psi = \xi\mathscr{E}/2$ and $\mathscr{E} = (2/\omega)\omega_a \dot{u}^a \neq 0$ for the case under consideration.

(ii) When ω^a is parallel to $\dot{u}^a$, the solution of (5.62) for ξ assumes a simple form. For suppose that

$$\dot{u}_a = \alpha n_a \quad (5.77)$$

for some scalar α. Then contracting (5.77) with n^a gives $\alpha = n^a \dot{u}_a = \frac{1}{2}\mathscr{E}$ by (5.51) and therefore from (5.77), $\dot{u}_a = \frac{1}{2}\mathscr{E} n_a$. Equation (5.62) for ξ therefore becomes

$$(\log \xi)_{,a} = \dot{u}_a. \quad (5.78)$$

But by (5.59),

$$\dot{u}^a = -(\log r)_{,a} - u_a (\log r)^{\cdot}; \quad (5.79)$$

we show that $(\log r)^{\cdot} = 0$. From (5.50),

$$(\log r)^{\cdot} = \frac{\dot{\mu}}{\mu + p}\frac{dp}{d\mu}, \quad (5.80)$$

and from the energy conservation equation for a perfect fluid, we have

$$\dot{\mu} + (\mu + p)\theta = 0. \quad (5.81)$$

Since $\theta = 0$ (see Ref. 22), it follows that $\dot{\mu} = 0$ and therefore by (5.80) that $(\log r)^{\cdot} = 0$. On substituting (5.79) with $(\log r)^{\cdot} = 0$ into (5.78), we obtain

$$(\log \xi r)_{,a} = 0, \quad (5.82)$$

and therefore $\xi = \beta/r$, where β is a constant. Thus $(1/r)(\omega^a/\omega)$ is a conformal Killing vector.

(iii) To determine when the conformal Killing vector is homothetic, consider $\psi_{,a}$. From (5.7), $\psi = \xi\mathscr{E}/2$ and hence with the aid of (5.62) we have

$$\psi_{,a} = (\xi/2)(\mathscr{E}_{,a} + \tfrac{1}{2}\mathscr{E}^2 n_a). \quad (5.83)$$

On substituting from (5.76) for $\mathscr{E}_{,a}$ into (5.83), we obtain[52]

$$\psi_{,a} = -(\xi/4)(\overset{*}{\mathscr{E}} + \tfrac{1}{2}\mathscr{E}^2)n_a + (\xi/4)(p - \mu - 2\Lambda)n_a. \quad (5.84)$$

To evaluate $\overset{*}{\mathscr{E}} + \frac{1}{2}\mathscr{E}^2$ we contract (5.76) with n^b; this gives

$$\overset{*}{\mathscr{E}} + \tfrac{1}{2}\mathscr{E}^2 = \tfrac{1}{3}(p - \mu - 2\Lambda), \quad (5.85)$$

and on substituting (5.85) into (5.84) we obtain

$$\psi_{,a} = (\xi/6)(p - \mu - 2\Lambda)n_a. \quad (5.86)$$

Thus $\psi_{,a} = 0$, and the conformal Killing vector is homothetic, if and only if $p = \mu + 2\Lambda$. It is a proper homothetic Killing vector because $\psi = \xi\mathscr{E}/2 \neq 0$ when $\omega_a \dot{u}^a \neq 0$ by (5.51).

(c) $\omega_a \dot{u}^a = 0$. To show that ω^a/ω is a Killing vector we first note that when $\omega_a \dot{u}^a = 0$, $\mathscr{E} = 0$ by (5.51). Conditions (5.3) and (5.5) with $\mathscr{E} = 0$ and (5.6) are satisfied as in part (b). Condition (5.4) reduces to $(\log \xi)_{,a} = 0$, which requires that $\xi = \text{const}$. We also have $\psi = \xi\mathscr{E}/2 = 0$. Hence ω^a/ω is a Killing vector. Finally, since $\mathscr{E} = 0 = \psi$, we see from either (5.76), (5.85), or (5.86) that $p = \mu + 2\Lambda$. □

When $\dot{u}^a = 0$, we obtain the Gödel solution generalized to include pressure.[22] We have observed that ω^a/ω is a Killing vector in the Gödel model. The foregoing results also serve to illustrate Theorem 3.7 in which necessarily $\mathscr{E}\mathscr{R} = 0$ for the congruence generated by a spacelike conformal Killing vector when $\overset{*}{n}{}^a = 0$. Equation (5.85) is the propagation equation for the expansion $\mathscr{E}$ along the congruence in the problem considered above, and it played a central role in determining the necessary and sufficient condition for a spacelike conformal motion to be homothetic. The general expression for the propagation equation for $\mathscr{E}$ has been derived[15] and it may be useful when considering more complicated problems of this kind.

As noted by Collins,[22] the results of Theorem 5.3 agree with those of McIntosh[49,50] on homothetic motions, who showed that in a perfect fluid if there is a *proper* homothetic Killing vector ξ^a orthogonal to u^a then necessarily $p = \mu + 2\Lambda$. (McIntosh set $\Lambda = 0$ but did not restrict consideration to the special case $\xi_a u^a = 0$.) If in a perfect fluid, ξ^a is a conformal Killing vector satisfying $\xi_a u^a = 0$, and if we define $F_{ab} = \xi_{[a;b]}$ and $J^a = F^{ab}{}_{;b}$, then it can be verified that

$$\psi_{,a} = \tfrac{1}{6}(p - \mu - 2\Lambda)\xi_a - \tfrac{1}{3}J_a, \quad (5.87)$$

and hence $\psi_{,a} = 0$ if and only if

$$J^a = \tfrac{1}{2}(p - \mu - 2\Lambda)\xi^a. \quad (5.88)$$

Thus J^a is parallel to ξ^a or vanishes, and in either case McIntosh showed further that

$$\psi J^a = 0. \quad (5.89)$$

Hence if ξ^a is a *proper* homothetic Killing vector ($\xi_a u^a = 0$), then it is necessary that $J^a = 0$ and therefore by (5.88), necessary that $p = \mu + 2\Lambda$. In Theorem 5.3, $J^a = 0$ since $F_{ab} = 0$ (see Ref. 22). The condition $J^a = 0$ is not in general sufficient to insure that $\psi \neq 0$; since $\psi = \xi\mathscr{E}/2$ we require $\mathscr{E} \neq 0$, i.e., by (5.5), $\xi_a \dot{u}^a \neq 0$.

The explicit dependence of $\mathscr{E}(\omega)$ and $\mathscr{S}_{ab}(\omega)$ on σ_{ab} and H_{ab} shows that σ_{ab} and H_{ab} will in general play an important role in determining if a fluid space-time admits a conformal Killing vector parallel to ω^a. Condition (5.5) with (5.46) and condition (5.3) with (5.47) place the following restrictions on σ_{ab} and H_{ab} in any fluid space-time that admits a conformal Killing vector parallel to ω^a:

$$\omega^a \sigma_{ab;c}\omega^{bc} + H_{ab}\omega^a\omega^b = 0, \quad (5.90)$$

$$p_a^c p_b^d (\eta_{(c}{}^{rst}\sigma_{d)r;s}u_t + H_{cd}) = 0, \quad (5.91)$$

where $p^{ab} = g^{ab} + u^a u^b - (\omega^a/\omega)(\omega^b/\omega)$. Both conditions are identically satisfied for the simple case considered in this subsection in which $\sigma_{ab} = 0$ and $H_{ab} = 0$. We see that the symmetry property of vanishing shear $\mathscr{S}_{ab}$ of the *spacelike* vortex congruence imposes a condition on the shear σ_{ab} of the *timelike* congruence of fluid particle world lines. If further $\overset{*}{n}{}^a = 0$, then it follows from Theorem 3.7 that for a *proper* conformal Killing vector parallel to ω^a, it is necessary that $\mathscr{R}_{ab}(\omega) = 0$, and therefore from (5.48) that

$$2\, p^c_a\, p^d_b \eta_{[c}{}^{rst}\sigma_{d]r;s}u_t + p^c_a\, p^d_b \eta_{cd}{}^{st}(\tfrac{2}{3}\theta_{,s} - q_s)u_t = 0\,. \quad (5.92)$$

There is no Newtonian analog of H_{ab}. For a vortex congruence in three-dimensional Euclidean space we have, corresponding to (5.46) to (5.48) (see Ref. 15),

$$\mathscr{C}(\omega) = -(1/\omega^3)\omega^{\beta}\sigma_{\beta\lambda,\tau}\omega^{\lambda\tau}\,, \quad (5.93)$$

$$\mathscr{S}_{\alpha\beta}(\omega) = (1/\omega)\, p^{\lambda}_{\alpha}\, p^{\mu}_{\beta}\eta_{(\lambda}{}^{\gamma\tau}\sigma_{\mu)\gamma,\tau} + (1/2\omega^3)(\omega^{\tau}\sigma_{\tau\nu,\lambda}\omega^{\nu\lambda})\, p_{\alpha\beta}\,, \quad (5.94)$$

$$\mathscr{R}_{\alpha\beta}(\omega) = (1/\omega)\, p^{\nu}_{\alpha}\, p^{\tau}_{\beta}\eta_{[\nu}{}^{\gamma\mu}\sigma_{\tau]\gamma,\mu} + (1/3\omega)\, p^{\nu}_{\alpha}\, p^{\tau}_{\beta}\eta_{\nu\tau}{}^{\mu}\theta_{,\mu}\,, \quad (5.95)$$

which can be derived by first establishing the following identity which corresponds to (5.49) (see Ref. 15):

$$\omega_{\alpha,\beta} = \tfrac{1}{3}\eta_{\alpha\beta}{}^{\tau}\theta_{,\tau} + \eta_{\alpha}{}^{\lambda\tau}\sigma_{\beta\lambda,\tau}\,. \quad (5.96)$$

We have, in place of Theorem 5.3, the following theorem.

Theorem 5.4: In three-dimensional Euclidean space, if the fluid is shear-free, i.e., if $\sigma_{\alpha\beta} = 0$, then ω^{α} is a Killing vector.

Proof: It follows immediately from (5.96) that if $\sigma_{\alpha\beta} = 0$ then $\omega_{(\alpha,\beta)} = 0$. Alternatively, the result can be established by considering conditions (3.30) and (3.31) of Theorem 3.4. When $\sigma_{\alpha\beta} = 0$, it follows from (5.93), (5.94), and (5.96) that $\mathscr{C} = 0$, $\mathscr{S}_{\alpha\beta} = 0$, and $\overset{*}{n}_{\alpha} = -(\log \omega)_{,\alpha}$; (3.30) is therefore satisfied and (3.31) reduces to

$$(\log(\xi/\omega))_{,\alpha} = 0\,, \quad (5.97)$$

which requires that $\xi = \beta\omega$, where β is a constant. Since also $\psi = \xi\mathscr{C}/2 = 0$, it follows that ω^{α} is a Killing vector. □

The foregoing result did not require $\theta = 0$. If further $\theta = 0$ then the flow is rigid and ω_{α} is a constant by (5.96). Killing's equation $\omega_{(\alpha,\beta)}$ is trivially satisfied.

Condition (3.30) with (5.94) places the following restriction on the shear of a fluid which admits a conformal Killing vector parallel to ω^{α}:

$$2\, p^{\lambda}_{\alpha}\, p^{\mu}_{\beta}\eta_{(\lambda}{}^{\gamma\tau}\sigma_{\mu)\gamma,\tau} + (1/\omega^2)(\omega^{\tau}\sigma_{\tau\nu,\lambda}\omega^{\nu\lambda})\, p_{\alpha\beta} = 0\,. \quad (5.98)$$

Unlike (5.91), to which it corresponds, (5.98) is identically satisfied in any shear-free fluid. Equation (5.90) has no analog in Newtonian theory because it is derived from the purely relativistic condition (5.5).

VI. MATERIAL CURVES IN ROTATIONAL FLUIDS

We have seen that vortex lines which are material curves play an important role in the theory of spacelike conformal motions. We conclude by examining the restrictions imposed on the flow of a thermodynamical perfect fluid by the requirement that the vortex lines are "frozen-in" to the fluid, and then to illustrate Theorem 5.2 (ii) we consider a rotational electrically conducting fluid with vanishing electric field; in such a fluid Maxwell's equations require the magnetic field lines to be material curves.[24]

A. Rotational thermodynamical perfect fluid

To investigate the restriction imposed on the flow by the requirement that vortex lines be material curves, we consider again the vorticity propagation equation[23,24]

$$h^{\,a}_{\,b}\dot{\omega}^b = u^a{}_{;b}\omega^b - \theta\omega^a + \tfrac{1}{2}\eta^{abcd}u_b\dot{u}_{c;d}\,. \quad (6.1)$$

On defining $n^a = \omega^a/\omega$, it follows that

$$h^{\,a}_{\,b}\dot{n}^b = \overset{*}{u}{}^a - (n_t\overset{*}{u}{}^t)n^a + (1/2\omega)\, p^a_t\eta^{tbcd}u_b\dot{u}_{c;d}\,, \quad (6.2)$$

where

$$p^a_t = g^a_t + u^a u_t - n^a n_t\,. \quad (6.3)$$

But from Appendix A, the necessary and sufficient condition for n^a ($n_a u^a = 0$) to be the unit tangent vector field to a material curve is that it satisfy the propagation equation

$$h^{\,a}_{\,b}\dot{n}^b = \overset{*}{u}{}^a - (n_t\overset{*}{u}{}^t)n^a\,; \quad (6.4)$$

hence vortex lines are material lines in the fluid if and only if

$$p^a_t\eta^{tbcd}u_b\dot{u}_{c;d} = 0\,. \quad (6.5)$$

The kinematical condition (6.5) applies to all fluids. Dynamics is introduced through the momentum conservation equation for $\dot{u}_c$.

We restrict consideration to a thermodynamical perfect fluid.[3,20,25] We assume that there exists an equation of state $e = e(p, 1/\rho)$, where e is the specific internal energy density and ρ is the particle rest-mass density, measured by u^a; we have $\mu = \rho(1+e)$, where μ is the total energy density measured by u^a. The temperature $T(p,1/\rho)$ and the specific entropy $S(p,1/\rho)$ can then be defined by the Gibbs equation

$$T\,dS = de + p\,d(1/\rho)\,, \quad (6.6)$$

which may be rewritten as

$$dp = \rho\,df - \rho T\,dS\,, \quad (6.7)$$

where $f = 1 + e + p/\rho = (\mu + p)/\rho$ is the index of the fluid.[25] [The remaining assumption that rest mass is conserved, which implies $(\rho u^a)_{;a} = 0$, will not be required.] With the aid of (6.7) the momentum conservation equation for a perfect fluid,

$$\dot{u}_a = -[1/(\mu + p)]h^{\,b}_{\,a}p_{,b}\,, \quad (6.8)$$

may be rewritten as

$$\dot{u}_a = -h^{\,b}_{\,a}(\log f)_{,b} + (T/f)h^{\,b}_{\,a}S_{,b}\,. \quad (6.9)$$

Thus

$$\eta^{tbcd}u_b\dot{u}_{c;d} = 2((T/f)\dot{S} - (\log f)^{\dot{}}\,)\omega^t + \eta^{tbcd}u_bS_{,c}(T/f)_{,d} \quad (6.10)$$

and since $n^t = \omega^t/\omega$ we have

$$p^a_t\eta^{tbcd}u_b\dot{u}_{c;d} = p^a_t\eta^{tbcd}u_bh^{\,p}_{\,c}S_{,p}h^{\,q}_{\,d}(T/f)_{,q}\,. \quad (6.11)$$

Thus by (6.5), vortex lines are material lines in a thermodynamical perfect fluid if and only if

$$p_t^a \eta^{tbcd} u_b h_c^{\ p} S_{,p} h_d^{\ q} (T/f)_{,q} = 0 . \tag{6.12}$$

There are three cases to consider.

Case (i). $h_a^{\ b} S_{,b} = 0$ *and/or* $h_a^{\ b}(T/f)_{,b} = 0$: Condition (6.12) is clearly satisfied. Further, there always exists an acceleration potential. For by (6.9), if $h_a^{\ b} S_{,b} = 0$ then f is an acceleration potential, while if $h_a^{\ b}(T/f)_{,b} = 0$ then $f \exp(-TS/f)$ is an acceleration potential. It is well established[24] that the existence of an acceleration potential is a sufficient, but not necessary, condition for vortex lines to be "frozen-in" to the fluid.

Case (ii). $h_a^{\ b} S_{,b} \neq 0$ *and* $h_a^{\ b}(T/f)_{,b} \neq 0$ *but parallel:* Condition (6.12) is again satisfied.

Case (iii). $h_a^{\ b} S_{,b} \neq 0$ *and* $h_a^{\ b}(T/f)_{,b} \neq 0$ *and not parallel:* Condition (6.12) is satisfied if and only if

$$S_{,a} n^a = 0 \text{ and } (T/f)_{,a} n^a = 0 , \quad \text{where } n^a = \omega^a/\omega . \tag{6.13}$$

To establish this result, suppose first that (6.12) is satisfied; then

$$\eta^{tbcd} u_b h_c^{\ p} S_{,p} h_d^{\ q} (T/f)_{,q} = \alpha n^t , \tag{6.14}$$

for some scalar $\alpha \neq 0$ and $n^t = \omega^t/\omega$. On contracting (6.14) in turn with $h_t^{\ r} S_{,r}$ and $h_t^{\ r}(T/f)_{,r}$, (6.13) is derived. Conversely, suppose that (6.13) is satisfied. At any one point expand $h_a^{\ b} S_{,b}$ and $h_a^{\ b}(T/f)_{,b}$ in terms of a local orthonormal tetrad consisting of u^a, n^a, r^a, and s^a, where r^a and s^a are mutually orthogonal unit vectors lying on the "screen" erected by observer u^a at the given point. By (6.13), $h_a^{\ b} S_{,b}$ and $h_a^{\ b}(T/f)_{,b}$ have components only in the directions r^a and s^a and therefore

$$\eta^{tbcd} u_b h_c^{\ p} S_{,p} h_d^{\ q} (T/f)_{,q} = \beta \eta^{tbcd} u_b r_c s_d , \tag{6.15}$$

for some scalar β. But $\eta^{tbcd} u_b r_c s_d$ is parallel to n^t and therefore contracting (6.15) with p_t^a gives (6.12).

In case (iii) the vortex lines are material lines if and only if S and T/f are constant along a given vortex line, i.e., if and only if the vortex lines lie along the intersection of the two-dimensional surfaces $S = \text{const}$ and $T/f = \text{const}$. If $h_a^{\ b} S_{,b}$ and $h_a^{\ b}(T/f)_{,b}$ are parallel [case (ii)], then these two-dimensional surfaces coincide. The vortex lines do not necessarily lie on these surfaces in this case. It is only in case (i) that an acceleration potential necessarily exists.

In Newtonian theory, the propagation equation for the vorticity vector ω^α in a self-gravitating fluid is[23]

$$\dot{\omega}^\alpha = v^\alpha_{\ ,\beta} \omega^\beta - \theta \omega^\alpha + \tfrac{1}{2} \eta^{\alpha\beta\gamma} a_{\beta,\gamma} , \tag{6.16}$$

where v^α denotes the velocity field of the fluid particles and a^α describes the combined effects of gravitational and inertial forces. It is found using (A10) of Appendix A that vortex lines are material lines if and only if

$$p_\tau^\alpha \eta^{\tau\beta\gamma} a_{\beta,\gamma} = 0 , \tag{6.17}$$

where $p_\tau^\alpha = h_\tau^{\ \alpha} - n^\alpha n_\tau$ and $n^\alpha = \omega^\alpha/\omega$. For a perfect fluid, the momentum conservation equation may be written as[23]

$$a_\alpha = -(1/\rho) p_{,\alpha} . \tag{6.18}$$

The Gibbs equation is valid in Newtonian theory, and corresponding to (6.12), the following necessary and sufficient condition for vortex lines to be material lines in a thermodynamical perfect fluid is obtained:

$$p_\tau^\alpha \eta^{\tau\beta\gamma} S_{,\beta} T_{,\gamma} = 0 . \tag{6.19}$$

If $S_{,\alpha}$ and $T_{,\alpha}$ are both nonzero and nonparallel, then vortex lines are "frozen-in" to a perfect fluid if and only if they lie along the intersection of the surfaces $S = \text{const}$ and $T = \text{const}$. The factor $1/f$ multiplying T in (6.12) is a relativistic effect due to the increase in the effective inertial-mass density from the Newtonian value ρ to $\mu + p$.

B. Magnetic field lines in a rotational fluid

Finally we discuss an example which illustrates Theorem 5.2 (ii). Consider a congruence of magnetic field lines in an electrically conducting fluid with nonzero vorticity. The electric and magnetic field four-vectors, E^a and H^a, measured by an observer with four-velocity u^a are defined by (5.2); both E^a and H^a are spacelike vectors since $E_a u^a = H_a u^a = 0$. We assume that (i) $E^a = 0$, (ii) $\omega^a \neq 0$, and (iii) space-time admits a spacelike conformal Killing vector field parallel to H^a. Physically the approximation of vanishing electric field can occur in a fluid for the idealized limit of infinite electric conductivity even when the anisotropy of the electric conductivity due to the magnetic field is taken into account.[53]

When $E_a = 0$, it is a direct consequence of Maxwell's equations that the magnetic field lines are material lines[24,54] and hence since $\omega^a \neq 0$ and space-time admits a spacelike conformal motion with symmetry vector parallel to H^a, it follows from Theorem 5.2 (ii) that the magnetic field lines must coincide with the vortex lines; thus

$$n^a = H^a/H = \omega^a/\omega . \tag{6.20}$$

This result was derived previously for the special case in which there exists a spacelike Killing vector field parallel to H^a (see Ref. 55). The property that the vortex and magnetic field lines coincide is a purely relativistic effect, the results of Theorem 5.2 having no direct analog in Newtonian theory.

Since the magnetic field lines are material lines, the vortex lines must also be "frozen-in" to the fluid. Unlike the magnetic field lines, which are material lines due to a physical property of the fluid ($E^a = 0$), the vortex lines are material lines due to a property of the flow. The flow must satisfy condition (6.5). If the fluid is a thermodynamical perfect fluid with uniform magnetic permeability λ and if the magnetic field is described by the Minkowski energy-momentum tensor with $E^a = 0$, then the momentum conservation equation may be expressed as[56]

$$\dot{u}_a = -h_a^{\ b}(\log f)_{,b} + \frac{T}{f} h_a^{\ b} S_{,b} + \frac{\lambda}{\rho f} \eta^{abcd} u_b \mathscr{J}_c H_d , \tag{6.21}$$

where $\mathscr{J}^a$ is the conduction current. ($\mathscr{J}^a = h_b^{\ a} J^b$, where J^b is the four-current density vector.) If $\mathscr{J}^a$ is parallel to H^a then $\eta^{abcd} u_b \mathscr{J}_c H_d = 0$ and the magnetic field is force-free; condition (6.5) reduces to (6.12) discussed in Sec. VI A. In general $\mathscr{J}^a$ will not be parallel to H^a and (6.5) with (6.21) will be complicated. We will not consider condition (6.5) further here, but it must be assumed that the flow is such that (6.5) is satisfied.

We now show that certain quantities are conserved along the magnetic field/vortex lines.

Theorem 6.1: If a fluid space-time admits a spacelike conformal motion with symmetry vector parallel to n^a, where

$$n^a = H^a/H = \omega^a/\omega\,, \tag{6.22}$$

and if $E^a = 0$, then

$$\text{(i)}\quad \frac{D}{dp}\left(\frac{\mathscr{R}^2}{H}\right) = 0\,, \tag{6.23}$$

$$\text{(ii)}\quad \frac{D}{dp}\left(\frac{\mathscr{R}}{\omega}\right) = 0\,, \tag{6.24}$$

$$\text{(iii)}\quad \frac{D}{dp}\left(\frac{\omega^2}{H}\right) = 0\,, \tag{6.25}$$

where p denotes arc length measured along a magnetic field/vortex line and $\mathscr{R}$ is the magnitude of the rotation of the spacelike congruence generated by n^a as measured by a comoving observer with four-velocity u^a.

Proof: (i) From definition (2.17),

$$\mathscr{E} = p^{ab} n_{a;b}\,, \tag{6.26}$$

where for a comoving observer, $p^{ab} = g^{ab} + u^a u^b - n^a n^b$. With $n^a = H^a/H$, (6.26) becomes

$$\mathscr{E} = (1/H) h^{ab} H_{a;b} - \overset{*}{H}/H\,. \tag{6.27}$$

But from Maxwell's equations, since $E^a = 0$ (see Ref. 24),

$$H^a{}_{;b} h^b_a = 0\,. \tag{6.28}$$

Using also (4.34) for $\mathscr{E}$, (6.27) assumes the form

$$\frac{D}{dp}(H\delta A) = 0\,. \tag{6.29}$$

But since space-time admits a conformal Killing vector parallel to n^a, we have, by (4.33),

$$\frac{D}{dp}(\mathscr{R}^2 \delta A) = 0\,. \tag{6.30}$$

Equation (6.23) follows immediately from (6.29) and (6.30).

(ii) Definition (6.26) with $n^a = \omega^a/\omega$ becomes

$$\mathscr{E} = (1/\omega) h^{ab}\omega_{a;b} - \overset{*}{(\omega/\omega)}\,. \tag{6.31}$$

But[23,24]

$$\omega^a{}_{;b} h^b_a = \omega_a \dot{u}^a\,, \tag{6.32}$$

and by (5.5), $\omega_a \dot{u}^a = \frac{1}{2}\omega\mathscr{E}$. Equation (6.31) therefore becomes

$$\tfrac{1}{2}\mathscr{E} = -\overset{*}{\omega}/\omega\,, \tag{6.33}$$

and hence with the aid of (4.34) for $\mathscr{E}$ we obtain

$$\frac{D}{dp}(\omega^2 \delta A) = 0. \tag{6.34}$$

Unlike (6.29), the conformal symmetry property was used in the derivation of (6.34). Equation (6.24) follows directly from (6.30) and (6.34).

(iii) Equation (6.25) is obtained from (6.23) and (6.24). □

An identity of the form (6.25) was derived by Prasad[10] without assuming the conformal symmetry property but assuming instead that the magnetic field and fluid vorticity are aligned, that $\sigma_{ab} H^b = 0$ and $\overset{*}{(H^a/H)} = 0$, and that the shear $\mathscr{S}_{ab}$ of the congruence of magnetic field/vortex lines vanishes.

In Theorem 6.1, the nature of the fluid was not specified, except for the requirement that $E^a = 0$. If the fluid is a thermodynamical perfect fluid then $\mathscr{R}, H, \omega$ can all be related to f, the index of the fluid.

Theorem 6.2: Suppose a fluid space-time admits a spacelike conformal motion with symmetry vector parallel to n^a, where n^a is given by (6.22), and that $E^a = 0$. If further $\dot{u}^a$ is given by (6.21) and

$$S_{,a} n^a = 0, \tag{6.35}$$

then

$$\text{(i)}\quad \frac{D}{dp}\left(\frac{\mathscr{R}}{f}\right) = 0, \tag{6.36}$$

$$\text{(ii)}\quad \frac{D}{dp}\left(\frac{H}{f^2}\right) = 0, \tag{6.37}$$

$$\text{(iii)}\quad \frac{D}{dp}\left(\frac{\omega}{f}\right) = 0, \tag{6.38}$$

where $\mathscr{R}$ is the rotation of the spacelike congruence generated by n^a as measured by u^a.

Proof: (i) By (5.5),

$$\mathscr{E} = 2 n_a \dot{u}^a, \tag{6.39}$$

and using (6.21) for $\dot{u}^a$ together with the assumption $S_{,a} n^a = 0$, we obtain

$$\mathscr{E} = -2(\log f)^*. \tag{6.40}$$

On relating $\mathscr{E}$ to δA through (4.34) we find that

$$\frac{D}{dp}(f^2 \delta A) = 0. \tag{6.41}$$

But by (4.33),

$$\frac{D}{dp}(\mathscr{R}^2 \delta A) = 0, \tag{6.42}$$

and (6.36) follows immediately from (6.41) and (6.42).

(ii) and (iii). Equations (6.37) and (6.38) follow directly from (6.23), (6.24), and (6.36). □

The assumption $S_{,a} n^a = 0$ of (6.35) is consistent with (6.13) which would apply through the "frozen-in" condition (6.5) if, for instance, the magnetic field were force-free. If the stronger assumption, S constant everywhere, is made, then f is directly proportional to r defined by (5.50). To see this, we note that when S is constant everywhere, the Gibbs equation (6.7) reduces to $dp = \rho\, df$ and hence since $f = (\mu + p)/\rho$, (5.50) becomes

$$r = \exp\left(\int_{f_0}^{f} \frac{df}{f}\right) = \frac{f}{f_0}, \tag{6.43}$$

where f_0 is a constant. In place of Eqs. (6.36)–(6.38) we therefore have for an isentropic thermodynamical perfect fluid [equivalently, a perfect fluid with barotropic equation of state $p = p(\mu)$]

$$\frac{D}{dp}\left(\frac{\mathscr{R}}{r}\right) = 0, \tag{6.44a}$$

$$\frac{D}{dp}\left(\frac{H}{r^2}\right) = 0, \tag{6.44b}$$

$$\frac{D}{dp}\left(\frac{\omega}{r}\right)=0. \tag{6.44c}$$

Equations (6.23), (6.24), (6.36), and (6.44a) were derived from a comoving observer at any one point. But since the magnetic field/vortex lines are material lines in the fluid, all other observers employed along the congruence will also be comoving if a comoving observer is employed at any point,[15] and hence these four equations are valid along the length of the congruence; $\mathscr{R}^2/H$, $\mathscr{R}/\omega$, $\mathscr{R}/f$, and $\mathscr{R}/r$ are therefore conserved along the congruence and any one may be taken as a measure of the strength of a flux tube formed by curves of the congruence. Also from (6.25), (6.38), and (6.44c), we have along a magnetic field line

$$\omega^2/H=\text{const}, \tag{6.45a}$$

$$\omega/f=\text{const}, \tag{6.45b}$$

$$\omega/r=\text{const}. \tag{6.45c}$$

All three results (6.45a)–(6.45c) were derived assuming that space-time admits a conformal Killing vector parallel to H^a and that $E^a=0$: otherwise (6.45a) is independent of the physical nature of the fluid, (6.45b) applies to a thermodynamical perfect fluid satisfying (6.35), and (6.45c) to a perfect fluid with equation of state $p=p(\mu)$. These three results in general relativity are similar in nature to Ferraro's law of isorotation[26] in nonrelativistic magnetohydrodynamics in which under certain circumstances the angular velocity of rotation of an electrically conducting fluid about an axis of symmetry is constant along a magnetic field line. Other analogs in general relativity of Ferraro's law of isorotation have been considered.[10,55,57,58]

VII. CONCLUDING REMARKS

As with a timelike conformal motion, we have seen that the properties of a spacelike conformal motion with symmetry vector parallel to a spacelike unit vector field n^a can usefully be studied in terms of quantities such as the expansion, shear, and rotation of the spacelike congruence generated by n^a. The essential properties of this spacelike congruence, which apply in both general relativity and Newtonian theory, are vanishing shear and a rotation which varies inversely as the square root of cross-sectional area along a flux tube formed by curves of the congruence. Purely relativistic effects are also important, especially in relation to material curves in a fluid space-time: if $n_a u^a=0$, then in an irrotational fluid the curves of the congruence generated by n^a must be material curves, while if the vorticity of the fluid is nonzero then the curves are material curves if and only if the congruence is a vortex congruence.

We have considered here applications of our results to fluid space-times only. For instance, we have not investigated applications to matter symmetries in the general relativistic kinetic theory of gases. The concept of a matter symmetry was introduced by Berezdivin and Sachs[59,60]: roughly, a matter symmetry will exist if there exists a vector field on eight-dimensional one-particle phase space that leaves the distribution function of matter unchanged. Berezdivin and Sachs established elementary properties and showed that a surface forming matter symmetry for a collision-free gas gives rise to a motion on space-time. The idea of a matter symmetry may be extended through the work of Iwai[61] and Oliver and Davis[62]; the latter two authors include theorems relating to conformal Kiling vectors. Oliver and Davis also considered two generalizations of the definition of matter symmetries in order to relate symmetry properties more general than motions to matter symmetries. There are many questions concerned with the relation between symmetries in one-particle phase space expressed as matter symmetries and symmetries in space-time that have not yet been settled.

Spacelike symmetries have been studied in connection with the Cauchy initial value problem.[3,63–65] If an initial spacelike hypersurface is endowed with a symmetry property, future spacelike hypersurfaces compatible with Einstein's equations do not necessarily preserve the symmetry. Berger[65] has derived the constraints imposed on initial data in a spacelike hypersurface due to the presence in space-time of a conformal Killing vector. For the special cases of a motion and a homothetic motion it can be shown that these constraints are preserved by the Einstein evolution equations in free space. For a proper conformal Killing vector this in general is not the case. Additional restrictions on the spacelike hypersurface have to be satisfied and in general these restrictions prevent a proper conformal Killing vector in spacelike initial data from being a space-time proper conformal Killing vector.

The level of symmetry considered in this paper, that of a conformal motion, represents a condition on the first derivatives of the metric tensor g_{ab}. This compares with, for instance, the affine collineation, which places a restriction on the second derivatives of g_{ab}, and the Ricci and curvature collineations, which place restrictions on the third derivatives of the metric tensor, which cannot in general be reduced to conditions on the first derivative (Oliver and Davies[3]). Also in the symmetry property inclusion diagram,[21] the Riemann curvature tensor does not appear at the level of symmetry of a conformal motion. Hence Einstein's equations (and the energy and momentum conservation equations) were not required in the derivation of many of our results, which are therefore essentially kinematical in nature. For instance, material curves result as a consequence of a restriction on the flow due to the conformal symmetry property and not due to the physical nature of the fluid.

ACKNOWLEDGMENTS

The authors would like to express their appreciation to the referee for useful comments. DPM would like to thank the members of the Department of Mathematics of the University of Queensland, Brisbane, Australia, for their hospitality during the second symester of 1983. MT would like to thank the members of the Department of Applied Mathematics of the University of the Witwatersrand for their hospitality during a recent visit.

DPM would also like to acknowledge the Council of Scientific and Industrial Research (CSIR), Pretoria, South Africa, for its support with an Overseas Research Bursary and the University of the Witwatersrand for an Anderson–Capelli-Convocation travel grant.

APPENDIX A: MATERIAL CURVES

This appendix is concerned with the equation governing the propagation along a fluid particle world line of the unit tangent vector to a material curve in a fluid.[16]

Theorem A.1: Let n^a denote the unit tangent vector to a spacelike curve in a fluid space-time such that $n_a u^a = 0$, where u^a is the unit four-velocity of the fluid. Then the curve is a material curve in the fluid if and only if n^a satisfies the propagation equation

$$h^a_b \dot{n}^b = \overset{*}{u}{}^a - (n_b \overset{*}{u}{}^b) n^a, \tag{A1}$$

where $h^a_b = g^a_b + u^a u_b$.

Proof: Consider a spacelike curve $\mathscr{C}$ with unit tangent vector n^a $(n_a u^a = 0)$.

(i) Suppose first that $\mathscr{C}$ is a material curve in the fluid. Then any two neighboring fluid particles initially on $\mathscr{C}$ always lie on $\mathscr{C}$ and hence the vector linking these particles always lies in the instantaneous direction of the unit tangent vector, n^a, to $\mathscr{C}$ at the particles. If δl is the distance as measured by u^a between the particles at any instant, then the vector $\delta l n^a$ links at all times these two particles and since also $n_a u^a = 0$, $\delta l n^a$ lies in the rest space of u^a and is therefore a relative position vector. But a relative position vector, $X^a_\perp$, satisfies the equation[23]

$$h^a_b \dot{X}^b_\perp = u^a{}_{;b} X^b_\perp; \tag{A2}$$

hence

$$h^a_b (\delta l n^b)^{\cdot} = u^a{}_{;b} (\delta l n^b). \tag{A3}$$

Contracting (A3) with n_a gives

$$(\delta l)^{\cdot} / \delta l = n_b \overset{*}{u}{}^b, \tag{A4}$$

and on substituting from (A4) for $(\delta l)^{\cdot}/\delta l$ into (A3) we obtain (A1).

(ii) Conversely, suppose that (A1) is satisfied. Define

$$\lambda(\tau; y^\alpha) = \exp\left(\int_0^\tau n_b \overset{*}{u}{}^b \, d\tau\right), \tag{A5}$$

where integration is performed along a fluid particle world line from some arbitrarily chosen space section S of space-time, τ is proper time measured along a fluid particle world line from S, and $\{y^\alpha\}$ $(\alpha = 1,2,3)$ are the coordinates of a fluid particle in S. Then

$$\dot{\lambda}/\lambda = n_b \overset{*}{u}{}^b, \tag{A6}$$

and (A1) may be rewritten as

$$h^a_b (\lambda n^b)^{\cdot} = u^a{}_{;b} (\lambda n^b). \tag{A7}$$

But since $n_a u^a = 0$, $\epsilon \lambda n^a$ for any small constant ϵ lies in the rest space of u^a and will link two neighboring fluid particles at $\tau = 0$. Hence there exists a relative position vector $X^a_\perp$ such that at $\tau = 0$, $X^a_\perp = \epsilon \lambda n^a$. Define $C^a = X^a_\perp - \epsilon \lambda n^a$; then at $\tau = 0$ we have $C^a = 0$, and also by multiplying (A7) by ϵ and subtracting from (A2), we have for, $\tau > 0$,

$$h^a_b \dot{C}^b = u^a{}_{;b} C^b. \tag{A8}$$

Hence $C^a = 0$, and $X^a_\perp = \epsilon \lambda n^a$, for all $\tau > 0$. Thus if $\epsilon \lambda n^a$ links two neighboring fluid particles initially it does so at all later times. The same fluid particles therefore always lie on $\mathscr{C}$ as the fluid evolves, and hence $\mathscr{C}$ is a material curve in the fluid. □

Since it was assumed in Theorem A1 that $n_a u^a = 0$, we have $n_b \overset{*}{u}{}^b = -u_b \overset{*}{n}{}^b$ and Eq. (A1) may be written equivalently as

$$h^a_b \dot{n}^b = \overset{*}{u}{}^a + (u_b \overset{*}{n}{}^b) n^a. \tag{A9}$$

For comparison we state the corresponding theorem in Newtonian theory.

Theorem A.2: Let n^α denote the unit tangent vector to a curve in a fluid and denote by v^α the velocity field of the fluid particles. Then the curve is a material curve in the fluid if and only if n^α satisfies the propagation equation

$$\dot{n}^\alpha = \overset{*}{v}{}^\alpha - (n_\beta \overset{*}{v}{}^\beta) n^\alpha, \tag{A10}$$

where the overhead dot denotes the convective time derivative. □

APPENDIX B: PROPAGATION EQUATION FOR ROTATION

In this appendix an outline is given of the derivation of the equation governing the propagation of $\mathscr{R}_{ab}$ along a curve of a spacelike congruence. Further details, as well as the propagation equations for $\mathscr{E}$ and $\mathscr{S}_{ab}$, are given by Tsamparlis and Mason.[15]

Theorem B.1: Let n^a be the unit tangent vector field to the curves of a spacelike congruence. The propagation equation for $\mathscr{R}_{ab}$ may be expressed in the following two equivalent forms:

$$\text{(i)}\quad p^c_a p^d_b (\overset{*}{\mathscr{R}}_{cd} - \overset{*}{n}_{[c;d]} + 2w^t n_{t;[c} \mathring{n}_{d]}) - 2\mathscr{S}^c{}_{[a} \mathscr{R}_{b]c} + \mathscr{E} \mathscr{R}_{ab} = 0, \tag{B1}$$

$$\text{(ii)}\quad \mathscr{L}_n \mathscr{R}_{ab} = p^c_a p^d_b (\overset{*}{n}_{[c;d]} - 2w^t n_{t;[c} \mathring{n}_{d]}). \tag{B2}$$

Proof: The Ricci identity for n^a is

$$n_{a;bc} - n_{a;cb} = R_{tabc} n^t. \tag{B3}$$

On contracting (B3) with n^c and noting that

$$n_{a;cb} n^c = \overset{*}{n}_{a;b} - n_{a;c} n^c{}_{;b}, \tag{B4}$$

we obtain

$$(n_{a;b})^* = \overset{*}{n}_{a;b} - n_{a;c} n^c{}_{;b} + R_{tabc} n^t n^c. \tag{B5}$$

Define

$$A_{ab} = p^c_a p^d_b n_{c;d} = \mathscr{R}_{ab} + \tfrac{1}{2}\mathscr{E} p_{ab} + \mathscr{S}_{ab}. \tag{B6}$$

On substituting (2.23) [with (B6)] into (B5), projecting on indices a and b with $p^c_a p^b_d$ and using the Greenberg transport law in the form (2.12), we obtain (renaming the free indices)

$$p^c_a p^d_b \overset{*}{A}_{cd} = -A_{ac} A^c{}_b + p^c_a p^d_b R_{tcds} n^t n^s + p^c_a p^d_b (\overset{*}{n}_{c;d} - \overset{*}{n}_c \overset{*}{n}_d + \mathring{n}_c \mathring{n}_d + 2\mathring{n}_c w^t n_{t;d}). \tag{B7}$$

The propagation equation for $\mathscr{R}_{ab}$ is the skew part of (B7). [The propagation equation for $\mathscr{E}$ is the trace of (B7) and the propagation equation for $\mathscr{S}_{ab}$ is the symmetric trace-free part of (B7).]

(i) The skew part of (B7) is

$$p_a^c\, p_b^d \overset{*}{\mathscr{R}}_{cd} = -A^c{}_{[b}A_{a]c} + p_a^c\, p_b^d R_{t\,[cd\,]s} n^t n^s + p_a^c\, p_b^d \big(\overset{*}{n}_{[c;d\,]} - 2w^t n_{t;[c}\mathring{n}_{d\,]}\big). \tag{B8}$$

Since

$$A^c{}_{[b}A_{a]c} = \mathscr{E}\mathscr{R}_{ab} - 2\mathscr{S}^c{}_{[a}\mathscr{R}_{b\,]c}, \tag{B9}$$

and

$$R_{t\,[cd\,]s} n^t n^s = 0, \tag{B10}$$

(B8) reduces to (B1).

(ii) Equation (B2) is derived from (B1) by showing, with the aid of the Greenberg transport law (2.24), that

$$p_a^c\, p_b^d \overset{*}{\mathscr{R}}_{cd} = \underset{n}{\mathscr{L}}\mathscr{R}_{ab} - \mathscr{E}\mathscr{R}_{ab} + 2\mathscr{S}^c{}_{[a}\mathscr{R}_{b\,]c}. \tag{B11}$$

□

[1] J. Ehlers, Akad.Wiss. Lit. (Mainz) Abh. Math. Naturwiss. Kl. **11**, 793 (1961).

[2] J. Ehlers, P. Geren, and R. K. Sachs, J. Math. Phys. **9**, 1344 (1968).

[3] D. R. Oliver and W. R. Davis, Gen. Relativ. Gravit. **8**, 905 (1977).

[4] P. J. Greenberg, J. Math. Anal. Appl. **30**, 128 (1970).

[5] C. D. Ciubotariu, An. Stiint. Univ. Al. I. Cuza Iasi **19**, 113 (1973).

[6] C. D. Ciubotariu, An. Stiint. Univ. Al. I. Cuza Iasi **19**, 121 (1973).

[7] C. D. Ciubotariu, Bul. Inst. Politeh. Iasi **22**, 93 (1976).

[8] T. H. Date, Ann. Inst. H. Poincaré **24**, 417 (1976).

[9] G. Prasad, Ann. Inst. H. Poincaré **30**, 17 (1979).

[10] G. Prasad, Gen. Relativ. Gravit. **11**, 177 (1979).

[11] G. Prasad, Inter. J. Theor. Phys. **18**, 329 (1979).

[12] G. Prasad and B. B. Sinha, Nuovo Cimento B **52**, 105 (1979).

[13] G. Prasad and B. B. Sinha, Nuovo Cimento B **54**, 401 (1979).

[14] G. Prasad and B. B. Sinha, Ann. Inst. H. Poincaré **32**, 15 (1980).

[15] M. Tsamparlis and D. P. Mason, J. Math. Phys. **24**, 1577 (1983).

[16] A material curve in a fluid is a curve which consists at all times of the same fluid particles and therefore it moves with the fluid as the fluid evolves; it is sometimes said to be "frozen-in" to the fluid.

[17] G. Williams, J. Math. Anal. Appl. **22**, 646 (1968).

[18] A. Trautman, in *Lectures in General Relativity*, Brandeis Summer Institute in Theoretical Physics, edited by S. Deser and K. Ford (Prentice–Hall, Englewood Cliffs, NJ, 1964), Vol. 1, Chap. 8, pp. 201–227.

[19] W. R. Davis, L. H. Green, and L. K. Norris, Nuovo Cimento B **34**, 256 (1976).

[20] D. R. Oliver and W. R. Davis, J. Math. Phys. **17**, 1790 (1976).

[21] L. K. Norris, L. H. Green, and W. R. Davis, J. Math. Phys. **18**, 1305 (1977).

[22] C. B. Collins, J. Math. Phys. **25**, 995 (1984).

[23] G. F. R. Ellis, in *General Relativity and Gravitation*, edited by R. K. Sachs (Academic, New York, 1971), pp. 104–182.

[24] G. F. R. Ellis, in *Cargèse Lectures in Physics*, edited by E. Schatzman (Gordon and Breach, New York, 1973), Vol. 6, pp. 1–60.

[25] A. Lichnerowicz, *Relativistic Magnetohydrodynamics* (Benjamin, Amsterdam, 1967), p. 28.

[26] V. C. A. Ferraro, Mon. Not. R. Astron. Soc. **97**, 458 (1937).

[27] Latin indices run over the four coordinates of space-time, Greek indices over the three spatial coordinates. A semicolon denotes covariant differentiation with respect to the metric tensor g_{ab} of space-time [signature $(-+++)$], and units are used for which the velocity of light and Einstein's gravitational constant are both unity. In Newtonian theory, the $\{x^\alpha\}$ are fixed curvilinear coordinates and a comma denotes covariant differentiation with respect to the flat three-space metric tensor $h_{\alpha\beta}$ [signature $(+++)$]; thus $h_{\alpha\beta,\gamma} = 0$.

[28] η^{abcd} is defined by

$$\eta^{abcd} = \eta^{[abcd\,]}, \quad \eta^{0123} = (-g)^{-1/2},$$

where $g = \det[g_{ab}]$. We make use of the following identities:

$$\eta^{abcd}\gamma_{arst} = -3!\delta_r^{[b}\delta_s^c\delta_t^{d\,]}, \quad \eta^{abcd}\eta_{abst} = -4\delta_s^{[c}\delta_t^{d\,]}.$$

[29] $\eta^{\alpha\beta\gamma}$ is defined by

$$\eta^{\alpha\beta\gamma} = \eta^{[\alpha\beta\gamma]}, \quad \eta^{123} = h^{-1/2},$$

where $h = \det[h_{\alpha\beta}]$. We make use of the following identities:

$$\eta^{\alpha\beta\gamma}\eta_{\alpha\lambda\mu} = 2\delta_\lambda^{[\beta}\delta_\mu^{\gamma]}, \quad \eta^{\alpha\beta\gamma}\eta_{\alpha\beta\mu} = 2\delta_\mu^\gamma.$$

[30] For a timelike congruence of curves with unit tangent vector v^a, the vorticity tensor ω_{ab}, the vorticity vector ω^a, the (rate of) expansion θ, the (rate of) shear σ_{ab}, and the acceleration $\dot{v}^a$ are defined by

$$\omega_{ab} = h_a^c h_b^d v_{[c;d\,]},$$
$$\omega^a = \tfrac{1}{2}\eta^{abcd}v_b v_{c;d},$$
$$\theta = h^{cd}v_{c;d},$$
$$\sigma_{ab} = h_a^c h_b^d v_{(c;d)} - (\theta/3)h_{ab} \quad (\sigma_a^a = 0),$$
$$\dot{v}^a = v^a{}_{;b}v^b,$$

where $h^{ab} = g^{ab} + v^a v^b$ is the projection tensor. The vorticity ω is the magnitude of $\omega^a : \omega^2 = \omega_a\omega^a = \frac{1}{2}\omega_{ab}\omega^{ab}$. An overhead dot denotes covariant differentiation along the world line of v^a, so for example $\dot{A}^a = A^a{}_{;b}v^b$. In a fluid space-time v^a is usually identified with u^a, the unit four-velocity of the fluid.

[31] G. F. R. Ellis, J. Math. Phys. **8**, 1171 (1967).

[32] A. J. White and C. B. Collins, J. Math. Phys. **25**, 332 (1984).

[33] C. B. Collins and A. J. White, J. Math. Phys. **25**, 1460 (1984).

[34] M. Born, Ann. Phys. **30**, 1 (1909).

[35] G. Herglotz, Ann. Phys. **31**, 393 (1910).

[36] F. Noether, Ann. Phys. **31**, 919 (1910).

[37] F. A. E. Pirani and G. Williams, Semin. Janet **5**, 8 (1962).

[38] R. H. Boyer, Proc. R. Soc. London Ser. A **283**, 343 (1964).

[39] The evolution of ω described by (4.18) is the same as in the shear-free expansion or contraction of a perfect fluid with equation of state $p = \mu/3$ (Ellis[24]). This is consistent with a result of Oliver and Davis[3] who showed that a shear-free isentropic, thermodynamical, perfect fluid admits a timelike conformal motion with symmetry vector parallel to the fluid four-velocity u^a if $(p - \mu/3)^{\cdot} = 0$.

[40] J. Serrin, in *Handbuch der Physik*, edited by S.Flügge (Springer, Berlin, 1959), Vol. 8(1), pp. 162 and 163.

[41] K. Gödel, Rev. Mod. Phys. **21**, 447 (1949).

[42] H. P. Robertson and T. W. Noonan, *Relativity and Cosmology* (Saunders, Philadelphia, 1968), pp. 330–334 and 410–413.

[43] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-time* (Cambridge U. P., Cambridge, 1973), pp. 168–170.

[44] R. Adler, M. Bazin, and M. Schiffer, *Introduction to General Relativity* (McGraw–Hill, New York, 1975), pp. 437–448.

[45] M. P. Ryan and L. C. Shepley, *Homogeneous Relativistic Cosmologies* (Princeton U.P., Princeton, NJ, 1975), pp. 128–131.

[46] J. R. Ray and J. C. Zimmerman, Nuovo Cimento B **42**, 183 (1977).

[47] Alternatively, $\mathscr{E}(\omega)$ and $\mathscr{S}_{ab}(\omega)$ can be obtained from the constraint equation for H_{ab} (see Refs. 23 and 24),

$$H_{ab} = 2\dot{u}_{(a}\omega_{b)} - h_a^s h_b^t(\omega_{(s}{}^{d;c} + \sigma_{(s}{}^{d;c})\eta_{t)fdc}u^f$$

[$\mathscr{E}(\omega)$ and $\mathscr{S}_{ab}(\omega)$ therefore do not depend on Einstein's field equations] and $\mathscr{R}_{ab}(\omega)$ can be obtained from the (0ν) components of Einstein's field equations[23,24]

$$h^{ab}(\tfrac{2}{3}\theta_{,b} - \sigma_{bc;d}h^{cd}) - \eta^{abcd}u_b(\omega_{c;d} + 2\omega_c\dot{u}_d) = q^a.$$

To evaluate $\overset{*}{n}{}^a$ where $n^a = \omega^a/\omega$, both the constraint equation for H_{ab} and the (0ν) field equations are required.

[48] We have expressed this result in terms of r for closer comparison with the spacelike results. For an isentropic thermodynamical perfect fluid it is shown in Sec. VI that r is directly proportional to the index of the fluid f defined by $f = (\mu + p)/\rho = 1 + e + p/\rho$; ρ is the rest-mass density and e is the specific internal energy density. Under the stated conditions in the text, $(1/f)u^a$ is a timelike conformal Killing vector, which is the result established by Oliver and Davis.[3]

[49] C. B. G. McIntosh, Phys. Lett. A **50**, 429 (1975).

[50] C. B. G. McIntosh, Gen. Relativ. Gravit. **7**, 199 (1976).

[51] By contracting (5.69) with p^{ab}, a general expression for $p^{ab}\mathscr{E}_{,b}$ can be

derived which corresponds to the (0ν) components of Einstein's field equations in relativistic fluid dynamics which yield an expression for $h^{ab}\theta_{,b}$ (see Ref. 47).

[52]That $p = \mu + 2\Lambda$ is a *necessary* condition for a homothetic motion now follows from (3.38) and (5.84).

[53]J. D. Bekenstein and E. Oron, Phys. Rev. D **18**, 1809 (1978).

[54]The magnetic field lines can be shown to be material curves when $E^a = 0$ by noting that Maxwell's propagation equation for H^a when $E^a = 0$ reduces to[24]

$$h^a_{\ b}\dot{H}^b = u^a_{\ ;b}H^b - \theta H^a;$$

thus, if $n^a = H^a/H$, then

$$h^a_{\ b}\dot{n}^b = \overset{*}{u}^a - (n_b\overset{*}{u}^b)n^a,$$

and since $n_a u^a = 0$ it follows from Theorem A.1 of Appendix A that n^a must be the unit tangent vector to a material curve.

[55]D. P. Mason, Gen. Relativ. Gravit. **8**, 871 (1977).

[56]To derive the magnetic force term in (6.21) we note that the Minkowski tensor is (Ref. 25, Chap. 4)

$$T_{\rm EM}^{\ \ ab} = F^{ac}H^b_{\ c} - \tfrac{1}{4}g^{ab}F_{cd}H^{cd}.$$

If the constitutive equations $D^a = \epsilon E^a$ and $B^a = \lambda H^a$ apply, where D^a and B^a are the electric displacement and magnetic induction four-vectors, respectively, and if $E^a = 0$, then

$$F^{ab} = -\lambda\eta^{abcd}u_cH_d, \quad H^{ab} = -\eta^{abcd}u_cH_d,$$

and $T_{\rm EM}^{\ \ ab}$ is *symmetric*. With the aid of Maxwell's equations

$$H^{ab}_{\ \ ;b} = J^a, \quad F_{[ab;c]} = 0,$$

where J^a is the four-current density vector, we find that if λ is constant,

$$T_{\rm EM}^{\ \ ab}{}_{;b} = -F^{ab}J_b.$$

Also J^a may be split by u^a into $J^a = \mathscr{J}^a + qu^a$, where $\mathscr{J}^a = h^a_{\ b}J^b$ and $q = -u_aJ^a$; thus

$$T_{\rm EM}^{\ \ ab}{}_{;b} = -\lambda\eta^{abcd}u_b\mathscr{J}_cH_d.$$

[57]P. Yodzis, Phys. Rev. D **3**, 2941 (1971).
[58]M. Tsamparlis and D. P. Mason, Gen. Relativ. Gravit. **14**, 717 (1982).
[59]R. Berezdivin and R. K. Sachs, in *Relativity*, edited by M. Carmeli, S. I. Fickler, and L. Witten (Plenum, New York, 1970), pp. 125–131.
[60]R. Berezdivin and R. K. Sachs, J. Math. Phys. **14**, 1254 (1973).
[61]T. Iwai, Tensor (N.S.) **31**, 98 (1977).
[62]D. R. Oliver and W. R. Davis, Ann. Inst. H. Poincaré **30**, 339 (1979).
[63]R. Berezdivin, J. Math. Phys. **15**, 1963 (1974).
[64]V. Moncrief, J. Math. Phys. **16**, 493 (1975).
[65]B. K. Berger, J. Math. Phys. **17**, 1268 (1976).

On the scalar-tetradic theory A—The Birkhoff theorem: Horizons and singularities

D. Sáez
Departamento de Mecánica y Astronomía, Facultad de Matemáticas, Burjasot, Valencia, Spain
J. V. Arnau
Departamento de Ecuaciones Funcionales, Facultad de Matemáticas, Burjasot, Valencia, Spain

(Received 18 February 1985; accepted for publication 26 April 1985)

In a recent paper a unique solution of the field equations of theory A, describing the outer field of the sun, has been obtained for each value of the coupling constant W. Here, this solution is studied in detail for any possible value of W (horizons, singularities,. . .) and the Birkhoff theorem is proved in the framework of theory A.

I. INTRODUCTION

The scalar-tetradic theory A (TA) is a generalization of Moller's theory of gravitation (MT).[1,2] In order to explain the interest in TA, let us mention the following facts.

(i) MT is a theory of gravitation giving satisfactory results in its applications to cosmology and gravitational radiation,[3,4] which has been proved to be formally identical[5] to the teleparallel limit of the guage theory formulated by Hehl *et al.*[6]

(ii) In TA, there is a scalar field playing the role of the gravitational constant just as in the theory proposed by Brans and Dicke[7] (BDT). This fact could be important in cosmology and astrophysics, where the small variations of the scalar field (gravitational constant) could lead to new results. It has suggested numerous investigations in BDT,[8–10] which can be extended to TA.

(iii) An energy-momentum complex leading to localized gravitational energy[11] has been defined in TA.[1]

(iv) Some satisfactory results in connection with cosmology and the gravitational radiation problem have been obtained in theory A (see Ref. 12); in particular the intergalactic energy assumption is not necessary to solve the problem of the missing matter of the universe.

(v) The PPN limit of TA appears to be identical to that of general relativity (GR) for any nonvanishing value of the coupling constant W of TA.[13]

Because of these considerations, we think TA is interesting and, consequently, we are developing this theory in a series of papers. This one deals with both (1) the Birkhoff theorem in the framework of TA and (2) the existence of horizons and singularities in the case of the three solutions of the equations of TA, called 1, 2, and 3 in a previous paper,[13] which are the unique solutions compatible with the PPN limit of TA ($\beta = \gamma = 1$). In Ref. 13 we proved that solution 1 corresponds to $W < 0$ or $W > 2$, solution 2 to $0 < W < 2$, and solution 3 to $W = 2$; therefore, we have a unique solution corresponding to a fixed value of W.

The Birkhoff theorem guarantees that the outer field of any spherical source is static and it does not depend on its inner structure, but only on its Newtonian mass.

The horizons and the singularities will be important in discussions both of (a) the occurrence of black holes, and (b) the explanation of the present experimental data about the binary x-ray sources.

We shall now describe briefly this work. TA is summarized in Sec. II. In the next section we describe the spherically symmetric case by using Schwarzschild coordinates. The Birkhoff theorm is proved in Sec. IV. The curvature of solutions 1, 2, and 3 is computed in Sec. V, getting a singularity in the case of the solution 1. In Sec. VI we study the radial free fall of a test particle in the space-time described by the above solutions. In Sec. VII we use the complex defined in TA to calculate the gravitational energy localized outside a spherical star of Newtonian mass m and radius R. Finally, in Sec. VIII we summarize our main conclusions.

II. THE THEORY A

The gravitational field is described by a tetrad field $h_{\bar n}$ and a scalar field ϕ with the dimensions of the inverse of the gravitational constant G.

The barred indices denote different vectors of the tetrad and the nonbarred ones are common tensorial indices. All the indices run from 1 to 4.

The partial derivative of arbitrary A with respect to the X^i coordinate will be denoted by $A_{,i}$ and the ordinary covariant derivative (Christoffel symbols) by $A_{;i}$.

The vectors $h_{\bar 1}$, $h_{\bar 2}$, and $h_{\bar 3}$ are assumed to be real and $h_{\bar 4}$ imaginary, in this way the metric

$$g^{ij} = h^i_{\bar n} h^j_{\bar n} \tag{1}$$

has Lorentz signature, and the symbols of connection

$$\Gamma^i{}_{jk} = h^i_{\bar n} h_{\bar n j,k} \tag{2}$$

are nonsymmetric; then, the tensor $S^i{}_{jk} = \frac{1}{2}(\Gamma^i{}_{jk} - \Gamma^i{}_{kj})$ allows us to define the vector $\Gamma_i = S^n{}_{in}$.

As has been proved,[1] the equations of motion of TA are the following:

$$T^{ij}{}_{;j} = 0, \tag{3}$$

where T^{ij} is the energy momentum tensor. These equations are a consequence of the field equations. They are formally identical to those of GR, MT, and BDT.

The field equations of TA can be written as follows (see Ref. 12):

$$\phi(G_{ij} + \overset{*}{\lambda} H_{ij}) - (W/\phi)[\phi_{,i}\phi_{,j} - \tfrac{1}{2}g_{ij}\phi_{,k}\phi^{,k}] + \phi_{,i}\Gamma_j + \phi_{,j}\Gamma_i - 2\phi_{,n}\Gamma^n g_{ij} - \phi^{,n}(S_{jin} + S_{ijn}) = 8\pi T_{ij}, \tag{4}$$

$$\overset{*}{\lambda}\,[\phi F_{ij} + \phi_{,s}(S_j{}^s{}_i - S_i{}^s{}_j - S^s{}_{ji})]$$

$$+\phi_{,i}\Gamma_j - \phi_{,j}\Gamma_i + \phi_{,s}S^s{}_{ij} = 0, \qquad (5)$$

$$R\text{-}\overset{*}{\lambda} H^i{}_i - 4\Gamma^i{}_{;i} + 2W\phi^{-1}\phi^{,i}{}_{;i} - W\phi^{-2}\phi_{,i}\phi^{,i} = 0, \qquad (6)$$

G_{ij} being the Einstein tensor, R the scalar curvature, $\overset{*}{\lambda}$ and W arbitrary dimensionless constants, and $\phi^{,n} = g^{nm}\phi_{,m}$. The tensors F_{ij} and H_{ij} are also involved in the field equations of MT,

$$G_{ij} + \overset{*}{\lambda} H_{ij} = 8\pi T_{ij}, \quad F_{ij} = 0, \qquad (7)$$

and their explicit form (see Ref. 12), will not be necessary here.

III. THE SPHERICALLY SYMMETRIC CASE

The isotropic coordinates used in a previous paper[1] are not adequate to prove the Birkhoff theorem. We will use Schwarzschild coordinates. In these coordinates we will consider the following tetrad:

$$\begin{aligned}
&h_{\bar 1 1} = e^{\beta}\cos\Phi\sin\theta, \quad h_{\bar 1 2} = r\cos\Phi\cos\theta,\\
&h_{\bar 1 3} = -r\sin\Phi\sin\theta,\\
&h_{\bar 2 1} = e^{\beta}\cos\theta, \quad h_{\bar 2 2} = -r\sin\theta, \quad h_{\bar 2 3} = 0,\\
&h_{\bar 3 1} = e^{\beta}\sin\theta\sin\Phi, \quad h_{\bar 3 2} = r\sin\Phi\cos\theta,\\
&h_{\bar 3 3} = r\sin\theta\cos\Phi,\\
&h_{\bar 1 4} = h_{\bar 2 4} = h_{\bar 3 4} = h_{\bar 4 1} = h_{\bar 4 2} = h_{\bar 4 3} = 0, \quad h_{\bar 4 4} = ie^{\alpha}
\end{aligned} \qquad (8)$$

(whose explicit form in isotropic coordinates has been already presented[1]). Since we treat the nonstatic case, the functions α and β depend on r and t.

The line element has the form

$$dS^2 = -e^{2\alpha}dt^2 + e^{2\beta}dr^2 + r^2(\sin^2\theta\, d\Phi^2 + d\theta^2) \qquad (9)$$

and the tensor $S^i{}_{jk}$ has the following nonvanishing components:

$$\begin{aligned}
&2S^2{}_{12} = -2S^2{}_{21} = r^{-1}(e^{\beta}-1) = 2S^3{}_{13} = -2S^3{}_{31},\\
&2S^4{}_{41} = -2S^4{}_{14} = \alpha', \quad 2S^1{}_{14} = -2S^1{}_{41} = \dot\beta,
\end{aligned} \qquad (10)$$

where an overdot denotes a derivative with respect to the coordinate t and a prime denotes a derivative with respect to r.

As a consequence of a paper by Schweizer *et al.*[14] it follows that the tensors F_{ij} and H_{ij} involved in the equations of MT (and TA) vanish in the case of the above tetrad [take into account the form of the tetrad (8) and the metric (9) in the isotropic coordinates X, Y, Z, t]. This fact also can be proved by a simple but lengthy calculation based on Eqs. (9) and (10) and the definitions of H_{ij} and F_{ij} (see Ref. 12).

It is worthwhile to notice that ϕ is a function of r and t.

The components of the Einstein tensor corresponding to the metric (9) are given in numerous references.[15]

IV. THE BIRKHOFF THEOREM

By using the data of Sec. III and the field equations of TA presented in Sec. II, we easily obtain the equations describing the evolution of a spherically symmetric source.

In the presence of matter as well as in the vacuum, the six equations (5) reduce to

$$\dot\phi = 0, \qquad (11)$$

therefore, in the spherical case, the scalar ϕ only depends on the variable r.

Taking into account Eq. (11) and by writing $\phi = e^{\gamma}$, Eqs. (4) become

$$e^{-2\alpha}(\dot\beta/r) = -(4\pi/\phi)T^4_1, \qquad (12)$$

$$\begin{aligned}
&r^{-2} + e^{-2\beta}[-(W/2)\gamma'^2 + 2r^{-1}\gamma'(e^{\beta}-1)\\
&\quad + (2\beta'/r) - r^{-2}] = -(8\pi/\phi)T^4_4,
\end{aligned} \qquad (13)$$

$$-r^{-2} + e^{-2\beta}[-(W/2)\gamma'^2 + 2r^{-1}\alpha' + r^{-2}] = (8\pi/\phi)T^1_1, \qquad (14)$$

$$\begin{aligned}
&-e^{-2\alpha}(\ddot\beta + \dot\beta^2 - \dot\alpha\dot\beta) + e^{-2\beta}\\
&\quad\times[\alpha'' + \alpha'^2 - \alpha'\beta' + (\alpha'-\beta')r^{-1}\\
&\quad + (W/2)\gamma'^2 - \gamma'\{(e^{\beta}-1)r^{-1} - \alpha'\}] = (8\pi/\phi)T^2_2,
\end{aligned} \qquad (15)$$

and Eqs. (6) and (11) lead to

$$\begin{aligned}
&-r^{-2} - 2e^{-2\alpha}(\ddot\beta + \dot\beta^2 - \dot\alpha\dot\beta) + e^{-2\beta}\\
&\quad\times\{W[\gamma'' + \gamma'(\alpha' - \beta' + 2r^{-1}\\
&\quad + \gamma'/2)] + 2\alpha'' + 2\alpha'^2\\
&\quad - 2\alpha'\beta' + 3r^{-2} + 6\alpha' r^{-1} - 4\beta' r^{-1}\\
&\quad - 2e^{\beta}r^{-2}(1 + a'r)\} = 0.
\end{aligned} \qquad (16)$$

The last equation can be obtained from Eqs. (11)–(15) due to the existence of the identies $T^{ij}{}_{;j} = 0$.

In the vacuum spherically symmetric case ($T^{ij} = 0$), we obtain the following results: (i) Eq. (12) becomes

$$\dot\beta = 0, \qquad (17)$$

hence, the function β only depends on r (as ϕ); and (ii) Eq. (14) can be written in the form

$$\alpha' = (1/2r)(e^{2\beta}-1) + Wr\gamma'^2. \qquad (18)$$

The right-hand side of the last equation only depends on r; therefore, an integration will give the following functional form of the solution:

$$\alpha = \alpha(r) + f(t), \qquad (19)$$

$f(t)$ being an arbitrary function. Then, a simple redefinition of the time coordinate leads to a static tetrad generating a static spherically symmetric metric, and Eqs. (13), (18), (15), and (16) reduce to the corresponding equations of the static spherically symmetric case. Thus, we conclude that the outer field (h_{ij},ϕ) of a spherically symmetric nonstatic source is always a static spherically symmetric field, which proves the Birkhoff theorem in the framework of TA.

In spite of the complicated form of the field equations of TA, the Birkhoff theorem holds as well as in GR.

V. SINGULARITIES

Since solutions 1, 2, and 3 have been proved to be very satisfactory[13] due to their agreement with present solar system data, we will study these solutions in detail. In this section we compute the curvature with the essential aim of studying the existence of singularities.

We have the line element (isotropic coordinates)

$$dS^2 = -e^{2\alpha}dt^2 + e^{2\beta}(dr^2 + r^2d\theta^2 + r^2\sin^2\theta\, d\Phi^2), \qquad (20)$$

and we define the one-forms

$$W^1 = e^\beta\, dr,\quad W^2 = re^\beta\, d\theta,$$
$$W^3 = r\sin\theta e^\beta\, d\Phi,\quad W^4 = e^\alpha\, dt, \tag{21}$$

then the two-forms of curvature are[15]

$$\mathscr{R}^1_2 = -\mathscr{R}^2_1 = H(W^1\wedge W^2),$$
$$\mathscr{R}^1_3 = -\mathscr{R}^3_1 = H(W^1\wedge W^3),$$
$$\mathscr{R}^2_3 = -\mathscr{R}^3_2 = P(W^2\wedge W^3),$$
$$\mathscr{R}^1_4 = -\mathscr{R}^4_1 = U(W^1\wedge W^4),$$
$$\mathscr{R}^2_4 = -\mathscr{R}^4_2 = V(W^2\wedge W^4),$$
$$\mathscr{R}^3_4 = -\mathscr{R}^4_3 = V(W^3\wedge W^4), \tag{22}$$

where

$$H = -(1/r)e^{-2\beta}(\beta' + r\beta''), \tag{23}$$
$$P = -(1/r)e^{-2\beta}(2\beta' + r\beta'^2), \tag{24}$$
$$U = -e^{-2\beta}(\alpha'' + \alpha'^2 - \alpha'\beta'), \tag{25}$$
$$V = -(1/r)e^{-2\beta}\alpha'(1 + r\beta'), \tag{26}$$

and the symbol $\wedge$ denotes the exterior product.

Now, we compute the quantities H, P, U, and V for solutions 1, 2, and 3.

Solution 1. $W<0$ or $W>2$: This solution is defined by

$$\alpha = (1/\delta)[\ln(1-\Delta) - \ln(1+\Delta)], \tag{27}$$
$$\beta = [(1/\delta^2) - (1/\delta)]\ln(1-\Delta) + [(1/\delta^2) + (1/\delta)]\ln(1+\Delta), \tag{28}$$

with

$$\delta = [(W-2)/W]^{1/2}, \tag{29}$$
$$\Delta = m\delta/2r. \tag{30}$$

Taking into account Eqs. (23)–(30), we easily get

$$H = (X/2)[\Delta^2 - (2\Delta/\delta) + 1], \tag{31}$$
$$P = X[(\Delta^2/\delta^2) - (\Delta/\delta) + 1 - \Delta^2]\,[(\Delta/\delta) - 1], \tag{32}$$
$$U = -X[(\Delta/\delta) - 1]^2, \tag{33}$$
$$V = (X/2)[1 - \Delta^2 + 2(\Delta/\delta)^2 - 2(\Delta/\delta)], \tag{34}$$

where

$$X = (-2m/r^3)(1-\Delta)^{[(2/\delta) - (2/\delta^2) - 2]} \times (1+\Delta)^{[(-2/\delta) - (2/\delta^2) - 2]}. \tag{35}$$

We see that the functions H, P, U, and V tend to ∞ as r tends to $m\delta/2$; hence, the curvature becomes singular at this value of r, which will be denoted by r_S.

Furthermore, from Eqs. (27) and (28), it follows that the metric is not well behaved at r_S.

Here r_S does not define a horizon (as the Schwarzschild radius in GR) but a true physical singularity. An infalling observer feels infinite tidal forces at r_S. This solution must be discarded if a sphere of fluid with a realistic equation of state can reach the singular radius in a finite time in accordance with the equations of TA.

Solution 1 tends to the Schwarzschild solution as $|W|$ tends to ∞ (r_S tends to the Schwarzschild radius); in this limit the singularity at r_S is removed and the functions H, P, U, and V take finite values at $r = m/2$ (as occurs in the Schwarzschild horizon).

The singular radius is small except when W is close to zero; therefore, the small values of $|W|$ must be discarded at all.

Solution 2. $0<W<2$: In this case we have

$$\alpha = \eta[-\pi + 2\tan^{-1}\zeta], \tag{36}$$
$$\beta = \eta[\pi - 2\tan^{-1}\zeta] - \eta^2\ln(1+\zeta^{-2}), \tag{37}$$

where

$$\eta = [W/(2-W)]^{1/2}, \tag{38}$$
$$\zeta = 2\eta r/m, \tag{39}$$

and the function $y = \tan^{-1}x$ takes values in the interval $[0,\pi/2]$.

We easily obtain

$$H = Y\{\zeta^2[(m/r) - 1] + 1\}, \tag{40}$$
$$P = Y[2\zeta^2 - 4\eta\zeta - (m/r)(1+\eta^2) + 2(1+2\eta^2)], \tag{41}$$
$$U = -2\eta rY[(4\eta/m) - (2\zeta/m) - (\eta/r)], \tag{42}$$
$$V = -Y[1 + \zeta^2 + 2\eta^2 - 2\eta\zeta], \tag{43}$$

where

$$\mathrm{Y} = (4\eta^2/mr)\zeta^{-4\eta^2}(1+\zeta^2)^{(2\eta^2-2)}e^{-2\eta[\pi - 2\tan^{-1}\zeta]}. \tag{44}$$

The functions H, P, U, and V take finite values for any value of r, except for $r=0$; therefore, in the case of solution 2, the curvature does not have any singularity at $r\neq 0$ whatever the value of the constant W ($0<W<2$) may be. Elsewhere we have seen that in the case of this solution there are no horizons.[13]

Solution 3. $W=2$: the functions α and β have the following form:

$$\alpha = -m/r, \tag{45}$$
$$\beta = (m/r) - (m^2/4r^2). \tag{46}$$

These expressions can be obtained from Eqs. (36) and (37) in the limit $W\to 2$ ($\eta\to\infty$), which proves that, in this limit, solution 2 tends to solution 3.

We easily find

$$H = \mathring{X}[1 - (m/r)], \tag{47}$$
$$P = \mathring{X}[-2 + (2m/r) - (m^2/r^2) + (m^3/4r^3)], \tag{48}$$
$$U = \mathring{X}[-2 + (2m/r) - (m^2/2r^2)], \tag{49}$$
$$V = \mathring{X}[1 - (m/r) + (m^2/2r^2)], \tag{50}$$

where

$$\mathring{X} = -(m/r^3)e^{-2(m/r) + (m^2/2r^2)}. \tag{51}$$

From Eqs. (47)–(51) we conclude that, in the case of solution 3, the curvature is not singular for any nonvanishing value of r. Furthermore, we already know that, in this case, no horizons exist.[13]

So far we have studied the singularities of solutions 1, 2, and 3.

Since solutions 1, 2, and 3 do not have any horizon of the Schwarzschild type, the black holes described in GR are not possible in TA and, consequently, the experimental data about binary x-ray sources require an explanation in accordance with this fact.

VI. THE RADIAL FREE FALL OF A TEST PARTICLE

In this section we will study the radial free fall of a point test particle in the static spherically symmetric space-time described by solutions 1, 2, and 3. We will compare our results with those of GR (Schwarzschild case).

The proper time of the infalling observers is denoted by τ.

We shall calculate the following quantities: (1) the radial acceleration $(d^2r/d\tau^2)_0$ at an arbitrary point $r = r_0$, where the velocity vanishes, and (2) the radial velocity of a point test particle, which starts from a point $r = r_0$ with vanishing velocity.

The radial acceleration $(d^2r/d\tau^2)_0$ describes the intensity of the gravitational field and the knowledge of the radial velocity allows us to determine the interval of proper time elapsed to reach the radius r starting from $r = r_0$ with $(dr/d\tau)_{r=r_0} = 0$. In this manner, we can study whether some special radius (as the singular radius r_S of solution 1) is reached in a finite time.

From the equations of the geodesics of the metric (20), we easily get

$$\left(\frac{d^2r}{d\tau^2}\right)_0 = -\alpha' e^{-2\beta}, \tag{52}$$

$$\frac{dr}{d\tau} = -[e^{2(\alpha_0-\alpha)} - 1]^{1/2} e^{-\beta}, \tag{53}$$

where $\alpha_0 = \alpha(r_0)$. These expressions will be studied in the three cases corresponding to solutions 1, 2, and 3.

Solution 1: From Eqs. (27), (28), (52), and (53) we get

$$\left(\frac{d^2r}{d\tau^2}\right)_0 = -\frac{4\Delta^2}{m\delta^2}(1-\Delta)^{-(2/\delta^2)+(2/\delta)-1} \times (1+\Delta)^{-(2/\delta^2)-(2/\delta)-1}, \tag{54}$$

$$\frac{dr}{d\tau} = -[e^{2\alpha_0}(1-\Delta)^{-2/\delta}(1+\Delta)^{2/\delta} - 1]^{1/2} \times (1-\Delta)^{(1/\delta)-(1/\delta^2)}(1+\Delta)^{-(1/\delta)-1/\delta^2}. \tag{55}$$

As follows from Eq. (54), the modulus of the radial acceleration increases when r decreases, tending to the corresponding Schwarzschild aceleration as $|W| \to \infty$ $(\delta \to 1)$.

Furthermore, in the limit $r \to r_S$, the radial acceleration tends to infinity in such a way that, when W is negative (positive), the order of this infinite is smaller (greater) than the order of the infinite corresponding to the Schwarzschild metric in the limit $r \to m/2$.

The radial velocity given by Eq. (55) does not vanish at any point; therefore, a free infalling particle reaches the singular radius r_S in a finite time. However, the pressure inside a fluid sphere could stop the fall before the radius r_S is reached; especially when the radial acceleration of the free fall increases more slowly than in GR $(W<0)$.

Solution 2: By using Eqs. (36), (37), (52), and (53) we obtain

$$\left(\frac{d^2r}{d\tau^2}\right)_0 = -4\eta^2 m[m^2(1+\zeta^2)]^{2\eta^2-1} \times (m\zeta)^{-4\eta^2} e^{-2\eta(\pi - 2\tan^{-1}\zeta)}, \tag{56}$$

$$\frac{dr}{d\tau} = -[e^{2\alpha_0} - e^{-2\eta(\pi-2\tan^{-1}\zeta)}]^{1/2}(1+\zeta^{-2})^{\eta^2}. \tag{57}$$

The modulus of the radial acceleration given by Eq. (56) increases when r decreases, tending to ∞ as r tends to zero, and the radial velocity defined by Eq. (57) does not vanish at any point; therefore, any radius is reached in a finite time (in the free fall).

The space-time described by solution 2 has neither horizons nor singularities; hence, the black holes are not possible, and, consequently, new concepts are necessary in order to explain the present observational evidences about binary x-ray sources. More study about this question is necessary. Now, it is worthwhile to notice that, in a similar space-time (Yilmaz exponential metric with $\alpha = -m/r$ and $\beta = m/r$), Clapp[16] defined a new concept, the "gray hole," which could be useful in order to explain the above evidence (see Ref. 17).

Solution 3: Equations (45), (46), (52), and (53) lead to

$$\left(\frac{d^2r}{d\tau^2}\right)_0 = -\frac{m}{r^2}e^{-(2m/r)+(m^2/2r^2)}, \tag{58}$$

$$\frac{dr}{d\tau} = -[e^{2\alpha_0+(2m/r)} - 1]^{1/2} e^{-(m/r)+(m^2/4r^2)}. \tag{59}$$

These expressions can be also obtained from Eqs. (56) and (57) in the limit $\eta \to \infty$ $(W \to 2,\ W<2)$. The qualitative behavior of the radial acceleration and the radial velocity given by Eqs. (58) and (59), respectively, is the same as in the case of solution 2. So, any radius is reached in a finite time.

VII. THE ENERGY LOCALIZED OUTSIDE A SPHERICAL SOURCE

An energy-momentum complex leading to localized gravitational energy was defined[1] in the framework of theory A. After some easy manipulations, this complex can be written in terms of $\Gamma^i{}_{jk}$ and $S^i{}_{jk}$. Its explicit form is the following:

$$\dot{T}^i_j = (1/8\pi)(-g)^{1/2}(A^i_j + t^i_j), \tag{60}$$

where

$$t^i_j = \Gamma^s{}_{nj}(S^n{}_s{}^i + S_s{}^{ni} - S^i{}_s{}^n) - 2\Gamma^n\Gamma^i{}_{nj} + 2\Gamma^i\Gamma^n{}_{nj} + \delta^i_j[2\Gamma_n\Gamma^n - S_{rsp}S^{srp} + \tfrac{1}{2}S_{rsp}S^{rps}] \tag{61}$$

and

$$A^i_j = -\tilde{\lambda} H^i_j + W\phi^{-2}[\phi^{,i}\phi_{,j} - (1/2)\delta^i_j\phi_{,k}\phi^{,k}] - \phi^{-1}[\phi^{,i}\Gamma_j + \phi_{,j}\Gamma^i - 2\phi_{,n}\Gamma^n\delta^i_j - \phi^{,n}(S_j{}^i{}_n + S^i{}_{jn}) - 8\pi T^i_j]. \tag{62}$$

The energy localized inside a region V is defined by

$$H_V = -\frac{1}{8\pi}\int_V \dot{T}^4_4\, dX^1\, dX^2\, dX^3, \tag{63}$$

and this energy does not depend on the spatial coordinates used in the evaluation of the integral.[1,11] Taking into account this fact and the Birkhoff theorem, we see that the evaluation of the energy localized outside a spherical source does not require the consideration of its evolutive state and its microscopic structure, but only the use of the outer field described by one of solutions 1, 2, and 3.

In isotropic coordinates, we can use Eqs. (20) and (60)–(63) and the nonvanishing components of $\Gamma^i{}_{jk}$ and $S^i{}_{jk}$ given in a previous paper,[1] to find the following expression giving the total energy localized outside a spherical source with Newtonian mass m and radius R (outer energy):

$$H = -\frac{1}{2}\int_R^{+\infty} r^2 e^{(\alpha+\beta)} \times\left[-\frac{W}{2}\gamma'^2 - 2\beta'\gamma' + \beta'^2 + 2\alpha'\beta'\right]dr. \tag{64}$$

In the case of solutions 1 and 2, taking into account Eqs. (27), (28), (36), and (37), the last formula can be written as follows:

$$H = \frac{1}{2}\int_R^{+\infty}\left(1+\frac{m^2}{4\xi r^2}\right)^{-(2+\xi)}\frac{m^2}{r^2}\left\{1+(1+\xi^{-1})\frac{m}{r} - \left(\frac{(1+2\xi)}{4\xi}\right)\frac{m^2}{r^2}\right\}dr, \tag{65}$$

where

$$\xi = W/(2-W), \quad W \neq 2 \text{ and } W \neq 0.$$

In the case of solution 3, by using Eqs. (45) and (46), we get the following expression giving the outer energy:

$$H = \frac{1}{2}\int_R^{+\infty} e^{-(m^2/4r^2)}\frac{m^2}{r^2}\left(1+\frac{m}{r}-\frac{m^2}{2r^2}\right)dr. \tag{66}$$

This equation can be also derived from Eq. (65) in the limit $|W| \to 2$ ($|\xi| \to \infty$).

Although other complexes can be defined in TA besides the complex (60), we have chosen this one in order to define H due to the following facts: (1) it leads to localized gravitational energy, and (2) it has given satisfactory results in its applications to the treatment of the gravitational radiation problem.[12] However, in spite of the satisfactory properties of this complex, the outer energy defined from it does not have any phyisical significance directly derived from its definition.

An integration of Eq. (65) gives

$$H(R) = (m/2)[(m/R) - 2][1 + (m^2/4\xi R^2)]^{-1-\xi} + m, \tag{67}$$

and taking the limit of the right-hand side of Eq. (67) as $|W| \to 2$ ($|\xi| \to \infty$), we get the following solution of Eq. (66):

$$H(R) = (m/2)[(m/R) - 2]e^{-(m^2/4R^2)}. \tag{68}$$

In the limit $|W| \to \infty$ (Schwarzschild case), Eq. (67) reduces to

$$H(R) = 2m^2/R. \tag{69}$$

This formula was already derived by Shah[18] by using the Møller complex[11] and the Schwarzschild solution in the framework of GR.

Taking into account Eq. (69), we see that when the Schwarzschild radius ($r = m/2$) is reached, all the energy is localized outside the source ($H = m$), i.e., the internal energy $m - H$ vanishes. On account of this fact, Shah[18] writes, "The energy considerations lead to a new interpretation of the event horizon."

Similarly, from Eqs. (67) and (68), it follows that the internal energy $m - H(R)$ vanishes at $R = m/2$ for any positive value of W (solutions 1, 2, and 3) and it does not vanish at any point of its interval of definition for $W < 0$ (solution 1). See Figs. 1–3. In theory A, the radius defined by the condition $H = m$ will be called the "energetic radius" r_e ($r_e = m/2$). Obviously, this radius is characterized by the same condition as the Schwarzschild radius in the Shah interpretation; however, in theory A, neither horizons nor singularities exist at r_e, and, consequently, the energetic radius can be reached in a physically consistent manner.

In accordance with the results obtained in Sec. VI, the radius r_e (as well as any finite radius) will be reached in a finite time by a freely infalling particle; however, in the case of a spherical star with a realistic equation of state, the effect of the pressure can be important in order to stop the fall. If,

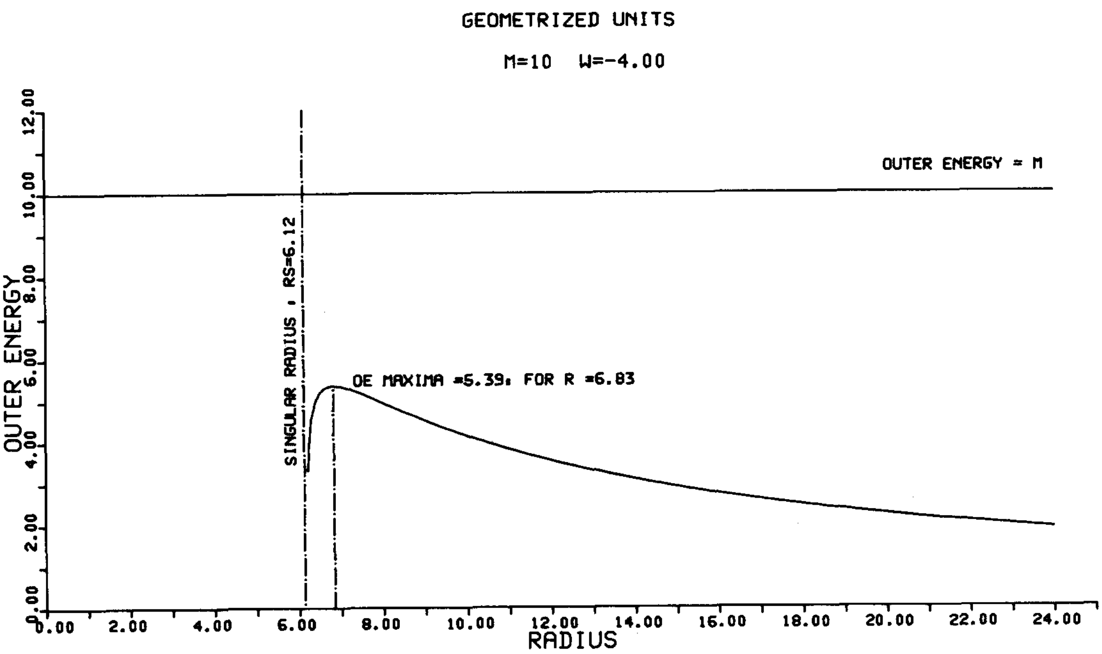


FIG. 1. Outer energy as a function of the isotropic radius R, for $W = -4$ and $M = 10$.

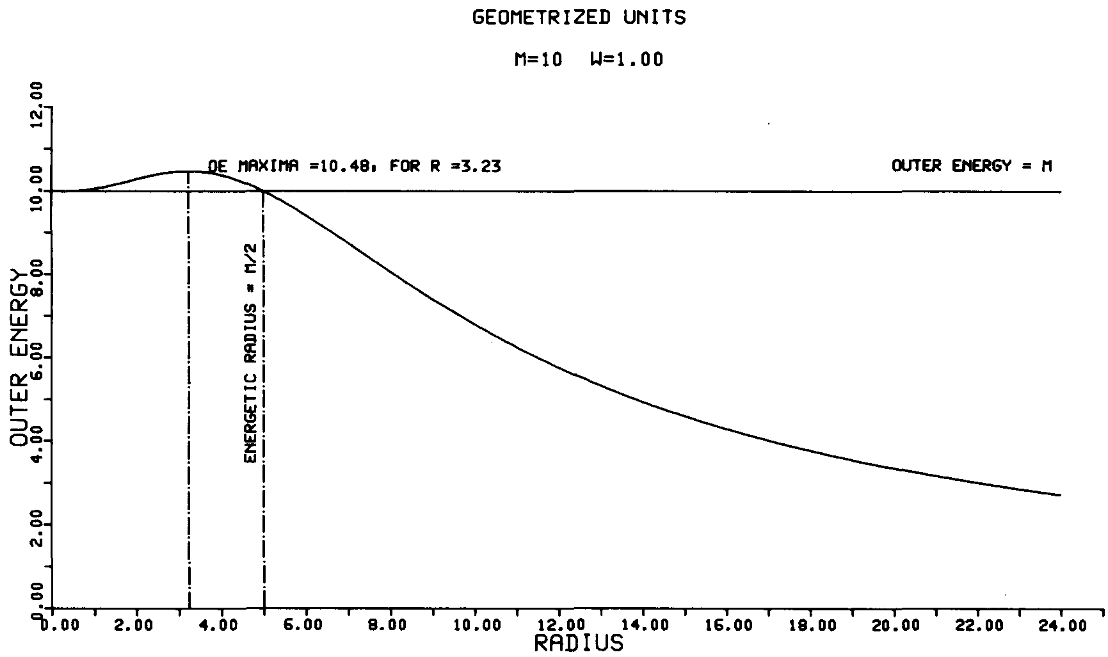


FIG. 2. Outer energy as a function of the isotropic radius R, for $W = 1$ and $M = 10$.

in the realistic case, either the energetic radius or another close radius are reached in a finite time, then, taking into account the similar definitions of the Schwarzschild and energetic radius ($H = m$), it is reasonable to hope the internal structure of a spherical star whose radius is close to the energetic one (in TA) will be similar to the internal structure of a object whose radius is close to the Schwarzschild one (in GR). Then, if the mass of the star is large enough, we will hope that it reaches the energetic radius in a finite time (as occurs with the Schwarzschild radius in GR) forming a small dense physically admissible object, which could belong to some binary x-ray source. However, these suggestions only can be confirmed or rejected by way of a direct study of the gravitational infalling of a realistic perfect fluid sphere.

The internal energy $m - H$ can be split as follows:

$$m - H = m_0 + E_i + H_i, \tag{70}$$

m_0 being the rest mass energy, E_i the nongravitational internal energy, and H_i the gravitational energy localized inside the spherical source. Since H_i and E_i cannot be calculated

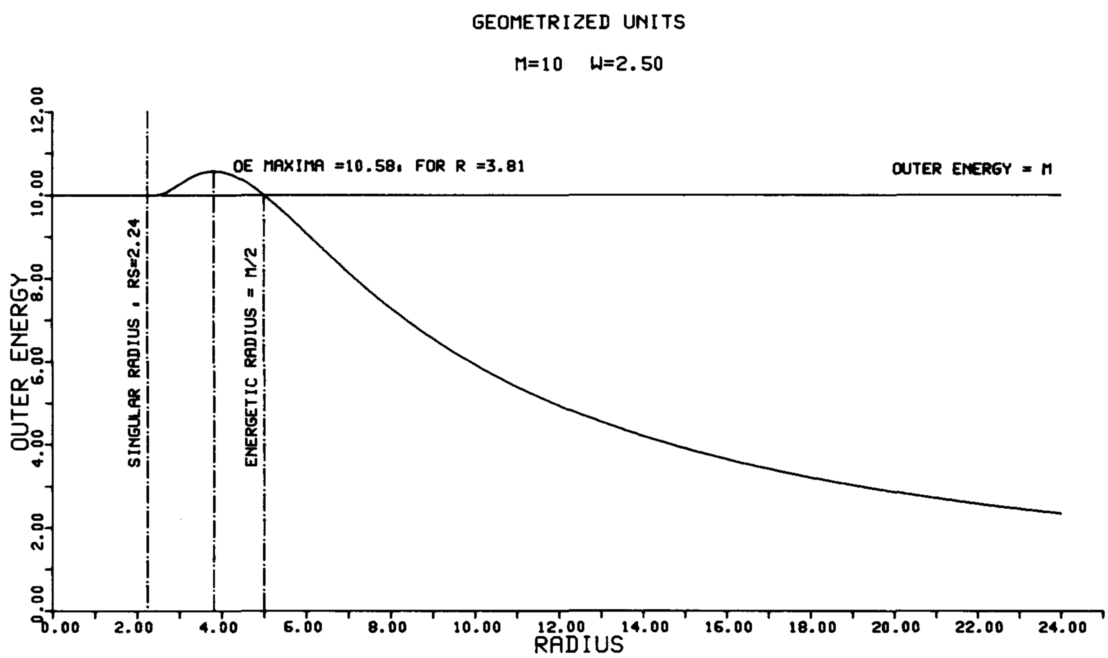


FIG. 3. Outer energy as a function of the isotropic radius R, for $W = 2.50$ and $M = 10$.

without considering the internal structure of the source, we cannot relate H (or $m - H$) with any physical magnitude [as the binding energy $m - m_0$ or the gravitational potential energy $H_i + H$ (see Ref. 15)] by only considering the outer field. Thus, we cannot derive definitive conclusions from the study of the curve $H = H(R)$ displayed in Figs. 1–3. Although these figures correspond to fixed values of m and W, Fig. 1 gives the qualitative behavior of $H = H(R)$ for arbitrary m and $W < 0$, Fig. 2 for any value of m and $0 < W < 2$, and Fig. 3 for arbitrary m and $W > 2$. In any of these figures we get a maximum corresponding to a radius r_M and in the particular case of Fig. 1 we get a vertical asymptote at r_S. Taking into account the above considerations, we cannot interpret the maximum and the asymptote in a definitive manner, but it is reasonable to hope that these elements will have a physical interpretation when the internal structure of the object generating the outer field is taken into account.

In this paper we are only concerned with the information and the prospects obtained from the study of the outer field of a spherical star. These prospects will be studied in some future papers (in progress).

VIII. CONCLUSIONS

We have proved the Birkhoff theorem in the scalar-tetradic theory A; thus, all vacuum spherically symmetric fields are static.

We have also obtained a physical singularity at $r_S \neq 0$ in the case of solution 1, whereas solutions 2 and 3 have been proved to be free of singularities.

The metric of solution 1 is well defined except at the singular radius r_S, while the metrics of solutions 2 and 3 are well defined at any radius $r \neq 0$. Therefore, there are no event horizons in the physical vacuum static spherically symmetric solutions of the field equations of TA.

As a consequence of the above results, we have discarded the values of W as being much too close to zero because they lead to large singular radii, and we have concluded that the black holes are not possible in TA.

Finally, the following prospects are motivated by the arguments presented in this paper.

(i) It would be interesting to study the gravitational infalling of a spherical star (by using an adequate equation of state and suitable boundary conditions) in order to discuss both (1) the validity of solution 1 for any value of W (singular radius), and (2) the physical significance of the energetic radius, the maximum of the curves $H = H(R)$, and the vertical asymptote displayed in Fig. 1.

(ii) It would be worthwhile to study the possible extension of the concept of a "gray hole" (or another similar concept) to TA, and its possible incidence in the explanation of the observational evidences about binary x-ray sources.

[1]D. Sáez, Phys. Rev. D **27**, 2839 (1983).
[2]C. Møller, K. Dan. Vidensk. Selsk. Mat. Fys. Medd. **39**, 1 (1978).
[3]D. Sáez, Gen. Relativ. Gravit. **16**, 501 (1984).
[4]D. Sáez, Gen. Relativ. Gravit. **17**, 39 (1984).
[5]H. Meyer, Gen. Relativ. Gravit. **14**, 531 (1982).
[6]F. W. Hehl, Y. Neemann, J. Nitsch, and P. von der Heyde, Phys. Lett. B **78**, 102 (1978).
[7]C. Brans and R. H. Dicke, Phys. Rev. **124**, 925 (1961).
[8]H. Nariai, Prog. Theor. Phys. **47**, 118 (1972).
[9]H. Nariai, Prog. Theor. Phys. **47**, 832 (1972).
[10]R. H. Dicke, Astrophys. J. **152**, 1 (1968).
[11]C. Møller, K. Dan. Vidensk. Selsk. Mat. Fys. Medd. **1**, 1 (1961).
[12]D. Sáez, "On the scalar-tetradic theories of gravitation: Cosmology and gravitational radiation," Gen. Relativ. Gravit. (in press).
[13]D. Sáez, "On the scalar-tetratic theory A. Some static spherically symmetric solutions," Phys. Rev. D. (submitted for publication).
[14]M. Schweizer, N. Straumann, and A. Wipf, Gen. Relativ. Gravit. **12**, 951 (1980).
[15]CH. W. Misner, K. S. Throne, and J. A. Wheleer, *Gravitation* (Freeman, San Francisco, 1973).
[16]R. E. Clapp, Phys. Rev. D **7**, 345 (1973).
[17]H. Yilmaz, Gen. Relativ. Gravit. **6**, 269 (1975).
[18]K. B. Shah, Proc. Indian Acad. Sci. Sect. A **64**, 327 (1966).

From spinor structure to magnetic monopoles

Krystyna Bugajska[a)]
Silesian University, Katowice, Poland

(Received 18 December 1979; accepted for publication 18 April 1985)

Starting with a spinor structure over a space-time, to which a more fundamental character is ascribed than to the vector structure, an affine-metric geometry of the space-time is defined. The U(1) bundle of electromagnetism appears naturally in this approach. A "sublocal" action of the U(1) on the spinor structure is found. Any field of timelike directions creates a "field" of magnetic monopoles.

I. INTRODUCTION

From the early days of general relativity, models of a unified field theory of gravitation and electromagnetism have been proposed. The trouble with Einstein's relativistic theory of the nonsymmetric field $g_{\mu\nu}$ as well as with Weyl's theory of the uniform change of scale are very well known.[1] In these an attempt was made to introduce the electromagnetic interaction by a change of the geometry of space-time itself. In the currently accepted theories, a good model for classical electromagnetism is provided by a connection on a principal U(1) bundle over a space-time M. However, we must consider this U(1) bundle as an abstractly given principle bundle.

In this paper we shall show that the electromagnetic interaction appears quite naturally when we start with a spinor structure of the space-time manifold. In this case, spinor fields can be treated as fundamental quantities, and geometrical properties of the universe can be determined by the geometry of the spinor structure. The key point is that a spinor connection, which in the most general case produces the Einstein, or the Einstein–Cartan model of space-time, has the SL(2,$\mathbb{C}$)$\times$U(1) group as its holonomy group. This is demonstrated in the Appendix. A simple method to handle spinor connections, developed in the Appendix, is applied there as well to classify vector connections as to consider the most general case, when connections on half-spinor spaces are unrelated.

This paper is organized as follows.

In Sec. II a simple derivation of the known[2] result of Ehlers, Pirani, and Schild is presented. We demonstrate also that exactly the same structure of space-time is implied by "observations" of free falls of classical spinning particles. This follows from the fact that a spinor connection produces a vector connection. In the physically most interesting case of the Einstein (or the Einstein–Cartan) model of the universe, the holonomy bundle of a spinor connection is the Whitney sum of two principal bundles: the SL(2,$\mathbb{C}$) bundle and the U(1) bundle. The former is considered[3] as the underlying structure for the gravitational gauge, and the latter we shall regard as the underlying structure for the electromagnetic gauge. Such a strong relation between the existence of a spinor structure M and the possibility of introducing the electromagnetic gauge seems physically justified, as all experimental data indicate that spin can be detected only by means of electromagnetic fields.

The natural way we have obtained the underlying structure of the electromagnetic gauge is in agreement with the spirit of investigations of Einstein and Weyl who wanted to relate the electromagnetic interaction with geometrical properties of space-time. The only difference is that the vector structure over M is replaced here by the spinor structure to which we ascribe a more fundamental character.

In Sec. III it is demonstrated that the interpretation of the U(1) bundle, obtained in Sec. II as the underlying structure for electromagnetism, implies a "sublocal" action of this U(1) on the spinor structure, that is to a U(1) fibering $\Sigma \to W_+$ of spinor space Σ over the upper part of the light cone. This leads in a straightforward way to the Hopf fibering $S_3 \to S_2$, realized as $\Sigma / R^+ \to W_+/R_+$, at any point of space-time. Now any timelike direction at a point of space-time determines a map $i: PW_+ \to W_+$, hence produces a natural connection, which on a physical ground is identified as a potential of a magnetic pole.[4,5] It should be noted that this pole is placed in the affine tangent space $A_m M$, $m \in M$, at a point different than (0,0,0,0).

There are a few possible physical interpretations of this formal result. We mention two of them.

(1) The timelike direction, which plays a crucial role in our construction, can be determined by a physical timelike vector field. Any such a field could then create a cloud of magnetic poles. If this vector field were the field of Weinberg mesons, we would obtain a kind of "geometrical coupling" between weak and electromagnetic interactions.

(2) The timelike direction could be related to an observer placed at the considered point of space-time. Thus any observer "dresses" the abstract Hopf fibering in a concrete numerical potential of a magnetic pole.

The magnetic pole is placed at (1,0,0,0) in the affine tangent space $A_m M$, $m \in M$. The appearance of the monopole at the point of $A_m M$ different than $m \in M$ could explain why magnetic monopoles cannot be observed.

II. SPINOR CONNECTION AND ELECTROMAGNETIC GAUGE

Let us assume that space-time is a smooth, connected, paracompact, Hausdorff four-dimensional manifold M. One can define vector and tensor fields on M, because the notion of tangent space at each point m of M is natural. However, to

[a)] Present address: Institut für Theoretische Physik, Technischen Universität Clausthan, 3392 Clausthal-Zellerfeld, Federal Republic of Germany.

differentiate a vector field, an affine structure (that is, a linear connection) is needed.

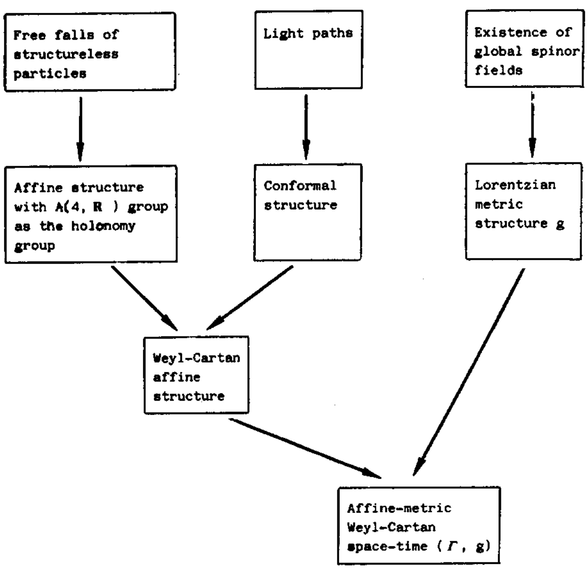
Ehlers, Pirani, and Schild proposed[2] to characterize the geometry of space-time by means of observations of paths of light rays and of structureless massive test particles.[6] The paths of light rays determine a light cone at each point. It is well known that the most general group that (beyond singularities) leaves invariant the light cone is the conformal group. The observations of the paths of structureless massive test particles determine the affine structure Γ on space-time. In the most general case, Γ has the affine group A(4,$\mathbb{R}$) as the holonomy group of its affine connection. Because the inhomogeneous Weyl group is the largest subgroup of A(4,$\mathbb{R}$) contained in the conformal group, we obtain that the light rays together with free falls of structureless particles define the Weyl–Cartan space as the most general geometrical framework for space-time.

On the other hand, the assumption of the existence of global spinor fields implies the existence, at any point m of M, of the two-dimensional spinor space $\Sigma(m)$ with SL(2,$\mathbb{C}$) as the symmetry group. The resulting vector bundle $\mathsf{E} = \cup_{m\in M}\Sigma(m)$ determines[7] some principal SL(2,$\mathbb{C}$) bundle over M. The last bundle can be viewed as a bundle of spinor frames over M. Making use of the known isomorphism

$$\chi : \Sigma(m) \otimes_H \overline{\Sigma}(m) \rightarrow T_m M \qquad (2.1)$$

between the Hermitian part of the tensor product $\Sigma(m) \otimes \overline{\Sigma}(m)$ and the tangent space at $m \in M$, we get the $\mathscr{L}_0$ bundle of linear frames, which in turn determines the Lorentzian metric structure g of the space-time.

Thus the geometrical structure of the space-time has to be the affine-metric[8] (Γ,g) one, with the most general Weyl–Cartan affine structure. In this case we deal with some special kind of nonmetricity, which causes a uniform change of scale in parallel transformed frames:



The existence of global spinor fields on the general space-time manifold M has much stronger consequences than would be suggested by the above considerations. To study them we shall assume throughout the remaining part of this paper the existence of the spinor bundle as the starting point. In other words we assume the following facts.

(1) Space-time is a smooth, connected, paracompact four-manifold M.

(2) There exists a two-dimensional complex vector bundle E over M equipped with a skew-symmetric inner product which we shall call ϵ structure, where

$$\epsilon = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

It means that the principal GL(2,$\mathbb{C}$) bundle $\xi_{\mathrm{GL}(2,\mathbb{C})}$, which is equivalent to the vector bundle E[10] (see Ref. 9), is reducible to the SL(2,$\mathbb{C}$) subbundle. This reduced principal bundle $\xi_{\mathrm{SL}(2,\mathbb{C})}$ can be viewed as a bundle of "canonical" spinor frames of E.

(3) There exists an isomorphism χ between the Hermitian part of the tensor product $\mathsf{E} \otimes \overline{\mathsf{E}}$ and the tangent bundle TM, i.e., $\forall m \in M$,

$$\chi_m : \Sigma(m) \otimes_H \overline{\Sigma}(m) \rightarrow T_m M,$$

where $\Sigma(m)$ is the fiber of E at $m \in M$.

Assumptions (2) and (3) imply a metric structure on M, i.e., determine a principal bundle $\xi_{\mathscr{L}_0}$ of orthonormal frames over M. Besides, we obtain immediately that the just-considered-in-(2) principal bundle $\xi_{\mathrm{SL}(2,\mathbb{C})}$ is exactly the prolongation of the bundle $\xi_{\mathscr{L}_0}$ in the Milnor–Lichnerowicz sense. For this reason we shall make some calculation in the Appendix in this more traditional approach (of Milnor–Lichnerowicz).

Of course we can make assumptions (2) and (3) only when the second Stiefel–Whitney class $w_2 \in H^2(M;Z_2)$ is equal to zero. The vanishing of w_2 is the property of the manifold M itself and not a metric structure on M. Hence for any four-manifold which admits Lorentz structure (i.e., which Euler class is equal to zero) and for which $w_2 = 0$ we can take a concrete two-dimensional complex vector bundle E with properties (2) and (3) instead of a concrete metric structure g on M.

Now, to make our considerations more clear, let us consider an affine-metric structure of M again. The metric g is equivalent to the bundle of orthonormal frames $\xi_{\mathscr{L}_0}$ over M. It means that the principal GL(4,R) bundle $\xi_{\mathrm{GL}(4,R)}$, which is equivalent[9] to the tangent bundle TM, is reducible to the Lorentz group or that TM can be given as associated bundle to $\xi_{\mathscr{L}_0}$; $TM = \xi_{\mathscr{L}_0}[R^4]$. Now we introduce, in addition to the metric structure g on M, an affine structure Γ (i.e., independently of the Levi-Civita connection $\left\{ {\mu \atop \rho\sigma} \right\}$ a linear connection Γ). In a general case with trace and traceless parts of nonmetricity[8] we obtain the whole $\xi_{\mathrm{GL}(4,R)}$ bundle as the holonomy bundle.

Our case of a spinor bundle E is exactly the same situation. The two-dimensional complex vector bundle E over M is equivalent to some principal GL(2,$\mathbb{C}$) bundle $\xi_{\mathrm{GL}(2,\mathbb{C})}$ over M. The skew inner product on E, i.e., ϵ structure is equivalent to the reduction of $\xi_{\mathrm{GL}(2,\mathbb{C})}$ to the spin bundle $\xi_{\mathrm{SL}(2,\mathbb{C})}$. Hence E can be written as $\mathsf{E} = \xi_{\mathrm{SL}(2,\mathbb{C})}^{[\mathbb{C}^2]}$. But again, in addition to the ϵ structure on E (which by χ produces a metric structure g on M), we can introduce a general connection on E (which by χ produces an affine structure Γ on M). Similarly as above, the most general connection that can be introduced on E has the $\xi_{GL(2,C)}$ bundle as its holonomy bundle. Because of this (using χ), we can produce a linear connection Γ on M only with the Weyl group as its holonomy group.

We shall assume the existence of spinor connections acting on the half-spinor spaces and interrelated by the complex conjugation [see (A11)–(A13)]. A spinor connection could be imagined as a result of "spin observations" of free falls of the classical spinning test particle.

The assumption that spinor structure of space-time is of a primary nature allows us to deduce the geometrical properties of space-time from the geometry of spinor structure.

Indeed, for every spinor connection on $E = \cup_{m \in M} \Sigma(m)$, the isomorphism χ [see (2.1), (A1), and (A2)] determines a vector connection (an affine structure Γ) on the space-time M. As we can see in the Appendix, the most general spinor connection produces the Weyl–Cartan vector connection. We conclude that the most general geometrical structure of space-time determined by the "observations" of classical free spin fields is the Weyl–Cartan affine-metric structure (Γ, g). This structure is exactly the same as the one obtained from observations of light paths, free structureless particles, and from the existence of global spinor fields. So, we have

$$\boxed{\begin{array}{l}\text{Spinor}\\ \text{structure}\end{array}} + \boxed{\begin{array}{l}\text{Spinor}\\ \text{connection}\end{array}} \rightarrow \boxed{\begin{array}{l}\text{Weyl–Cartan}\\ \text{affine-metric}\\ \text{space-time } (\Gamma, g)\end{array}} \; .$$

Let us notice that the Einstein (or Einstein–Cartan) model of space-time can be obtained from a spinor connection with the holonomy group equal to $SL(2,\mathbb{C}) \times U(1)$ (see the Appendix). The appearance of U(1) is caused by the simple fact that if we transform the space $\Sigma(m)$ by an element $a(m) \in U(1)$ before applying the isomorphism χ we obtain exactly the same result as when we apply the isomorphism χ alone. It means that the diagram

$$\begin{tikzcd} \Sigma(m) \otimes_H \bar{\Sigma}(m) \arrow[rr, "a"] \arrow[dr, "\chi"'] & & \Sigma(m) \otimes_H \bar{\Sigma}(m) \arrow[dl, "\chi"] \\ & T_m M & \end{tikzcd}$$

is commutative [the action of $a \in U(1)$ is given by (A4) and (A5)]. Obviously, we can choose different $a(m) \in U(1)$ in different points of the space-time M, but we must multiply all elements of $\Sigma(m)$ by the same $a(m)$.

The holonomy bundle of the spinor connection may be treated as a Whitney sum of two principal bundles: an $SL(2,\mathbb{C})$ bundle and a U(1) bundle. The former is considered as an underlying structure for the gravitational gauge, and the latter we shall regard as the underlying structure for the electromagnetic gauge. Such a strong relation between the existence of spinor structure on M and the possibility of introducing the electromagnetic gauge seems physically justified as all experimental data indicate that spin can be detected only by means of electromagnetic fields.

III. MAGNETIC MONOPOLES

It has been shown in the Appendix that the holonomy group of a spinor connection, which determines the Einstein model of space-time, is the direct product of the groups $SL(2,\mathbb{C})$ and U(1). So the spinor holonomy bundle is the Whitney sum of two principal bundles: the $SL(2,\mathbb{C})$ bundle and the U(1) bundle. This latter will be treated as an underlying structure of the electromagnetic gauge. As we have seen, this U(1) principal bundle appears naturally, because the relation between the Hermitian part of $\Sigma(m) \otimes \bar{\Sigma}(m)$ and $T_m M$, as given by χ, does not change as we multiply all spinors of $\Sigma(m)$ by the same element $a(m)$ of U(1).

Now let us restrict our attention to the diagonal part $\mathscr{D}$ of $\Sigma(m) \otimes_H \bar{\Sigma}(m)$. Obviously, $\mathscr{D} \simeq \Sigma(m)$. So we have

$$\Sigma(m) \ni u \xrightarrow{j} u \otimes u^* \xrightarrow{\chi} x \in W_+(m), \tag{3.1}$$

where $W_+(m)$ denotes the upper part of the light cone of $T_m M$. Because $\chi|_{\mathscr{D}}$ is invariant under the "sublocal" (that is dependent on $x \in W_+$) transformations $a(x) \in U(1)$, it is easy to see that the space $\Sigma(m)$ has a structure of the U(1) principal bundle over $W_+(m)$ with projection $\pi: \Sigma(m) \to W_+$ equal to

$$\pi = \chi|_{\mathscr{D}} \circ j. \tag{3.2}$$

Indeed, let $\{e_0, e_1, e_2, e_3\}$ be the orthonormal frame of $T_m M$ defined as in the Appendix, and let

$$\{\rho, \sigma\} = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$$

be the corresponding canonical spinor frame of $\Sigma(m)$. Also, let

$$\zeta = e_0 + e_3 = \chi(2\sigma \otimes \sigma^*), \quad \eta = e_0 - e_3 = \chi(2\sigma \otimes \sigma^*). \tag{3.3}$$

The vectors ζ and η belong to $W_+ \subset T_m M$. Now let $U_\eta \subset W_+$ denote the image of spinors $\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$ for $z_1 \neq 0$, and let $U_\zeta \subset W_+$ denote the image of spinors $\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$ for $z_2 \neq 0$, under the projection π.

It is easy to see from (A2) that each spinor

$$u = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \in \Sigma(m)$$

determines a light vector $x \in W_+$ with components (x_0, x_1, x_2, x_3) in the base $\{e_0, e_1, e_2, e_3\}$ of $T_m M$ given by

$$\frac{1}{2} \begin{pmatrix} x_0 + x_3 & x_2 - ix_1 \\ x_2 - ix_1 & x_0 - x_3 \end{pmatrix} = \begin{pmatrix} z_1 \bar{z}_1 & z_1 \bar{z}_2 \\ \bar{z}_1 z_2 & z_2 \bar{z}_2 \end{pmatrix}. \tag{3.4}$$

We see that $U_\zeta \cup U_\eta = W_+$, and that local trivializations of the U(1) bundle $\Sigma(m)$ over W_+ are given by local cross sections

$$h_\zeta(x) = \sqrt{\frac{x_0 + x_3}{2}} \begin{pmatrix} 1 \\ (x_2 + ix_1)/(x_0 + x_3) \end{pmatrix}, \tag{3.5}$$

$$h_\eta(x) = \sqrt{\frac{x_0 - x_3}{2}} \begin{pmatrix} (x_2 - ix_1)/(x_0 - x_3) \\ 1 \end{pmatrix}, \tag{3.6}$$

and

$$\begin{aligned} &h_\zeta : U_\zeta \to \pi^{-1}(U_\zeta) \subset \Sigma(m), \\ &h_\eta : U_\eta \to \pi^{-1}(U_\eta) \subset \Sigma(m). \end{aligned} \tag{3.7}$$

For $x \in U_\zeta \cap U_\eta$, we have

$$h_\eta(x) = h_\zeta(x)(x_2 - ix_1)/\sqrt{x_1^2 + x_2^2}, \tag{3.8}$$

so the transition function $g_{\eta\zeta}(x)$ belongs to U(1).

Now let us consider the projective space of null directions PW_+ and the injective map $i:PW_+ \to W_+$, defined by a section of W_+ by the spacelike hyperplane $x_0 = 1$. *The map* i allows us to obtain immediately the induced U(1) bundle $i^*(\Sigma(m))$ over the space of null directions.

It is easy to see that the bundle space of $i^*(\Sigma(m))$ is formed by the set of spinors $\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \in \Sigma(m)$ with the property

$$|z_1|^2 + |z_2|^2 = 1. \tag{3.9}$$

Thus $i^*(\Sigma(m))$ can be viewed as the three-dimensional sphere S_3. The group U(1) acts on S_3 by

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} a = \begin{pmatrix} z_1 a \\ z_2 a \end{pmatrix}, \text{ for } a \in \mathrm{U}(1), \tag{3.10}$$

hence we obtain the Hopf fiber bundle

$$S_3 \to S_2. \tag{3.11}$$

The base space of this bundle is the set of null vectors which form a sphere S_2 with center at the point (1,0,0,0) with respect to m. (Coordinates of these vectors satisfy the relation $x_1^2 + x_2^2 + x_3^2 = 1$.)

The bundle space of the Hopf fibering $i^*(\Sigma)$ lies in $\Sigma(m)$. The Bacry–Kihlberg[10] isomporphism provides a parametrization of this bundle space by means of the Euler angles $(\varphi, \vartheta, \psi)$. Indeed, a general element Λ of SL(2,C) can be represented as

$$\Lambda = e^{-i\varphi L_{12}} e^{-i\vartheta L_{31}} e^{-i\psi L_{12}} e^{ipL_{03}} e^{-it_1(L_{01} + L_{31})} e^{-it_2(L_{02} - L_{23})}, \tag{3.12}$$

where L_{01}, L_{02}, L_{03} are boosts, i.e., generators of proper Lorentz transformations, while L_{23}, L_{31}, L_{12} are operators of angular momentum, i.e., generators of SU(2).

Now, if we take into account that $\Sigma(m) \simeq \mathrm{SL}(2,\mathrm{C})/\mathscr{C}$, where $\mathscr{C}$ is a two-dimensional group parametrized by t_1 and t_2 (the Crumeyrolle group),[11] we obtain the parametrization of spinors $u = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \in \Sigma(m)$:

$$\begin{aligned} z_1 &= e^{p/2} e^{(i/2)(\psi + \varphi)} \cos(\vartheta/2), \\ z_2 &= e^{p/2} e^{(i/2)(\psi - \varphi)} \sin(\vartheta/2), \end{aligned} \tag{3.13}$$

where

$$e^{p/2} = \sqrt{z_1 \bar{z}_1 + z_2 \bar{z}_2}. \tag{3.14}$$

For a spinor belonging to the considered bundle space S_3 we have

$$\begin{aligned} z_1 &= e^{i/2}(\psi + \varphi) \cos(\vartheta/2), \\ z_2 &= e(i/2)(\psi - \varphi) \sin(\vartheta/2). \end{aligned} \tag{3.15}$$

This is the needed parametrization by the Euler angles. Thus the natural connection carried by the Hopf fibering may be conveniently expressed in terms of theEuler angles. Because the Riemannian line element on the sphere S_3 is given by

$$\begin{aligned} dl^2 &= d\bar{z}_1\, dz_1 + d\bar{z}_2\, dz_2 \\ &= \tfrac{1}{4}[d\vartheta^2 + \sin^2\vartheta\, d\varphi^2 + (d\psi + \cos\vartheta\, d\varphi)^2], \end{aligned} \tag{3.16}$$

we see that the form

$$\alpha = \tfrac{1}{2}(d\psi + \cos\vartheta\, d\varphi) \tag{3.17}$$

defines a connection on S_3 considered as a circle bundle over S_2.

It is known, that the natural connection carried by the Hopf fibering $i^*(\Sigma): S_3 \to S_2$ describes the magnetic monopole.[4] The singularities of the potential of the magnetic pole are due to the nontrivial character of the bundle $S_3 \to S_2$. The curvature F of the connection α, $F = \frac{1}{2}\sin\vartheta\, d\varphi \wedge d\vartheta$, extended to the Minkowski space $A_m M$ is the "electromagnetic field" of a magnetic pole of strength $q = \frac{1}{2}$, placed at the point (1,0,0,0).

Let us set

$$V_\xi = U_\xi \cap S_2, \quad V_\eta = U_\eta \cap S_2, \tag{3.18}$$

where U_ξ and U_η were defined earlier. The corresponding local sections are

$$\begin{aligned} R_\xi(x) &= h_\xi / V_\xi(x) \\ &= \sqrt{\frac{1 + x_3}{2}} \begin{pmatrix} 1 \\ (x_2 + ix_1)/(1 + x_3) \end{pmatrix} \\ &= \begin{pmatrix} \cos(\vartheta/2) \\ e^{-i\varphi} \sin(\vartheta/2) \end{pmatrix} \end{aligned} \tag{3.19}$$

(where $\vartheta \neq \pi$, or equivalently $x_3 \neq -1$), and

$$\begin{aligned} R_\eta(x) &= h_\eta / V_\eta(x) \\ &= \sqrt{\frac{1 - x_3}{2}} \begin{pmatrix} (x_2 - ix_1)/(1 - x_0) \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} e^{i\varphi} \cos(\vartheta/2) \\ \sin(\vartheta/2) \end{pmatrix} \end{aligned} \tag{3.20}$$

(where $\vartheta \neq 0$ or equivalently $x_3 \neq -1$). The potential A in the gauge R_ξ is

$$(A)_\xi = R^*(\alpha) = -\tfrac{1}{2}(1 - \cos\vartheta) d\varphi \tag{3.21}$$

and its essential component with respect to an orthonormal affine frame $\{p; e_0, e_1, e_2, e_3\}$ of the space-time manifold M [$p = (1,0,0,0)$ in $\{0; e_0, e_1, e_2, e_3\}$] is

$$(A_\varphi)_\xi = (1 - \cos\vartheta)/2r \sin\vartheta). \tag{3.22}$$

For the gauge R_η we obtain similarly

$$(A)_\eta = \tfrac{1}{2}(1 + \cos\vartheta) d\varphi \tag{3.23}$$

and

$$(A_\varphi)_\eta = (1 + \cos\vartheta)/(2r \sin\vartheta). \tag{3.24}$$

It is known that formulas (3.22) and (3.24) describe the field if it is a magnetic monopole.

Let us summarize the surprising picture we have just obtained.

The fundamental character we have ascribed to the spinor structure on the space-time manifold M has produced the additional U(1) bundle over M. The appearance of this bundle is caused by the natural possibility of multiplying spinors of $\Sigma(m)$ by the same phase factor without influencing the relation between pairs of spinors and vectors described by χ of (2.1), (A1), and (A2). Thus the transformations of U(1) act locally on the spinor structure E. If we limit the domain of χ to get only light vectors [see (3.1)] we may multiply spinors of $\Sigma(m)$ by different phase factors without any change of the mentioned relation given by χ. Thus we obtain a "sublocal" action of U(1) on E. Now any field of timelike directions determines maps $i{:}PW_+ \to W_+$, at each point $m \in M$, which leads in a straightforward way to the Hopf fibering $i^*(\Sigma)$.

If we identify the U(1) bundle on M with the underlying structure for electromagnetism, then we can understand the "sublocal" action of U(1) as well the resulting Hopf fibering as a "substructure" of the electromagnetic interaction. This seems to suggest the interpretation of fields A_φ of (3.22) and (3.24) as the field of magnetic monopole placed in $A_m M$ at (1,0,0,0) with respect to m. The appearance of the monopole at the point of $A_m M$ different than $m \in M$ could explain why magnetic monopoles cannot be observed.

ACKNOWLEDGMENTS

I wish to thank Professor S. Sternberg for his kind hospitality at Harvard University, as well as for reading the manuscript in the first draft of this paper.

I would also like to thank Professor P. Budinich for his hospitality at the Scuola Internazionale Superiore di Studi Avanzati, Trieste.

APPENDIX: SPINOR CONNECTION

Let M be a four-dimensional pseudo-Riemannian manifold admitting a spinor structure. To differentiate vector fields one needs a connection. It is known, that a linear vector connection Γ on M defines a spinor connection only if this linear connection is a metric one. It has been shown that a unique spinor connection is determined by the requirement that it should be compatible with the spinor inner product.

On the other hand, two-component spinors (half-spinors) may be used to form vectors. Hence we can start with a spinor connection, and then produce a vector connection from it. A detailed description of this subject can be found in the papers of Luehr and others.[12]

Here we shall treat this problem in a slightly different way. Owing to this we can (1) classify in a very simple manner those vector connections which are determined by spinor connection, and (2) consider the most general case, when connections on the involved half-spinor spaces are unrelated.

To begin with, let us recall some known facts. The isomorphism

$$\chi{:}\Sigma \otimes \overline{\Sigma} \to E \tag{A1}$$

is given by

$$e_0 = \chi(\rho \otimes \rho^* + \sigma \otimes \sigma^*), \quad e_1 = (1/i)\chi(\sigma \otimes \rho^* - \rho \otimes \sigma^*),$$
$$e_2 = \chi(\sigma \otimes \rho^* + \rho \otimes \sigma^*), \quad e_3 = \chi(\rho \otimes \rho^* - \sigma \otimes \sigma^*), \tag{A2}$$

where e_0, e_1, e_2, e_3 is an orthonormal frame of the Minkowski space E, and $\{\rho,\sigma\}$, $\{\rho^*,\sigma^*\}$ are the canonical spinor bases of Σ and $\overline{\Sigma}$, respectively.

Any vector $x \in E$ can be uniquely described by a Hermitian matrix $\hat{x}$:

$$\hat{x} = \begin{pmatrix} x_{\rho\otimes\rho^*} & x_{\rho\otimes\sigma^*} \\ x_{\sigma\otimes\rho^*} & x_{\sigma\otimes\sigma^*} \end{pmatrix} = \frac{1}{2}\begin{pmatrix} x_0 + x_3 & x_2 - ix_1 \\ x_2 + ix_1 & x_2 - x_3 \end{pmatrix}. \tag{A3}$$

An element s of GL(2,C) acts in different ways on the spaces Σ and $\overline{\Sigma}$:

$$su = \Lambda u, \quad \text{for } u \in \Sigma,$$
$$su^* = \overline{\Lambda} u, \quad \text{for } u \in \overline{\Sigma}, \tag{A4}$$

according to

$$(su)^* = su^*. \tag{A5}$$

(Here Λ is a nonsingular complex matrix 2×2.) When we transform Σ and $\overline{\Sigma}$ by $s \in \mathrm{GL}(2,\mathbb{C})$, then the appropriate transformation of $x \in E$ is given by

$$\hat{x} \to \hat{x}' = \Lambda \hat{x} \Lambda^+ = (\hat{x}')^+. \tag{A6}$$

Now let us see what happens when the space Σ is transformed by $s_1 \in \mathrm{GL}(2,\mathbb{C})$ and $\overline{\Sigma}$ is transformed by $s_2 \in \mathrm{GL}(2,\mathbb{C})$. We obtain

$$\hat{x}' = \Lambda_1 \hat{x} \Lambda_2^+, \tag{A7}$$

so transformations s_1 of Σ and s_2 of $\overline{\Sigma}$ determine the map

$$f_{s_1 s_2}{:}E \to E^c \tag{A8}$$

of the Minkowski space E into its complexification E^c. It is easy to see that $f_{s_1 s_2}$ defines a linear transformation of the Minkowski space iff $\Lambda_1 \hat{x} \Lambda_2^+$ is a Hermitian matrix. Taking into account that

$$\Lambda_1 \hat{x} \Lambda_2^+ = \Lambda_1 \hat{x} \Lambda_1^+ (\Lambda_1^+)^{-1} \Lambda_2^+, \tag{A9}$$

it is easy to calculate that ths condition is equivalent to

$$\Lambda_2 = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} \Lambda_1, \tag{A10}$$

with $a \in \mathbb{R} \setminus \{0\} = R^*$.

Coming back to our main subject, let us assume that the connection acting on complex conjugate spinors $\overline{\mathsf{E}} = U_{m\in M} \overline{\Sigma}(m)$ (even half-spinors) is given by the complex conjugate of the spinor connection acting on $\mathsf{E} = U_{m\in M} \Sigma(m)$ (odd half-spinors)

$$\overline{\nabla}_X \overline{\psi} = \overline{\nabla_X \psi}, \tag{A11}$$

where X is a vector field on M, ψ is an odd half-spinor field, and $\overline{\psi}$ is the corresponding even half-spinor field.

Let us fix[13] a global section $r_0(m)$ of the bundle of orthonormal frames over space-time M. It allows us to determine uniquely spinor spaces $\Sigma(m)$ and $\overline{\Sigma}(m)$ at each point $m \in M$, and global fields of their canonical spinor frames $\mathscr{S}(m)$ and $\mathscr{S}^*(m)$ [of course $(\mathscr{S}(m))^* = \mathscr{S}^*(m)$, where $*$ is the anti-isomorphism $*{:}\ \Sigma \to \overline{\Sigma}$].

Let us take some path $\tau_{AB} \subset M$, $A,B \in M$. The parallel displacement of the canonical spinor frame $\mathscr{S}(A)$ to the point B along τ_{AB} gives the spinor frame $\mathscr{S}_\tau(B)$. In the general case,

$$\mathscr{S}_\tau(B) = \mathscr{S}(B)\cdot s, \quad \text{with } s \in \mathrm{GL}(2,\mathbb{C}). \tag{A12}$$

The assumption (A11) is equivalent to the fact that the parallel displacement along τ_{AB} joins the spinor $\mathscr{S}^*(A) = (\mathscr{S}(A))^*$ with the spinor frame

$$\mathscr{S}^*_\tau(B) = \mathscr{S}^*(B)\cdot s, \tag{A13}$$

and s is the same as in (A12). [We recall (A1) and (A5).]

We can determine in a usual manner the spinor holonomy bundle through $\mathscr{S}(A)$ as the set of spinor frames of E, which can be joined with $\mathscr{S}(A)$ by a parallel displacement along any piecewise differential curve of the class C^∞ in M.

Because spinors are combined to form vectors by (A1) and (A2), every spinor frame $\mathscr{S}_\tau$ at $B \in M$ determines a linear frame $r(B)$. Thus we see that the parallel displacement of the spinor frame $\mathscr{S}(A)$ along τ_{AB} determines a parallel displacement of the orthonormal frame $r_0(A)$ along τ_{AB}. Taking into account (A12), (A13), and (A6) we see that this displacement defines a linear frame r_τ at B,

$$r_\tau(B) = r_0(B) \cdot g, \tag{A14}$$

with a linear transformation g.

The group GL(2,C) is the direct product of $SL(2,C) \times U(1) \times R^+$, hence we see from (A6) that in the most general case of spinor connection with GL(2,C) as the holonomy group, the element g of (A14) has to belong to the Weyl group. This means, that the most general spinor connection produces a vector connection, which holonomy group can be at most equal to the Weyl group.

Moreover, we see that when s of (A12) and (A13) is an element of $SL(2,C) \times U(1)$ then the parallel displacement of the orthogonal frame $r_0(A)$ is determined by the transformation $g = \kappa(\lambda) \in \mathscr{L}_0$ [cf. (A14)], where λ is the element of SL(2,C) corresponding to s and κ is the covering map of SL(2,C) onto $\mathscr{L}_0$. Thus we obtain that a spinor connection with the holonomy group equal to $SL(2,C) \times U(1)$ produces the Einstein (torsionless) or the Einstein–Cartan structure of space-time.

This approach allows us to consider the case of spinor connections of E and $\overline{\Sigma}$ when the assumption (A11) is rejected. Now the parallel displacement of $\mathscr{S}(A)$ and $\mathscr{S}^*(A) = (\mathscr{S}(A))^*$ along τ_{AB} determines spinor frames $\mathscr{S}_\tau(B)$ and $\mathscr{S}^*_\tau(B)$, but $\mathscr{S}^*_\tau(B) \neq (\mathscr{S}_\tau(B))^*$. Thus, instead of relations (A12) and (A13) we obtain

$$\left.\begin{aligned} &\mathscr{S}_\tau(B) = \mathscr{S}(B) \cdot s, \\ &\mathscr{S}^*_\tau(B) = \mathscr{S}^*(B) \cdot s', \end{aligned}\right\} \quad \text{with } s, s' \in GL(2,C),\ s \neq s'. \tag{A15}$$

If we make use (as previously) of the isomorphism χ [see (A1) and (A2)] to produce a vector connection, we see [from (A8) and (A10)] that it proves possible only if

$$s' = as, \quad \text{with } a \in R^*. \tag{A16}$$

The holonomy group of this connection will be equal to the Weyl group in the general case. If the condition (A16) is not satisfied, then it is easy to check from (A1), (A2), and (A15) that the "parallel displacement" along τ_{AB} induced by the spinor connections transfers real vectors into complex ones.

[1] R. J. Adler, M. J. Bazin, and M. Schiffer, *Introduction to General Relativity* (McGraw–Hill, New York, 1975), 2nd ed.

[2] J. Ehlers, F. A. E. Pirani, and A. Schild, in *General Relativity*, edited by L. O'Raifeartaigh (Claredon, Oxford, 1973), pp. 63–84.

[3] P. G. Bergmann, Phys. Rev. **107**, 624 (1957); M. Carmeli, J. Math. Phys. **11**, 2728 (1970); N. Yang, Phys. Rev. Lett. **33**, 445 (1974); C. J. Isham, A. Salam, and J. Strathdee, Lett. Nuovo Cimento **5**, 969 (1972).

[4] A. Trautman, Int. J. Theor. Phys. **16**, 561 (1977).

[5] T.T. Wu and C. N. Yang, Phys. Rev. D **12**, 3845 (1975).

[6] Trajectories of free spinning particles are not geodetics.

[7] R. M.Switzer, *Algebraic Topology, Homotopy and Homology* (Springer, Berlin, 1970).

[8] F.W. Hehl, G. D. Kerlick, and P. van der Heyde; Z. Naturforsch. **31a**, 111 (1975).

[9] D. Husemoller, *Fibre Bundles* (Springer, Berlin 1970), p. 71; F. Hirzebruch, *Topological Methods in Algebraic Geometry* (Springer, Berlin, 1966), p. 61.

[10] H. Bacry and A. Kihlberg, J. Math. Phys. **10**, 2132 (1969).

[11] A. Crumeyrolle, Ann. Inst. H. Poincaré Sec. A **14**, 309 (1971).

[12] C. P. Luehr and M. Rosenbaum, J. Math. Phys. **15**, 1120 (1974); C. P. Leuhr, M. Rosenbaum, M. P. Ryan, Jr., and L. C. Shepley, J. Math. Phys. **18**, 965 (1977).

[13] The existence of such section has been shown by R. Geroch, J. Math. Phys. **9**, 1739 (1968).

Bianchi type-VI$_0$ solutions in modified Brans–Dicke cosmology

Shri Ram
Applied Mathematics Section, Institute of Technology, Banaras Hindu University, Varanasi 221005, India

(Received 4 March 1985; accepted for publication 17 May 1985)

The spatially homogeneous and anisotropic Bianchi type-VI$_0$ cosmological dust solution for the modified Brans–Dicke theory is presented. Some properties of the model are discussed.

I. INTRODUCTION

As a consequence of great discoveries in the field of astrophysics, cosmology, and elementary particles, several scalar-tensor theories of gravitation have been proposed which result in a time variability of the gravitational constant in conformity with Mach's principle. Amongst the various modifications of the general theory of relativity, the scalar-tensor theory of Brans–Dicke[1] (BD) is most widely accepted. It has also been found that the cosmological constant Λ in Einstein's theory, which arises from spontaneous symmetry breaking, is not a constant.

The investigation of particle physics within the context of the BD Lagrangian has stimulated the study of the cosmological term with a modified BD Lagrangian in cosmology and elementary particle physics. Further, Bergmann[2] and Wagoner[3] have suggested that the cosmological term should be a function of a scalar field ϕ. Endo and Fukui[4] have obtained the Brans–Dicke field equations modified by $\Lambda(\phi)$. Recently, Banerjee and Santos[5] have obtained cosmological dust solutions for a Bianchi type-I homogeneous space in this modified scalar-tensor theory. Since Bianchi type-I models are a very special set of spatially homogeneous models, it is of considerable interest to consider the more general Bianchi type-VI$_0$ space-times to study the large-scale dynamics of the universe.[6]

In this paper the field equations for dust in the Brans–Dicke modified theory in a Bianchi type-VI$_0$ homogeneous space are discussed and a particular solution is obtained representing a spatially homogeneous cosmological model admitting anisotropic expansions. Some properties of the model are discussed.

II. FIELD EQUATIONS

The field equations for the modified Brans–Dicke theory with the introduction of $\Lambda(\phi)$ obtained by Endo and Fukui are[5]

$$G_{ij} + g_{ij}\Lambda = -(k/\phi)T_{ij} - (\omega/\phi^2)(\phi_{,i}\phi_{,j} - \tfrac{1}{2}g_{ij}\phi_{,a}\phi^{,a}) - (1/\phi)(\phi_{;ij} - g_{ij}\Box\phi), \tag{1}$$

$$-\Lambda + \phi\frac{\partial\Lambda}{\partial\phi} = \frac{k}{2\phi}T - \frac{2\omega+3}{2\phi}\Box\phi. \tag{2}$$

Here T_{ij} is the energy-momentum tensor for matter alone and we consider the distributions for dust

$$T_{ij} = \rho V_i V_j, \tag{3}$$

where ρ is the mass density and V^i is the four-velocity. We assume the coordinates to be comoving so that

$$V^1 = V^2 = V^3 = 0, \quad V^4 = 1. \tag{4}$$

It has been further assumed that the matter and scalar fields are related through

$$\Box\phi = k\mu T/(2\omega + 3), \tag{5}$$

where the constant μ shows how much this theory including $\Lambda(\phi)$ deviates from that of Brans–Dicke and as usual ω is the coupling constant. Substituting (5) into (2) we obtain

$$\Lambda - \phi\frac{\partial\Lambda}{\partial\phi} = a\frac{\Box\phi}{\phi}, \tag{6}$$

where a is a constant defined by

$$a = [(2w+3)/2](1/\mu - 1). \tag{7}$$

If Λ is a function of ϕ only, we conclude from (6) that $\Box\phi = f(\phi)$. Following Banerjee and Santos[5] we further assume that the functional relation $f(\phi)$ is of the form $f(\phi) = m\phi^n$, where m and n are arbitrary constants. Then Eq. (6) reduces to

$$\phi\frac{\partial\Lambda}{\partial\phi} - \Lambda = -am\phi^{n-1}. \tag{8}$$

From (8) the solution for Λ is as follows:

$$\Lambda = [am/(2-n)]\phi^{n-1} + D_1\phi, \quad n \neq 2, \tag{9}$$

and

$$\Lambda = -am\ln\phi + D_2\phi, \quad n = 2,$$

D_1 and D_2 being integration constants.

III. SOLUTION OF FIELD EQUATIONS

The line element for the spatially homogeneous Bianchi type-VI$_0$ can be written as

$$ds^2 = -dt^2 + A^2(t)dx^2 + B^2(t)e^{-2qx}dy^2 + C^2(t)e^{2qx}dz^2, \tag{10}$$

where A, B, and C are cosmic scale functions, and q is a nonzero constant. We number the coordinates x, y, z, and t as 1, 2, 3, and 4, respectively. The nonzero components of the field equations (1) for (9) are

$$G^1_1 = \frac{B_{44}}{B} + \frac{C_{44}}{C} + \frac{B_4C_4}{BC} + \frac{q^2}{A^2} = -\Lambda - \frac{\omega}{2}\left(\frac{\phi_4}{\phi}\right)^2 + \frac{A_4}{A}\frac{\phi_4}{\phi} + \frac{\Box\phi}{\phi}, \tag{11}$$

$$G^2_2 = \frac{A_{44}}{A} + \frac{C_{44}}{C} + \frac{A_4C_4}{AC} - \frac{q^2}{A^2} = -\Lambda - \frac{\omega}{2}\left(\frac{\phi_4}{\phi}\right) + \frac{B_4}{B}\frac{\phi_4}{\phi} + \frac{\Box\phi}{\phi}, \tag{12}$$

 0022-2488/85/112915-03$02.50

$$G_3^3=\frac{A_{44}}{A}+\frac{B_{44}}{B}+\frac{A_4B_4}{AB}-\frac{q^2}{A^2}$$
$$=-\Lambda-\frac{\omega}{2}\Big(\frac{\phi_4}{\phi}\Big)+\frac{C_4}{C}\frac{\phi_4}{\phi}+\frac{\Box\phi}{\phi}, \tag{13}$$
$$G_4^4=\frac{A_4B_4}{AB}+\frac{B_4C_4}{BC}+\frac{A_4C_4}{AC}-\frac{q^2}{A^2}$$
$$=-\Lambda+\frac{k}{\phi}\rho+\frac{\omega}{2}\Big(\frac{\phi_4}{\phi}\Big)^2+\frac{\phi_{44}}{\phi}+\frac{\Box\phi}{\phi}, \tag{14}$$
$$G_1^4=-q\Big(\frac{B_4}{B}-\frac{C_4}{C}\Big), \tag{15}$$
where the subscript 4 denotes ordinary differentiation with respect to t.

Equation (15) readily gives
$$B=\lambda C, \tag{16}$$
λ being an integration constant. Without loss of any generality we take $\lambda=1$. From the conservation equation $T^j_{i;j}=0$, we obtain
$$\rho_4=-\rho(A_4/A+2B_4/B), \tag{17}$$
which leads, after integration, to
$$\rho=c/AB^2, \tag{18}$$
where c is an integration constant. From (5), (6), (8), and (18) we obtain
$$1/AB^2=-(m/d)\phi^n, \tag{19}$$
where
$$d=k\mu c/(2\omega+3).$$
From (18) and (19) the density can be written in terms of the scalar field
$$\rho=-(mc/d)\phi^n. \tag{20}$$
In order to treat Eqs. (11)–(14) we introduce new variables α, β, and T by
$$A=e^\alpha,\quad B=e^\beta,\quad dt=AB^2\,d\tau, \tag{21}$$
and differentiation with respect to τ is denoted by a prime. Then Eq. (19) gives
$$e^{\alpha+2\beta}=-(d/m)\phi^{-n}. \tag{22}$$
From (22) it follows that
$$e^{-\alpha}=-(m/d)\phi^ne^{2\beta}. \tag{23}$$
We can express $\Box\phi=m\phi^n$ by
$$\phi''=-(d^2/m)\phi^{-n}. \tag{24}$$
Substituting (21)–(24) into Eqs. (11)–(14) we obtain
$$-\beta''+3\beta'^2+(2n-1)\beta'(\phi'/\phi)-q^2e^{4\beta}$$
$$=\Big(-\frac{ad^2}{m(2-n)}+\frac{d^2}{m}-\frac{nd^2}{m}\Big)\phi^{-(n+1)}$$
$$-\frac{D_1d^2}{m^2}\phi^{-(2n-1)}-\Big(\frac{\omega}{2}+n\Big)\Big(\frac{\phi'}{\phi}\Big)^2, \tag{25}$$
$$2\beta''+3\beta'^2+(2n+2)\beta'(\phi'/\phi)+q^2e^{4\beta}$$
$$=\Big(-\frac{ad^2}{m(2-n)}+\frac{d^2}{m}\Big)\phi^{-(n+1)}-\frac{D_1d^2}{m^2}\phi^{-(2n-1)}$$
$$-\Big(\frac{\omega}{2}+n\Big)\Big(\frac{\phi'}{\phi}\Big)^2, \tag{26}$$
$$3\beta'^2+2n\beta'(\phi'/\phi)+q^2e^{4\beta}$$
$$=\Big(\frac{ad^2}{m(2-n)}+\frac{kcd}{m}\Big)\phi^{-(n+1)}+\frac{D_1d^2}{m^2}\phi^{-(2n-1)}$$
$$-\Big(\frac{\omega}{2}+n\Big)\Big(\frac{\phi'}{\phi}\Big)^2. \tag{27}$$
Subtraction of (25) from (26) gives
$$3\beta''+3\beta'(\phi'/\phi)+2q^2e^{4\beta}=(nd^2/m)\phi^{-(n+1)}. \tag{28}$$
Eliminating β'' from (25) and (28) we obtain
$$3\beta'+2n\beta'(\phi'/\phi)-\tfrac{1}{3}q^2e^{4\beta}$$
$$=\Big(-\frac{ad^2}{m(2-n)}+\frac{d^2}{m}-\frac{2}{3}\frac{nd^2}{m}\Big)\phi^{-(n+1)}$$
$$-\frac{D_1d^2}{m^2}\phi^{-(2n-1)}\Big(\frac{\omega}{2}+n\Big)\Big(\frac{\phi'}{\phi}\Big)^2. \tag{29}$$
Subtracting (29) from (27) we get
$$\frac{4}{3}q^2e^{4\beta}=\Big(\frac{2ad^2}{m(2-n)}+\frac{kcd}{m}+\frac{2}{3}\frac{nd^2}{m}-\frac{d^2}{m}\Big)\phi^{-(n+1)}$$
$$+\frac{2D_1d^2}{m^2}\phi^{-(2n-1)}. \tag{30}$$
As the general solution of the above equations is rather difficult we consider the case $D_1=0$. Then, (30) can be written in the form
$$q^2e^{4\beta}=P\phi^{-(n+1)}, \tag{31}$$
where
$$P=\frac{3}{4}\Big(\frac{2ad^2}{m(2-n)}+\frac{kcd}{m}+\frac{2}{3}\frac{nd^2}{m}-\frac{d^2}{m}\Big).$$
Differentiation of (31) gives
$$\beta'=-[(n+1)/4](\phi'/\phi) \tag{32}$$
and
$$\beta''=\frac{(n+1)d^2}{4m}\phi^{-(n+1)}+\frac{n+1}{4}\Big(\frac{\phi'}{\phi}\Big)^2. \tag{33}$$
Equation (28) is satisfied if
$$k=[d/2c(2-n)](n^2-3n+4a-2). \tag{34}$$
Substituting (32) and (33) into (29) we obtain
$$(\phi'/\phi)^2=(1/Q^2)\phi^{-(n+1)}, \tag{35}$$
where
$$Q^2=m(3+8\omega+14n-5n^2)/2(7-5n)d^2.$$
The general solution of (35) is
$$\phi=[(n+1)(\tau+E)/2Q]^{2/(n+1)}, \tag{36}$$
where E is an integration constant and may be chosen such that at $\tau=0$ one has $\phi=0$. From (23), (31), and (36) we finally have the solution for α and β as follows:
$$e^{2\alpha}=\frac{1}{P}\Big(\frac{qd}{m}\Big)^2\Big[\frac{(n+1)(\tau+E)}{2Q}\Big]^{-2(n-1)/(n+1)}, \tag{37}$$
$$e^{4\beta}=\frac{P}{q^2}\Big[\frac{(n+1)(\tau+E)}{2Q}\Big]^{-2}. \tag{38}$$
The cosmological term given (8) with (36) is
$$\Lambda=[am/(2-n)]\phi^{n-1}. \tag{39}$$

The density can be written as a function of the cosmological term

$$\rho = -(mc/d)([(2-n)/am]\Lambda)^{n/(n-1)}. \tag{40}$$

We can also express α and β in terms of ϕ as

$$e^{2\alpha} = \frac{1}{P}\left(\frac{qd}{m}\right)^2 \phi^{-(n-1)}, \quad e^{4\beta} = \frac{P}{q^2}\phi^{-(n+1)}. \tag{41}$$

IV. DISCUSSIONS

In the last section we have obtained a cosmological dust solution of modified Brans–Dicke field equations in a Bianchi type-VI_0 homogeneous space. From (36) it is easily seen that ϕ is an increasing function of τ if $n<0$ and a decreasing function if $n>0$. The later case is physically unrealistic. The spatial volume $V=AB^2$ is given by

$$V=(d/m)\phi^{-n}. \tag{42}$$

If we consider an expanding universe, V is increasing with time which in view of (36) and (42) fixes the value of n always less than zero. So in the case $n<0$ and the epoch $\phi \to 0$ we have $V \to 0$ and $\rho \to \infty$. In course of time the model expands and has infinite volume $V \to \infty$ and $\rho \to 0$ as $\phi \to \infty$.

[1]C. Brans and R. H. Dicke, Phys. Rev. **124**, 925 (1951).
[2]P. G. Bergmann, Int. J. Theor. Phys. **1**, 25 (1968).
[3]R. V. Wagoner, Phys. Rev. D **1**, 3209 (1970).
[4]M. Endo and T. Fukui, Gen. Relativ. Gravit. **8**, 833 (1977).
[5]A. Banerjee and N. O. Santos, Gen. Relativ. Gravit. **14**, 559 (1982).
[6]M. P. Ryan and L. C. Shepley, *Homogeneous Relativistic Cosmologies* (Princeton U. P., Princeton, NJ, 1975).

Global existence and asymptotic behavior for the discrete velocity models of the Boltzmann equation

G. Toscani
Dipartimento di Matematica, Università di Pavia, Pavia, Italy

(Received 17 November 1984; accepted for publication 12 June 1985)

By using fixed point techniques, some theorems which supply local and global existence, uniqueness, and asymptotic behavior of the solution of a general discrete velocity model of the Boltzmann equation in three dimensions are given.

I. INTRODUCTION

The solution of the Cauchy problem for the discrete velocity models of the Boltzmann equation exists, globally in time, when the initial values satisfy a suitable "smallness" condition. This result, obtained by Illner,[1] holds for initial densities which depend on one or more spatial variables, and for whatever discrete velocity model.

Recently the author, with the aid of a fixed point property of the collisional integral, obtained at first for the $2r$-velocities plane regular model,[2] then for the full Broadwell model,[3] sufficient conditions both on the global existence and on the asymptotic stability of the solution.

In this paper, with a proper use of the methods of Ref. 3, we give, for a general discrete velocity model, results both on the local and global existence, and on the asymptotic behavior of the solution.

More in detail, in Sec. III we will prove a local existence theorem, which points out the existence of an upper time-dependent bound in the local evolution of the gas densities. In fact, we find that for bounded initial values, there is at least an interval of time in which the dominant process in the gas is the free flow.

In Sec. IV, we investigate the possibility to extend the result of Sec. III, in order to obtain the existence, globally in time, of the solution. If the initial data are "small" in a suitable L_1 norm, we deduce that the free flow dominates the evolution of the densities for all times. With our analysis, the asymptotic behavior of the solution is found, together with the global existence.

We think that our method of proof clarifies the rule of the collisional integral in the evolution of the densities; in fact, the main Lemma points out that the gain part of this integral, thanks to the geometry of the collisions, remains opportunely bounded if so are the initial data.

It has to be noted that our results are in accord with those of Hamdache,[4] who, independently of the author and generalizing a result of Tartar[5] to the three-dimensional case, obtained results on the asymptotic behavior of the global solution.

Finally, it is worth mentioning that the analysis of the existence and uniqueness of the solution, globally in time, for the initial value problem, and of its asymptotic behavior, is relevant in order to supply a physicomathematical validation of the discrete velocity Boltzmann equation as a mathematical model in molecular gas dynamics.

II. THE DISCRETE VELOCITY MODEL

The formulation of the initial value problem for a discrete velocity model of the Boltzmann equation is the following:

$$\frac{\partial f_i}{\partial t} + \mathbf{v}_i \cdot \nabla_x f_i = G_i(\underset{\sim}{f},\underset{\sim}{f}) - f_i L_i(\underset{\sim}{f}),$$
$$f_i(\mathbf{x}, 0) = \varphi_i(\mathbf{x}), \quad i = 1, 2, \ldots, p, \tag{2.1}$$

where $\{\mathbf{v}_i\}$, $i = 1, 2, \ldots, p$ is the set of the admissible velocities, and $\underset{\sim}{f}(\mathbf{x}, t) = \{f_1(\mathbf{x}, t), f_2(\mathbf{x}, t), \ldots, f_p(\mathbf{x}, t)\}$ is the p vector whose ith component $f_i(\mathbf{x}, t)$ represents the density of the gas with velocity $\mathbf{v}_i$ in the position $\mathbf{x}$ and at time t. Both vectors $\mathbf{x}$ and $\mathbf{v}_i$ are referred to an inertial frame of reference S.

In system (2.1) G_i and L_i are defined in the following fashion:

$$G_i(\underset{\sim}{f},\underset{\sim}{f})(\mathbf{x}, t) = \frac{1}{2} \sum_{j,k,m} A^{ij}_{km} f_k(\mathbf{x}, t) f_m(\mathbf{x}, t),$$
$$L_i(\underset{\sim}{f})(\mathbf{x}, t) = \frac{1}{2} \sum_{j,k,m} A^{km}_{ij} f_j(\mathbf{x}, t), \tag{2.2}$$
$$i, j, k, m = 1, 2, \ldots, p.$$

The quantities A^{km}_{ij} are non-negative constants, connected with the probability that two particles with velocities $\mathbf{v}_i$ and $\mathbf{v}_j$ collide and are scattered after the collision with velocities $\mathbf{v}_k$ and $\mathbf{v}_m$. As usual, the A^{km}_{ij} obey the reversibility hypothesis $A^{km}_{ij} = A^{ij}_{km}$, and the physical consistency $A^{ii}_{km} = A^{ij}_{kk} = A^{ij}_{mm} = A^{jj}_{km} = 0$.

Thanks to classical results, if the initial values $\varphi_i(\mathbf{x})$ are non-negative, $i = 1, 2, \ldots, p$, the solution $\underset{\sim}{f}^*(\mathbf{x}, t)$ of the Cauchy problem (2.1) has non-negative components in its interval of existence.

Keeping this property in mind, in what follows we will look for a solution of the modified problem

$$\frac{\partial f_i}{\partial t} + \mathbf{v}_i \cdot \nabla_x f_i = G_i(\underset{\sim}{f},\underset{\sim}{f}),$$
$$f_i(\mathbf{x}, 0) = \varphi_i(\mathbf{x}), \quad i = 1, 2, \ldots, p, \tag{2.3}$$

whose solution $\underset{\sim}{f}(\mathbf{x}, t)$, when it exists, trivially satisfies the conditions $f_i(\mathbf{x}, t) \geq f^*_i(\mathbf{x}, t)$, $i = 1, 2, \ldots, p$. The analysis of system (2.3) will be based on the concept of mild solution.

For every $T > 0$, let $C_b(\mathbf{R}^3)$ be the space of all bounded continuous functions on $\mathbf{R}^3$, and let $\widetilde{C}_b(\mathbf{R}^3)$ be the space of all

 0022-2488/85/112918-04$02.50

functions $\varphi \in C_b(\mathbf{R}^3)$ that go to zero as $\mathbf{x}$ goes to infinity.

With obvious notations, let us introduce, in analogy with Ref. 3, the Banach space

$$\mathbf{B}_T = \{C_b([0, T] \oplus \mathbf{R}^3)\}^p, \tag{2.4}$$

equipped with the supremum norm

$$\|\underset{\sim}{f}\| = \max_{i \leqslant p} \sup_{(\mathbf{x}, t) \in \mathbf{R}^3 \oplus [0, T]} |f_i(\mathbf{x}, t)|. \tag{2.5}$$

At this point the definition of mild solution can be proposed.

Definition: Let $\underset{\sim}{f} \in \mathbf{B}_T$. Then f is said to be a mild solution of the Cauchy problem (2.3) if for all $\mathbf{x} \in \mathbf{R}^3$ and $t \in [0, T]$ the equations

$$f_i(\mathbf{x} + \mathbf{v}_i t, t) = \varphi_i(\mathbf{x}) + \int_0^t G_i(\underset{\sim}{f}, \underset{\sim}{f})(\mathbf{x} + \mathbf{v}_i s, s) ds,$$

$$i = 1, 2, \ldots, p$$

are satisfied.

III. LOCAL EXISTENCE

We will prove in this section some results on the local existence of the solutions for the Cauchy problem (2.3). These results are quite different from the previous one. In fact, we obtain some particular bounds for the local solution, depending on the point $(\mathbf{x}, t)$, derived *a priori* from the given initial values, in such a way as the local existence can imply the global existence.

In the following, as specified in the Introduction, we will prove that for bounded initial data there is a time in which the dominant process in the evolution of the densities should be the free flow. To this purpose, let us define on $\mathbf{B}_T$ the operator A by components, as

$$(A\underset{\sim}{f})_i(\mathbf{x} + \mathbf{v}_i t, t) = \int_0^t G_i(\underset{\sim}{f}, \underset{\sim}{f})(\mathbf{x} + \mathbf{v}_i s, s) ds, \qquad i = 1, 2, \ldots, p. \tag{3.1}$$

By virtue of Definition (2.2) it is routine to verify that A maps $\mathbf{B}_T$ in $\mathbf{B}_T$.

Let $\alpha(s)$: $\mathbf{R}^+ \to \mathbf{R}^+$ be a nonincreasing bounded function for which

$$\alpha^* = \sup_{s > 0} \frac{\alpha((\sqrt{2}/2)s)}{\alpha(s)} < \infty,$$

and let us define

$$\lambda(t, \alpha) = \alpha^* \max_{i \leqslant p} \sum_{j,k,m} A_{km}^{ij} \times \int_0^t \alpha(\inf\{|\mathbf{v}_i - \mathbf{v}_k|, |\mathbf{v}_i - \mathbf{v}_m|\} \cdot s) ds. \tag{3.2}$$

Finally, for given constant $K > 0$, let us consider the following closed convex subsets of $\mathbf{B}_T$:

$$\mathbf{N}_T(K\alpha) = \{\underset{\sim}{f} \in \mathbf{B}_T;\ 0 \leqslant f_i(\mathbf{x}, + \mathbf{v}_i t, t) \leqslant K\alpha(|\mathbf{x}|);\ i = 1, 2, \ldots, p\}. \tag{3.3}$$

At this point, the following result can be proposed.

Lemma: Let $\underset{\sim}{f} \in \mathbf{N}_T(K\alpha)$. Then $A\underset{\sim}{f} \in \mathbf{N}_T\ (\lambda(T/2, \alpha) K^2 \alpha)$.

Proof: Let $\underset{\sim}{f} \in \mathbf{N}_T(K\alpha)$. Then

$$(A\underset{\sim}{f})_i(\mathbf{x} + \mathbf{v}_i t, t) = \frac{1}{2} \sum_{j,k,m} A_{km}^{ij} \int_0^t (f_k f_m)(\mathbf{x} + \mathbf{v}_i s, s) ds$$
$$\leqslant \frac{1}{2} K^2 \sum_{j,k,m} A_{km}^{ij} \int_0^t \alpha(|\mathbf{x} + (\mathbf{v}_i - \mathbf{v}_k)s|) \times \alpha(|\mathbf{x} + (\mathbf{v}_i - \mathbf{v}_m)s|) ds.$$

Thanks to the property of physical consistency, we have contribution to the sum only when $i \neq j \neq k \neq m$, that is, when $\mathbf{v}_i \neq \mathbf{v}_k$ and at the same time $\mathbf{v}_i \neq \mathbf{v}_m$.

Moreover, by virtue of the geometry of the collision, two particles with velocities $\mathbf{v}_i$ and $\mathbf{v}_j$ are scattered into $\mathbf{v}_k$ and $\mathbf{v}_m$ in such a way that the two vectors $\mathbf{v}_i - \mathbf{v}_k$ and $\mathbf{v}_i - \mathbf{v}_m$ are orthogonal. This simple property plays a fundamental role, since the problem is at this point reduced to majorized quantities as

$$\int_0^t \alpha(|\mathbf{x} + \mathbf{u}s|) \alpha(|\mathbf{x} + \mathbf{v}s|) ds,$$

where $|\mathbf{u}|, |\mathbf{v}| > 0$ and $\mathbf{u} \cdot \mathbf{v} = 0$.

Since the above integral is invariant with respect to a change of the reference frame S, we can evaluate it with respect to the frame S', defined by the orthogonal triad $\mathbf{i}, \mathbf{j}, \mathbf{k}$ such that

$$\mathbf{u} = \frac{|\mathbf{u}|}{\sqrt{2}}(\mathbf{i} + \mathbf{j}), \quad \mathbf{v} = \frac{|\mathbf{v}|}{\sqrt{2}}(\mathbf{i} - \mathbf{j}). \tag{3.4}$$

With respect to S', let us consider the vector $\mathbf{x} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$. At first, let us suppose that $x > 0$ and $y > 0$. Then

$$|\mathbf{x} + \mathbf{u}s|^2 = (x + (|\mathbf{u}|/\sqrt{2})s)^2 + (y + (|\mathbf{u}|/\sqrt{2})s)^2 + z^2 \geqslant \tfrac{1}{2}(x + y + \sqrt{2}|\mathbf{u}|s)^2$$

and

$$|\mathbf{x} + \mathbf{v}s|^2 = (x + (|\mathbf{v}|/\sqrt{2})s)^2 + (y - (|\mathbf{v}|/\sqrt{2})s)^2 + z^2 \geqslant \tfrac{1}{2}|\mathbf{x}|^2.$$

Since $\alpha(s)$ is not increasing, it follows that

$$\alpha(|\mathbf{x} + \mathbf{u}s|)\alpha(|\mathbf{x} + \mathbf{v}s|) \leqslant \alpha^* \cdot \alpha(|\mathbf{x}|)\alpha(|(x + y)/\sqrt{2} + |\mathbf{u}|s|).$$

If now $x < 0$ and $y < 0$, we obtain

$$\alpha(|\mathbf{x} + \mathbf{u}s|) \leqslant \alpha(|(|x| + |y|)/\sqrt{2} - |\mathbf{u}|s|),$$
$$\alpha(|\mathbf{x} + \mathbf{v}s|) \leqslant \alpha^* \alpha(|\mathbf{x}|),$$

and, if $x > 0$ and $y < 0$

$$\alpha(|\mathbf{x} + \mathbf{u}s|) \leqslant \alpha^* \alpha(|\mathbf{x}|),$$
$$\alpha(|\mathbf{x} + \mathbf{v}s|) \leqslant \alpha(|(x + |y|)/\sqrt{2}) + |\mathbf{v}|s|).$$

In the last case, that is, for $x < 0$ and $y > 0$,

$$\alpha(|\mathbf{x} + \mathbf{u}s|) \leqslant \alpha^* \alpha(|\mathbf{x}|),$$
$$\alpha(|\mathbf{x} + \mathbf{v}s|) \leqslant \alpha(|(|x| + y)/\sqrt{2} - |\mathbf{v}|s|).$$

Now, since $\alpha(s)$ is not increasing, the following inequalities are easily derived:

$$\int_0^t \alpha(|\,|\mathbf{x}| + |\mathbf{u}|s|)ds \leqslant 2\int_0^{t/2} \alpha(|\mathbf{u}|s)ds,$$
$$\int_0^t \alpha(|\,|\mathbf{x} - \mathbf{u}|s|)ds \leqslant 2\int_0^{t/2} \alpha(|\mathbf{u}|s)ds. \tag{3.5}$$

Thanks to (3.5), if $t \leqslant T$,

$$\int_0^t \alpha(|\mathbf{x}| + |\mathbf{u}s|)\alpha(|\mathbf{x} + \mathbf{v}s|)ds$$
$$\leqslant \alpha^*\alpha(|\mathbf{x}|)\sup_{|\mathbf{u}|,|\mathbf{v}|} \times \left\{\int_0^t \alpha\left(\left|\frac{|x| + |y|}{\sqrt{2}} \pm |\mathbf{u}|s\right|\right)ds; \int_0^t \alpha\left(\left|\frac{|x| + |y|}{\sqrt{2}} \pm |\mathbf{v}|s\right|\right)ds\right\}$$
$$\leqslant \alpha^*\alpha(|\mathbf{x}|)2 \sup_{|\mathbf{u}|,|\mathbf{v}|}\left\{\int_0^{T/2} \alpha(|\mathbf{u}|s)ds; \int_0^{T/2} \alpha(|\mathbf{v}|s)ds\right\}$$
$$\leqslant \alpha^*\alpha(|\mathbf{x}|)\cdot 2 \cdot \int_0^{T/2} \alpha(\inf(|\mathbf{u}|, |\mathbf{v}|)\cdot s)ds.$$

Let us retake into account the components of $A\underset{\sim}{f}$:

$$(A\underset{\sim}{f})_i(\mathbf{x} + \mathbf{v}_i t, t)$$
$$\leqslant \frac{1}{2}K^2 \sum_{j,k,m} A_{km}^{ij} \times \int_0^t \alpha(|\mathbf{x} + (\mathbf{v}_i - \mathbf{v}_k)s|)\alpha(|\mathbf{x} + (\mathbf{v}_i - \mathbf{v}_m)s|)ds$$
$$\leqslant \alpha(|\mathbf{x}|)\cdot \alpha^* K^2 \sum_{j,k,m} A_{km}^{ij} \times \int_0^{T/2} \alpha(\inf\{|\mathbf{v}_i - \mathbf{v}_k|; |\mathbf{v}_i - \mathbf{v}_m|\}s)ds,$$

so that, in virtue of Definition (3.2), for $i \leqslant p$

$$(A\underset{\sim}{f})_i(\mathbf{x} + \mathbf{v}_i t, t) \leqslant \lambda(T/2, \alpha)\cdot K^2 \cdot \alpha(|\mathbf{x}|),$$

and, since trivially $(A\underset{\sim}{f})_i(\mathbf{x} + \mathbf{v}_i t, t) \geqslant 0$, $i = 1, 2, \ldots, p$, we conclude that

$$A\underset{\sim}{f} \in \mathbb{N}_T(\lambda(T/2, \alpha)K^2\alpha). \qquad \square$$

With the aid of the Lemma, a theorem on the existence and unicity of the local mild solution of the Cauchy problem (2.3) is immediately derived.

Let us define, given non-negative initial data $\varphi_i(\mathbf{x}) \in C_b(\mathbb{R})^3$, $i \leqslant p$, the following:

$$\phi(s) = \max_{i \leqslant p} \sup_{|\mathbf{x}| > s} \varphi_i(\mathbf{x}). \tag{3.6}$$

From Definition (3.6) it follows that $\phi(s)$ is a bounded nonincreasing function which maps $\mathbb{R}^+$ in $\mathbb{R}^+$.

Let us suppose that the following condition is verified:

$$1 < \phi^* = \sup_{s > 0} \frac{\phi((\sqrt{2}/2)s)}{\phi(s)} < \infty, \tag{3.7}$$

and let T_0 be the time for which

$$\lambda(T_0/2, \phi) = \tfrac{1}{4}. \tag{3.8}$$

Theorem 3.1: Let $0 \leqslant \varphi_i(\mathbf{x}) \in C_b(\mathbb{R}^3)$, $i \leqslant p$, and let $\phi^* < \infty$. Then the Cauchy problem (2.3) has a unique non-negative mild solution $\underset{\sim}{f}(\mathbf{x}, t)$ in the interval $[0, T_0]$, with T_0 defined by (3.8). Moreover in this interval, for every $\mathbf{x} \in \mathbb{R}^3$

$$0 \leqslant f_i(\mathbf{x}, t) \leqslant 2\phi(|\mathbf{x} - \mathbf{v}_i t|), \quad i = 1, 2, \ldots, p. \tag{3.9}$$

Proof: Let us introduce on $\mathbb{B}_{T_0}$ the operator V whose components are defined by

$$(V\underset{\sim}{f})_i(\mathbf{x}, t) = \varphi_i(\mathbf{x} - \mathbf{v}_i t) + (A\underset{\sim}{f})_i(\mathbf{x}, t), \quad i = 1, 2, \ldots, p.$$

Thanks to the hypotheses made on the initial values, if $\underset{\sim}{f} \in \mathbb{B}_{T_0}$, then $V\underset{\sim}{f} \in \mathbb{B}_{T_0}$.

Now let $\underset{\sim}{f} \in \mathbb{N}_{T_0}(2\phi)$. Recalling the result of the Lemma, when $K = 2$, we obtain that $A\underset{\sim}{f} \in \mathbb{N}_{T_0}(\lambda(T_0/2, \phi)4\phi)$, and, since T_0 is defined through (3.8), $A\underset{\sim}{f} \in \mathbb{N}_{T_0}(\phi)$. Therefore, since the vector with components $\varphi_1(\mathbf{x} - \mathbf{v}_1 t), \varphi_2(\mathbf{x} - \mathbf{v}_2 t), \ldots, \varphi_p(\mathbf{x} - \mathbf{v}_p t)$ is a vector of $\mathbb{N}_{T_0}(\phi)$, $V\underset{\sim}{f} \in \mathbb{N}_{T_0}(2\phi)$. Finally, let $\underset{\sim}{f}, \underset{\sim}{g} \in \mathbb{N}_{T_0}(2\phi)$. Then

$$|(V\underset{\sim}{f})_i(\mathbf{x} + \mathbf{v}_i t, t) - (V\underset{\sim}{g})_i(\mathbf{x} + \mathbf{v}_i t, t)|$$
$$\leqslant |(A\underset{\sim}{f})_i(\mathbf{x} + \mathbf{v}_i t, t) - (A\underset{\sim}{g})_i(\mathbf{x} + \mathbf{v}_i t, t)|$$
$$\leqslant \frac{1}{2}\sum_{j,k,m} A_{km}^{ij}\int_0^t |f_k f_m - g_k g_m|(\mathbf{x} + \mathbf{v}_i s, s)ds$$
$$\leqslant \frac{1}{2}\sum_{j,k,m} A_{km}^{ij}\int_0^t \|\underset{\sim}{f} - \underset{\sim}{g}\|(f_m + g_k)(\mathbf{x} + \mathbf{v}_i s, s)ds$$
$$\leqslant \|\underset{\sim}{f} - \underset{\sim}{g}\| \cdot \frac{1}{2}\sum_{j,k,m} A_{km}^{ij}\int_0^t 2\{\phi(|\mathbf{x} + (\mathbf{v}_i - \mathbf{v}_m)s|) + \phi(|\mathbf{x} + (\mathbf{v}_i - \mathbf{v}_k)s|)\}ds$$
$$\leqslant \|\underset{\sim}{f} - \underset{\sim}{g}\| \cdot \sum_{j,k,m} A_{km}^{ij}\int_0^t \{\phi(|\,|\mathbf{x}| - |\mathbf{v}_i - \mathbf{v}_m|s|) + \phi(|\,|\mathbf{x}| - |\mathbf{v}_i - \mathbf{v}_k|s|)\}ds.$$

An application of inequalities (3.5) gives

$$\int_0^t \phi(|\,|\mathbf{x}| - |\mathbf{v}_i - \mathbf{v}_k|s|)ds + \int_0^t \phi(|\,|\mathbf{x}| - |\mathbf{v}_i - \mathbf{v}_k|s|)ds$$
$$\leqslant 4\int_0^{t/2} \phi(\inf\{|\mathbf{v}_i - \mathbf{v}_k|; |\mathbf{v}_i - \mathbf{v}_m|\}s)ds,$$

so that

$$\sum_{i,k,m} A_{km}^{ij}\int_0^t \{\phi(|\,|\mathbf{x}| - |\mathbf{v}_i - \mathbf{v}_k|s|) + \phi(|\,|\mathbf{x}| - |\mathbf{v}_i - \mathbf{v}_m|s|)\}ds$$
$$\leqslant 4\max_{i \leqslant p}\sum_{j,k,m} A_{km}^{ij}\int_0^{T_0/2} \phi(\inf\{|\mathbf{v}_i - \mathbf{v}_k|; |\mathbf{v}_i - \mathbf{v}_m|\}s)ds$$
$$= 4\cdot\lambda(T_0/2), \phi)\cdot\phi^{*-1} < 1.$$

Taking the supremum over $\mathbf{x} \in \mathbb{R}^3$ and $t \in [0, T_0]$ we obtain finally

$$\|V\underset{\sim}{f} - V\underset{\sim}{g}\| < \|\underset{\sim}{f} - \underset{\sim}{g}\|.$$

Thanks to the contraction mapping principle, the result follows. Since the solution lies in $\mathbb{N}_{T_0}(2\phi)$, we have also proved inequality (3.9). $\square$

Remark: Let us briefly comment on condition (3.7). It seems that Theorem 3.1 does not give answers if $\phi^* = 1$ or $\phi^* = \infty$.

In the first case, we have only to restrict the interval of existence, for example, by choosing T_1 so that $\lambda(T_1/2, \phi) = \frac{1}{3}$. In this way, the operator V is a contraction in $\mathbf{N}_{T_1}(2\phi)$. In the second case, typical of initial data that vanish outside of a finite volume, we have only to consider a bounded nonincreasing function $\phi_1(s) \geqslant \phi(s)$ for which $\phi_1^* < \infty$, and then apply the contraction principle on $\mathbf{N}_{T_0}(2\phi_1)$.

IV. GLOBAL EXISTENCE AND ASYMPTOTIC BEHAVIOR

From Theorem 3.1 follows a global existence theorem, if the initial values are "small" in a suitable sense. We have in fact the following.

Theorem 4.1: Let $0 \leqslant \varphi_i(\mathbf{x}) \in \widetilde{C}_b(\mathbf{R}^3)$, $i \leqslant p$, and let $\phi^* < \infty$. If in addition $\phi(s) \in L_1(\mathbf{R}^+)$, and $\lambda(\infty, \phi) \leqslant \frac{1}{4}$, the Cauchy problem (2.3) has a unique non-negative global mild solution $\underset{\sim}{f}(\mathbf{x}, t)$. Moreover, for every $\mathbf{x} \in \mathbf{R}^3$ and $t > 0$

$$0 \leqslant f_i(\mathbf{x}, t) \leqslant 2\phi(|\mathbf{x} - \mathbf{v}_i t|), \quad i = 1, 2, \ldots, p. \tag{4.1}$$

Proof: The result is a simple consequence of the fact that, for $\phi \in L_1(\mathbf{R}^+)$ with the L_1 norm of ϕ opportunely small, $\lambda(t, \phi) \leqslant \frac{1}{4}$ for all $t > 0$. □

Let us examine now briefly the asymptotic behavior of the global solution of the Cauchy problem (2.1).

From inequality (4.1) follows the asymptotic decay to zero of the solution of the Cauchy problem (2.3), and therefore the asymptotic decay to zero of the solution of the Cauchy problem (2.1). In fact, since under the hypothesis of Theorem 4.1, $\phi(s) \in \widetilde{C}_b(\mathbf{R}^+)$, from (4.1) we deduce

$$\lim_{t \to \infty} f_i(\mathbf{x}, t) \leqslant \lim_{t \to \infty} 2\phi(|\mathbf{x} - \mathbf{v}_i t|) = 0.$$

Moreover, let $t_1 < t_2$. Then

$$|f_i(\mathbf{x} + \mathbf{v}_i t_2, t_2) - f_i(\mathbf{x} + \mathbf{v}_i t_1, t_1)| \leqslant \int_{t_1}^{t_2} |G_i(\underset{\sim}{f}, \underset{\sim}{f}) - f_i L_i(\underset{\sim}{f})|(\mathbf{x} + \mathbf{v}_i s, s) ds.$$

Now, proceeding as in the Lemma,

$$\int_{t_1}^{t_2} G_i(\underset{\sim}{f}, \underset{\sim}{f})(\mathbf{x} + \mathbf{v}_i s, s) ds \leqslant 4\phi(|\mathbf{x}|)\lambda((t_2 - t_1)/2, \phi)$$

and

$$\int_{t_1}^{t_2} f_i L(\underset{\sim}{f})(\mathbf{x} + \mathbf{v}_i s, s) \leqslant 2\phi(|\mathbf{x}|) \int_{t_1}^{t_2} \sum_{j,k,m} A_{ij}^{km} \phi(||\mathbf{x}| - |\mathbf{v}_i - \mathbf{v}_j|s|) ds \leqslant 4\phi(|\mathbf{x}|)\lambda_1((t_2 - t_1)/2, \phi),$$

where

$$\lambda_1(t, \phi) = \max_{i \leqslant p,\, i \neq j} \sum_{j,k,m} \int_0^t A_{ij}^{km} \phi(|\mathbf{v}_i - \mathbf{v}_j|s) ds.$$

Therefore

$$|f_i(\mathbf{x} + \mathbf{v}_i t_2, t_2) - f_i(\mathbf{x} + \mathbf{v}_i t_1, t_1)| \leqslant 4\phi(|\mathbf{x}|)\{\lambda((t_2 - t_1)/2, \phi) + \lambda_1(t_2 - t_1/2, \phi)\}.$$

Since the function $\phi(s)$ is bounded, and both $\lambda(t, \phi)$, $\lambda_1(t, \phi)$ are continuous with respect to t, the existence of $\lim_{t \to \infty} f_i(\mathbf{x} + \mathbf{v}_i t, t)$, $i = 1, 2, \ldots, p$ follows.

[1]R. Illner, "Global existence results for discrete velocity models of the Boltzmann equation in several dimensions," J. Méc. Theor. Appl. **1**, (1982).

[2]G. Toscani, "On the discrete velocity models of the Boltzmann equation in several dimensions," Ann. Mat. Pura Appl. (4) **138** (1984).

[3]G. Toscani, "On the asymptotic behavior and stability of the solution for the Broadwell model of the Boltzmann equation in three dimensions," Math. Methods Appl. Sci. (to appear).

[4]K. Hamdache, "Existence globale et comportement asymptotique pour l'équation de Boltzmann à répartitio discrète de vitesses," preprint.

[5]L. Tartar, "Some existence theorem for semi-linear hyperbolic system in one space variable," preprint, University of Wisconsin, Madison, 1980.

Skeleton inequalities and mean field properties for lattice spin systems

Takashi Hara
Institute of Physics, College of Arts and Sciences, University of Tokyo, Komaba, Meguro-ku, Tokyo 153, Japan
Tetsuya Hattori
Department of Physics, Gakushuin University, Mejiro, Toshima-ku, Tokyo 171, Japan
Hal Tasaki
Department of Physics, Faculty of Science, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan

(Received 12 December 1984; accepted for publication 10 May 1985)

We present a proof of skeleton inequalities for ferromagnetic lattice spin systems with potential $V(\varphi^2) = (a/2)\varphi^2 + \Sigma_{n=2}^{M}\{\lambda_{2n}/(2n)!\}\varphi^{2n}$ (a real, $\lambda_{2n} > 0$) generalizing the Brydges–Fröhlich–Sokal and Bovier–Felder methods. As an application of the inequalities, we prove that, for sufficiently soft systems in $d > 4$ dimensions, critical exponents γ, α, and Δ_4 take their mean-field values (i.e., $\gamma = 1$, $\alpha = 0$, and $\Delta_4 = \frac{3}{2}$).

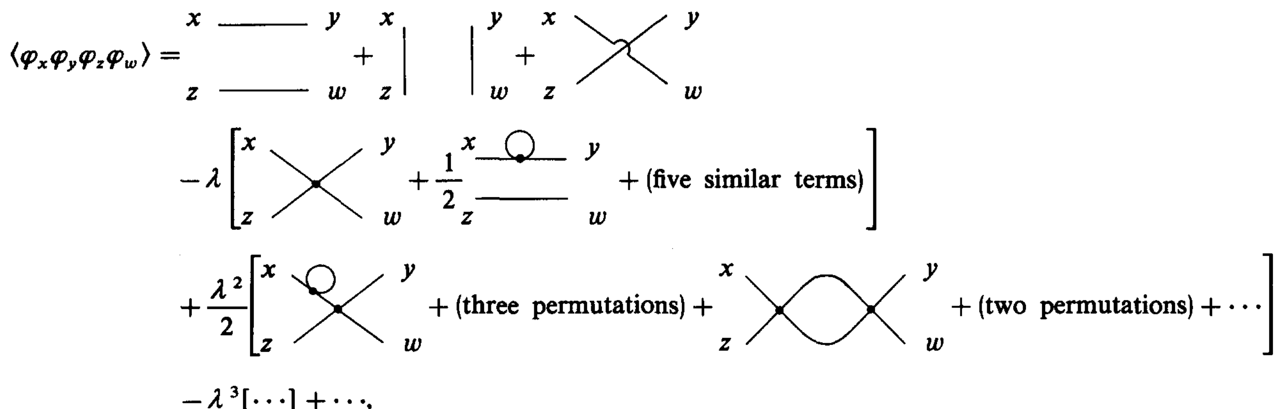
I. INTRODUCTION

Recently, Brydges, Fröhlich, and Sokal[1] introduced a new set of quite interesting inequalities called skeleton inequalities into lattice φ^4 systems. Making full use of these inequalities, they have succeeded in a simple and elegant construction of (weakly coupled) nontrivial $\lambda\varphi^4{}_3$ and $\lambda\varphi^4{}_2$ continuum field theories.[2,3] This, with the result of $\varphi^4{}_d$ triviality ($d > 4$) due to Aizenman and Fröhlich,[4–9] are the two outstanding achievements in the rigorous analysis of lattice field theories by the random-walk methods.

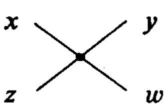
Let us briefly review Brydges–Fröhlich–Sokal skeleton inequalities. We first describe the notion of *skeleton expansion*. Consider the $2p$-point function of a lattice φ^4 system:

$$\langle \varphi_{x_1}\varphi_{x_2}\cdots\varphi_{x_{2p}}\rangle \equiv \frac{\int \Pi_{x\in\Lambda}\, d\varphi_x (\varphi_{x_1}\varphi_{x_2}\cdots\varphi_{x_{2p}})\exp\{\frac{1}{2}\Sigma_{x,y}J_{xy}\varphi_x\varphi_y - \Sigma_x((a/2)\varphi_x{}^2 + (\lambda/4!)\varphi_x{}^4)\}}{\int \Pi_{x\in\Lambda}\, d\varphi_x \exp\{\frac{1}{2}\Sigma_{x,y}J_{xy}\varphi_x\varphi_y - \Sigma_x((a/2)\varphi_x{}^2 + (\lambda/4!)\varphi_x{}^4)\}}.$$

If one expands the term $\exp\{-(\lambda/4!)\varphi_x{}^4\}$ in a formal Taylor series, the $2p$-point function can be formally expressed in terms of an infinite series of the Gaussian propagators $G_{xy}{}^{(0)} = \langle\varphi_x\varphi_y\rangle_{\lambda=0}$. This is nothing but the well-known formal perturbation expansion. For example, the four-point function is represented graphically as

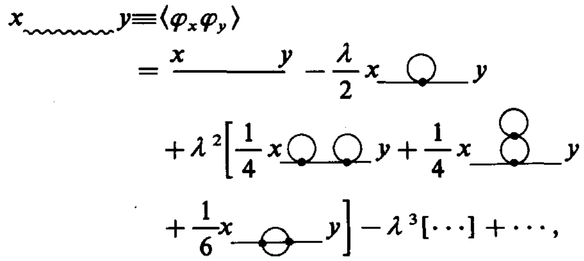


where a line stands for a Gaussian propagator $G_{xy}{}^{(0)}$, and a dot is a shorthand for a summation over the lattice sites, for example,



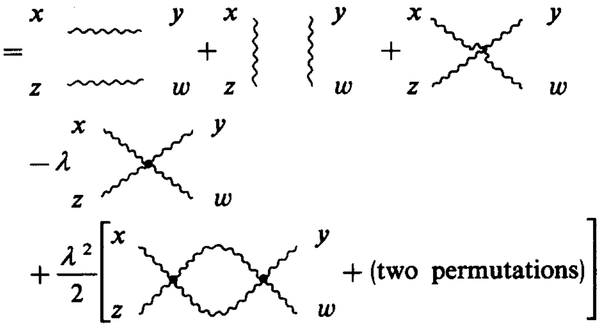
$$= \sum_u G_{xu}{}^{(0)}G_{yu}{}^{(0)}G_{zu}{}^{(0)}G_{wu}{}^{(0)}.$$

Using the similar formal series for the two-point function



one can rewrite the $2p$-point function in terms of the (non-Gaussian) two-point functions. If one takes into account the topologies and coefficients of the graphs carefully, one will observe that this expression for the $2p$-point function is represented as a series of *skeleton graphs*, i.e., graphs without self-energy part, with exactly the same combinatorial coefficients as those in the formal perturbation expansion:

$\langle \varphi_x \varphi_y \varphi_z \varphi_w \rangle$



$-\lambda^3[\cdots] + \cdots.$

We call such an expression of the $2p$-point function a skeleton expansion.

Skeleton inequalities provide us with a rigorous version of the perturbation theory in the following sense. Consider a partial summation, up to some order $(-\lambda)^n$, of the skeleton expansion for the $2p$-point function. Then this quantity gives a rigorous upper (resp. lower) bound for the $2p$-point function if the order n is an even (resp. odd) integer. See Eqs. (3.9)–(3.11) for the precise forms of the inequalities.

Brydges, Fröhlich, and Sokal proved the skeleton inequalities for the four-point function by construction up to second order.[1] A complete proof of the inequalities for the φ^4 system to all orders in λ was given by Bovier and Felder,[10] who devised an elegant argument relying on the asymptoticity of the formal perturbation expansion.

In the present paper, we will push their idea further and study general continuous spin systems with potential of the form

$$V_x(\varphi_x{}^2) = \frac{a_x}{2}\varphi_x{}^2 + \sum_{n=2}^{M_x} \frac{\lambda_{2n,x}}{(2n)!}\varphi_x{}^{2n}, \quad \lambda_{2m} \geqslant 0.$$

We can again define *skeleton expansion* in the same way as in the φ^4 system, and prove the skeleton inequalities for the $2p$-point functions.

In Sec. II, which is the main part of the present paper, we carry out to the full extent the ideas of Bovier and Felder, and prove these generalized skeleton inequalities. We first prove the existence of alternating upper and lower bounds expressed as summations over the graphs. Each term in the summations is a product of some coefficient and graph amplitude, which is written in terms of the two-point functions. The crucial point is that these coefficients are independent of the system parameters. Then, making full use of this independence, we determine the explicit forms of the bounds, which actually coincide with partial summations of the skeleton expansion.

The difference in our method as opposed to the original method of Bovier and Felder is in the first step of the proof. Bovier and Felder exhibited that the summations over graphs extend only to the skeleton graphs. We here omit this procedure, and leave everything to the second step, which gives us sufficient information.

In Sec. III, we discuss some applications of these inequalities to the statistical mechanical theory of spin systems.[11] We prove that in a "sufficiently soft" system ("soft" in the sense that the system is weakly coupled), the critical exponents γ, α, and Δ_4 are exactly identical with those values predicted by the mean-field theory, provided that the dimensionality of the lattice is greater than 4.

II. PROOF OF SKELETON INEQUALITIES

A. Preliminaries and main result

First we describe our model systems. Let Λ be an arbitrary finite lattice. To each site $x \in \Lambda$, we associate a spin variable $\varphi_x \in R$, with *a priori* measure

$$d\nu_x(\varphi_x) \equiv \exp(-V_x(\varphi_x{}^2))d\varphi_x. \tag{2.1}$$

We consider a potential $V_x(\varphi_x{}^2)$ of order $2M_x$ $(M_x \geqslant 2)$,

$$V_x(\varphi_x{}^2) \equiv \frac{a_x}{2}\varphi_x{}^2 + \lambda \sum_{n=2}^{M_x} \frac{g_{2n,x}}{(2n)!}\varphi_x{}^{2n}, \tag{2.2}$$

where a_x is real, $g_{2n,x} \geqslant 0$, and $g_{2M_x,x} > 0$. Constants M_x, a_x, and $g_{2n,x}$ can be site dependent, while the expansion parameter λ is a site-independent positive constant.

Thermal expectation of our system is defined as[12]

$$\langle \cdots \rangle \equiv Z^{-1} \int \prod_x d\nu_x(\varphi_x)(\cdots)e^{-\mathscr{H}}, \tag{2.3}$$

where Z is the normalization factor (partition function), and $\mathscr{H}$ is the Hamiltonian

$$\mathscr{H} \equiv -\frac{1}{2}\sum_{x,y} J_{xy}\varphi_x\varphi_y, \tag{2.4}$$

with

$$J_{xy} = J_{yx} \geqslant 0, \quad J_{xx} = 0.$$

Next we recall some useful formulas from the random-walk representation theory.[1,13,14] The following "integration by parts formula" plays a central role:

$$\langle \varphi_x F(\{\varphi\}) \rangle = \sum_{y\in\Lambda} \sum_{\omega: x\to y} J^\omega \int d\nu_\omega(t) \mathscr{Z}(t) \left\langle \frac{\partial F}{\partial \varphi_y} \right\rangle_t. \tag{2.5}$$

The second sum ranges over all random walks,

$$\omega \equiv (\omega(0), \omega(1), \ldots, \omega(n)), \quad \omega(i) \in \Lambda,$$

starting at $x = \omega(0)$ and ending at $y = \omega(n)$. Here J^ω stands for the product

$$J^\omega \equiv J_{\omega(0)\omega(1)} J_{\omega(1)\omega(2)} \cdots J_{\omega(n-1)\omega(n)}.$$

The integrations over "local times" $\{t_x\}_{x\in\Lambda}$ are performed with measures

$$d\nu_\omega(t) \equiv \prod_{x\in\Lambda} d\nu_{n(x,\omega)}(t_x), \tag{2.6}$$

where

$$d\nu_n(s) \equiv \begin{cases} \delta(s)ds & (n = 0), \\ \theta(s)\dfrac{s^{n-1}}{(n-1)!}ds & (n \geqslant 1), \end{cases} \tag{2.6'}$$

and $n(x,\omega)$ is the number of times that ω visits the site x. Z_t and $\langle \cdots \rangle_t$ are defined by replacing each $V_x(\varphi_x{}^2)$ in (2.3) by $V_x(\varphi_x{}^2 + 2t_x)$, i.e.,

$$\langle \cdots \rangle_t \equiv Z_t^{-1} \int \prod_x \{ d\varphi_x \, e^{-V_x(\varphi_x^2 + 2t_x)} \} (\cdots) e^{-\mathscr{H}}$$

and

$$\mathscr{Z}(t) \equiv Z_t / Z.$$

This representation provides us with convenient expressions for correlation functions of the system. If we define unsymmetrized $2p$-point functions F_{2p} by[10]

$$F_{2p}(x_1,y_1;x_2,y_2;\ldots;x_p,y_p) \equiv \sum_{\substack{\omega_i: x_i \to y_i \\ (i=1,2,\ldots,p)}} \left(\prod_{i=1}^{p} J^{\omega_i} \right) \int \prod_{i=1}^{p} d\nu_{\omega_i}(t^i) \mathscr{Z}\left(\sum_{i=1}^{p} t^i \right), \tag{2.7}$$

then $2p$-point functions can be written as

$$\langle \varphi_{x_1} \varphi_{x_2} \cdots \varphi_{x_{2p}} \rangle = \sum_{\pi} F(x_{\pi(1)}, x_{\pi(2)}; \cdots; x_{\pi(2p-1)}, x_{\pi(2p)}), \tag{2.8}$$

where the summation runs over all the pairings of the $2p$-points $x_1, x_2, \ldots, x_{2p}$. In particular, two-point functions have a simple expression:

$$\langle \varphi_x \varphi_y \rangle = \sum_{\omega: x \to y} J^{\omega} \int d\nu_{\omega}(t) \mathscr{Z}(t).$$

Finally, we describe some graphical concepts. A graph G consists of a set of external points G_e, sets of internal points of order $2n$ $G_{i,2n}$ ($n = 2,3,\ldots,M$), and a set of lines G_l. We suppose that the set G_e is ordered. Moreover, each line is connecting two of the (external or internal) points in such a way that (i) only one line attaches to each external point, and (ii) $2n$ different lines attach to each internal point of order $2n$. The graph G is nothing but the graph which appears in the formal perturbation expansion in the $\lambda \sum_{n=2}^{M} g_{2n} \varphi^{2n}$ theory. We denote by $\mathscr{G}_{2p}$ the set of all graphs with $2p$ external points.

A graph G is called a skeleton graph if no internal point of G is separated from the external points by cutting arbitrary one or two internal lines of G (in other words, if G contains no "self-energy parts").

With a graph $G \in \mathscr{G}_{2p}$, and $2p$ sites $x_1, x_2, \ldots, x_{2p}$ of the lattice, we associate its amplitude $\mathscr{A}^{(\lambda)}(G)_{(x_1,x_2,\ldots,x_{2p})}$ written in terms of the two-point functions of the system. Substitute site y_q of the lattice Λ for each point q in G_i or G_e. (For the external points $q \in G_e$, we fix $y_q = x_q$, where $q = 1,2,\ldots,2p$.) Then the amplitude is defined as

$$\mathscr{A}^{(\lambda)}(G)_{(x_1,x_2,\ldots,x_{2p})} \equiv \sum_{\substack{y_q \in \Lambda \\ (\text{all } q \in G_i)}} \left(\prod_{q \in G_i} g_{(\text{order of } q), y_q} \right) \times \prod_{l \in G_l} \langle \varphi_{y_{l_1}} \varphi_{y_{l_2}} \rangle, \tag{2.9}$$

where l_1 and l_2 denote endpoints of a line l.

Now we can state our main theorem using the following shorthand notation due to Bovier and Felder.[10]

Definition 2.1: Given $f(\lambda)$ and $\{ fk(\lambda) \}_{k=0,1,\ldots}$ as functions of λ, we write

$$f(\lambda) \lessgtr \sum^{N} (-\lambda)^k f_k(\lambda)$$

if

$$f(\lambda) \leqslant \sum_{k=0}^{2n} (-\lambda)^k f_k(\lambda)$$

and

$$f(\lambda) \geqslant \sum_{k=0}^{2m+1} (-\lambda)^k f_k(\lambda)$$

hold for all $\lambda > 0$ and all non-negative integers n and m satisfying $2n \leqslant N$ and $2m+1 \leqslant N$. We write

$$f(\lambda) \lessgtr \sum (-\lambda)^k f_k(\lambda)$$

if the above relations hold for all $N \geqslant 0$.

Theorem 2.2: For our models defined by Eqs. (2.1)–(2.4), we have the following alternating bounds:

$$\langle \varphi_{x_1} \varphi_{x_2} \cdots \varphi_{x_{2p}} \rangle \lessgtr \sum (-\lambda)^k \sum_{G \in \mathscr{S}_{2p}^{(k)}} C_{\text{pert}}(G) \mathscr{A}^{(\lambda)}(G)_{(x_1,\ldots,x_{2p})}, \tag{2.10}$$

where $\mathscr{S}_{2p}^{(k)}$ is the set of all skeleton graphs with k-internal points that arise in the formal perturbation expansion of the $2p$-point function, and $C_{\text{pert}}(G)$ is the corresponding non-negative combinatorial factor in the perturbation expansion.

The proof of this theorem is given in the following two subsections. In Sec. II B, we prove the existence of the alternating bounds of the form of (2.10), where, however, the coefficients $C_{\text{pert}}(G)$ are replaced by (unknown) universal constants $C^{(k)}(G)$ and the summations extend also to non-skeleton graphs $G \in \mathscr{G}_{2p}$. Then in Sec. II C, using the asymptoticity of the formal perturbation series (for a suitable choice of the parameters), we determine the unknown coefficients $C^{(k)}(G)$, and get the theorem.

Remarks: (1) As is noted in Refs. 1, 10, and 14, the skeleton inequalities are also valid for the two-component systems[15,16] described by the potential (2.2) with φ_x^2 replaced by $|\varphi_x|^2$.

(2) The skeleton inequalities are also valid in the infinite volume limit, as long as the summation on the right-hand side has a well-defined limit. This is often the case if the system is in the high-temperature region.

B. Existence of alternating bounds

Here we will prove the existence of alternating bounds for $2p$-point functions.

Proposition 2.3: For our models defined by Eqs. (2.1)–(2.4), the relation

$$\langle \varphi_{x_1} \varphi_{x_2} \cdots \varphi_{x_{2p}} \rangle \lessgtr \sum (-\lambda)^k \sum_{G \in \mathscr{G}_{2p}} C^{(k)}(G) \mathscr{A}^{(\lambda)}(G)_{(x_1,\ldots,x_{2p})} \tag{2.11}$$

holds. Here $C^{(k)}(G)$ are some non-negative constants depending on the graph G, and being independent of the parameters $x_1, x_2, \ldots, x_{2p}$, Λ, a_x, $g_{2n,x}$, λ, M_x, and J_{xy}. Moreover, $C^{(k)}(G)$ are nonvanishing only for a finite number of G's for fixed values of k.

Proof: We will prove the following three relations for all values of N. Then the second relation (2.13) with the formula (2.8) proves the proposition.

Relation 1:

$$\langle\varphi_x\varphi_y\rangle_t-\langle\varphi_x\varphi_y\rangle \leqslant\sum^{N}(-\lambda)^k\sum_{l\geqslant 1}\sum_{\substack{m_i,n_i\geqslant 1\\(i=1,\ldots,l)}}\sum_{G\in\mathscr{G}_{2(1+\Sigma n_i)}}C_1^{(k)}(l,\{m_i\},\{n_i\};G)\times\sum_{\substack{x_i\in\Lambda\\(i=1,\ldots,l)}}\prod_{i=1}^{l}g_{2m_i+2n_i,x_i}(t_{x_i})^{m_i}\mathscr{A}^{(\lambda)}(G)_{(x,y,\underbrace{x_1,\ldots,x_1}_{2n_1},\ldots,\underbrace{x_l,\ldots,x_l}_{2n_l})}. \tag{2.12}$$

On the right-hand side of Eq. (2.12), the $k=0$ term is absent.

Relation 2:

$$F_{2p}(x_1,y_1;\ldots;x_p,y_p)\leqslant\sum^{N}(-\lambda)^k\sum_{G\in\mathscr{G}_{2p}}C_2^{(k)}(G)\mathscr{A}^{(\lambda)}(G)_{(x_1,y_1,\ldots,x_p,y_p)}. \tag{2.13}$$

Relation 3:

$$\langle\varphi_x\varphi_y;\varphi_z^{\,2m}\rangle\equiv\langle\varphi_x\varphi_y\varphi_z^{\,2m}\rangle-\langle\varphi_x\varphi_y\rangle\langle\varphi_z^{\,2m}\rangle\leqslant\sum^{N}(-\lambda)^k\sum_{G\in\mathscr{G}_{2m+2}}C_3^{(k)}(G)\mathscr{A}^{(\lambda)}(G)_{(x,y,\underbrace{z,\ldots,z}_{2m})}. \tag{2.14}$$

Here $C_1^{(k)}$, $C_2^{(k)}$, and $C_3^{(k)}$ are non-negative constants depending only on G (and integers l,m_i,n_i). We give the proof by an induction in N, using three key identities (which are the simple consequences of the definitions and random-walk representation) and a simple lemma described below. We start with Relation 1 for $N=0$ and proceed to Relation 2 for $N=0$, Relation 3 for $N=0$, Relation 1 for $N=1$, and so on, as is indicated in Fig. 1.

Key 1:

$$\langle\varphi_x\varphi_y\rangle_t-\langle\varphi_x\varphi_y\rangle=-\lambda\sum_{k,l\geqslant 0}^{k+l<M}\frac{2^k k}{(2k+2l)!}\,{}_{l+k}C_k\sum_z(g_{2k+2l,z})(t_z)^k\times\int_0^1 d\alpha\,\alpha^{k-1}\langle\varphi_x\varphi_y;\varphi_z^{\,2l}\rangle_{\alpha t}. \tag{2.15}$$

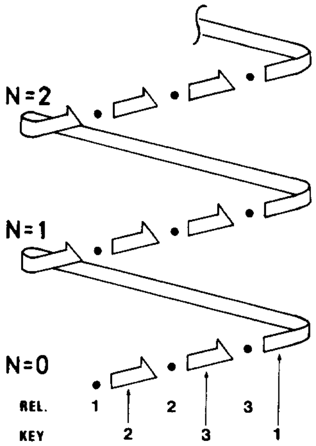


FIG. 1. How to carry out the inductive proof.

Key 2:

$$F_{2p}(x_1,y_1;\ldots;x_p,y_p)=\sum_{\substack{\omega_i:x_i\to y_i\\(i=1,\ldots,p-1)}}J^{\omega_1}\cdots J^{\omega_{p-1}}\int\prod_{i=1}^{p-1}d\nu_{\omega_i}(t^i)\times\mathscr{Z}\left(\sum_{i=1}^{p-1}t^i\right)\langle\varphi_{x_p}\varphi_{y_p}\rangle_{\Sigma_{i=1}^{p-1}t^i}. \tag{2.16}$$

Key 3:

$$\langle\varphi_x\varphi_y;\varphi_z^{\,2m}\rangle=2m(2m-1)!!F_{2m+2}(x,z;y,z;z,z;\ldots;z,z)+(2m-1)!!\sum_{\substack{\omega_i:z\to z\\(i=1,\ldots,m)}}J^{\omega_1}\cdots J^{\omega_m}\int\prod_{i=1}^{m}d\nu_{\omega_i}(t^i)\times\mathscr{Z}\left(\sum_{i=1}^{m}t^i\right)\{\langle\varphi_x\varphi_y\rangle_{\Sigma_{i=1}^{m}t^i}-\langle\varphi_x\varphi_y\rangle\}. \tag{2.17}$$

Lemma 2.4 (Bovier and Felder[10]): Let

$$f(\lambda)\leqslant\sum^{N}(-\lambda)^k f_k(\lambda),$$

$$f_k(\lambda)\leqslant\sum^{N}(-\lambda)^l f_{kl}(\lambda).$$

Then

$$f(\lambda)\leqslant\sum^{N}(-\lambda)^k h_k(\lambda),$$

with

$$h_k(\lambda)\equiv\sum_{l+m=k}f_{lm}(\lambda).$$

Since our inductive proof is a natural extension of those given by Brydges, Fröhlich, and Sokal[1] and by Bovier and Felder,[10] we only sketch the outlines of each step.

Proof of Relation 1 for $N=0$: This can be proved constructively by the monotonicity inequality

$$\langle\varphi_x\varphi_y\rangle_t\leqslant\langle\varphi_x\varphi_y\rangle,$$

which is a consequence of the first key identity (2.15) and the Griffiths II inequality.[16–18]

Proof of Relation 2 for N, given Relation 1 for N and all Relations for $N-1$, $N-2$,...: This is done by an induction in $2p$ (number of points). For $p=1$, the trivial bound

$$\langle\varphi_x\varphi_y\rangle\leqslant\sum^{N}(-\lambda)^k f_k(\lambda),$$

with $f_0=\langle\varphi_x\varphi_y\rangle$, $f_n=0$ $(n\geqslant 1)$ establishes the relation.

To prove the relation for $2p$ (given the relations for $2p$-2, $2p-4$,...), insert Relation 1 for N into the right-hand side of the second key identity (2.16). For the term originating from $\langle\varphi_x\varphi_y\rangle$, we can use the assumed Relation 2 about F_{2p-2} for N to get the desired alternating bound. As for the remaining terms, a tedious calculation using the "path splitting formula[1]" reduces the quantity to a combination of F_{2q}'s and two-point functions. Substituting the assumed alternating bounds (up to order $N-1$) for F_{2q}'s and rearranging the inequalities by Lemma 2.4, we obtain alternating bounds. We can then check that the resulting bounds are of the right form (e.g., the factor $g_{2n,x}$ appears corresponding to each internal point x of order $2n$), and consequently obtain the desired relation.[19]

Proof of Relation 3 for N, given Relations 1 and 2 for N, and all Relations for N−1, N−2,...: This can be done in exactly the same way as the previous step, once we take advantage of the third key identity (2.17).

Proof of Relation 1 for N+1, given all relations for N, N−1,...: Observe that on the right-hand side of the first key identity (2.15), the zeroth-order term in $(-\lambda)$ does not exist. This fact enables us to lift the relations of order N to those of order $N+1$.

Since the relations are independent of the specific values of the system parameters, Relation 3 for N is valid for the quantity

$$\langle \varphi_x \varphi_y ; \varphi_z{}^{2m} \rangle_{\alpha t},$$

with

$$\tilde{g}_{2n} \equiv \sum_{k=0}^{M} \frac{g_{2n+2k}(2t)^k{}_{n+k} C_k (2n)!}{(2n+2k)!}$$

playing the role of g_{2n}. Substituting this relation into the right-hand side of the first key identity (2.15), we obtain $(N+1)$-th-order bounds for $\langle \varphi_x \varphi_y \rangle_t - \langle \varphi_x \varphi_y \rangle$ in terms of the two-point functions $\langle \varphi_u \varphi_v \rangle_{\alpha t}$. But, if we apply Relation 1 for N, these unusual two-point functions can be bounded alternatingly by the usual two-point functions. We obtain the desired relation by Lemma 2.4. ■

C. Skeleton inequalities

Here we will refine the argument given in the previous subsection and show that all the coefficients in the alternating bounds (2.11) are exactly equal to those which appear in the skeleton expansion.

Proposition 2.5: Coefficients $C^{(k)}(G)$ in Eq. (2.11) satisfy

$$C^{(k)}(G) = \begin{cases} C_{\text{pert}}(G), & \text{if } G \in \mathscr{S}_{2p}{}^{(k)}, \\ 0, & \text{otherwise.} \end{cases} \tag{2.18}$$

Proof: Since the coefficients $C^{(k)}(G)$ are universal (i.e., independent of all the parameters, $\{J_{xy}\}$, Λ, $\{x_i\}$, $\{a_x\}$, and $\{g_{2n,x}\}$), we only have to determine $C^{(k)}(G)$ for some conveniently chosen lattice system. Here we let the parameters a_x be sufficiently large so that the system with $\lambda = 0$ (the Gaussian model) is well defined.

Existence of the Gaussian model, together with the Griffiths II inequality, implies the following bound for $2p$-point functions:

$$0 \leqslant \langle \varphi_{x_1} \varphi_{x_2} \cdots \varphi_{x_{2p}} \rangle \leqslant \text{const}_1 < \infty, \tag{2.19}$$

where the const_1 is independent of λ and $\{x_i\}$. This bound, applied to the $(-\lambda)^{N+1}$ term on the right-hand side of Eq. (2.11), immediately leads to the following estimate:

$$\Big| \langle \varphi_{x_1} \cdots \varphi_{x_{2p}} \rangle - \sum_{k=0}^{N} (-\lambda)^k \sum_{G \in \mathscr{G}_{2p}} C^{(k)}(G) \mathscr{A}^{(\lambda)}(G)_{(x_1,\ldots,x_{2p})} \Big| \leqslant \lambda^{N+1} \text{const}_2. \tag{2.20}$$

On the other hand, we can prove the asymptoticity of the formal perturbation series of $2p$-point functions, using the Schwinger–Dyson equation (as a generator of the series) and the bound (2.19). Partial resummations of the series yield the following lemma.

Lemma 2.6: In our model, for all $N > 0$,

$$\Big| \langle \varphi_{x_1} \cdots \varphi_{x_{2p}} \rangle - \sum_{k=0}^{N} (-\lambda)^k \sum_{G \in \mathscr{S}_{2p}{}^{(k)}} C_{\text{pert}}(G) \mathscr{A}^{(\lambda)}(G)_{(x_1,\ldots,x_{2p})} \Big| \leqslant \lambda^{N+1} \text{const}_3 \tag{2.21}$$

hold. Here $\mathscr{S}_{2p}{}^{(k)}$ is the set of all skeleton graphs with k vertices and $2p$-external points that appear in the perturbation expansion of the $2p$-point function. (For completeness, we will prove this lemma at the end of this section.)

Combining Eqs. (2.20) and (2.21), we obtain our basic estimate

$$\Big| \sum_{k=0}^{N} (-\lambda)^k \sum_{G \in \mathscr{G}_{2p}} C^{(k)}(G) \mathscr{A}^{(\lambda)}(G)_{(x_1,\ldots,x_{2p})} - \sum_{k=0}^{N} (-\lambda)^k \sum_{G \in \mathscr{S}_{2p}{}^{(k)}} C_{\text{pert}}(G) \mathscr{A}(G)_{(x_1,\ldots,x_{2p})} \Big| \leqslant \lambda^{N+1} \text{const}_4. \tag{2.22}$$

Now we prove the proposition $C^{(k)}(G) = C_{\text{pert}}(G)$ (for $G \in \mathscr{S}^{(k)}$) by an induction in N. Assume the proposition for all $k \leqslant N-1$. Then, on the right-hand side of (2.22), all the terms with $k \leqslant N-1$ exactly cancel. Dividing the consequent inequality by λ^N and letting $\lambda \to 0$, we are led to the *equality*

$$\sum_{G \in \mathscr{G}_{2p}} C^{(k)}(G) \mathscr{A}^{(0)}(G)_{(x_1,\ldots,x_{2p})} = \sum_{G \in \mathscr{S}_{2p}{}^{(k)}} C_{\text{pert}}(G) \mathscr{A}^{(0)}(G)_{(x_1,\ldots,x_{2p})}, \tag{2.23}$$

where $\mathscr{A}^{(0)}$ stands for amplitudes obtained from the Gaussian propagators $\langle \varphi_x \varphi_y \rangle_{\lambda=0}$. Note that the Gaussian propagator has the explicit expression

$$C_{xy} \equiv \langle \varphi_x \varphi_y \rangle_{\lambda=0} = (A^{-1})_{xy},$$

where a $|\Lambda| \times |\Lambda|$ matrix A is defined as $A_{xy} = -J_{xy}$ for $x \neq y$ and $A_{xx} = a_x$. This implies that we can choose every C_{xy} at will[20] independently. If the (finite) lattice Λ is sufficiently large (compared with N), successive differentiations of Eq. (2.23) by various C_{xy} yield the desired equality.

To start the inductive proof, we only have to note that no terms with inverse powers of $(-\lambda)$ exist in Eq. (2.22). This (formally) corresponds to the desired statement for $N = -1$. ■

Proof of Lemma 2.6: We first define

$$G_{xy}{}^{(\lambda,N)} \equiv \sum_{k=0}^{N} (-\lambda)^k \sum_{G \in \mathscr{G}_2{}^{(k)}} C_{\text{pert}}(G) \mathscr{A}^{(0)}(G)_{(x,y)}, \tag{2.24}$$

where the second summation runs over all the kth-order graphs that appear in the formal perturbation series for $\langle \varphi_x \varphi_y \rangle$. By the asymptoticity of the formal perturbation series, we have (if we write $G_{xy}{}^{(\lambda)} = \langle \varphi_x \varphi_y \rangle$)

$$|G_{xy}{}^{(\lambda)} - G_{xy}{}^{(\lambda,N)}| \leqslant \lambda^{N+1} \text{const}_5. \tag{2.25}$$

Thus, if we denote by $\mathscr{A}^{(\lambda,N)}$ the amplitude obtained by replacing $G_{xy}{}^{(\lambda)}$ by $G_{xy}{}^{(\lambda,N)}$ in the definition of $\mathscr{A}^{(\lambda)}$, Eq. (2.9), we obtain

$$\Bigg| \sum_{k=0}^{N} (-\lambda)^k \sum_{G \in \mathscr{S}_{2p}{}^{(k)}} C_{\text{pert}}(G) \mathscr{A}^{(\lambda)}(G)_{(x_1,\ldots,x_{2p})} - \sum_{k=0}^{N} (-\lambda)^k \sum_{G \in \mathscr{S}_{2p}{}^{(k)}} C_{\text{pert}}(G) \mathscr{A}^{(\lambda,N)}(G)_{(x_1,\ldots,x_{2p})} \Bigg| \leqslant \lambda^{N+1} \text{const}_6. \tag{2.26}$$

Next substitute the definition (2.24) for $G^{(\lambda,N)}$ in $\mathscr{A}^{(\lambda,N)}$, and rewrite everything in terms of the Gaussian propagator $G_{xy}{}^{(0)}$. Then the one-to-one correspondence of the resulting graphs up to order N leads to the following "skeleton expansion vs formal expansion estimate":

$$\Bigg| \sum_{k=0}^{N} (-\lambda)^k \sum_{G \in \mathscr{S}_{2p}{}^{(k)}} C_{\text{pert}}(G) \mathscr{A}^{(\lambda,N)}(G)_{(x_1,\ldots,x_{2p})} - \sum_{k=0}^{N} (-\lambda)^k \sum_{G \in \varphi_{2p}{}^{(k)}} C_{\text{pert}}(G) \mathscr{A}^{(0)}(G)_{(x_1,\ldots,x_{2p})} \Bigg| \leqslant \lambda^{N+1} \text{const}_7. \tag{2.27}$$

The desired equation [(2.21)] follows from Eqs. (2.26) and (2.27). ■

III. APPLICATION TO CRITICAL PHENOMENA

We discuss some results obtained by applying the skeleton inequalities to statistical mechanical problems. For this purpose, we consider a translation invariant system with nearest neighbor interaction on a d-dimensional hypercubic lattice Z^d $(d>4)$. The interaction $\{J_{xy}\}$ satisfies $J_{xy} = J > 0$ if $|x-y| = 1$, and $J_{xy} = 0$ otherwise, and the potential $V(\varphi_x{}^2)$ is given by

$$V(\varphi_x{}^2) \equiv \frac{a}{2}\varphi_x{}^2 + \sum_{n=2}^{M} \frac{\lambda_{2n}}{(2n)!}\varphi_x{}^{2n}, \quad \lambda_{2n} \geqslant 0, \tag{3.1}$$

where the constants a, M, and λ_{2n} $(=\lambda g_{2n}$, in Sec. II) are now site independent. Moreover, we require an additional condition $a>0$.

Thermal expectation of the system is then constructed as an infinite volume limit[12] $\Lambda \to Z^d$ of the finite volume expectation (2.3), where Λ is a rectangular-parallelepiped sublattice of Z^d with periodic boundary condition. We denote such a limit expectation by $\langle \cdots \rangle$.

We will investigate the behavior of the following three thermodynamic quantities:

$$\chi \equiv \sum_{x \in Z^d} \langle \varphi_0 \varphi_x \rangle, \tag{3.2}$$

$$C \equiv \sum_{\substack{|x|=1 \\ |y-y'|=1}} \{\langle \varphi_0 \varphi_x \varphi_y \varphi_{y'} \rangle - \langle \varphi_0 \varphi_x \rangle \langle \varphi_y \varphi_{y'} \rangle\}, \tag{3.3}$$

$$|\tilde{u}_4| \equiv \sum_{x,y,z \in Z^d} |u_4(0,x,y,z)|, \tag{3.4}$$

where the truncated four-point function u_4 is defined as

$$u_4(x_1,x_2,x_3,x_4) \equiv \langle \varphi_{x_1}\varphi_{x_2}\varphi_{x_3}\varphi_{x_4} \rangle - \langle \varphi_{x_1}\varphi_{x_2} \rangle \langle \varphi_{x_3}\varphi_{x_4} \rangle - \langle \varphi_{x_1}\varphi_{x_3} \rangle \langle \varphi_{x_2}\varphi_{x_4} \rangle - \langle \varphi_{x_1}\varphi_{x_4} \rangle \langle \varphi_{x_2}\varphi_{x_3} \rangle,$$

and χ and C are the susceptibility and specific heat, respectively.

For the lattice dimensionality $d \geqslant 2$, it has been rigorously established that, these systems possess critical points[12,21–23] characterized by[2,24] (among other things)

$$\chi \to \infty \quad \text{as } J \to J_c - 0, \tag{3.5}$$

with finite and nonzero J_c. Accordingly, we define[25] the critical exponents γ, α, and Δ_4, which characterize the critical behavior of the quantities (3.2)–(3.4):

$$\gamma \equiv \lim_{J \to J_c - 0} -\ln \chi / \ln(J_c - J), \tag{3.6}$$

$$\alpha \equiv \lim_{J \to J_c - 0} -\ln C / \ln(J_c - J), \tag{3.7}$$

$$2\Delta_4 + \gamma \equiv \lim_{J \to J_c - 0} -\ln|\tilde{u}_4| / \ln(J_c - J). \tag{3.8}$$

These definitions are formal ones, since the existence of the limits is not known. But for some suitable systems in high dimensions, we have the following theorem.

Theorem 3.1: In the system (3.1) with $d>4$, critical exponents γ, α, and Δ_4 exist and satisfy the equalities

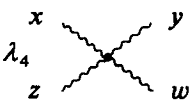
$$\gamma = 1, \quad \alpha = 0, \quad \Delta_4 = \tfrac{3}{2},$$

if the couplings λ_{2n} are sufficiently small.[27]

Proof: We first illustrate the proof of the equality $\Delta_4 = \frac{3}{2}\gamma$ in $\lambda_4\varphi^4$ systems (i.e., $M=2$). In this case, the first three skeleton inequalities applied to $u_4(x,y,z,w)$ are

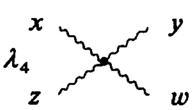
$$u_4(x,y,z,w) \leqslant 0, \tag{3.9}$$

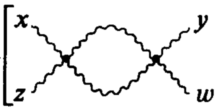
$$u_4(x,y,z,w) \geqslant -\lambda_4 \sum_u \langle \varphi_x \varphi_u \rangle \langle \varphi_y \varphi_u \rangle \langle \varphi_z \varphi_u \rangle \langle \varphi_w \varphi_u \rangle$$

$= -\lambda_4$ 

, (3.10)

$$u_4(x,y,z,w) \leqslant -\lambda_4 \sum_u \langle \varphi_x \varphi_u \rangle \langle \varphi_y \varphi_u \rangle \langle \varphi_z \varphi_u \rangle \langle \varphi_w \varphi_u \rangle + \frac{(\lambda_4)^2}{2} \sum_{u,v} [\langle \varphi_x \varphi_u \rangle \langle \varphi_y \varphi_u \rangle \langle \varphi_u \varphi_v \rangle^2 \times \langle \varphi_v \varphi_z \rangle \langle \varphi_v \varphi_w \rangle + (\text{two permutations})]$$

$= -\lambda_4$ 

$+ \dfrac{(\lambda_4)^2}{2}\Bigg[$ 

$+ \text{(two permutations)}\Bigg]$, (3.11) where, in the graphical descriptions, a wavy line stands for the two-point function and a dot is a shorthand for a summation over the lattice sites.

From the first-order skeleton inequality (3.10) and the definitions (3.4) and (3.2), one can easily see[28]

$$|\tilde{u}_4| < \lambda_4 \sum_{x,y,z,u} \langle\varphi_0\varphi_u\rangle\,\langle\varphi_x\varphi_u\rangle\,\langle\varphi_y\varphi_u\rangle\,\langle\varphi_z\varphi_u\rangle = \lambda_4\chi^4. \tag{3.12}$$

Similarly, from the second-order skeleton inequality (3.11), we have

$$|\tilde{u}_4| > \lambda_4\chi^4 - \frac{3}{2}(\lambda_4)^2\sum_x\langle\varphi_0\varphi_x\rangle^2\chi^4 = \lambda_4\Big(1 - \frac{3}{2}\lambda_4\sum_x\langle\varphi_0\varphi_x\rangle^2\Big)\chi^4. \tag{3.13}$$

If the coefficient of χ^4 on the right-hand side is finite and nonzero as $J \to J_c - 0$, we can conclude

$$|\tilde{u}_4| \sim \chi^4 \quad \text{as } J \to J_c - 0,$$

and, from Eqs. (3.5) and (3.7), we can obtain a critical exponent equality $\Delta_4 = \frac{3}{2}\gamma$, which, with $\gamma = 1$, gives the desired equality $\Delta_4 = \frac{3}{2}$.

To estimate the coefficient in question, we use infrared bounds of Fröhlich, Simon, and Spencer,[21,22]

$$\begin{aligned}\sum_x\langle\varphi_0\varphi_x\rangle^2 &= \sum_x\Big\{(2\pi)^{-d/2}\int d^dk\, G(k)e^{ikx}\Big\}^2 \\ &= \int d^dk\,|G(k)|^2 \\ &< \frac{1}{J^2}\int_{-\pi}^{\pi}\frac{d^dk}{(2\pi)^d}\Big[2\sum_{i=1}^d(1-\cos k_i)\Big]^{-2} = \frac{1}{J^2}C_2(d).\end{aligned}$$

Note that $C_2(d)$ is a finite constant in $d > 4$ dimensions. Combining this with the bound of the critical temperature[31] $J_c > a/(2d)$, we can bound the coefficient in Eq. (3.13) (in the limit $J \to J_c$) as

$$1 - \frac{3}{2}\frac{\lambda_4}{a^2}(2d)^2C_2(d) < \Big(1 - \frac{3}{2}\lambda_4\sum_x\langle\varphi_0\varphi_x\rangle^2\Big) < 1.$$

Now, it is easily observed that the left-hand side of the above inequality can be made strictly positive by letting λ_4 become sufficiently small compared with a.

Proofs of the equalities $\gamma = 1$ and $\alpha = 0$ can be given in a similar way.[5,26] Representations

$$\frac{d}{dJ}(\chi^{-1}) = -\frac{1}{2\chi^2}\sum_{x,|y-y'|=1} u_4(0,x,y,y') - 2d, \tag{3.14}$$

$$C = \sum_{\substack{|x|=1\\ |y-y'|=1}}\{u_4(0,x,y,y') + 2\langle\varphi_0\varphi_y\rangle\,\langle\varphi_x\varphi_{y'}\rangle\}, \tag{3.15}$$

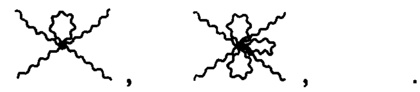
together with the zeroth- and first-order skeleton inequalities, (3.9) and (3.10), imply (again for $d > 4$ and sufficiently small λ_4)

$$\frac{d}{dJ}(\chi^{-1}) \sim \text{const}, \quad C \sim \text{const}, \quad \text{as } J \to J_c - 0,$$

which reduce to the exponent equalities $\gamma = 1$ and $\alpha = 0$.

Extensions of the above proofs to the general $\Sigma\lambda_{2n}\varphi^{2n}$-systems can be done almost automatically, if one takes care of the following few points.

In the first-order skeleton inequality corresponding to Eq. (3.10), one now finds the terms like



which contain bubbles

$$= \langle\varphi_0^{\,2}\rangle.$$

But a bubble can be controlled by infrared bounds as

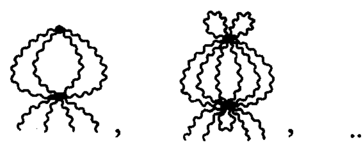
$$\langle\varphi_0^{\,2}\rangle < \frac{1}{J}\int\frac{d^dk}{(2\pi)^d}\frac{1}{2\Sigma_{i=1}^d(1-\cos k_i)},$$

where the integral converges if $d > 2$.

In the second-order skeleton inequality [corresponding to Eq. (3.11)], along with the bubbled "fish" diagrams (with "spinals")

one has to deal with "squids"

and "quadrupuses"



These can be again controlled by infrared bounds in $d > 4$ dimensions. ■

Finally, we note that most of the critical exponent (in) equalities discussed in the present section have been obtained for various systems through different correlation inequalities. We summarize these results in Table I, where the ingredients of the proofs and the validities of the inequalities are listed.

Remarks: (1) Theorem 3.1 is also valid for the two-component systems with the same potential.

(2) The condition $\lambda_{2n} \geq 0$ is not always necessary for the proof of Theorem 3.1. For example, in a suitable φ^8 system with $\lambda_4, \lambda_8 > 0$ and $\lambda_6 < 0$, we can prove (by brute force) first- and second-order *skeletonlike* inequalities and Theorem 3.1. It may be interesting to extend skeleton inequalities to a much wider range of models, though the simple extension of our method does not seem to work.

(3) To prove the inequality $\gamma \leq 1$ from the Fröhlich inequality (see Table I), one has to employ numerical evaluations of the infrared integrals (see Ref. 8).

ACKNOWLEDGMENTS

We would like to thank Professor M. Kato for careful reading of the manuscript and useful discussions, Professor

TABLE I. Ingredients of the proof of critical exponent inequalities for various systems ($d>4$). IR = Infrared bounds,[21,22] Leb = Lebowitz inequality,[32,14] A = Aizenman inequality,[5] F = Fröhlich inequality,[8,9] AG = Aizenman–Graham inequality,[6] Sk = Skeleton inequalities,[1,13,10] Gr = Griffiths inequalities.[17,18]

	"soft" $V(\varphi^2)$	$\lambda\varphi^4$ $\lambda>0$	Ising	$V(\varphi^2)$ with $V''(x)>0$
$\gamma>1$		Leb[a]		
$\gamma<1$	Sk,IR[b]	(i) AG,IR (ii) F,IR	A,IR[c]	F,IR[b]
$\alpha<0$		Leb,IR[d]		
$\alpha>0$		Gr,IR[e]		
$\Delta_4<\frac{3}{2}\gamma$	(i) Sk (ii) F	(i) A (ii) F (iii) Sk[f]	A	F
$\Delta_4>\frac{3}{2}\gamma$	Sk,IR	...	...	...

[a] Reference 33.
[b] Reference 8.
[c] Reference 5.
[d] Reference 26.
[e] Reference 25.
[f] Reference 3.

M. Suzuki, Professor T. Eguchi, Doctor I. Ichinose, and K.-I. Kondo for valuable discussions, and the referee for instructive comments and suggestions. We would also like to thank Professor D. Brydges, Professor J. Fröhlich, Professor A. D. Sokal, Doctor A. Bovier, and Doctor G. Felder for sending us copies of their results prior to publication.

[1]D. Brydges, J. Fröhlich, and A. D. Sokal, Commun. Math. Phys. **91**, 117 (1983).

[2]D. Brydges, J. Fröhlich, and A. D. Sokal, Commun. Math. Phys. **91**, 141 (1983).

[3]A. D. Sokal, Ann. Inst. H. Poincaré A **37**, 317 (1982).

[4]M. Aizenman, Phys. Rev. Lett. **47**, 1 (1981).

[5]M. Aizenman, Commun. Math. Phys. **86**, 1 (1982).

[6]M. Aizenman and R. Graham, Nucl. Phys. B **225** [FS9], 261 (1983).

[7]C. Aragão De Carvalho, S. Caracciolo, and J. Fröhlich, Nucl. Phys. B **215** [FS7], 209 (1983).

[8]J. Fröhlich, Nucl. Phys. B **200** [FS4], 281 (1982).

[9]T. Hattori, J. Math. Phys. **24**, 2200 (1983).

[10]A. Bovier and G. Felder, Commun. Math. Phys. **93**, 259 (1984).

[11]The inequalities are also useful for the construction of continuum field theories in three dimensions with bare Lagrangian containing terms like $\rho_{2n}\varphi^{2n}$. Here the bare couplings ρ_{2n} are adjusted to order ϵ^{n-2}, where ϵ is the lattice spacing. (For the details, see T. Hara, "Construction of a Nontrivial Field Theory in 3 dimensions Starting from a Lagrangian of φ^6-type" University of Tokyo-Komaba 84-15.)

[12]D. Ruelle, *Statistical Mechanics—Rigorous Results* (Benjamin, New York, 1969).

[13]D. Brydges, Field theories and Symanzik's polymer representation," in *Gauge Theories—Fundamental Interactions and Rigorous Results* (Proceedings of the 1981 Poiana Brasov Summer School) edited by G. Dita, V. Georgescu, and R. Purice (Birkhäuser, Boston, 1982).

[14]D. Brydges, J. Fröhlich, and T. Spencer, Commun. Math. Phys. **83**, 123 (1982).

[15]F. Dunlop and C. M. Newman, Commun. Math. Phys. **44**, 223 (1975).

[16]J. Ginibre, Commun. Math. Phys. **16**, 310 (1970).

[17]R. B. Griffiths, J. Math. Phys. **8**, 478, 484 (1967).

[18]G. S. Sylvester, J. Stat. Phys. **15**, 327 (1976).

[19]Here, and in each following step, we construct not only the Nth-order inequality, but (formally) all the inequalities of order $M<N-1$, so that these $(N+1)$ inequalities together form a set of alternating bounds $\cdots\leq\Sigma(-\lambda)^k\cdots$. These inequalities for $M<N-1$ actually turn out to be identical to those used in the inductive proof. See Sec. II C.

[20]Of course this choice must be done without breaking the ferromagnetic condition $J_{xy}>0$ $(x\neq y)$. This will be satisfied if (for example) we first set $J_{xy}=J>0$ for all $x\neq y$ and $a_x=a$ sufficiently large, and vary the C_{xy}'s in some bounded region including the initial values.

[21]J. Fröhlich, B. Simon, and T. Spencer, Commun. Math. Phys. **50**, 79 (1976).

[22]J. Fröhlich, R. Israel, E. H. Lieb, and B. Simon, Commun. Math. Phys. **62**, 1 (1978).

[23]J. Fröhlich, R. Israel, E. H. Lieb, and B. Simon, J. Stat. Phys. **22**, 297 (1980).

[24]J. Rosen, Adv. Appl. Math. **1**, 37 (1980).

[25]As for the present definition[26] of α, which reflects the behavior of full specific heat rather than its singular part, the inequality $\alpha>0$ is valid for arbitrary systems in Z^d $(d>3)$ with Griffiths-class *a priori* measure. To prove the inequality, assume the converse, i.e., $C(J)\to 0$ as $J\to J_c$. Then from the Griffiths II inequalities, we have $\langle\varphi_0\varphi_1\rangle_{\rm crit}=\langle\varphi_0{}^2\rangle_{\rm crit}$. (Hereafter $\langle\cdots\rangle_{\rm crit}$ denotes $\lim_{J\to J_c}\langle\cdots\rangle_J$.) Inserting this into the Gell-Mann–Low representation[3] $\langle\varphi_0\varphi_{(n,0,0,\ldots)}\rangle_{\rm crit}=(\Phi_0\Omega,T^n\Phi_0\Omega)$ (T is a Hermitian operator with $|T|<1$), we have $T\Phi_0\Omega=\Phi_0\Omega$ and thus $\langle\varphi_0\varphi_{(n,0,0,\ldots)}\rangle_{\rm crit}=\langle\varphi_0{}^2\rangle_{\rm crit}$ for all n. This contradicts with the infrared bounds[3,21,22] $\langle\varphi_0\varphi_x\rangle_{\rm crit}<{\rm const}\,|x|^{2-d}$. We are grateful to the referee for suggesting to us the present proof.

[26]A. D. Sokal, Phys. Lett. **71**, 451 (1979).

[27]One can write down this smallness condition in the form

$$1>\sum_{n=2}^{M}{\rm const}(M,n,d)\frac{\lambda_{2n}}{a^n}.$$

[28]Actually, Eq. (3.12) [and all the other similar relations such as Eq. (3.14)] must be proved in a finite lattice (with periodic boundary condition) for the corresponding quantities such as $\chi_\Lambda\equiv\Sigma_{x\in\Lambda}\langle\varphi_0\varphi_x\rangle_\Lambda$. If one "works hard[2]" using some correlation inequalities (e.g., the Simon–Lieb inequality[14,29,30] and the first-order skeleton inequality), one can prove that these finite volume quantities converge to the infinite volume quantities (3.2)–(3.4) in the high-temperature region (characterized by $\chi<\infty$), and the relations carry over to the infinite systems.

[29]E. H. Lieb, Commun. Math. Phys. **77**, 127 (1980).

[30]B. Simon, Commun. Math. Phys. **77**, 111 (1980).

[31]To prove this bound note that if $J<a/(2d)$, a Gaussian system defined by letting $\lambda_{2n}=0(n>2)$ is well defined. Then the Griffiths II inequality implies $\chi<\chi_{\rm Gaussian}<\infty$ and the system is in its high-temperature region.

[32]J. L. Lebowitz, Commun. Math. Phys. **35**, 87 (1974).

[33]J. Glimm and A. Jaffe, Phys. Rev. D **10**, 536 (1974).

A multitype random sequential process. III. The case of constant target area

B. Mellein
INIFTA, División Química Teórica, Sucursal 4, Casilla de Correo 16, 1900 La Plata, Argentina

(Received 15 January 1985; accepted for publication 26 April 1985)

We consider the static properties of a sequential process where the compartments of a $1\times n$ lattice space are filled irreversibly with particles of integral random length α (α-bell particles, α-mers; $1 \leqslant q \leqslant \alpha \leqslant r, r \geqslant 2$). While, in a previous model, filling was assumed to be random on the occasionally accessible (yet unoccupied) part of the lattice (shrinking target area), particle placing is now assumed to be random on the *entire* array at *any time* (constant target area) and subject to the condition of no overlap, i.e., particles striking already filled sites will be rejected. The occupation statistics of the lattice in the jammed state is analyzed by means of three random variables, (i) the total number of empty sites, (ii) the number of α-bell particles forming part of the saturation coverage ($\alpha = q,...,r$), and (iii) the number of vacancies of m sites ($m = 0,1,...,q-1$). Recursion relationships are obtained for the expectation values of these random variables and their behavior for $n\to\infty$ is studied. The results are used to describe the size distribution of adsorbed particles on infinite arrays.

I. INTRODUCTION

In a previous paper[1] two distinct multitype random sequential processes, designated model I and model II, have been suggested as possible generalizations of the classical one-type sequential filling process.[2–6] The common feature of both models is as follows. A finite one-dimensional array of equivalent compartments is to be filled with particles of random lengths subject to the following conditions: (i) once a particle has been placed, its position remains permanently fixed, and (ii) no two particles overlap. Addition of particles takes place sequentially and randomly but is carried out in two different ways: In model I, in any attempt to place a particle, the entire lattice space is assumed to be target area. Consequently, due to the condition of no overlap, a particle of length k, say, will stick only if striking a stretch of k contiguous yet unoccupied compartments; otherwise it will be rejected. In model II filling instructions square with the condition of no overlap beforehand: In a given attempt to place a particle only the still accessible part of the lattice space is taken as target area.

The random sequential process in the setup of model I may be thought of as describing the complete monolayer adsorption of multicomponent gas or liquid mixtures into parallel troughs of suitable crystal surfaces.[1–3] Since a molecule (generally believed to contact the surface in a spatially random manner) will be adsorbed only if landing on a vacant trough segment, the filling procedure as postulated in model I seems to fit in with the real situation particularly well. Clearly, the sequential approach is only reasonable if the molecules interact with the surface so strongly that they stick without diffusion.

Another example of model I type we have in mind is that of the reaction kinetics [7–9] of a long chain molecule (array) carrying reactive substituents or groups (compartments), which, when exposed to a multicomponent mixture of reagents, may suffer various kinds of reactions (bonding). A certain type of reaction (assumed to affect a specific number of substituents) is represented by the placement of a particle of respective size. If reactions are irreversible the sequential approach appears to be justified while the assumption of randomness obviously requires the absence of inhibitory or activitory neighboring-group effects.[10,11] However, reacted chain segments, on obstructing potential reactions (those which would be possible if the participation of already reacted groups in later reactions were allowed for), will have some influence on the activities (and hence on further reactions) of the different solute species in bulk solution. Since this fact is taken into consideration in model I, the placement conception of model I seems to be preferable to that of model II.

Contrarily, with reference to the modeling of crystallization of linear chains[12,13] and finite cascade processes[14] one would choose model II rather than model I, since placing a particle may then be identified with subdividing the array into two independent subarrays.

Most interesting two-dimensional random sequential placing problems arise when considering particles in cell membranes[15] and protein adsorption on surfaces such as glass or metals.[16–18] However, in two or more dimensions[5,19] analytical results are scarce and most of our information stems from numerical simulations.

As pointed out earlier,[1] the two models under consideration not only differ in their (manifestly unlike) kinetic behaviors but also in the static properties of their occupation configurations in the jamming limit (see Fig. 2 and Table I of Ref. 1). Quantities which give some valuable insight into the final state configuration of the lattice space are, e.g., the total number of occupied sites (extent of reaction), the total number of particles of some given kind finally placed, and the total number of gaps (stretches of vacant compartments) of some specified length ultimately present.

In model II, after placing the first particle, the initial problem is reduced to two independent problems of the same type. On account of this fact it is rather easy to derive fairly simple difference equations for the variables of interest.[1,20] In model I, owing to the unchanging target area, the final occupation configuration of some subarray will not only depend on the subarray's length but also on the extension of the entire lattice space the subarray belongs to. It still will be possible to derive difference equations for the quantities we are interested in, but they will depend on the length of the initial lattice space (see Secs. III and V); this complicates an asymptotic analysis.

Certain aspects of the above-mentioned random variables have been studied already within the framework of model II.[1,20] In the present paper we will be interested in a similar analysis of model I. The results obtained will be compared with the corresponding ones of model II. In Sec. II a precise description of model I will be given and results, deduced in subsequent sections (Secs. V–VII), will be put together in Sec. III. In Sec. IV we present some examples and illustrations.

II. THE MODEL

The formal description of model I is as follows. We consider an initial one-dimensional array of n equivalent compartments (sites). From the probability distribution $P=\{p_q,...,p_r\}$ on $\{q,q+1,...,r-1,r\}$ with $q\geqslant 1$, $r\geqslant 2$, $p_q>0$, $p_{q+1}\geqslant 0,..., p_r\geqslant 0$, we sample integers $\alpha_1,\alpha_2,...$ and proceed to place an α_1-bell particle (a particle that occupies α_1 adjacent compartments) on the $1\times n$ array at random, i.e., the particle's left-hand end point occupies any of the sites $1,2,...,n-\alpha_1+1$ with equal probabilities $1/(n-\alpha_1+1)$. We then make a random attempt (independent of the first) to place an α_2-bell particle. Its placement, however, will only be realized if the condition of no overlap (with the already placed α_1-bell particle) is fulfilled. Otherwise we discard the particle and try to place (independently and at random) an α_3-bell particle instead. We thus continue until the α_1-bell and some α_k-bell particle, $k=2,3,...$, have no sites in common. A random number of placement attempts then follow until for the first time some α_{k+j}-bell particle, $j=1,2,...$, intersects neither the α_1-bell nor the secondly placed α_k-bell particle. The process of placing nonintersecting random particles thus goes on and is finished when no further particle fits, i.e., when all holes are made up of less than q sites. At this stage we say that the lattice space is saturated or in the jammed state.

Clearly, at the beginning of this "trial and error" filling process, attempts at inserting a particle will be highly successful but as the coverage of the lattice space increases checks will become more and more frequent. The adsorption dynamics of model I has been studied by McQuistan and Lichtman[6] in the most simple (one-type) case of dumbbells, while the kinetics of related continuous-time models for polymer reactions has been considered by various authors.[7,8,10,11,21]

III. NOTATIONS AND RESULTS

In the present paper we will analyze the static properties of model I by means of the following random variables, which refer to the *jammed* state of an initially unoccupied $1\times n$ array: A_n, the total number of vacant compartments; B^i_n, the total number of i-bell particles placed, $i=q,...,r$; and C^m_n, the total number of m-gaps (runs of *exactly* m vacant sites), $m=0,1,...,q-1$. These random variables are obviously connected through the relations

$$n-\sum_{i=q}^{r} iB^i_n=A_n=\sum_{m=0}^{q-1} mC^m_n, \tag{3.1}$$

and the main interest will be in their expectation values, denoted by a_n, b^i_n, and c^m_n, respectively.

Setting

$$\pi^j_{k,n}=\begin{cases}\left.\begin{array}{ll}\dfrac{p_j/(n-j+1)}{\Sigma^k_{\nu=q}p_\nu(k-\nu+1)/(n-\nu+1)}, & \text{if } j=q,\dots,k\\ 0, & \text{if } j=k+1,\dots,r\end{array}\right\} & k=q,\dots,r-1,\\ \dfrac{p_j/(n-j+1)}{\Sigma^r_{\nu=q}p_\nu(k-\nu+1)/(n-\nu+1)}, \quad \text{if } j=q,\dots,r, & k=r,\dots,n,\end{cases} \tag{3.2}$$

and defining recursively (here and in the sequel an empty sum is given the value zero)

$$a_{k,n}=\begin{cases}k, & \text{if } k=1,\dots,q-1,\\ 2\displaystyle\sum_{\nu=q}^{r}\pi^\nu_{k,n}\sum_{j=1}^{k-\nu}a_{j,n}, & \text{if } k=q,\dots,n;\end{cases} \tag{3.3}$$

$$b^i_{k,n}=\begin{cases}0, & \text{if } k=1,\dots,i-1,\\ 2\displaystyle\sum_{\nu=q}^{r}\pi^\nu_{k,n}\sum_{j=i}^{k-\nu}b^i_{j,n}+(k-i+1)\pi^i_{k,n}, & \text{if } k=i,\dots,n;\end{cases} \tag{3.4}$$

and

$$c^m_{k,n}=\left.\begin{cases}\delta_{k,m}, & \text{if } k=0,\dots,m\\ 2\displaystyle\sum_{\nu=q}^{r}\pi^\nu_{k,n}\sum_{j=m}^{k-\nu}c^m_{j,n}, & \text{if } k=m+1,\dots,n\end{cases}\right\}\quad m=0,\dots,q-1 \tag{3.5}$$

($\delta_{k,m}$ denotes the Kronecker delta), we shall see in Sec. V that

$$a_n=a_{n,n}, \tag{3.6}$$

$$b^i_n=b^i_{n,n},\quad i=q,\dots,r, \tag{3.7}$$

and

$$c^m_n=c^m_{n,n},\quad m=0,1,\dots,q-1. \tag{3.8}$$

As they stand, Eqs. (3.6)–(3.8) look somewhat mysterious but their origin is rather easy to understand since $a_{k,n}$, $b^{\,i}_{k,n}$, and $c^m_{k,n}$, $k = 1,\dots,n$, prove to be nothing else but the total average number of *unoccupied sites, i-bell particles*, and *m-gaps*, respectively, of a k-gap in the terminal state, having been filled as part of a $1\times n$ array.

The relative frequency of i-bell particles among the particles forming the saturation coverage of a $1\times n$ lattice space,

$$p^*_{i,n} = \frac{b^{\,i}_{n}}{\Sigma^r_{j=q} b^{\,j}_n}, \quad i = q,\dots,r, \tag{3.9}$$

and the relative probability of an m-gap in a saturated $1\times n$ array,

$$\rho_{m,n} = \frac{c^m_n}{\Sigma^{q-1}_{j=0} c^j_n}, \quad m = 0,\dots,q-1, \tag{3.10}$$

are, by virtue of (3.7) and (3.8), computable from the recurrence relations (3.4) and (3.5), respectively, and provide, together with Eqs. (3.3) and (3.6), a fairly satisfactory description of the mean saturation configuration of a finite array.

The mean occupation statistics of a saturated $1\times n$ lattice space, where n is large, may be described by means of two different limit processes. (i) We may let n tend to infinity and request information on the asymptotic behavior of a_n, $b^{\,i}_n$, and c^m_n. (ii) Considering a k-gap we may first let n and then k tend to infinity and ask about the asymptotic behavior of $a_{k,n}$, $b^{\,i}_{k,n}$, and $c^m_{k,n}$.

The latter procedure may be of particular interest in practical situations, e.g., experimental conditions may be such that only a relatively small section of the entire target area (chain molecule, trough) is accessible to observation. Data will then be available only for array segments.

From physical considerations we may expect that the terminal occupation statistics of a large array and a large gap within an even larger lattice space will not be really different. It is therefore not surprising that both limiting procedures lead to a unique limit value. More precisely, it will be seen that

$$\lim_{n\to\infty} a_n/n = \lim_{k\to\infty} a_{k,\infty}/k, \tag{3.11}$$

$$\lim_{n\to\infty} b^{\,i}_n/n = \lim_{k\to\infty} b^{\,i}_{k,\infty}/k, \quad i = q,\dots,r, \tag{3.12}$$

$$\lim_{n\to\infty} c^m_n/n = \lim_{k\to\infty} c^m_{k,\infty}/k, \quad m = 0,\dots,q-1, \tag{3.13}$$

where

$$a_{k,\infty} = \lim_{n\to\infty} a_{k,n}, \tag{3.14}$$

$$b^{\,i}_{k,\infty} = \lim_{n\to\infty} b^{\,i}_{k,n}, \quad i = q,\dots,r, \tag{3.15}$$

and

$$c^m_{k,\infty} = \lim_{n\to\infty} c^m_{k,n}, \quad m = 0,\dots,q-1. \tag{3.16}$$

However, it turns out that the mean values $a_{k,\infty}$, $b^{\,i}_{k,\infty}$, and $c^m_{k,\infty}$, in what refers to an asymptotic analysis ($k\to\infty$), are easier to handle than a_n, $b^{\,i}_n$, and c^m_n for $n\to\infty$. The reason is rather obvious. On account of Eqs. (3.6)–(3.8) the sequences $(a_n)_n$, $(b^{\,i}_n)_n$, and $(c^m_n)_n$ are in fact (diagonal) *double* sequences of a quite complicated structure [see Eqs. (3.3)–(3.5)]. Owing to this fact there seems to be no way to establish explicit analytic forms of their respective generating functions. On the contrary, this may be managed in the case of the sequences $(a_{k,\infty})_k$, $(b^{\,i}_{k,\infty})_k$, and $(c^m_{k,\infty})_k$ and gives rise to fairly accurate approximation formulas [see Eqs. (3.25)–(3.27) below]. Unfortunately, we do not see how these could be used to improve (if actually possible) the comparatively poor asymptotic results [Eqs. (3.31)–(3.33)] on $(a_n)_n$, $(b^{\,i}_n)_n$, and $(c^m_n)_n$.

To state explicit expressions for the common limit values in Eqs. (3.11)–(3.13) we must introduce some notations. Let $p_1 = p_2 = \cdots = p_{q-1} = 0$ and set

$$\theta = \sum_{j=q}^{r} jp_j, \tag{3.17}$$

$$\xi(x) = \sum_{j=1}^{r-1} \frac{x^j(1-\Sigma^j_{i=q}p_i)}{j}, \tag{3.18}$$

$$g_i(x) = x^{r-1}\left\{\sum_{j=1}^{r-i-1} p_j + \sum_{j=1}^{i+1} p_{r-i-1+j}\, x^j\right\},$$
$$i = 0,\dots,r-2, \tag{3.19}$$

$$t_i(x) = p_i x^{r-1}\{r-i-(r-i-1)x\}, \quad i = q,\dots,r, \tag{3.20}$$

$$R(x) = \sum_{k=1}^{r-2} a_{k,\infty}\, g_k(x), \tag{3.21}$$

$$T_i(x) = \sum_{k=i}^{r-2} b^{\,i}_{k,\infty}\, g_k(x), \quad i = q,\dots,r, \tag{3.22}$$

$$Q_m(x) = \sum_{k=m}^{r-2} c^m_{k,\infty}\, g_k(x), \quad m = 0,\dots,q-1. \tag{3.23}$$

From these definitions it is easily checked that

$$\lim_{x\to 0} R(x)x^{1-\theta} = 0. \tag{3.24}$$

Like relations hold for t_i, T_i, and Q_m. The existence of the integrals (lower limit) appearing in Eqs. (3.28)–(3.30) is therefore guaranteed.

In Sec. VI we shall obtain the following asymptotic forms. For any $0<\epsilon<p = 1/(r-1)$, as $k\to\infty$,

$$a_{k,\infty} = (k+\theta)a + O(k^{-k(p-\epsilon)}), \tag{3.25}$$

$$b^{\,i}_{k,\infty} = (k+\theta)b(i) - p_i + O(k^{-k(p-\epsilon)}),$$
$$i = q,\dots,r, \tag{3.26}$$

$$c^m_{k,\infty} = (k+\theta)c(m) + O(k^{-k(p-\epsilon)}), \quad m = 0,\dots,q-1, \tag{3.27}$$

with

$$a \equiv a(p_q,\dots,p_r) = \left\{2\int_0^1 (1-x)x^{-\theta}R(x)e^{2\xi(x)}\,dx + (r-1)\delta_{q,r}\right\}e^{-2\xi(1)}, \tag{3.28}$$

$$b(i) \equiv b(i,p_q,\dots,p_r) = e^{-2\xi(1)}\int_0^1 x^{-\theta}\{2(1-x)T_i(x) + t_i(x)\}e^{2\xi(x)}\,dx, \tag{3.29}$$

$$c(m) \equiv c(m,p_q,\dots,p_r) = \left\{2\int_0^1 (1-x)x^{-\theta}Q_m(x)e^{2\xi(x)}\,dx + \delta_{q,r}\delta_{m,r-1}\right\}e^{-2\xi(1)}. \tag{3.30}$$

Clearly, as a consequence of (3.11) and (3.25), (3.12) and (3.26), and (3.13) and (3.27) we have

$$\lim_{n\to\infty} a_n/n = a, \tag{3.31}$$

$$\lim_{n\to\infty} b^i_n/n = b(i), \quad i = q,\dots,r, \tag{3.32}$$

and

$$\lim_{n\to\infty} c^m_n/n = c(m), \quad m = 0,\dots,q-1, \tag{3.33}$$

respectively.

Concerning the variances of A_n, B^i_n, and C^m_n, we mention that it may be shown that, as $n\to\infty$,

$$\operatorname{Var} A_n \equiv \langle A^2_n\rangle - \langle A_n\rangle^2 = O(n), \tag{3.34}$$

$$\operatorname{Var} B^i_n = O(n), \quad i = q,\dots,r, \tag{3.35}$$

$$\operatorname{Var} C^m_n = O(n), \quad m = 0,\dots,q-1. \tag{3.36}$$

Just as in the case of the means, the random variables referring to gaps in *infinite* lattice spaces can be treated more accurately. For example, the variance of $A_{k,n}$, the number of unoccupied sites of a k-gap (in the jamming limit) filled as part of a $1\times n$ array, exhibits the following asymptotic behavior. There exist constants $d_1 > 0$ and d_2 such that, for every $0<\epsilon<p = 1/(r-1)$,

$$\lim_{n\to\infty} \operatorname{Var} A_{k,n} = d_1 k + d_2 + O(k^{-k(p-\epsilon)}), \quad \text{as } k\to\infty. \tag{3.37}$$

The proof of Eq. (3.37) is similar to that presented in Sec. V C of Ref. 1. Equations (3.34)–(3.36) may be deduced adopting the argument given in the second part of the following section (VI). As a general rule, the proofs are rather heavy and for that reason will not be given.

What is the probability that a particle chosen *at random* from those forming the saturation coverage of a $1\times n$ array, is an i-bell particle? In such an experiment, letting I_n denote the length of the selected particle, we are then interested in

$$P(I_n = i) = \left\langle \frac{B^i_n}{\sum_{j=q}^r B^j_n} \right\rangle, \quad i = q,\dots,r. \tag{3.38}$$

Generally, given two random variables X and Y, $\langle X/Y\rangle \neq \langle X\rangle/\langle Y\rangle$; in the present case, since B^i_n and $\sum_{j=q}^r B^j_n$ are obviously dependent, the equality [recall Eq. (3.9)] of $P(I_n = i)$ and $p^*_{i,n}$ cannot be expected either. Indeed, consider, for example, the two-type case $q = 2$, $r = 3$, and take $n = 4$ and $i = 3$. Then, obviously, $P(I_4 = 3) = p_3$ and $p^*_{3,4} = p_3/(p_3 + \frac{2}{3}p_2) \neq p_3$ if $p_2 \neq 0 \neq p_3$. For finite n, the expectation value on the right-hand side of (3.38) is difficult to determine. It is therefore helpful to know that the *limiting* distribution ($n\to\infty$) of I_n may be obtained from the relative frequencies $p^*_{i,n}$; more precisely,

$$\lim_{n\to\infty} P(I_n = i) = \lim_{n\to\infty} p^*_{i,n} \equiv p^*_i, \quad i = q,\dots,r. \tag{3.39}$$

Similarly, defining J_n as the length of a gap chosen *randomly* from those present on a $1\times n$ array in the terminal state, we shall see in Sec. VII that [see Eq. (3.10)]

$$\lim_{n\to\infty} P(J_n = m) = \lim_{n\to\infty} \rho_{m,n} \equiv \rho_m, \quad m = 0,\dots,q-1. \tag{3.40}$$

TABLE I. The relative frequencies $p^*_{i,n}$ and $\rho_{m,n}$ [Eqs. (3.9) and (3.10), respectively] corresponding to the three-type model (model I) $q = 2$, $r = 4$, $p_2 = p_3 = p_4 = \frac{1}{3}$, for various values of n.

n	$\rho_{0,n}$	$\rho_{1,n}$	$p^*_{2,n}$	$p^*_{3,n}$	$p^*_{4,n}$
3	0.7500	0.2500	0.5000	0.5000	0.0000
4	0.7500	0.2500	0.4545	0.2727	0.2727
5	0.6783	0.3217	0.5286	0.2571	0.2143
6	0.7358	0.2642	0.5687	0.2485	0.1828
10	0.7197	0.2803	0.5436	0.2685	0.1879
20	0.7181	0.2819	0.5330	0.2733	0.1937
50	0.7173	0.2827	0.5256	0.2763	0.1981
100	0.7170	0.2830	0.5230	0.2773	0.1997
200	0.7169	0.2831	0.5217	0.2778	0.2005
1000	0.7168	0.2832	0.5206	0.2783	0.2011
∞	0.7168	0.2832	0.5204	0.2784	0.2013

In Table I we list $p^*_{i,n}$ and $\rho_{m,n}$, corresponding to the three-type model $q = 2$, $r = 4$, $p_2 = p_3 = p_4 = \frac{1}{3}$, for some values of n. It is seen that for $n \geqslant 100$ all entries differ from the respective limit values by no more than 0.8%.

A final question we ask is the following. Choosing *at random* a site on a filled $1\times n$ array, what is the probability, $\varphi_{i,n}$, that it is occupied by an i-bell particle, $i = q,\dots,r$, or that it belongs to an i-gap, $i = 1,\dots,q-1$? In finite lattice spaces, end effects[2] (e.g., the qth compartment from either end of the array will always be occupied) complicate an answer seriously. However, as $n\to\infty$, end effects clearly disappear and, obviously [recall Eqs. (3.1), (3.32), and (3.33)],

$$\varphi_i \equiv \lim_{n\to\infty} \varphi_{i,n} = \begin{cases} ic(i), & i = 1,\dots,q-1,\\ ib(i), & i = q,\dots,r. \end{cases} \tag{3.41}$$

IV. ILLUSTRATIONS

In this section we specialize to some cases of particular interest and present various examples of models that include up to nine types of particles. Given some symbol defined within the present model I, x say, let us agree to write ${}_{\mathrm{I}}x$ instead, in order to distinguish it from the corresponding quantity of model II,[1,20] which will be denoted ${}_{\mathrm{II}}x$.

As pointed out earlier,[1] the two models do not differ in their static properties if filling is carried out with just one kind of object. Thus, as a check on our results we must meet with well-known formulas when specializing to the case $q = r$.

A. One-type filling

Let $q = r$, $p_r = 1$, i.e., $\theta = r$. It then follows from (3.18) and (3.19) that

$$\xi(x) = \sum_{j=1}^{r-1} \frac{x^j}{j}, \quad g_i(x) = x^{i+r}, \quad i = 0,1,\dots,r-2. \tag{4.1}$$

Since $a_{k,\infty} = k$, $k = 1,\dots,r-1$, we find from (3.21) and (3.28) that

$$a=\left\{r-1+2\int_0^1(1-x)\sum_{k=1}^{r-2}kx^ke^{2\xi(x)}\,dx\right\}e^{-2\xi(1)}. \tag{4.2}$$

This representation (one of various[1] possible) of the uncovered fraction of a saturated infinite lattice space has been stated in Ref. 22. Another one[1,3,22] comes up on observing (3.1) and making use of Eqs. (3.4), (3.20), (3.22), (3.29), (3.31), (3.32), and (4.1),

$$a=1-rb(r)=1-re^{-2\xi(1)}\int_0^1e^{2\xi(x)}\,dx. \tag{4.3}$$

Turning to gaps, Eqs. (3.5), (3.23), (3.30), and (4.1) yield the expressions first stated by Mackenzie[3]

$$c(m)=2e^{-2\xi(1)}\int_0^1(1-x)x^me^{2\xi(x)}\,dx,\quad m=0,1,\dots,r-1. \tag{4.4}$$

Note that $c(r-1)$ admits of the alternative representation[8,11,22]

$$c(r-1)=e^{-2\xi(1)}. \tag{4.5}$$

Furthermore, observe that substitution of (4.4) and (4.5) into $a=\sum_{m=0}^{r-1}mc(m)$ [this relation is a consequence of Eq. (3.1)] leads to (4.2) again.

B. The three-type model $q=2, r=4$

Recall that $q=2$ implies that $p_2>0$. Hence $a_{2,\infty}=0$ and since $a_{1,\infty}=1$ it follows from (3.19) and (3.21) that $R(x)=g_1(x)=x^3\{p_2+p_3x+p_4x^2\}$. Consequently,

$$\begin{aligned}{}_{\mathrm{I}}a&={}_{\mathrm{I}}a(p_2,p_3,p_4)\\&=2e^{-2\xi(1)}\int_0^1(1-z)\{p_2z+p_3z^2+p_4z^3\}\\&\quad\times z^{-p_3-2p_4}e^{2\xi(z)}\,dz\end{aligned} \tag{4.6}$$

with

$$2\xi(x)=2x+(p_3+p_4)x+\tfrac{2}{3}p_4x^3. \tag{4.7}$$

Figures 1 and 2 show ${}_{\mathrm{I}}a$ and ${}_{\mathrm{II}}a$ (see Ref. 1) as functions of p_3, keeping fixed $p_4=0$ and $p_2=0.15$, respectively. In the two-type situation of Fig. 1 (dimers and trimers only) we observe that there is no difference in the average saturation coverage of models I and II if (a) $p_2=1$, $p_3=0$ and (b) $p_2=0+$, $p_3=1-$. This is to be expected since (a) corresponds to the one-type case considered above and since (b) refers to the limiting case where filling is in stages: The regime "First trimers and then dimers" rules out the effect of particle competition peculiar to model I which puts the trimers at a disadvantage (with respect to model II trimers) and causes, if $0<p_3<1$, more isolated vacancies in model I than in model II (see Fig. 1). Not surprisingly, in the present three-type model, the average saturation coverage takes on its maximum value [equal to that of model II (see Ref. 1)] when filling is in stages: First 4-mers, then trimers, and finally dimers. More precisely,

$$\lim_{p_3\downarrow0}\Big(\lim_{p_2\downarrow0}{}_{\mathrm{I}}a(p_2,p_3,p_4)\Big)=2e^{-11/3}\int_0^1(1-z)z\times\exp\{2z+z^2+\tfrac{2}{3}z^3\}dz\simeq0.0505.$$

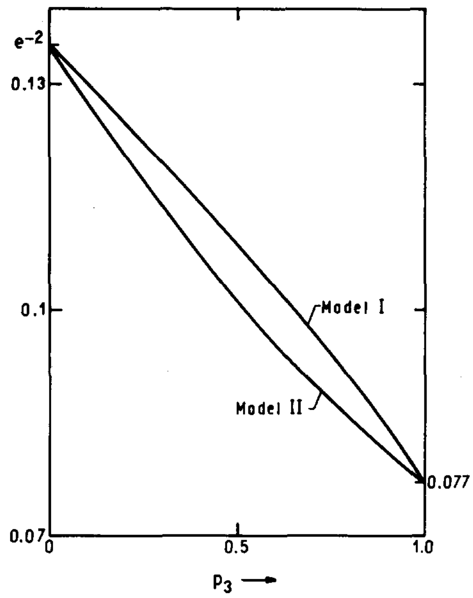


FIG. 1. In the two-type model $q=2$, $r=3$, the average uncovered fractions, ${}_{\mathrm{I}}a$ and ${}_{\mathrm{II}}a$ (model I and model II, respectively), of an infinite lattice space in the saturation limit, as functions of p_3.

From Eq. (3.29) we get the following expressions for the relative numbers of i-bell particles:

$$\begin{aligned}{}_{\mathrm{I}}b(i)&={}_{\mathrm{I}}b(i,p_2,p_3,p_4)\\&=e^{-2\xi(1)}\int_0^1f_i(x)x^{1-p_3-2p_4}e^{2\xi(x)}\,dx,\\i&=2,3,4,\end{aligned} \tag{4.8}$$

where

$$f_2(x)=(2-x)p_2+2x(1-x)[p_2+p_3x+p_4x^2],$$
$$f_3(x)=p_3,\ f_4(x)=p_4x,$$

and where $\xi(x)$ is stated explicitly in (4.7).

It is rather obvious from our model assumptions that whenever $0<p_2,p_4<1$ then ${}_{\mathrm{I}}b(2)>{}_{\mathrm{II}}b(2)$ and ${}_{\mathrm{I}}b(4)<{}_{\mathrm{II}}b(4)$, confirming once more the fact that model I favors the short particles. Concerning trimers, the situation is most interesting. Depending on the activity parameters p_2, p_3, and p_4,

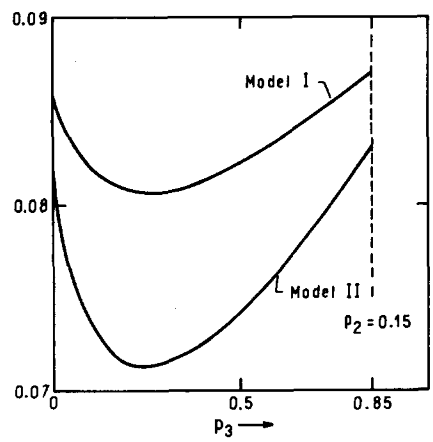


FIG. 2. In the three-type model $q=2$, $r=4$, $p_2=0.15$, the average uncovered fraction, ${}_{\mathrm{I}}a$ (${}_{\mathrm{II}}a$), of an infinite array filled due to model I (model II), as function of p_3. [${}_{\mathrm{I}}a$ is given explicitly in Eq. (4.6).]

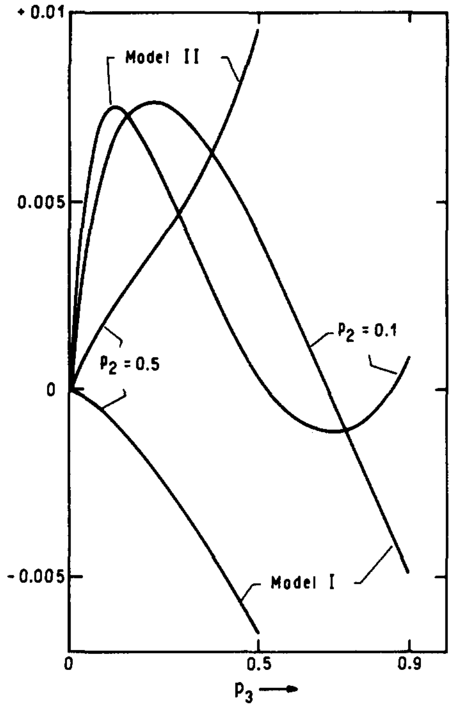


FIG. 3. In the three-type model $q = 2$, $r = 4$, $p_2 = 0.1$, the *difference between* ${}_{\rm I}b(3) = {}_{\rm I}b(3,p_3)$ [${}_{\rm II}b(3) = {}_{\rm II}b(3,p_3)$], the relative number of model I (model II) trimers that form part of the saturation coverage of an infinite lattice *and* $0.28p_3$, $0 \leqslant p_3 \leqslant 0.9$; in the three-type model $q = 2$, $r = 4$, $p_2 = 0.5$, the *difference between* ${}_{\rm I}b(3,p_3)$ [${}_{\rm II}b(3,p_3)$] *and* $0.29p_3$, $0 \leqslant p_3 \leqslant 0.5$. [${}_{\rm I}b(3)$ is stated explicitly in Eq. (4.8); see also text.]

their relative number may be greater in model I than in model II and vice versa: Figure 3 shows plots of [the following notation is introduced in Eq. (4.8)] $b(3,0.1,p_3,0.9\text{-}p_3)\text{-}0.28p_3$ and $b(3,0.5,p_3,0.5\text{-}p_3)\text{-}0.29p_3$, which reveal that (a) ${}_{\rm I}b(3)$ and ${}_{\rm II}b(3)$ differ very little (see scale) and (b) ${}_{\rm I}b(3) > {}_{\rm II}b(3)$ only if $p_2 + \epsilon_1 < p_3 < 1 - p_2 - \epsilon_2$ for some *positive* ϵ_1, ϵ_2. Evidently, in *both* models, there is a hierarchy of particles. Dimers are in the most favorable position and trimers predominate over 4-mers. In model II, this is exclusively due to the fact that dimers fit on 2-gaps, which are not accessible to both trimers and 4-mers, and trimers may stick to 3-gaps whereas 4-mers

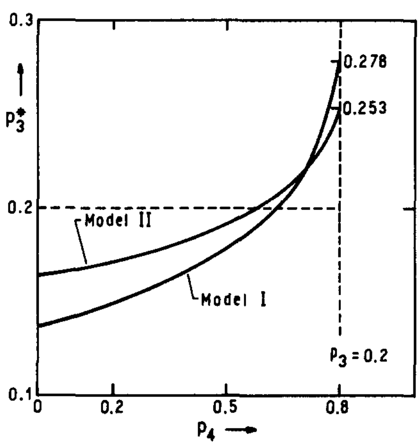


FIG. 4. In the three-type model $q = 2$, $r = 4$, $p_3 = 0.2$, the relative frequency, ${}_{\rm I}p_3^*({}_{\rm II}p_3^*)$, of model I (model II) trimers among the particles composing the saturation coverage of an infinite array, as function of p_4. [See Eq. (3.39).]

cannot. Thus (note that model II particles—as long as they may be accommodated—are treated in an equitable manner: a *chosen* particle *is placed*), hierarchy becomes effective only in the late stages of occupation when the lattice is close to saturation. On the contrary, in model I, hierarchy is pertinent to *any* placement attempt (except the first), since for particles, the shorter they are, the better possibilities of access they have. Taking this in mind, case (b) is not as strange as it looks: ${}_{\rm I}b(3) > {}_{\rm II}b(3)$ if trimers, being *more* active than dimers, gain an advantage (from the very beginning of the model I filling process) at the expense of 4-mers (in Fig. 3, for $p_2 = 0.1$, if $0.15 \leqslant p_3 \leqslant 0.73$, i.e., $0.17 \leqslant p_4 \leqslant 0.75$). However, this effect clearly disappears when 4-mers become rare (in Fig. 3, for $p_2 = 0.1$, if $p_4 < 0.17$, i.e., $0.73 < p_3 \leqslant 0.9$); then the (relative) superiority of (model I) dimers to (model I) trimers prevails and once more ${}_{\rm I}b(3) < {}_{\rm II}b(3)$.

Recalling (3.9), (3.32), and (3.39) and utilizing (4.8) the probabilities p_i^*, $i = 2,3,4$, may be calculated. Figure 4 shows ${}_{\rm I}p_3^*$ and ${}_{\rm II}p_3^*$ as functions of p_4 for fixed $p_3 = 0.2$. In both models, $p_3^* \geqslant p_3 = 0.2$ if dimers are sufficiently rare ($p_2 \leqslant 0.17$ in model I and $p_2 \leqslant 0.22$ in model II). Due to the effect described above, if 4-mers are predominant in number ($0.705 < p_4 \leqslant 0.8$), model I trimers take profit and ${}_{\rm I}p_3^* > {}_{\rm II}p_3^*$.

C. Some nine-type models

The characteristic features of some nine-type models are shown diagrammatically in Figs. 5–7. In (a) is given the graph of the particle ("input," "activity") distribution $P_q = \{p_q,\ldots,p_{q+8}\}$, (b) and (c) show the probabilities [see Eq. (3.41)] ${}_{\rm I}\varphi_i$ and ${}_{\rm II}\varphi_i$, $i = 1,\ldots,q+8$, respectively, and in (d) and (e) are plotted the particle "output" distributions [see Eq. (3.39)] ${}_{\rm I}p_k^*$ and ${}_{\rm II}p_k^*$, $k = q,\ldots,q+8$, respectively. In Figs. 5 and 6, $q = 2$, whereas $q = 10$ in Fig. 7.

The "input" distribution P_2 in Fig. 6 (ii) is chosen such that the "output" distribution ${}_{\rm I}P_2^* = \{{}_{\rm I}p_2^*,\ldots,{}_{\rm I}p_{10}^*\}$ is nearly uniform.

It is particularly worthwhile to compare Fig. 5 (i) with Fig. 7. In both, the "input" distribution is symmetrically binomial, with centers in 6 and 14, respectively. But while in the former case "input" and "output" distributions (related to model I as well as model II) are rather unlike, they resemble each other in shape in the latter situation. Clearly, in the model of Fig. 5 (i) particles are very much unlike (the longest exceeds in length the smallest by a factor 5), while in that of Fig. 7 differences in particle lengths are relatively small (the factor referring to extremes is now 1.8 only), making hierarchy effects more tenuous.

V. RECURSION RELATIONSHIPS

The proofs of Eqs. (3.6)–(3.8) are all in the same spirit and it will therefore suffice to establish Eq. (3.7), say.

Before tackling the general case let us consider the following simple situation. Let $n = 5$, $q = 2$, $r = 3$, and $p_2 = p_3 = \frac{1}{2}$, i.e., a 1×5 lattice is to be saturated by equally active dimers and trimers. Remaining a 3-gap to be filled (suppose, e.g., that compartments 1 and 2 are already occupied by a dimer), will the two types of particles have equal chances to get stuck? Because of their coinciding striking

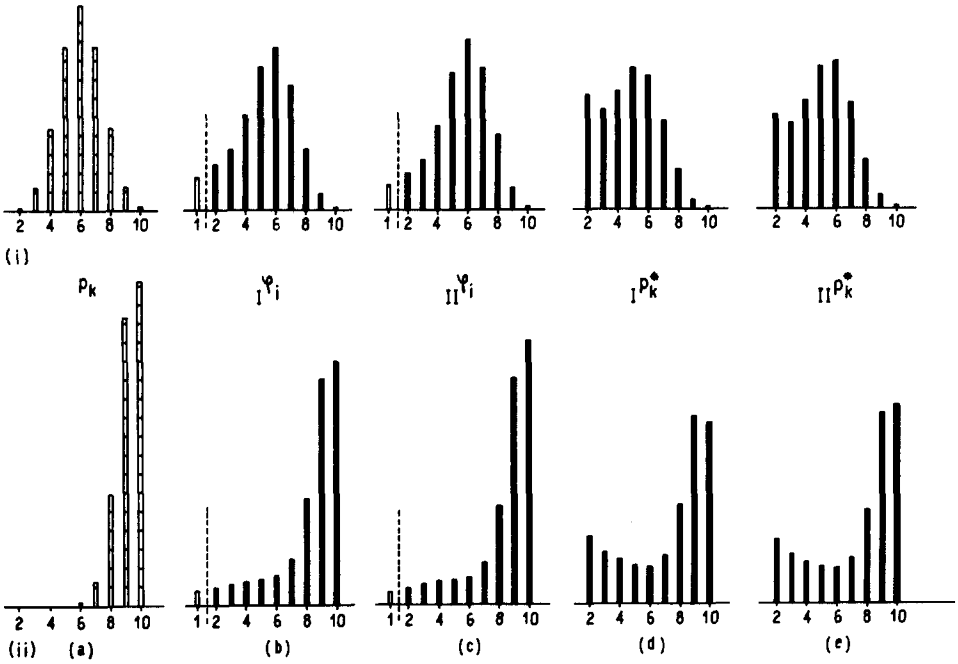


FIG. 5. Two nine-type models with $q = 2$ and $r = 10$. The particle ("input") distribution [diagram (a)] is in (i) binomial with parameter $p = 0.5$ and in (ii) binomial with parameter $p = 0.9$. In (b) and (c) are shown the probabilities φ_i [see Eq. (3.41)], referring to models I and II, respectively, and in (d) and (e) the "output" probabilities p_k^* [see Eq. (3.39)] of model I and model II, respectively.

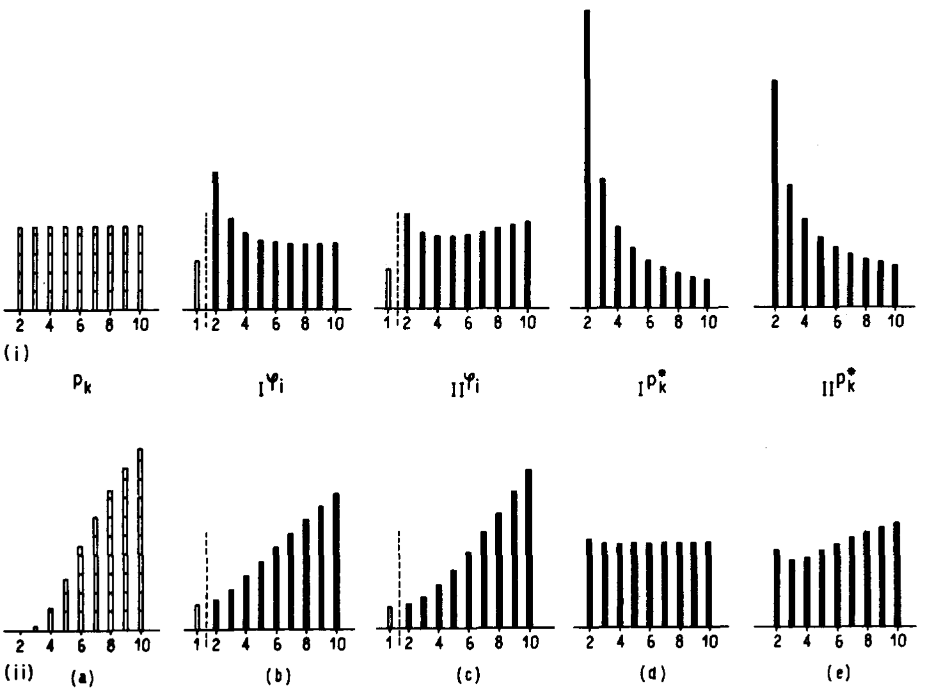


FIG. 6. The nine-type model $q = 2$, $r = 10$, with uniform "input" distribution in (i) (a) and (approximately) uniform "output" distribution (model I) in (ii) (d). (See symbol explanation in Fig. 5.)

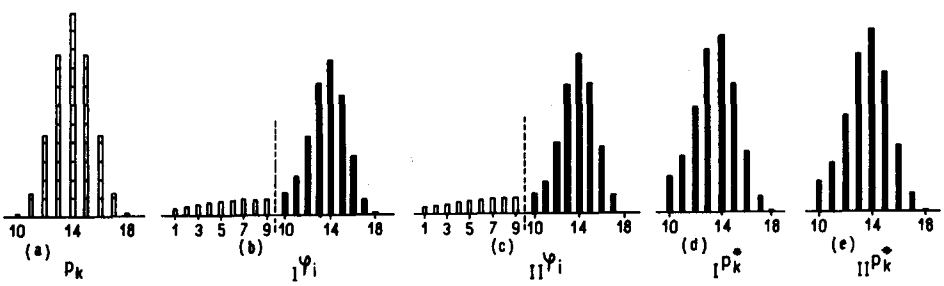


FIG. 7. The nine-type model $q = 10$, $r = 18$, with symmetrical binomial particle ("input") distribution in (a). (See symbol explanation in Fig. 5.)

frequencies one might give an affirmative answer. Equation (5.1) confirms the opposite, however. Clearly, dimers and trimers are equally likely to take part in a given trial, but while a trimer sticks to the triple vacancy only with probability $\frac{1}{3}$, a dimer does with probability $\frac{1}{2}$ (when striking sites 3 and 4 or sites 4 and 5). Obviously, these probabilities refer to *individual* strokes (involving dimers and trimers, respectively), but they suggest that the trimers are in an unfavorable condition also as final occupation is concerned. Indeed, the probability of the event

E: "A trimer (finally) lands in the 3-gap" is less than $\frac{1}{2}$, namely

$$P(E)=\tfrac{2}{5}=\tfrac{1}{2}\cdot\tfrac{1}{3}/(\tfrac{1}{2}\cdot\tfrac{1}{3}+\tfrac{1}{2}\cdot\tfrac{2}{4}). \tag{5.1}$$

To see the validity of (5.1) we may decompose the event $E=U_{j=1}^{\infty}E_j$ into the mutually exclusive events

E_j: "A trimer sticks to the 3-gap in the jth attempt," whose probabilities are given by

$$P(E_j)=\tfrac{1}{2}\cdot\tfrac{1}{3}(\tfrac{1}{2}\cdot\tfrac{2}{3}+\tfrac{1}{2}\cdot\tfrac{2}{4})^{j-1},\quad j=1,2,\dots\ . \tag{5.2}$$

Since $P(E)=\Sigma_{j=1}^{\infty}P(E_j)$, Eq. (5.1) follows from Eq. (5.2). Another argument goes as follows. Due to our model assumptions any attempt which results in the rejection of a particle does not change the odds of E, i.e., relevant to the occurrence or nonoccurrence of E are only those attempts in which a particle becomes fixed. Using the relative frequency of these trials [which equals the denominator in the quotient on the rhs of (5.1)] as a standard of measurement and taking into account that outcomes (there is only one) favorable to E occur with probability $\frac{1}{2}\cdot\frac{1}{3}$ we get (5.1) again.

In the above example, if we shift the 3-gap to another position within the 1×5 array (e.g., assuming that sites 1 and 5 are not accessible), Eq. (5.1) remains unchanged. However, if we imagine the 3-gap to be within a $1\times n$ array instead, then

$$P(E)=\frac{\frac{1}{2}\cdot[1/(n-2)]}{\frac{1}{2}\cdot[1/(n-2)]+\frac{1}{2}\cdot[1/(n-1)]},\quad n=3,4,\dots, \tag{5.3}$$

i.e., the odds of E change. Thus, contrary to the situation in model II, the occupation statistics of an initial $1\times k$ array and a k-gap situated within a larger $1\times n$ lattice space will be different. It is this point that makes the recursion schemes behind Eqs. (3.6)–(3.8) somewhat complicated. Clearly, still accessible small-size gaps will be not infrequent in a large array whose coverage is slightly below the jamming limit. It is therefore worthwhile to note that the expression on the rhs of (5.3), as $n\to\infty$, tends to $\frac{1}{2}$, i.e., the inferiority of the trimers then becomes even more pronounced.

Now turning to the general case, fix $i\in\{q,\dots,r\}$ and $k\in\{1,\dots,n\}$. Consider a k-gap (its exact position does not matter) within a $1\times n$ lattice space and number its compartments from left to right $1,2,\dots,k$. This k-gap will be destroyed if a first ν-bell particle [$\nu=q,\dots,\min(k,r)$] sticks to any of its $k-\nu+1$ ν-fold vacancies. Since particles are chosen independently according to the probability distribution $P=\{p_q,\dots,p_r\}$ and since placements are directed independently and randomly to the *entire* $1\times n$ array this will happen—in a given attempt—with probability

$$\sum_{\nu=q}^{\min(k,r)}p_\nu\frac{k-\nu+1}{n-\nu+1}$$

(which is zero if $k<q$). Hence, for any $m=1,\dots,k-j+1$ the probability of the event

$E_{k,n}^{j,m}$: "The first particle sticking to the k-gap in question is a j-bell particle occupying positions $m,m+1,\dots,m+j-1$"

is seen to be given by $\pi_{k,n}^{j}$ as defined by Eq. (3.2). Defining $B_{k,n}^{i}$ as the number of i-bell particles that have been placed on the k-gap when the latter (as part of the $1\times n$ array) has become saturated, we therefore find

$$b_{k,n}^{i}\equiv\langle B_{k,n}^{i}\rangle=\sum_{\nu=q}^{r}\sum_{j=1}^{k-\nu+1}\langle B_{k,n}^{i}|E_{k,n}^{\nu,j}\rangle\pi_{k,n}^{\nu}. \tag{5.4}$$

Since the conditional expectation

$$\langle B_{k,n}^{i}|E_{k,n}^{\nu,j}\rangle=b_{j-1,n}^{i}+b_{k-j-\nu-1,n}^{i}+\delta_{i,\nu},$$
$$j=1,\dots,k-\nu+1, \tag{5.5}$$

and since obviously $b_{k,n}^{i}=0$, $k=1,\dots,i-1$, Eq. (3.4) follows. The observation that $B_{n,n}^{i}=B_n^{i}$ then yields (3.7).

VI. ASYMPTOTICS OF THE MEANS

Recalling (3.14) we conclude from Eqs. (3.2) and (3.3) that

$$a_{k,\infty}\sum_{j=q}^{r}(k-j+1)p_j$$
$$=2\sum_{\nu=q}^{r}p_\nu\sum_{j=1}^{k-\nu}a_{j,\infty},\quad k=r-1,r,\dots, \tag{6.1}$$

which, on introducing the generating function

$$G(s)=\sum_{k=r-1}^{\infty}a_{k,\infty}s^k, \tag{6.2}$$

may be converted into the first-order linear differential equation

$$sG'(s)=G(s)\left\{\theta-1+\frac{2sg(s)}{1-s}\right\}+\frac{2}{1-s}R(s), \tag{6.3}$$

where θ and R are as defined in (3.17) and (3.21), respectively, and

$$g(s)=\sum_{j=q}^{r}p_js^{j-1}. \tag{6.4}$$

Substituting [see Eq. (3.18) for the definition of ξ]

$$G(s)=s^{\theta-1}e^{-2\xi(s)}H(s)/(1-s)^2 \tag{6.5}$$

into (6.3) and utilizing the fact that[1]

$$\xi'(x)=\{1-g(x)\}/(1-x),\quad x\neq 1, \tag{6.6}$$

yields the even simpler equation

$$H'(s)=2(1-s)s^{-\theta}e^{2\xi(s)}R(s). \tag{6.7}$$

As seen from (3.3), (6.2), and (6.5), the initial condition to be imposed is

$$H(0)=\begin{cases}0, & \text{if } q<r,\\ r-1, & \text{if } q=r,\end{cases} \tag{6.8}$$

and it leads to the solution

$$H(s) = (r-1)\delta_{q,r} + 2\int_0^s (1-x)x^{-\theta}R(x)e^{2\xi(x)}\,dx. \quad (6.9)$$

Recalling (3.24) and furthermore noting that θ is integral if $q = r$, $s^{\theta-1}e^{-2\xi(s)}H(s)$ is seen to be an integral function[23] of order at most $r-1$ (it *is* of order $r-1$ if $p_r > 0$). Additionally observing that $g'(1) = \theta - 1 = \xi'(1)$ and reasoning just as in Ref. 20 we obtain Eq. (3.25), with $a = H(1)e^{-2\xi(1)}$, as a consequence[23] of (6.2) and (6.5).

Equations (3.26) and (3.27) may be proved analogously. We omit the details.

To establish (3.11) [and hence, by (3.25), Eq. (3.31)] we shall apply the concept of uniform convergence. Let $n \gg r$ and introduce the probability distribution $\{p_q(n),...,p_r(n)\}$ by means of the definition

$$p_j(n) = \frac{p_j/(n-j+1)}{\Sigma_{\nu=q}^r p_\nu/(n-\nu+1)}, \quad j = 1,...,r. \quad (6.10)$$

Replace the probability masses p_i in Eqs. (3.17)–(3.19) by $p_i(n)$ and call the arising quantities θ_n, ξ_n, and $g_{i,n}$, respectively. Furthermore put

$$R_n(x) = \sum_{k=1}^{r-2} a_{k,n} g_{k,n}(x) \quad (6.11)$$

and define the modified generating function H_n [compare with (6.5)] by

$$H_n(s)s^{\theta_n - 1} = (1-s)^2 \exp\{2\xi_n(s)\} \sum_{k=r-1}^{\infty} a_{k,n}s^k. \quad (6.12)$$

Rewriting Eq. (3.3) in the form

$$a_{k,n} \sum_{j=q}^{r} p_j(n)(k-j+1) = 2\sum_{\nu=q}^{r} p_\nu(n) \sum_{j=1}^{k-\nu} a_{j,n}, \quad k = r-1,..., \quad (6.13)$$

and following the line of thought adopted at the beginning of this section, one finds from (6.12) and (6.13) that

$$H_n(s) = (r-1)\delta_{q,r} + 2\int_0^s (1-x)x^{-\theta_n}R_n(x) \times \exp\{2\xi_n(x)\}dx. \quad (6.14)$$

Thus,[23] by (6.12), for some sufficiently small $\rho > 0$ and as $k \to \infty$

$$a_{k,n} = [\theta_n + k]H_n(1)\exp\{-2\xi_n(1)\} + O(k^{-k\rho}), \quad n \gg r, \quad (6.15)$$

and, consequently,

$$a(n) \equiv \lim_{k\to\infty} a_{k,n}/k = H_n(1)\exp\{-2\xi_n(1)\}, \quad n \gg r. \quad (6.16)$$

Clearly, the constant implied in the O term of Eq. (6.15) *a priori* is not independent of n. However, revising the equations defining $p_j(n)$ and θ_n [see Eqs. (3.17) and (6.10)] the coefficients of the polynomials ξ_n and R_n are seen to be uniformly bounded; since furthermore $s^{\theta_n - 1}H_n(s)$, $n = r, r+1,...$, and $s^{\theta-1}H(s) = \lim_{n\to\infty} s^{\theta_n - 1}H_n(s)$ are all integral functions of the same *finite* order we conclude that (6.15) is valid as it stands and hence that $a_{k,n}/k \to a(n)$ uniformly in n. (Observe that this follows from $a_{k,n} = [\theta_n + k]a(n) + O(1)$, $k \to \infty$, only.) Since the iterated limit

$$\lim_{n\to\infty}(\lim_{k\to\infty} a_{k,n}/k) = \lim_{n\to\infty} a(n) = H(1)\exp\{-2\xi(1)\} = a, \quad (6.17)$$

the double limit of $(a_{k,n}/k)_{k,n}$ also exists and is equal to a. Particularly, $\lim_{n\to\infty} a_{n,n}/n = a$, proving (3.11). Equations (3.12) and (3.13) follow by similar arguments.

To investigate the error term μ_n in the representation $a_{n,n} = an + \mu_n$, one would have to examine, as suggested by (6.15), the speed of convergence of $H_n(1)\exp\{-2\xi_n(1)\}$. Since, as $n \to \infty$, $\theta_n = O(1/n)$ only, it is not clear at all if the asymptotic behavior of μ_n is similar to that of the corresponding error term in (3.25). In Table II we give $a_n \equiv {}_{\rm I}a_n = a_{n,n}$ (as well as the corresponding values ${}_{\rm II}a_n$ for model II) for some values of n in the two-type case $q = 2$, $r = 3$, $p_2 = p_3 = \frac{1}{2}$. Differences in the behaviors of ${}_{\rm I}a_n$ and ${}_{\rm II}a_n$ are anything but significant. Observe that $n = 1000$ is small, however.

TABLE II. The average uncovered fractions in the saturation limit, ${}_{\rm I}a_n/n$ and ${}_{\rm II}a_n/n$, of a $1\times n$ array filled sequentially and randomly (due to model I and model II, respectively) with equally frequent dimers and trimers (i.e., $q = 2$, $r = 3$, $p_2 = p_3 = 0.5$), for various values of n.

n	${}_{\rm I}a_n/n$	${}_{\rm II}a_n/n$
5	0.1467	0.1417
10	0.1335	0.1265
20	0.1210	0.1138
50	0.1133	0.1062
100	0.1108	0.1037
200	0.1095	0.1024
500	0.1087	0.1017
1000	0.1085	0.1014
∞	0.1082	0.1012

VII. LIMITING DISTRIBUTION OF A PARTICLE CHOSEN AT RANDOM

To prove (3.39) we shall modify an idea of Bánkövi used in his paper[24] on the gap distribution in Rényi's "parking problem."[25]

Fix $i \in \{q,...,r\}$ and let $S_n = \Sigma_{j=q}^r B_n^j$, $s_n = \Sigma_{j=q}^r b_n^j$, and $Q_n^i = B_n^i/S_n$. Equation (3.39) then reads [recall Eqs. (3.9) and (3.38)]

$$\lim_{n\to\infty}\langle Q_n^i\rangle = \lim_{n\to\infty} b_n^i/s_n = p_i^*, \quad i = q,...,r, \quad (7.1)$$

and since $0 < Q_n^i < 1$, $n \gg q$, it suffices to show that $(Q_n^i)_n$ converges to p_i^* in probability (stochastically), i.e., for any $\epsilon > 0$,

$$\lim_{n\to\infty} P(|Q_n^i - p_i^*| > \epsilon) = 0, \quad i = q,...,r. \quad (7.2)$$

To this end let D_n^i and D_n denote the dispersions of B_n^i and S_n, respectively, and note that, due to Chebyshev's inequality, for any positive λ_1 and λ_2,

$$P(|B_n^i - b_n^i| < \lambda_1 D_n^i \text{ and } |S_n - s_n| < \lambda_2 D_n) > 1 - \lambda_1^{-2} - \lambda_2^{-2}. \quad (7.3)$$

By virtue of (3.32) and (3.35) we can choose n_0 such that $s_n - \lambda_2 D_n > 0$ for $n \geqslant n_0$. Rewriting Q_n^i in the form $Q_n^i = (B_n^i - b_n^i + b_n^i)/(S_n - s_n + s_n)$ we find that

$$P\left(\frac{b_n^i(1-\lambda_1 D_n^i/b_n^i)}{s_n(1+\lambda_2 D_n/s_n)} \leqslant Q_n^i \leqslant \frac{b_n^i(1+\lambda_1 D_n^i/b_n^i)}{s_n(1-\lambda_2 D_n/s_n)}\right)$$

constitutes, for any $n \geqslant n_0$, an upper bound for the probability on the lhs of Eq. (7.3). On observing once more (3.32) and (3.35) we therefore conclude that, for any $n \geqslant n_0$,

$$P(b_n^i(1+\epsilon_1)/s_n \leqslant Q_n^i \leqslant b_n^i(1+\epsilon_2)/s_n) > 1 - \lambda_1^{-2} - \lambda_2^{-2}, \tag{7.4}$$

where ϵ_1 and ϵ_2 depend on i,n,λ_1, and λ_2 and $|\epsilon_1| \to 0$, $|\epsilon_2| \to 0$ as $n \to \infty$. Hence

$$P(|Q_n^i - b_n^i/s_n| > \max(|\epsilon_1|,|\epsilon_2|)b_n^i/s_n) < \lambda_1^{-2} + \lambda_2^{-2}, \quad n \geqslant n_0 \tag{7.5}$$

and (7.2) follows since λ_1 and λ_2 are arbitrary.

Equation (3.40) may be derived quite analogously. We desist from giving the details.

VIII. DISCUSSION

We have considered a process where a one-dimensional lattice space is filled sequentially and irreversibly with particles of random length. Filling may be random both on the occasionally accessible part of the lattice (changing target area; model II) and on the entire array (constant target area; model I). In the latter case particles are adsorbed only if they do not intersect previously placed particles. This condition of no overlap, apart from leading to a "trial and error" process where checks become more and more frequent as time goes on, establishes a hierarchy of particles: The shorter a particle is, the better are its chances of adsorption—clearly, *once* the particle *is chosen* to be placed. As a "measure of benefit" may serve $d_k = p_k^* - p_k$, the difference between p_k, the "input" probability of a k-bell particle (its activity coefficient or frequency), and p_k^*, the probability of choosing such a particle randomly from among the particles making up the saturation coverage of an infinite lattice. See Figs. 5–7 to get an idea of the magnitude of d_k and its sign. Also, compare d_k to the corresponding quantity in model II where "benefit" is only a "late" phenomenon in the filling procedure. (See Sec. IV B.)

The saturation configuration of a finite as well as an infinite array has been analyzed by means of the most important mean total number of occupied sites (extent of reaction), the distribution of isolated vacancies and the above-mentioned "output" or "response" distribution $\{p_k^*, k = q,...,r\}$. Recursion relationships constitute the solution for finite lattices, while a generating function technique is used to get a picture of the terminal occupation statistics of an infinite lattice space.

The two models, coinciding (in what refers to their static properties) in the one-type case, are different in the multitype case. As emphasized earlier,[1] model II is particularly well-adapted for being applied to problems such as polymer crystallization[12] and finite cascade processes.[14] On the contrary, model I should be taken into account when treating filling problems, which are controlled by external conditions like molecule adsorption[3] and polymer reactions.[7–10,21]

ACKNOWLEDGMENTS

The author wishes to express his appreciation to Dr. E. E. Mola for some valuable discussions.

INIFTA is sponsored by the Universidad Nacional de La Plata, the Consejo Nacional de Investigaciones Científicas y Técnicas, and the Comisión de Investigaciones Científicas (Provincia de Buenos Aires).

[1]B. Mellein and E. E. Mola, J. Math. Phys. **26**, 514 (1985).
[2]E. S. Page, J. R. Stat. Soc. B **21**, 364 (1959).
[3]J. K. Mackenzie, J. Chem. Phys. **37**, 723 (1962).
[4]B. Widom, J. Chem. Phys. **58**, 4043 (1973).
[5]B. E. Blaisdell and H. Solomon, J. Appl. Probab. **7**, 667 (1970).
[6]R. B. McQuistan and D. Lichtman, J. Math. Phys. **9**, 1680 (1968).
[7]T. H. K. Barron and E. A. Boucher, Trans. Faraday Soc. **66**, 2320 (1970).
[8]E. A. Boucher, Faraday Trans. II **69**, 1839 (1973).
[9]C. W. Lewis, J. Polymer Sci. A2 **10**, 377 (1972).
[10]J. J. González, P. C. Hemmer, and J. S. Høye, J. Chem. Phys. **3**, 228 (1974).
[11]N. O. Wolf, J. W. Evans, and D. K. Hoffman, J. Math. Phys. **25**, 2519 (1984).
[12]F. Gornick and J. L. Jackson, J. Chem. Phys. **39**, 1150 (1963).
[13]M. Gordon and I. H. Hillier, J. Chem. Phys. **38**, 1376 (1963).
[14]P. Ney, Ph.D. thesis, Columbia University, 1960.
[15]L. Finegold and J. T. Donnell, Nature **278**, 443 (1979).
[16]J. Feder, J. Theor. Biol. **87**, 237 (1980).
[17]Y. Pomeau, J. Phys. A: Math. Gen. **13**, L193 (1980).
[18]R. H. Swendsen, Phys. Rev. A **24**, 504 (1981).
[19]B. E. Blaisdell and H. Solomon, J. Appl. Probab. **19**, 382 (1982).
[20]B. Mellein, J. Math. Phys. **26**, 1769 (1985).
[21]B. Mellein and J. L. Vicente, J. Appl. Probab. **23** (1986) (in press).
[22]B. Mellein, A. Maltz, and E. E. Mola, in *Stochastic Processes Applied to Physics and Other Related Fields*, edited by B. Gómez, S. Moore, A. M. Rodríguez-Vargas, and A. Rueda (World Scientific, Singapore, 1983), Vol. 1, p. 199.
[23]E. C. Titchmarsh, *The Theory of Functions* (Oxford U. P., Oxford, 1939).
[24]G. Bánkövi, Publ. Math. Inst. Hung. Acad. Sci. **7**, 395 (1962).
[25]A. Rényi, Sel. Translat. Math. Stat. Probab. **4**, 205 (1963).

Perturbation theory for polyacetylene-type kink dynamics model with acoustic phonon effects

Y. Leblanc, H. Matsumoto, and H. Umezawa
Department of Physics, The University of Alberta, Edmonton, Alberta T6G 2J1, Canada

(Received 24 April 1984; accepted for publication 19 June 1985)

Using a quasirealistic generalized continuum Takayama–Liu–Maki (TLM) model including acoustic phonon effects, perturbation theory is applied and a formal method for computing the kink of the TLM model with its perturbative corrections due to acoustic phonon interaction effects is presented. It is found that acoustic effects induce the motion of the soliton. The kink is then constrained to propagate along the molecular chain with a fixed velocity of magnitude close to that of the acoustic phonon velocity. No static solution is consistent with perturbation theory.

I. INTRODUCTION

The basic aim of this paper is to present a formal computational method dealing with topological soliton solutions of apparently complicated scalar–spinor field systems by making use of perturbation theory. As a practical presentation of the method, we introduce a physically relevant quasirealistic generalization of the Takayama–Liu–Maki (TLM) model[1] for the polyacetylene molecule taking into account acoustic phonon interaction effects. The relevance of such a model lies on the possibility that it might predict new physics about polyacetylene on the experimental level, if acoustic phonon effects would be observed. On the other hand, on a more theoretical and mathematical basis, our generalized TLM model falls into the class of soliton systems with nontrivial dynamics treated perturbatively and studied recently by various authors.[2–6] A candidate for such systems is, for example, the modified sine–Gordon equation,[5] which is used to model the Josephson junction. In this case, however, the perturbational interaction dissipates energy, while in our case the acoustic phonon interaction does not violate the energy conservation law.

A completely realistic generalized continuum TLM model can be formally derived from the discrete SSH Lagrangian model,[7] introduced few years ago by Su, Schrieffer, and Heeger, by keeping terms up to the square of the lattice spacing thereby including acoustic effects. Under a suitable limiting procedure, one recovers the TLM theory. This realistic continuum model, however, contains also certain derivative coupling terms in the optical phonon–electron interaction, which slightly modify the homogeneous theory. One, therefore, expects a deviation from the TLM gap equation, which resembles the superconduction BCS-type equation. Under the assumption that this deviation from the TLM gap equation is small, we can modify in a somewhat *ad hoc* manner the optical phonon–electron derivative couplings in order to simplify the perturbational analysis. Such a modified theory will be called quasirealistic. The explicit derivation from the discrete model is left to Appendix A.

Calculating the acoustic effects on the inhomogeneous sector of the quasirealistic model, our result is going to show that although these effects modify slightly the soliton kink profile of the TLM solution, the single kink becomes dependent of time. To understand this surprising situation it may be helpful to recall that the polyacetylene Lagrangian is not Lorentz invariant and that the continuum model under consideration is obtained from the SSH model for the reference system in which the lattice points are at rest. We will find that acoustic effects not only change the shape of the kink but also induce a uniform motion of the soliton in the lattice reference system. The velocity of this motion is controlled by the acoustic phonon velocity v. Since it is expected that the presence of the soliton deforms the displacement field (i.e., the acoustic phonon field) inducing deformations of lattices around it, we are not surprised by the creation of a second-order parameter, that is a soliton of an acoustic phonon field. This companion excitation to the optical phonon order parameter is moving with the same velocity as the one for the optical phonon soliton.

Our computation is organized as follows. First, we apply perturbation theory to the mean-field equations of our quasirealistic model, where the solution is assumed to be expressed in terms of the square of the acoustic phonon velocity (v^2) since the additional coupling in the interaction is proportional to v^2. At each order we obtain a set of differential equations, which we solve by means of the formal method of asymptotic expansion, where the solution is displayed as a power series of the asymptotic form for the unperturbed part of the soliton (the TLM kink). Since the asymptotic form of the TLM kink is an exponentially damping one indicating how fast the solution approaches the homogeneous theory at spatial infinity, it is identical in shape to the so-called boson function and the asymptotic expansion scheme then becomes very similar to the boson transformation method,[8] where the condensation of bosons in the vacuum gives rise to the whole soliton structure. In this computation, however, we restrict ourselves to first-order acoustic perturbational effects, that is, to second order (v^2) in acoustic phonon velocity.

Whenever we consider perturbative modifications to soliton systems, we meet the serious question of what is the best choice for the unperturbed state. In the present case we easily find that a static kink solution does not exist. To look for a time-dependent solution, the natural choice for the unperturbed part is the boosted TLM kink because the TLM model is Lorentz invariant. We then are led to an equation that determines the soliton velocity $v_{\rm sol}$. Self-consistency of the perturbation expansion therefore requires that the soliton be constrained to propagate at a fixed velocity, which is propor-

 0022-2488/85/112940-16$02.50

tional to v, along the molecular chain. The perturbation to the soliton shape is computed up to the third power of the boson function. An algorithm is then developed that provides a quick way of calculating coefficients up to any desired order. The result can be formally summed up and we can bring the complete kink solution into the following elegant closed form (up to v^2 order):

$$\phi(X) = [1 - v^2 Q\,\mathrm{sech}^2(X/2)]\tanh(X/2), \tag{1.1}$$

where Q is a constant and X is given as

$$X = (2\Delta_0/v_F)(\sqrt{1 - v_{sol}^2/v^2})^{-1}[x_1 \pm v_{sol}x_0 - \bar{x}], \tag{1.2}$$

with

$$\sigma^2 \equiv \frac{v_{sol}^2}{v^2} = \begin{cases} (\lambda^2 + 2\lambda)^{1/2} - \lambda, & \lambda \geqslant 0, \\ \pm(\lambda^2 + 2\lambda)^{1/2} - \lambda, & \lambda \leqslant -2. \end{cases} \tag{1.3}$$

Here λ is a dimensionless constant, which specifies the strength of the v^2 correction in the optical-phonon–electron coupling (see Appendix A). Note that $2\Delta_0$ is the Peierls energy gap.

The acoustic phonon soliton is obtained as (at lower order)

$$\langle \xi \rangle(X) = R\tanh(X/2) + \text{arbitrary constant}, \tag{1.4}$$

where R is a constant.

The limit $v \to 0$ leads back to the static TLM kink. A complete proof of the results (1.1)–(1.4) can be obtained by the following self-consistent consideration: first, assuming (1.1)–(1.4), calculate the wave function of electrons, and then recalculate the phonon order parameters by means of the electron wave functions thus obtained. When the calculated order parameters become (1.1) and (1.4), the proof is completed. A brief sketch of such a proof is given in Appendix B. A more detailed account of this proof together with the major physical results were already published in Ref. 9. However, the derivation of the result (1.1)–(1.4) is not at all an easy task; it requires quite a tedious series of steps. The main purpose of this paper is to present the method used in the derivation of (1.1)–(1.4).

In physical situations, the above analytical solution may help us in studying the behavior of the propagating soliton. The free parameter λ remains, however, unknown in the present context. However, Eq. (1.3) tells us that a vanishing value for λ brings us back to the static case. We must mention that a preliminary study of the completely realistic model described in Appendix A has shown that a static soliton is not a solution of the field equations. Therefore the case $\lambda = 0$ in the context of our quasirealistic model does not correspond to a physical situation.

Our results seem in accordance (at least qualitatively) with numerical computations applied to the discrete SSH model; Su and Schrieffer[10] showed that the velocity of the single kink was of the same order of magnitude as the acoustic phonon velocity. Furthermore Bishop *et al.*[11] recently found an upper limit of about $2.7v$, which corresponds to $\lambda \simeq -4.22$ in our analysis.

Our preliminary study of the completely realistic model has shown that the correct time dependence of the soliton may be much more complicated than a constant motion. Oscillation may also be present as suggested by recent numerical results in Ref. 11. The region $-2 < \lambda < 0$ may well correspond to such a kinetic zone forbidding uniform translational motion. It should also be pointed out that even the quasirealistic model studied might have more solutions different from the one presented in this paper, although no static solution is permitted.

As is well known, the TLM model contains other types of soliton solutions.[12] However these are not topological kink solutions. The TLM kink is the same as the one for the $\lambda\phi^4$ theory[1,12–17] The question asking how far the Yukawa-type TLM model is similar to the $\lambda\phi^4$ theory has recently been answered.[17] It has been shown that a set of solutions for the static optical phonon order parameter $\Delta(x)$ can be classified by the equation

$$[\Delta_0^2 - \Delta^2(x)]\Delta(x) + \frac{1}{2}v_F^2\hbar^2\left[\frac{\partial^2\Delta}{\partial x^2} - n\frac{1}{\Delta(x)}\left(\frac{\partial\Delta}{\partial x}\right)^2\right] = 0, \tag{1.5}$$

with number n. The solution with $n = 0$ satisfies the equation for the $\lambda\phi^4$ model because (1.5) with $n = 0$ coincides with the equation for static $\lambda\phi^4$ theory. A solution with $n = \frac{3}{2}$ yields the polaron state. Mancini presented also an integral form for solutions of (1.5) with arbitrary n. However, the topological kink solution is of more physical importance since its topological properties render the dimerized state of polyacetylene stable. Therefore, in this paper we study only the TLM kink solution modified by the acoustic phonon effect.

II. THE (1 + 1)-DIMENSIONAL QUASIREALISTIC MODEL

The bare Lagrangian density describing our quasirealistic generalized continuum model for transpolyacetylene is written as

$$\mathscr{L} = \mathscr{L}_0 + \mathscr{L}_I, \tag{2.1}$$

where

$$\mathscr{L}_0 = \psi^\dagger\left[i\frac{\partial}{\partial t} - \mu + iv_F\tau_3\frac{\partial}{\partial x}\right]\psi + \frac{1}{2}[\dot{\xi}^2 - v^2\xi'^2] + \frac{1}{2}[\dot{\Phi}^2 + v^2\Phi'^2 - m^2\Phi^2] \tag{2.2}$$

and

$$\mathscr{L}_I = g\psi^\dagger\tau_1\psi\Phi + g(v^2/m^2)[-i(\psi^\dagger\tau_3\psi' - \psi^{\dagger\prime}\tau_3\psi)\xi' + \lambda(\psi^\dagger\tau_1\psi)'\Phi']. \tag{2.3}$$

As usual, ψ describes the two-component electron field while Φ and ξ are the optical and acoustic phonon fields, respectively. The parameters v_F, v, μ, and m are the Fermi velocity, the acoustic phonon velocity, the quasielectron mass, and the optical phonon mass, respectively, while g denotes the phonon–electron coupling constant. The nondimensional parameter λ is a constant which specifies the strength of v^2 correction in the optical-phonon–electron coupling. (See Appendix A.)

This model has been obtained from the natural continuum limit of the discrete SSH model where terms up to the square of the lattice spacing have been kept. This is reflected by the appearance of terms of order v^2 in the continuum model.

The bare field equations obtained from the above Lagrangian are

$$\left[-\frac{\partial^2}{\partial t^2}-v^2\frac{\partial^2}{\partial x^2}-m^2\right]\Phi=-g\left[1-\frac{\lambda v^2}{m^2}\frac{\partial^2}{\partial x^2}\right]\psi^\dagger\tau_1\psi, \tag{2.4}$$

$$\left[-\frac{\partial^2}{\partial t^2}+v^2\frac{\partial^2}{\partial x^2}\right]\xi = -ig\frac{v^2}{m^2}\left[\psi^\dagger\tau_3\left(\frac{\partial^2\psi}{\partial x^2}\right)-\left(\frac{\partial^2\psi^\dagger}{\partial x^2}\right)\tau_3\psi\right], \tag{2.5}$$

$$\left[i\frac{\partial}{\partial t}-\mu+iv_F\tau_3\frac{\partial}{\partial x}\right]\psi = g\left[\left(-1+\frac{\lambda v^2}{m^2}\frac{\partial^2}{\partial x^2}\right)\Phi\tau_1 + \frac{iv^2}{m^2}\left(\xi''+2\xi'\frac{\partial}{\partial x}\right)\tau_3\right]\psi. \tag{2.6}$$

Now let us write the vacuum expectation value of acoustic and optical phonon fields as follows:

$$\langle 0|\xi|0\rangle\equiv\langle\xi\rangle(x,t), \tag{2.7}$$

$$\langle 0|\Phi|0\rangle\equiv(\Delta_0/g)\phi(x,t), \tag{2.8}$$

where $2\Delta_0$ is the Peierls energy gap of the quasielectron field.

Neglecting the boson-excitation modes as well as their quantum corrections, the set of equations (2.4)–(2.6) leads to the following mean-field equations:

$$\left[i\frac{\partial}{\partial x_0}-\mu+iv_F\tau_3\frac{\partial}{\partial x_1}\right]\psi = \left[\left(-1+\frac{\lambda v^2}{m^2}\frac{\partial^2}{\partial x_1^2}\right)\Delta_0\phi\tau_1 + i\frac{gv^2}{m^2}\left(\langle\xi\rangle''+2\langle\xi\rangle'\frac{\partial}{\partial x_1}\right)\tau_3\right]\psi, \tag{2.9}$$

$$\left[\frac{\partial^2}{\partial x_0^2}+v^2\frac{\partial^2}{\partial x_1^2}+m^2\right]\frac{\Delta_0\phi}{g^2}=\left(1-\frac{\lambda v^2}{m^2}\frac{\partial^2}{\partial x_1^2}\right)\langle\psi^\dagger\tau_1\psi\rangle, \tag{2.10}$$

$$\left[-\frac{\partial^2}{\partial x_0^2}+v^2\frac{\partial^2}{\partial x_1^2}\right]\langle\xi\rangle = -ig\frac{v^2}{m^2}\langle 0|\left[\psi^\dagger\tau_3\left(\frac{\partial^2\psi}{\partial x_1^2}\right)-\left(\frac{\partial^2\psi^\dagger}{\partial x_1^2}\right)\tau_3\psi\right]|0\rangle, \tag{2.11}$$

where (x_0,x_1) stands for (t,x) and x denotes from now on a space-time dependence in general.

Equations (2.9)–(2.11) constitute a set of three coupled nonlinear differential equations. Equation (2.9), however, enables us to compute the quasielectron two-point function, which, in turn, under a suitable limiting procedure, enables the explicit computation of the currents on the right-hand side of (2.10) and (2.11). The limiting procedure is the so-called point splitting method.

The quasielectron two-point function is defined as

$$iG_{\alpha\beta}(x,y)\equiv\langle 0|T\psi_\alpha(x)\psi_\beta^\dagger(y)|0\rangle. \tag{2.12}$$

Now, since

$$\langle\psi^\dagger\tau_1\psi\rangle=\operatorname{tr}\tau_1\langle\psi^\dagger(x)\psi(x)\rangle \tag{2.13}$$

and

$$\langle 0|\left[\psi^\dagger\tau_3\left(\frac{\partial^2\psi}{\partial x_1^2}\right)-\left(\frac{\partial^2\psi^\dagger}{\partial x_1^2}\right)\tau_3\psi\right]|0\rangle = \operatorname{tr}\tau_3\langle 0|\left[\psi^\dagger\left(\frac{\partial^2\psi}{\partial x_1^2}\right)-\left(\frac{\partial^2\psi^\dagger}{\partial x_1^2}\right)\psi\right]|0\rangle, \tag{2.14}$$

one can rewrite the latter local currents in terms of the quasielectron Green's function as follows:

$$\langle\psi^\dagger(x)\psi(x)\rangle=\lim_{\substack{y\to x\\(y_0>x_0)}}-iG(x,y)\equiv-iG(x,x) \tag{2.15}$$

and

$$\langle 0|\left[\psi^\dagger\left(\frac{\partial^2\psi}{\partial x_1^2}\right)-\left(\frac{\partial^2\psi^\dagger}{\partial x_1^2}\right)\psi\right]|0\rangle = \lim_{\substack{y\to x\\(y_0>x_0)}}-i\left(\frac{\partial^2}{\partial x_1^2}-\frac{\partial^2}{\partial y_1^2}\right)G(x,y). \tag{2.16}$$

Using the definitions

$$S(x,y)\equiv\tau_1G(x,y), \tag{2.17}$$

$$\Box(x,y)\equiv\tau_3G(x,y), \tag{2.18}$$

and

$$\hat\phi\equiv\phi-1, \tag{2.19}$$

and making use of Eqs. (2.13)–(2.16), one rewrites the field equations (2.9)–(2.11) as

$$\left[i\frac{\partial}{\partial x_0}-\mu+iv_F\tau_3\frac{\partial}{\partial x_1}+\Delta_0\tau_1\right]\psi = -\left[\left(1-\frac{\lambda v^2}{m^2}\frac{\partial^2}{\partial x_1^2}\right)\Delta_0\hat\phi\tau_1 - i\frac{gv^2}{m^2}\left(\langle\xi\rangle''+2\langle\xi\rangle'\frac{\partial}{\partial x_1}\right)\tau_3\right]\psi, \tag{2.20}$$

$$\left[\frac{\partial^2}{\partial x_0^2}+v^2\frac{\partial^2}{\partial x_1^2}+m^2\right]\frac{\Delta_0}{g^2}\hat\phi = -i\left(1-\frac{\lambda v^2}{m^2}\frac{\partial^2}{\partial x_1^2}\right)\operatorname{tr}S(xx)-\frac{m^2}{g^2}\Delta_0, \tag{2.21}$$

$$\left[-\frac{\partial^2}{\partial x_0^2}+v^2\frac{\partial^2}{\partial x_1^2}\right]\langle\xi\rangle = -\frac{gv^2}{m^2}\lim_{y\to x}\left(\frac{\partial^2}{\partial x_1^2}-\frac{\partial^2}{\partial y_1^2}\right)\operatorname{tr}\Box(x,y). \tag{2.22}$$

From (2.20) one directly obtains an iterated form of the two-point function for the quasielectron field

$$\begin{aligned}G(12)&=G_0(12)-G_0(11')\Sigma(1')G_0(1'2)\\&\quad+G_0(11')\Sigma(1')G_0(1'2')\Sigma(2')G_0(2'2)\\&\quad-G_0(11')\Sigma(1')G_0(1'2')\Sigma(2')\\&\quad\times G_0(2'3')\Sigma(3')G_0(3'2)+-\cdots,\end{aligned} \tag{2.23}$$

where $G(12)$ stands for $G(xy)$ and primed integers stand for internal space-time coordinates being integrated over. This notation will be very useful in future computations. Here $\Sigma(x)$ is defined as

$$\Sigma(x) \equiv \Delta_0\left(1 - \frac{\lambda v^2}{m^2}\frac{\partial^2}{\partial x_1^2}\right)\hat{\phi}\tau_1 - i\frac{gv^2}{m^2}\left[\langle\xi\rangle'' + 2\langle\xi\rangle'\frac{\partial}{\partial x_1}\right]\tau_3. \tag{2.24}$$

As usual $G_0(x,y)$ satisfies

$$\left[i\frac{\partial}{\partial x_0} - \mu + iv_F\tau_3\frac{\partial}{\partial x_1} + \Delta_0\tau_1\right]G_0(x,y) = \delta^{(2)}(x-y) \tag{2.25}$$

and its Fourier form is

$$G_0(x,y) = \frac{1}{(2\pi)^2}\int d^2p\, e^{-ip(x-y)}G_0(p_0p_1), \tag{2.26}$$

where

$$G_0(p_0p_1) = \frac{(p_0-\mu) - \Delta_0\tau_1 + v_Fp_1\tau_3}{[(p_0-\mu)^2 - E_p^2] + i\epsilon}, \tag{2.27}$$

with E_p defined as

$$E_p^2 = v_F^2p_1^2 + \Delta_0^2\,. \tag{2.28}$$

Now using the gap equation

$$(m^2/g^2)\Delta_0 + i\,\mathrm{tr}\,S_0(xx) = 0, \tag{2.29}$$

and realizing that

$$\frac{\partial^2}{\partial x_1^2}S_0(xx) = 0, \tag{2.30}$$

as well as

$$\lim_{y\to x}\left(\frac{\partial^2}{\partial x_1^2} - \frac{\partial^2}{\partial y_1^2}\right)\square_0(xy) = 0, \tag{2.31}$$

one casts the field equations (2.21) and (2.22) into the following final forms:

$$\left[\frac{\partial^2}{\partial x_0^2} + v^2\frac{\partial^2}{\partial x_1^2} + m^2\right]\frac{\Delta_0}{g^2}\hat{\phi} = -i\left(1 - \frac{\lambda v^2}{m^2}\frac{\partial^2}{\partial x_1^2}\right)\mathrm{tr}\,\hat{S}(xx) \tag{2.32}$$

for the optical phonon soliton and

$$\left[-\frac{\partial^2}{\partial x_0^2} + v^2\frac{\partial^2}{\partial x_1^2}\right]\langle\xi\rangle = -g\frac{v^2}{m^2}\lim_{y\to x}\left(\frac{\partial^2}{\partial x_1^2} - \frac{\partial^2}{\partial y_1^2}\right)\mathrm{tr}\,\hat{\square}(xy) \tag{2.33}$$

for the acoustic phonon order parameter. The two-point functions with a hat in (2.32) and (2.33) are obtained from

$$\hat{G}(x,y) \equiv G(xy) - G_0(xy). \tag{2.34}$$

The set of equations (2.32) and (2.33) together with the expansion (2.23) for the two-point functions can now be solved perturbatively in terms of powers of the square of the acoustic phonon velocity v^2.

III. A PERTURBATIONAL ANALYSIS

We now proceed to solve the combined equations (2.32) and (2.33) using perturbation theory. But before doing so let us rewrite Eq. (2.32) as follows:

$$m^2\Delta_0\hat{\phi} + ig^2\,\mathrm{tr}\,\hat{S}(xx) = v^2\left[\frac{ig^2}{m^2}\frac{\lambda\partial^2}{\partial x_1^2}\mathrm{tr}\,\hat{S}(xx) - \frac{\Delta_0\partial^2}{\partial x_1^2}\hat{\phi}\right] - \frac{\partial^2}{\partial x_0^2}\Delta_0\hat{\phi}. \tag{3.1}$$

In analogy with the study of the modified sine–Gordon equation presented by Salerno *et al.*,[5] we should carefully choose the unperturbed solution. The celebrated hyperbolic tangent profile corresponds to the static solution of (3.1) with $v = 0$, which makes the left-hand side (lhs) of (3.1) vanishing. Furthermore, the time-dependent solution of the vanishing lhs of (3.1) is obtained from the static one by a Lorentz boost since the static equation with $v = 0$ can be considered as a static situation of a Lorentz invariant system with the quasielectron satisfying a Dirac-type equation. This suggests that a reasonable choice for the unperturbed state is the boosted kink and the perturbation is given by the rhs of (3.1). This then implies that the "anomalous" term $\Delta_0(\partial^2/\partial x_0^2)\hat{\phi}$ in the rhs of (3.1) is of order $O(v^2)$, suggesting that the velocity of the soliton is of the order of the acoustic phonon velocity v. As a matter of fact we tried a perturbational analysis made under the constraint that the soliton be static for finite λ and found that this led us to an inconsistency, suggesting that the soliton be time dependent. The origin of this inconsistency will be pointed out at a later stage in our calculation.

Let us now assume the following expansions in which (x_0,x_1) appears in the configuration (1.2), where $v_{\rm sol}$ depends on v,

$$\hat{\phi} = \hat{\phi}_0 + v^2\hat{\phi}_1 + \cdots, \tag{3.2}$$

$$\langle\xi\rangle = \xi_0 + v^2\xi_1 + \cdots, \tag{3.3}$$

$$\Sigma = \Sigma_0 + v^2\Sigma_1 + \cdots. \tag{3.4}$$

The expansions (3.2)–(3.4) enable us to define

$$\hat{G}(x,y) \equiv \hat{G}^{(0)}(x,y) + v^2G^{(1)}(x,y) + \cdots. \tag{3.5}$$

The velocity $v_{\rm sol}$ will be determined self-consistently. The following calculation will show that the choice (1.2) for X works for this model. In other cases, one may need more complex X (see Refs. 2–6). Inserting the latter expansions into (2.24) gives directly

$$\Sigma_0 = \Delta_0\tau_1\hat{\phi}_0 \tag{3.6}$$

and

$$\Sigma_1 = \Delta_0\tau_1\hat{\phi}_1 - \frac{\lambda\Delta_0\tau_1}{m^2}\hat{\phi}_0'' - \frac{ig\tau_3}{m^2}\left[\xi_0'' + 2\xi_0'\frac{\partial}{\partial x_1}\right]. \tag{3.7}$$

Plugging the latter results into (2.23) and making use of (3.5) gives, in turn,

$$\begin{aligned}\hat{G}^{(0)}(12) = &-G_0(11')\Sigma_0(1')G_0(1'2) + G_0(11')\Sigma_0(1')G_0(1'2')\Sigma_0(2')G_0(2'2)\\ &- G_0(11')\Sigma_0(1')G_0(1'2')\Sigma_0(2')G_0(2'3')\Sigma_0(3')G_0(3'2) + -\cdots\end{aligned} \tag{3.8}$$

and

$$
\begin{aligned}
G^{(1)}(12) = & - G_0(11')\Sigma_1(1')G_0(1'2) + G_0(11')\Sigma_0(1')G_0(1'2')\Sigma_1(2')G_0(2'2) \\
& + G_0(11')\Sigma_1(1')G_0(1'2')\Sigma_0(2')G_0(2'2) \\
& - G_0(11')\Sigma_0(1')G_0(1'2')\Sigma_0(2')G_0(2'3')\Sigma_1(3')G_0(3'2) \\
& - G_0(11')\Sigma_0(1')G_0(1'2')\Sigma_1(2')G_0(2'3')\Sigma_0(3')G_0(3'2) \\
& - G_0(11')\Sigma_1(1')G_0(1'2')\Sigma_0(2')G_0(2'3')\Sigma_0(3')G_0(3'2) \\
& + - \cdots.
\end{aligned}
\tag{3.9}
$$

One notes that $G^{(1)}(xy)$ is linear in Σ_1. Inserting Eqs. (3.2), (3.3), and (3.5) into the field equations (2.32) and (2.33) gives finally

$$\Delta_0\hat{\phi}_0(x) = -(ig^2/m^2)\mathrm{tr}\,\hat{S}^{(0)}(xx), \tag{3.10}$$

$$\Delta_0\hat{\phi}_1(x) = -(ig^2/m^2)\mathrm{tr}\,S^{(1)}(xx) - \frac{\Delta_0}{m^2}\left[\frac{\partial^2}{\partial(vx_0)^2} + (1+\lambda)\frac{\partial^2}{\partial x_1^2}\right]\hat{\phi}_0(x), \tag{3.11}$$

for the optical soliton and

$$\xi_0(x) = -\frac{g}{m^2}\left[\frac{\partial^2}{\partial x_1^2} - \frac{\partial^2}{\partial(vx_0)^2}\right]^{-1} \times \lim_{y\to x}\left(\frac{\partial^2}{\partial x_1^2} - \frac{\partial^2}{\partial y_1^2}\right)\mathrm{tr}\,\hat{\square}^{(0)}(xy) \tag{3.12}$$

for the acoustic phonon order parameter. Note that ξ_1 gives a contribution or order v^4 to the soliton solution and is therefore disregarded as well as higher-order terms in the expansions (3.2)–(3.5) since one restricts oneself to second-order effects. The "anomalous" term in (3.1) is responsible for the appearance of time derivative v-dependent terms in the expressions (3.11) and (3.12) for $\hat{\phi}_1$ and ξ_0. However, when the condition of the soliton velocity being of order $O(v)$ will be implemented, these terms will become independent of v.

Equation (3.10) together with (3.6) and (3.7) give the unperturbed part of the solution, which leads to the boosted kink profile

$$\hat{\phi}_0 = \tanh\frac{M}{2}\left(\frac{x_1 \pm v_F u x_0 - \bar{x}}{\sqrt{1-u^2}}\right) - 1, \tag{3.13}$$

where uv_F is the soliton velocity ($v_{\rm sol}$), $\bar{x}$ its center at initial time, and M is defined as

$$M \equiv 2\Delta_0/v_F\,. \tag{3.14}$$

Equation (3.12) together with (3.6) and (3.8) tells us that the knowledge of the unperturbed part $\hat{\phi}_0$ determines ξ_0 completely. Once ξ_0 has been obtained, Eqs. (3.6), (3.7), and (3.9) together with (3.11) as well as (3.13) determines $\hat{\phi}_1$ uniquely. The complete soliton solution is then known up to v^2.

In order to obtain the explicit forms for ξ_0 and $\hat{\phi}_1$, we use the asymptotic expansion scheme.

Defining

$$X \equiv (M/\sqrt{1-u^2})(x_1 \pm uv_F x_0 - \bar{x}), \tag{3.15}$$

Eq. (3.13) can then be expanded as

$$\hat{\phi}_0 = \sum_{n=1}^{\infty} b_n f^n(X) \quad (X>0), \tag{3.16}$$

where

$$f(X) \equiv e^{-X} \tag{3.17}$$

and

$$b_n \equiv 2(-1)^n = (-1)^{n+1} b_1. \tag{3.18}$$

The $f(X)$ in the expansion (3.16) is the so-called boson function.[2] This expansion scheme is called asymptotic since the function $f(X)$ is just the asymptotic form for $\hat{\phi}_0$ and therefore indicates how fast the TLM soliton approaches the homogeneous theory at spatial infinity. Expressions for ξ_0 and $\hat{\phi}_1$ will then be obtained as power series of this function. Since (3.16) is defined only in the region of positive X, the complete solution for the soliton for the entire range of X will be determined from topological considerations at the end of the computation. One then expands ξ_0 and $\hat{\phi}_1$ as

$$\xi_0 = \sum_{n=0}^{\infty} c_n f^n(X) \tag{3.19}$$

and

$$\hat{\phi}_1 = \sum_{n=0}^{\infty} a_n f^n(X). \tag{3.20}$$

The coefficients c_n and a_n are to be determined self-consistently from (3.6), (3.8), (3.12), and (3.16) and (3.6), (3.7), (3.9), (3.11), and (3.16), respectively. Now in order to eliminate the v dependence of Eqs. (3.11) and (3.12) and therefore render the perturbation method meaningful, one writes

$$v_F u = v\sigma, \tag{3.21}$$

where σ is a pure number also to be self-consistently determined through the calculation afterward. This redefines the "generalized coordinate" X as

$$X = (M/\sqrt{1-\sigma^2 v^2/v_F^2})(x_1 \pm \sigma v x_0 - \bar{x}). \tag{3.22}$$

A convenient form for the boson function is the following rotated expression into the complex plane:

$$f(x) = Ke^{-ikx} \equiv Ke^{ik_1x_1 - ik_0x_0}, \tag{3.23}$$

where

$$k_1 \equiv (\beta/\sqrt{1-\sigma^2v^2/v_F^2}) \simeq \beta + O(v^2), \tag{3.24}$$

$$k_0 \equiv \pm\sigma v k_1 \simeq \pm\sigma v\beta + O(v^3), \tag{3.25}$$

with

$$\beta \equiv iM \tag{3.26}$$

and

$$K \equiv \exp(M\bar{x}/\sqrt{1-\sigma^2v^2/v_F^2}). \tag{3.27}$$

The function $f(X)$ is rotated back to the real axis at the end of the computation. This method is just a trick enabling us to treat the boson function as an external leg in the computation of the forecoming Feynman diagrams.

IV. COMPUTATION OF ξ_0

As a first step toward the establishment of the soliton solution, one must determine the acoustic phonon order parameter ξ_0, which we calculate using the asymptotic expansion method. From (2.18), (3.6), and (3.8) one obtains $\hat{\Box}^{(0)}(xy)$ as

$$\begin{aligned}\hat{\Box}^{(0)}(12) = & -\Box_0(11')\tau_1\tau_3\Box_0(1'2)[\Delta_0\hat{\phi}_0(1')] \\ & + \Box_0(11')\tau_1\tau_3\Box_0(1'2')\tau_1\tau_3\Box_0(2'2)[\Delta_0^2\hat{\phi}_0(1')\hat{\phi}_0(2')] \\ & - \Box_0(11')\tau_1\tau_3\Box_0(1'2')\tau_1\tau_3\Box_0(2'3')\tau_1\tau_3\Box_0(3'2)[\Delta_0^3\hat{\phi}_0(1')\hat{\phi}_0(2')\hat{\phi}_0(3')] \\ & + - \cdots. \end{aligned} \tag{4.1}$$

Making use of the expansion (3.16) for $\hat{\phi}_0$ gives, up to the third power of the boson function,

$$\begin{aligned}\hat{\Box}^{(0)}(12) = & -\Box_0(11')\tau_1\tau_3\Box_0(1'2)\{\Delta_0 b_1 f(1') + \Delta_0 b_2 f^2(1') + \Delta_0 b_3 f^3(1') + \cdots\} \\ & + \Box_0(11')\tau_1\tau_3\Box_0(1'2')\tau_1\tau_3\Box_0(2'2)\{\Delta_0^2 b_1^2 f(1')f(2') + \Delta_0^2 b_1 b_2[f(1')f^2(2') + f^2(1')f(2')] + \cdots\} \\ & - \Box_0(11')\tau_1\tau_3\Box_0(1'2')\tau_1\tau_3\Box_0(2'3')\tau_1\tau_3\Box_0(3'2)\{\Delta_0^3 b_1^3 f(1')f(2')f(3') + \cdots\} \\ & + - \cdots. \end{aligned} \tag{4.2}$$

Evaluation of the rhs of (3.12) when using (4.2) gives a local power series of the boson function. By virtue of (3.19) the coefficients c_n are then determined and so is ξ_0. In the following we compute the coefficients up to second order. To that purpose let us define the following quantities:

$$\begin{aligned} & I_1(\partial)f^n(x) \\ & \equiv \left[\frac{\partial^2}{\partial x_1^2} - \frac{\partial^2}{\partial(vx_0)^2}\right]^{-1} \lim_{y\to x}\left(\frac{\partial^2}{\partial x_1^2} - \frac{\partial^2}{\partial y_1^2}\right) \\ & \quad \times \mathrm{tr}\int d^2z\, \Box_0(xz)\tau_1\tau_3\Box_0(zy)f^n(z), \end{aligned} \tag{4.3}$$

$$\begin{aligned} & I_2(\partial)f^2(x) \\ & \equiv \left[\frac{\partial^2}{\partial x_1^2} - \frac{\partial^2}{\partial(vx_0)^2}\right]^{-1} \lim_{y\to x}\left(\frac{\partial^2}{\partial x_1^2} - \frac{\partial^2}{\partial y_1^2}\right) \\ & \quad \times \mathrm{tr}\int d^2z\, d^2w\, \Box_0(xz)\tau_1\tau_3\Box_0(zw)\tau_1\tau_3\Box_0(wy)f(z)f(w), \end{aligned} \tag{4.4}$$

where

$$I_i(\partial)e^{-inkx} \equiv I_i(nk)e^{-inkx}. \tag{4.5}$$

From (2.26) and (3.23), it follows that

$$\begin{aligned} & I_1(\partial)f^n(x) \\ & = K^n\left[\frac{\partial^2}{\partial x_1^2} - \frac{\partial^2}{\partial(vx_0)^2}\right]^{-1} \lim_{y\to x}\left(\frac{\partial^2}{\partial x_1^2} - \frac{\partial^2}{\partial y_1^2}\right)\frac{1}{(2\pi)^4} \\ & \quad \times \int d^2z \int d^2p\, d^2q\, e^{-ip(x-z) - iq(z-y) - inkz} \\ & \quad \times \mathrm{tr}\,\Box_0(p_0 p_1)\tau_1\tau_3\Box_0(q_0 q_1) \end{aligned} \tag{4.6}$$

and

$$\begin{aligned} & I_2(\partial)f^2(x) \\ & = K^2\left[\frac{\partial^2}{\partial x_1^2} - \frac{\partial^2}{\partial(vx_0)^2}\right]^{-1} \lim_{y\to x}\left(\frac{\partial^2}{\partial x_1^2} - \frac{\partial^2}{\partial y_1^2}\right)\frac{1}{(2\pi)^6} \\ & \quad \times \int d^2z\, d^2w \int d^2p\, d^2q\, d^2l \\ & \quad \times e^{-ip(x-z) - iq(z-w) - il(w-y)} e^{-ik(z+w)} \\ & \quad \times \mathrm{tr}\,\Box_0(p_0 p_1)\tau_1\tau_3\Box_0(q_0 q_1)\tau_1\tau_3\Box_0(l_0 l_1). \end{aligned} \tag{4.7}$$

Performing all the derivatives and limits in the displayed order and then integrating over internal coordinates yields

$$\begin{aligned} I_1(nk) = & [1-\sigma^2]^{-1}\frac{1}{nk_1}\frac{\mathrm{tr}}{(2\pi)^2}\int d^2p(2p_1 - nk_1) \\ & \times \Box_0(p_0 p_1)\tau_1\tau_3\Box_0(p_0 - nk_0, p_1 - nk_1) \end{aligned} \tag{4.8}$$

and

$$\begin{aligned} I_2(2k) = & [1-\sigma^2]^{-1}\frac{1}{k_1}\frac{\mathrm{tr}}{(2\pi)^2} \\ & \times \int d^2p(p_1 - k_1)\Box_0(p_0 p_1)\tau_1\tau_3 \\ & \times \Box_0(p_0 - k_0, p_1 - k_1)\tau_1\tau_3\Box_0(p_0 - 2k_0, p_1 - 2k_1), \end{aligned} \tag{4.9}$$

where integrations over δ functions created by internal coordinates' integrations have been performed. One must realize however that the above expressions in fact lead to expansions in powers of v. This is so because of (3.24) and (3.25). Since ξ_0 contributes to the soliton solution $\hat{\phi}$ in the second order of v, any v-dependent terms in (4.8) and (4.9) will give higher-order corrections to the soliton form. One can, therefore, drop them at this stage. Equations (4.8) and (4.9) are then evaluated in the static limit, that is,

$$k_1 \to \beta, \quad k_0 \to 0. \tag{4.10}$$

This prescription applies as well to the computation of the coefficients for $\hat{\phi}_1$. One must be careful, however, to perform any time derivative before taking this limit.

Using the explicit form (2.27) for the Green's function, taking the trace, and applying the above prescription lead to

$$I_1(n\beta) = -\frac{2v_F\Delta_0[1-\sigma^2]^{-1}}{n\beta(2\pi)^2} \times \int d^2p\frac{(2p_1 - n\beta)^2}{(p_0^2 - E_p^2)(p_0^2 - E_{p-n\beta}^2)} \tag{4.11}$$

and

$$I_2(2\beta) = \frac{2v_F[1-\sigma^2]^{-1}}{\beta(2\pi)^2} \times \int d^2p\frac{(p_1-\beta)^2[p_0^2 + 3\Delta_0^2 - v_F^2 p_1(p_1 - 2\beta)]}{(p_0^2 - E_p^2)(p_0^2 - E_{p-\beta}^2)(p_0^2 - E_{p-2\beta}^2)}. \tag{4.12}$$

Carrying out the integrations and making use of the identity

$$4\Delta_0^2 + v_F^2\beta^2 \equiv 0, \tag{4.13}$$

which follows straightforwardly from (3.14) and (3.26), one obtains finally

$$I_1(n\beta) = (1/n)[S - m^2/g^2][1-\sigma^2]^{-1} \tag{4.14}$$

and

$$I_2(2\beta) = -(1/4\Delta_0)[S - m^2/g^2][1-\sigma^2]^{-1}. \tag{4.15}$$

In gaining the above expressions, use has been made of the gap equation (2.29), which can be written, after integration, in the following form:

$$\frac{m^2}{g^2} = \frac{1}{\pi v_F}\sinh^{-1}\left(\frac{v_F\Lambda}{\Delta_0}\right), \tag{4.16}$$

where Λ is a high momentum cutoff. Note that in (4.14) and (4.15) a surface term denoted by S ($S = 1/\pi v_F$) has appeared while integrating. It originates from the choice of the high momentum cutoff regularization scheme and can be shown to disappear by a suitable symmetrization procedure. However, from now on, we deliberately keep it since it will be shown to help discover the structure of the computation in later sections. From Eq. (3.12) for ξ_0 as well as (4.2) and the definitions (4.3) and (4.4), one obtains the following coefficients:

$$c_0 = \text{(arbitrary)}, \tag{4.17}$$

$$c_1 = -(g/m^2)[-\Delta_0 b_1 I_1(\beta)], \tag{4.18}$$

$$c_2 = -(g/m^2)[-\Delta_0 b_2 I_1(2\beta) + \Delta_0^2 b_1^2 I_2(2\beta)]. \tag{4.19}$$

The expressions (4.14) and (4.15) together with (3.18) finally lead to

$$c_1 = (g\Delta_0/m^2)[S - m^2/g^2][1-\sigma^2]^{-1}b_1, \tag{4.20}$$

$$c_2 = (g\Delta_0/m^2)[S - m^2/g^2][1-\sigma^2]^{-1}b_2. \tag{4.21}$$

This suggests that, in general,

$$c_n = (g\Delta_0/m^2)[S - m^2/g^2][1-\sigma^2]^{-1}b_n \quad (n\geqslant 1), \tag{4.22}$$

so that

$$\xi_0(X) = (g\Delta_0/m^2)(S - m^2/g^2)(1-\sigma^2)^{-1}\phi_0(X) + \text{arbitrary constant}. \tag{4.23}$$

A complete proof for this expression requires the self-consistent calculation, which was mentioned in the Introduction, and involving fermion wave functions.

V. A DIFFERENTIAL EQUATION FOR $\hat{\phi}_1$

Equations (3.6), (3.7), and (3.9) together with (2.17) and (2.18) lead to the following expansion for $S^{(1)}(xx)$:

$$\begin{aligned}
S^{(1)}(11) = &-S_0(11')S_0(1'1)\{\Delta_0\hat{\phi}_0(1') + \nabla(1')\} - S_0(11')[\Omega(1')]\Box_0(1'1)\\
&+ S_0(11')S_0(1'2')S_0(2'1)\{\Delta_0^2[\hat{\phi}_0(1')\hat{\phi}_1(2') + \hat{\phi}_1(1')\hat{\phi}_0(2')] + \Delta_0[\hat{\phi}_0(1')\nabla(2') + \nabla(1')\hat{\phi}_0(2')]\}\\
&+ S_0(11')[\Delta_0\hat{\phi}_0(1')]S_0(1'2')[\Omega(2')]\Box_0(2'1) + S_0(11')[\Omega(1')]\Box_0(1'2')[\Delta_0\hat{\phi}_0(2')]S_0(2'1)\\
&- S_0(11')S_0(1'2')S_0(2'3')S_0(3'1)\{\Delta_0^3[\hat{\phi}_0(1')\hat{\phi}_0(2')\hat{\phi}_1(3') + \hat{\phi}_0(1')\hat{\phi}_1(2')\hat{\phi}_0(3')\\
&+ \hat{\phi}_1(1')\hat{\phi}_0(2')\hat{\phi}_0(3')] + \Delta_0^2[\hat{\phi}_0(1')\hat{\phi}_0(2')\nabla(3') + \hat{\phi}_0(1')\nabla(2')\hat{\phi}_0(3') + \nabla(1')\hat{\phi}_0(2')\hat{\phi}_0(3')]\}\\
&- S_0(11')[\Delta_0\hat{\phi}_0(1')]S_0(1'2')[\Delta_0\hat{\phi}_0(2')]S_0(2'3')[\Omega(3')]\Box_0(3'1)\\
&- S_0(11')[\Delta_0\hat{\phi}_0(1')]S_0(1'2')[\Omega(2')]\Box_0(2'3')[\Delta_0\hat{\phi}_0(3')]S_0(3'1)\\
&- S_0(11')[\Omega(1')]\Box_0(1'2')[\Delta_0\hat{\phi}_0(2')]S_0(2'3')[\Delta_0\hat{\phi}_0(3')]S_0(3'1)\\
&+ - \cdots,
\end{aligned} \tag{5.1}$$

where

$$\nabla(x) \equiv -(\lambda\Delta_0\hat{\phi}_0''/m^2)(x) \tag{5.2}$$

and

$$\Omega(x) \equiv -\frac{ig}{m^2}\left[\xi_0'' + 2\xi_0'\frac{\partial}{\partial x_1}\right]. \tag{5.3}$$

In Eq. (3.11) we then substitute (5.1). Extracting linear terms with respect to $\hat{\phi}_1$, $\hat{\phi}_0$, and ξ_0, we define the derivative operators D_0, D_1, D_2, as

$$D_0(\partial)\Delta_0\hat{\phi}_1(1) \equiv i\,\mathrm{tr}\,S_0(11')S_0(1'1)[\Delta_0\hat{\phi}_1(1')] - (m^2\Delta_0/g^2)\hat{\phi}_1(1), \tag{5.4}$$

$$D_1(\partial)\Delta_0\hat{\phi}_0(1) \equiv -i\,\mathrm{tr}\,S_0(11')S_0(1'1)[\nabla(1')] + (\Delta_0/g^2)[(1+\lambda)\hat{\phi}_0''(1) + \ddot{\hat{\phi}}_0/v^2], \tag{5.5}$$

$$D_2(\partial)\xi_0 \equiv -i\,\mathrm{tr}\,S_0(11')[\Omega(1')]\Box_0(1'1), \tag{5.6}$$

where

$$D_i(\partial)e^{-inkx} \equiv D_i(nk)e^{-inkx}. \tag{5.7}$$

Then (3.11) and (5.1) yield the following differential equation for $\hat{\phi}_1$:

$$
\begin{aligned}
D_0(\partial)\Delta_0\phi_1(1) = {} & D_1(\partial)\Delta_0\hat{\phi}_0(1) + D_2(\partial)\xi_0(1) + i\,\mathrm{tr}\{S_0(11')S_0(1'2')S_0(2'1)\{\Delta_0^2[\hat{\phi}_0(1')\hat{\phi}_1(2') + \hat{\phi}_1(1')\hat{\phi}_0(2')] \\
& + \Delta_0[\hat{\phi}_0(1')\nabla(2') + \nabla(1')\hat{\phi}_0(2')] + S_0(11')[\Delta_0\hat{\phi}_0(1')]S_0(1'2')[\Omega(2')]\Box_0(2'1) \\
& + S_0(11')[\Omega(1')]\Box_0(1'2')[\Delta_0\hat{\phi}_0(2')]S_0(2'1) - S_0(11')S_0(1'2')S_0(2'3')S_0(3'1)\{\Delta_0^3[\hat{\phi}_0(1)\hat{\phi}_0(2')\hat{\phi}_1(3') \\
& + \hat{\phi}_0(1')\hat{\phi}_1(2')\hat{\phi}_0(3') + \hat{\phi}_1(1')\hat{\phi}_0(2')\hat{\phi}_0(3')] + \Delta_0^2[\hat{\phi}_0(1')\hat{\phi}_0(2')\nabla(3') + \hat{\phi}_0(1')\nabla(2')\hat{\phi}_0(3') + \nabla(1')\hat{\phi}_0(2')\hat{\phi}_0(3')]\} \\
& - S_0(11')[\Delta_0\hat{\phi}_0(1')]S_0(1'2')[\Delta_0\hat{\phi}_0(2'2)]S_0(2'3)[\Omega(3')]\Box_0(3') \\
& - S_0(11')[\Delta_0\hat{\phi}_0(1')]S_0(1'2')[\Omega(2')]\Box_0(2'3')[\Delta_0\hat{\phi}_0(3')]S_0(3'1) \\
& - S_0(11')[\Omega(1')]\Box_0(1'2')[\Delta_0\hat{\phi}_0(2')]S_0(2'3')[\Delta_0\hat{\phi}_0(3')]S_0(3'1) \\
& + - \cdots\}.
\end{aligned} \tag{5.8}
$$

The differential operator $D_0(\partial)$ is the operator that determines the mass shell condition (3.14), giving information on the speed rate of the exponential damping of the asymptotic form of ϕ_0 and ϕ_1 in the static limit. In that limit it is obtained as

$$D_0(n\beta) = \frac{-\sqrt{4\Delta_0^2 + n^2 v_F^2 \beta^2}}{n\pi v_F^2 \beta} \sinh^{-1}\left(\frac{n v_F \beta}{2\Delta_0}\right). \tag{5.9}$$

When rotating back to the real axis, one must be careful in choosing the proper branch for (5.9). However, one shall not be concerned with this problem since $D_0(n\beta)$ will disappear self-consistently throughout the calculation. The operator $D_1(\partial)$ is now easily calculated. To that purpose let us write (5.4) and (5.5) as

$$[D_0(nk) + m^2/g^2]e^{-inkx} = i\,\mathrm{tr}\,S_0(xz)S_0(zx)e^{-inkz} \tag{5.10}$$

and

$$
\begin{aligned}
D_1(nk)e^{-inkx} &= -(n^2k_1^2/m^2)\{i\lambda\,\mathrm{tr}\,S_0(xz)S_0(zx)e^{-inkz} \\
&\quad + (m^2/g^2)[(1+\lambda)+\sigma^2]e^{-inkx}\}.
\end{aligned} \tag{5.11}
$$

In (5.11) we made use of (5.2) as well as the following relations:

$$
\begin{aligned}
f^{n'}(x) &= ink_1 f^n(x), \\
f^{n''}(x) &= -n^2k_1^2 f^n(x), \\
\ddot{f}^n(x) &= -n^2k_1^2\sigma^2v^2 f^n(x).
\end{aligned} \tag{5.12}
$$

Inserting (5.10) into (5.11) and taking the static limit lead to

$$D_1(n\beta) = 4\frac{n^2\Delta_0^2}{v_F^2m^2}\left[\frac{m^2}{g^2}((2\lambda+1)+\sigma^2) + \lambda D_0(n\beta)\right]. \tag{5.13}$$

Now one writes equation (5.6) as

$$
\begin{aligned}
D_2(nk)e^{-inkx} &= \frac{g^2}{m^2}\,\mathrm{tr}\,S_0(xz)\left[n^2k_1^2 - 2ink_1\frac{\partial}{\partial z_1}\right]\Box_0(zx)e^{-inkz},
\end{aligned} \tag{5.14}
$$

where (5.3) and (5.12) have been used. In the Fourier representation, (5.14) becomes

$$
\begin{aligned}
D_2(nk)e^{-inkx} &= \frac{g}{m^2}\frac{\mathrm{tr}}{(2\pi)^4}\int d^2z\int d^2p\,d^2q\,e^{-ip(x-z)-inkz} \\
&\quad \times S_0(p_0p_1)\Box_0(q_0q_1)\left[n^2k_1^2 - 2ink_1\frac{\partial}{\partial z_1}\right]e^{-iq(z-x)},
\end{aligned} \tag{5.15}
$$

which leads to

$$
\begin{aligned}
D_2(nk)e^{-inkx} &= \frac{nk_1g}{m^2(2\pi)^4}\int d^2z\int d^2p\,d^2q\,e^{-ip(x-z)-iq(z-x)-inkz} \\
&\quad \times [2q_1 + nk_1]\,\mathrm{tr}\,S_0(p_0p_1)\Box_0(q_0q_1).
\end{aligned} \tag{5.16}
$$

Integrating over internal coordinates and over the δ function thus created yields

$$
\begin{aligned}
D_2(nk) &= \frac{n\,k_1g}{m^2(2\pi)^2}\int d^2p(2p_1 - nk_1) \\
&\quad \times \mathrm{tr}\,S_0(p_0p_1)\Box_0(p_0 - nk_0, p_1 - nk_1).
\end{aligned} \tag{5.17}
$$

Making use of (2.27) and taking the trace in the static limit gives directly

$$D_2(n\beta) = -\frac{2n\Delta_0\beta g}{m^2(2\pi)^2}\int d^2p\frac{(2p_1 - n\beta)^2}{(p_0^2 - E_p^2)(p_0^2 - E_{p-n\beta}^2)}. \tag{5.18}$$

Comparing this result with (4.11) and (4.14) yields finally

$$D_2(n\beta) = -(4n\Delta_0^2g/v_F^2m^2)[S - m^2/g^2]. \tag{5.19}$$

Once again the static limit has been used to determine the explicit form for $D_0(n\beta)$, $D_1(n\beta)$, and $D_2(n\beta)$. Otherwise computed these quantities would have given higher-order corrections to the soliton profile. The knowledge of the three latter operators enables us in turn to obtain full information about the linear terms of the expansion (5.8). Using (3.16), (3.19), and (3.20) and comparing linear f terms in (5.8), one is led to

$$D_0(\beta)\Delta_0a_1 = D_1(\beta)\Delta_0b_1 + D_2(\beta)c_1. \tag{5.20}$$

Note that a_0 in (3.20) is trivially obtained as zero. According to (3.14) and (3.26), we have $4\Delta_0^2 + v_F^2\beta^2 = 0$ in (5.9). Thus, the contribution from a_1 in (5.20) vanishes. Equation (5.20) together with the result (4.22) gives the following condition on the soliton velocity:

$$\frac{m^2}{g^2}[(2\lambda+1)+\sigma^2] - \frac{g^2}{m^2}\left[S - \frac{m^2}{g^2}\right]^2[1-\sigma^2]^{-1} = 0, \tag{5.21}$$

where the explicit forms (5.13) and (5.19) have been used. The latter equation can be brought into the form

$$\sigma^4 + 2\sigma^2\lambda - 2\lambda = 0, \tag{5.22}$$

in which the surface term S has been set equal to zero. The reason for this has been mentioned previously.

Equation (5.22) is solved as

$$\sigma^2 = \begin{cases} (\lambda^2 + 2\lambda)^{1/2} - \lambda, & \lambda > 0, \\ \pm(\lambda^2 + 2\lambda)^{1/2} - \lambda, & \lambda < -2. \end{cases} \tag{5.23}$$

The critical condition (5.21) will be shown to reappear self-consistently in higher power computations.

Since the case $\lambda = 0$ is equivalent to neglecting altogether the optical-phonon–electron derivative coupling, that is, recovering the TLM coupling, it is easy to see from (5.21) how a perturbational analysis made on the constraint that the soliton be static for $\lambda \neq 0$ leads to an inconsistency, as mentioned in the beginning of Sec. III. The static case corresponds to $\sigma = 0$. Inserting this into (5.23) for nonvanishing λ leads us to a dead end.

From (5.20), the coefficient a_1 in (3.20) is left undetermined. It will be obtained from boundary conditions on the full solution at the end of the calculation. Higher-order coefficients will be obviously a_1-dependent.

VI. COMPUTATION OF THE SECOND-ORDER COEFFICIENT FOR $\hat{\phi}_1$

When the expansions (3.16), (3.19), and (3.20) for $\hat{\phi}_0$, ξ_0, and $\hat{\phi}_1$ are used, Eq. (5.8) leads to the following relation for the second-order coefficient for $\hat{\phi}_1$:

$$\begin{aligned} &D_0(2\beta)\Delta_0 a_2 f^2(1) \\ &\quad = D_1(2\beta)\Delta_0 b_2 f^2(1) + D_2(2\beta)c_2 f^2(1) + i \operatorname{tr} S_0(11')S_0(1'2')S_0(2'1)f(1')f(2')\{2\Delta_0^2\, b_1 a_1 + 2\Delta_0^2\lambda(\beta^2/m^2)b_1^2\} \\ &\qquad - i\frac{g}{m^2}\Delta_0 b_1 c_1 \Big\{ i \operatorname{tr} S_0(11')S_0(1'2')\Big[-\beta^2 + 2i\beta\frac{\partial}{\partial 2'}\Big]\Box_0(2'1)f(1')f(2') + i \operatorname{tr} S_0(11')\Big[-\beta^2 + 2i\beta\frac{\partial}{\partial 1'}\Big] \\ &\qquad \times \Box_0(1'2')S_0(2'1)f(1')f(2')\Big\}, \end{aligned} \tag{6.1}$$

where $\partial/\partial 1'$ and $\partial/\partial 2'$ mean the space derivative with respect to the internal coordinates $1'$ and $2'$, respectively. In the above formula, (5.2), (5.3), and (5.12) were used as well as the static limit taken.

Let us define the following operators:

$$A_0(\partial)f^2(1) \equiv i \operatorname{tr} S_0(11')S_0(1'2')S_0(2'1)f(1')f(2'), \tag{6.2}$$

$$A_1(\partial)f^2(1) \equiv i \operatorname{tr} S_0(11')S_0(1'2')\Big[-\beta^2 + 2i\beta\frac{\partial}{\partial 2'}\Big] \times \Box_0(2'1)f(1')f(2'), \tag{6.3}$$

and

$$A_2(\partial)f^2(1) \equiv i \operatorname{tr} S_0(11')\Big[-\beta^2 + 2i\beta\frac{\partial}{\partial 1'}\Big] \times \Box_0(1'2')S_0(2'1)f(1')f(2'), \tag{6.4}$$

where

$$A_i(\partial)f^2 \equiv A_i(2\beta)f^2. \tag{6.5}$$

One can show that $A_0(2\beta)$ is related to $D_0(2\beta)$ of Eq. (5.9) by the relation

$$A_0(2\beta) = B(2\beta)D_0(2\beta), \tag{6.6}$$

where $B(2\beta)$ is found to be

$$B(2\beta) = 1/2\Delta_0. \tag{6.7}$$

Now using again the Fourier representation for the two-point functions and performing as usual the appropriate integrations over internal coordinates and momenta enables us to solve (6.3) and (6.4) for $A_1(\partial)$ and $A_2(\partial)$ as

$$A_1(2\beta) = -i\beta\frac{\operatorname{tr}}{(2\pi)^2}\int d^2p(2p_1 - 3\beta)S_0(p_0p_1)S_0(p_0p_1 - \beta)\Box_0(p_0p_1 - 2\beta) \tag{6.8}$$

and

$$A_2(2\beta) = -i\beta\frac{\operatorname{tr}}{(2\pi)^2}\int d^2p(2p_1 - \beta)S_0(p_0p_1)\Box_0(p_0p_1 - \beta)S_0(p_0p_1 - 2\beta). \tag{6.9}$$

Taking the trace of both equations and summing them up give explicitly

$$A_1(2\beta) + A_2(2\beta) = -\frac{8iv_F}{(2\pi)^2}\int d^2p\left\{\frac{p_0^2[p_1^2 - 2\beta(p_1 - \beta)] + (p_1 - \beta)^2[3\Delta_0^2 - v_F^2 p_1(p_1 - 2\beta)]}{(p_0^2 - E_p^2)(p_0^2 - E_{p-\beta}^2)(p_0^2 - E_{p-2\beta}^2)}\right\}. \tag{6.10}$$

Carrying out the integration yields, after much algebra,

$$A_1(2\beta) + A_2(2\beta) = -(4i\Delta_0/v_F^2)[D_0(2\beta) + (S - m^2/g^2)]. \tag{6.11}$$

After insertion of all the above $A_i(2\beta)$'s, (6.1) leads to the relation, among the coefficients a_i, b_i and c_i,

$$\begin{aligned} D_0(2\beta)\Delta_0 a_2 = {} & \frac{16\Delta_0^2}{v_F^2 m^2}\Big[\frac{m^2}{g^2}((2\lambda + 1) + \sigma^2) + \lambda D_0(2\beta)\Big]\Delta_0 b_2 - 8\frac{\Delta_0^2 g}{v_F^2 m^2}\Big[S - \frac{m^2}{g^2}\Big]c_2 + \Delta_0\Big[a_1 b_1 - \frac{4\lambda\Delta_0^2}{v_F^2 m^2}b_1^2\Big]D_0(2\beta) \\ & - 4\frac{\Delta_0^2 g}{v_F^2 m^2}b_1 c_1\Big[D_0(2\beta) + \Big(S - \frac{m^2}{g^2}\Big)\Big]. \end{aligned} \tag{6.12}$$

Inserting (3.18) and (4.23), (6.12) becomes

$$D_0(2\beta)a_1 = \frac{32\Delta_0^2}{v_F^2 m^2}\left\{\frac{m^2}{g^2}((2\lambda+1)+\sigma^2) - \frac{g^2}{m^2}\left(S - \frac{m^2}{g^2}\right)^2(1-\sigma^2)^{-1}\right\}$$

$$+\left\{\frac{16\Delta_0^2}{v_F^2 m^2}\left[\lambda - \frac{g^2}{m^2}\left(S-\frac{m^2}{g^2}\right)(1-\sigma^2)^{-1}\right] - 2a_1\right\}D_0(2\beta). \quad (6.13)$$

Remembering the condition (5.21), a_2 is then determined as

$$a_2 = 8Q - 2a_1, \quad (6.14)$$

where Q is defined as

$$Q \equiv \frac{2\Delta_0^2}{v_F^2 m^2}\left[\lambda - \frac{g^2}{m^2}\left(S - \frac{m^2}{g^2}\right)(1-\sigma^2)^{-1}\right]. \quad (6.15)$$

Obviously, a third-order calculation is required to shed more light on the perturbed solution. The calculation in the next section shows that a_3 is indeed a linear combination of Q and a_1, as in the present case for a_2.

VII. COMPUTATION OF THE THIRD-ORDER COEFFICIENTS FOR $\hat{\phi}_1$

The equation of the third-order coefficient for $\hat{\phi}_1$ is obtained from (5.8) as (note that $b_1c_2 = b_2c_1$)

$$\begin{aligned}
D_0(3\beta)\Delta_0 a_3 f^3(1) &= D_1(3\beta)\Delta_0 b_3 f^3(1) + D_2(3\beta)c_3 f^3(1) + i\,\mathrm{tr}\Big\{S_0(11')S_0(1'2')S_0(2'1)[f^2(1')f(2') + f(1')f^2(2')] \\
&\times\Big\{\Delta_0^2[a_1b_2 + a_2b_1] + 5\frac{\lambda\Delta_0^2\beta^2}{m^2}b_1b_2\Big\} - i\frac{g\Delta_0 b_1}{m^2}c_2\Big\{S_0(11')S_0(1'2')\Big[-\beta^2 + 2i\beta\frac{\partial}{\partial 2'}\Big]\Box_0(2'1)f^2(1')f(2') \\
&+ S_0(11')S_0(1'2')\Big[-4\beta^2 + 4i\beta\frac{\partial}{\partial 2'}\Big]\Box_0(2'1)f(1')f^2(2') + S_0(11')\Big[-4\beta^2 + 4i\beta\frac{\partial}{\partial 1'}\Big] \\
&\times\Box_0(1'2')S_0(2'1)f^2(1')f(2') + S_0(11')\Big[-\beta^2 + 2i\beta\frac{\partial}{\partial 1'}\Big]\Box_0(1'2')S_0(2'1)f(1')f^2(2')\Big\} \\
&- S_0(11')S_0(1'2')S_0(2'3')S_0(3'1)f(1')f(2')f(3')\Big\{3\Delta_0^3\Big[b_1^2a_1 + \frac{\lambda\beta^2 b_1^3}{m^2}\Big]\Big\} \\
&+ i\frac{g}{m^2}\Delta_0^2 b_1^2 c_1 f(1')f(2')f(3')\Big\{S_0(11')S_0(1'2')S_0(2'3')\Big[-\beta^2 + 2i\beta\frac{\partial}{\partial 3'}\Big]\Box_0(3'1) \\
&+ S_0(11')S_0(1'2')\Big[-\beta^2 + 2i\beta\frac{\partial}{\partial 2'}\Big]\Box_0(2'3')S_0(3'1) + S_0(11')\Big[-\beta^2 + 2i\beta\frac{\partial}{\partial 1'}\Big]\Box_0(1'2')S_0(2'3')S_0(3'1)\Big\}\Big\},
\end{aligned} \quad (7.1)$$

where (5.2), (5.3), and (5.12) have been used again.

Now let us introduce the following set of operators:

$$\begin{aligned}
J_1(\partial)f^3(1) &\equiv i\,\mathrm{tr}\,S_0(11')S_0(1'2')S_0(2'1)[f^2(1')f(2') + f(1')f^2(2')] \\
&= 2\,i\,\mathrm{tr}\,S_0(11')S_0(1'2')S_0(2'1)f(1')f^2(2'),
\end{aligned} \quad (7.2)$$

$$J_2(\partial)f^3(1) \equiv i\,\mathrm{tr}\,S_0(11')S_0(1'2')S_0(2'3')S_0(3'1)f(1')f(2')f(3'), \quad (7.3)$$

$$\begin{aligned}
L_0(\partial)f^3(1) &\equiv i\,\mathrm{tr}\,\Big\{S_0(11')S_0(1'2')S_0(2'3')\Big[-\beta^2 + 2i\beta\frac{\partial}{\partial 3'}\Big]\Box_0(3'1) + S_0(11')S_0(1'2')\Big[-\beta^2 + 2i\beta\frac{\partial}{\partial 2'}\Big]\Box_0(2'3')S_0(3'1) \\
&+ S_0(11')\Big[-\beta^2 + 2i\beta\frac{\partial}{\partial 1'}\Big]\Box_0(1'2')S_0(2'3')S_0(3'1)\Big\}f(1')f(2')f(3'),
\end{aligned} \quad (7.4)$$

$$\begin{aligned}
L_1(\partial)f^3(1) &\equiv i\,\mathrm{tr}\,\Big\{S_0(11')S_0(1'2')\Big[-\beta^2 + 2i\beta\frac{\partial}{\partial 2'}\Big]\Box_0(2'1)f^2(1')f(2') \\
&+ S_0(11')\Big[-\beta^2 + 2i\beta\frac{\partial}{\partial 1'}\Big]\Box_0(1'2')S_0(2'1)f(1')f^2(2')\Big\},
\end{aligned} \quad (7.5)$$

and

$$\begin{aligned}
L_2(\partial)f^3(1) &\equiv i\,\mathrm{tr}\,\Big\{S_0(11')S_0(1'2')\Big[-4\beta^2 + 4i\beta\frac{\partial}{\partial 2'}\Big]\Box_0(2'1)f(1')f^2(2') \\
&+ S_0(11')\Big[-4\beta^2 + 4i\beta\frac{\partial}{\partial 1'}\Big]\Box_0(1'2')S_0(2'1)f^2(1')f(2')\Big\},
\end{aligned} \quad (7.6)$$

where

$$J_i(\partial)f^3 \equiv J_i(3\beta)f^3 \tag{7.7}$$

and

$$L_i(\partial)f^3 \equiv L_i(3\beta)f^3. \tag{7.8}$$

One notes again that, similarly to $A_0(2\beta)$ in (6.6), $J_1(3\beta)$ and $J_2(3\beta)$ can be related to $D_0(3\beta)$ by the following relations:

$$J_1(3\beta) = [R(3\beta)/B(2\beta)]D_0(3\beta) \tag{7.9}$$

and

$$J_2(3\beta) = C(3\beta)D_0(3\beta), \tag{7.10}$$

where $B(2\beta)$ and $D_0(3\beta)$ have been defined by the relations (6.7) and (5.9), respectively, while $R(3\beta)$ and $C(3\beta)$, after tedious algebraic work, can be obtained as

$$R(3\beta) = \frac{2D_0^{-1}(3\beta)}{3\pi v_F^4\beta^3}\left[\frac{1}{2}\sqrt{4\Delta_0^2 + qv_F^2\beta^2}\sinh^{-1}\left(\frac{3v_F\beta}{2\Delta_0}\right) - \sqrt{\Delta_0^2 + v_F^2\beta^2}\sinh^{-1}\left(\frac{v_F\beta}{\Delta_0}\right)\right] \tag{7.11}$$

and

$$C(3\beta) = \frac{648\Delta_0^2 D_0^{-1}(3\beta)}{\pi v_F^6(3\beta)^5}\sqrt{\Delta_0^2 + \mathrm{v}_F^2\beta^2}\sinh^{-1}\left(\frac{v_F\beta}{\Delta_0}\right). \tag{7.12}$$

However, (5.9) yields the following relations:

$$\sqrt{4\Delta_0^2 + 9v_F^2\beta^2}\sinh^{-1}(3v_F\beta/2\Delta_0) = -3\pi v_F^2\beta D_0(3\beta) \tag{7.13}$$

and

$$\sqrt{\Delta_0^2 + v_F\beta^2}\sinh^{-1}(v_F\beta/\Delta_0) = -\pi v_F^2\beta D_0(2\beta). \tag{7.14}$$

Inserting the relations (7.11)–(7.14) as well as (6.7) into (7.9) and (7.10) yields finally

$$J_1(3\beta) = (1/\Delta_0)[\tfrac{1}{2}D_0(3\beta) - \tfrac{1}{3}D_0(2\beta)] \tag{7.15}$$

and

$$J_2(3\beta) = -(1/6\Delta_0^2)D_0(2\beta), \tag{7.16}$$

where (3.14) together with (3.26) have been used at will.

Again going to the Fourier representation and carrying out the appropriate differentiations and integrations yield the following results for $L_0(3\beta)$, $L_1(3\beta)$, and $L_2(3\beta)$:

$$\begin{aligned} L_0(3\beta) = -\frac{i\beta\,\mathrm{tr}}{(2\pi)^2}\int d^2p\{&(2p_1-\beta)S_0(p_0p_1)\Box_0(p_0p_1-\beta)S_0(p_0p_1-2\beta)S_0(p_0p_1-3\beta)\\ &+(2p_1-3\beta)S_0(p_0p_1)S_0(p_0p_1-\beta)\Box_0(p_0p_1-2\beta)S_0(p_0p_1-3\beta)\\ &+(2p_1-5\beta)S_0(p_0p_1)S_0(p_0p_1-\beta)S_0(p_0p_1-2\beta)\Box_0(p_0p_1-3\beta)\}, \end{aligned} \tag{7.17}$$

$$\begin{aligned} L_1(3\beta) = -\frac{i\beta\,\mathrm{tr}}{(2\pi)^2}\int d^2p\{&(2p_1-\beta)S_0(p_0p_1)\Box_0(p_0p_1-\beta)S_0(p_0p_1-3\beta)\\ &+(2p_1-5\beta)S_0(p_0p_1)S_0(p_0p_1-2\beta)\Box_0(p_0p_1-3\beta)\}, \end{aligned} \tag{7.18}$$

and

$$\begin{aligned} L_2(3\beta) = -\frac{4i\beta\,\mathrm{tr}}{(2\pi)^2}\int d^2p\{&(p_1-\beta)S_0(p_0p_1)\Box_0(p_0p_1-2\beta)S_0(p_0p_1-3\beta)\\ &+(p_1-2\beta)S_0(p_0p_1)S_0(p_0p_1-\beta)\Box_0(p_0p_1-3\beta)\}. \end{aligned} \tag{7.19}$$

Carrying out the above integrations yields finally, after very tedious manipulations,

$$L_0(3\beta) = (2i/v_F^2)[\tfrac{11}{3}D_0(2\beta) - D_0(3\beta)], \tag{7.20}$$

$$L_1(3\beta) + L_2(3\beta) = -(12i\Delta_0/v_F^2)[D_0(3\beta) - D_0(2\beta) + (S - m^2/g^2)]. \tag{7.21}$$

Inserting the explicit expressions for the $J_i(3\beta)$'s and $L_i(3\beta)$'s as well as Eqs. (5.13) and (5.19), together with (4.23), into Eq. (7.9) for a_3 yields the following:

$$D_0(3\beta)\Delta_0 a_3 = \frac{36\Delta_0^3}{v_F^2 m^2}\left[\frac{m^2}{g^2}((2\lambda+1)+\sigma^2)+\lambda D_0(3\beta)\right]b_3 - \frac{12g^2\Delta_0^3}{v_F^2 m^4}\left[S-\frac{n^2}{g^2}\right]^2(1-\sigma^2)^{-1}b_3$$
$$+\left[\Delta_0(a_1b_2+a_2b_1)-20\frac{\lambda\Delta_0^3}{v_F^2m^2}b_1b_2\right]\left[\frac{1}{2}D_0(3\beta)-\frac{1}{3}D_0(2\beta)\right]$$
$$-\frac{12g^2\Delta_0^3}{v_F^2m^4}b_1b_2\left[S-\frac{m^2}{g^2}\right][1-\sigma^2]^{-1}\left[D_0(3\beta)-D_0(2\beta)+\left(S-\frac{m^2}{g^2}\right)\right]$$
$$+\frac{\Delta_0}{2}\left[b_1^2a_1-\frac{4\lambda\Delta_0^2b_1^3}{v_F^2m^2}\right]D_0(2\beta)-\frac{2g^2\Delta_0^3b_1^3}{v_F^2m^4}\left[S-\frac{m^2}{g^2}\right][1-\sigma^2]^{-1}\left[\frac{11}{3}D_0(2\beta)-D_0(3\beta)\right]. \quad (7.22)$$

Using (3.18) together with (6.14) and (6.15), one rewrites the above relation as

$$D_0(3\beta)a_3 = -\frac{72\Delta_0^2}{v_F^2m^2}\left\{\frac{m^2}{g^2}((2\lambda+1)+\sigma^2)-\frac{g^2}{m^2}\left(S-\frac{m^2}{g^2}\right)^2(1-\sigma^2)^{-1}\right\}$$
$$+\left\{3a_1-\frac{48\Delta_0^2}{v_F^2m^2}\left[\lambda-\frac{g^2}{m^2}\left(S-\frac{m^2}{g^2}\right)(1-\sigma^2)^{-1}\right]\right\}D_0(3\beta). \quad (7.23)$$

Terms containing $D_0(2\beta)$ cancel among each other. Remembering the condition (5.21) for the soliton velocity, one therefore determines a_3 as

$$a_3 = 3a_1 - 24Q, \quad (7.24)$$

where Q was defined by Eq. (6.15).

VIII. A CLOSED FORM FOR $\hat{\phi}_1$

The results (6.14) and (7.24) indicate that, in general, each coefficient a_n is indeed separated into two parts, one proportional to a_1 and another proportional to Q. Let us therefore write

$$a_n = A_nQ - B_na_1 \quad (n \geqslant 1). \quad (8.1)$$

Equations (6.14) and (7.24) are consistent with the following:

$$B_n = (-1)^n n, \quad (8.2)$$

$$A_1 = 0. \quad (8.3)$$

In order to discover the sequence A_n, we will inspect the structure of Q as well as Eqs. (6.12) and (7.22) for a_2 and a_3, respectively. The key factor is the quantity $(S - m^2/g^2)$.

The characteristic form under which the calculation is displayed for each order in the asymptotic expansion is the following:

$$D_0(n\beta)a_n = 4n^2(\Delta_0^2/v_F^2m^2)[\text{Eq. (5.21)}]b_n + [a_nQ - B_na_1]D_0(n\beta) \quad (n=1,2,3), \quad (8.4)$$

which is merely a restatement of (8.1) since (5.21) is the condition for the soliton velocity and vanishes identically. The factor $4n^2$ in the first term comes from the expression (5.13) for $D_1(n\beta)$. At an earlier stage in the calculation, however, one has

$$D_0(n\beta)a_n = 4n^2\frac{\Delta_0^2}{v_F^2m^2}\left[\frac{m^2}{g^2}((2\lambda+1)+\sigma^2)\right]b_n - 4n\frac{\Delta_0 g}{v_F^2m^2}\left[S-\frac{m^2}{g^2}\right]c_n - \frac{\Delta_0^2}{v_F^2m^2}$$
$$\times\left[\delta_n^{(1)}\frac{g^2}{m^2}\left(S-\frac{m^2}{g^2}\right)(1-\sigma^2)^{-1}D_0(n\beta)+\delta_n^{(2)}\frac{g^2}{m^2}\left(S-\frac{m^2}{g^2}\right)^2(1-\sigma^2)^{-1}\right]+\gamma_nD_0(n\beta) \quad (n=1,2,3). \quad (8.5)$$

The first two terms are contributions from $D_1(n\beta)$ given by (5.13) and $D_2(n\beta)$ given by (5.19), respectively. The third term containing linear and quadratic forms for $(S - m^2/g^2)$ with respective coefficients $\delta_n^{(1)}$ and $\delta_n^{(2)}$ is obtained from computation of terms involving the acoustic phonon order parameter ξ_0. The remaining contribution is simply denoted by γ_n. Now assuming the validity of the general result (4.23) for c_n one rewrites (8.5) as

$$D_0(n\beta)a_n = \frac{\Delta_0^2}{v_F^2m^2}\left\{4n^2\left[\frac{m^2}{g^2}((2\lambda+1)+\sigma^2)\right]b_n - 4n\left[\frac{g^2}{m^2}\left(S-\frac{m^2}{g^2}\right)^2(1-\sigma^2)^{-1}\right]b_n\right.$$
$$\left.-\delta_n^{(1)}\left[\frac{g^2}{m^2}\left(S-\frac{m^2}{g^2}\right)(1-\sigma^2)^{-1}\right]D_0(n\beta)-\delta_n^{(2)}\left[\frac{g^2}{m^2}\left(S-\frac{m^2}{g^2}\right)^2(1-\sigma^2)^{-1}\right]\right\}+\gamma_nD_0(n\beta) \quad (n=1,2,3). \quad (8.6)$$

Noticing that the factor $(S - m^2/g^2)$ apears only linearly in the expression (6.15) for Q, one therefore obtains, by comparison with (8.4),

$$A_n = \delta_n^{(1)}/2 \quad (n=1,2,3). \quad (8.7)$$

The factor $\frac{1}{2}$ comes in because Q has been defined with a factor 2. The quantity $(S - m^2/g^2)$, however, appears only quadratically in the condition (5.21) for the soliton velocity. Comparing (8.4) and (8.6), one obtains

$$\delta_n^{(2)} = (4n^2 - 4n)b_n \quad (n = 1,2,3). \tag{8.8}$$

The next critical step is based on the following nontrivial observation:

$$\delta_n^{(1)} = \delta_n^{(2)} \quad (n = 1,2,3). \tag{8.9}$$

Upon the assumption that the above considerations remain valid for any order n, Eqs. (8.7)–(8.9) lead to

$$A_n = 2n(n-1)b_n \quad (n \geqslant 1). \tag{8.10}$$

From (3.18) one finally obtains

$$A_n = (-1)^n 4n(n-1) \quad (n \geqslant 1). \tag{8.11}$$

This generates the sequence

$$A_n = \{0, 8, -24, 48, -80, 120, - + \cdots\}, \tag{8.12}$$

which agrees with (6.14), (7.24), and (8.3). Therefore (8.1) becomes

$$a_n = (-1)^n [4n(n-1)Q - na_1], \tag{8.13}$$

which in turn determines $\hat{\phi}_1$ as

$$\hat{\phi}_1 = 4Q \sum_{n=1}^{\infty} (-1)^n n(n-1) e^{-nX} - a_1 \sum_{n=1}^{\infty} (-1)^n n e^{-nX} \quad [X > 0]. \tag{8.14}$$

Now, since

$$\tanh \frac{X}{2} = 1 + 2 \sum_{n=1}^{\infty} (-1)^n e^{-nX} \quad [X > 0] \tag{8.15}$$

and

$$\operatorname{sech} \frac{X}{2} = 2 \sum_{n=0}^{\infty} (-1)^n e^{-(2n+1)/2X} \quad [X > 0], \tag{8.16}$$

one obtains

$$\operatorname{sech}^2 \frac{X}{2} = -4 \sum_{n=1}^{\infty} (-1)^n n e^{-nX} \quad [X > 0] \tag{8.17}$$

and

$$\operatorname{sech}^2 \frac{X}{2} \left[\tanh \frac{X}{2} - 1 \right] = -4 \sum_{n=1}^{\infty} (-1)^n n(n-1) e^{-nX} \quad [X > 0]. \tag{8.18}$$

Inserting (8.17) and (8.18) into (8.14) gives finally

$$\hat{\phi}_1 = \frac{a_1}{4} \operatorname{sech}^2 \frac{X}{2} - Q \operatorname{sech}^2 \frac{X}{2} \left[\tanh \frac{X}{2} - 1 \right], \tag{8.19}$$

where X was defined in (3.22).

The rather exotic argument presented in this section enabled us to bring $\hat{\phi}_1$ into the very appealing closed form (8.19). The complete form for the perturbed soliton, up to v^2 order, can now be written as

$$\phi(X) = \tanh \frac{X}{2} + v^2 \left[\left(\frac{a_1}{4} + Q \right) \operatorname{sech}^2 \frac{X}{2} - Q \operatorname{sech}^2 \frac{X}{2} \tanh \frac{X}{2} \right]. \tag{8.20}$$

At any finite time, the following topological properties should be realized:

$$\phi(+\infty) = -\phi(-\infty) = 1. \tag{8.21}$$

This implies that, for the same time, the soliton obeys the following condition at its center:

$$\phi(0) = 0. \tag{8.22}$$

The latter boundary condition determines the coefficient a_1 in (8.20). It is easily found to be

$$a_1 = -4Q. \tag{8.23}$$

Therefore one is led finally to

$$\phi(X) = \left[1 - v^2 Q \operatorname{sech}^2 \frac{X}{2} \right] \tanh \frac{X}{2}, \tag{8.24}$$

which has the required topological properties.

The soliton profile (8.24) is constrained to move along the molecular chain with the velocity

$$v_{\text{soliton}} = \sigma v, \tag{8.25}$$

where σ is given by (5.23).

Again a complete proof for the soliton profile (8.24) requires the self-consistent computation mentioned in the Introduction. Such a proof is briefly sketched in Appendix B.

IX. CONCLUSION

An important feature of the solution (8.24) is that the space-time coordinates always appear in the configuration X given by (3.22). It therefore has the form of a purely boosted solution. However, since the full generalized model is not Lorentz invariant, as opposed to the TLM model, one expects an additional constraint on the solution (8.24). This constraint chooses a preferred frame, which is given by (8.25). The appearance of a special configuration X for space-time coordinates in soliton systems treated perturbatively is not a feature restricted to our model. Other soliton systems, like the modified sine–Gordon equation,[5] which is used to model the Josephson junction, have also a special configuration X.

The determination of a suitable X is closely related to the problem of choosing the unperturbed state for the perturbation calculations. The explicit form for the choice of the configuration coordinate X is determined by the physical properties of the soliton system. A boosted like form seems to be appropriate for nondissipative systems as in our case while a more complex form may be suitable for systems violating the energy conservation law as in the soliton model for the Josephson junction. The dissipative (nondissipative) quality of a soliton system seems to be reflected by the nonintegrability (integrability) property of its solution.

Now let us make some comments on the observability of the features of the present quasirealistic model in accordance with previous numerical computations.[10] However, although our quasirealistic model predicts that the shape of the kink remains independent of time for the range of λ allowed by (5.23), solutions to the truly realistic model may be more complicated than a constant motion. Oscillations may also be present.[11] The region $-2 < \lambda < 0$ in our case may well correspond to a forbidden zone for constant translational motion. Note that a value $\lambda \simeq -4.22$ yields the maximum velocity $v_{\text{sol}} \simeq 2.7v$ of Ref. 11. For positive λ, the soliton velocity is always smaller than v. Finally, as will be seen in Appendix B, acoustic effects do not destroy the fermionic zero-energy mode of the TLM model. This mode, however, should be properly called a zero mode only in the frame moving with the soliton. As a result the well-known charge fractionalization mechanism[1,16,18-20] should remain an observable consequence of the model.

To put even more emphasis on the above results, we can mention the fact that it has been shown that a static soliton solution is also inconsistent with the truly realistic continuum model. Work on this model is now in progress.

ACKNOWLEDGMENTS

This work was supported by the Natural Sciences and Engineering Research Council, Canada, and the Dean of Science, Faculty of Science, The University of Alberta, Edmonton, Alberta, Canada. One of the authors (YL) is supported by a postgraduate NSERC Scholarship.

APPENDIX A: DERIVATION OF THE QUASIREALISTIC MODEL

In this appendix we formally derive the completely realistic generalized TLM model from the discrete SSH model and indicate the modification which leads to the quasirealistic model.

Su, Schrieffer, and Heeger[7] proposed the following model for a linear chain of trans-$(\mathrm{CH})_X$, the polyacetylene modecule, as equal to

$$\begin{aligned}\sum_n C_n^\dagger\Big(i\frac{\partial}{\partial t}-\mu\Big)C_n &+\sum_n\frac{1}{2}\Big[\rho\Big(\frac{\partial u_n}{\partial t}\Big)^2-K(u_{n+1}-u_n)^2\Big]\\ &+\sum_n t_{n,n+1}(C_nC_{n+1}+C_{n+1}^\dagger C_n^\dagger),\end{aligned}\tag{A1}$$

where C_n is the destruction operator of the electron at the lattice point r_n, u_n is the displacement field of the lattice r_n, μ is the chemical potential for the electron, ρ is the mass of the lattice atom, and K is the spring constant. The hopping matrix element $t_{n,n+1}$ is given by

$$t_{n,n+1}=t_0-\alpha(u_{n+1}-u_n).\tag{A2}$$

The operators u_n and C_n satisfy the following canonical commutation relations:

$$\Big[u_n,\frac{\partial}{\partial t}u_n\Big]=\frac{i\hbar}{\rho}\delta_{nm},\tag{A3}$$

$$[C_n,C_m^\dagger]_+=\delta_{nm}.\tag{A4}$$

In order to go to the continuum model, it is convenient to introduce suitable Fourier transforms. For an operator A_n associated with the lattice point r_n, we define its Fourier transform as

$$A[k]=\sqrt{\frac{a}{2\pi}}\sum_n e^{-ikr_n}A_n,\tag{A5}$$

where a is the lattice spacing. The inverse transform is expressed by

$$A_n=\sqrt{\frac{a}{2\pi}}\int_{-\pi/a}^{\pi/a}dk\,e^{ikr_n}A[k].\tag{A6}$$

Dividing the Brillouin zone into half, we define two fields $\psi_1[k]$, $\psi_2[k]$ as well as acoustic phonon $\xi[q]$ and optical phonon $\Phi[q]$ as

$$\psi_1[k]\equiv C[k+\pi/2a],\quad -\pi/2a<k<\pi/2a,\tag{A7}$$

$$\psi_2[k]\equiv -iC[k-\pi/2a],\quad -\pi/2a<k<\pi/2a,\tag{A8}$$

and

$$\xi[q]\equiv u[q],\quad -\pi/2a<q<\pi/2a,\tag{A9}$$

$$\Phi[q]\equiv u[q+\pi/a],\quad -\pi/2a<q<\pi/2a.\tag{A10}$$

By making use of the relations

$$\frac{a}{2\pi}\sum_n\exp(ikr_n)=\sum_N\delta\Big(k+\frac{2\pi N}{a}\Big),\tag{A11}$$

$$\frac{a}{2\pi}\int_{-\pi/a}^{\pi/a}dk\exp[ik(r_n-r_m)]=\delta_{nm},\tag{A12}$$

as well as the definition

$$\psi[k]\equiv\begin{pmatrix}\psi_1[k]\\ \psi_2[k]\end{pmatrix},\tag{A13}$$

we can express the Lagrangian (A1) in terms of Fourier amplitudes and expand it in powers of a. Defining the fields in configuration space as

$$\psi(x)=\int\frac{dk}{\sqrt{2\pi}}\psi[k]e^{ikx},\tag{A14}$$

$$\xi(x)=\sqrt{\rho}\int\frac{dq}{\sqrt{2\pi}}\xi[q]e^{ikx},\tag{A15}$$

$$\Phi(x)=\sqrt{\rho}\int\frac{dq}{\sqrt{2\pi}}\Phi[q]e^{ikx},\tag{A16}$$

we have, up to order a^2,

$$L=\int dx\,\mathscr{L}(x),\tag{A17}$$

where

$$\begin{aligned}\mathscr{L}=\psi^\dagger\Big(i\frac{\partial}{\partial t}-\mu+iv_F\tau_3\frac{\partial}{\partial x}\Big)\psi&+\frac{1}{2}\Big[\dot\xi^2-v^2\Big(\frac{\partial\xi}{\partial x}\Big)^2\Big]\\ &+\frac{1}{2}\Big[\dot\Phi^2+v^2\Big(\frac{\partial\Phi}{\partial x}\Big)^2-m^2\Phi\Big]+g\psi^\dagger\tau_1\psi\Phi\\ &+g\frac{v^2}{m^2}\Big[-i\Big(\psi^\dagger\tau_3\frac{\partial\psi}{\partial x}-\frac{\partial\psi^\dagger}{\partial x}\tau_3\psi\Big)\frac{\partial\xi}{\partial x}\\ &+\Big(\psi^\dagger\tau_1\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi^\dagger}{\partial x^2}\tau_1\psi\Big)\Phi\Big],\end{aligned}\tag{A18}$$

in which

$$v_F\equiv 2at_0,\quad v^2\equiv(K/\rho)a^2,\quad m^2\equiv 4K/\rho,\quad g=4\alpha\sqrt{a/\rho}.\tag{A19}$$

The Lagrangian density (A18) describes the completely realistic generalized continuum model for polyacetylene. The passage to the quasirealistic model consistent with an acoustic-effects-free BCS-type gap equation is simply obtained by replacing the derivative coupling term involving only the optical phonon by the following correction:

$$\mathscr{L}_{\mathrm{corr}}=g\frac{v^2}{m^2}\frac{\partial}{\partial x}(\psi^\dagger\tau_1\psi)\frac{\partial\Phi}{\partial x},\tag{A20}$$

where λ is a dimensionless parameter. This *ad hoc* modification yields the quasirealistic model as described by (2.1)–(2.3).

The interaction (A20) term is the only possible form that respects both the total degree of derivative of the original interaction term as well as the BCS-type condition imposed on the gap equation. The constant λ is introduced to correct roughly the change of effective coupling. This parameter remains unknown in the present context and is to be determined from experimental results.

APPENDIX B: SELF-CONSISTENT PROOF FOR SOLITON SOLUTIONS

In this appendix we sketch briefly, without going into the computational details, a self-consistent consideration that presents a complete proof for the soliton solutions (acoustic and optical phonon order parameters) obtained in this paper.

The consistency of our solutions with the mean-field equations (2.9)–(2.11) is most easily shown by assuming from the start the validity of the results of this paper. They are summarized as

$$\phi = \phi_0(X) - v^2\phi_1(X) + \cdots \tag{B1}$$

and

$$\langle \xi \rangle = \xi_0(X) + \cdots, \tag{B2}$$

where the dots stand for higher-order corrections and

$$\phi_0(X) = \tanh(X/2), \tag{B3}$$

$$\phi_1(X) = Q\,\mathrm{sech}^2(X/2)\tanh(X/2) = Q(\phi_0 - \phi_0^3), \tag{B4}$$

and

$$\xi_0(X) = R\phi_0(X) + \text{arbitrary constant.} \tag{B5}$$

The "generalized coordinate" X is specified by Eq. (1.2) together with the constraint (1.3) on the soliton velocity, while the constants Q and R have been obtained as

$$Q = (2\Delta_0^2/v_F^2 m^2)[\lambda + (1-\sigma^2)^{-1}], \tag{B6}$$

$$R = -(\Delta_0/g)(1-\sigma^2)^{-1}. \tag{B7}$$

Using the relations (B1)–(B7) we can now explicitly calculate the fermion wave functions from the field equation (2.9). Once they are obtained, these wave functions are then inserted back into the source terms of (2.10) and (2.11), thus checking the consistency of (B1)–(B7). To that purpose, one must, however, expand the electron wave functions in powers of the acoustic phonon velocity as in (B1) and (B2). But since the unperturbed soliton part is the boosted TLM kink, one must be careful to implement such a perturbation expansion in the frame moving with the soliton and thereafter extract the Lorentz invariant part as the unperturbed state. Expanding the electron field in such a frame (X_0, X_1) as

$$\psi(X_0X_1) = \psi_0(X_0X_1) + v^2\psi_1(X_0X_1) + \cdots$$
$$= \left[\begin{pmatrix} U_0(X_1) \\ V_0(X_1)\end{pmatrix} + v^2\begin{pmatrix} U_1(X_1) \\ V_1(X_1)\end{pmatrix} + \cdots\right] \times e^{-i\omega_k/v_F X_0 + ikX_1}, \tag{B8}$$

one obtains the following set of linearized field equation (up to the correct order):

$$(\omega - v_F k)U_0 + iv_F U_0' + \Delta_0\phi_0 V_0 = 0, \tag{B9}$$

$$(\omega + v_F k)V_0 - iv_F V_0' + \Delta_0\phi_0 U_0 = 0, \tag{B10}$$

$$(m^2\Delta_0\phi_0/g^2)(X_1) = \langle 0|\psi_0^\dagger(X_0X_1)\tau_1\psi_0(X_0X_1)|0\rangle, \tag{B11}$$

for the unperturbed part (which is the TLM model in the boosted frame) and

$$(\omega - v_F k)U_1 + iv_F U_1' + \Delta_0\phi_0 V_1 = \Delta_0\left(\phi_1 + \frac{\lambda\phi_0''}{m^2}\right)V_0 + \frac{ig}{m^2}(\xi_0'' + 2ik\xi_0')U_0 + 2\frac{ig}{m^2}\xi_0' U_0', \tag{B12}$$

$$(\omega + v_F k)V_1 - iv_F V_1' + \Delta_0\phi_0 U_1 = \Delta_0\left(\phi_1 + \frac{\lambda\phi_0''}{m^2}\right)U_0 - \frac{ig}{m^2}(\xi_0'' + 2ik\xi_0')V_0 - 2\frac{ig}{m^2}\xi_0' V_0', \tag{B13}$$

and

$$[1-\sigma^2]\xi_0''(X_1) = -\frac{ig}{m^2}\langle 0|[\psi_0^\dagger\tau_3\psi_0'' - \psi_0^{\dagger\prime\prime}\tau_3\psi_0]|0\rangle, \tag{B14}$$

$$[(\lambda+1)+\sigma^2]\frac{\Delta_0\phi_0''}{g^2}(X_1) - m^2\frac{\Delta_0}{g^2}\phi_1(X_1) = \langle 0|[\psi_0^\dagger\tau_1\psi_1 + \psi_1^\dagger\tau_1\psi_0]|0\rangle, \tag{B15}$$

for the perturbation. Of course the boosted spatial coordinate X_1 is related to the configuration X at the origin by

$$X = (2\Delta_0/v_F)X_1. \tag{B16}$$

The electron energy ω_k is given as

$$\omega_k = \pm\sqrt{v_F^2k^2 + \Delta_0^2}. \tag{B17}$$

The solutions to the unperturbed system (B9) and (B10) are well known[1,16] and a detailed review has been given by Nakahara.[16] The results are

$$U_0 = \frac{A_k}{2}\left[\frac{(\omega_k + v_F k) + i\Delta_0\phi_0}{\omega_k}\right], \tag{B18}$$

$$V_0 = \frac{A_k}{2}\left[\frac{-i(\omega_k - v_F k) - \Delta_0\phi_0}{\omega_k}\right], \tag{B19}$$

where A_k is a normalization factor. Note that Eqs. (B9) and (B10) also have a zero-frequency solution ($\omega = 0$),

$$U_{0_B} = -iV_{0_B} = \tfrac{1}{2}\sqrt{\Delta_0/2v_F}\sqrt{1-\phi_0^2}. \tag{B20}$$

Although such a bound state zero mode does not contribute in shaping the soliton profile as opposed to the scattered modes (B18) and (B19), it yields nevertheless the so-called charge fractionalization mechanism. This zero mode, however, should be properly called as such only in the frame moving with the soliton.

Using the relations

$$\phi_0' = (\Delta_0/v_F)(1-\phi_0^2), \tag{B21}$$

$$\phi_0'' = -(2\Delta_0^2/v_F^2)(\phi_0 - \phi_0^3), \tag{B22}$$

one easily realizes that the perturbed scattered wave functions solutions of (B12) and (B13) are now given as polynomials in ϕ_0. They are readily obtained as

$$U_1 = \frac{g\Delta_0 R}{v_F^2 m^2}\left\{U_0 + A_k\left[\frac{i(\omega_k + v_F k)}{\Delta_0}\phi_0 - \frac{(\omega_k + 3v_F k)}{2\omega_k}\phi_0^2 - \frac{3i\Delta_0}{2\omega_k}\phi_0^3\right]\right\}, \tag{B23}$$

$$V_1 = \frac{g\Delta_0 R}{v_F^2 m^2}\left\{V_0 + A_k\left[-\frac{(\omega_k + v_F k)}{\Delta_0}\phi_0 + i\frac{(\omega_k + 3v_F k)}{2\omega_k}\phi_0^2 + \frac{3\Delta_0}{2\omega_k}\phi_0^3\right]\right\}. \tag{B24}$$

The correction to the zero mode is obtained as

$$U_{1_B} = -iV_{1_B} = (g\Delta_0 R / v_F^2 m^2)[1 - 3\phi_0^2]U_{0_B}. \tag{B25}$$

Since the normalization of the fermion number obtained from (B25) is not altered by these acoustic corrections, so is the charge fractionalization mechanism.

Rewriting (B11), (B14), and (B15) as

$$\frac{m^2\Delta_0}{g^2}\phi_0 = \sum_k [U_0^* V_0 + \text{c.c.}], \tag{B26}$$

$$(1-\sigma^2)\xi_0'' = -\frac{ig}{m^2}\sum_k [U_0^* U_0'' - V_0^* V_0'' - \text{c.c.}], \tag{B27}$$

$$((1+\lambda)+\sigma^2)\frac{\Delta_0}{g^2}\phi_0'' - m^2\frac{\Delta_0}{g^2}\phi_1 = \sum_k [U_0^* V_1 + V_0^* U_1 + \text{c.c.}], \tag{B28}$$

one can now insert back the wave functions (B18), (B19), (B23), and (B24) into the rhs's of (B26) and (B28). Doing so Eq. (B26) yields the gap equation

$$\frac{m^2}{g^2} = -\sum_k \frac{A_k^2}{\omega_k}, \tag{B29}$$

while (B27) together with the above gap equation yields the results (B5) and (B7) for the acoustic phonon order parameter. Finally insertion into (B28) of the electron wave functions as well as the use of (B4), (B6), (B21), (B22), and the gap equation (B29) yields the following condition for the soliton velocity:

$$(2\lambda + 1) + \sigma^2 - (1-\sigma^2)^{-1} = 0. \tag{B30}$$

The physical root of this equation is just Eq. (1.3). One has proven therefore that our soliton solutions to the present model are consistent with its mean-field equations if and only if the condition (1.3) on the soliton velocity is implemented.

[1]H. Takayama, Y. R. Lin-Liu, and K. Maki, Phys. Rev. B **21**, 2388 (1980).
[2]D. W. McLaughlin and A. C. Scott, Phys. Rev. A **18**, 1652 (1978).
[3]V. I. Karpman and V. V. Solov'ev, "The influence of external perturbations on solitons in Josephon junctions," IZMIRAN Preprint No. 28, Moscow, 1980, p. 294.
[4]M. Salerno and A. C. Scott, Phys. Rev. B **26**, (1982).
[5]M. Salerno, M. P. Soerensen, O. Skovgaard, and P. L. Christiansen, Wave Motion **5**, 49 (1983).
[6]V. I. Karpman and V. V. Solov'ev, Physica D **3**, 487 (1981).
[7]W. P. Su, J. R. Schrieffer, and A. J. Heeger, Phys. Rev. Lett. **42**, 1698 (1979); Phys. Rev. B **22**, 2099 (1980); Phys. Rev. Lett. **46**, 738 (1981).
[8]H. Matsumoto, N. J. Papastamatiou, and H. Umezawa, Nucl. Phys. B **82**, 45 (1974); Nucl. Phys. B **97**, 90 (1975); G. Oberlechler, N. Umezawa, and C. Zelses, Lett. Nuovo Cimento **23**, 641 (1978); L. Wadati, H. Matsumoto, and H. Umezawa, Phys. Lett. B **73**, 448 (1978).
[9]Y. Leblanc, H. Matsumoto, H. Umezawa, and F. Mancini, Phys. Rev. B **30**, 5958 (1984).
[10]W. P. Su and J. R. Schrieffer, Proc. Natl. Acad. Sci. USA **77**, 5626 (1980).
[11]A. R. Bishop, D. K. Campbell, P. S. Lomdahl, B. Horowitz, and S. R. Phillpot, Phys. Rev. Lett. **52**, 671 (1984).
[12]D. K. Campbell and A. R. Bishop, Nucl. Phys. B **200** [FS4] 2, 297 (1982).
[13]S. A. Brazovskii, JETP Lett. **28**, 606 (1978) [Pis'ma Zh. Eksp. Teor. Fiz. **28**, 656 (1978)]; Sov. Phys. JETP **51**, 342 (1980) [Zh. Eksp. Teor. Fiz. **78**, 677 (1980)].
[14]S. A. Brazovskii and N. Kirova, Pis'ma Zh. Eksp. Teor. Fiz. **33**, 6 (1981); D. K. Campbell and A. R. Bishop, Phys. Rev. B **24**, 4859 (1981).
[15]M. Nakahara and K. Maki, Phys. Rev. B **25**, 7789 (1982).
[16]M. Nakahara, Ph.D. thesis, University of Kyoto, Japan, 1981.
[17]G. Rella, S. Mancini, M. Marimiro, and G. Scarpetta, Phys. Lett. A **100**, 482 (1984).
[18]J. Goldstone and R. Jackiw, Phys. Rev. D **11**, 1486 (1975); R. Jackiw and C. Rebbi, Phys. Rev. D **13**, 3398 (1976).
[19]R. Jackiw and J. R. Schrieffer, Nucl. Phys. B **190** [FS3], 253 (1981).
[20]G. Semenoff, H. Matsumoto, and H. Umezawa, Phys. Rev. D **25**, 1054 (1982).

Unbounded representations of symmetry groups in gauge quantum field theory. I. Confinement and differentiation

A. H. Völkel
Institut für Mathematik III, Freie Universität Berlin, 1000 Berlin 33, West Germany

(Received 27 March 1985; accepted for publication 12 June 1985)

Symmetry groups and especially the covariance (substitution rules) of the basic fields in a gauge quantum field theory of the Wightman–Gårding type are investigated. By means of the continuity properties hidden in the substitution rules it is shown that every unbounded form–isometric representation U of a Lie group has a form-skew-symmetric differential ∂U with dense domain in the unphysical Hilbert space. Necessary and sufficient conditions for the existence of the closures of U and ∂U as well as for the isometry of U are derived. It is proved that a class of representations of the translation group enforces a relativistic confinement mechanism, by which some or all basic fields are confined but certain mixed products of them are not.

I. INTRODUCTION

In recent years there has been increasing evidence that gauge quantum field theories may be the only kind of field theories relevant to elementary particle physics. According to theorems of Strocchi and Wightman[1–4] and references quoted by them, it seems to be unavoidable for a consistent formulation of a gauge quantum field theory to introduce an indefinite metric formalism.[5] This formalism makes use of three Hilbert spaces $H_0 \subset H \subset \mathscr{H}$ with scalar product $(\cdot,\cdot)$ and a Hermitian sesquilinear form $\langle \cdot,\cdot \rangle = (\cdot, \eta\, \cdot)$, which is 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 $\mathscr{H}_{\rm ph}$. As a consequence there emerge a variety of structural differences between gauge quantum field theories and the classical Wightman theories.[6–8] We mention only a few which are relevant for the consideration below: Physically interesting quantities (like transition amplitudes, etc.) have to be computed in terms of the sesquilinear form $\langle \cdot,\cdot \rangle$ whereas the basic fields of the theory are operators in $\mathscr{H}$ rather than in $\mathscr{H}_{\rm ph}$. Observables have to be symmetric with respect to the form $\langle \cdot,\cdot \rangle$ (in the following called η symmetry) but not necessarily essentially self-adjoint in $\mathscr{H}$. Similarly symmetry groups are represented by operators in $\mathscr{H}$ which are isometric with respect to the form $\langle \cdot,\cdot \rangle$ (in the following called η isometry), but in general they are neither unitary nor even bounded in $\mathscr{H}$. The best-known example for such an unbounded representation is probably the Lorentz group in the Gupta–Bleuler formulation of the free electromagnetic vector potential (Ref. 5, Sec. III).

Within the classical Wightman quantum field theories the unitarity of a representation of a Lie group establishes the existence of its differential; this means a representation of its Lie algebra in terms of essentially skew-adjoint operators (Ref. 9, Chap. 4). Conversely, under a variety of well-known conditions (See Ref. 9, Chap.4, and Refs. 10–18) a representation of a Lie algebra by linear (skew-adjoint) operators can be integrated to a continuous (unitary) representation of the universal covering group of a corresponding Lie group. In a gauge quantum field theory this connection is an open problem in both directions. *A priori* the differential may fail to exist because the representation of the Lie group is not continuous. All integrability conditions known (to the author) from the literature lead to continuous group representations. Hence in general they are not directly applicable in the present case. On the other hand in a gauge quantum field theory the connection between the representations of a Lie group and its Lie algebra is of the same fundamental importance for the physical interpretation as in the Wightman theory, since the η-skew symmetric differentials of a symmetry group are the candidates for the observables of the theory. Moreover conditions for the integrability of an η-skew symmetric representation of a Lie algebra, respectively, for its failure are an important aspect in the deeper understanding of symmetry breaking mechanisms. Thus a systematic investigation of the differentiation and integration of unbounded representations of Lie groups, respectively, Lie algebras in gauge quantum field theories seems to be of vital interest.

The present note is concerned with the differentiation of unbounded representations of groups satisfying the η isometry and especially the covariance condition (substitution rule for the basic fields) in $\mathscr{H}$. It will be shown that in gauge quantum field theories (over a quite general class of countably normed test function spaces including the Schwartz as well as the Jaffe spaces[19]), which satisfy the Wightman-Gårding axioms[5] with the possible exception of locality, Lorentz covariance, and spectrum condition, any such representation of a Lie group possesses an η-skew-symmetric differential with dense domain in $\mathscr{H}$ (Sec. V). Sufficient conditions for the existence of the closure for the representatives of both the original group representation and its differential are proved (Sec. IV, respectively, Sec. V). As a by-product we derive a necessary and sufficient condition for the unitarity of the group representation in terms of special covariance properties of the basic fields for those theories in which the vacuum state is an eigenstate of the metric operator η (Sec. IV). This condition shows that in these theories the "usual" one-dimensional representations of the translation group are always unitary. The key to all our investigations and their results is the observation that via the covariance of the basic fields the unbounded representations in the Hilbert space $\mathscr{H}$ are closely related to continuous representations on the underlying countably normed test function space and the com-

 0022-2488/85/112956-18$02.50

pleted tensor products of them (Sec. III). For the latter the differentiation is a well-understood operation (Ref. 9, Chap. 4). In order to pave the ground for the investigation of the inverse operation (the integration of an η-skew-symmetric representation of a Lie algebra) an infinitesimal characterization for the subspaces of $\mathscr{C}^{\infty}$ vectors of the bounded representations in the countably normed test function spaces is derived (Sec. VI). For the integration itself will be treated in a separate note. For almost two decades the confinement of the quarks, etc., has been the central problem of elementary particle physics. In Sec. VIII we prove that a relativistic confinement mechanism is enforced by a certain class of representations of the translation group $(\mathbb{R}_4, +)$. It is self-evident that this program has to be started with a record of the axioms for a gauge quantum field theory and the discussion of symmetry groups in the necessary and unavoidable minuteness of detail (Sec. II).

II. GAUGE QUANTUM FIELD THEORIES AND SYMMETRY GROUPS

As indicated in the Introduction a distinguished role will be played by the test function spaces. We have to specify them first. In the following, $S(\mathbb{R}_{4L},\mathbb{C}_N)$ ($L,N \in \mathbb{N}$, natural numbers) denotes a separable and nuclear Frechét space (i.e., locally convex, metrizable, and complete) of complex $\mathscr{C}^{\infty}$ functions $f:\mathbb{R}_{4L}\to\mathbb{C}_N$, $x\to f(x)=(f^1(x),...,f_N(x))$ in $4L$ real variables $x=(x_1^0,...,x_1^3,...,x_L^0,...,x_L^3)$. The topology is given by a countable set of pairwise compatible norms $\|f\|_p$ ($p\in\mathbb{N}^0=\mathbb{N}\cup\{0\}$) (Ref. 20, Chap. I). Finally, it is assumed that the nuclear theorem holds. It roughly says that any multilinear functional defined on all product functions

$$\left(\otimes_{i=1}^{L} f_i\right)(x)=\left(\prod_{i=1}^{L} f_i^1(x_i),...,\prod_{i=1}^{L} f_i^N(x_i)\right)$$

$[x_j=(x_j^0,...,x_j^3)]$ and continuous in each variable f_j separately has a unique extension to a continuous linear functional on $S(\mathbb{R}_{4L},\mathbb{C}_N)$. Three important properties which will be needed below are (Ref. 20, Chap. I, Secs. 3–6, and Ref. 21, Part III, Proposition 50.2) the following.

(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 exists a $c\in\mathbb{R}^+$ (positive real numbers) and a minimal $p_0\in\mathbb{N}^0$ such that $|F(f)|\leqslant c\|f\|_{p_0}$.

Physically important examples of such spaces are, besides the Schwartz space of strongly decreasing $\mathscr{C}^{\infty}$ functions, all strictly localizable spaces of Jaffe $S^g(\mathbb{R}_{4L},\mathbb{C}_N)=\mathscr{L}(\mathbb{R}_{4L})\bar{\otimes}\,\mathbb{C}_N$ ($\bar{\otimes}$ denotes the completed tensor product).[19] The countable set of norms generating their topologies is given by

$$\begin{aligned}\|f\|^g_{\nu(m,\rho)} &= \|f\|^g_{m,\rho} := \sup\Bigg\{\prod_{t=1}^{L} g\Big(\rho\sum_{s=0}^{3}(p_t^s)^2\Big) \\ &\times\left[\prod_{j=1}^{L}\prod_{i=0}^{3}(1+|p_j^i|)^m\right]\left|\int d^{4L}x\, f^{\mu}(x)x^{|r|}\right. \\ &\times\exp\left[-i\sum_{k=1}^{L}(p_k\cdot x_k)\right]\Bigg|\ \Bigg|\ p\in\mathbb{R}_{4L};\ r_j^i\in\mathbb{N}^0, \\ &\sum_{t=1}^{L}\sum_{s=0}^{3} r_t^s<m;\ \mu=1,...,N\Bigg\},\end{aligned}\tag{2.1}$$

with $m\in\mathbb{N}^0$, $\rho\in\mathbb{N}$, and

$$(p\cdot x):=p^0x^0-\sum_{\alpha=1}^{3}p^{\alpha}x^{\alpha},\quad x^{|r|}:=\prod_{j=1}^{L}\prod_{i=0}^{3}(x_j^i)^{r_j^i}.\tag{2.2}$$

Furthermore, ν denotes a bijection from $\mathbb{N}^0\times\mathbb{N}$ onto $\mathbb{N}^0$. Here, $g:\mathbb{C}\to\mathbb{R}$ is some entire function which is positive and monotonically growing on $\mathbb{R}^+\cup\{0\}$ and satisfies the condition (strictly localizability)

$$\int_1^{\infty} dt\, t^{-2}\ln g(t^2)<+\infty.$$

Note that for $g\equiv1$ we get back the Schwartz spaces.

With these preparations we can proceed to the formulation of the axioms for a gauge quantum field theory (GQFT).[2,5]

A.I: Field operators: Let $\mathscr{H}$ denote a Hilbert space with elements Ψ, Φ,..., scalar product (Ψ,Φ), and norm $\|\Psi\|_{\mathscr{H}}=(\Psi,\Phi)^{1/2}$, D a dense linear subset of $\mathscr{H}$ and $\mathbf{T}$ an at most countable set of (multi-) indices $\Gamma,\Delta,...$. Then for every $f\in S(\mathbb{R}_4,\mathbb{C})$ and $\Gamma\in\mathbf{T}$ there exists a linear operator $\varphi_\Gamma(f)$ with domain $D(\varphi_\Gamma(f))$ such that (a) $D\subseteq D(\varphi_\Gamma(f))$ and $\varphi_\Gamma(f)D\subseteq D$; (b) if $\varphi_{\Gamma^*}(\bar f)=\varphi_{\underset{\Gamma}{*}}(\bar f)=\varphi_\Gamma(f)^*$ ($\bar f$ complex conjugate of f) denotes the adjoint operator of $\varphi_\Gamma(f)$, then $\Gamma^*\in\mathbf{T}$ for every $\Gamma\in\mathbf{T}$; 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 $S(\mathbb{R}_4,\mathbb{C})$.

Since all field operators together with their adjoint ones possess a dense domain, they are all closable. For the sake of notational simplicity we assume they are closed; that is, $\varphi_{\Gamma^{**}}=\varphi_\Gamma$.

For notational convenience we denote the mapping $f\to\mathrm{id}_{\mathscr{H}}\int d^4x\, f(x)$ [$f\in S(\mathbb{R}_4,\mathbb{C})$ and $\mathrm{id}_{\mathscr{H}}$ the identity operator in $\mathscr{H}$] in misuse of the phrase field operator by $\varphi_0(f)$ and assume $0,0^*\in\mathbf{T}$. Of course φ_0 possesses all properties of a field operator.

A.II: Metric operator and physical Hilbert space: There exists a linear, bounded, and Hermitian operator η with $\eta D\subseteq D$ which generates a nontrivial and nonpositive semidefinite sesquilinear form $\langle\cdot,\cdot\rangle:=(\cdot,\eta\cdot)$ on $\mathscr{H}$. Furthermore, there exists a nontrivial and maximal linear subspace $H\subset\mathscr{H}$ such that for all $\Psi\in H$, $\langle\Psi,\Psi\rangle\geqslant0$. 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 Hilbert space of physical states $\mathscr{H}_{\mathrm{ph}}$.

Let us remark in passing that we do not demand the form $\langle\cdot,\cdot\rangle$ to be nondegenerate. Hence η is not necessarily invertible. On the other hand our continuity assumption A.I (c) is stronger than that in Refs. 2 and 5, where only the continuity of the matrix elements $\langle\Psi,\varphi_\Gamma(f)\Phi\rangle$ in f is assumed. If η has an inverse, however, then both are equivalent.

Throughout this paper we mean by a representation R of the group $(G,\cdot)$ on a vector space E a homomorphism $R:G\to\mathrm{Aut}\,E$, $g\to R(g)$ of $(G,\cdot)$ into the automorphism group of E and by a representation W of a Lie algebra $(\mathfrak{g},\otimes,\odot,[\ ,\ ])$ on E a homomorphism $W:\mathfrak{g}\to\mathrm{End}\,E$, $X\to W(X)$ of the vector space $(\mathfrak{g},\otimes,\odot)$ into the vector space of the endomorphisms of E such that for all X,Y, $\in\mathfrak{g}:W([X,Y])=W(X)W(Y)-W(Y)W(X)$.

A. III: Translational symmetry and the vacuum: There exists a representation T of the vector group of $\mathbb{R}_4$ on a dense linear subspace $D_T\supseteq D$ which leaves D invariant and has the following properties.

(a) (η isometry): For all $y\in\mathbb{R}_4$; $\Psi,\Phi,\in D_T$

$$\langle T(y)\Psi,T(y)\Phi\rangle=\langle\Psi,\Phi\rangle. \tag{2.3}$$

(b) (Vacuum): There exists a unique state $\Psi_0\in H$ (called the vacuum) such that $\langle\Psi_0,\Psi_0\rangle=1$ and for all $y\in\mathbb{R}_4:T(y)\Psi_0=\Psi_0$.

(c) (Covariance): There is a decomposition of T into a countable union $\mathbf{T}=\cup_{\mathscr{T}\in I_T}\mathbf{T}_T(\mathscr{T})\times\{\mathscr{T}\}$ of pairwise disjoint finite subsets $\mathbf{T}_T(\mathscr{T})\times\{\mathscr{T}\}$ such that for all $y\in\mathbb{R}_4$; $\mathscr{T}_i\in I_T$, $\mu_i\in\mathbf{T}_T(\mathscr{T}_i)$ $(i=1,\ldots,L)$, $L\in\mathbb{N}$, and $\Psi\in D$ the substitution rule holds:

$$\eta T(y)\prod_{i=1}^{L}\varphi_{\mu_i,\mathscr{T}_i}(f_i)T(y)^{-1}\Psi$$
$$=\eta\sum_{\mu_1\in\mathbf{T}_T(\mathscr{T}_1)}\cdots\sum_{\mu_L\in\mathbf{T}_T(\mathscr{T}_L)}\prod_{i=1}^{L}\varphi_{\nu_i,\mathscr{T}_i}(T^{\mathscr{T}_i}((y\cdot\partial_{x_i}))^{\nu_i}{}_{\mu_i}f_{i,y})\Psi, \tag{2.4}$$

with $f_{i,y}(x):=f_i(x-y)$ and $(y\cdot\partial_x):=\Sigma_{\rho=0}^{3}y^\rho(\partial/\partial x^\rho)$.

Here $T^{\mathscr{T}}((y\cdot\partial_x))$ is a $\mathring{\mathscr{T}}=|\mathbf{T}_T(\mathscr{T})|$-dimensional indecomposable (matrix) representation of $(\mathbb{R}_4,+)$ on $S(\mathbb{R}_4;\mathbb{C}_{\mathring{\mathscr{T}}})$ and continuous in y. Moreover, it is assumed $\mathring{\mathscr{T}}^*=\mathring{\mathscr{T}}$.

Since in the covariance condition we have admitted "unusual" higher-dimensional representations, we have to make some comments.

(i) The decomposition of $\mathbf{T}$ just means a splitting of the multi-indices $\Gamma=(\mu,\mathscr{T})$ into a (multi-) index $\mathscr{T}$ which characterizes besides other properties of $\varphi_\Gamma=\varphi_{\mu,\mathscr{T}}$ a special representation to which φ_Γ belongs and an index $\mu\in\mathbf{T}_T(\mathscr{T})$ counting the $\mathring{\mathscr{T}}=|\mathbf{T}_T(\mathscr{T})|$ components which belong to the representation $\mathscr{T}$.

(ii) $T(y)$ is neither assumed to be unitary nor even a bounded operator on D_T. In theories in which Ψ_0 is not an eigenvector of η the representation $T(y)$ $(y\neq0)$ cannot be unitary. For if it would be unitary, than (a) and (b) imply $T(y)\eta\Psi_0=\eta\Psi_0$ and therefore $\eta\Psi_0=d\Psi_0$ with $d\in\mathbb{R}^+$.

(iii) From Theorems 3.1, 3.2, and 7.1 below it follows that the continuity of $T^{\mathscr{T}}((y\cdot\partial_x))$ in y implies the continuity of the matrix elements $\langle\Phi,T(y)\Psi\rangle$ in y for all quasilocal states Ψ,Φ.

(iv) In classical Wightman theories ($\eta=\mathrm{id}_{\mathscr{H}}$, identity operator) one admits only one-dimensional representations [$\mathring{\mathscr{T}}=1$ and $T^{\mathscr{T}}((y\cdot\partial_x))=1$]. In case Ψ_0 is an eigenvector of the metric operator η this would imply here that $T(y)$ has a unique extension to $\mathscr{H}$ (see Theorem 4.2 below). On the other hand it has been argued in Ref. 3 that charge confinement may be closely connected with the unboundedness of the translation operators $T(y)$. In spite of the fact that we present in Sec. VIII a confinement mechanism which does not necessarily require higher-dimensional representations $T^{\mathscr{T}}((y\cdot\partial_x))$ with $\mathring{\mathscr{T}}>1$, the latter can *a priori* not be excluded. If $q=q(\mathscr{T})$ denotes an arbitrary element from $\mathbb{C}_4$ then any indecomposable representation T *on* $S(\mathbb{R}_4,\mathbb{C}_{\mathring{\mathscr{T}}})$ of dimension $m=\mathring{\mathscr{T}}$ and continuous in y is equivalent to the (operator) matrix representation (Ref. 22, Chap. V, Sec. 9)

$$t_q^m(y\cdot\partial_x)=\exp(y\cdot q(\mathscr{T}))\begin{pmatrix}1 & (y\cdot\partial_x) & 1/2(y\cdot\partial_x)^2 & \cdots & 1/(m-2)!(y\cdot\partial_x)^{m-2} & 1/(m-1)!(y\cdot\partial_x)^{m-1}\\ 0 & 1 & (y\cdot\partial_x) & \cdots & 1/(m-3)!(y\cdot\partial_x)^{m-3} & 1/(m-2)!(y\cdot\partial_x)^{m-2}\\ \vdots & \vdots & \vdots & & \vdots & \vdots\\ 0 & 0 & 0 & & 1 & (y\cdot\partial_x)\\ 0 & 0 & 0 & & 0 & 1\end{pmatrix}. \tag{2.5}$$

If in the case of Jaffe spaces $S^{g}(\mathbb{R}_4,\mathbb{C}_m)$ the real part of $q(\mathscr{T})$ is unequal to zero for some $\mathscr{T}$ then the corresponding fields $\varphi_{\mu,\mathscr{T}}(f)$ together with all polynomials of them can create only states of zero η norm from the vacuum Ψ_0

$$\forall L\in\mathbb{N},\quad\forall\mu_i,\ldots,\mu_L\in\mathbf{T}_T(\mathscr{T}),\quad\forall f_1,\ldots,f_L\in S^{g}(\mathbb{R}_4,\mathbb{C}_m),$$
$$\prod_{i=1}^{L}\varphi_{\mu_i,\mathscr{T}}(f_i)\Psi_0\in\mathscr{H}_0\supset H_0.$$

Moreover for a *mixed* product of two (or more) such fields $\varphi_{\mu,\mathscr{T}_1}(f)$ and $\varphi_{\mu,\mathscr{T}_2}(h)$ this need not be true. By a suitable choice of $q(\mathscr{T}_1)$ and $q(\mathscr{T}_2)$ the mixed product of them can create states in $\mathscr{H}\setminus\mathscr{H}_0$. This offers an attractive explanation of the confinement mechanism for quark fields, etc., in spite of the fact that there seems to be some vinegar in this wine; the fields $\varphi_{\mu,\mathscr{T}}$ can be Lorentz but not Poincaré covariant. The latter requires $q(\mathscr{T})=0$. However, since they are unobservable, they do not need to be Poincaré covariant themselves but only those *mixed* products (of different ones) of them which create states in $\mathscr{H}\setminus\mathscr{H}_0$ from the vacuum. The important point is that the latter products are Poincaré covariant if their constituents are covariant under Lorentz transformations and translations separately. The detailed derivation of these results will be given in Sec. VIII.

A.IV: Completeness: The vacuum Ψ_0 is cyclic with respect to the polynomial * algebra $\mathscr{P}(\varphi)$ over $\mathbb{C}$ with basis $\{\varphi_\Gamma(f)|\Gamma\in\mathbf{T},f\in S(\mathbb{R}_4,\mathbb{C})\}$ and $\mathscr{P}(\varphi)\Psi_0\cap H$ dense in H.

The completeness assumption means that the linear subspace (LH $\sim$linear hull)

$$\mathscr{D}_{\Pi} := \mathscr{P}(\varphi)\Psi_0$$
$$= \mathrm{LH}\left\{ \prod_{i=1}^{n} \varphi_{\Gamma_i}(f_i)\Psi_0 \middle| f_i \in S(\mathbf{R}_4,\mathbf{C});\Gamma_i \in \mathbf{T}, n \in \mathbf{N} \right\} \tag{2.6}$$

in dense in $\mathscr{H}$. As a consequence $\mathscr{H}$ is separable, since $S(\mathbf{R}_4,\mathbf{C})$ is separable and $\mathbf{T}$ countable. Since the algebraic tensor product $\otimes^L S(\mathbf{R}_4,\mathbf{C}) = S(\mathbf{R}_4,\mathbf{C}) \otimes \cdots \otimes S(\mathbf{R}_4,\mathbf{C})$ (L times) is dense in $S(\mathbf{R}_{4L},\mathbf{C})$ we may use the nuclear theorem, the completeness of $S(\mathbf{R}_{4L},\mathbf{C})$, and the continuity of the field operators to define on D the more general operators $\left[\overline{\otimes}_{i=1}^{L}\right]\varphi_{\Gamma_i}(f)$, $\Gamma_i \in \mathbf{T}$, $f \in S(\mathbf{R}_{4L},\mathbf{C})$. They are obtained in exactly the same as in the Wightman theories (Ref. 23, Chap. III, Sec. 1) by means of the strong limits in $\mathscr{H}$

$$\left[\overline{\otimes}_{i=1}^{L}\right]\varphi_{\Gamma_i}(f)\Phi := \operatorname*{s-lim}_{n\to\infty} \sum_{j=1}^{n} \prod_{i=1}^{L} \varphi_{\Gamma_i}(f_i^j)\Phi, \quad \Phi \in D, \tag{2.7}$$

for an arbitrary sequence $\left(\sum_{j=1}^{n} \otimes_{i=1}^{L} f_i^j\right)_n \mathbf{N}$ converging in $S(\mathbf{R}_{4L},\mathbf{C})$ of f. The dense linear subspace defined by the linear hull

$$\mathscr{D}_{\mathrm{QL}} = \mathrm{LH}\Big\{\left[\overline{\otimes}_{i=1}^{L}\varphi_{\Gamma_i}\right](f)\Psi_0 \Big| f \in S(\mathbf{R}_{4L},\mathbf{C});$$
$$\Gamma_1,\ldots,\Gamma_L \in \mathbf{T};\ L \in \mathbf{N}\Big\} \tag{2.8}$$

is called the set of quasilocal states. It obviously belongs to the domain of every (basic) field operator $\varphi_\Gamma(f)$, which means $\mathscr{D}_{\mathrm{QL}} \subseteq D$. Moreover we assume $\mathscr{D}_\Pi$ and $\mathscr{D}_{\mathrm{QL}}$ to be invariant under the metric operator η.

A. V.: η stability:

$$\eta\mathscr{D}_\Pi \subseteq \mathscr{D}_\Pi \wedge \eta\mathscr{D}_{\mathrm{QL}} \subseteq \mathscr{D}_{\mathrm{QL}}. \tag{2.9}$$

For the sake of completeness we only mention the two remaining axioms of a gauge quantum field theory without spelling them out in detail,[2,5] because we do not use them.

A. VI: Spectrum condition.

A. VII: Locality (Einstein causality).

These two additional axioms of course would reduce the class of admissible test function spaces to the strictly localizable ones of Jaffe.[19]

Finally, in analogy to the translational symmetry in axiom A.III we define a global symmetry or a symmetry group of a gauge quantum field theory by the following.

Definition 2.1 (global symmetry): A Hausdorff topological group $\mathscr{G}$ is called a symmetry group, if there exists a representation V of $\mathscr{G}$ on D_G with $\mathscr{D}_{\mathrm{QL}} \subseteq D \subseteq D_G$, $V(g)\mathscr{D}_{\mathrm{QL}} \subseteq \mathscr{D}_{\mathrm{QL}}$, and the further following properties.

(a) (Invariance of the vacuum): $\forall g \in G,\quad V(g)\Psi_0 = \Psi_0$.

(b) (η isometry): $\forall g \in G$, $\forall \Psi,\Phi \in D_G$, $\langle V(g)\Psi, V(g)\Phi\rangle = \langle\Psi,\Phi\rangle$.

(c) (Covariance): There exists a decomposition of $\mathbf{T}$ into a countable union $\mathbf{T} = \cup_{\mathscr{A} \in I_G} \mathbf{T}_G(\mathscr{A}) \times \{\mathscr{A}\}$ of pairwise disjoint finite subsets $\mathbf{T}_G(\mathscr{A}) \times \{\mathscr{A}\}$ with $\mathring{\mathscr{A}} := |\mathbf{T}_G(\mathscr{A})| = \mathring{\mathscr{A}}^* = |\mathbf{T}_G(\mathscr{A}^*)|$ and for every $\mathscr{A} \in I_G$ a continuous representation $R^{\mathscr{A}}$ of G on $S(\mathbf{R}_4,\mathbf{C}_{\mathring{\mathscr{A}}})$ such that for all $g \in G$, $\mathscr{A}_i \in I_G$, $\mu_i \in \Pi_G(\mathscr{A}_i)$, and $\Psi \in \mathscr{D}_{\mathrm{QL}}$ the substitution rule holds:

$$\eta V(g) \prod_{i=1}^{L} \varphi_{\mu_i,\mathscr{A}_i}(f_i) V(g)^{-1}\Psi$$
$$= \eta \sum_{\nu_1 \in \mathbf{T}_G(\mathscr{A}_1)} \cdots \sum_{\nu_L \in \mathbf{T}_G(\mathscr{A}_L)} \prod_{j=1}^{L} \varphi_{\nu_j,\mathscr{A}_j}(R^{\mathscr{A}_j}(g)^{\nu_j}{}_{\mu_j} f_j)\Psi. \tag{2.10}$$

(G,V) is called a *strict* global symmetry if $V(g)(D_\Pi \cap H) \subseteq (D_\Pi \cap H)$.

We have to add here a series of comments.

(i) By a continuous representation $R^{\mathscr{A}}$ on $S(\mathbf{R}_4;\mathbf{C}_{\mathring{\mathscr{A}}})$ we mean a homomorphism into the group of topological automorphisms of $S(\mathbf{R}_4;\mathbf{C}_{\mathring{\mathscr{A}}})$ such that for every fixed $f \in S(\mathbf{R}_4;\mathbf{C}_{\mathring{\mathscr{A}}})$ the map $G \to S(\mathbf{R}_4,\mathbf{C}_{\mathring{\mathscr{A}}})$, $g \to R^{\mathscr{A}}(g)f$ is continuous. If G is locally compact and countable at infinity then this implies that the map $G \times S(\mathbf{R}_4,\mathbf{C}_{\mathring{\mathscr{A}}}) \mapsto S(\mathbf{R}_4,\mathbf{C}_{\mathring{\mathscr{A}}})$, $(g,f) \mapsto R^{\mathscr{A}}(g)f$ is continuous (Ref. 9, Chap. 4.1).

(ii) By means of the principle of uniform boundedness the continuity of $R^{\mathscr{A}}$ implies that $R^{\mathscr{A}}$ is uniformly bounded on some neighborhood $\mathscr{N}(e)$ of the unit element e of G. This means for every $p \in \mathbf{N}^0$ there exist $p' \in \mathbf{N}^0$ and $r(\mathscr{A},p) \in \mathbf{R}^+$ such that for all $f \in S(\mathbf{R}_4,\mathbf{C}_{\mathring{\mathscr{A}}})$ and $g \in \mathscr{N}(e)$ we have

$$\|R^{\mathscr{A}}(g)f\|_p < r(\mathscr{A},p)\|f\|_{p'}. \tag{2.11}$$

According to Theorems 3.1 and 3.2 below this together with the continuity of $R^{\mathscr{A}}(g)$ in g implies the continuity of the matrix elements $\langle\Psi, V(g)\Phi\rangle$ in g for all $\Psi \in \mathscr{H}$ and $\Phi \in \mathscr{D}_{\mathrm{QL}}$.

(iii) In Sec. VII it will be shown that for the proper orthochronous Poincaré group, all representations

$$(R^{\mathscr{A}}(a,a)f)^\mu(x)$$
$$:= \sum_{\rho,\nu \in \mathbf{T}_{P_+^1}(\mathscr{A})} T_q((a\cdot\partial_x))^\mu{}_\rho M^{\mathscr{A}}(\alpha)^\rho{}_\nu$$
$$\times f^\nu(\Lambda(\alpha^{-1})(x-a)) \quad (q=0), \tag{2.12}$$

where $T_{\mathring{0}}^{\mathscr{A}}$ is equivalent to a direct sum of the matrices (2.5) and $M^{\mathscr{A}}$ is an $\mathring{\mathscr{A}}$-dimensional matrix representation of SL(2,C), possess all the continuity and boundedness properties of Definition 2.1 on every Jaffe space $S^{g}(\mathbf{R}_4;\mathbf{C}_{\mathring{\mathscr{A}}})$.

(iv) In analogy with the discussion of the Poincaré symmetry in Ref. 5, pp. 137–142 and 151, the general form of the substitution rule (2.10) (inclusive of the factor η) is "dictated by physics." However, when compared with the substitution rules of a classical Wightman theory there is a characteristic difference which is closely related to the nonunitarity of V. Exactly as in the case of a Wightman theory the unitarity of V would enforce a close connection between the substitution rules of a field $\varphi_{\mu,\mathscr{A}}$ and its adjoint $\varphi_{\mu,\mathscr{A}^*}$. Essentially (see Theorem 4.2) it reads $R^{\mathscr{A}^*}(g)f = \overline{R^{\mathscr{A}}(g)\overline{f}}$. For the Lorentz group this means that the adjoint field transforms with the complex conjugate of that matrix according to which the field itself transforms. In a gauge quantum field theory this connection in general breaks down, which in turn causes the nonunitarity and even the unboundedness of V. For example, in the Gupta–Bleuler formulation of the free electromagnetic field the substitution rules for the vector potentials $\varphi_{\mu,\mathscr{A}}$ and their adjoints $\varphi_{\mu,\mathscr{A}^*}$ under Lorentz transformations read (Ref. 5, Part III)

$$V(\Lambda)\varphi_{\mu,\mathscr{A}}(f)V(\Lambda^{-1})\Psi = \sum_{\nu=0}^{3} \varphi_{\nu,\mathscr{A}}(\Lambda^{\nu}{}_{\mu} f_{\Lambda})\Psi,$$
$$V(\Lambda)\varphi_{\mu,\mathscr{A}^*}(f)V(\Lambda^{-1})\Psi = \sum_{\nu=0}^{3} \varphi_{\nu,\mathscr{A}^*}((G\Lambda G^{-1})^{\nu}{}_{\mu} f_{\Lambda})\Psi, \tag{2.13}$$

with $f_{\Lambda}(x) = f(\Lambda^{-1}x)$ and $G = (g_{\lambda\rho})$ the Minkowski metric tensor. Here, $G\Lambda G^{-1}$ is the transposed inverse of Λ and not the complex conjugate. On the other hand, the η isometry of V connects the substitution rule of $\varphi_{\mu,\mathscr{A}}$ under $V(g)$ with that of the adjoint field $\varphi_{\mu,\mathscr{A}^*}$ under the adjoint and inverse representation operator $V(g^{-1})^*$ in the same way as stated before in the case of unitarity. If in addition η has an inverse this leads to the same connection as above between the substitution rules under $V(g)$ for $\varphi_{\mu,\mathscr{A}}$ and the "η-adjoint" operator $(\eta^{-1}\varphi_{\mu,\mathscr{A}^*}\eta)$ but not $\varphi_{\mu,\mathscr{A}^*}$ itself.

For the remainder it will be of considerable advantage to introduce instead of the one-component fields $\varphi_{\Gamma}(f)$, $\Gamma \in \mathrm{T}$ the multicomponent Wightman–Gårding fields with respect to the symmetry group G defined by[5]

$$\phi^{\mathscr{A}}(f) := \sum_{\mu \in \mathrm{T}_G(\mathscr{A})} \varphi_{\mu,\mathscr{A}}(f^{\mu}), \quad f \in S(\mathbb{R}_4, \mathbb{C}_{\dot{\mathscr{A}}}). \tag{2.14}$$

More general, let $S(\mathscr{A}_1,...,\mathscr{A}_L)$ denote the L-fold completed tensor product of the nuclear Frechét spaces $S(\mathbb{R}_4, \mathbb{C}_{\dot{\mathscr{A}}_i})$ (Ref. 21, III, Chap. 50 ff; Ref. 24, Secs. 41–44; and Ref. 9, Appendix 2)

$$S(\mathscr{A}_1,...,\mathscr{A}_L) = \bar{\otimes}_{i=1}^{L} S(\mathbb{R}_4, \mathbb{C}_{\dot{\mathscr{A}}_i}) = S(\mathbb{R}_{4L}, \bar{\otimes}_{i=1}^{L} \mathbb{C}_{\dot{\mathscr{A}}_i}). \tag{2.15}$$

Then we define for all $f \in S(\mathscr{A}_1,...,\mathscr{A}_L)$

$$\phi^{\mathscr{A}_1,...,\mathscr{A}_L}(f) := \sum_{\mu_1 \in \mathrm{T}_G(\mathscr{A}_1)} \cdots \sum_{\mu_L \in \mathrm{T}_G(\mathscr{A}_L)} \left[\bar{\otimes}_{i=1}^{L} \varphi_{\mu_i,\mathscr{A}_i}\right](f^{\mu_1,...,\mu_L}). \tag{2.16}$$

By virtue of the equality (2.15) 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 (S.II) and so on. Especially, it follows from Eq. (2.7) for all $\Psi \in \mathscr{H}, \Phi \in D$ and any sequence $(\Sigma_{j=1}^{n} \otimes_{i=1}^{L} f_i^{(j)})_{n \in \mathbb{N}}$ converging in $S(\mathscr{A}_1,...,\mathscr{A}_L)$ to f,

$$\begin{aligned}(\Psi,\phi^{\mathscr{A}_1,...,\mathscr{A}_L}(f)\Phi) &= \lim_{n\to\infty}\left(\Psi,\phi^{\mathscr{A}_1,...,\mathscr{A}_L}\left(\sum_{j=1}^{n} \bar{\otimes}_{i=1}^{L} f_i^{(j)}\right)\Phi\right) \\ &= \lim_{n\to\infty}\left(\Psi, \sum_{j=1}^{n} \prod_{i=1}^{L} \phi^{\mathscr{A}_i}(f_i^{(j)})\phi\right).\end{aligned} \tag{2.17}$$

The equivalence of the sets $\mathrm{T}_G(\mathscr{A})$ and $\mathrm{T}_G(\mathscr{A}^*)$ $(\mathring{\mathscr{A}} = \mathring{\mathscr{A}}^*)$ implies $\phi^{\mathscr{A}}(f)^* = \phi^{\mathscr{A}^*}(\bar{f})$. In order to generalize this relation to the product fields (2.16) we introduce the antilinear and isometric (with respect to all norms $\|\cdots\|_p, p \in \mathbb{N}$) bijections $\mathscr{C}'_L = \mathscr{C}'_L(\mathscr{A}_1,...,\mathscr{A}_L)$, $L \in \mathbb{N}$ by

$$\mathscr{C}'_L : \otimes_{i=1}^{L} S(\mathbb{R}_4, \mathbb{C}_{\dot{\mathscr{A}}_i}) \mapsto \otimes_{i=L}^{1} S(\mathbb{R}_4, \mathbb{C}_{\dot{\mathscr{A}}_i}),$$
$$\otimes_{i=1}^{L} f_i(x_i) \mapsto \otimes_{i=L}^{1} \overline{f_i(x_i)}. \tag{2.18}$$

Note that $\mathscr{C}'_1 = \mathscr{C}'^{-1}_1$ acts simply as complex conjugation of the components of f. If $\mathscr{C}_L$ denotes the unique isometric extension of $\mathscr{C}'_L$ onto the completed tensor product $S(\mathscr{A}_1,...,\mathscr{A}_L)$, then it follows by means of (2.17)

$$\phi^{\mathscr{A}_1,...,\mathscr{A}_L}(f)^* = \phi^{\mathscr{A}_L^*,...,\mathscr{A}_1^*}(\mathscr{C}_L f). \tag{2.19}$$

Finally the linear subspaces $\mathscr{D}_{\Pi}$, respectively, $\mathscr{D}_{\mathrm{QL}}$ read in terms of the Wightman–Gårding fields

$$\begin{aligned}\mathscr{D}_{\Pi} &= \mathrm{LH}\left\{\prod_{i=1}^{L} \phi^{\mathscr{A}_i}(f_i)\Psi_0 \middle| f_i \in S(\mathbb{R}_4, \mathbb{C}_{\dot{\mathscr{A}}_i}); \; \mathscr{A}_1,...,\mathscr{A}_L \in I_G;\; L \in \mathbb{N}\right\}, \\ \mathscr{D}_{\mathrm{QL}} &= \mathrm{LH}\left\{\phi^{\mathscr{A}_1,...,\mathscr{A}_L}(f)\Psi_0 \middle| f \in S(\mathscr{A}_1,...,\mathscr{A}_L); \; \mathscr{A}_1,...,\mathscr{A}_L \in I_G;\; L \in \mathbb{N}\right\}.\end{aligned} \tag{2.20}$$

III. CONTINUITY IN THE GROUP ELEMENTS

Our Definition 2.1 of a symmetry group G contains no explicit assumption about continuity properties of $V(g)\Psi$ in its dependence on g for fixed $\Psi \in D_G$. The aim of this section is to construct in three steps a representation U of G on $\mathscr{D}_{\mathrm{QL}}$ which is strongly continuous in the group elements g and in case V is a strict global symmetry physically equivalent to V. The latter means that both U and V generate on $\mathscr{H}_{\mathrm{ph}}$ one and the same unitary representation of G. This continuity is necessary for the differentiation in Sec. V. Furthermore, for the remaining considerations we have to get rid of the factor η in the substitution rule at least on $\mathscr{D}_{\mathrm{QL}}$.

Theorem 3.1: If G is a symmetry group, then for all $\mathscr{A}_1,...,\mathscr{A}_L \in I_G$, $L \in \mathbb{N}$ the extended tensor product $R^{\mathscr{A}_1,...,\mathscr{A}_L}(g) = \bar{\otimes}_{i=1}^{L} R^{\mathscr{A}_i}(g)$ is a continuous representation of G on $S(\mathscr{A}_1,...,\mathscr{A}_L)$.

Proof: If G is locally compact and countable at infinity, Theorem 3.1 follows directly from Proposition 4.1.2.4 in Ref. 9. In order to cover also the gauge groups of the second kind we do not assume this here and in the following section. However, it is well known (Ref. 24, Secs. 41, 5) that the tensor product $\otimes_{i=1}^{L} R^{\mathscr{A}_i}(g)$ of the topological automorphisms $R^{\mathscr{A}_i}(g)$ has a unique extension to a topological automorphism $R^{\mathscr{A}_1,...,\mathscr{A}_L}(g) = \bar{\otimes}_{i=1}^{L} R^{\mathscr{A}_i}(g)$ of $S(\mathscr{A}_1,...,\mathscr{A}_L)$. It plainly satisfies the group property. It remains to prove the continuity in g. Since the topology of $S(\mathscr{A}_1,...,\mathscr{A}_L)$ is equivalent to the projective one, the uniform boundedness of $R^{\mathscr{A}}(g)$ on $\mathscr{N}(e)$ and Propositions 43.1 and 43.2 in Ref. 21 imply the following inequalities. For every

$$h = \sum_j \otimes_{i=1}^{L} h_i^j \in \otimes_{i=1}^{L} S(\mathbb{R}_4, \mathbb{C}_{\dot{\mathscr{A}}_i}),$$

every $g \in \mathscr{N}(e)$, every $p \in \mathbb{N}^0$, and some $p' = p'(p) > p_2 > p_1 > p$,

$$\begin{aligned}\left\|\otimes_{i=1}^{L} R^{\mathscr{A}_i}(g)h\right\|_p &< C_1 \inf\left\{\sum_j \prod_{i=1}^{L} \|R^{\mathscr{A}_i}(g)\tilde{h}_i^j\|_{p_1} \middle| \sum_j \bar{\otimes}_{i=1}^{L} \tilde{h}_i^j = h\right\} \\ &< C_1 \rho_1(\mathscr{A}_1,...,\mathscr{A}_L; p_1) \\ &\quad \times \inf\left\{\sum_j \prod_{i=1}^{L} \|\tilde{h}_i^j\|_{p_2} \middle| \sum_j \otimes_{i=1}^{L} \tilde{h}_i^j = h\right\} \\ &< C_1 C_2 \rho_1(\mathscr{A}_1,...,\mathscr{A}_L; p')\|h\|_{p'}.\end{aligned} \tag{3.1}$$

Now let f be from $S(\mathscr{A}_1,...,\mathscr{A}_L)$ and $(h^{(n)})_{n\in\mathbb{N}}$ a sequence with elements from $\otimes_{i=1}^{L} S(\mathbb{R}_4,\mathbb{C}_{\dot{\mathscr{A}}_i})$ converging to f; then (3.1) implies that for every $p\in\mathbb{N}^0$ there exists a $p'\in\mathbb{N}^0$ such that for all $n\in\mathbb{N}$ and $g\in\mathscr{N}(e)$ we have

$$\begin{aligned}\|R^{\mathscr{A}_1,...,\mathscr{A}_L}(g)f\|_p &\leqslant \rho(\mathscr{A}_1,...,\mathscr{A}_L;p)\|h^{(n)}\|_{p'} \\ &\quad + \|R^{\mathscr{A}_1,...,\mathscr{A}_L}(g)f - \otimes_{i=1}^{L} R^{\mathscr{A}_i}(g)h^{(n)}\|_p \\ &\leqslant \rho(\mathscr{A}_1,...,\mathscr{A}_L;p)(\|f\|_{p'} + \|f-h^{(n)}\|_{p'}) \\ &\quad + \|R^{\mathscr{A}_1,...,\mathscr{A}_L}(g)(f-h^{(n)})\|_p. \end{aligned} \tag{3.2}$$

Taking the limit $n\to\infty$ and observing the continuity of $R^{\mathscr{A}_1,...,\mathscr{A}_L}(g)$ as an operator on $S(\mathscr{A}_1,...,\mathscr{A}_L)$ we obtain for all $p\in\mathbb{N}^0$, $g\in\mathscr{N}(e)$, $f\in S(\mathscr{A}_1,...,\mathscr{A}_L)$, and some $p'(p)\in\mathbb{N}^0$:

$$\|R^{\mathscr{A}_1,...,\mathscr{A}_L}(g)f\|_p \leqslant \rho(\mathscr{A}_1,...,\mathscr{A}_L;p)\|f\|_{p'}. \tag{3.3}$$

Now we use this uniform boundedness of $R^{\mathscr{A}_1,...,\mathscr{A}_L}$ on the neighborhood $\mathscr{N}(e)$ of the unit element to derive the strong continuity of $R^{\mathscr{A}_1,...,\mathscr{A}_L}(g)$ in g. We restrict ourselves to the case $L=2$, since the general case then easily follows via complete induction from the equation $S(\mathscr{A}_1,...,\mathscr{A}_L) = S(\mathscr{A}_1,...,\mathscr{A}_{L-1})\bar{\otimes} S(\mathbb{R}_4,\mathbb{C}_{\dot{\mathscr{A}}_L})$ by a literal repetition of the arguments below. We first observe [Ref. 24, Sec. 41.4(6)] that for every $f\in S(\mathscr{A}_1,\mathscr{A}_2)$ there exist null sequences $(h_i^n)_{n\in\mathbb{N}}$ in $S(\mathbb{R}_4,\mathbb{C}_{\dot{\mathscr{A}}_i})$, $i=1,2$ and in addition a sequence $(\lambda_n)_{n\in\mathbb{N}}$ of complex numbers with $\Sigma_{n=1}^{\infty}|\lambda_n|\leqslant 1$ such that

$$f = \lim_{n\to\infty}\sum_{j=1}^{n}\lambda_j h_1^j\otimes h_2^j = \lim_{n\to\infty} f^{(n)}. \tag{3.4}$$

Let $(g_\alpha)_{\alpha\in I(<)}$ be an arbitrary net converging to the unit element of G. Then for every $\epsilon>0$ and $p\in\mathbb{N}^0$ there exist a natural number $N(\epsilon,p)$ and a $\beta(\epsilon)\in I(<)$ such that for all $\alpha\in I(<)$ with $\beta<\alpha$ and all $n\geqslant N(\epsilon,p)$ we have $g_\alpha\in\mathscr{N}(e)$, respectively,

$$\left|\left| f - \sum_{j=1}^{n}\lambda_j h_1^j\otimes h_2^j\right|\right|_p < \epsilon. \tag{3.5}$$

This together with the inequality (3.3) implies that for any $p\in\mathbb{N}^0$ there exists a $p'\in\mathbb{N}^0$ such that for all $\beta(\epsilon)<\alpha$ and $n\geqslant\bar{N}(\epsilon,p) := \max\{N(\beta,p),N(\beta,p')\}$ we have

$$\begin{aligned}\|R^{\mathscr{A}_1\mathscr{A}_2}(g_\alpha)f - f\|_p &\leqslant \|f-f^{(n)}\|_p + \|R^{\mathscr{A}_1\mathscr{A}_2}(g_\alpha)(f-f^{(n)})\|_p \\ &\quad + \|(R^{\mathscr{A}_1}(g_\alpha)\otimes R^{\mathscr{A}_2}(g))f^{(n)} - f^{(n)}\|_p \\ &< \epsilon\cdot(1+\rho(\mathscr{A}_1,\mathscr{A}_2;p)) + \sum_{j=1}^{n}|\lambda_j| \\ &\quad \times\|R^{\mathscr{A}_1}(g_\alpha)h_1^j\otimes R^{\mathscr{A}_2}(g_\alpha)h_2^j - h_1^j\otimes h_2^j\|_p. \end{aligned} \tag{3.6}$$

Each term in the last sum can be rewritten in the form

$$\begin{aligned}\|R^{\mathscr{A}_1}(g_\alpha)h_1^j\otimes R^{\mathscr{A}_2}(g_\alpha)h_2^j - h_2^j\otimes h_2^j\|_p &= \|R^{\mathscr{A}_1}(g_\alpha)h_1^j - h_1^j\|_p\|h_2^j\|_p \\ &\quad + \|R^{\mathscr{A}_1}(g_\alpha)h_1^j\|_p\,\|R^{\mathscr{A}_2}(g_\alpha)h_2^j - h_2^j\|_p. \end{aligned} \tag{3.7}$$

Since $R^{\mathscr{A}_i}(g)$ is strongly continuous in g it follows at once from the last two relations for every $\epsilon>0$ and $f\in S(\mathscr{A}_1,\mathscr{A}_2)$

$$\lim_{\alpha}\|R^{\mathscr{A}_1\mathscr{A}_2}(g_\alpha)f - f\|_p < \epsilon\cdot(1+\rho(\mathscr{A}_1,...,\mathscr{A}_2;p)). \tag{3.8}$$

This proves the continuity in $g\in G$. □

Theorem 3.2: If G is a symmetry group, then there exists a representation U of G *on* $\mathscr{D}_{\rm QL}$ with the following properties.

(1) For all $g\in G$; $\mathscr{A}_1,...,\mathscr{A}_L\in I_G$; $L\in\mathbb{N}$; and $\Psi\in\mathscr{D}_{\rm QL}$: $U(g)\Psi_0=\Psi_0$,

$U(g)\phi^{\mathscr{A}_1,...,\mathscr{A}_L}(f)U(g)^{-1}\Psi = \phi^{\mathscr{A}_1,...,\mathscr{A}_L}(R^{\mathscr{A}_1,...,\mathscr{A}_L}(g)f)\Psi.$

(2) $U(g)$ is strongly continuous in g; this means that for every fixed $\Psi\in\mathscr{D}_{\rm QL}$ the map $G\to H$, $g\to U(g)\Psi$ is strongly continuous.

(3) (η-coincidence). For all $g\in G$, $\Psi\in\mathscr{D}_{\rm QL}$: $\eta U(g)\Psi = \eta V(g)\Psi$.

(4) (η-isometry). For all $g\in G$, $\Psi\in\mathscr{D}_{\rm QL}$: $\langle U(g)\Psi,U(g)\Phi\rangle = \langle\Phi,\Psi\rangle$.

Proof: On $\mathscr{D}_{\rm QL}$ we define the linear operators $U(g)$ in the following way:

$$\begin{aligned} &U(g)\psi_0 := \Psi_0, \\ &U(g)\phi^{\mathscr{A}_1,...,\mathscr{A}_L}(f)\Psi_0 := \phi^{\mathscr{A}_1,...,\mathscr{A}_L}(R^{\mathscr{A}_1,...,\mathscr{A}_L}(g)f)\Psi_0, \end{aligned} \tag{3.9}$$

and on the remaining states by linear extension. Plainly U is a representation of G on $\mathscr{D}_{\rm QL}$ since all $R^{\mathscr{A}_1,...,\mathscr{A}_L}$ are representations of G on $S(\mathscr{A}_1,...,\mathscr{A}_L)$. The second part of statement (1) is a trivial consequence of the definition (3.9).

In order to prove statement (2) it suffices to prove the continuity in g on the states of the form $\phi^{\mathscr{A}_1,...,\mathscr{A}_L}(f)\Psi_0$. For them it follows from the definition (3.9) by means of Eq. (2.19)

$$\begin{aligned}&\|U(g)\phi^{\mathscr{A}_1,...,\mathscr{A}_L}(f)\Psi_0 - \phi^{\mathscr{A}_1,...,\mathscr{A}_L}(f)\Psi_0\|^2_{\mathscr{H}} \\ &\quad = (\Psi_0,\phi^{\mathscr{A}_L^\dagger,...,\mathscr{A}_1^\dagger}(\mathscr{C}_L[R^{\mathscr{A}_1,...,\mathscr{A}_L}(g)f-f]) \\ &\qquad \times\phi^{\mathscr{A}_1,...,\mathscr{A}_L}(R^{\mathscr{A}_1,...,\mathscr{A}_L}(g)f-f)\Psi_0) \\ &\quad = (\Psi_0,\phi^{\mathscr{A}_L^\dagger,...,\mathscr{A}_1^\dagger\mathscr{A}_1,...,\mathscr{A}_L}([\mathscr{C}_L(R^{\mathscr{A}_1,...,\mathscr{A}_L}(g)f-f)] \\ &\qquad \otimes[R^{\mathscr{A}_1,...,\mathscr{A}_L}(g)f-f])\Psi_0). \end{aligned} \tag{3.10}$$

Since the matrix element in the last row is a continuous linear functional on $S(\mathscr{A}_L^\dagger,...,\mathscr{A}_L)$, it is of finite order. Hence there exist a smallest number $p_0\in\mathbb{N}^0$ and a positive real number $K(\mathscr{A}_1,...,\mathscr{A}_L)$ independent of g such that

$$\begin{aligned}&\|U(g)\phi^{\mathscr{A}_1,...,\mathscr{A}_L}(f)\Psi_0 - \phi^{\mathscr{A}_1,...,\mathscr{A}_L}(f)\Psi_0\|^2_{\mathscr{H}} \\ &\quad \leqslant K(\mathscr{A}_1,...,\mathscr{A}_L)\|\mathscr{C}_L[R^{\mathscr{A}_1,...,\mathscr{A}_L}(g)f-f]\|_{p_0} \\ &\qquad \times\|R^{\mathscr{A}_1,...,\mathscr{A}_L}(g)f-f\|_{p_0}. \end{aligned} \tag{3.11}$$

This inequality, the isometry of $\mathscr{C}_L$, and Theorem 3.1 imply the continuity of $U(g)$ in g. From Definitions 2.1 (a) and 2.1 (c) and Eqs. (3.9), the η coincidence (3) follows trivially for all $\Psi\in\mathscr{D}_{\rm II}\subset\mathscr{D}_{\rm QL}$. Let f be from $S(\mathscr{A}_1,...,\mathscr{A}_L)$ and $(\Sigma_{j=1}^{n}\otimes_{i=1}^{L} h_i^{(j)})_{n\in\mathbb{N}}$ a sequence from $\otimes_{i=1}^{L} S(\mathbb{R}_4,\mathbb{C}_{\dot{\mathscr{A}}_i})$ converging to f; then we deduce by means of Theorem 3.1 and the η isometry of the original representation V for all $\Theta\in D_G\supseteq\mathscr{D}_{\rm QL}$

$$(\Theta,\eta V(g)\phi^{\mathscr{A}_1,\ldots,\mathscr{A}_L}(f)\Psi_0)$$
$$= \lim_{n\to\infty}\Big(V(g)^{-1}\Theta,\eta\phi^{\mathscr{A}_1,\ldots,\mathscr{A}_L}\Big(\sum_{j=1}^{n}\otimes_{i=1}^{L} h_i^{(j)}\Big)\Psi_0\Big)$$
$$= \lim_{n\to\infty}\Big(\Theta,\eta V(g)\sum_{j=1}^{n}\prod_{i=1}^{L}\phi^{\mathscr{A}_i}(h_i^{(j)})\Psi_0\Big)$$
$$= \lim_{n\to\infty}\Big(\Theta,\eta\phi^{\mathscr{A}_1,\ldots,\mathscr{A}_L}\Big(\sum_{j=1}^{n}\otimes_{i=1}^{L} R^{\mathscr{A}_i}(g)h_i^{(j)}\Big)\Psi_0\Big)$$
$$= (\Theta,\eta\phi^{\mathscr{A}_1,\ldots,\mathscr{A}_L}(R^{\mathscr{A}_1,\ldots,\mathscr{A}_L}(g)f)\Psi_0). \tag{3.12}$$

In view of Eq. (3.9) this proves (3) for all $\Psi\in\mathscr{D}_{\mathrm{QL}}$. Finally, the proof of the η isometry for U is simply an application of (4). Take $\Psi,\Phi\in\mathscr{D}_{\mathrm{QL}}$ and apply the η coincidence twice:

$$(U(g)\Psi,\eta U(g)\Phi)$$
$$=(U(g)\Psi,\eta V(g)\Phi)=(\eta U(g)\Psi,V(g)\Phi)$$
$$=(\eta V(g)\Psi,V(g)\Phi)=\langle\Psi,\Phi\rangle. \qquad \square$$

It remains to prove the physical equivalence of the representations V and U. At a first glance (but only at a first one) it seems to be a trivial consequence of the properties (3) and (4) of Theorem 3.2. Unfortunately, due to the unboundedness of both representations there are some hidden intrinsic problems concerning domains and their invariance which force our steps right into the boring details and require an additional assumption on η. On the other hand, the theorem below has some considerable interest of its own. In the following $\overline{B}$ denotes the completion of a linear subspace B in the norm of $\mathscr{H}$ and $\overleftrightarrow{A}$ the completion of a linear subspace A of a factor space, for instance H/H_0, in the natural scalar product $([\Psi],[\Phi])_H:=\langle\Psi,\Phi\rangle$.

Corollary 3.2: If the metric operator satisfies the conditions $\ker\eta\cap\mathscr{D}_\Pi\subseteq H_0$ and $V(g)(\mathscr{D}_\Pi\cap H)\subseteq\mathscr{D}_\Pi\cap H$, then $\mathscr{D}_{\mathrm{QL}}\cap H$ is an invariant subspace for the representation U of Theorem 3.2.

Proof: If Φ is from $\mathscr{D}_\Pi\subseteq H$, then $V(g)\Phi\in\mathscr{D}_\Pi\cap H$, $U(g)\Phi\in\mathscr{D}_\Pi$, and in view of the η-coincidence,

$$U(g)\Phi=V(g)\Phi+\Psi(g,\Phi),$$

with $\Psi(g,\Phi)\in\ker\eta\cap\mathscr{D}_\Pi$. Since $\ker\eta\cap\mathscr{D}_\Pi\subseteq H_0$, this means $U(g)(\mathscr{D}_\Pi\cap H)\subseteq\mathscr{D}_\Pi\cap H$.

By means of Theorem 3.2(1), the continuity of the representations $R^{\mathscr{A}_1,\ldots,\mathscr{A}_L}$, and the continuity of the matrix elements of the field operators it follows in the weak topology of $\mathscr{H}$ for any sequence $(f_n)_{n\in\mathrm{N}}$ converging in $S(\mathscr{A}_1,\ldots,\mathscr{A}_L)$ to f

$$\operatorname*{w-lim}_{n\to\infty} U(g)\phi^{\mathscr{A}_1,\ldots,\mathscr{A}_L}(f_n)\Psi_0=U(g)\phi^{\mathscr{A}_1,\ldots,\mathscr{A}_L}(f)\Psi_0. \tag{3.13}$$

By assumption A.IV $\mathscr{D}_\Pi\cap H$ is dense in $\mathscr{D}_{\mathrm{QL}}\cap H$. Hence by definition of $\mathscr{D}_{\mathrm{QL}}$ any $\Psi\in(\mathscr{D}_{\mathrm{QL}}\cap H)\setminus\{\Psi_0\}$ is the limit of a sequence $(\Psi_n)_{n\in\mathrm{N}}$ from $\mathscr{D}_\Pi\cap H$ of the form

$$\Psi_n=\sum_{N=1}^{L}\sum_{i_1,\ldots,i_N}\sum_{k=1}^{n}\prod_{j=1}^{N}\phi^{\mathscr{A}_{i_j}}(f_{i_j}^{(k)})\Psi_0, \tag{3.14}$$

with $(\sum_{k=1}^{n}\otimes_{j=1}^{N}f_{i_j}^{(k)})_{n\in\mathrm{N}}$ some Cauchy sequence in $S(\mathscr{A}_{i_1},\ldots,\mathscr{A}_{i_N})$.

From the last two equations we obtain

$$\operatorname*{w-lim}_{n\to\infty} U(g)\Psi_n=U(g)\Psi.$$

Since $U(g)\Psi_n\in\mathscr{D}_\Pi\cap H$ for all $n\in\mathrm{N}$ and H is weakly complete, it follows $U(g)\Psi\in\mathscr{D}_{\mathrm{QL}}\cap H$. $\square$

Theorem 3.3: Let $E^i,B^i=E^i\cap H$, respectively, $B_0^i=E^i\cap H_0$, $(i=1,2)$ be linear subspaces of $\mathscr{H}$ with $B^1\cap B^2$ dense in H, G a group, and $W_i:G\to\operatorname{Aut}E_i$, $g\to W_i(g)$ η-isometric representations of G on E^i. Furthermore, asume $W_i(g)B^i\subseteq B^i$ and η coincidence on $B^1\cap B^2$

$$\forall g\in G,\quad \forall\Psi\in B^1\cap B^2,$$
$$\eta W_1(g)\Psi=\eta W_2(g)\Psi.$$

Then we have (α) $\overleftrightarrow{B^i/B_0^i}$ is unitary equivalent to $\mathscr{H}_{\mathrm{ph}}=\overleftrightarrow{H/H_0}$; this means there exists an unitary mapping $\rho_i:\overleftrightarrow{B^i/B_0^i}\to\mathscr{H}_{\mathrm{ph}}(i=1,2)$; and (β) W_1 and W_2 generate one and the same unitary representation $\mathscr{W}:G\to\operatorname{Aut}\mathscr{H}_{\mathrm{ph}}$ of G on the physical Hilbert space.

Plainly, if we identify (E^1,W_1) with $(\mathscr{D}_{\mathrm{QL}},U)$ and (E^2,W_2) with $(\mathscr{D}_\Pi\cap H,V\restriction(\mathscr{D}_\Pi\cap H))$ we get the desired physical equivalence provided the condition of Corollary 3.2 holds for η, i.e., if (G,V) is a strict global symmetry.

Proof: (α) Let $[\Psi]:=\Psi+H_0\in H/H_0$ and $[\Phi]_i=\Phi+B_0^i\in B^i/B_0^i$ denote the equivalence classes generated by $\Psi\in H$, respectively, $\Phi\in B^i$. Then the ordered pair $\{B^i/B_0^i;([\cdot]_i,[\cdot]_i)_{B_i}:=\langle\cdot,\cdot\rangle\}$ is like $\{H/H_0;([\cdot],[\cdot])_H:=\langle\cdot,\cdot\rangle\}$ a pre-Hilbert space. It has been shown by Mintchev *et al.* (Ref. 25, Lemma 1 and Theorem 2) that there exists a linear, injective, and isometric mapping

$$\tilde\rho_i:B^i/B_0^i\to H/H_0,\quad [\Psi]_i\to\tilde\rho_i([\Psi]_i)=[\Psi], \tag{3.15}$$

with the important property that the range of $\tilde\rho_i$ is dense in H/H_0 if $\overline{B^i}=H$. Therefore $\tilde\rho_i$ has a unique extension to a unitary mapping ρ_i from $\overleftrightarrow{B^i/B_0^i}$ onto $\overleftrightarrow{H/H_0}=\mathscr{H}_{\mathrm{ph}}$. (Notice that in the proofs of Lemma 1 and Theorem 2 in Ref. 25 the nondegeneracy of η has not been used.)

(β) Since $W_i:G\to\operatorname{Aut}E^i$ is an η-isometric representation of G, the linear mapping

$$\widetilde W_i(g):B^i/B_0^i\to B^i/B_0^i,\quad [\Psi]_i\to[W_i(g)\Psi]_i \tag{3.16}$$

defines for any $g\in G$ an isometric automorphism of the pre-Hilbert space $\{B^i/B_0^i;([\cdot]_i,[\cdot]_i)_{B^i}\}$, which plainly possesses a unique unitary extention $\mathring W_i(g)$ onto $\overleftrightarrow{B^i/B_0^i}$. Moreover, the mapping $\mathring W_i:G\to\operatorname{Aut}\overleftrightarrow{B^i/B_0^i}$, $g\to\mathring W_i(g)$ is a unitary representation of G. If ρ_i is the unitary mapping of part (α), then for any $g\in G$ the linear mapping

$$\mathscr{W}_i(g):\mathscr{H}_{\mathrm{ph}}\to\mathscr{H}_{\mathrm{ph}},$$
$$[\Psi]\to(\rho_i\circ\mathring W_i(g)\circ\rho_i^{-1})([\Psi])=\rho_i(\mathring W_i(g)\rho_i^{-1}([\Psi])) \tag{3.17}$$

defines a unitary operator on $\mathscr{H}_{\mathrm{ph}}$ with $\mathscr{W}_i(g_1\cdot g_2)=\mathscr{W}_i(g_1)\mathscr{W}_i(g_2)$. Hence $\mathscr{W}_i:G\to\operatorname{Aut}\mathscr{H}_{\mathrm{ph}}$, $g\to\mathscr{W}_i(g)$ is a unitary representation of G. According to part (α) the set $\{\rho_i([\Phi])_i)=[\Phi]\in\mathscr{H}_{\mathrm{ph}}\mid\Phi\in B^1\cap B^2\}$ is dense in $\mathscr{H}_{\mathrm{ph}}$. On the other hand, we obtain from (3.16), (3.17), and a unitarity of ρ_i via a straightforward calculation for all $[\Psi]\in\mathscr{H}_{\mathrm{ph}}$, $\Phi\in B^1\cap B^2$, and $i\in\{1,2\}$

$$([\Psi],\mathscr{W}_i(g)[\Phi])_H = (\rho_i([\Psi]_i),\rho_i(\mathring{W}_i(g)\rho_i^{-1}([\Phi])))_H$$
$$= ([\Psi]_i,\mathring{W}_i(g)\rho_i^{-1}([\Phi]))_{B^i}$$
$$= ([\Psi]_i,\widetilde{W}_i(g)[\Phi])_{B^i}$$
$$= ([\Psi]_i,[W_i(g)\Phi]_i)_{B^i}$$
$$= (\Psi,\eta W_i(g)\Phi)_{\mathscr{H}}.$$

Therefore the η coincidence on $B^1 \cap B^2$ implies $\mathscr{W}_1(g) = \mathscr{W}_2(g)$. □

After we have established the physical equivalence of the representations V and U we will drop the former and use exclusively the latter one in the remainder of this note. By virtue of Theorems 3.1 and 3.2 it is not difficult to construct the differential of U. However, we will postpone it until Sec. V. First we will investigate the existence of the closure $\overline{U}$ of U and the connection of the η isometry and the substitution rules with the unitarity of $\overline{U}$ in $\mathscr{H}$ which we have mentioned in comment (iv) to Definition 2.1. The reader who is mainly interested in the confinement mechanism may directly precede to Secs. VII and VIII.

IV. THE CLOSURE OF U AND ITS UNITARITY IN $\mathscr{H}$

With one exception all results obtained in the present section are true only under the additional assumption that the vacuum Ψ_0 is an eigenvector of the metric operator η

$$\eta\Psi_0 = a\Psi_0, \quad a\in\mathbb{R}^+, \tag{4.1}$$

and they are immediate consequences of the following lemma.

Lemma 4.1: Let G be a symmetry group and U the η-isometric representation of Theorem 3.2. Assume Ψ_0 is an eigenvector of η. Then we have for all $g, g'\in G$ (i) $\mathscr{D}_{\rm QL} \subseteq D(U(g)^*)$ and $U(g)^*\mathscr{D}_{\rm QL} = \mathscr{D}_{\rm QL}$; (ii) $U(g)^* \times U(g')^*\Psi = U(g\cdot g')^*\Psi, \Psi\in\mathscr{D}_{\rm QL}$; and (iii) for all $\mathscr{A}_1,\ldots,\mathscr{A}_L \in I_G; f\in S(\mathscr{A}_1,\ldots,\mathscr{A}_L)$; $L\in\mathbb{N}$; and $\Psi\in\mathscr{D}_{\rm QL}$:

$$U(g)^*\phi^{\mathscr{A}_1,\ldots,\mathscr{A}_L}(f)U(g^{-1})^*\Psi$$
$$= \phi^{\mathscr{A}_1,\ldots,\mathscr{A}_L}(\mathscr{C}_L^{-1}R^{\mathscr{A}_1^\dagger,\ldots,\mathscr{A}_L^\dagger}(g^{-1})\mathscr{C}_L f)\Psi.$$

Proof: For an arbitrary $\Psi\in\mathscr{D}_{\rm QL}$ it follows from Eq. (2.19)

$$(U(g)\Psi,\phi^{\mathscr{A}_1,\ldots,\mathscr{A}_L}(f)\Psi_0) = (\phi^{\mathscr{A}_1^\dagger,\ldots,\mathscr{A}_L^\dagger}(\mathscr{C}_L f)U(g)\Psi,\Psi_0). \tag{4.2}$$

By means of Eq. (4.1), the invariance of Ψ_0 under $U(g)$, the η isometry of $U(g)$, the substitution rule in Theorem 3.2(1), and last but not least Eq. (2.19) we easily get from (4.2)

$$(U(g)\Psi,\phi^{\mathscr{A}_1,\ldots,\mathscr{A}_L}(f)\Psi_0)$$
$$= a^{-1}(\phi^{\mathscr{A}_1^\dagger,\ldots,A_L^\dagger}(\mathscr{C}_L f)U(g)\Psi,\eta U(g)\Psi_0)$$
$$= a^{-1}(U(g)^{-1}\phi^{\mathscr{A}_1^\dagger,\ldots,\mathscr{A}_L^\dagger}(\mathscr{C}_L f)U(g)\Psi,\eta\Psi_0)$$
$$= (\phi^{\mathscr{A}_1^\dagger,\ldots,\mathscr{A}_L^\dagger}(R^{\mathscr{A}_1^\dagger,\ldots,\mathscr{A}_L^\dagger}(g^{-1})\mathscr{C}_L f)\Psi,\Psi_0)$$
$$= (\Psi,\phi^{\mathscr{A}_1,\ldots,\mathscr{A}_L}(\mathscr{C}_L^{-1}R^{\mathscr{A}_1^\dagger,\ldots,\mathscr{A}_L^\dagger}(g^{-1})\mathscr{C}_L f)\Psi_0). \tag{4.3}$$

This proves (ii) since $\mathscr{C}_L, \mathscr{C}_L^{-1}$, and $R^{\mathscr{A}_1^\dagger,\ldots,\mathscr{A}_L^\dagger}(g^{-1})$ are antilinear, respectively, linear bijections and every element of $\mathscr{D}_{\rm QL}$ is a linear combination of states of the form $\phi^{\mathscr{A}_1,\ldots,\mathscr{A}_L}(f)\Psi_0$. Moreover the action of $U(g)^*$ reads explicitly

$$U(g)^*\phi^{\mathscr{A}_1,\ldots,\mathscr{A}_L}(f)\Psi_0$$
$$= \phi^{\mathscr{A}_1,\ldots,\mathscr{A}_L}(\mathscr{C}_L^{-1}R^{\mathscr{A}_1^\dagger,\ldots,\mathscr{A}_L^\dagger}(g^{-1})\mathscr{C}_L f)\Psi_0. \tag{4.4}$$

Now (ii) is a trivial consequence of Eq. (4.4) since $R^{\mathscr{A}_1^\dagger,\ldots,\mathscr{A}_L^\dagger}$ is a representation of G (Theorem 3.1). Finally the proof of (iii) reduces by virtue of the Eqs. (2.18) and (2.19) and the tensor product structure of $R^{\mathscr{A}_1,\ldots,\mathscr{A}_L}$ (Theorem 3.1) to a simple calculation

$$U(g)^*\phi^{\mathscr{A}_1,\ldots,\mathscr{A}_L}(f)U(g^{-1})^*\phi^{B_1,\ldots,B_N}(h)\Psi_0$$
$$= U(g)^*\phi^{\mathscr{A}_1,\ldots,\mathscr{A}_L,B_1,\ldots,B_N}(f\otimes[\mathscr{C}_N^{-1} \times R^{B_1^\dagger,\ldots,B_N^\dagger}(g)\mathscr{C}_N h])\Psi_0$$
$$= \phi^{\mathscr{A}_1,\ldots,B_N}(\mathscr{C}_{L+N}^{-1}R^{B_1^\dagger,\ldots,B_N^\dagger,\mathscr{A}_1^\dagger,\ldots,\mathscr{A}_L^\dagger}(g^{-1})\mathscr{C}_{L+N} \times(f\otimes[\mathscr{C}_N^{-1}R^{B_1^\dagger,\ldots,B_N^\dagger}(g)\mathscr{C}_N h]))\Psi_0$$
$$= \phi^{\mathscr{A}_1,\ldots,B_N}(\mathscr{C}_{L+N}^{-1}R^{B_1^\dagger,\ldots,B_N^\dagger,\mathscr{A}_1^\dagger,\ldots,\mathscr{A}_L^\dagger}(g^{-1}) \times[(R^{B_1^\dagger,\ldots,B_N^\dagger}(g)\mathscr{C}_N h)\otimes\mathscr{C}_L f]))\Psi_0$$
$$= \phi^{\mathscr{A}_1,\ldots,B_N}(\mathscr{C}_{L+N}^{-1}[\mathscr{C}_N h\otimes(R^{\mathscr{A}_1^\dagger,\ldots,\mathscr{A}_L^\dagger}(g^{-1})\mathscr{C}_L f)])\Psi_0$$
$$= \phi^{\mathscr{A}_1,\ldots,\mathscr{A}_L}(\mathscr{C}_L^{-1}R^{A_1^\dagger,\ldots,\mathscr{A}_L^\dagger}(g^{-1})\mathscr{C}_L f)\phi^{B_1,\ldots,B_N}(h)\Psi_0. \tag{4.5}$$

□

With this preparation it is quite trivial to show that every $U(g)$, $g\in G$ is a closable operator.

Theorem 4.1: Let G be a symmetry group and U the η-isometric representation of Theorem 3.2. If either Ψ_0 is an eigenvector of η or η has an inverse (or both are valid), then $U(g)$ is a closable operator in $\mathscr{H}$ for every $g\in G$.

Proof: Case 1: Ψ_0 is an eigenvector of η: Then Theorem 4.1 is a trivial consequence of Lemma 4.1, since the domain of $U(g)^*$ contains the dense set of quasilocal states $\mathscr{D}_{\rm QL}$.

Case 2: η has an inverse: From the η isometry of $U(g)$ we deduce that $D(U(g)^*)\supseteq\eta\mathscr{D}_{\rm QL}$. But if η is invertible then $\eta\mathscr{D}_{\rm QL}$ is dense in $\mathscr{H}$, since $\mathscr{D}_{\rm QL}$ itself has this property. For if $\eta\mathscr{D}_{\rm QL}$ is not dense in $\mathscr{H}$, then there exists a vector $\Phi\neq\mathscr{O}$ such that for all $\Psi\in\mathscr{D}_{\rm QL}$ we have $(\Psi,\eta\Phi)=\mathscr{O}$ and thus $\eta\Phi=\mathscr{O}$ in contradiction to the assumption that η has an inverse. □

In order to formulate and prove the announced necessary and sufficient condition for the unitarity of the closure $\overline{U(g)}$ [which implies for instance the commutativity of $\overline{U(g)}$ and η] we need a further lemma.

Lemma 4.2: Let $K(\mathscr{A})$ denote the closed linear subspace of $S(\mathbb{R}_4, \mathbb{C}_{\dot{\mathscr{A}}})$ defined by

$$K(\mathscr{A}) := \{h\in S(\mathbb{R}_4,\mathbb{C}_{\dot{\mathscr{A}}})\,|\,\forall\theta\in\mathscr{H}, \quad \forall\Psi\in\mathscr{D}_{\rm QL},$$
$$(\theta,\phi^{\mathscr{A}}(h)\Psi)=0\}.$$

Then $K(\mathscr{A})$ has the following properties: (a) $K(\mathscr{A})$ is invariant under the conjugation $\mathscr{C}_1$: $h\to\mathscr{C}_1 h=\overline{h}$; (b) $K(\mathscr{A}) = K(\mathscr{A}^*)$; and (c) If $D(U(g)^*)$ is dense in $\mathscr{H}$, then $K(\mathscr{A})$ is invariant under both $R^{\mathscr{A}}(g)$ and $R^{\mathscr{A}^*}(g)$.

An example of a nontrivial space $K(\mathscr{V})$ is provided by a conserved current $j_{\mu,\mathscr{V}}(f)$. Then all elements of the form $(\partial^0 h, \partial^1 h, \partial^2 h, \partial^3 h)$ with h arbitrary from $S(\mathbb{R}_4,\mathbb{C})$ belong to $K(\mathscr{V})$.

Proof: Let h_1 and h_2 be the real, respectively, the imaginary part of an arbitrary $h\in K(\mathscr{A})$

$$||\phi^{\mathscr{A}}(h_1+ih_2)\Psi||^2_{\mathscr{H}} = ||\phi^{\mathscr{A}}(h_1)\Psi||^2_{\mathscr{H}} + ||\phi^{\mathscr{A}}(h_2)\Psi||^2_{\mathscr{H}} + i[(\phi^{\mathscr{A}}(h_1)\Psi,\phi^{\mathscr{A}}(h_2)\Psi) - (\phi^{\mathscr{A}}(h_2)\Psi,\phi^{\mathscr{A}}(h_1)\Psi)] = 0. \quad (4.6)$$

This equation plainly implies (a). However, (b) then follows directly from $\phi^{\mathscr{A}}(\bar{h})^* = \phi^{\mathscr{A}^*}(h)$. Finally, for $\Theta \in D(U(g)^*)$ and $\Psi \in \mathscr{D}_{\mathrm{QL}}$ we have $(\Theta, \phi^{\mathscr{A}}(R(g)h)\Psi) = (U(g)^*\Theta, \phi^{\mathscr{A}}(h) \times U(g^{-1})\Psi)$. Thus $K(\mathscr{A})$ is invariant under $R^{\mathscr{A}}(g)$ and correspondingly $K(\mathscr{A}^*)$ under $R^{\mathscr{A}^*}(g)$. In combination with (b) this proves (c). □

Theorem 4.2: [Unitarity of $U(g)$]: Let G be a symmetry group and U the η-isometric representation of Theorem 3.2. Assume that Ψ_0 is an eigenvector of the metric operator η.

(i) The closure $\overline{U(g)}$ is an isometric operator on $\mathscr{H}$ if and only if for all $\mathscr{A} \in I_G$ and $f \in S(\mathbb{R}_4, \mathbb{C}_{\dot{\mathscr{A}}})$

$$R^{\mathscr{A}}(g)f - \overline{R^{\mathscr{A}^*}(g)\bar{f}} \in K(\mathscr{A}). \quad (4.7)$$

(ii) $\overline{U}: G \to \mathrm{Aut}\,\mathscr{H}, g \to \overline{U(g)}$ is a strongly continuous unitary representation if and only if the relation (4.7) holds for all $g \in G$.

For the proper orthochronous Poincaré group $P^{\uparrow}_{+}$

$$(R^{\mathscr{A}}(\alpha,a)f)^{\mu}(x) = \sum_{\nu \in T_{P^{\uparrow}_{+}}(\mathscr{A})} M^{\mathscr{A}}(\alpha,a)^{\mu}{}_{\nu} f^{\nu}(\Lambda(\alpha^{-1})(x-a)),$$

the condition (4.7) says that the matrix $M^{\mathscr{A}^*}(\alpha,a)$ according to which the adjoint of $\varphi_{\mu,\mathscr{A}}$ transforms has to be essentially equal to the complex conjugate of the matrix $M^{\mathscr{A}}(\alpha,a)$ with which the field $\varphi_{\mu,\mathscr{A}}$ itself transforms.

It should be pointed out that the η isometry is a vital assumption for the proof that (4.7) is a sufficient condition.

Proof: (i) (1)⟹: Let Ψ, ϕ be from $\mathscr{D}_{\mathrm{QL}}$. The isometry of $\overline{U}(g)$ and Theorem 3.2 imply

$$\begin{aligned}(\Psi, U(g)\phi^{\mathscr{A}^*}(f)U(g^{-1})\Phi) &= (U(g)\phi^{\mathscr{A}}(\bar{f})U(g^{-1})\Psi,\Phi)\\ &= (\phi^{\mathscr{A}}(R(g)\bar{f})\Psi,\Phi) = (\Psi,\phi^{\mathscr{A}^*}(\overline{R^{\mathscr{A}}(g)\bar{f}})\Phi),\end{aligned}$$

and therefore

$$R^{\mathscr{A}^*}(g)f - \overline{R^{\mathscr{A}}(g)\bar{f}} \in K(\mathscr{A}).$$

(2)⟸: Consider $\Psi \in \mathscr{D}_{\mathrm{QL}}$ and $\Phi = \phi^{\mathscr{A}_1,\dots,\mathscr{A}_L}(f)\Psi_0$. Then Theorem 3.2(1) and Lemma 4.1 (iii) imply

$$\begin{aligned}(U(g)\Psi, U(g)\phi^{\mathscr{A}_1,\dots,\mathscr{A}_L}(f)\Psi_0) &= (\Psi, U(g)^*\phi^{\mathscr{A}_1,\dots,\mathscr{A}_L}(R^{\mathscr{A}_1,\dots,\mathscr{A}_L}(g)f)\Psi_0)\\ &= (\Psi,\phi^{\mathscr{A}_1,\dots,\mathscr{A}_L}(\mathscr{C}_L^{-1} R^{\mathscr{A}_1^{\dagger},\dots,\mathscr{A}_L^{\dagger}}(g^{-1})\\ &\quad \times \mathscr{C}_L R^{\mathscr{A}_1,\dots,\mathscr{A}_L}(g)f)\Psi_0). \end{aligned} \quad (4.8)$$

Let $(\sum_{j=1}^{n} \otimes_{i=1}^{L} h_i^{(j)})_{n\in\mathbb{N}}$ be a sequence from $\otimes_{i=1}^{L} S(\mathbb{R}_4, \mathbb{C}_{\dot{\mathscr{A}}_i})$ converging to f. Then from Eqs. (4.8) and (2.18), the continuity of the matrix elements of $\phi^{\mathscr{A}_1,\dots,\mathscr{A}_L}(f)$, the tensor product structure and continuity of $R^{\mathscr{A}_1,\dots,\mathscr{A}_L}(g)$ (Theorem 3.1), and last but not least from the condition (4.7) and Lemma 4.2 we obtain

$$\begin{aligned}(U(g)\Psi, U(g)\phi^{\mathscr{A}_1,\dots,\mathscr{A}_L}(f)\Psi_0) &= \lim_{n\to\infty}\Big(\Psi,\phi^{\mathscr{A}_1,\dots,\mathscr{A}_L}(\mathscr{C}_L^{-1}R^{\mathscr{A}_1^{\dagger},\dots,\mathscr{A}_L^{\dagger}}(g^{-1})\\ &\quad\times\sum_{j=1}^{n}\prod_{i=1}^{L} R^{\mathscr{A}_i}(g)h_i^{(j)})\Psi_0\Big)\\ &= \lim_{n\to\infty}\Big(\Psi,\sum_{j=1}^{n}\prod_{i=1}^{L}\phi^{\mathscr{A}_i}(R^{\mathscr{A}_i^{\dagger}}(g^{-1})\,\overline{R^{\mathscr{A}_i}(g)h_i^{(j)}})\Psi_0\Big)\\ &= \lim_{n\to\infty}\Big(\Psi,\sum_{j=1}^{n}\prod_{i=1}^{L}\phi^{\mathscr{A}_i}(h_i^{(j)})\Psi_0\Big) = (\Psi,\phi^{\mathscr{A}_1,\dots,\mathscr{A}_L}(f)\Psi_0).\end{aligned}$$

Thus $U(g)$ is isometric on the dense subspace $\mathscr{D}_{\mathrm{QL}}$ and therefore its closure $\overline{U(g)}$ is an isometric operator on $\mathscr{H}$.

(ii) If condition (4.7) holds for all $g \in G$ the $\overline{U(g)}$ and $\overline{U(g)}^{-1} = \overline{U(g^{-1})}$ are both isometric operators on $\mathscr{H}$ and therefore unitary.

It remains to extend the strong continuity in g from the operators $U(g)$ [Theorem 3.2(2)] to their closures $\overline{U(g)}$. In view of the unitarity of $\overline{U}(g)$ it suffices to prove the weak continuity in g. Let $(g_\alpha)_{\alpha\in D(<)}$ be a net which converges in the topology of G to the unit element e. Then once again the unitarity of $\overline{U(g)}$ implies that for every $\Psi \in \mathscr{H}$ the net $(\overline{U(g_\alpha)}\Psi)_{\alpha\in I(<)}$ is bounded (in the norm) by $||\Psi||_{\mathscr{H}}$. Therefore according to Lemma 1.31 in Ref. 26, Chap. III, it suffices to prove $\lim_\alpha(\Phi, \overline{U(g_\alpha)}\Psi) = (\Phi,\Psi)$ for every $\Psi \in \mathscr{H}$ and all Φ from a fundamental subset of $\mathscr{H}$. The set of all vectors of the form $\phi^{\mathscr{A}_1,\dots,\mathscr{A}_L}(f)\Psi_0$ represents such a fundamental set. Finally, by means of Theorem 3.2(1), Theorem 3.1, and the continuity of the matrix elements of the field operators it follows for all $\Psi \in \mathscr{H}$

$$\begin{aligned}\lim_\alpha(\phi^{\mathscr{A}_1,\dots,\mathscr{A}_L}(f)\Psi_0, \overline{U(g_\alpha)}\Psi) &= \lim_\alpha(\phi^{\mathscr{A}_1,\dots,\mathscr{A}_L}(R^{\mathscr{A}_1,\dots,\mathscr{A}_L}(g_\alpha^{-1})f)\Psi_0,\Psi)\\ &= (\phi^{\mathscr{A}_1,\dots,\mathscr{A}_L}(f)\Psi_0,\Psi).\end{aligned}$$

□

V. THE DIFFERENTIAL OF AN UNBOUNDED REPRESENTATION U

In the present and the following section we assume the symmetry group G to be a Lie group in its strictest sense (analytic manifold) and moreover countable at infinity (a countable union of compact subspaces). The main burden in the construction of the differential ∂U of an unbounded representation U (this means an η-skew-symmetric representation ∂U of the Lie algebra $\mathfrak{g}$ of G induced by U) has already been unloaded in Sec. III. Since the differentiation of continuous representations on countably normed spaces is a well-understood operation (Ref. 9, Chaps. 4.1–4.4), the explicit connection between the unbounded representation U in $\mathscr{H}$ and the continuous representations $R^{\mathscr{A}_1,\dots,\mathscr{A}_L}$ on the countably normed spaces $S(\mathscr{A}_1,\dots,\mathscr{A}_L)$ established in Sec. III practically dictates the promising path of argumentation. All we have to do is to introduce some suitable definitions, collect the relevant results about continuous representations on countably normed spaces from the literature,[9] and transmit them by means of Theorems 3.1 and 3.2 to the pair $(U,\mathscr{H})$.

If $\mathbb{O}$ is an open set of $\mathbb{R}_n$ then $\mathscr{C}^\infty(\mathbb{O};S(\mathscr{A}_1,...,\mathscr{A}_L))$ denotes the vector space of all infinitely often differentiable functions

$$h:\quad \mathbb{O}\times\mathbb{R}_{4L}\to\bar{\otimes}_{i=1}^{L}\mathbb{C}^{\circ}_{\mathscr{A}_i},\quad (y,x)\to h(y,x), \tag{5.1}$$

such that for every relatively compact open set $\Omega\subseteq\mathbb{O}$ and every $p\in\mathbb{N}^0$, $q\in\mathbb{N}_n^0$ the supremum

$$\sup\left\{\left|\left|\prod_{i=1}^{n}\frac{\partial^{q_i}}{(\partial y_i)^{q_i}}h(y,\cdot)\right|\right|_p\middle| y\in\Omega\right\}<+\infty \tag{5.2}$$

exists. Hence for every fixed $y\in\mathbb{O}$ the function h together with all its derivatives in the first n variables y_i $(i=1,...,n)$ is an element from $S(\mathscr{A}_1,...,\mathscr{A}_L)$. More general, if G is a Lie group, a function $h:G\times\mathbb{R}_{4L}\to\bar{\otimes}_{i=1}^{L}\mathbb{C}^{\circ}_{\mathscr{A}_i}$ is said to be from $\mathscr{C}^\infty(G;S(\mathscr{A}_1,...,\mathscr{A}_L))$ if for any $g\in G$ and a local chart (O_α, V_α) of a maximal atlas with $g\in O_\alpha$ the function

$$h\circ v_\alpha^{-1}:\quad v_\alpha[O_\alpha]\times\mathbb{R}_{4L}\to\bar{\otimes}_{i=1}^{L}\mathbb{C}^{\circ}_{\mathscr{A}_i};\quad (y,x)\to h(v_\alpha^{-1}(y),x) \tag{5.3}$$

is an element from $\mathscr{C}^\infty(v_\alpha[O_\alpha];S(\mathscr{A}_1,...,\mathscr{A}_L))$. Finally, an element $f\in S(\mathscr{A}_1,...,\mathscr{A}_L)$ is called a *differentiable* or $\mathscr{C}^\infty$ *vector* for $R^{\mathscr{A}_1,...,\mathscr{A}_L}$ if the mapping

$$f_R:\quad G\to S(\mathscr{A}_1,...,\mathscr{A}_L),\quad g\to f_R(g)=R^{\mathscr{A}_1,...,\mathscr{A}_L}(g)f \tag{5.4}$$

is from $\mathscr{C}^\infty(G;S(\mathscr{A}_1,...,\mathscr{A}_L))$. The set of all $\mathscr{C}^\infty$ vectors for $R^{\mathscr{A}_1,...,\mathscr{A}_L}$ is denoted by $\sigma^\infty(\mathscr{A}_1,...,\mathscr{A}_L)$:

$$\sigma^\infty(\mathscr{A}_1,...,\mathscr{A}_L) := \{f\in S(\mathscr{A}_1,...\mathscr{A}_L)|f_R\in\mathscr{C}^\infty(G;S(\mathscr{A}_1,...,\mathscr{A}_L))\}. \tag{5.5}$$

If $\mathfrak{g}$ denotes the Lie algebra of G, $\mathscr{G}$ the universal enveloping algebra of its complexification $\mathfrak{g}_c=\mathfrak{g}+i\mathfrak{g}$, and exp the exponential mapping from $\mathfrak{g}$ into G, then the following facts are known about the differentiation of a continuous representation $R^{\mathscr{A}_1,...,\mathscr{A}_L}$ (Theorem 3.1) on a countable normed space $S(\mathscr{A}_1,...,\mathscr{A}_L)$.

(I) The set $\sigma^\infty(\mathscr{A}_1,...,\mathscr{A}_L)$ of $\mathscr{C}^\infty$ vectors is a dense linear subspace of $S(\mathscr{A}_1,...,\mathscr{A}_L)$ and invariant under the application of $R^{\mathscr{A}_1,...,\mathscr{A}_L}(g)$, $g\in G$.

(II) For every $X\in\mathfrak{g}$ and $f\in\sigma^\infty(\mathscr{A}_1,...,\mathscr{A}_L)$ the limit

$$\partial R^{\mathscr{A}_1,...,\mathscr{A}_L}(X)f:=\lim_{t\to0}t^{-1}[R^{\mathscr{A}_1,...,\mathscr{A}_L}(\exp tX)f-f]$$

exists in the topology of $S(\mathscr{A}_1,...,\mathscr{A}_L)$ and defines a linear operator from $\sigma^\infty(\mathscr{A}_1,...,\mathscr{A}_L)$ into $\sigma^\infty(\mathscr{A}_1,...,\mathscr{A}_L)$.

(III) The differential

$$\partial R^{\mathscr{A}_1,...,\mathscr{A}_L}:\quad \mathfrak{g}\to\mathrm{End}\,\sigma^\infty(\mathscr{A}_1,...,\mathscr{A}_L),\quad X\to\partial R^{\mathscr{A}_1,...,\mathscr{A}_L}(X)$$

is a representation of $\mathfrak{g}$ on $\sigma^\infty(\mathscr{A}_1,...,\mathscr{A}_L)$. It has a unique extension to a representation $\partial_{\mathscr{G}}R^{\mathscr{A}_1,...,\mathscr{A}_L}$ of $\mathscr{G}$ on $\sigma^\infty(\mathscr{A}_1,...,\mathscr{A}_L)$.

(IV) For all $h\in\sigma^\infty(B_1,...,B_N)$, $f\in\sigma^\infty(\mathscr{A}_1,...,\mathscr{A}_L)$, and $X\in\mathfrak{g}$ we have

$$\partial R^{\mathscr{A}_1,...,\mathscr{A}_L,B_1,...,B_N}(X)(f\otimes h) = (\partial R^{\mathscr{A}_1,...,\mathscr{A}_L}(X)f)\otimes h+f\otimes(\partial R^{B_1,...,B_N}(X)h). \tag{5.6}$$

The first three statements can be found in Ref. 9, pp. 252–254. The last one is a direct consequence of the inclusion $\sigma^\infty(\mathscr{A}_1,...,\mathscr{A}_L)\otimes\sigma^\infty(B_1,...,B_N)\subset\sigma^\infty(\mathscr{A}_1,...,\mathscr{A}_L,B_1,...,B_N)$, the statement (II), and the following identity which is easily obtained from the tensor product structure of $R^{\mathscr{A}_1,...,B_N}(g)$:

$$\begin{aligned}
&t^{-1}[R^{\mathscr{A}_1,...,\mathscr{A}_L,B_1,...,B_N}(\exp tX)(f\otimes h)-(f\otimes h)]\\
&\quad-[(\partial R^{\mathscr{A}_1,...,\mathscr{A}_L}(X)f)\otimes h+f\otimes(\partial R^{B_1,...,B_N}(X)h)]\\
&\quad=(t^{-1}[R^{\mathscr{A}_1,...,\mathscr{A}_L}(\exp tX)f-f]\\
&\quad-\partial R^{\mathscr{A}_1,...,\mathscr{A}_L}(X)f)\otimes h\\
&\quad+f\otimes(t^{-1}[R^{B_1,...,B_N}(\exp tX)h-h]\\
&\quad-\partial R^{B_1,...,B_N}(X)h)\\
&\quad+(t^{-1}[R^{\mathscr{A}_1,...,\mathscr{A}_L}(\exp tX)f-f])\\
&\quad\otimes(R^{B_1,...,B_N}(\exp tX)h-h).
\end{aligned} \tag{5.7}$$

Let $\mathscr{D}^\infty(U)$ denote the dense linear subspace obtained from $\mathscr{D}_{\mathrm{QL}}$ by restricting the functions f in Eq. (2.20) to the dense subspace $\sigma^\infty(\mathscr{A}_1,...,\mathscr{A}_L)$ of $S(\mathscr{A}_1,...,\mathscr{A}_L)$. Explicitly it is given by the linear hull

$$\mathscr{D}^\infty(U)=\mathrm{LH}\{\phi^{\mathscr{A}_1,...,\mathscr{A}_L}(f)\Psi_0|f\in\sigma^\infty(\mathscr{A}_1,...,\mathscr{A}_L;\ \mathscr{A}_1,...,\mathscr{A}_L\in I_G;\ L\in\mathbb{N}\}. \tag{5.8}$$

With these preparations it is not hard to demonstrate the existence of (what we call in possible misuse of the phrase) the differential ∂U of U.

Theorem 5.1: Let U and $R^{\mathscr{A}_1,...,\mathscr{A}_L}$; $\mathscr{A}_1,...,\mathscr{A}_L\in I_G$; $L\in\mathbb{N}$ be the representations of a symmetry (Lie) group G described in Theorem 3.2, respectively, Theorem 3.1. Then for all $X\in\mathfrak{g}$ and $\Psi\in\mathscr{D}^\infty(U)$ the strong limit $(t\in\mathbb{R})$

$$\partial U(X)\Psi=\operatorname*{s-lim}_{t\to0}t^{-1}[U(\exp tX)\Psi-\Psi] \tag{5.9}$$

exists in $\mathscr{H}$ and defines an unbounded linear operator from $\mathscr{D}^\infty(U)$ to $\mathscr{D}^\infty(U)$ with the following properties.

(i) For every $f\in\sigma^\infty(\mathscr{A}_1,...,\mathscr{A}_l)$; $\mathscr{A}_1,...,\mathscr{A}_L\in I_G$; $L\in\mathbb{N}$, respectively, $\Psi\in\mathscr{D}^\infty(U)$ the operator $\partial U(X)$ satisfies the equations

$$\partial U(X)\Psi_0=\mathscr{O}\quad(\mathscr{O}\text{ is the zero vector in }\mathscr{H}), \tag{5.10}$$

$$[\partial U(X),\phi^{\mathscr{A}_1,...,\mathscr{A}_L}(f)]_-\Psi =\phi^{\mathscr{A}_1,...,\mathscr{A}_L}(\partial R^{\mathscr{A}_1,...,\mathscr{A}_L}(X)f)\Psi. \tag{5.11}$$

(ii) For every $X\in\mathfrak{g}$, $\partial U(X)$ is η-skew symmetric; this means

$$\forall\Psi,\Phi\in\mathscr{D}^\infty(U),\quad\langle\partial U(X)\Psi,\Phi\rangle=-\langle\Psi,\partial U(X)\Phi\rangle. \tag{5.12}$$

(iii) The mapping ∂U: $\mathfrak{g}\to\mathrm{End}\,\mathscr{D}^\infty(U)$, $X\to\partial U(X)$ is a representation of the Lie algebra $\mathfrak{g}$ on $\mathscr{D}^\infty(U)$ which has a unique extension to a representation $\partial_{\mathscr{G}}U$ of the universal enveloping algebra $\mathscr{G}$ of $\mathfrak{g}_c$ on $\mathscr{D}^\infty(U)$.

Proof: The existence of the limit (5.9) for $\Psi=\Psi_0$ and Eq. (5.10) are trivial consequences of the invariance of Ψ_0 under U. Hence in order to prove the existence of the operator $\partial U(X)$ and at the same time Eq. (5.11) for $\Psi\neq\Psi_0$ it suffices to show that for all finite linear combinations of the states $\phi^{\mathscr{A}_1,...,\mathscr{A}_L}(f)\Psi_0$, which we abbreviate by $(lc\phi)^{\cdots}(f)\Psi_0$, the following limit exists:

$$\lim_{t\to 0} \| t^{-1}[U \exp tX)(lc\phi)^{\cdot\cdot}(f)\Psi_0 - (lc\phi)^{\cdot\cdot}(f)\Psi_0]$$
$$- (lc\phi)^{\cdot\cdot}(\partial R^{\cdot\cdot}(X)f)\Psi_0\|_{\mathscr{H}} = 0. \tag{5.13}$$

Due to Theorem 3.2(2) this is certainly true, if for all $f\in\sigma^\infty(\mathscr{A}_1,...,\mathscr{A}_L)$; $\mathscr{A}_1,...,\mathscr{A}_L\in I_G$; $L\in\mathbb{N}$ we can show

$$\lim_{t\to 0} \|\phi^{\mathscr{A}_1,...,\mathscr{A}_L}(t^{-1}[R^{\mathscr{A}_1,...,\mathscr{A}_L}(\exp tX)f - f]$$
$$- \partial R^{\mathscr{A}_1,...,\mathscr{A}_L}(X)f)\Psi_0\|_{\mathscr{H}} = 0. \tag{5.14}$$

Since $\|\phi^{\mathscr{A}_1,...,\mathscr{A}_L}(h)\Psi_0\|^2_{\mathscr{H}}$ is a continuous linear functional on $S(\mathscr{A}_L^*,...,\mathscr{A}_1^*)\bar{\otimes} S(\mathscr{A}_1,...,\mathscr{A}_L)$ there exists a smallest number $p\in\mathbb{N}^0$ and a positive real number $\rho = \rho(\mathscr{A}_1,...,\mathscr{A}_L)$ such that

$$\begin{aligned} &\|\phi^{\mathscr{A}_1,...,\mathscr{A}_L}(t^{-1}[R^{\mathscr{A}_1,...,\mathscr{A}_L}(\exp tX)f - f] \\ &\quad - \partial R^{\mathscr{A}_1,...,\mathscr{A}_L}(X)f)\Psi_0\|^2_{\mathscr{H}} \\ &\quad \leqslant \rho\|t^{-1}[R^{\mathscr{A}_1,...,\mathscr{A}_L}(\exp tX)f - f] \\ &\quad - \partial R^{\mathscr{A}_1,...,\mathscr{A}_L}(X)f\|_p \\ &\quad \times \|\mathscr{C}_L(t^{-1}[R^{\mathscr{A}_1,...,\mathscr{A}_L}(\exp tX)f - f] \\ &\quad - \partial R^{\mathscr{A}_1,...,\mathscr{A}_L}(X)f)\|_p. \end{aligned} \tag{5.15}$$

Here $\mathscr{C}_L$ is the isometric bijection defined at the end of Sec. II. Now Eq. (5.14) follows from the inequality (5.15), the continuity of $\mathscr{C}_L$, and statement (II) above. By means of statement (IV) and

$$\phi^{\mathscr{A}_1,...,\mathscr{A}_L,B_1,...,B_N}(f\otimes h)\Psi_0 = \phi^{\mathscr{A}_1,...,\mathscr{A}_L}(f)\phi^{B_1,...,B_N}(h)\Psi_0,$$

Eq. (5.11) can for any $\Psi\in\mathscr{D}^\infty(U)$ be traced back to that for $\Psi = \Psi_0$. Furthermore (iii) is a simple consequence of (i) and statement (III). Finally, the η isometry of U leads directly to the η-skew symmetry of $\partial U(X)$. Indeed for all $\Psi,\Phi\in\mathscr{D}^\infty(U)$ we deduce from Eq. (5.9) and the continuity of η

$$\begin{aligned} &\langle \partial U(X)\Phi,\Psi\rangle \\ &\quad = \lim_{t\to 0} t^{-1}[(U(\exp tX)\Phi,\eta\Psi) - (\Phi,\eta\Psi)] \\ &\quad = \lim_{t\to 0} t^{-1}[(\Phi,\eta U(\exp t(-X))\Psi) - (\Phi,\eta\Psi)] \\ &\quad = \langle\Phi,\partial U(-X)\Psi\rangle = -\langle\Phi,\partial U(X)\Psi\rangle \qquad \square \end{aligned}$$

If U is a continuous representation on $\mathscr{H}$ then it is well known[10,27] that the operators $\partial U(X)$ are closable. In the present case of noncontinuous representations U it turns out that the same restrictions of the metric operator η (i.e., either η invertible or Ψ_0 an eigenvector of η; see Theorem 4.1) which implied the operators $U(g)$ to be closable also assure the existence of the closure $\overline{\partial U(X)}$ of $\partial U(X)$ for every $X\in\mathfrak{g}$.

Theorem 5.2: Let U be the representation of a symmetry (Lie) group G described in Theorem 3.2 and ∂U its differential. Assume that either Ψ_0 is an eigenvector of η or η is invertible (or both). Then $\partial U(X)$ is closable for every $X\in\mathfrak{g}$.

Proof: Since the domain $D(\partial U(X)) = \mathscr{D}^\infty(U)$ is dense in $\mathscr{H}$, the adjoint operator $[\partial U(X)]^*$ exists. It has to be shown that its domain $D([\partial U(X)]^*)$ is dense in $\mathscr{H}$. The η-skew symmetry of $\partial U(X)$ implies for all $\Psi\in\mathscr{D}^\infty(U)$

$$[\partial U(X)]^*\eta\Psi = -\eta\partial U(X)\Psi, \tag{5.16}$$

and via Theorem 5.1(i) for the vacuum

$$[\partial U(X)]^*\eta\Psi_0 = \mathscr{O}. \tag{5.17}$$

Case 1: η invertible: By literally the same arguments as in the proof of Theorem 4.1 (case 2) it follows that $\mathscr{D}^\infty(U)$ dense in $\mathscr{H}$ implies $\eta\mathscr{D}^\infty(U)$ dense in $\mathscr{H}$.

Case 2: $\eta\Psi_0 = a\Psi_0$, $a\in\mathbb{R}^+$: Then Eq. (5.17) says that Ψ_0 is in the domain of $[\partial U(X)]^*$ and moreover

$$[\partial U(X)]^*\Psi_0 = \mathscr{O}. \tag{5.18}$$

Consider $\Psi\in\mathscr{D}^\infty(U)$ and $f\in\sigma^\infty(\mathscr{A}_L^*,...,\mathscr{A}_1^*)$. Then $\mathscr{C}_L^{-1}f$ is from $S(\mathscr{A}_1,...,\mathscr{A}_L)$ and we obtain by means of Theorem 5.1(i) and Eq. (5.18)

$$\begin{aligned} &(\partial U(X)\Psi,\phi^{\mathscr{A}_1,...,\mathscr{A}_L}(\mathscr{C}_L^{-1}f)\Psi_0) \\ &\quad = -(\phi^{\mathscr{A}_L^*,...,\mathscr{A}_1^*}(\partial R^{\mathscr{A}_L^*,...,\mathscr{A}_1^*}(X)f)\Psi,\Psi_0) \\ &\qquad + (\partial U(X)\phi^{\mathscr{A}_L^*,...,\mathscr{A}_1^*}(f)\Psi,\Psi_0) \\ &\quad = -(\Psi,\phi^{\mathscr{A}_1,...,\mathscr{A}_L}(\mathscr{C}_L^{-1}\partial R^{\mathscr{A}_L^*,...,\mathscr{A}_1^*}(X)f)\Psi_0). \end{aligned} \tag{5.19}$$

Therefore $D([\partial U(X)]^*)$ contains the linear space

$$\begin{aligned} V := \mathrm{LH}\{&\phi^{\mathscr{A}_1,...,\mathscr{A}_L}(\mathscr{C}_L^{-1}f)\Psi_0 | f\in\sigma^\infty(\mathscr{A}_L^*,...,\mathscr{A}_1^*); \\ &\mathscr{A}_1,...,\mathscr{A}_L\in I_G;\ L\in\mathbb{N}\}. \end{aligned} \tag{5.20}$$

Since $\sigma^\infty(\mathscr{A}_L^*,...,\mathscr{A}_1^*)$ is dense in $S(\mathscr{A}_L^*,...,\mathscr{A}_1^*)$ and $\mathscr{C}_L^{-1}$ is a antilinear isometric bijection of this space onto $S(\mathscr{A}_1,...,\mathscr{A}_L)$, the linear space $\mathscr{C}_L^{-1}\sigma^\infty(\mathscr{A}_L^*,...,\mathscr{A}_1^*)$ is dense in $S(\mathscr{A}_1,..., \mathscr{A}_L)$. But then V is dense in $\mathscr{D}^\infty(U)$ and therefore in $\mathscr{H}$. $\square$

VI. INFINITESIMAL CHARACTERIZATION OF $\mathscr{C}^\infty$ VECTORS

For the construction of the differential ∂U in $\mathscr{H}$ the subspaces $\sigma^\infty(\mathscr{A}_1,...,\mathscr{A}_L)$ of $\mathscr{C}^\infty$ vectors for the continuous representations $R^{\mathscr{A}_1,...,\mathscr{A}_L}$ on the test function spaces $S(\mathscr{A}_1,...,\mathscr{A}_L)$ and the corresponding differentials $\partial R^{\mathscr{A}_1,...,\mathscr{A}_L}$ played a fundamental part. Even more, the results of Sec. III represent a first important step for the integration of a representation of a Lie algebra to an unbounded representation of the universal covering of a corresponding Lie group in $\mathscr{H}$. By means of Theorems 3.1 and 3.2 this problem is completely reduced to the construction of a bounded representation in the countably normed spaces $S(\mathscr{A}_1,...,\mathscr{A}_L)$ and hence to the generalization of the well-known Banach space results[9–18] to countably normed spaces.

From the experience with the Banach space representations it is obvious that a characterization of the subspaces $\sigma^\infty(\mathscr{A}_1,...,\mathscr{A}_L)$ of $\mathscr{C}^\infty$ vectors in terms of the differentials $\partial R^{\mathscr{A}_1,...,\mathscr{A}_L}$ will be an unavoidable ingredient for the integration. Our aim in this section is to derive such an infinitesimal characterization. The remaining steps of the integration will be presented in a separate note.

Since the indices $\mathscr{A}_1,...,\mathscr{A}_L$ remain fixed throughout this section we replace them by dots. Moreover, the notions of a closed or closable operator T, of the closure $\overline{T}$, of a T-convergent sequence, and of the graph of T in a countably normed space are literally taken over from the corresponding definitions in Banach spaces (Ref. 26. Chap. III, Sec. 5) by replacing the norm of the latter by the countable set of norms of the former. For instance, consider the direct product $S(\cdots)\times S(\cdots)$ consisting of all ordered pairs $\{f;h\}$ with $f,h\in S(\cdots)$. By the standard definition of the linear operations it becomes a vector space over $\mathbb{C}$. Futhermore it is a countable normed space (separable, metrizable, complete, etc.; see Ref. 20, Chap. I.3) if the countable set of pairwise compatible norms is defined by $\|\{f;h\}\|_p := (\|f\|_p^2 + \|h\|_p^2)^{1/2}$, $p\in\mathbb{N}^0$. Now the graph of an operator T with domain $D(T)$ is the subset

$$\Upsilon(T) := \{\{f;Tf\}\,|\,f\in D(T)\}. \tag{6.1}$$

Exactly as in the case of a Banach space, T is closed if $\Upsilon(T)$ is a closed linear subspace of $S(\cdots)\times S(\cdots)$ and it is closable if the closure $\overline{\Upsilon(T)}$ [of the set $\Upsilon(T)$] is a graph. Finally if T is closable then $\Upsilon(\overline{T}) = \overline{\Upsilon(T)}$.

For fixed $X\in\mathfrak{g}$ the mapping

$$\rho_X^{\cdots}:\quad \mathbb{R}\to\mathrm{Aut}\,S(\cdots),\quad t\to\rho_X^{\cdots}(t) = R^{\cdots}(\exp tX) \tag{6.2}$$

generates a strongly continuous one-parameter group of operators in $S(\cdots)$ whose infinitesimal generator $dR^{\cdots}(X)$ is defined by

$$D(dR^{\cdots}(X))$$

$$:= \{f\in S(\cdots)\,|\,\lim_{t\to 0} t^{-1}[R^{\cdots}(\exp tX)f - f]\ \text{exists}\}, \tag{6.3}$$

$$dR^{\cdots}(X)f := \lim_{t\to 0} t^{-1}[R^{\cdots}(\exp tX)f - f],$$

$$f\in D(dR^{\cdots}(X)). \tag{6.4}$$

From these definitions and the statement (II) in Sec. V it follows at once that

$$dR^{\cdots}(X)\supseteq\partial R^{\cdots}(X) \tag{6.5}$$

and therefore

$$\forall n\in\mathbb{N},\quad X\in\mathfrak{g},$$

$$\sigma^\infty(\cdots)\subseteq D(dR^{\cdots}(X)^n). \tag{6.6}$$

In order to sharpen these relations we will need the following lemma which represents the bridge to the results of Goodman[11] for Banach spaces.

Lemma 6.1: Let $S_p(\cdots)$ ($p\in\mathbb{N}^0$) denote the completion of $S(\cdots)$ with respect to the norm $\|\cdots\|_p$, $R_p^{\cdots}(g)$ the unique continuous extension of $R^{\cdots}(g)$ to $S_p(\cdots)$ and $\sigma_p^\infty(\cdots)$, $\partial R_p^{\cdots}(X)$, respectively, $dR_p^{\cdots}(X)$ the corresponding space of $\mathscr{C}^\infty$ vectors, the differential of $R_p^{\cdots}$, respectively, the infinitesimal generator of $R_p^{\cdots}(\exp tX)$, $t\in\mathbb{R}$. Then the following statements hold for all $p\in\mathbb{N}^0$:

(a) $R_{p+1}^{\cdots}(g) = R_p^{\cdots}(g)\restriction S_{p+1}(\cdots)$ ($\restriction$ ~restriction),

(b) $\sigma_{p+1}^\infty(\cdots)\subseteq\sigma_p^\infty(\cdots)$,

(c) $\partial R_{p+1}^{\cdots}(X) = \partial R_p^{\cdots}(X)\restriction\sigma_{p+1}^\infty(\cdots)$,

(d) $dR_{p+1}^{\cdots}(X)\subseteq dR_p^{\cdots}(X)$,

(e) $\sigma^\infty(\cdots) = \bigcap_{n\in\mathbb{N}^0}\sigma_n^\infty(\cdots)$,

(f) $dR_p^{\cdots}(X) = \overline{\partial R_p^{\cdots}(X)}^p$, where $\overline{T}_p^p$ denotes the closure of the operator T_p in the Banach space $S_p(\cdots)$.

Proof: By definition we have for all $p\in\mathbb{N}^0$

$$s_{p+1}(\cdots)\subseteq S_p(\cdots), \tag{6.7}$$

and due to the statement (S.I) in Sec. II, respectively, Eq. (2.15)

$$S(\cdots) = \bigcap_{n\in\mathbb{N}^0} S_n(\cdots). \tag{6.8}$$

Then statement (c) obviously follows from (a) and (b). Statement (d) is a direct consequence of (a), the definition [analogous to (6.3) and (6.4)] of the infinitesimal generator $dR_p^{\cdots}(X)$, and the pairwise compatibility of the norms $\|\cdots\|_p$; $p\in\mathbb{N}^0$. Next observe that $\sigma_p^\infty(\cdots)$ is dense in the Banach space $S_p(\cdots)$ and stable with respect to $R_p^{\cdots}(\exp tX)$, $t\in\mathbb{R}$ [Banach space analog of statement (I) in Sec. V]. But then (f) follows from a result of Poulsen (Ref. 27, Corollary 1.3), since the one-parameter group $(\{R_p^{\cdots}(\exp tX)|t\in\mathbb{R}\},\cdot)$ is strongly continuous and $\sigma_p^\infty(\cdots)\subseteq D(dR_p^{\cdots}(X))$.

It remains to prove (a), (b), and (e). We begin with the proof for (a). Let $h\in S_{p+1}(\cdots)$ and $(h_n)_{n\in\mathbb{N}}$ be an arbitrary sequence with $h_n\in S(\cdots)$ which converges in $S_{p+1}(\cdots)$ to h. Since the norms are nondecreasing ($\|f\|_p\leqslant\|f\|_{p+1}$) this sequence also converges in $S_p(\cdots)$ to h. However, from the validity of

$$\lim_{n\to\infty}\|R_k^{\cdots}(g)h - R^{\cdots}(g)h_n\|_k = 0 \tag{6.9}$$

for both $k=p$ and $k=p+1$ it follows that for every $\epsilon>0$ there exists an $N(\epsilon)\in\mathbb{N}$ such that for all $n > N(\epsilon)$ we have

$$\begin{aligned}
&\|R_{p+1}^{\cdots}(g)h - R_p^{\cdots}(g)h\|_{p+1}\\
&\quad\leqslant\|R_{p+1}^{\cdots}(g)h - R^{\cdots}(g)h_n\|_{p+1}\\
&\qquad+\|R_p^{\cdots}(g)h - R^{\cdots}(g)h_n\|_{p+1}\\
&\quad<\epsilon/2 + \|R_p^{\cdots}(g)h - R^{\cdots}(g)h_n\|_{p+1}.
\end{aligned}\tag{6.10}$$

The sequence $(R_p^{\cdots}(g)h - R^{\cdots}(g)h_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in both norms $\|\cdots\|_p$ and $\|\cdots\|_{p+1}$ and converges to zero in the first one. Due to the pairwise compatibility of the norms it has to converge to zero also in the second one. Hence for every $\epsilon>0$ there exists an $\widetilde{N}(\epsilon)\in\mathbb{N}$ such that for all $n>\widetilde{N}(\epsilon)$ we have

$$\|R_p^{\cdots}(g)h - R^{\cdots}(g)h_n\|_{p+1} < \epsilon/2. \tag{6.11}$$

But the inequalities (6.10) and (6.11) together imply (a).

Analogous to the definitions at the beginning of Sec. V, f is from $\sigma_k^\infty(\cdots)$ if and only if for any local chart $(\mathscr{O}_\alpha^n, v_\alpha)$ of a maximal atlas for G we have

$$\forall q\in\mathbb{N}_n^0,$$

$$\sup\left\{\left|\left|\left(\prod_{i=1}^n\frac{\partial^{q_i}}{(\partial y_i)^{q_i}}\right)R_k^{\cdots}(v_\alpha^{-1}(y_1,\ldots,y_n))f\right|\right|_k\,\right| \tag{6.12}$$

$$(y_1,\ldots,y_n)\in v_\alpha[\mathscr{O}_\alpha^n]\Bigg\} < +\infty.$$

However, this statement in combination with (a) and the inequality $\|f\|_p\leqslant\|f\|_{p+1}$ for $f\in S_{p+1}(\cdots)$ implies (b).

Finally, f is from $\sigma^\infty(\cdots)$ if and only if the relation (6.12) holds for all $k\in\mathbb{N}^0$. Hence (e) is a simple consequence of (b) and the definition of $\mathscr{C}^\infty$ vectors in the various normed, respectively, countably normed spaces. □

With these preparations we are now able to formulate and prove the announced infinitesimal characterization of the subspace of $\mathscr{C}^\infty$ vectors for the representation $R^{\mathscr{A}_1,\dots,\mathscr{A}_L}$ of G.

Theorem 6.1: Let $X_1,\dots,X_r$ be a basis for the Lie algebra $\mathfrak{g}$ of G. Then for any subset $\{\mathscr{A}_1,\dots,\mathscr{A}_L\}\subseteq I_G$ and $L\in\mathbb{N}$ the following relations hold:

(i) $dR^{\mathscr{A}_1,\dots,\mathscr{A}_L}(X)=\overline{\partial R^{\mathscr{A}_1,\dots,\mathscr{A}_L}(X)},\quad X\in\mathfrak{g},$

(ii) $$\bigcap_{j\in\{1,\dots,r\}}\bigcap_{n\in\mathbb{N}} D(dR^{\mathscr{A}_1,\dots,\mathscr{A}_L}(X_j)^n) = \sigma^\infty(\mathscr{A}_1,\dots,\mathscr{A}_L) \subseteq \bigcap_{X\in\mathfrak{g}}\bigcap_{n\in\mathbb{N}} D(dR^{\mathscr{A}_1,\dots,\mathscr{A}_L}(X)^n).$$

Among other things the first part of (ii) together with (i) says that the subspace of $\mathscr{C}^\infty$ vectors for the representation $R^{\mathscr{A}_1,\dots,\mathscr{A}_L}$ of the Lie group G is equal to the intersection of the domains of the closures of its differential operators $\overline{\partial R^{\mathscr{A}_1,\dots,\mathscr{A}_L}(X_j)}$, $j\in\{1,\dots,r\}$ for a basis of the corresponding Lie algebra $\mathfrak{g}$ and all their (positive integer) powers. This will be the starting point for the integration of a representation of $\mathfrak{g}$. In passing we mention another interpretation of (ii) due to Goodman,[11] which can also be taken over to the domain $\mathscr{D}^\infty(U)$: f is a $\mathscr{C}^\infty$ vector for $R^{\mathscr{A}_1,\dots,\mathscr{A}_L}$ if and only if f is a common $\mathscr{C}^\infty$ vector for its restrictions $R^{\mathscr{A}_1,\dots,\mathscr{A}_L}\restriction G_k$, $k\in\{1,\dots,r\}$ to the one-parameter subgroups $G_k=\{\exp(tX_k)|\,t\in\mathbb{R}\}$ of G generated by the elements of a basis of $\mathfrak{g}$.

Proof (of Theorem 6.1): (i) Let $(f_n)_{n\in\mathbb{N}}$ be a $dR^{\cdots}(X)$-convergent sequence. Then there exist $f,h\in S(\cdots)$ such that for all $p\in\mathbb{N}^0$

$$\lim_{n\to\infty}\|f_n-f\|_p=0, \tag{6.13}$$

$$\lim_{n\to\infty}\|dR^{\cdots}(X)f_n-h\|_p=0. \tag{6.14}$$

From the definitions of the domains $D(dR^{\cdots}_{\cdots}(X))$ and Lemma 6.1(d) we deduce

$$D(dR^{\cdots}(X))=\bigcap_{p\in\mathbb{N}^0} D(dR_p^{\cdots}(X)), \tag{6.15}$$

and moreover for every $k\in D(dR^{\cdots}(X))$ and $p\in\mathbb{N}^0$

$$dR^{\cdots}(X)k=dR_p^{\cdots}(X)k. \tag{6.16}$$

By virtue of Lemma 6.1(f) the last four equations imply

$$f\in D(dR^{\cdots}(X))\quad\text{and}\quad h=dR^{\cdots}(X)f. \tag{6.17}$$

This proves the closedness of $dR^{\cdots}(X)$ and in view of the inclusion (6.6) also

$$dR^{\cdots}(X)\supseteq\overline{\partial R^{\cdots}(X)}. \tag{6.18}$$

In order to complete the proof of (i) it suffices to verify the following equality of the graphs:

$$\Upsilon(dR^{\cdots}(X))=\Upsilon(\overline{\partial R^{\cdots}(X)}). \tag{6.19}$$

Let $\overline{\Upsilon(\overline{\partial R^{\cdots}(X)})}^p$ denote the completion of $\Upsilon(\overline{\partial R^{\cdots}(X)})$ with respect to the (single) norm $\|\{\,;\,\}\|_p$. It is a Banach space.

Moreover, from

$$\Upsilon(\overline{\partial R^{\cdots}(X)})=\overline{\Upsilon(\partial R^{\cdots}(X))} \tag{6.20}$$

we trivially obtain

$$\overline{\Upsilon(\overline{\partial R^{\cdots}(X)})}^p=\overline{\Upsilon(\partial R^{\cdots}(X))}^p, \tag{6.21}$$

and thus for every $p\in\mathbb{N}^0$

$$\overline{\Upsilon(\partial R^{\cdots}(X))}\subseteq\overline{\Upsilon(\partial R^{\cdots}(X))}^{p+1}\subseteq\overline{\Upsilon(\partial R^{\cdots}(X))}^p. \tag{6.22}$$

Since $\overline{\Upsilon(\partial R^{\cdots}(X))}$ is a complete, countably normed space the relations (6.20) and (6.22) imply (Ref. 20, Chap. I, Sec. 3.2)

$$\Upsilon(\overline{\partial R^{\cdots}(X)})=\bigcap_{p\in\mathbb{N}^0}\overline{\Upsilon(\partial R^{\cdots}(X))}^p. \tag{6.23}$$

On the other hand, let $\Upsilon_p(T)$ denote the graph of the operator T in the Banach space $S_p(\cdots)$. Since all infinitesimal generators $dR^{\cdots}(X)$ and $dR_p^{\cdots}(X)$ are closed operators, their graphs are complete, countably normed, respectively, normed spaces which due to Lemma 6.1(d) and Eq. (6.15) satisfy the inclusion relations

$$\Upsilon(dR^{\cdots}(X))\subseteq\Upsilon_{p+1}(dR_{p+1}^{\cdots}(X))\subseteq\Upsilon_p(dR_p^{\cdots}(X)). \tag{6.24}$$

Hence once again we obtain from Chap. I, Sec. 3.2 in Ref. 20

$$\Upsilon(dR^{\cdots}(X))=\bigcap_{p\in\mathbb{N}^0}\Upsilon_p(dR^{\cdots}(X)). \tag{6.25}$$

$\sigma^\infty(\cdots)$ is a dense linear subspace of the Banach space $S_p(\cdots)$ and $dR_p^{\cdots}(X)$ is the closed infinitesimal generator of a strongly continuous one-parameter group which leaves $\sigma^\infty(\cdots)$ invariant. In addition, from Lemma 6.1 we obtain the two further properties

$$\sigma^\infty(\cdots)\subseteq\sigma_p^\infty(\cdots)\subseteq D(dR_p^{\cdots}(X)), \tag{6.26}$$

$$\partial R^{\cdots}(X)=dR_p^{\cdots}(X)\restriction\sigma^\infty(\cdots). \tag{6.27}$$

These statements together are just the presumptions of Corollary 1.3 in Ref. 27, from which it therefore follows:

$$\overline{\partial R^{\cdots}(X)}^p=\overline{dR_p^{\cdots}(X)\restriction\sigma^\infty(\cdots)}^p=dR_p^{\cdots}(X). \tag{6.28}$$

Translated into the corresponding graphs this means

$$\overline{\Upsilon(\partial R^{\cdots}(X))}^p=\Upsilon_p(dR_p^{\cdots}(X)). \tag{6.29}$$

Equations (6.23), (6.25), and (6.29) imply Eq. (6.19) and therefore statement (i).

(ii) In view of the relation (6.6) it suffices to show that

$$\bigcap_{j\in\{1,\dots,r\}}\bigcap_{n\in\mathbb{N}} D(dR^{\cdots}(X_j)^n)\subseteq\sigma^\infty(\cdots). \tag{6.30}$$

From Lemma 6.1(e) and the result of Goodman (Ref. 11, Theorem 1.1) in the case of Banach spaces we deduce

$$\sigma^\infty(\cdots)=\bigcap_{\substack{n\in\mathbb{N}\\ p\in\mathbb{N}^0}}\bigcap_{j\in\{1,\dots,r\}} D(dR_p^{\cdots}(X_j)^n). \tag{6.31}$$

If e denotes the unit element of G it follows by means of the continuity of $R^{\cdots}(g)$ that $f\in D(dR^{\cdots}(X_j)^n)\subseteq S(\cdots)$ if and only if

$$\lim_{t_1\to0}\lim_{t_2\to0}\cdots\lim_{t_n\to0}\left[\prod_{j=1}^n t_j^{-1}(R^{\cdots}(\exp t_jX)-R^{\cdots}(e))\right]f$$

exists in the norm of every Banach space $S_p(\cdots)$, $p\in\mathbb{N}^0$. In combination with the equation

$$R^{\cdots}(g)=R_p^{\cdots}(g)\restriction S(\cdots), \tag{6.32}$$

this implies $f\in D(dR_p^{\cdots}(X_j)^n)$ for every $p\in\mathbb{N}^0$, or equivalently,

$$D(dR^{\cdots}(X_j)^n)\subseteq\bigcap_{p\in\mathbb{N}^0} D(dR_p^{\cdots}(X_j)^n). \tag{6.33}$$

However, from the two relations (6.31) and (6.33) we just get (6.30). □

VII. THE POINCARÉ GROUP

As an explicit example for the general investigations we present a class of representations for the universal covering group $P^{\uparrow}_{+} = \mathrm{SL}(2,\mathbb{C}) \otimes \mathbb{R}_4$ of the proper orthochronous Poincaré group, which on any strictly localizable space $S^g(\mathbb{R}_4,\mathbb{C}_{\mathring{\mathcal{A}}})$ has all continuity and boundedness properties required in Definition 2.1. This class contains the (up to now) physically most important representations including the "usual" ones, for which the restrictions to the subgroup of translations $(\mathbb{R}_4,+)$ are all one dimensional.

We start with some standard notations: $P^{\uparrow}_{+}$ consists of pairs (α,a) with $a \in R_4$ and

$$\alpha = \begin{pmatrix} \alpha_1^1 & \alpha_1^2 \\ \alpha_2^1 & \alpha_2^2 \end{pmatrix} \in \mathrm{SL}(2,\mathbb{C}).$$

The composition law reads $(\alpha,a)(\beta,b) = (\alpha\beta, a + \Lambda(\alpha)b)$. Here $\Lambda(\alpha)$ is the real 4×4 matrix with elements

$$\Lambda(\alpha)^{\mu}{}_{\nu} = \frac{1}{2}\sum_{\lambda=0}^{3} g_{\mu\lambda}\, \mathrm{tr}(\tau^{\mu}\alpha\tau^{\lambda}\overline{\alpha}^{t}). \tag{7.1}$$

Here, τ^{μ}, $(\mu = 0,...,3)$ denote the Pauli matrices and

$$(g^{\mu\nu}) = (g_{\mu\nu}) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

the Minkowski metric tensor. Finally, $\overline{\alpha}^{t}$ is the complex conjugate and transposed of the matrix α. The topology of $P^{\uparrow}_{+}$ can be defined in terms of the following neighborhood basis of an element (α,a):

$$\mathfrak{N}(\alpha,a) := \{E_r(\alpha) \times F_n(a) \mid r,n \in \mathbb{N}\}, \tag{7.2}$$

with

$$\begin{aligned} E_r(\alpha) &:= \left\{\beta \in \mathrm{SL}(2,\mathbb{C}) \,\Big|\, \left(\sum_{i,j=1}^{2} |\alpha_j^i - \beta_j^i|^2\right)^{1/2} < \frac{1}{r}\right\}, \\ F_n(a) &:= \left\{b \in \mathbb{R}_4 \,\Big|\, \left(\sum_{i=0}^{3} |a^i - b^i|^2\right)^{1/2} < \frac{1}{n}\right\}. \end{aligned} \tag{7.3}$$

Let $M^{\mathcal{A}}$ denote an $\mathring{\mathcal{A}}$-dimensional matrix representation of $\mathrm{SL}(2,\mathbb{C})$ and $T^{\mathcal{A}}$ a representation of $(\mathbb{R}_4,+)$ of the form

$$T_0^{\mathcal{A}}(a\cdot\partial_x) := W(\mathcal{A})\left[\bigoplus_{i=1}^{k} t_0^{\tau_i}(a\cdot\partial_x)\right]W(\mathcal{A})^{-1}, \tag{7.4}$$

with $\Sigma_{i=1}^{k}\tau_i = \mathring{\mathcal{A}}$. The matrices $t_0^{\tau}(a\cdot\partial_x)$ in the direct sum are explicitly given in Eq. (2.5) with $q=0$ and $W(\mathcal{A})$ is an $\mathring{\mathcal{A}}$-dimensional nonsingular matrix independent of a. Now by straightforward calculations one easily verifies that the mappings

$$R^{\mathcal{A}}(\alpha,a)\colon S^g(\mathbb{R}_4,\mathbb{C}_{\mathring{\mathcal{A}}}) \to S^g(\mathbb{R}_4,\mathbb{C}_{\mathcal{A}}), \quad f \to R^{\mathcal{A}}(\alpha,a)f, \tag{7.5}$$

with

$$(R^{\mathcal{A}}(\alpha,a)f)^{\mu}(x) := \sum_{\rho,\nu\in \mathbb{T}_{P^{\uparrow}_{+}}(\mathcal{A})} T_0^{\mathcal{A}}(a\cdot\partial_x)^{\mu}{}_{\rho} M^{\mathcal{A}}(\alpha)^{\rho}{}_{\nu} f^{\nu}(\Lambda(\alpha^{-1})(x-a)), \tag{7.6}$$

define a representation of $P^{\uparrow}_{+}$ on $S^g(\mathbb{R}_4,\mathbb{C}_{\mathring{\mathcal{A}}})$ if and only if for all $\alpha\in\mathrm{SL}(2,\mathbb{C})$ and $a\in\mathbb{R}_4$ we have

$$M^{\mathcal{A}}(\alpha)T_0^{\mathcal{A}}(a\cdot\partial_x) = T_0^{\mathcal{A}}(a\cdot\partial_x)M^{\mathcal{A}}(\alpha). \tag{7.7}$$

Hence unless $T_0^{\mathcal{A}}(a\cdot\partial_x)$ is the unit matrix $(\tau_1 = \tau_2 = \cdots = \tau_{\mathcal{A}} = 1$; "usual" representation) the representation $M^{\mathcal{A}}$ has to be reducible. Thus its general form reads

$$M^{\mathcal{A}}(\alpha) = B(\mathcal{A})\left[\bigoplus_{i=1}^{r} S_i(\alpha^{-1t})D^{n_i}(\alpha^{-1t})\right]B(\mathcal{A})^{-1}, \tag{7.8}$$

with $\Sigma_{i=1}^{r} n_i = \mathring{\mathcal{A}}$. Here again $B(\mathcal{A})$ is an $\mathring{\mathcal{A}}$-dimensional nonsingular matrix independent of α; D^n denotes an n-dimensional irreducible matrix representation[28] and S_i a one-dimensional representation of $\mathrm{SL}(2,\mathbb{C})$. The only important point for the following is that the elements of the matrices $D^n(\alpha)$ are polynomials in the parameters α_k^i, $(i,k = 1,2)$ (Ref. 28 and Ref. 29, Chaps. 10.8 and 10.16). Therefore $M^{\mathcal{A}}(\alpha)$ is continuous in the topology above if the functions $S_i(\alpha)$ $(i = 1,...,r)$ are continuous. The latter will be assumed in the following without further comment.

Theorem 7.1: Every homomorphism $R^{\mathcal{A}}$ $: P^{\uparrow}_{+} \to \mathrm{Aut}\, S^g(\mathbb{R}_4;\mathbb{C}_{\mathring{\mathcal{A}}})$, $(\alpha,a) \mapsto R^{\mathcal{A}}(\alpha,a)$ defined in Eqs. (7.4)–(7.8) is a continuous representation of $P^{\uparrow}_{+}$.

Proof: (1) In the first step we show that $R^{\mathcal{A}}(\alpha,a)$ is a continuous automorphism of $S^g(\mathbb{R}_4,\mathbb{C}_{\mathring{\mathcal{A}}})$.

Using the abbreviations

$$\begin{aligned} K(\alpha) &= \mathring{\mathcal{A}}^2 \max\{|M(\alpha)^{\mu}{}_{\nu}| \,|\, \mu,\nu \in \mathbb{T}_{P^{\uparrow}_{+}}(\mathcal{A})\}, \\ Z &= \mathring{\mathcal{A}}^2[\max\{|W(\mathcal{A})^{\mu}{}_{\nu}|, \\ &\qquad |(W(\mathcal{A})^{-1})^{\mu}{}_{\nu}| \,|\, \mu,\nu\in\mathbb{T}_{P^{\uparrow}_{+}}(\mathcal{A})\}]^2, \end{aligned} \tag{7.9}$$

$$D^{|r|}(p) = \frac{\partial^{r^0+\cdots+r^3}}{(\partial p^0)^{r^0}\cdots(\partial p^3)^{r^3}}, \tag{7.10}$$

we deduce via the Leibniz rule and tedious but elementary estimates from Eqs. (2.1) and (7.4)–(7.7) and the explicit matrices (2.5) (with $q=0$) the following bound:

$$\begin{aligned} &\|R^{\mathcal{A}}(\alpha,a)f\|^{g}_{m,\rho} \\ &\leqslant K(\alpha)\cdot Z\cdot \sup\Bigg\{g\left(\rho\sum_{t=0}^{3}(p^t)^2\right) \\ &\quad\times\left[\prod_{j=0}^{3}(1+|p^j|)^m(1+|a^j|)^m\right] \\ &\quad\times(2\mathring{\mathcal{A}})^m(1+|(a\cdot p)|)^{\mathring{\mathcal{A}}-1}|D^{|r|}(p) \\ &\quad\times\int d^4x\, f^{\mu}(\Lambda(\alpha^{-1}(x-a)))\exp[-i(p\cdot x)]\,|\,|p\in\mathbb{R}_4; \\ &\quad r^0,...,r^3\in\mathbb{N}^0, \sum_{s=0}^{3} r^s \leqslant m; \mu\in\mathbb{T}_{P^{\uparrow}_{+}(\mathcal{A})}\Bigg\}. \end{aligned} \tag{7.11}$$

The application of an obvious change of variables, the Leibniz rule, and the chain rule leads to the further estimate

$$\begin{aligned} &\max\Bigg\{|D^{|r|}(p)\int d^4x\, f^{\mu}(\Lambda(\alpha^{-1})(x-a)) \\ &\quad\times\exp[-i(p\cdot x)]\,|\,|\,r^0,...,r^3\in\mathbb{N}^0, \sum_{j=0}^{3} r^j \leqslant m\Bigg\} \\ &\leqslant d(\alpha)^m\left[\prod_{s=0}^{3}((1+|a^s|)^m\right]\max\Bigg\{|D^{|r|}(q(p)) \\ &\quad\times\int d^4x\, f^{\mu}(x)\exp[-i(q(p)\cdot x)]\,|\,| \\ &\quad r^0,...,r^3\in\mathbb{N}^0, \sum_{j=0}^{3} r^j\leqslant m\Bigg\}. \end{aligned} \tag{7.12}$$

Here we have used the shorthand notations $q(P) = \Lambda(\alpha^{-1})P$ and

$$d(\alpha) = 4\max\{1, \Lambda(\alpha^{-1})^{\mu}{}_{\nu} | \mu,\nu \in \mathbf{T}_{P^{\uparrow}_{+}}(\mathscr{A})\}. \tag{7.13}$$

Feeding the inequality (7.12) back into (7.11) and observing the inequalities

$$(1 + |(a\cdot p)|)^{\dot{\mathscr{A}}} \leq \prod_{j=0}^{3} (1 + |p^j|)^{\dot{\mathscr{A}}} (1 + |a^j|)^{\dot{\mathscr{A}}}, \tag{7.14a}$$

$$(1 + |(\Lambda(\alpha)q)^j|^m \leq d(\alpha^{-1})^m \prod_{i=0}^{3} (1 + |q^i|)^m, \tag{7.14b}$$

$$[(\Lambda(\alpha)q)^j]^2 \leq 4d(\alpha^{-1})^2 \sum_{\nu=0}^{3} (q^{\nu})^2, \tag{7.14c}$$

then by means of the monotonic growth of g we obtain the simple estimate

$$\|R^{\mathscr{A}}(\alpha,a)f\|^{g}_{m,\rho} \leq (2\mathring{\mathscr{A}})^m \left[d(\alpha^{-1})^4 d(\alpha) \prod_{i=0}^{3} (1 + |a^i|)^2 \right]^{m+\dot{\mathscr{A}}-1} \times Z\cdot K(\alpha) \|f\|^{g}_{4(m+\dot{\mathscr{A}}),[\alpha]\rho}. \tag{7.15}$$

Here $[\alpha]$ denotes the smallest natural number larger or equal to $16d(\alpha^{-1})^2$. The relation (7.15) plainly implies the automorphism $R^{\mathscr{A}}(\alpha,a)$ to be a continuous one.

(2) In order to prove the continuity of $R^{\mathscr{A}}(\alpha,a)f$ as a function of the group elements (α,a) it suffices to show that

$$\lim_{(\alpha,a)\to(\epsilon,0)} \|R^{\mathscr{A}}(\alpha,a)f - f\|^{g}_{m,\rho} = 0,$$
$$\text{for all } f \in S^{g}(\mathbf{R}_4, \mathbf{C}_{\dot{\mathscr{A}}}) \quad \text{and} \quad (m,\rho) \in \mathbf{N}^0 \times \mathbf{N}.$$

Introducing the shorthand notation

$$f_{(\alpha,a)}(x) = f(\Lambda(\alpha^{-1})(x-a)), \tag{7.16}$$

it trivially follows from Eqs. (7.4) and (7.7)

$$\begin{aligned} \|R^{\mathscr{A}}(\alpha,a)f - f\|^{g}_{m,\rho} &\leq \|M^{\mathscr{A}}(\alpha)T^{\mathscr{A}}_0(a\cdot\partial_x)f_{(\alpha,a)} - T^{\mathscr{A}}_0(a\cdot\partial_x)f_{(\alpha,a)}\|^{g}_{m,\rho} \\ &\quad + \|T^{\mathscr{A}}_0(a\cdot\partial_x)f_{(\alpha,a)} - f_{(\alpha,a)}\|^{g}_{m,\rho} + \|f_{(\alpha,a)} - f\|^{g}_{m,\rho}. \end{aligned} \tag{7.17}$$

We are going to show that every term on the right-hand side of (7.17) vanishes in the limit $(\alpha,a)\to(\epsilon,0)$. From the explicit expression (2.1) for the norms we get at once

$$\|M^{\mathscr{A}}(\alpha)T^{\mathscr{A}}_0(a\cdot\partial_x)f_{(\alpha,a)} - T^{\mathscr{A}}_0(a\cdot\partial_x)f_{(\alpha,a)}\|^{g}_{m,\rho} \leq \|T^{\mathscr{A}}_0(a\cdot\partial_x)f_{(\alpha,a)}\|^{g}_{m,\rho} \times \sum_{\mu,\nu\in \mathbf{T}_{P^{\uparrow}_{+}}(\mathscr{A})} |M^{\mathscr{A}}(\alpha)^{\mu}{}_{\nu} - M^{\mathscr{A}}(\epsilon)^{\mu}{}_{\nu}|. \tag{7.18}$$

On the other hand, by literally the same arguments as applied in the derivation of the inequality (7.15) we obtain

$$\|T^{\mathscr{A}}_0(a\cdot\partial_x)f_{(\alpha,a)}\|^{g}_{m,\rho} \leq \left[d(\alpha^{-1})^4 d(\alpha) \prod_{i=0}^{3} (1 + |a^i|)^2 \right]^{m+\dot{\mathscr{A}}-1} \times (2\mathring{\mathscr{A}})^m Z \|f\|^{g}_{4(m+\dot{\mathscr{A}}),[\alpha]\rho}. \tag{7.19}$$

Consider the arbitrary but fixed neighborhood

$$\mathscr{N}_{\kappa,\delta}(\epsilon,0) = \left\{ (\alpha,a) \in P^{\uparrow}_{+} \,\Big|\, \left(\sum_{i,j=1}^{2} |\alpha^i_j - \epsilon^i_j|^2 \right)^{1/2} < \kappa, \left(\sum_{\nu=0}^{3} |a^{\nu}|^2 \right)^{1/2} < \delta \right\}. \tag{7.20}$$

Then from Eqs. (7.13) and (7.1) it follows at once for all $(\alpha,a) \in \mathscr{N}_{\kappa,\delta}(\epsilon,0)$,

$$d(\alpha) \leq 32(1+\kappa)^2, \quad d(\alpha^{-1}) \leq 32(1+\kappa)^2, \tag{7.21}$$

and moreover if $n(\kappa)$ denotes a natural number larger than $[16(1+\kappa)]^4$,

$$\|f\|^{g}_{4(m+\dot{\mathscr{A}}),[\alpha]\rho} \leq \|f\|^{g}_{4(m+\dot{\mathscr{A}}),n(\kappa)\rho}. \tag{7.22}$$

Now the upper bounds (7.19), (7.21), and (7.22) imply the first factor on the right-hand side of the inequality (7.18) to be a bounded function on $N_{\kappa,\delta}(\epsilon,0)$. Therefore from the continuity of $M^{\mathscr{A}}(\alpha)^{\mu}{}_{\nu}$ in α we deduce

$$\lim_{(\alpha,a)\to(\epsilon,0)} \|M^{\mathscr{A}}(\alpha)T^{\mathscr{A}}_0(a\cdot\partial_x)f_{(\alpha,a)} - T^{\mathscr{A}}_0(a\cdot\partial_x)f_{(\alpha,a)}\|^{g}_{m,\rho} = 0. \tag{7.23}$$

It is obvious from Eqs. (7.4) and (2.5) that in the difference $T^{\mathscr{A}}_0(a\cdot\partial_x) - T^{\mathscr{A}}_0(0\cdot\partial_x)$ there do not occur any nonzero elements independent of a. This observation leads along the same lines as before to an improved upper bound compared to (7.19)

$$\begin{aligned} \|T^{\mathscr{A}}_0(a\cdot\partial_x)f_{(\alpha,a)} - f_{(\alpha,a)}\|^{g}_{m,\rho} &\leq \max\{|a^{\mu}| \,|\, \mu = 0,\dots,3\}(2\mathring{\mathscr{A}})^m Z \Big[d(\alpha^{-1})^4 d(\alpha) \\ &\quad \times \prod_{i=0}^{3} (1 + |a^i|^2) \Big]^{m+\dot{\mathscr{A}}-1} \|f\|^{g}_{4(m+\dot{\mathscr{A}}),[\alpha]\rho}. \end{aligned} \tag{7.24}$$

Hence, due to the first factor on the right-hand side it follows by means of the inequalities (7.21) and (7.22)

$$\lim_{(\alpha,a)\to(\epsilon,0)} \|T^{\mathscr{A}}_0(a\cdot\partial_x)f_{(\alpha,a)} - f_{(\alpha,a)}\|^{g}_{m,\rho} = 0. \tag{7.25}$$

There remains the third term on the right-hand side of the inequality (7.17). Let $(\alpha_n, a_n)_{n\in\mathbf{N}}$ be an arbitrary sequence which converges in $P^{\uparrow}_{+}$ to the unit element $(\epsilon,0)$. Then from the continuity of $f(x)$ and $\Lambda(\alpha)$ we deduce that the sequence $(f_{(\alpha_n,a_n)} - f)_{n\in\mathbf{N}}$ converges in $\mathbf{C}_{\dot{\mathscr{A}}}$ pointwise to zero; this means for every fixed $x \in \mathbf{R}_4$ we have in the Euclidean topology of $\mathbf{C}_{\dot{\mathscr{A}}}$

$$\lim_{n\to\infty} (f(\Lambda(\alpha_n^{-1})(x - a_n)) - f(x)) = 0. \tag{7.26}$$

Once again by literally the same arguments as applied in the derivation of the upper bound (7.15) we obtain

$$\|f_{(\alpha,a)}\|^{g}_{m,\rho} \leq \left[d(\alpha^{-1})^4 d(\alpha) \prod_{i=1}^{3} (1 + |a^i|) \right]^m \|f\|^{g}_{4m,[\alpha]\rho}. \tag{7.27}$$

Hence in combination with the relations (7.20)–(7.22) this implies that the set $\{f_{(\alpha_n,a_n)} - f \,|\, n \in \mathbf{N}\}$ is a bounded subset of $S^{g}(\mathbf{R}_4, \mathbf{C}_{\dot{\mathscr{A}}})$. According to our presumption (S.0) in Sec. II it is relatively sequentially compact. Therefore our sequence has at least one limit point in the topology of $S^{g}(\mathbf{R}_4, \mathbf{C}_{\dot{\mathscr{A}}})$. Assume it possesses a limit point $h \neq 0$. Then there exists a subsequence $(f_{(\alpha_{\nu(n)}, a_{\nu(n)})} - f)_{n\in\mathbf{N}}$ such that for all $(m,\rho) \in \mathbf{N}^0 \times \mathbf{N}$,

$$\lim_{n\to\infty} \| f_{(\alpha_{\nu(n)},a_{\nu(n)})} - f - h \|_{m,\rho} = 0, \tag{7.28}$$

in contradiction to Eq. (7.26), since convergence in the topology of $S^g(\mathbb{R}_4,\mathbb{C}_{\mathscr{A}})$ implies pointwise convergence in $\mathbb{C}_{\mathscr{A}}$. Thus we obtain for all $(m,\rho)\in \mathbb{N}^0\times\mathbb{N}$

$$\lim_{(\alpha,a)\to(\epsilon,0)} \| f_{(\alpha,a)} - f \|_{m,\rho} = 0. \tag{7.29}$$

VIII. CONFINEMENT

By now we have gained the necessary insight into the structure of unbounded representations for symmetry groups and especially the Poincaré group in order to present the details of the confinement mechanism which we shortly indicated in comment (iv) to axiom A.III. Throughout this section we assume that the test function space is a Jaffe space and the basic fields transform under translations with the representations

$$\begin{aligned} T_q^{\mathscr{T}}(y\cdot\partial_x) &= W(\mathscr{T})\left[\bigoplus_{i=1}^{k} t_q^{\tau_i}(y\cdot\partial_x)\right]W(\mathscr{T})^{-1} \qquad (8.1)\\ &= [\exp(y\cdot q(\mathscr{T}))]\,W(\mathscr{T})\left[\bigoplus_{i=1}^{k} t_0^{\tau_i}(y\cdot\partial_x)\right]W(\mathscr{T})^{-1}\\ &= [\exp(y\cdot q(\mathscr{T}))]\,T_0^{\mathscr{T}}(y\cdot\partial_x), \end{aligned}$$

where $t_q^{\tau}(y\cdot\partial_x)$ is explicitly given in Eq. (2.5) and $\mathring{\mathscr{T}}=\Sigma_{i=1}^{k}\tau_i$. We denote by $q_R(\mathscr{T})$ the real part of the complex quadruple $q(\mathscr{T})$. Plainly, we assume $q(\mathscr{O})=0$ for the trivial field $\varphi_{\mathscr{O}}(f)=\mathrm{id}_{\mathscr{H}}\int d^4x f(x)$.

Theorem 8.1: (Confinement): For any subset $\{\mathscr{T}_1,\ldots,\mathscr{T}_L\}\subseteq I_T$ with the property $\Sigma_{i=1}^{L} q_R(\mathscr{T}_i)\neq 0$ we have

$$\forall f\in S^g(\mathscr{T}_1,\ldots,\mathscr{T}_L),$$
$$\phi^{\mathscr{T}_1,\ldots,\mathscr{T}_L}(f)\Psi_0\in\mathscr{H}_0 := \{\Psi\in\mathscr{H}\,|\,\langle\Psi,\Psi\rangle=0\}\supset H_0.$$

Proof: If $|\eta|_{\mathrm{OP}}$ denotes the operator norm of η and $\widetilde{T}$ the representation of Theorem 3.2 corresponding to T of the translation group, it follows from the η isometry of $\widetilde{T}$, the substitution rule Theorem 3.2(1), and Eq. (8.1) for all $y\in\mathbb{R}_4$

$$\begin{aligned} &|\langle\phi^{\mathscr{T}_1,\ldots,\mathscr{T}_L}(f)\Psi_0,\phi^{\mathscr{T}_1,\ldots,\mathscr{T}_L}(f)\Psi_0\rangle|\\ &= \exp\left[2\sum_{i=1}^{L}(y\cdot q_R(\mathscr{T}_i))\right]\Big|\langle\phi^{\mathscr{T}_1,\ldots,\mathscr{T}_L}\\ &\quad\times(\bar{\otimes}_{i=1}^{L} T_0^{\mathscr{T}_i}(y\cdot\partial_{x_i})f)\Psi_0,\phi^{\mathscr{T}_1,\ldots,\mathscr{T}_L}\\ &\quad\times(\bar{\otimes}_{j=1}^{L} T_0^{\mathscr{T}_j}(y\cdot\partial_{x_j})f)\Psi_0\rangle\Big|\\ &\leqslant |\eta|_{\mathrm{OP}}\exp\left[2\sum_{i=1}^{L}(y\cdot q_R(\mathscr{T}_i))\right]\|\phi^{\mathscr{T}_1,\ldots,\mathscr{T}_L}\\ &\quad\times(\bar{\otimes}_{i=1}^{L} T_0^{\mathscr{T}_i}(y\cdot\partial_{x_i})f)\Psi_0\|^2_{\mathscr{H}}. \end{aligned} \tag{8.2}$$

Since the Hilbert space norm is a continuous linear functional on $S^g(\mathscr{T}_L^*,\ldots,\mathscr{T}_1^*,\mathscr{T}_1,\ldots,\mathscr{T}_L)$ there exist a pair $(m_0,\rho_0)\in\mathbb{N}^0\times\mathbb{N}$ and a positive real number $\kappa=\kappa(\mathscr{T}_1,\ldots,\mathscr{T}_L)$ such that

$$\begin{aligned} &\left\|\phi^{\mathscr{T}_1,\ldots,\mathscr{T}_L}\left(\bar{\otimes}_{i=1}^{L} T_0^{\mathscr{T}_i}(y\cdot\partial_{x_i})f\right)\Psi_0\right\|^2_{\mathscr{H}}\\ &\leqslant\kappa\left\|\mathscr{C}_L\left[\bar{\oplus}_{i=1}^{L} T_0^{\mathscr{T}_i}(y\cdot\partial_{x_i})f\right]\right\|^g_{m_0,\rho_0}\\ &\quad\times\left\|\bar{\otimes}_{j=1}^{L} T_0^{\mathscr{T}_j}(y\cdot\partial_{x_j})f\right\|^g_{m_0,\rho_0}\\ &=\kappa\left(\left\|\bar{\otimes}_{i=1}^{L} T_0^{\mathscr{T}_i}(y\cdot\partial_{x_i})f\right\|^g_{m_0,\rho_0}\right)^2. \end{aligned} \tag{8.3}$$

In the last step we have used the isometry of the conjugation operator $\mathscr{C}_L$. Finally, by literally the same arguments as applied in the derivation of the inequality (7.15) we obtain the upper bound

$$\begin{aligned} &\left\|\bar{\otimes}_{i=1}^{L} T_0^{\mathscr{T}_i}(y\cdot\partial_{x_i})f\right\|^g_{m_0,\rho_0}\\ &\quad\leqslant Z^L 2^{mL}\|f\|^g_{4(Lm_0+|\mathring{\mathscr{T}}|),\rho}\left[\prod_{i=1}^{3}(1+|y^i|^2)\right]^{Lm_0+|\mathring{\mathscr{T}}|},\\ &|\mathring{\mathscr{T}}|=\sum_{i=1}^{L}\mathring{\mathscr{T}}_i. \end{aligned} \tag{8.4}$$

From the three inequalities above, Theorem 8.1 follows immediately by taking y to infinity in a suitable direction. □

Theorem 8.1 offers a sufficient condition for states of $\mathscr{D}_{\mathrm{QL}}$ to be from $\mathscr{H}_0$, respectively, a necessary condition to be from $\mathscr{H}\setminus\mathscr{H}_0$.

If we take $\mathscr{T}_1=\mathscr{T}_2=\cdots=\mathscr{T}_L=\mathscr{T}$ then in case $q_R(\mathscr{T})\neq 0$ Theorem 8.1 says that neither $\phi^{\mathscr{T}}(h)$ for any $h\in S^g(\mathbb{R}_4,\mathbb{C}_{\mathscr{T}})$ nor any polynomial of it can create states with nonzero η norm from the vacuum. On the other hand, due to the completeness axiom A.IV the products $\Pi_{i=1}^{N}\phi^{\mathscr{T}_i}(h_i)$ of different field operators with $q_R(\mathscr{T}_i)\neq 0$ but $\Sigma_{i=1}^{L} q_R(\mathscr{T}_i)=0$ create states from the vacuum which are in $\mathscr{H}\setminus\mathscr{H}_0$ and hence lead to observable states. Of course if $q(\mathscr{T})\neq 0$ for some $\mathscr{T}\in I_T$ then the corresponding field $\phi^{\mathscr{T}}$ cannot be Poincaré covariant. Hence Poincaré symmetry in the sense of Definition 2.1 cannot hold in $\mathscr{H}_0$ and therefore not in the entire Hilbert space $\mathscr{H}$. However, in order to save the Poincaré invariance in the physical Hilbert space $\mathscr{H}_{\mathrm{ph}}=\overline{H/H_0}$ it suffices that Lorentz symmetry holds in the sense of Definition 2.1 and in addition Poincaré symmetry on the linear subspace

$$\begin{aligned} \mathscr{D}_C := \mathrm{LH}\Big\{&\phi^{\mathscr{T}_1,\ldots,\mathscr{T}_L}(f)\Psi_0\,|\,f\in S^g(\mathscr{T}_1,\ldots,\mathscr{T}_L);\\ &\mathscr{T}_1,\ldots,\mathscr{T}_L\in I_T;\\ &L\in\mathbb{N}\text{ and }\sum_{i=1}^{L} q(\mathscr{T}_i)=0\Big\}, \end{aligned} \tag{8.5}$$

which is dense in $\mathscr{H}\setminus\mathscr{H}_0$. To be precise, in order to describe a gauge quantum field theory with confinement of some or all basic fields and Poincaré invariance in the physical Hilbert space $\mathscr{H}_{\mathrm{ph}}$ we replace the axiom A.III by the following one, in which the representations $T_q^{\mathscr{T}}$ of the translation group are explicitly given by Eqs. (8.1) and (2.5).

A.III.C: *Confinement and Poincaré quasisymmetry:* There exist η-isometric representations T of $(\mathbb{R}_4,+)$ and $\mathscr{L}$ of SL(2,C) on a dense linear subspace $D_P\supseteq\mathscr{D}_{\mathrm{QL}}$ which leave $\mathscr{D}_{\mathrm{QL}}\cap H$ invariant and share the following properties.

(1) (Invariance of the Vacuum): There exists a unique vacuum state Ψ_0 such that $\langle\Psi_0,\Psi_0\rangle = 1$ and

$$\forall y \in \mathbb{R}_4, \quad T(y)\Psi_0 = \Psi_0$$
$$\forall \alpha \in \mathrm{SL}(2,\mathbb{C}), \quad \mathscr{L}(\alpha)\Psi_0 = \Psi_0. \tag{8.6}$$

(2) (Covariance): There exists a decomposition of $\mathbb{T}$ into a countable union $\mathbb{T} = \cup_{\mathscr{T} \in I_p} \mathbb{T}_{I_p}(\mathscr{T}) \times \{\mathscr{T}\}$ of pairwise disjoint finite subsets $\mathbb{T}_{I_p}(\mathscr{T}) \times \{\mathscr{T}\}$ and for every $\mathscr{T} \in I_p$ a $\mathring{\mathscr{T}}$-dimensional matrix representation $M^{\mathscr{T}}$ of SL(2,$\mathbb{C}$), respectively, $T_q^{\mathscr{T}}(y\cdot\partial_x) = \exp[(y\cdot q(\mathscr{T}))]T_0^{\mathscr{T}}(y\cdot\partial_x)$ of $(\mathbb{R}_4, +)$ on $S^{\mathring{\mathscr{T}}}(\mathbb{R}_{4,\mathscr{T}})$ with $\mathring{\mathscr{T}} = \mathring{\mathscr{T}}^*$ and

$$M^{\mathscr{T}}(\alpha)T_0^{\mathscr{T}}(y\cdot\partial_x) = T_0^{\mathscr{T}}(y\cdot\partial_x)M^{\mathscr{T}}(\alpha), \tag{8.7}$$

such that for all $\mathscr{T}_i \in I_p$, $(i = 1,...,L)$, $L \in \mathbb{N}$, the substitution rules hold:

$$T(y)\prod_{i=1}^{L}\phi^{\mathscr{T}_i}(f_i)\Psi_0 = \prod_{i=1}^{L}\phi^{\mathscr{T}_i}(T_q^{\mathscr{T}_i}(y\cdot\partial_{x_i})f_{i,(\epsilon,y)})\Psi_0, \tag{8.8}$$

$$\mathscr{L}(\alpha)\prod_{i=1}^{L}\phi^{\mathscr{T}_i}(f_i)\Psi_0 = \prod_{i=1}^{L}\phi^{\mathscr{T}_i}(M^{\mathscr{T}_i}(\alpha)f_{i,(\alpha,0)})\Psi_0, \tag{8.9}$$

with $q = q(\mathscr{T}) \in \mathbb{C}_4$, $f_{i,(\alpha,y)}(x) := f_i(\Lambda(\alpha^{-1})(x-y))$ and ϵ the unit element of SL(2,$\mathbb{C}$).

Now the Poincaré invariance in the physical Hilbert space $\mathscr{H}_{\mathrm{ph}}$ is easily verified.

Theorem 8.2: Let $\tilde{T}$ and $\tilde{\mathscr{L}}$ denote the representations on $\mathscr{D}_{\mathrm{QL}}$ of Theorem 3.2 corresponding to T, respectively, $\mathscr{L}$. Then there exists an η-isometric representation U of the proper orthochronous Poincaré group $P_+^{\uparrow}$ on $\mathscr{D}_C$

$$U: \quad P_+^{\uparrow} \to \mathrm{Aut}\,\mathscr{D}_C, \quad (\alpha,y) \mapsto U(\alpha,y) = \tilde{T}(y)\tilde{\mathscr{L}}(\alpha),$$

which leaves Ψ_0 invariant and lifts to a strongly continuous unitary representation $\mathscr{U}$ of $P_+^{\uparrow}$ on $\mathscr{H}_{\mathrm{ph}}$.

Proof: Since $\tilde{T}$ and $\tilde{\mathscr{L}}$ are representations of $(\mathbb{R}_4, +)$, respectively, SL(2,$\mathbb{C}$) on $\mathscr{D}_{\mathrm{QL}}$ the linear mapping

$$\tilde{U}(\alpha,y): \mathscr{D}_{\mathrm{QL}} \to \mathscr{D}_{\mathrm{QL}}, \quad \Psi \to \tilde{T}(y)\tilde{\mathscr{L}}(\alpha)\Psi \tag{8.10}$$

is for every $y \in \mathbb{R}_4$ and $\alpha \in \mathrm{SL}(2,\mathbb{C})$ an automorphism of $\mathscr{D}_{\mathrm{QL}}$ and by virtue of Theorem (3.2)(2) strongly continuous in (α, y). Explicitly we obtain from Theorem 3.1 and Theorem 3.2

$$\tilde{U}(\alpha,y)\Psi_0 = \Psi_0, \tag{8.11}$$

$$\tilde{U}(\alpha,y)\phi^{\mathscr{T}_1,...,\mathscr{T}_L}(f)\Psi_0 = \exp\left[\sum_{i=1}^{L}(y\cdot q(\mathscr{T}_i))\right]\phi^{\mathscr{T}_1,...,\mathscr{T}_L}\times\left(\bigotimes_{j=1}^{L}\left[T_0^{\mathscr{T}_j}(y\cdot\partial_{x_j})M^{\mathscr{T}_j}(\alpha)\right]f_{(\alpha,y)}\right)\Psi_0. \tag{8.12}$$

According to Theorems 7.1, 3.2, and 3.1, without the exponential factor on the right-hand side of Eq. (8.12), $\tilde{U}$ would be an η-isometric representation of $P_+^{\uparrow}$ on $\mathscr{D}_{\mathrm{QL}}$. Therefore its restriction to $\mathscr{D}_C$

$$U: \quad P_+^{\uparrow} \to \mathrm{Aut}\,\mathscr{D}_C, \quad (\alpha,y) \mapsto U(\alpha,y) = \tilde{U}(\alpha,y) \upharpoonright \mathscr{D}_C \tag{8.13}$$

is an η-isometric representation of $P_+^{\uparrow}$ on $\mathscr{D}_C$ and strongly continuous in (α, y).

Consider the vector spaces $B^{\mathrm{QL}} := \mathscr{D}_{\mathrm{QL}} \cap H$ and $B_0^{\mathrm{QL}} := \mathscr{D}_{\mathrm{QL}} \cap H_0$. Then obviously B_0^{QL} is invariant under every $\tilde{U}(\alpha,y)$ and B^{QL} can be represented as the vector space sum

$$B^{\mathrm{QL}} = (\mathscr{D}_C \cap H) \dotplus B_0^{\mathrm{QL}}. \tag{8.14}$$

Let the elements and the natural scalar product of the factor space $B^{\mathrm{QL}}/B_0^{\mathrm{QL}}$ be denoted by $[\Psi]_{\mathrm{QL}} = \Psi + B_0^{\mathrm{QL}}$, respectively, $([\cdot]_{\mathrm{QL}},[\cdot]_{\mathrm{QL}})_{\mathrm{QL}} := \langle\cdot,\cdot\rangle$. Due to Eq. (8.14), for every $\Psi \in B^{\mathrm{QL}}$ there exist $\Psi_C \in (\mathscr{D}_C \cap H)$ and $\theta \in B_0^{\mathrm{QL}}$ with $\Psi = \Psi_C + \theta$ and therefore $[\Psi]_{\mathrm{QL}} = [\Psi_C]_{\mathrm{QL}}$. Hence in view of the η isometry of U the mappings

$$[\Psi]_{\mathrm{QL}} \to W(\alpha,y)[\Psi]_{\mathrm{QL}} := [\tilde{U}(\alpha,y)\Psi]_{\mathrm{QL}} = [U(\alpha,y)\Psi_C]_{\mathrm{QL}}$$

generate an isometric, strongly continuous representation W: $P_+^{\uparrow} \to \mathrm{Aut}\,B^{\mathrm{QL}}/B_0^{\mathrm{QL}}, (\alpha, y) \mapsto W(\alpha,y)$ of the Poincaré group on the pre-Hilbert space $(B^{\mathrm{QL}}/B_0^{\mathrm{QL}};(\cdot,\cdot)_{\mathrm{QL}})$ which has a unique unitary extension $\overline{W}$ onto the Hilbert space $\overline{B^{\mathrm{QL}}/B_0^{\mathrm{QL}}}$. Moreover, in literally the same way as in the proof of Theorem 4.2(ii) it follows that $\overline{W}$ is strongly continuous. Finally, according to Theorem 3.3(α) the Hilbert space $\overline{B^{\mathrm{QL}}/B_0^{\mathrm{QL}}}$ is unitarily equivalent to $\mathscr{H}_{\mathrm{ph}}$. If ρ denotes the corresponding unitary mapping then the representation $\mathscr{U}$ is just given by $\rho\overline{W}\rho^{-1}$. □

Finally, let us make some short remarks about the application of the confinement mechanism above to quantum chromodynamics. In this model in addition to the Poincaré quasisymmetry the color group $\mathrm{SU(3)_{col}}$ should be a symmetry group (or at least a quasisymmetry). Hence instead of $P_+^{\uparrow}$ the fundamental group G is a combination of $P_+^{\uparrow}$ and $\mathrm{SU(3)_{col}}$. Last but not least all physical states are color singlets. Therefore all basic quark and gluon fields together with all those products of them, which do not belong to $\mathrm{SU(3)_{col}}$ singlets, must be confined $[q_R(\mathscr{T}) \neq 0$ for all $\mathscr{T} \in I_G]$. Moreover the covariance condition A.III.C(2) has to include the color group in such a way that by a suitable choice for the set of quadruples $\{g(\mathscr{T}) \neq 0 | \mathscr{T} \in I_G\}$ the subspace $\mathscr{D}_C$ contains only Poincaré covariant color singlet states. Indeed for an SU(3) model with a quark triplet, an antiquark triplet, and a self-conjugate gluon octet as basic fields in which the four vectors $q_R(\mathscr{T})$ are essentially determined by the color charge of the basic fields, it can be shown[30] that (i) the basic quark, antiquark, and gluon fields are confined; (ii) every vector in the physical Hilbert space is an SU(3)-color singlet state; and (iii) Poincaré invariance holds in the physical Hilbert space.

ACKNOWLEDGMENT

It is a pleasure to thank Dr. D. Schmersau for several encouraging and enlightening discussions. Without them the author would certainly have terminated the work on the subjects of this article.

[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 (1978).
[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]G. Warner, *Harmonic Analysis on Semi-Simple Lie Group I.* (Springer, Berlin, 1972).
[10]E. Nelson, Ann. Math. **70**, 572 (1959).
[11]R. Goodman, Trans. Am. Math. Soc. **143**, 55 (1969).
[12]R. T. Moore, Bull. Am. Math. Soc. **71**, 903 (1965).
[13]E. B. Davis, Commun. Math. Phys. **23**, 159 (1971).
[14]M. Flato, J. Simon, H. Snellman, and D. Sternheimer, Ann. Sci. Ec. Norm. Sup. **5**, 423 (1972).
[15]J. Simon, Commun. Math. Phys. **28**, 39 (1972).
[16]M. Flato and J. Simon, J. Funct. Anal. **13**, 268 (1973).
[17]H. Snellman, J. Math. Phys. **15**, 1054 (1974).
[18]J. Fröhlich, Commun. Math. Phys. **54**, 135 (1977).
[19]A. M. Jaffe, Phys. Rev. **158**, 1454 (1967).
[20]I. M. Gelfand and G. E. Shilov, *Generalized Functions* (Academic, New York, 1968), Vol. II.
[21]F. Treves, *Topological Vector Spaces, Distributions and Kernels* (Academic, New York, 1967).
[22]H. Boerner, *Darstellungen von Gruppen* (Springer, Berlin, 1967).
[23]A. H. Völkel, *Lecture Notes in Physics* (Springer, Berlin, 1977), Vol. 66.
[24]G. Köthe, *Topological Vector Space* (Springer, Berlin, 1979), Vol. II.
[25]M. Mintchev and E. d'Emilio, J. Math. Phys. **22**, 1267 (1981).
[26]T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1966).
[27]N. S. Poulsen, J. Funct. Anal. **9**, 87 (1972).
[28]H. Joos, Fortschr. Phys. **10**, 65 (1962).
[29]A. Erdélyi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. II.
[30]A. H. Völkel, J. Math. Phys. **26**, 1057 (1985).

The geometrical setting of Utiyama's interaction theory

Renato Grassini
Dipartimento di Matematica ed Applicazioni, Università di Napoli, via Mezzacannone, 8, 80134 Napoli, Italy

(Received 27 February 1985; accepted for publication 24 May 1985)

A mathematical deduction of the fiber bundle theory of interaction from Utiyama's theory is drawn, in order to get some deeper insights into the geometry of classical gauge theories.

I. INTRODUCTION

Utiyama's theory[1] deals with the invariance properties of interacting particle and gauge fields, defined as local fields on space-time up to gauge transformations.

The current formulation of this theory[2] embodies all of its results within a geometrical framework where the interacting fields are defined as geometrical objects on a principal fiber bundle, assumed to be the basic space of the theory.

The transition from the former to the latter formulation—formally suggested by the behavior of the fields under gauge transformations—is only inductively explained through qualitative motivations, which would presume to force the above fiber bundle approach upon us as it were due to our own perception of nature.[3]

Surviving to such motivations is, however, in our opinion, the question whether this approach is merely superimposed on, or rather ingrained in the physical theory—in fact, as long as the transition is not a mathematical deduction, one can exhibit only a possible but not any essential role of the involved geometrical structures.

The aim of this paper is then to answer the question, by drawing a mathematical deduction of the current interaction theory from Utiyama's theory.

To this end we start (Sec. II) from a coordinate setting of Utiyama's theory, implied by the invariant theoretical results of a previous paper.[4]

Hence we infer (Sec. III) the view that Utiyama's theory is nothing else but the passive viewpoint of a geometrical gauge theory, deduced from the former through a quotient operation defined by gauge transformations.

Then we develop (Sec. IV) the active viewpoint of this gauge theory as a consequence of the passive one, so embodying the current fiber bundle approach to interaction.[5]

Finally we remark (Sec. V) how the above geometrical theory naturally includes the classical and relativistic space-time theories of gravitational interaction.

II. UTIYAMA'S THEORY

Let

$$E_0 = M \times F^n$$

be the Cartesian product of a four-dimensional, oriented space-time manifold M and an n-dimensional vector space F^n over $F = R$ or C.[6]

(i) Let $(g_{\alpha\beta})$ be a maximal cocycle, over an open covering (U_α) of M, of G-valued transition mappings

$$g_{\alpha\beta}: U_\alpha \cap U_\beta \to G,$$

G being a Lie group acting on F^n as a closed subgroup of $\mathrm{GL}(n,F)$.

A coordinate particle field is any section ψ_α of local phase space E_0 defined on an open set U_α of the above covering, up to gauge transformations

$$\psi_\beta = g_{\beta\alpha} \cdot \psi_\alpha \tag{1}$$

on nonempty intersections $(U_\alpha \cap U_\beta)$.

(ii) Let (Γ_α) be an atlas, over (U_α), of $\mathfrak{g}$-valued one-forms

$$\Gamma_\alpha: U_\alpha \to L(TM,\mathfrak{g})$$

($\mathfrak{g}$ being the Lie algebra of G), obeying the adjoint, pseudotensorial gauge transformation law

$$\Gamma_\beta = \mathrm{ad}(g_{\alpha\beta}^{-1}) \cdot \Gamma_\alpha + g_{\alpha\beta}^{*}\theta. \tag{2}$$

The jet extension of a coordinate particle field ψ_α into local jet space

$$\overline{E}_0 = E_0 \oplus L(TM,E_0)$$

is the section $j_\alpha \psi_\alpha$ of $\overline{E}_0$, the image of ψ_α under the map

$$j_\alpha = \mathrm{id} \oplus D_\alpha,$$

where

$$D_\alpha = d + \Gamma_\alpha$$

is the differential operator characterized by Γ_α.

The jet extension is gauge covariant,[7] that is,

$$j_\beta \psi_\beta = g_{\beta\alpha} \cdot j_\alpha \psi_\alpha \tag{2'}$$

[whenever Eqs. (1) and (2) hold true].

(iii) Let

$$\mathscr{L}: \overline{E}_0 \to R$$

be a G-invariant real function on $\overline{E}_0$—i.e., it is invariant under a (local) vertical automorphism of $\overline{E}_0$ iff this is characterized by a G-valued transition mapping on M.

The action density of a coordinate particle field ψ_α, interacting with coordinate coupling field Γ_α, is the real function on M

$$J(\psi_\alpha,\Gamma_\alpha) = \mathscr{L} \circ j_\alpha \psi_\alpha$$

defined by local Lagrangian $\mathscr{L}$ and jet extension map j_α.

Owing to the G invariance of $\mathscr{L}$, the action density is gauge invariant,[8] that is,

$$J(\psi_\beta,\Gamma_\beta) = J(\psi_\alpha,\Gamma_\alpha) \tag{3}$$

[whenever Eqs. (1) and (2) hold true].

III. GEOMETRICAL THEORY: PASSIVE VIEWPOINT

Let

$E = E_{0/\sim}$

be the quotient of E_0—under the equivalence relation $\sim$ defined by cocycle $(g_{\alpha\beta})$— regarded as a (nontrivial) vector bundle over M.[9]

(i) Let

$\eta: PE \to \mathrm{GL}(n,F)_{/G}$

be the Higgs metric on E defined by $(g_{\alpha\beta})$.[10]

Coordinate particle fields (ψ_α) related to each other by gauge transformations (1) are all pullbacks

$\psi_\alpha = \Phi_\alpha^{-1} \circ \psi,$

through sections

$\Phi_\alpha : E_{0|U_\alpha} \to E_{|U_\alpha}$

of G-principal fiber bundle $\ker(\eta)$, of a unique section ψ of E.

Then any section ψ of phase space E is called a particle field, undergoing, after changes of sections (or gauges) in $\ker(n)$, passive gauge transformations (1).

(ii) Let

$\omega: PE \to L(T(PE), \mathcal{gl}(n,F))$

be the Yang–Mills connection on E, compatible with Higgs metric η, defined by (Γ_α) (Ref. 11) and then undergoing passive gauge transformations (2).

Jet extensions $(j_\alpha \psi_\alpha)$ of coordinate fields (ψ_α) corresponding to a particle field ψ, are all pullbacks

$j_\alpha \psi_\alpha = \overline{\Phi}_\alpha^{-1} \circ j\psi,$

through the covariant extensions of (Φ_α), of a unique section $j\psi$ of

$\overline{E} = E \oplus L(TM,E),$

the image of ψ under the map

$j = \mathrm{id} \oplus D^\omega,$

where D^ω is the covariant derivative defined by ω in E.[12]

Then $j\psi$ is called the jet extension of particle field ψ into jet space $\overline{E}$.

Owing to Eq. (2′), jet extension $j\psi$ is covariant under passive gauge transformations.

(iii) Let

$L: \overline{E} \to R$

be the real function on $\overline{E}$ whose restrictions to (U_α) are

$L_{|U_\alpha} = \mathcal{L} \circ \overline{\Phi}_\alpha^{-1}$

[L is well defined, i.e., the above restrictions agree on intersections $(U_\alpha \cap U_\beta)$, for sections (Φ_α) are related to each other by $(g_{\alpha\beta})$ and $\mathcal{L}$ is G invariant].

Action densities $(J(\psi_\alpha, \Gamma_\alpha))$—$\psi_\alpha$ and Γ_α undergoing gauge transformations—are all restrictions to (U_α) of a unique real function $J(\psi,\omega)$ on M, given by

$J(\psi,\omega) = L \circ j\psi.$

Then $J(\psi,\omega)$ is called the action density of particle field ψ, interacting with coupling field ω, defined by Lagrangian L and jet extension map j.

Owing to Eq. (3), the action density is invariant under passive gauge transformations.

IV. GEOMETRICAL THEORY: ACTIVE VIEWPOINT

Let

$g: E \to E$

be a vertical vector bundle automorphism of E, also characterized by a vertical principal bundle automorphism of PE

$g: PE \to PE.$

The pullback of Higgs metric η, defined by

$g^*\eta = \eta \circ g,$

is a Higgs metric, too [due to the variance properties of g and η under the action of $\mathrm{GL}(n,F)$ on PE].

If, in particular, η is preserved

$g^*\eta = \eta,$

then g is also characterized by a vertical principal bundle automorphism of $\ker(\eta)$—the restriction $\tilde{g}$ of g to $\ker(\eta)$.

(i) Let g be a vertical vector bundle automorphism of E which preserves η (fiber isometry).

Then the pullback of a particle field ψ, defined by

$g^*\psi = g^{-1} \circ \psi,$

is an active gauge transformation of ψ.

If the particle field is described by means of a Higgs zero-form[13]

$\tilde{\psi}: \ker(\eta) \to F^n,$

then the above transformation corresponds to the ordinary pullback

$\tilde{g}^*\tilde{\psi} = \tilde{\psi} \circ \tilde{g}.$

(ii) Pullback

$g^*\omega = \omega \cdot Tg$

of Yang–Mills connection ω under a fiber isometry g of E, defines an active gauge transformation of ω; it is a Yang–Mills connection, too, compatible with η, which yields the jet extension map

$g^*j = \mathrm{id} \oplus D^{g^*\omega}.$

Then jet extension $j\psi$ is covariant under active gauge transformations, that is,

$(g^*j)(g^*\psi) = g^*(j\psi)$

(where, on the right-hand side, g^* denotes the pullback action on $j\psi$ of the covariant extension $\overline{g}$ of g).

In fact, if we introduce the connection one-form on $\ker(\eta)$

$\tilde{\omega} = \omega_{|\ker(\eta)},$

the above covariance law can be given, in terms of Higgs forms, the following known[14] expression:

$\tilde{g}^*\tilde{\psi} \oplus D^{\tilde{g}^*\tilde{\omega}}(\tilde{g}^*\tilde{\psi}) = \tilde{g}^*(\tilde{\psi} \oplus D^{\tilde{\omega}}\tilde{\psi}).$

(iii) Lagrangian L is invariant under fiber isometries of E, due to the G invariance of its coordinate expression $\mathscr{L}$.

Consequently, for any isometry g,

$$L \circ (g^*j)(g^*\psi) = L \circ g^*(j\psi) = L \circ \bar{g}^{-1} \circ j\psi = L \circ j\psi.$$

Then action density is invariant under active gauge transformations

$$J(g^*\psi, g^*\omega) = J(\psi,\omega).$$

V. CONCLUDING REMARK

The above general theory of interaction is then based upon a G-invariant Lagrangian L defined on the jet extension of the phase space $E(\eta,\omega)$ of a particle field—a vector bundle E over space-time M, carrying a Higgs metric η [which spontaneously breaks, through $\ker(\eta)$, the structure group of E down to G] and a Yang–Mills connection ω [which is assumed to be reducible to a connection one-form on $\ker(\eta)$].

In particular, the Lagrangian of a free particle is defined—as is well known—on the jet extension of the tangent bundle $E = TM$ of space-time.

In this case, η and ω come out to be a G structure and a G connection (i.e., a generalized metric structure and a compatible linear connection) on space-time.

The choice of G structure η determines the classical or relativistic kind of theory,[15] whereas G connection ω is the geometrical description of the gravitational field interacting with a free particle.

ACKNOWLEDGMENT

This work was performed under the auspices of the Italian Research Council, Consiglio Nazionale delle Ricerche (CNR).

[1]R. Utiyama, "Invariant theoretical interpretation of interaction," Phys. Rev. **101**, 1597 (1956).

[2]D. Bleecker, *Gauge Theory and Variational Principles* (Addison–Wesley, Reading, MA, 1981), Chap. 3.

[3]See Ref. 2, pp. xiii–xiv.

[4]R. Grassini, J. Math. Phys. **26**, 1345 (1985).

[5]Y. Choquet-Bruhat, C. DeWitt-Morette, and M. Dillard-Bleick, *Analysis, Manifolds and Physics* (North-Holland, Amsterdam, 1982), p. 406.

[6]Manifolds and mappings will be assumed to be C^∞. For standard geometrical definitions and notations, see S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Wiley, New York, 1963) and R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin/Cummings, Cambridge, MA, 1978).

[7]R. Grassini, J. Math. Phys. **26**, 109 (1985).

[8]Conversely ($\mathscr{L}$ being G invariant), the gauge invariance of the action density forces $(g_{\alpha\beta})$ to take values in $G \subset \mathrm{GL}(n,F)$ and (Γ_α) to obey the adjoint, pseudotensorial transformation law—whereas the additional minimality requirement of being tensorial component-free forces (Γ_α) to take values in $\mathscr{g} \subset \mathscr{gl}(n,F)$ (See Ref. 4).

[9]C. Godbillon, *Géométrie Différentielle et Mécanique Analitique* (Hermann, Paris, 1969), p. 46.

[10]See Ref. 4. A Higgs metric on E is a map from the principal fiber bundle PE of vertical linear frames to homogeneous space $\mathrm{GL}(n,F)_{/G}$, equivariant with respect to the action of $\mathrm{GL}(n,F)$.

[11]See Refs. 4 and 7. A Yang–Mills connection on E is a connection one-form on PE.

[12]See Ref. 7.

[13]A. Trautman, "The geometry of gauge fields," Czech. J. Phys. B **29**, 107 (1979).

[14]See Ref. 2, Theorem 3.3.6.

[15]A. Trautman, "A classification of space-time structures," Rep. Math. Phys. **10** (1976).

Recurrence relations for the analytic calculation of Feynman integrals in the axial gauge: The case of massless particles

H. A. Slim
Department of Physics, University of Western Australia, Nedlands, Western Australia 6009

(Received 27 February 1985; accepted for publication 26 April 1985)

An algorithm is described which enables the calculation of Feynman integrals in the axial gauge. The algorithm is based on recurrence relations, which are used to simplify integrals associated with both massless and massive particles. Detailed formulas are presented for the case of massless particles. They can be used to tabulate all Feynman integrals that are relevant for the calculation of propagators in axial gauge quantum chromodynamics and quantum gravity.

I. INTRODUCTION

The Feynman integrals for nontrivial field theories are known analytically only in the simplest cases of one or, at most, a few loops. If the field theory is a non-Abelian gauge theory then the necessary effort for the analytic computation crucially depends on the choice of gauge. Although the class of Lorentz covariant gauges introduces fictitious particles, the propagators and vertices get a simple form. The axial gauge is a competitor of the covariant gauge because the ghost particles decouple.[1] However the propagator for the gauge field has a much more complicated structure involving two scalar functions instead of one, together with a dependence on the axial vector n^μ.

The axial gauge has been used especially in nonperturbative studies of the infrared (IR) behavior of the gluon propagator in quarkless quantum chromodynamics.[2–5] One assumes a reasonable ansatz for the triple gluon vertex, consistent with the Slavnov–Taylor identity which attains its simplest form in the axial gauge. This ansatz is inserted into the Dyson–Schwinger equation for the gluon propagator and, in the first instance, the quadruple gluon vertex is dropped. The result is a closed, nonlinear integral equation for the gluon propagator in momentum space, the BBZ (Baker, Ball, Zachariasen) equation.[3] The solution of this equation for momenta p^μ for which $p^2 \to 0$ gives the IR behavior of the gluon propagator. If the propagator is very singular for $p^2 \to 0$, then its Fourier transform, which is the Born approximation to the Wilson-loop potential, leads to a color confining force.

Analytic calculations are cumbersome in this approximation scheme because of the complex nonlinear nature of the BBZ equation. They have mostly been done with the major simplifying restriction of $p \cdot n = 0$ (see Ref. 4). One calculation has been done with $p \cdot n \neq 0$, in order to study the self-consistency of the BBZ equation in the IR limit.[5] The results of the present paper were used in that analysis.

Another area of application is quantum gravity. The ultraviolet (UV) divergent part of the one-loop self-energy for the graviton, in the axial gauge, has been calculated explicitly.[6,7] The result is quite remarkable; not only do the necessary counterterms depend on the axial vector, indicating the nonrenormalizability of quantum gravity, but the infinite part of the self-energy also fails to be transverse. It has been shown that the imaginary part of the graviton's self energy is transverse.[7] How this works out for the combination of the axial gauge and the background field quantization method is still an open question.

With regard to these problems, the purpose of this paper is to describe a systematic procedure for the analytic computation of single-loop Feynman integrals, as they occur in the calculation of propagators in the axial gauge.[5–8] In particular we will consider integrals of the form

$$I^{\mu_1 \cdots \mu_N}_{\alpha_1,\alpha_2,l}(p;m_1,m_2,\chi) = i\int \frac{d^Dq}{(2\pi)^D} \frac{q^{\mu_1}\cdots q^{\mu_N}}{((q+p)^2 - m_1^2 + i\epsilon)^{\alpha_1}(q^2 - m_2^2 + i\epsilon)^{\alpha_2}} \times \mathscr{P}\frac{1}{(q\cdot n + \chi)^l}. \tag{1.1}$$

The integration is carried out over D-dimensional momentum space with the integration variable q, which has one time and D-1 space components. Here, p^μ is an arbitrary vector in this D-dimensional momentum space on which the scalar product, $p\cdot q$, is defined by

$$p\cdot q = p^\mu q^\nu g_{\mu\nu} = p_0 q_0 - \sum_{i=1}^{D} p_i q_i .$$

The axial vector, characterizing the axial gauge, is denoted as n^μ and it will always be assumed to be a spacelike vector, $n^2 < 0$, so that

$$(p\cdot n)^2/n^2 < 0. \tag{1.2}$$

In formula (1.1) m_1 and m_2 are mass parameters whereas α_1 and α_2 are positive integers. The parameter l is a non-negative integer and the singularity $q\cdot n + \chi = 0$, present in the integrand for $l \geqslant 1$, is regularized by taking the principal value, defined as[1]

$$\mathscr{P}\frac{1}{(q\cdot n + \chi)^l} = \lim_{\eta\downarrow 0} \frac{1}{2}\left\{\frac{1}{(q\cdot n + \chi + i\eta)^l} + \frac{1}{(q\cdot n + \chi - i\eta)^l}\right\}. \tag{1.3}$$

For $l = 0$ the rhs of (1.3) is replaced by the value 1. For the parameter χ usually the equalities $\chi = 0$ or $\chi = p\cdot n$ hold.

Because (1.1) is a tensor on the D-dimensional momentum space, the index N will be called the rank of the integral and for $N = 0$, it will be understood as

$$I_{\alpha_1,\alpha_2,l}(p;m_1,m_2,\chi) = i\int \frac{d^Dq}{(2\pi)^D} \frac{1}{((q+p)^2 - m_1^2 + i\epsilon)^{\alpha_1}(q^2 - m_2^2 + i\epsilon)^{\alpha_2}} \times \mathscr{P}\frac{1}{(q\cdot n + \chi)^l}.$$

In Sec. II we will show that the computation of (1.1) for $N \geqslant 1$ can be simplified via certain recurrence relations to integrals of rank $N = 0$, with $l = 0$ or $l = 1$ only, i.e.,

$$I^{\mu_1\cdots\mu_N}_{\alpha_1,\alpha_2,l}(p;m_1,m_2;\chi) = \sum_\alpha \sum_{l'=0,1} \sum_{n \geqslant 0} T^{\mu_1\cdots\mu_N}_{\alpha,l,n}(p,n,g;\chi,D) \times \int_0^1 dx\, x^n U_{\alpha,l'}(-xp;xm_1^2 + (1-x)m_2^2 - xp^2,\chi), \tag{1.4}$$

where $T^{\mu_1\cdots\mu_N}_{\alpha,l,n}(p,n,g;\chi,D)$ is a tensor of rank N, which in general depends on p^μ, n^μ, $g^{\mu\nu}$ and scalars formed from combination of the vectors p^μ and n^μ, and on χ and D. The function $U_{\alpha,l}$ is given by Eqs. (2.9b) and (2.10b).

There remains the integration over the Feynman variable, x. This will be taken up in the third section, for the case of massless integrals. There formula (1.4) will be simplified further and the finite part will be expressed in terms of a single, nontrivial, transcendental function.

The last section contains conclusions and an outlook on further applications of the techniques developed in this paper.

II. DERIVATION OF THE RECURRENCE RELATIONS

In this section we will derive formulas that enable the calculation of the Feynman integral (1.1) recursively. Instead of dealing directly with (1.1) we will consider a more general type of Feynman integral. This new form is in fact essential for the derivation of the desired recurrence relations.

The integrals (1.1) can be rewritten with the Feynman trick

$$\frac{1}{A^{\alpha_1}B^{\alpha_2}} = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\int_0^1 dx \frac{x^{\alpha_1-1}(1-x)^{\alpha_2-1}}{(Ax + B(1-x))^{\alpha_1+\alpha_2}},$$

in the form

$$I^{\mu_1\cdots\mu_N}_\alpha(p;m_1,m_2;\chi) = \frac{1}{2^{\alpha_1+\alpha_2}\Gamma(\alpha_1)\Gamma(\alpha_2)(4\pi)^{D/2}}\int_0^1 dx\, x^{\alpha_1-1}(1-x)^{\alpha_2-1} \times F^{\mu_1\cdots\mu_N}_{\alpha_1+\alpha_2,l}(-xp;xm_1^2 + (1-x)m_2^2 - xp^2,\chi). \tag{2.1}$$

Here the following class of generalized Feynman integrals has been defined as

$$F^{\mu_1\cdots\mu_N}_{\alpha,l}(\tilde p;C,\chi) = 2^\alpha\Gamma(\alpha)i\int \frac{d^Dq}{\pi^{D/2}} \frac{q^{\mu_1}\cdots q^{\mu_N}}{(q^2 - 2\tilde p\cdot q - C + i\epsilon)^\alpha}\mathscr{P}\frac{1}{(q\cdot n + \chi)^l}, \tag{2.2}$$

where $\tilde p^\mu$ is again an arbitrary, D-dimensional, momentum vector, and C and χ are arbitrary parameters, not necessarily related to $\tilde p^\mu$.

Instead of considering (2.2) for the few positive integer values of D where the integral might exist, we adopt the principle of dimensional regularization and consider D as a complex variable. Then the right-hand side (rhs) of (2.2) exists if D, N, α, and l satisfy the inequality $\operatorname{Re} D < 2\alpha + l - N$. However, in the following we will assume that

$$\operatorname{Re} D < 2\alpha_{\min} - 1 \tag{2.3}$$

holds where $\alpha_{\min} = \alpha - [N/2]$, $[N/2]$ defined as the greatest integer equal to, or smaller than $N/2$. Once the left-hand side (lhs) of (2.2) is known for these values of D, the principle of analytic continuation can be used to find the value of $F^{\mu_1\cdots\mu_N}_{\alpha,l}$ in the entire complex D plane and to locate its singularities. These will be interpreted as the UV divergences of the integral (2.2).

Now we state the recurrence relations for the functions (2.2) which express an integral of rank $N \geqslant 2$ in terms of integrals of rank $N-1$ and $N-2$. Two separate cases are considered below. The case $l = 0$ is given by Eq. (2.4) whereas the case of general $l \geqslant 1$ is covered by Eq. (2.5). The bar over a Lorentz index μ, i.e., $\bar\mu$ in the following formulas, indicates the absence of that Lorentz index.

Lemma: For $l = 0$, we have, for $N = 1$,

$$F^{\mu_1}_{\alpha,0}(\tilde p;C,\chi) = \tilde p^{\mu_1}F^{\bar\mu_1}_{\alpha,0}(\tilde p;C,\chi), \tag{2.4a}$$

and, for $N \geqslant 2$,

$$F^{\mu_1\cdots\mu_N}_{\alpha,l}(\tilde p;C,\chi) = \tilde p^{\mu_1}F^{\bar\mu_1,\mu_2\cdots\mu_N}_{\alpha,l}(\tilde p;C,\chi) + \sum_{i=2}^N g^{\mu_1\mu_i}F^{\bar\mu_1\mu_2\cdots\bar\mu_i\cdots\mu_N}_{\alpha-1,l}(\tilde p;C,\chi). \tag{2.4b}$$

For $l \geqslant 1$, $N = 1$,

$$F^{\mu_1}_{\alpha,l}(\tilde p;C,\chi) = (\tilde p^{\mu_1} - [(\tilde p\cdot n + \chi)/n^2]n^{\mu_1})F^{\bar\mu_1}_{\alpha,l}(\tilde p;C,\chi) + (n^{\mu_1}/n^2)F^{\bar\mu_1}_{\alpha,l-1}(\tilde p;C,\chi), \tag{2.5a}$$

and, for $N \geqslant 2$,

$$F^{\mu_1\cdots\mu_N}_{\alpha,l}(\tilde p;C,\chi) = (\tilde p^{\mu_1} - [(\tilde p\cdot n + \chi)/n^2]n^{\mu_1})F^{\bar\mu_1\mu_2\cdots\mu_N}_{\alpha,l}(\tilde p;C,\chi) + (n^{\mu_1}/n^2)F^{\bar\mu_1\mu_2\cdots\mu_N}_{\alpha,l-1}(\tilde p;C,\chi) + \sum_{i=2}^N\left(g^{\mu_1\mu_i} - \frac{n^{\mu_1}n^{\mu_i}}{n^2}\right)F^{\bar\mu_1\mu_2\cdots\bar\mu_i\cdots\mu_N}_{\alpha-1,l}(\tilde p;C,\chi). \tag{2.5b}$$

The seemingly special role, played by the index μ_1, is illusory, since (2.2) is totally symmetric under permutation of the indices.

Proof of the Lemma: The proof of the lemma is based on the following two simple identities, which are true for any $N \geqslant 0$:

$$F^{\mu_1\cdots\mu_N\mu}_{\alpha+1,l}(\tilde p;C,\chi) = \frac{\partial}{\partial\tilde p_\mu}F^{\mu_1\cdots\mu_N}_{\alpha,l}(\tilde p;C,\chi) \tag{2.6}$$

and

$$F^{\mu_1\cdots\mu_N}_{\alpha,l}(\tilde p;C,\chi) = -\frac{1}{2}\int_C^\infty dC'\, F^{\mu_1\cdots\mu_N}_{\alpha+1,l}(\tilde p;C',\chi). \tag{2.7}$$

Here it has been assumed that C and χ are independent of $\tilde p$.

Although the index α is changed by the two operations (2.6) and (2.7), the combination of both leaves α invariant; only the value of the rank N is changed by 1, as is shown in the formula

$$F_{\alpha,l}^{\mu_1\cdots\mu_N\mu}(\tilde{p};C,\chi) = -\frac{1}{2}\int_C^\infty dC'\frac{\partial}{\partial\tilde{p}_\mu}F_{\alpha,l}^{\mu_1\cdots\mu_N}(\tilde{p};C',\chi). \tag{2.8}$$

These three identities, (2.6)–(2.8), will now be exploited for the proof of (2.4) and (2.5).

First the relations (2.4b) and (2.5b) may be proved by induction. Suppose they hold true for $2 \leqslant N \leqslant \mathscr{N}$. Then the integral (2.2) of rank $N=\mathscr{N}+1$, and with given α and l, can be expressed in terms of the integral of rank $\mathscr{N}$ with the same α and l, via Eq. (2.8). Substituting the rhs of (2.4b) into the rhs of (2.8), and using (2.7) shows that Eq. (2.4b) again holds, but now for $N=\mathscr{N}+1$. This proves (2.4b) for any $N \geqslant 2$, once the case $N=2$ has been demonstrated. The proof of (2.5b) carries through in a similar fashion. It therefore remains to consider (2.4b) and (2.5b) with $N=2$. These equations readily follow from Eq. (2.4a) and (2.5a), respectively, by using the identities (2.8) and (2.7) once more.

For the proof of (2.4a) and (2.5a), the explicit form of the integrals (2.2), for $N=0$, has to be used. For $l=0$, the integral is well known and the answer is

$$F_{\alpha,0}(\tilde{p};C,\chi) = -U_{\alpha,0}(\tilde{p};C,\chi), \tag{2.9a}$$

with

$$U_{\alpha,0}(\tilde{p};C,\chi) = (-2)^\alpha\Gamma(\alpha - D/2)(\tilde{p}^2 + C)^{-\alpha+D/2}. \tag{2.9b}$$

Substituting (2.9) into the rhs of Eq. (2.8) then gives (2.4a) for $l=0$.

The evaluation of $F_{\alpha,1}(\tilde{p};C,\chi)$ is more elaborate, using either the Feynman or the Schwinger representation, together with the principle value prescription (1.3). In both cases we get the same answer, that is

$$F_{\alpha,1}(\tilde{p};C,\chi) = -[(\tilde{p}\cdot n+\chi)/n^2]\,U_{\alpha,1}(\tilde{p};C,\chi), \tag{2.10a}$$

with

$$U_{\alpha,1}(\tilde{p};C,\chi) = U_{\alpha+1,0}(\tilde{p};C,\chi)\,{}_2F_1\left(\alpha+1-\frac{D}{2},1;\frac{3}{2};\frac{(\tilde{p}\cdot n+\chi)^2}{(\tilde{p}^2+C)n^2}\right), \tag{2.10b}$$

where ${}_2F_1(a,b;c;z)$ is the hypergeometric function. To prove (2.5a) for $l=1$ we first rewrite (2.10), using the integral representation of the hypergeometric function, as

$$\begin{aligned} F_{\alpha,1}(\tilde{p};C,\chi) &= (-2)^\alpha\Gamma(\alpha+1-D/2)(\tilde{p}\cdot n+\chi)/n^2 \\ &\quad\times\int_0^1 dt\,(1-t)^{-1/2} \\ &\quad\times(\tilde{p}^2+C-t(\tilde{p}\cdot n+\chi)^2/n^2)^{-\alpha-1+D/2}. \end{aligned} \tag{2.11}$$

Substituting this into the rhs of the identity (2.8), returning to the ${}_2F_1$ notation and using the relation

$${}_2F_1(a+1,1;c+1;z) = (c/az)(-1+{}_2F_1(a,1;c;z)) \tag{2.12}$$

finally gives (2.5a) for $l=1$.

To complete the proof of (2.5a) we use the following relation, which holds for any $N \geqslant 0$ and for $l \geqslant 2$:

$$F_{\alpha,l}^{\mu_1\cdots\mu_N}(\tilde{p};C,\chi) = \frac{(-1)^{l-1}}{\Gamma(l)}\frac{\partial^{l-1}}{\partial\chi^{l-1}}F_{\alpha,1}^{\mu_1\cdots\mu_N}(\tilde{p};C,\chi). \tag{2.13}$$

This formula enables us to express $F_{\alpha,l}(\tilde{p};C,\chi)$ in terms of integrals with l replaced by $l-1$ and $l-2$ and also α changed in value. Differentiating $F_{\alpha,l}(\tilde{p};C,\chi)$ as written in the form (2.11) with respect to χ, returning to the ${}_2F_1$ notation, and using (2.12) again gives

$$\frac{\partial}{\partial\chi}F_{\alpha,1}(\tilde{p};C,\chi) = (1/n^2)(F_{\alpha+1,0}(\tilde{p};C,\chi) - (\tilde{p}\cdot n+\chi)F_{\alpha+1,1}(\tilde{p};C,\chi)),$$

so in general, for $l \geqslant 2$,

$$\frac{\partial^l}{\partial\chi^l}F_{\alpha,1}(\tilde{p};C,\chi) = -\frac{1}{n^2}\left\{(l-1)\frac{\partial^{l-2}}{\partial\chi^{l-2}}F_{\alpha+1,1}(\tilde{p};C,\chi) + (\tilde{p}\cdot n+\chi)\frac{\partial^{l-1}}{\partial\chi^{l-1}}F_{\alpha+1,1}(\tilde{p};C,\chi)\right\}$$

and with (2.13), for $N=0$, we finally get

$$\begin{aligned} F_{\alpha,l}(\tilde{p};C,\chi) &= [1/(l-1)n^2]\{-F_{\alpha+1,l-2}(\tilde{p};C,\chi) \\ &\quad + (\tilde{p}\cdot n+\chi)F_{\alpha+1,l-1}(\tilde{p};C,\chi)\}. \end{aligned} \tag{2.14}$$

Now Eq. (2.5a) can be proved for $l \geqslant 2$ by induction, making use of (2.7), (2.8), (2.14), and (2.5a) for $l=0$ and $l=1$. This completes the proof of the lemma.

One can extend the formula (2.14) to integrals of rank $N \geqslant 1$. Using (2.8) we find, for $l \geqslant 2$, another recurrence relation,

$$\begin{aligned} F_{\alpha,l}^{\mu_1\cdots\mu_N}(\tilde{p};C,\chi) &= \frac{1}{(l-1)n^2}\left\{\sum_{i=1}^N n^{\mu_i}F_{\alpha,l-1}^{\mu_1\cdots\hat{\mu}_i\cdots\mu_N}(\tilde{p};C,\chi)\right. \\ &\quad + (\tilde{p}\cdot n+\chi)F_{\alpha+1,l-1}^{\mu_1\cdots\mu_N}(\tilde{p};C,\chi) \\ &\quad \left. - F_{\alpha+1,l-2}^{\mu_1\cdots\mu_N}(\tilde{p};C,\chi)\right\}. \end{aligned} \tag{2.15}$$

We want to make a final comment to the proof. For a precise derivation of the results it is actually correct to use a finite upper limit of integration in (2.7) and (2.8). So f.i. instead of (2.8) we have to consider the difference

$$F_{\alpha,l}^{\mu_1\cdots\mu_N\mu}(\tilde{p};C,\chi) - F_{\alpha,l}^{\mu_1\cdots\mu_N\mu}(\tilde{p};\Lambda,\chi) = -\frac{1}{2}\int_C^\Lambda dC'\frac{\partial}{\partial\tilde{p}_\mu}F_{\alpha,l}^{\mu_1\cdots\mu_N}(\tilde{p};C',\chi).$$

The formulas (2.4) and (2.5) also hold for this difference. After using these iteratively, we get eventually for (2.2) the expression

$$F_{\alpha,l}^{\mu_1\cdots\mu_N}(\tilde{p};C,\chi)$$
$$=F_{\alpha,l}^{\mu_1\cdots\mu_N}(\tilde{p};\Lambda,\chi)$$
$$+\sum_{\alpha'>\alpha_{\min}}\sum_{l'=0,1}S_{\alpha',l'}^{\mu_1\cdots\mu_N}(\tilde{p},n,g;\chi,D)$$
$$\times\{U_{\alpha',l'}(\tilde{p};C,\chi)-U_{\alpha',l'}(\tilde{p};\Lambda,\chi)\},$$

where $\alpha_{\min}$ has already been defined by Eq. (2.3). The explicit form of $F_{\alpha,l}^{\mu_1\cdots\mu_N}(\tilde{p};\Lambda,\chi)$, as given by (2.9), (2.10), and (2.14), shows that, for Re $D<2\alpha_{\min}-1$, this function vanishes for $\Lambda\to\infty$. Therefore the Λ-dependent terms can be dropped in the above formula, and using (2.1), (2.9a), and (2.10a), Eq. (1.4) follows.

The results derived so far are valid for the general case of integrals that carry Lorentz indices, such as (1.1). In addition they are also valid for the case where some of the vectors $q^{\mu_1},\ldots,q^{\mu_N}$ are contracted with an external vector or the metric tensor. These integrals frequently occur in the computation of propagators.[5–8] Therefore it is more useful to work without explicit Lorentz indices and to saturate them all by contraction of (1.1) with the external momentum p^μ, the axial vector n^μ, and the metric tensor $g^{\mu\nu}$. So we are led to consider integrals of the following type:

$$\mathscr{F}_{\alpha,l}(\tilde{p};C,\chi;r,s,t)$$
$$=n_{\lambda_1}\cdots n_{\lambda_r}p_{\nu_1}\cdots p_{\nu_s}$$
$$\times g_{\rho_1\sigma_1}\cdots g_{\rho_t\sigma_t}F_{\alpha,l}^{\lambda_1\cdots\lambda_r\nu_1\cdots\nu_s\rho_1\cdots\rho_t\sigma_1\cdots\sigma_t}(\tilde{p};C,\chi), \tag{2.16}$$

where always $r\geqslant 0$, $s\geqslant 0$, $t\geqslant 0$, and $N=r+s+t$.

Recurrence relations for $\mathscr{F}_{\alpha,l}(\tilde{p};C,\chi;r,s,t)$ can be deduced from (2.4) and (2.5). For the derivation we assume that p^μ and $\tilde{p}^\mu$ are proportional,

$$\tilde{p}^\mu=Rp^\mu, \tag{2.17}$$

where R is a scalar. This relation is certainly satisfied for the cases in which we are interested, i.e., $R=-x$ in (1.1).

To be able to use (2.4) and (2.5) we assume that $N\geqslant 1$ and consider different cases. There are three possibilities, corresponding to either $r\geqslant 1$ and $s+t\geqslant 0$, $s\geqslant 1$ and $r+t\geqslant 0$, or $t\geqslant 1$ and $r+s\geqslant 0$. We also distinguish between the cases $l=0$ and $l\geqslant 1$.

The recurrence relations are generated by substituting the rhs of (2.4) and (2.5) into the rhs of (2.6) with the Lorentz index μ_1 in (2.4) and (2.5) replaced by λ_1, ν_1, and ρ_1, corresponding to, respectively, $r\geqslant 1$, $s\geqslant 1$, and $t\geqslant 1$. For example, if $r\geqslant 1$, $s+t\geqslant 0$ we write (2.4b) in the form

$$F_{\alpha,0}^{\lambda_1\cdots\lambda_r\nu_1\cdots\nu_s\rho_1\cdots\rho_t\sigma_1\cdots\sigma_t}(\tilde{p};C,\chi)$$
$$=\tilde{p}^{\lambda_1}F_{\alpha,0}^{\bar{\lambda}_1\lambda_2\cdots\lambda_r\nu_1\cdots\nu_s\rho_1\cdots\rho_t\sigma_1\cdots\sigma_t}(\tilde{p};C,\chi)$$
$$+\sum_{i=2}^{r}g^{\lambda_1\lambda_i}F_{\alpha-1,0}^{\bar{\lambda}_1\lambda_2\cdots\bar{\lambda}_i\cdots\lambda_r\nu_1\cdots\nu_s\rho_1\cdots\rho_t\sigma_1\cdots\sigma_t}(\tilde{p};C,\chi)$$
$$+\sum_{i=1}^{s}g^{\lambda_1\nu_i}F_{\alpha-1,0}^{\bar{\lambda}_1\lambda_2\cdots\lambda_r\nu_1\cdots\bar{\nu}_i\cdots\nu_s\rho_1\cdots\rho_t\sigma_1\cdots\sigma_t}(\tilde{p};C,\chi)$$
$$+\sum_{i=1}^{t}g^{\lambda_1\rho_i}F_{\alpha-1,0}^{\bar{\lambda}_1\lambda_2\cdots\lambda_r\nu_1\cdots\nu_s\rho_1\cdots\bar{\rho}_i\cdots\rho_t\sigma_1\cdots\sigma_t}(\tilde{p};C,\chi)$$
$$+\sum_{i=1}^{t}g^{\lambda_1\sigma_i}F_{\alpha-1,0}^{\bar{\lambda}_1\lambda_2\cdots\lambda_r\nu_1\cdots\nu_s\rho_1\cdots\rho_t\sigma_1\cdots\bar{\sigma}_i\cdots\sigma_t}(\tilde{p};C,\chi).$$

Substituting this into the rhs of (2.16) gives

$$\mathscr{F}_{\alpha,0}(\tilde{p};C,\chi;r,s,t)$$
$$=Rp\cdot n\mathscr{F}_{\alpha,0}(\tilde{p};C,\chi;r-1,s,t)$$
$$+(r-1)n^2\mathscr{F}_{\alpha-1,0}(\tilde{p};C,\chi;r-2,s,t)$$
$$+sp\cdot n\mathscr{F}_{\alpha-1,0}(\tilde{p};C,\chi;r-1,s-1,t)$$
$$+2t\mathscr{F}_{\alpha-1,0}(\tilde{p};C,\chi;r,s,t-1). \tag{2.18}$$

Similarly, for $l=0$, $s\geqslant 1$, and $r+t\geqslant 0$, we get

$$\mathscr{F}_{\alpha,0}(\tilde{p};C,\chi;r,s,t)$$
$$=Rp^2\mathscr{F}_{\alpha,0}(\tilde{p};C,\chi;r,s-1,t)$$
$$+rp\cdot n\mathscr{F}_{\alpha-1,0}(\tilde{p};C,\chi;r-1,s-1,t)$$
$$+(s-1)p^2\mathscr{F}_{\alpha-1,0}(\tilde{p};C,\chi;r,s-2,t)$$
$$+2t\mathscr{F}_{\alpha-1,0}(\tilde{p};C,\chi;r,s,t-1), \tag{2.19}$$

and for $l=0$, $t\geqslant 1$, and $r+s\geqslant 0$, we get

$$\mathscr{F}_{\alpha,0}(\tilde{p};C,\chi;r,s,t)$$
$$=R\mathscr{F}_{\alpha,0}(\tilde{p};C,\chi;r,s+1,t-1)$$
$$+(r+s+2t+D-2)\mathscr{F}_{\alpha-1,0}(\tilde{p};C,\chi;r,s,t-1). \tag{2.20}$$

The case $l\geqslant 1$ is treated in exactly the same way, now using (2.5b). It is not necessary to consider $r\geqslant 1$ because $\mathscr{F}_{\alpha,l}(\tilde{p};C,\chi;r,s,t)$ can be expressed in terms of $\mathscr{F}_{\alpha,0}(\tilde{p};C,\chi;r,s,t)$ with $r\geqslant 0$ and in $\mathscr{F}_{\alpha,l}(\tilde{p};C,\chi;0,s,t)$ with $l\geqslant 1$.

For $l=1$, $r=0$, $s\geqslant 1$, and $t=0$, we get

$$\mathscr{F}_{\alpha,1}(\tilde{p};C,\chi;0,s,t)$$
$$=(Rp^2-(Rp\cdot n+\chi)p\cdot n/n^2)$$
$$\times\mathscr{F}_{\alpha,1}(\tilde{p};C,\chi;0,s-1,t)$$
$$+[(p\cdot n)/n^2]\mathscr{F}_{\alpha,0}(\tilde{p};C,\chi;0,s-1,t)$$
$$+(s-1)(p^2-(p\cdot n)^2/n^2)$$
$$\times\mathscr{F}_{\alpha-1,1}(\tilde{p};C,\chi;0,s-2,t)$$
$$+2t\mathscr{F}_{\alpha-1,1}(\tilde{p};C,\chi;0,s,t-1)$$
$$-2t[(p\cdot n)/n^2]\mathscr{F}_{\alpha-1,0}(\tilde{p};C,\chi;0,s-1,t-1)$$
$$+2t[(p\cdot n\chi)/n^2]\mathscr{F}_{\alpha-1,1}(\tilde{p};C,\chi;0,s-1,t-1), \tag{2.21}$$

and for $r=0$, $s=0$, and $t\geqslant 1$, we get

$$\mathscr{F}_{\alpha,1}(\tilde{p};C,\chi;0,s,t)$$
$$=\frac{1}{n^2}\{Rn^2\mathscr{F}_{\alpha,1}(\tilde{p};C,\chi;0,s+1,t-1)$$
$$-(Rp\cdot n+\chi)\mathscr{F}_{\alpha,0}(\tilde{p};C,\chi;0,s,t-1)$$
$$+(Rp\cdot n+\chi)\chi\mathscr{F}_{\alpha,1}(\tilde{p};C,\chi;0,s,t-1)$$
$$+\mathscr{F}_{\alpha,0}(\tilde{p};C,\chi;1,s,t-1)$$
$$+sn^2\mathscr{F}_{\alpha-1,1}(\tilde{p};C,\chi;0,s,t-1)$$
$$-sp\cdot n\mathscr{F}_{\alpha-1,0}(\tilde{p};C,\chi;0,s-1,t-1)$$
$$+sp\cdot n\chi\mathscr{F}_{\alpha-1,1}(\tilde{p};C,\chi;0,s-1,t-1)$$
$$+(2t+D-3)n^2\mathscr{F}_{\alpha-1,1}(\tilde{p};C,\chi;0,s,t-1)$$
$$-2(t-1)\mathscr{F}_{\alpha-1,0}(\tilde{p};C,\chi;1,s,t-2)$$
$$+2(t-1)\chi\mathscr{F}_{\alpha-1,0}(\tilde{p};C,\chi;0,s,t-2)$$
$$-2(t-1)\chi^2\mathscr{F}_{\alpha-1,1}(\tilde{p};C,\chi;0,s,t-2)\}. \tag{2.22}$$

The analog of (2.15) becomes, for $l \geqslant 2$ and $r=0$,

$$\begin{aligned}
&\mathscr{F}_{\alpha,l}(\tilde{p}; C,\chi; 0,s,t)\\
&\quad= \frac{1}{(l-1)n^2}\{sp\cdot n\mathscr{F}_{\alpha,l-1}(\tilde{p}; C,\chi; 0,s-1,t)\\
&\qquad+2t\mathscr{F}_{\alpha,l-2}(\tilde{p}; C,\chi; 0,s,t-1)\\
&\qquad-2t\chi\mathscr{F}_{\alpha,l-1}(\tilde{p}; C,\chi; 0,s,t-1)\\
&\qquad+(\tilde{p}\cdot n+\chi)\mathscr{F}_{\alpha+1,l-1}(\tilde{p}; C,\chi; 0,s,t)\\
&\qquad-\mathscr{F}_{\alpha+1,l-2}(\tilde{p}; C,\chi; 0,s,t)\},
\end{aligned}\tag{2.23}$$

and this relation holds for $s+t \geqslant 0$. It expresses $\mathscr{F}_{\alpha,l}(\tilde{p}; C,\chi; 0,s,t)$ for $l \geqslant 2$ in corresponding functions with l replaced by $l-1$ and $l-2$, and with different indices α, s, and t.

A practical strategy would be to calculate $\mathscr{F}_{\alpha,0}(\tilde{p}; C,\chi; r,s,t)$ for a given range of values of r, s, and t, using (2.18)–(2.20). Next the calculation of $\mathscr{F}_{\alpha,1}(\tilde{p}; C,\chi; 0,s,t)$ can be done with (2.21) and (2.22) for the appropriate values of s and t. Finally $\mathscr{F}_{\alpha,l}(\tilde{p}; C,\chi; 0,s,t)$ can be calculated for $l \geqslant 2$ with Eq. (2.23). In this way any integral of type (2.16) can be broken up into a sum of terms which only contain the two basic integrals (2.9) and (2.10), i.e.,

$$\mathscr{F}_{\alpha,0}(\tilde{p}; C,\chi; 0,0,0) = -U_{\alpha,0}(\tilde{p}; C,\chi)$$

and

$$\mathscr{F}_{\alpha,1}(\tilde{p}; C,\chi; 0,0,0) = -[(\tilde{p}\cdot n+\chi)/n^2]U_{\alpha,1}(\tilde{p}; C,\chi).$$

To treat the remaining integrals of (1.4) in a uniform way, we want to reexpress the functions $U_{\alpha,l}(\tilde{p}; C,\chi)$ in terms of $U_{\alpha_D,l}(\tilde{p}; C,\chi)$, where α_D is a finite number. Since eventually we are interested in the physical case $D=4$, the preferred choice for α_D is $\alpha_D=2$. This is motivated by the fact that in the calculation of propagators one usually encounters (1.1) with $\alpha_1=\alpha_2=1$. In fact, $\alpha=2$ is an exceptional point because for $\alpha \leqslant 2$, $U_{\alpha,0}(\tilde{p}; C,\chi)$ has a sequence of poles left of, and including, the point $D=4$. This is also the case for $U_{\alpha,1}(\tilde{p}; C,\chi)$ if $\alpha \leqslant 1$. The pole at $D=4$ will be interpreted as the usual UV divergence of the integral (2.2).

The relation between $U_{\alpha,l}(\tilde{p}; C,\chi)$ and $U_{\alpha_D,l}(\tilde{p}; C,\chi)$ follows, for $l=0$, from Eq. (2.9b), i.e.,

$$\begin{aligned}
&U_{\alpha,0}(\tilde{p}; C,\chi)\\
&\quad= -[1/(2\alpha-D)](\tilde{p}^2+C)U_{\alpha+1,0}(\tilde{p}; C,\chi).
\end{aligned}\tag{2.24}$$

The corresponding recurrence relation in the case $l=1$ follows from (2.10b), (2.24), and another property of the hypergeometric function,

$$\begin{aligned}
&{}_2F_1(a,1;c;z)\\
&\quad=[1/(a+1-c)]\{1-c+a(1-z){}_2F_1(a+1,1;c;z)\}.
\end{aligned}$$

The result is

$$\begin{aligned}
&U_{\alpha,1}(\tilde{p}; C,\chi)\\
&\quad= -[1/(2\alpha+1-D)]\{U_{\alpha+1,0}(\tilde{p}; C,\chi)\\
&\qquad+(\tilde{p}^2+C-(\tilde{p}\cdot n+\chi)^2/n^2)U_{\alpha+1,1}(\tilde{p}; C,\chi)\}.
\end{aligned}\tag{2.25}$$

Conversely, the relations to lower the index α are

$$U_{\alpha,0}(\tilde{p}; C,\chi) = -\frac{2\alpha-2-D}{\tilde{p}^2+C}U_{\alpha-1,0}(\tilde{p}; C,\chi) \tag{2.26}$$

and

$$\begin{aligned}
&U_{\alpha,1}(\tilde{p}; C,\chi)\\
&\quad= -[\tilde{p}^2+C-(\tilde{p}\cdot n+\chi)^2/n^2]^{-1}\\
&\qquad\times\{U_{\alpha,0}(\tilde{p}; C,\chi)+(2\alpha-1-D)U_{\alpha-1,1}(\tilde{p}; C,\chi)\}.
\end{aligned}\tag{2.27}$$

Equations (2.24) and (2.25) can be used to simplify the rhs of Eq. (1.4) and to reexpress it in integrals of the form

$$\int_0^1 dx\, x^n U_{\alpha,l}(-xp; xm_1^2+(1-x)m_2^2-xp^2,\chi),$$

with $n \geqslant 0$ and $\alpha \geqslant \alpha_D$. The remaining exercise is to simplify this integral and express it in terms of one or perhaps more basic transcendental functions.

In the next section this program will be carried out for the simpler case of integrals associated with massless particles.

III. FEYNMAN INTEGRALS FOR ZERO MASS PARTICLES

In this section we start with (1.4) and set $m_1=m_2=0$, so the quantities $\tilde{p}$, C, and χ of Sec. II now become

$$\tilde{p}^\mu=-xp^\mu,\quad C=-xp^2,\quad \chi=0 \text{ or } \chi=p\cdot n. \tag{3.1}$$

Besides UV divergences, the integral $I^{\mu_1\cdots\mu_N}_{\alpha_1,\alpha_2,l}(p;0,0,\chi)$ can also have IR divergences and, according to power counting, it exists in a domain of the complex D plane, determined by

$$2\alpha_2+l-N<\operatorname{Re}D<2\alpha_1+2\alpha_2+l-N.$$

For these values of D the integral can be expressed as

$$\begin{aligned}
&I^{\mu_1\cdots\mu_N}_{\alpha_1,\alpha_2,l}(p;0,0,\chi)\\
&\quad=\sum_{\alpha \geqslant \alpha_{\min}}\sum_{l'=0,1}\sum_{n \geqslant 0}T^{\mu_1\cdots\mu_N}_{\alpha,l',n}(p,n,g;\chi,D)\\
&\qquad\times\int_0^1 dx\, x^n\, U_{\alpha,l'}(-xp;\ -xp^2,\chi),
\end{aligned}\tag{3.2}$$

where the upper limits of summation for α and n, although not specified, can be determined and are finite. Outside its domain of existence, $I^{\mu_1\cdots\mu_N}_{\alpha_1,\alpha_2,l}(p;0,0,\chi)$ will be defined by the analytic continuation of the rhs of (3.2) in D.

An important simplification of (3.2) is possible if we compare the integrals over x with $\chi=0$ and $\chi=p\cdot n$ for $l'=1$. According to (2.9b) and (2.10b) we may write

$$\begin{aligned}
&\int_0^1 dx\, x^n U_{\alpha,1}(-xp;\ -xp^2, p\cdot n)\\
&\quad=(-2)^{\alpha+1}\Gamma\Big(\alpha+1-\frac{D}{2}\Big)(-p^2)^{-\alpha-1+D/2}\\
&\qquad\times\int_0^1 dx\, x^n(x(1-x))^{-\alpha-1+D/2}\\
&\qquad\times{}_2F_1\Big(\alpha+1-\frac{D}{2},1;\frac{3}{2};-\frac{(1-x)(p\cdot n)^2}{xp^2n^2}\Big).
\end{aligned}$$

If we now replace the integration variable x by $1-x$ then this integral becomes

$$(-2)^{\alpha+1}\Gamma\Big(\alpha+1-\frac{D}{2}\Big)(-p^2)^{-\alpha-1+D/2}$$
$$\times\int_0^1 dx(1-x)^n(x(1-x))^{-\alpha-1+D/2}$$
$$\times{}_2F_1\Big(\alpha+1-\frac{D}{2},1,\frac{3}{2};-\frac{(1-x)(p\cdot n)^2}{xp^2n^2}\Big)$$
$$=\int_0^1 dx(1-x)^n U_{\alpha,1}(-xp;-xp^2,0).$$

This corresponds to the shift of integration variable $q\to q-p$ in (1.4). So we find the relation

$$\int_0^1 dx\,x^n U_{\alpha,1}(-xp;-xp^2,p\cdot n)$$
$$=\int_0^1 dx(1-x)^n U_{\alpha,1}(-xp;-xp^2,0), \tag{3.3}$$

and therefore the integrals on the rhs of (3.2) only have to be evaluated for $\chi=0$.

If $\alpha_{\min}<2$ then the lower range of summation for α can be restricted to $\alpha\geqslant 2$ with the use of (2.24) and (2.25). Since eventually we want to separate that part of $I^{\mu_1\cdots\mu_N}_{\alpha_1,\alpha_2,l}(p;0,0,\chi)$ that contains a divergence if $D\to 4$, it will be convenient to define the complex variable

$$\epsilon=2-D/2 \tag{3.4}$$

and to study the singularites of the rhs of (3.2) as a function of ϵ. With this notation and (3.1), Eqs. (2.24) and (2.25) can be written as, respectively,

$$U_{\alpha,0}(-xp;-xp^2,0)$$
$$=[x(1-x)p^2/2(\alpha-2+\epsilon)]\,U_{\alpha+1,0}(-xp;-xp^2,0) \tag{3.5}$$

and

$$U_{\alpha,1}(-xp;-xp^2,0)$$
$$=-[1/(2\alpha-3+2\epsilon)]\{U_{\alpha+1,0}(-xp;-xp^2,0)$$
$$-x(1-\eta x)p^2U_{\alpha+1,1}(-xp;-xp^2,0)\}, \tag{3.6}$$

where

$$\eta=1-\xi,\quad \xi=(p\cdot n)^2/p^2n^2. \tag{3.7}$$

With the aid of Eq. (3.3) and the three formulas above, Eq. (3.2) can be brought into the form

$$I^{\mu_1\cdots\mu_N}_{\alpha_1,\alpha_2,l}(p;0,0,\chi)$$
$$=\sum_{\alpha\geqslant 2}\sum_{l'=0,1}\sum_{n\geqslant 0}\tilde T^{\mu_1\cdots\mu_N}_{\alpha,l',n}(p,n,g;\xi,\epsilon)$$
$$\times\int_0^1 dx\,x^nU_{\alpha,l'}(-xp;-xp^2,0) \tag{3.8}$$

and we only have to consider integrals of the type

$$W_{\alpha,l,n}(p^2,\xi)=\int_0^1 dx\,x^nU_{\alpha,l}(-xp;-xp^2,0), \tag{3.9}$$

for $\alpha\geqslant 2$, $l=0$ or $l=1$, and $n\geqslant 0$. Explicit formulas for the tensors $\tilde T^{\mu_1\cdots\mu_N}_{\alpha,l',n}$ will not be given as they follow from the recurrence relations of Sec. II. So we will be contented with studying (3.9) in the remaining part of this section.

If $\alpha=2$ then, with (2.9b), (3.1), and (3.4) we get

$$W_{2,0,n}(p^2,\xi)=4\Gamma(\epsilon)(-p^2)^{-\epsilon}B(n+1-\epsilon,1-\epsilon),$$

where the symbol B represents the beta function. It has the following Taylor series expansion around $\epsilon=0$:

$$B(n+1-\epsilon,1-\epsilon)$$
$$=\frac{1}{n+1}\Big\{1+\epsilon\Big(\frac{2}{n+1}+\sum_{l=1}^{n}\frac{1}{l}\Big)+\mathcal{O}(\epsilon^2)\Big\}.$$

Thus, for $n\geqslant 0$,

$$W_{2,0,n}(p^2,\xi,\epsilon)$$
$$=\frac{4}{n+1}(-p^2)^{-\epsilon}\Big\{\Gamma(\epsilon)+\frac{2}{n+1}+\sum_{l=1}^{n}\frac{1}{l}+\mathcal{O}(\epsilon)\Big\} \tag{3.10}$$

The occurrence of $\Gamma(\epsilon)$ with its pole at $\epsilon=0$ is interpreted as a UV divergence.

Next we consider $W_{2,1,n}$. Using (2.9b) and (2.10b), we may write

$$W_{2,1,n}(p^2,\xi,\epsilon)=(8/p^2)(-p^2)^{-\epsilon}\Phi_{n-1}(\xi,\epsilon), \tag{3.11}$$

where we have defined

$$\Phi_n(\xi,\epsilon)=\int_0^1 dx\,x^nh(x,\xi,\epsilon), \tag{3.12}$$

with

$$h(x,\xi,\epsilon)=\Gamma(1+\epsilon)x^{-\epsilon}(1-x)^{-1-\epsilon}$$
$$\times{}_2F_1(1+\epsilon,1;\tfrac{3}{2};-x\xi/(1-x)), \tag{3.13}$$

for integer n. The integral (3.12) exists for $\epsilon\to 0$, only if $n\geqslant 0$, and, in that case we define

$$\Phi_n(\xi)=\Phi_n(\xi,0)=\int_0^1 dx\,x^nh(x,\xi), \tag{3.14}$$

where now

$$h(x,\xi)=h(x,\xi,0)$$
$$=\frac{1}{2\sqrt{x(1-\eta x)\xi}}\log\frac{1+\sqrt{x\xi/(1-\eta x)}}{1-\sqrt{x\xi/(1-\eta x)}}. \tag{3.15}$$

For $n\geqslant 1$ the function $\Phi_n(\xi)$ can be related to $\Phi_0(\xi)$, or

$$\Phi_0(\xi)=\int_0^1 dt\frac{1}{\xi+t^2(1-\xi)}\log\frac{1+t}{1-t}, \tag{3.16}$$

by using the following recurrence relation:

$$\Phi_1(\xi)=(1/2\eta)(\Phi_0(\xi)-\log(4\xi)), \tag{3.17a}$$
$$\Phi_n(\xi)=(1/2n\eta)\{[2n-1+2(n-1)\eta]\Phi_{n-1}(\xi)$$
$$-(2n-3)\Phi_{n-2}(\xi)-1/(n-1)\},\quad n\geqslant 2. \tag{3.17b}$$

For $n<0$ and $\operatorname{Re}\epsilon<-n+1$ the following recurrence relation holds:

$$\Phi_n(\xi,\epsilon)$$
$$=-\frac{1}{2n+1-2\epsilon}\Big\{\Gamma(1+\epsilon)B(n+1-\epsilon,1-\epsilon)$$
$$-[2n+3-2\epsilon+2\eta(n+1-2\epsilon)]\Phi_{n+1}(\xi,\epsilon)$$
$$+2\eta(n+2-\epsilon)\Phi_{n+2}(\xi,\epsilon)\Big\}. \tag{3.18}$$

With this formula, $\Phi_n(\xi,\epsilon)$ can be expressed as a sum of beta functions and the functions $\Phi_0(\xi,\epsilon)$ and $\Phi_1(\xi,\epsilon)$. This enables the analytic continuation of $\Phi_n(\xi,\epsilon)$ to

$\mathrm{Re}\,\epsilon > -n+1$, where its singularities are determined by the beta functions. The singularity at the point $\epsilon = 0$ is a pole which is interpreted as an IR divergence.

We next consider the case $\alpha \geqslant 3$ which is also relevant for calculations in nonperturbative quantum chromodynamics and quantum gravity.[5,8] As a shorthand notation we define the function

$$g(x,\epsilon) = \Gamma(1+\epsilon)x^{-\epsilon}(1-x)^{-\epsilon}. \tag{3.19}$$

Then (2.26) and (2.27) for $\chi = 0$, together with (2.9b), (2.10b), (3.1), and (3.12), give the following expressions for $W_{\alpha,l,n}$:

$$\begin{aligned} W_{\alpha,0,n}(p^2,\xi) &= (-p^2)^{-\epsilon}(p^2)^{-\alpha+2}2^{\alpha}\Gamma(\alpha-2+\epsilon) \\ &\quad \times B(n-\alpha+3-\epsilon,\ -\alpha+3-\epsilon), \\ &\quad \alpha \geqslant 2, \end{aligned} \tag{3.20}$$

and

$$\begin{aligned} W_{\alpha,1,n}(p^2,\xi) &= (-p^2)^{-\epsilon}(p^2)^{-\alpha+1}2^{\alpha}\frac{\Gamma(\alpha-\frac{3}{2}+\epsilon)}{\Gamma(\frac{1}{2}+\epsilon)}\int_0^1 dx\left\{\sum_{k=0}^{\alpha-3}\frac{\Gamma(\frac{1}{2}+\epsilon)\Gamma(\alpha-k-2+\epsilon)}{\Gamma(\alpha-\frac{3}{2}-k+\epsilon)}\right. \\ &\quad \left.\times x^{-\alpha+1}(1-x)^{-\alpha+k+2}(1-\eta x)^{-k-1}g(x,\epsilon) + 2x^{-\alpha+2}(1-\eta x)^{-\alpha+2}h(x,\xi,\epsilon)\right\}, \quad \alpha \geqslant 3. \end{aligned} \tag{3.21}$$

The integrand of (3.21) depends on the integration variable x through terms that are not the product of a rational function of x and either $g(x,\epsilon)$ or $h(x,\xi,\epsilon)$. The rational functions are of the form $x^n(1-x)^{-m_1}(x-1/\eta)^{-m_2}$, where m_1 and m_2 are non-negative integers and n can take on integer values. They can be rewritten as sums of partial fractions, so that $W_{\alpha,1,n}(p^2,\xi)$ becomes a linear combination of integrals of the following type:

(a) $\displaystyle\int_0^1 dx\, x^n g(x,\epsilon) = \Gamma(1+\epsilon)(-p^2)^{-\epsilon}B(n+1-\epsilon,\ 1-\epsilon),$

(b) $\displaystyle\int_0^1 dx(1-x)^{-m}g(x,\epsilon) = \Gamma(1+\epsilon)(-p^2)^{-\epsilon}B(1-\epsilon,\ -m+1-\epsilon),$

(c) $\displaystyle\int_0^1 dx\left(x-\frac{1}{\eta}\right)^{-m}g(x,\epsilon) = G_m(\xi,\epsilon),$

(d) $\displaystyle\int_0^1 dx\, x^n h(x,\xi,\epsilon) = \Phi_n(\xi,\epsilon),$

(e) $\displaystyle\int_0^1 dx\left(x-\frac{1}{\eta}\right)^{-m}h(x,\xi,\epsilon) = H_m(\xi,\epsilon).$

This list exhausts all possible integrals that occur in the process of decomposing the rational part of the intergrand of $W_{\alpha,1,n}(p^2,\xi)$ into partial fractions. Furthermore, m is always a positive integer and n an integer. The function $\Phi_n(\xi,\epsilon)$ has been studied before so the only new types of integrals are given by (c) and (e). Since $m>0$ the functions $G_m(\xi,\epsilon)$ and $H_m(\xi,\epsilon)$ are well defined in the limit $\epsilon \to 0$. The former is given by

$$\begin{aligned} &G_1(\xi,0) = \log\xi, \\ &G_m(\xi,0) = [(-1)^m/(m-1)]\eta^{m-1}(\xi^{-m+1}-1), \end{aligned} \tag{3.22}$$

for $m \geqslant 2$.

The latter, $H_m(\xi,0)$, can again be calculated via a recurrence relation, which we will now describe.

Define

$$H_m(\xi,0) = (-1)^m\eta^m\Psi_m(\xi)/\xi, \tag{3.23}$$

where

$$\Psi_m(\xi) = \int_0^1 dt\left[1+\left(\frac{1}{\xi}-1\right)t^2\right]^{m-1}\log\frac{1+t}{1-t}, \tag{3.24}$$

then the following properties of this function can be proved:

$$\Psi_0(\xi) = \xi\Phi_0(\xi), \tag{3.25a}$$

$$\Psi_1(\xi) = \log 4, \tag{3.25b}$$

and, for $m \geqslant 2$,

$$\begin{aligned} &\Psi_m(\xi)[1/(2m-1)]\{((2m-3)/\xi+2(m-1)) \\ &\quad \times\Psi_{m-1}(\xi) - [2(m-2)/\xi]\Psi_{m-2}(\xi) \\ &\qquad + [1/(m-1)](\xi^{-m+1}-1)\}. \end{aligned} \tag{3.25c}$$

The collection of formulas (3.9)–(3.25) provides an adequate algorithm for the calculation of the function $W_{\alpha,l,n}(p^2,\xi)$, defined by (3.9), with α and n integers, satisfying the conditions $\alpha \geqslant 2$, $n \geqslant 0$, and with the integrand given by (2.9b) for $l=0$ and (2.10b) for $l=1$. We conclude that the finite part of the integral (3.2), after application of this algorithm, is expressed in terms of p^μ, n^μ, $g^{\mu\nu}$, the scalars $p^2, p\cdot n, n^2$, the dimensionless quantity $\xi = (p\cdot n)^2/p^2n^2$, and the functions $\log\xi$ and $\Phi_0(\xi)$. The singularities of (3.2) at $\epsilon = 0$ are either of the form $\Gamma(\xi)$ or $B(-\epsilon, 1-\epsilon)$, corresponding to, respectively, UV and IR divergences.

IV. CONCLUSIONS

In this paper I have described an algorithm for the calculation of Feynman integrals related to the propagators of non-Abelian gauge theories in axial gauge. Besides giving in Sec. III a complete treatment of integrals associated with massless particles, generally useful recurrence relations also were given in Sec. II. The latter also apply to the case of massive particles, i.e., both m_1 and m_2 unequal to zero in formula (1.1) and an algorithm, similar to that of Sec. III for massive Feynman integrals in the axial gauge can be found.[8]

As the algorithms are based on recurrence relations they can be optimally used if implemented in a computer program for algebraic computation.[9,10]

After completion of this work I was informed about similar results, claimed by Lee and Milgram.[11] These authors have given explicit formulas for massless Feynman integrals in the form of truncated and infinite power series. For practical calculations, which should preferably be done on a computer, the use of recurrence relations seems to be advantageous above an explicit formula in the form of power series. The hard part of the calculation of a large set of integrals (1.1) is the simplification to sums of basic integrals, i.e., (1.4) or (3.16), which cannot be done analytically anyhow.

The foregoing argument will hold even more for the case where all particles are massive and for which explicit formulas, as in Ref. 11, have not been given yet. Here recurrence relations may provide the only viable way to tackle Feynman integrals.[8]

ACKNOWLEDGMENTS

I wish to thank Associate Professor B. G. Kenny for reading the manuscript, Dr. E. L. Heck for useful discussions, and Professor R. Delbourgo for bringing Ref. 11 to my attention.

This work is supported by a grant from the Australian Research Grants Scheme (A.R.G.S.).

[1]R. Delbourgo, A. Salam, and J. Strathdee, Nuovo Cimento A **23**, 237 (1974); W. Kummer, Acta. Phys. Austriaca **41**, 315 (1975).
[2]R. Delbourgo, J. Phys. G **5**, 603 (1979); J. Phys. A **14**, 753 (1981).
[3]M. Baker, J. S. Ball, and F. Zachariasen, Nucl. Phys. B **186**, 531, 560 (1981), and references therein; D. Atkinson, P. W. Johnson, W. J. Schoenmaker, and H. A. Slim, Nuovo Cimento A **77**, 197 (1983).
[4]A. I. Alekseev, Yad. Fiz. **33**, 516 (1981) [Sov. J. Nucl. Phys. **33**, 269 (1981)]; W. J. Schoenmaker, Nucl. Phys. B **194**, 535 (1982); M. Baker, J. S. Ball, and F. Zachariasen, Nucl. Phys. B **226**, 455 (1983); D. Atkinson and P. W. Johnson, Nucl. Phys. B **241**, 189 (1984).
[5]E. L. Heck and H. A. Slim, "On the infra-red behaviour of the gluon vacuum polarization in axial gauge," Nuovo Cimento A (to appear).
[6]D. M. Capper and G. Leibbrandt, Phys. Rev. D **25**, 1009 (1982); R. Delbourgo, J. Phys. A **14**, L235, 3123 (1981).
[7]B. Winter, Ph.D. thesis, University of Tasmania, 1984.
[8]H. A. Slim (to appear).
[9]M. Veltman, computercode SCHOONSCHIP; H. Strubbe, Comput. Phys. Commun. **8**, 1 (1974).
[10]A. C. Hearn, computercode REDUCE.
[11]H. C. Lee and M. S. Milgram, Phys. Lett. B **132**, 397 (1983); **133**, 320 (1983); Ann. Phys. **157**, 408 (1984).

Character table for the 1080-element point-group-like subgroup of SU(3)

Henrik Flyvbjerg
Service de Physique Théorique, Centre d'Etudes Nucléaires, 91191 Gif-sur-Yvette Cedex, France

(Received 13 December 1984; accepted for publication 19 April 1985)

The character table for $S(1080)$, the largest point-group-like subgroup of SU(3), is presented. So is the reduction of the first 28 SU(3) characters to $S(1080)$ characters. These tables are needed in a recently proposed systematic description of SU(3)-invariant lattice gauge theories by effective $S(1080)$-invariant theories. Problems encountered in an alternative systematics, using only *local* $S(1080)$-invariant theories, are discussed.

I. INTRODUCTION

The color symmetry group of QCD, SU(3), has a largest point-group-like subgroup $S(1080)$ with 1080 elements.[1,2] It has recently been suggested that numerical simulations of QCD on a lattice might be facilitated by the use of $S(1080)$ instead of SU(3), when the differences between the two groups, the degrees of freedom residing in the coset SU(3)/$S(1080)$, are properly accounted for.[3,4] In doing so, arguments based upon invariance with respect to $S(1080)$ are extensively used. Hence, a knowledge of the irreducible representations of $S(1080)$ is necessary. So is a knowledge of the character expansion in $S(1080)$ of SU(3) characters restricted to $S(1080)$. Both are presented here.

The present study of $S(1080)$ is useful not only for a systematic decimation of SU(3) to $S(1080)$ along the lines advocated in Refs. 3 and 4. Starting with a local gauge theory with fields in SU(3), the outcome of such a decimation is at best a quasilocal gauge theory with fields in $S(1080)$. The latter theory may be preferable to the former in numerical simulations, because it drastically reduces memory requirements, and increases the speed of multiplication. But the quasilocality of the effective action is an unattractive feature, which reduces the advantages of the effective $S(1080)$ theory. Therefore it may be useful to notice that a given local gauge theory with a continuous gauge group may be well approximated by a gauge theory with a discrete subgroup for a whole range of effective actions.[5] This universality with respect to change of both action *and* group is just an extension of the much-studied universality amongst lattice actions for fixed gauge groups.[7–14] With this universality in mind, we may try to construct the one-plaquette effective action of $S(1080)$, which gives the best possible approximation to a given SU(3)-invariant theory. If this approximation is not found satisfactory, we may proceed to search amongst the simplest actions involving also two plaquettes. The building blocks for such construction are the characters of $S(1080)$, and the resolutions of SU(3) characters on them. This is the other reason for the present study.

In Sec. II the characters of $S(1080)$ are found and their table given. In Sec. III the restrictions to $S(1080)$ of the 28 lowest-lying representations of SU(3) are expanded on $S(1080)$ characters and the problems encountered in any attempt to describe a local SU(3)-invariant theory by a *local* effective $S(1080)$-invariant theory are discussed.

II. THE CHARACTERS OF $S(1080)$

The group $S(1080)$ covers three times a 360-element subgroup of SU(3)/Z_3, which is isomorphic with A_6, the alternating group of six letters.[1,2] Hence $S(1080)$ has four generators. In the three-dimensional representation corresponding to the fundamental representation of SU(3), the generators may be chosen as[2]

$$A=\begin{Bmatrix}1 & 0 & 0\\ 0 & -1 & 0\\ 0 & 0 & -1\end{Bmatrix}, \tag{2.1}$$

$$E=\begin{Bmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ 1 & 0 & 0\end{Bmatrix}, \tag{2.2}$$

$$W=-\frac{1}{2}\begin{Bmatrix}1 & \mu_2 & \mu_1\\ \mu_2 & \mu_1 & 1\\ \mu_1 & 1 & \mu_2\end{Bmatrix}, \tag{2.3}$$

$$B=\begin{Bmatrix}-1 & 0 & 0\\ 0 & 0 & \omega^*\\ 0 & \omega & 0\end{Bmatrix}, \tag{2.4}$$

where

$$\mu_1\equiv(1-\sqrt{5})/2,\quad \mu_2\equiv(1+\sqrt{5})/2, \tag{2.5}$$

and

$$\omega\equiv\exp(i\pi/3)=(1+i\sqrt{3})/2. \tag{2.6}$$

We remark that A, E, and W are real, and generate the 60-element icosahedral group in its defining representation. Since $S(1080)$ has only 1080 elements it is easily generated and studied on a computer. It has 17 different classes, shown with the classes of SU(3) in Fig. 1. The classes of $S(1080)$ were found by sorting its elements according to the values of their traces, and checking for transitivity of the adjoint action of $S(1080)$ on elements with the same trace values. Only the elements with trace zero do not form a single class. They fall in two classes, each containing 120 elements.

Any irreducible representation of SU(3), $\mathscr{D}_{(\lambda,\mu)}$, $\lambda,\mu\in\{0,1,2,3,\dots\}$, is also a representation of $S(1080)$, though not necessarily an irreducible one. Hence, in terms of irreducible representations $\mathscr{D}'_r$ of $S(1080)$, we may write

$$\mathscr{D}_{(\lambda,\mu)}=\bigoplus_{r\in(\lambda,\mu)}\mathscr{D}'_r, \tag{2.7}$$

 0022-2488/85/112985-05$02.50

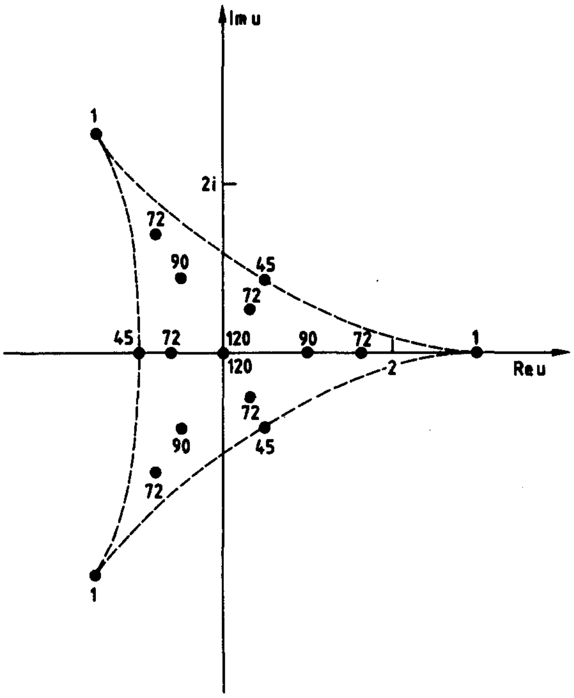


FIG. 1. The number of elements in each class of $S(1080)$ is shown next to the complex value of trace u, for u belonging to the class. There are 17 classes, two of them having vanishing trace. The values of trace U, $U \in SU(3)$, fall within the dashed lines.

where $r \in (\lambda,\mu)$ is a shorthand notation for $\mathscr{D}'_r$ being contained in $\mathscr{D}_{(\lambda,\mu)}$. Taking the trace on both sides, we obtain

$$\chi_{(\lambda,\mu)} = \sum_r m_{(\lambda,\mu),r} \chi'_r, \tag{2.8}$$

where $m_{(\lambda,\mu),r}$ is a non-negative integer equal to the number of times $\mathscr{D}'_r$ occurs in $\mathscr{D}_{(\lambda,\mu)}$. We introduce the scalar product amongst class functions on $S(1080)$,

$$(\chi,\psi) = \frac{1}{1080} \sum_{\text{classes } c} n_c \chi^*(c) \psi(c), \tag{2.9}$$

where n_c is the number of elements in the class c. In this product

$$(\chi_{(\lambda,\mu)}, \chi_{(\lambda',\mu')}) = \sum_r m_{(\lambda,\mu),r} m_{(\lambda',\mu'),r}. \tag{2.10}$$

These products are shown in Table I for the 28 irreducible representations $\mathscr{D}_{(\lambda,\mu)}$ of SU(3) having $\lambda + \mu \leqslant 6$.

From Table I we see that the eight irreducible representations of SU(3) characterized by

$$(\lambda,\mu) = (0,0),\ (1,0),\ (0,1),\ (1,1),\ (2,0),$$
$$(0,2),\ (2,1),\ \text{and}\ (1,2), \tag{2.11}$$

also are irreducible representations of $S(1080)$. So are $\mathscr{D}_{(3,0)}$ and $\mathscr{D}_{(0,3)}$, but they form one, self-conjugate representation, when considered as representations of $S(1080)$. These nine representations $\mathscr{D}'_1,\ldots,\mathscr{D}'_9$, and their corresponding characters $\chi'_1,\ldots,\chi'_9$, are the only ones identical for SU(3) and $S(1080)$, as we shall see.

Reading on in Table I, we find

$$(\chi_{(2,2)}, \chi_{(2,2)}) = \sum_r m^2_{(2,2),r} = 4. \tag{2.12}$$

Hence $\mathscr{D}_{(2,2)}$ is reducible, either to four different irreducible representations of $S(1080)$, or to one irreducible representation occurring twice. The second possibility is ruled out by the fact that $\mathscr{D}_{(2,2)}$ is 27 dimensional, i.e., has odd dimen-

TABLE I. Scalar products of Eq. (2.10).

(λ,μ)	(0,0)	(1,0)	(0,1)	(1,1)	(2,0)	(0,2)	(2,1)	(1,2)	(3,0)	(0,3)	(2,2)	(3,1)	(1,3)	(4,0)	(0,4)	(3,2)	(2,3)	(4,1)	(1,4)	(5,0)	(0,5)	(3,3)	(4,2)	(2,4)	(5,1)	(1,5)	(6,0)	(0,6)
(0,0)	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
(1,0)	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0
(0,1)	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0
(1,1)	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0	1	1
(2,0)	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	2	0	0	1	0	0
(0,2)	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2	1	0	0	0
(2,1)	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	2	0	0	0	0	1	0	0	2	2	0	0	0
(1,2)	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	2	0	0	1	0	0	2	0	0	2	0	0
(3,0)	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	1	0	0	2	0	0	0	0	0	0
(0,3)	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	1	0	0	2	0	0	0	0	0	0
(2,2)	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	2	2	0	0	5	0	0	0	0	3	3
(3,1)	0	0	0	0	0	0	0	1	0	0	0	2	0	0	1	0	3	0	0	1	0	0	4	0	0	3	0	0
(1,3)	0	0	0	0	0	0	1	0	0	0	0	0	2	1	0	3	0	0	0	0	1	0	0	4	3	0	0	0
(4,0)	0	0	0	0	0	1	0	0	0	0	0	0	1	2	0	1	0	0	0	0	0	0	0	4	2	0	0	0
(0,4)	0	0	0	0	1	0	0	0	0	0	0	1	0	0	2	0	1	0	0	0	0	0	4	0	0	2	0	0
(3,2)	0	0	0	0	0	0	2	0	0	0	0	0	3	1	0	6	0	0	0	0	3	0	0	6	5	0	0	0
(2,3)	0	0	0	0	0	0	0	2	0	0	0	3	0	0	1	0	6	0	0	3	0	0	6	0	0	5	0	0
(4,1)	0	0	0	1	0	0	0	0	1	1	2	0	0	0	0	0	0	4	4	0	0	6	0	0	0	0	2	2
(1,4)	0	0	0	1	0	0	0	0	1	1	2	0	0	0	0	0	0	4	4	0	0	6	0	0	0	0	2	2
(5,0)	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	3	0	0	3	0	0	2	0	0	3	0	0
(0,5)	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	3	0	0	0	0	3	0	0	2	3	0	0	0
(3,3)	0	0	0	1	0	0	0	0	2	2	5	0	0	0	0	0	0	6	6	0	0	12	0	0	0	0	5	5
(4,2)	0	0	0	0	2	0	0	2	0	0	0	4	0	0	4	0	6	0	0	2	0	0	12	0	0	8	0	0
(2,4)	0	0	0	0	0	2	2	0	0	0	0	0	4	4	0	6	0	0	0	0	2	0	0	12	8	0	0	0
(5,1)	0	1	0	0	0	1	2	0	0	0	0	0	3	2	0	5	0	0	0	0	3	0	0	8	7	0	0	0
(1,5)	0	0	1	0	1	0	0	2	0	0	0	3	0	0	2	0	5	0	0	3	0	0	8	0	0	7	0	0
(6,0)	1	0	0	1	0	0	0	0	0	0	3	0	0	0	0	0	0	2	2	0	0	5	0	0	0	0	5	5
(0,6)	1	0	0	1	0	0	0	0	0	0	3	0	0	0	0	0	0	2	2	0	0	5	0	0	0	0	5	5

sion. So we have encountered four more irreducible representations of $S(1080)$, $\mathscr{D}'_{10}, \mathscr{D}'_{11}, \mathscr{D}'_{12}$ and $\mathscr{D}'_{13}$, in

$$\mathscr{D}_{(2,2)} = \mathscr{D}'_{10} + \mathscr{D}'_{11} + \mathscr{D}'_{12} + \mathscr{D}'_{13}. \tag{2.13}$$

Two additional irreducible representations, a conjugate pair, occur in $\mathscr{D}_{(3,1)}$ and $\mathscr{D}_{(1,3)}$. These 24-dimensional representations are both reducible to two representations of $S(1080)$, one of which is $\mathscr{D}'_9 = \mathscr{D}_{(1,2)}$, respectively, $\mathscr{D}'_8 = \mathscr{D}_{(2,1)}$. The latter representations are 15 dimensional, hence we have found a conjugate pair of nine-dimensional representations $\mathscr{D}'_{14}$ and $\mathscr{D}'_{15}$. The corresponding characters are

$$\chi'_{14} = \chi_{(3,1)} - \chi_{(1,2)}, \tag{2.14}$$

$$\chi'_{15} = \chi_{(1,3)} - \chi_{(2,1)}. \tag{2.15}$$

Reading on in Table I, we find

$$\mathscr{D}_{(4,0)} = \mathscr{D}'_6 \oplus \mathscr{D}'_{15}, \tag{2.16}$$

$$\mathscr{D}_{(0,4)} = \mathscr{D}'_5 \oplus \mathscr{D}'_{14}, \tag{2.17}$$

i.e., both representations are reducible to representations of $S(1080)$ already encountered. New irreducible representations occur in $\mathscr{D}_{(3,2)}$ and $\mathscr{D}_{(2,3)}$

$$\mathscr{D}_{(3,2)} = \mathscr{D}'_7 \oplus \mathscr{D}'_7 \oplus \mathscr{D}'_{15} \oplus \mathscr{D}', \tag{2.18}$$

where $\mathscr{D}' = \mathscr{D}'_{16}$, and $\mathscr{D}'_{16}$ and its conjugate representation $\mathscr{D}'_{17}$, contained in $\mathscr{D}_{(2,3)}$, are three dimensional. The corresponding characters are

$$\chi'_{16} = \chi_{(3,2)} - \chi_{(2,1)} - \chi_{(1,3)}, \tag{2.19}$$

$$\chi'_{17} = \chi_{(2,3)} - \chi_{(1,2)} - \chi_{(3,1)}. \tag{2.20}$$

Now we have encountered all 17 irreducible representations of $S(1080)$. But we still have to find expressions for the four characters $\chi'_{10}, \chi'_{11}, \chi'_{12}$, and χ'_{13} making up $\chi_{(2,2)}$. Table I shows that $\mathscr{D}_{(4,1)}$, which equals $\mathscr{D}_{(1,4)}$, contains two of the four irreducible representations contained in $\mathscr{D}_{(2,2)}$. So does $\mathscr{D}_{(3,3)}$, and it contains one of them twice. A small calculation based on the table yields that they are the same two representations—let us call them $\mathscr{D}'_{12}$ and $\mathscr{D}'_{13}$, that they are eight and nine dimensional, and that the corresponding characters may be expressed as

$$\chi'_{12} = \chi_{(3,0)} + \chi_{(2,2)} + \chi_{(4,1)} - \chi_{(3,3)}, \tag{2.21}$$

$$\chi'_{13} = -\chi_{(1,1)} - 2\chi_{(3,0)} - \chi_{(2,2)} + \chi_{(3,3)}. \tag{2.22}$$

The two characters remaining to be found, χ'_{10} and χ'_{11}, cannot be expressed in terms of SU(3) characters. Only their *sum*

$$\chi'_{10} + \chi'_{11} = \chi_{(2,2)} - \chi'_{12} - \chi'_{13} \tag{2.23}$$

is a class function also on SU(3). As all such functions, including the hitherto-determined characters of $S(1080)$, it takes the same value on both classes of $S(1080)$ elements having zero trace. This is so, because the traceless elements in SU(3) form *one* class within SU(3). But since $\chi'_{10} - \chi'_{11}$ is the last class function to be determined, it is uniquely characterized as being orthogonal to $\chi'_1, \ldots, \chi'_9$, $\chi'_{10} + \chi'_{11}, \chi'_{12}, \ldots, \chi'_{17}$ with norm $\sqrt{2}$ and a non-negative, integer value on the group's unit element.

The character table of $S(1080)$ is given in Table II. In the seven representations $\mathscr{D}'_1, \mathscr{D}'_4, \mathscr{D}'_9, \mathscr{D}'_{10}, \mathscr{D}'_{11}, \mathscr{D}'_{12}$, and $\mathscr{D}'_{13}$ the center Z_3 of $S(1080)$ is represented by the unit matrix. Hence, they are also representations of $S(1080)/Z_3$, and the seven corresponding characters have already been given in Table IV of Ref. 2.

TABLE II. Character table for $S(1080)$. The first column numbers the characters as in the text. The first row numbers the classes according to decreasing values of the real part of the character in the fundamental representation, i.e., χ'_2. The second row shows the number of elements in each class, together with the kind of cycle generated by those elements. $\mu_1 = (1-\sqrt{5})/2$; $\mu_2 = (1+\sqrt{5})/2$; $\omega = (1+i\sqrt{3})/2$; $\omega^* = (1-i\sqrt{3})/2$.

CLASSES

CLASS	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
r	1E	$72C_5$	$90C_4$	$45C_6$	$45C'_6$	$72C_{15}$	$72C'_{15}$	$120C_3$	$120C'_3$	$90C_{12}$	$90C'_{12}$	$72C'_5$	$72C_{15}$	$72C'_{15}$	$45C_2$	$1C''_3$	$1C'''_3$
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	3	μ_2	1	ω	ω^*	$-\mu_1\omega$	$-\mu_1\omega^*$	0	0	$-\omega^*$	$-\omega$	μ_1	$-\mu_2\omega^*$	$-\mu_2\omega$	-1	$-3\omega^*$	-3ω
3	3	μ_2	1	ω^*	ω	$-\mu_1\omega^*$	$-\mu_1\omega$	0	0	$-\omega$	$-\omega^*$	μ_1	$-\mu_2\omega$	$-\mu_2\omega^*$	-1	-3ω	$-3\omega^*$
4	8	μ_2	0	0	0	μ_1	μ_1	-1	-1	0	0	μ_1	μ_2	μ_2	0	8	8
5	6	1	0	$-2\omega^*$	-2ω	$-\omega^*$	$-\omega$	0	0	0	0	1	$-\omega$	$-\omega^*$	2	-6ω	$-6\omega^*$
6	6	1	0	-2ω	$-2\omega^*$	$-\omega$	$-\omega^*$	0	0	0	0	1	$-\omega^*$	$-\omega$	2	$-6\omega^*$	-6ω
7	15	0	-1	ω	ω^*	0	0	0	0	ω^*	ω	0	0	0	-1	$-15\omega^*$	-15ω
8	15	0	-1	ω^*	ω	0	0	0	0	ω	ω^*	0	0	0	-1	-15ω	$-15\omega^*$
9	10	0	0	-2	-2	0	0	1	1	0	0	0	0	0	-2	10	10
10	5	0	-1	1	1	0	0	-1	2	-1	-1	0	0	0	1	5	5
11	5	0	-1	1	1	0	0	2	-1	-1	-1	0	0	0	1	5	5
12	8	μ_1	0	0	0	μ_2	μ_2	-1	-1	0	0	μ_2	μ_1	μ_1	0	8	8
13	9	-1	1	1	1	-1	-1	0	0	1	1	-1	-1	-1	1	9	9
14	9	-1	1	$-\omega^*$	$-\omega$	ω^*	ω	0	0	$-\omega$	$-\omega^*$	-1	ω	ω^*	1	-9ω	$-9\omega^*$
15	9	-1	1	$-\omega$	$-\omega^*$	ω	ω^*	0	0	$-\omega^*$	$-\omega$	-1	ω^*	ω	1	$-9\omega^*$	-9ω
16	3	μ_1	1	ω	ω^*	$-\mu_2\omega$	$-\mu_2\omega^*$	0	0	$-\omega^*$	$-\omega$	μ_2	$-\mu_1\omega^*$	$-\mu_1\omega$	-1	$-3\omega^*$	-3ω
17	3	μ_1	1	ω^*	ω	$-\mu_2\omega^*$	$-\mu_2\omega$	0	0	$-\omega$	$-\omega^*$	μ_2	$-\mu_1\omega$	$-\mu_1\omega^*$	-1	-3ω	$-3\omega^*$

CHARACTERS

III. REDUCTION OF SU(3) CHARACTERS TO $S(1080)$ CHARACTERS

Having found the characters of $S(1080)$, we can expand the SU(3) characters restricted to $S(1080)$ in this orthonormal basis

$$\chi_{(\lambda,\mu)} = \sum_{r=1}^{17} c_{(\lambda,\mu),r}\chi'_r, \quad \text{in } S(1080). \tag{3.1}$$

The expansion coefficients $c_{(\lambda,\mu),r}$ are given in Table III. This table is useful in the construction of effective $S(1080)$-invariant lattice gauge theories designed to approximate SU(3)-invariant ones. The problem to be circumvented by such an effective theory is the following.

Any SU(3)-invariant lattice gauge theory with a local Boltzmann weight exp s may have its Boltzmann weight expanded on SU(3) characters

$$\exp s(U) = \sum_{(\lambda,\mu)} \beta_{(\lambda,\mu)}\chi_{(\lambda,\mu)}(U), \quad U \in \mathrm{SU}(3). \tag{3.2}$$

Typically, $\beta_{(\lambda,\mu)}$ is positive and decreasing sufficiently fast with increasing $\lambda+\mu$ so that only a limited number of coefficients $\beta_{\lambda,\mu}$ need be considered.[15] Restricted to $S(1080)$, Eq. (3.2) reads

$$\exp s(n) = \sum_{n=1}^{17} \beta'_r\chi'_r(u), \quad u \in S(1080), \tag{3.3}$$

with

$$\beta'_r = \sum_{(\lambda,\mu)} \beta_{(\lambda,\mu)} c_{(\lambda,\mu),r}. \tag{3.4}$$

From Table III we read, for example, that

$$\beta'_1 = \beta_{(0,0)} + 2\beta_{(6,0)} + \cdots, \tag{3.5}$$

$$\beta'_2 = \beta'_3 = \beta_{(1,0)} + \beta_{(0,5)} + \beta_{(5,1)} + \cdots. \tag{3.6}$$

The first term in the series expansion of the free energy of the SU(3)-invariant theory is proportional to $\ln \beta_{(0,0)}$.[15] For the $S(1080)$-invariant theory the equivalent term is $\ln \beta'_1$. Hence, the $S(1080)$-invariant theory obtained by substituting $S(1080)$ for SU(3) in an SU(3)-invariant theory, with *no* accompanying change in the Boltzmann weight, has for this first term

$$\ln \beta'_1 = \ln(\beta_{(0,0)} + 2\beta_{(6,0)} + \cdots). \tag{3.7}$$

The discrepancy between this result, and the desired one $\ln \beta_{(0,0)}$, is of order six in β in the case of Wilson's action

$$s = \beta\tfrac{1}{6}(\chi_{(1,0)} + \chi_{(0,1)}). \tag{3.8}$$

The relative discrepancy between the desired result and the one obtained increases for higher coefficients; for example, everywhere, where $\beta_{(1,0)}$ is desired, $\beta_{(1,0)} + \beta_{(0,5)} + \cdots$ is obtained, i.e., a discrepancy of order 4 in β.

For the lower coefficients β'_r, the cure seems obvious. One should use an action for the approximate $S(1080)$ theory, which is *defined* by having $\beta'_1 = \beta_{(0,0)}, \beta'_2 = \beta_{(1,0)}$, etc., for the first eight coefficients. This strategy does lead to an $S(1080)$-invariant theory, which reproduces the series expansion of the SU(3)-invariant theory, that it is to approximate, to all 16 orders known.[15] For the present it is not known, however, how one optimally assigns coefficients β'_r to the

TABLE III. Coefficients $c_{(\lambda,\mu),r}$, $\lambda+\mu<6$, $r=1,2,\ldots,17$, defined in Eq. (3.1).

SU(3) \ $S(1080)$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
(0,0)	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
(1,0)	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
(0,1)	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
(1,1)	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
(2,0)	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
(0,2)	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
(2,1)	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
(1,2)	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
(3,0)	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
(0,3)	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
(2,2)	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0
(3,1)	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0
(1,3)	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0
(4,0)	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
(0,4)	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0
(3,2)	0	0	0	0	0	0	2	0	0	0	0	0	0	0	1	1	0
(2,3)	0	0	0	0	0	0	0	2	0	0	0	0	0	1	0	0	1
(4,1)	0	0	0	1	0	0	0	0	1	0	0	1	1	0	0	0	0
(1,4)	0	0	0	1	0	0	0	0	1	0	0	1	1	0	0	0	0
(5,0)	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
(0,5)	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0
(3,3)	0	0	0	1	0	0	0	0	2	1	1	1	2	0	0	0	0
(4,2)	0	0	0	0	2	0	0	2	0	0	0	0	0	2	0	0	0
(2,4)	0	0	0	0	0	2	2	0	0	0	0	0	0	0	2	0	0
(5,1)	0	1	0	0	0	1	2	0	0	0	0	0	0	0	1	0	0
(1,5)	0	0	1	0	1	0	0	2	0	0	0	0	0	1	0	0	0
(6,0)	1	0	0	1	0	0	0	0	0	1	1	0	1	0	0	0	0
(0,6)	1	0	0	1	0	0	0	0	0	1	1	0	1	0	0	0	0

last nine characters in the character expansion of the Boltzmann weight. So, whereas the problem just posed is well defined, the solution is not, at present. Some tools necessary for an attack on the problem have been presented here.

ACKNOWLEDGMENTS

It is a pleasure to thank J. B. Zuber for discussions and J. Raynal for reading the manuscript.

Partial support from the French Foreign Ministry and the Danish Research Council is acknowledged.

[1]G. A. Miller, H. F. Dickson, and L. E. Blichfeldt, *Theory and Applications of Finite Groups* (Stechert, New York, 1938), Chap. XII.

[2]W. M. Faibairn, T. Fulton, and W. H. Klink, J. Math. Phys. **5**, 1038 (1964).

[3]H. Flyvbjerg, Nucl. Phys. B **243**, 350 (1984).

[4]H. Flyvbjerg, Nucl. Phys. B **240**, 481 (1984).

[5]For example, SU(2) in four dimensions with Wilson's action and $\beta < 3$ is extremely well approximated by its 120-element icosahedral subgroup with the same action. But it is also well approximated by a quasilocal action obtained by a systematic decimation in Ref. 3. Presumably, it is well approximated by a multiparameter set of action containing the two just mentioned. Similar results have been otained for $Z(N)$-invariant theories which approximate a $U(1)$-invariant gauge theory.[4,6]

[6]H. Flyvbjerg, unpublished higher-order calculations extending results in Ref. 4.

[7]Yu. M. Makeenko and M. I. Polikarpov, Nucl. Phys. B **205**, 386 (1982).

[8]T. L. Chen, C. I. Tan, and Z. T. Zheng, Phys. Lett. B **109**, 383 (1982).

[9]S. Samuel, Phys. Lett. B **112**, 237 (1982).

[10]B. Grossman and S. Samuel, Phys. Lett. B **120**, 237 (1983).

[11]Yu. M. Makeenko, M. I. Polikarpov, and A. V. Zhelonkin, Phys. Lett. B **126**, 82 (1983).

[12]J. Jurkiewicz, C. P. Korthals Altes, and J. M. Dash, Nucl. Phys. B **233**, 457 (1984).

[13]B. Berg, Cargèse lectures 1983, DESY 84-012.

[14]References 7–12 are based on large-N arguments. For references to Monte Carlo studies, Symanzik's, and Wilson's actions, see the review by Berg.[13]

[15]J. M. Drouffe and J. B. Zuber, Phys. Rep. **102**, 1 (1983).

Soliton solutions for self-dual SU(3) gauge field theory

Demetrios Papadopoulos
Department of Astronomy, University of Thessaloniki, Thessaloniki, Greece

(Received 28 November 1984; accepted for publication 17 May 1985)

The Belinskii–Zakharov technique is applied to find a soliton solution for self-dual SU(3) gauge fields with axial symmetry on a Euclidean four-dimensional flat space. We find that for the special case of solutions generated from a special class of diagonal seed solutions the obtained soliton solutions reduce to quadratures.

I. INTRODUCTION

Yang[1] has written the self-dual SU(2) gauge field equations in complexified Euclidean space in the R gauge in terms of one real variable (ϕ) and two complex variables (ρ and $\bar\rho$). Prasad[2] has examined the self-dual SU(3) gauge fields in the R gauge in terms of two real variables (ϕ_1 and ϕ_2) and six complex variables ($\rho_1, \bar\rho_1, \rho_2, \bar\rho_2, \rho_3$, and $\bar\rho_3$).

The self-dual equations[3–5] may be derived by appropriate variations from Lagrangian densities for the SU(2) and SU(3) Lie groups. These are, respectively,

$$\mathscr{L}[\mathrm{SU}(2)] = \int d^4x \frac{1}{\phi^2} \times [(\phi_\mu \phi_\nu + \rho_\mu \bar\rho_\nu) g^{\mu\nu} + \epsilon^{\mu\nu} \rho_\mu \bar\rho_\nu], \tag{1.1}$$

$$\begin{aligned}\mathscr{L}[\mathrm{SU}(3)] &= \int d^4x \left\{ \frac{1}{\phi_1^2} [(\phi_{1\mu}\phi_{1\nu} + \phi_2 \rho_{1\mu} \bar\rho_{1\nu}) g^{\mu\nu} + \epsilon^{\mu\nu} \phi_2 \rho_{1\mu} \bar\rho_{1\nu}] \right.\\ &+ \frac{1}{\phi_2^2}[(\phi_{2\mu}\phi_{2\nu} + \phi_1 \rho_{2\mu}\bar\rho_{2\nu}) g^{\mu\nu} + \epsilon^{\mu\nu}\phi_1 \rho_{2\mu}\bar\rho_{2\nu}] \\ &+ \frac{1}{\phi_1\phi_2}[(\rho_{3\mu} - \rho_2\rho_{1\mu})(\bar\rho_{3\nu} - \bar\rho_2 \bar\rho_{1\nu})(g^{\mu\nu} + \epsilon^{\mu\nu}) \\ &\left. - \phi_{1\mu}\phi_{2\nu} g^{\mu\nu}] \right\},\end{aligned} \tag{1.2}$$

where the only nonvanishing elements in the metric $g_{\mu\nu}$ are

$$g_{y\bar y} = g_{\bar y y} = g_{z\bar z} = g_{\bar z z} = 1, \tag{1.3}$$

$$\epsilon^{\mu\nu} = \epsilon^{\mu\nu\lambda\sigma}(\hat y_\lambda \hat{\bar y}_\sigma + \hat z_\lambda \hat{\bar z}_\sigma),$$
$$\text{with } \epsilon^{\mu\nu\lambda\sigma} = 1, \quad \text{when } \mu\nu\lambda\sigma = y\bar y z \bar z, \tag{1.4}$$

and

$$d^4x \equiv dy\, d\bar y\, dz\, d\bar z. \tag{1.5}$$

The explicit forms of the self-dual field equations for the Euclidean SU(3) theory are

$$\begin{aligned}&(\partial_y \partial_{\bar y} + \partial_z \partial_{\bar z}) \ln \phi_1 + (\phi_2/\phi_1)(\rho_{1y}\bar\rho_{1\bar y} + \rho_{1z}\bar\rho_{1\bar z}) \\ &+ (1/\phi_1\phi_2)[(\rho_{3y} - \rho_2\rho_{1y})(\bar\rho_{3\bar y} - \bar\rho_2\bar\rho_{1\bar y}) \\ &+ (\rho_{3z} - \rho_2\rho_{1z})(\bar\rho_{3\bar z} - \bar\rho_2\bar\rho_{1\bar z})] = 0,\end{aligned} \tag{1.6a}$$

$$\begin{aligned}&(\partial_y \partial_{\bar y} + \partial_z \partial_{\bar z}) \ln(\phi_1/\phi_2) + (\phi_2/\phi_1)(\rho_{1y}\bar\rho_{1\bar y} + \rho_{1z}\bar\rho_{1\bar z}) \\ &- (\phi_1/\phi_2^2)(\rho_{2y}\bar\rho_{2\bar y} + \rho_{2z}\bar\rho_{2\bar z}) = 0,\end{aligned} \tag{1.6b}$$

$$\left[\frac{\rho_{3y} - \rho_2\rho_{1y}}{\phi_1\phi_2}\right]_{\bar y} + \left[\frac{\rho_{3z} - \rho_2\rho_{1z}}{\phi_1\phi_2}\right]_{\bar z} = 0, \tag{1.6c}$$

$$\left[\frac{\bar\rho_{3\bar y} - \bar\rho_2\bar\rho_{1\bar y}}{\phi_1\phi_2}\right]_y + \left[\frac{\bar\rho_{3\bar z} - \bar\rho_2\bar\rho_{1\bar z}}{\phi_1\phi_2}\right]_z = 0, \tag{1.6d}$$

$$\begin{aligned}&\left(\frac{\phi_2\rho_{1y}}{\phi_1^2}\right)_{\bar y} + \left(\frac{\phi_2\rho_{1z}}{\phi_1^2}\right)_{\bar z} \\ &- (i/\phi_1\phi_2)[\bar\rho_{2\bar y}(\rho_{3y} - \rho_2\rho_{1y}) \\ &+ \bar\rho_{3\bar z}(\rho_{3z} - \rho_2\rho_{3z})] = 0,\end{aligned} \tag{1.6e}$$

$$\begin{aligned}&\left(\frac{\phi_2\bar\rho_{1\bar y}}{\phi_2^2}\right)_y + \left(\frac{\phi_2\bar\rho_{1\bar z}}{\phi_1^2}\right)_z \\ &- (1/\phi_1\phi_2)[\rho_{2y}(\bar\rho_{3\bar y} - \bar\rho_2\bar\rho_{1\bar y}) \\ &+ \rho_{2z}(\bar\rho_{3\bar z} - \bar\rho_2\bar\rho_{1\bar z})] = 0,\end{aligned} \tag{1.6f}$$

$$\begin{aligned}&\left(\frac{\phi_2\rho_{2y}}{\phi_2^2}\right)_{\bar y} + \left(\frac{\phi_1\rho_{2z}}{\phi_2^2}\right)_{\bar z} \\ &+ (1/\phi_1\phi_2)[\bar\rho_{1\bar y}(\rho_{3y} - \rho_2\rho_{1y}) \\ &+ \bar\rho_{1\bar z}(\rho_{3z} - \rho_2\rho_{1z})] = 0,\end{aligned} \tag{1.6g}$$

$$\begin{aligned}&\left(\frac{\phi_1\bar\rho_{2\bar y}}{\phi_2^2}\right)_y + \left(\frac{\phi_1\bar\rho_{2\bar z}}{\phi_2^2}\right)_z \\ &+ (1/\phi_1\phi_2)[\rho_{1y}(\bar\rho_{3\bar y} - \bar\rho_2\bar\rho_{1\bar y}) \\ &+ \rho_{1z}(\bar\rho_{3\bar z} - \bar\rho_2\bar\rho_{1\bar z})] = 0,\end{aligned} \tag{1.6h}$$

where the subscripts denote partial differentiation and

$$\begin{aligned}\sqrt2 y &= x_1 + ix_2, \quad \sqrt2 z = x_3 - ix_4, \\ \sqrt2\, \bar y &= x_1 - ix_2, \quad \sqrt2\, \bar z = x_3 + ix_4,\end{aligned} \tag{1.7}$$

whereas x_μ ($\mu = 1,2,3,4$) are the complexified Cartesian coordinates.

Recently, Letelier[6] has found soliton solutions for self-dual SU(2) gauge fields extending the Belinskii–Zakharov[7] solution-generating technique used in general relativity. The purpose of this paper is to find explicit pure soliton solutions to Eqs. (1.6a)–(1.6h) for self-dual SU(3) gauge fields. Thus in Sec. II we present the Belinskii–Zakharov new solution-generating algorithm for the SU(3) group. In Sec.III we find explicitly the function Ψ_0 specializing the g_0 as in Eqs. (3.8). In Secs. IV and V we derive the one- and two-soliton solutions, respectively.

II. THE BELINSKII–ZAKHAROV NEW SOLUTION-GENERATING ALGORITHM

We restrict $\phi_1, \phi_2, \rho_1, \bar\rho_1, \rho_2, \bar\rho_2, \rho_3$, and $\bar\rho_3$ to be functions only of $r = \sqrt{2y\bar y}$ and $\omega = (1/\sqrt2)(z + \bar z)$ and we find that Eqs. (1.6) read as

$$\phi_1(\phi_{1rr}+(1/r)\phi_{1r}+\phi_{1\omega\omega})-\phi_{1r}^2-\phi_{1\omega}^2$$
$$+\phi_2(\rho_{1r}\bar\rho_{1r}+\rho_{1\omega}\bar\rho_{1\omega})$$
$$+(\phi_1/\phi_2)[(\rho_{3r}-\rho_2\rho_{1r})(\bar\rho_{3r}-\bar\rho_2\bar\rho_{1r})$$
$$+(\rho_{3\omega}-\rho_2\rho_{1\omega})(\bar\rho_{3\omega}-\bar\rho_2\bar\rho_{1\omega})]=0, \tag{2.1a}$$
$$(1/\phi_1^2)[\phi_1(\phi_{1rr}+\phi_{1r}/r+\phi_{1\omega\omega})-\phi_{1r}^2-\phi_{1\omega}^2$$
$$+\phi_2(\rho_{1r}\bar\rho_{1r}+\rho_{1\omega}\bar\rho_{1\omega})]$$
$$-(1/\phi_2^2)[\phi_2(\phi_{2rr}+(1/r)\phi_{2r}+\phi_{2\omega\omega})-\phi_{2r}^2-\phi_{2\omega}^2$$
$$+\phi_1(\rho_{2r}\bar\rho_{2r}+\rho_{2\omega}\bar\rho_{2\omega})]=0, \tag{2.1b}$$
$$\frac{2}{r\phi_1\phi_2}(\rho_{3r}-\rho_2\rho_{1r})+r\left[\frac{\rho_{3r}-\rho_2\rho_{1r}}{r\phi_1\phi_2}\right]_r$$
$$+\left[\frac{\rho_{3\omega}-\rho_2\rho_{1\omega}}{\phi_1\phi_2}\right]_\omega=0, \tag{2.1c}$$
$$\frac{2}{r\phi_1\phi_2}(\bar\rho_{3r}-\bar\rho_2\bar\rho_{1r})+r\left[\frac{\bar\rho_{3r}-\bar\rho_2\bar\rho_{1r}}{r\phi_1\phi_2}\right]_r$$
$$+\left[\frac{\bar\rho_{3\omega}-\bar\rho_2\bar\rho_{1\omega}}{\phi_1\phi_2}\right]_\omega=0, \tag{2.1d}$$
$$\frac{2\phi_2\rho_{1r}}{r\phi_1^2}+r\left(\frac{\phi_2\rho_{1r}}{r\phi_1^2}\right)_r+\left(\frac{\phi_2\rho_{1\omega}}{\phi_1^2}\right)_\omega$$
$$-(1/\phi_1\phi_2)[\bar\rho_{2r}(\rho_{3r}-\rho_2\rho_{1r})$$
$$+\bar\rho_{2\omega}(\rho_{3\omega}-\rho_2\rho_{1\omega})]=0, \tag{2.1e}$$
$$\frac{2\phi_2\bar\rho_{1r}}{r\phi_1^2}+r\left(\frac{\phi_2\bar\rho_{1r}}{r\phi_1^2}\right)_r+\left(\frac{\phi_2\bar\rho_{1\omega}}{\phi_1^2}\right)_\omega$$
$$-(1/\phi_1\phi_2)[\rho_{2r}(\bar\rho_{3r}-\bar\rho_2\bar\rho_{1r})$$
$$+\rho_{2\omega}(\bar\rho_{3\omega}-\bar\rho_2\bar\rho_{1\omega})]=0, \tag{2.1f}$$
$$\frac{2\phi_1\rho_{2r}}{r\phi_1^2}+r\left(\frac{\phi_1\rho_{2r}}{r\phi_2^2}\right)_r+\left(\frac{\phi_1\rho_{1\omega}}{\phi_2^2}\right)_\omega$$
$$+(1/\phi_1\phi_2)[\bar\rho_{1r}(\rho_{3r}-\rho_2\rho_{1r})$$
$$+\bar\rho_{1\omega}(\rho_{3\omega}-\rho_2\rho_{1\omega})]=0, \tag{2.1g}$$
$$\frac{2\phi_1\bar\rho_{2r}}{r\phi_2^2}+r\left(\frac{\phi_1\bar\rho_{2r}}{r\phi_2^2}\right)_r+\left(\frac{\phi_1\bar\rho_{1\omega}}{\phi_2^2}\right)_\omega$$
$$+(1/\phi_1\phi_2)[\rho_{1r}(\bar\rho_{3r}-\bar\rho_2\bar\rho_{1r})$$
$$+\rho_{1\omega}(\bar\rho_{3\omega}-\bar\rho_2\bar\rho_{1\omega})]=0. \tag{2.1h}$$

The system of equations (2.1) can be written as follows:

$$\partial_r U+\partial_\omega V=0, \tag{2.2}$$

where

$$U\equiv rg_r g^{-1},\quad V\equiv rg_z g^{-1}, \tag{2.3}$$

and

$$g=\frac{r^{2/\eta}}{\phi_1}\times\begin{pmatrix}1 & \bar\rho_1 & \bar\rho_3\\ \rho_1 & \rho_1\bar\rho_1+\phi_1^2/\phi_2 & \rho_1\bar\rho_3+\bar\rho_2(\phi_1^2/\phi_2)\\ \rho_3 & \bar\rho_1\rho_3+\rho_2(\phi_1^2/\phi_2) & \rho_2\bar\rho_2(\phi_1^2/\phi_2)+\phi_1\phi_2\end{pmatrix}. \tag{2.4}$$

From the definition (2.4) we have $g=g^{+1}$ and $\det g=r^{6/\eta}$. Note that in order to satisfy Eq. (2.2) we need $n=3$, e.g., $\det g=r^2$. However, we have included the parameter η to facilitate the comparison of the SU(3) case with the SU(2) one; we shall return to this point in the last section.

The Belinskii–Zakharov (BZ) method for finding soliton solutions of Eq. (2.1) focuses attention on the problem of finding a solution of g appearing in Eqs. (2.2) and (2.3). The key point is the association to the system (2.1) of the linear "eigenvalue" problem

$$D_r\Psi=[(rU+\lambda V)/(\lambda^2+r^2)]\Psi, \tag{2.5a}$$
$$D_\omega\Psi=[(rV-\lambda U)/(\lambda^2+r^2)]\Psi, \tag{2.5b}$$

where

$$D_r\equiv\partial_r+\frac{2\lambda r}{\lambda^2+r^2}\partial_\lambda,\quad D_\omega\equiv\partial_\omega-\frac{2\lambda^2}{\lambda^2+r^2}\partial_\lambda, \tag{2.6}$$

λ is a complex spectral parameter, and $\Psi(\lambda,r,\omega)$ is a three-dimensional complex matrix function, which satisfies the condition

$$\Psi(\lambda=0,r,\omega)=g_0. \tag{2.7}$$

The solitonic character of the solutions of Eqs. (2.2) is associated with solutions of the form

$$\Psi=\chi\Psi_0, \tag{28}$$

with

$$\chi\equiv I+\sum_{k=1}^N\frac{R_k}{\lambda-\mu_k}, \tag{2.9}$$

where Ψ_0 is a solution to Eqs. (2.5) for a particular seed solution to Eqs. (2.2). The R_k are complex matrix functions of r and ω only, and the μ_k are scalar complex functions of r and ω only. The number of solitons appearing in the solution depends on the number of poles existing in the matrix χ. Note that for $\lambda=0$, Eq. (2.8) gives

$$g=\chi(\lambda=0,r,\omega)g_0. \tag{2.10}$$

The condition $g^+=g$ is insured for the present case regarding the expression

$$g=\chi(-r^2/\bar\lambda,r,\omega)g_0[\chi(\lambda,r,\omega)]^+. \tag{2.11}$$

The last condition tells us that if $g_0=g_0^+$, then $g=g^+$ when g is given by (2.7).

Hence, knowing the Ψ_0, a solution Ψ can be generated by purely algebraic operations if one assumes that Ψ is the product of a three-dimensional matrix, with n simple poles in the complex λ plane and Ψ_0. Equation (2.11) shows that an n-soliton solution $g(r,\omega)$ can then be found. The explicit procedure is as follows: From Eqs. (2.5)–(2.11) one obtains

$$\mu_{k,r}=2r\mu_k/(r^2+\mu_k^2),\quad \mu_{k,\omega}=-2r^2/(r^2+\mu_k^2). \tag{2.12}$$

The solution of Eq. (2.12) is the so-called "pole trajectories" μ_k, which are

$$\mu_k=a_k-\omega\pm\sqrt{(a_k-\omega)^2+r^2},\quad k=1,2,..., \tag{2.13}$$

where the a_k are the arbitrary complex constants.

The new solution generated by this method is completely determined by the function Ψ_0 as follows:

$$g=\left(I-\sum_{k=1}^N\frac{R_k}{\mu_k}\right)g_0, \tag{2.14}$$

where $(R_k)_{ab}\equiv\eta_a^{(k)}m_b^{(k)}$, $\eta_a^{(k)}$ can be obtained from

$$\eta_a^{(k)}\equiv D_{lk}(N_a^{(l)}/\bar\mu_l), \tag{2.15}$$

where

$$N_a^{(l)}\equiv m_c^{(l)}(g_0)_{ca},\quad D_{lk}=(\Gamma_{lk})^{-1}, \tag{2.16}$$

$$\Gamma_{kl} \equiv (m^{(k)} \cdot \overline{m}^{(l)})/(r^2 + \mu_k \overline{\mu}_l) = \overline{\Gamma}_{lk}, \tag{2.17}$$

with

$$m^{(k)} \cdot \overline{m}^{(l)} \equiv m_a^{(k)} (g_0)_{ab} \overline{m}_b^{(l)} \tag{2.18}$$

and

$$m_a^{(k)} \equiv m_{oc}^{(k)} [\Psi_0^{-1}(\mu_k, r, \omega)]_{ca}. \tag{2.19}$$

Finally one can derive the matrix g by

$$g_{ab} = (g_0)_{ab} - \sum_{k,l} \frac{\overline{N}_a^{(k)} D_{kl} N_b^{(l)}}{\mu_k \overline{\mu}_l}. \tag{2.20}$$

Equation (2.20) is a solution to Eq. (2.2) but it is not a solution to the field equations because the matrix whose elements are (2.20) is Hermitian in general, and it is not possible to cast it in the form (2.4) since $\det g = r^2$. It can be shown that the determinant of g is

$$\det g = (-1)^N r^{2N} \left(\prod_{k=1}^{N} \frac{1}{\mu_k^2} \right) \det g_0. \tag{2.21}$$

The problem can be remedied by defining a new matrix

$$g^{\text{ph}} \equiv r g (\det g)^{-1/2}, \tag{2.22}$$

that satisfies the conditions $(g^{\text{ph}})^+ = g^{\text{ph}}$ and $\det g^{\text{ph}} = r^2$. One can prove that g^{ph} satisfies Eq. (2.2) whenever g is a solution.

Expression (2.21) shows that only for N even will the signature of g_0 and g coincide. For N odd one will need a nonphysical seed to obtain a matrix with the physical signature; e.g.,

$$g_0' = \text{diag}\{ r^{2/n} \phi_1^{-1}, \quad r^{2/n} \phi_1 \phi_2^{-1}, \quad -r^{2/n} \phi_2 \}, \tag{2.23}$$

with

$$\det g_0' = -r^2 \quad (n = 3). \tag{2.24}$$

III. THE FUNCTION Ψ_0

The function Ψ_0 obeys the differential equations (2.5) with g replaced by g_0, i.e.,

$$D_r \Psi_0 = [(rU_0 + \lambda V_0)/(\lambda^2 + r^2)] \Psi_0, \tag{3.1a}$$

$$D_\omega \Psi_0 = [(rV_0 - \lambda U_0)/(\lambda^2 + r^2)] \Psi_0, \tag{3.1b}$$

and

$$\Psi_0|_{\lambda = 0} = g_0. \tag{3.2}$$

Recently, Letelier has found exact solutions for Ψ_0 in the case of the SU(2) group when g_0 is a diagonal metric.[6] Also, Letelier[8] showed that for a general diagonal seed solution the integration of (3.1) along the pole's trajectories reduces to a single quadrature.

In this section we study the system of equations (3.1) with the boundary condition (3.2) when g_0 is diagonal for the SU(3) group.

The set of equations [(3.1) and (3.2)] is equivalent to the equations

$$D_r \Psi_0 = [(rU_0 + \lambda V_0)/(\lambda^2 + r^2)] \Psi_0, \tag{3.3a}$$

$$\lambda \partial_r \Psi_0 + r \partial_\omega \Psi_0 = V_0 \Psi_0, \tag{3.3b}$$

and

$$\Psi_0|_{\lambda = 0} = g_0. \tag{3.4}$$

Assuming that g_0 and Ψ_0 are diagonal matrices, Eqs. (3.3) and (3.4) yield

$$D_r \ln(\det \Psi_0) = 2r/(\lambda^2 + r^2), \tag{3.5a}$$

$$(\lambda \partial_r + r \partial_\omega) \ln(\det \Psi_0) = 0, \tag{3.5b}$$

with

$$(\det \Psi_0)_{\lambda = 0} = r^2. \tag{3.6}$$

A solution to Eqs. (3.5) with the condition (3.6) is

$$\det \Psi_0 = r^2 - 2\lambda\omega - \lambda^2. \tag{3.7}$$

Furthermore, we specialize g_0 as follows:

$$(g_0)_{11} = r^{2(1-b)/n} e^{X_1}, \tag{3.8a}$$

$$(g_0)_{22} = r^{2(1+b)/n} e^{X_2}, \tag{3.8b}$$

$$(g_0)_{33} = r^k e^{X_3} \quad (k = \text{const}), \tag{3.8c}$$

and

$$(g_0)_{22} = (g_0)_{21} = (g_0)_{13} = (g_0)_{31} = (g_0)_{23} = (g_0)_{32} = 0, \tag{3.3d}$$

where $X_i = -[a_i \omega + c_i (r^2/2 - \omega^2)]$ with $i = 1,2,3$ and a_i, c_i arbitrary real constants. Note that in this case Eq. (2.2) reduces to

$$(rX_{i,r})_r + (rX_{i,\omega})_\omega = 0, \tag{3.9}$$

i.e., to the usual Laplace equation in cylindrical coordinates. The condition $\det g = r^{6/n}$ $(n = 3)$ gives $\Sigma_{i=1}^3 X_i = 0$, which implies

$$\sum_{i=1}^3 a_i = \sum_{i=1}^3 c_i = 0. \tag{3.10}$$

Combining Eqs. (3.8), (3.3), and (3.4), one obtains

$$D_r \ln(\Psi_0)_{11} = \frac{2(1-b)\eta^{-1} r - c_1 r^3 - \lambda (a_1 r - 2c_1 \omega r)}{\lambda^2 + r^2}, \tag{3.11a}$$

$$(\lambda \partial_r + r \partial_\omega) \ln(\Psi_0)_{11} = -r(a_1 - 2c_1 \omega), \tag{3.11b}$$

$$\ln(\Psi_0)_{11}|_{\lambda = 0} = \ln g_0. \tag{3.12}$$

A straightforward computation gives

$$\begin{aligned}(\Psi_0)_{11} = {} & (r^2 - 2\lambda\omega - \lambda^2)^{(1-b)/n} \\ & \times \exp\{ -[a_1(\omega + \lambda/2) \\ & + c_1(r^2/2 - (\omega + \lambda/2)^2)] \}.\end{aligned} \tag{3.13}$$

Using a similar procedure and with the aid of Eq. (3.6) we obtain

$$\begin{aligned}(\Psi_0)_{22} = {} & (r^2 - 2\lambda\omega - \lambda^2)^{(1+b)/n} \\ & \times \exp\{ -[a_2(\omega + \lambda/2) \\ & + c_2(r^2/2 - (\omega + \lambda/2)^2)] \}.\end{aligned} \tag{3.14a}$$

and

$$\begin{aligned}(\Psi_0)_{33} = {} & (r^2 - 2\lambda\omega - \lambda^2)^{k/2} \\ & \times \exp\{ -[a_3(\omega + \lambda/2) \\ & + c_3(r^2/2 - (\omega + \lambda/2)^2)] \}.\end{aligned} \tag{3.14b}$$

In the case that g_0 has a negative determinant one can take the following expression for g_0' and Ψ_0':

$$(g_0')_{11} = (g_0)_{11}, \quad (g_0')_{22} = (g_0)_{22}, \quad (g_0')_{33} = -(g_0)_{33}, \tag{3.15}$$

and

$$(\Psi_0')_{11} = (\Psi_0)_{11}, \quad (\Psi_0')_{22} = (\Psi_0)_{22}, \quad (\Psi_0')_{33} = -(\Psi_0)_{33}. \tag{3.16}$$

IV. THE ONE-SOLITON SOLUTION IN SU(3)

According to the theory already established in Secs. II and III the one-soliton solution in SU(3) for odd number solitons can be written as the following:

$$g_{11} = -\frac{r^{2/n}}{\phi_1}\frac{|\mu|^2[|m^{(1)}_{02}\Psi_1\Psi_3|^2\phi_1\phi_2 - |m^{(1)}_{03}\Psi_1\Psi_2|^2 r^{k-2/n}(\phi_1^2/\phi_2)] + r^2|m^{(1)}_{01}\Psi_1\Psi_3|^2}{|\mu|^2\Delta'}, \tag{4.1}$$

$$g_{12} = \bar g_{21} = r^{2/n} m^{(1)}_{01}\bar m^{(1)}_{02}\frac{r^2+|\mu|^2}{|\mu|^2\Delta'}\frac{2a\bar\mu\phi_2}{\Psi_3}, \tag{4.2}$$

$$g_{13} = \bar g_{31} = -r^k m^{(1)}_{01}\bar m^{(1)}_{03}\frac{r^2+|\mu|^2}{|\mu|^2\Delta'}\frac{2a\bar\mu\phi_1}{\Psi_2\phi_2}, \tag{4.3}$$

$$g_{22} = r^{2/n}\phi_2\frac{|\mu|^2[(m^{(1)}_{01}\Psi_2\Psi_3|^2 + r^{k-2/n}|m^{(1)}_{03}\Psi_1\Psi_2|^2(\phi_1^2/\phi_2) - r^2\phi_1\phi_2|m^{(1)}_{02}\Psi_1\Psi_3|^2]}{|\mu|^2\Delta'}, \tag{4.4}$$

$$g_{23} = \bar g_{32} = r^k m^{(1)}_{02} m^{(1)}_{03}\frac{r^2+|\mu|^2}{|\mu|^2\Delta'}\frac{2a\bar\mu\phi_1^2}{\Psi_1}, \tag{4.5}$$

$$g_{33} = \frac{r^k\phi_1}{|\mu|^2\Delta'\phi_2}\{|\mu|^2[|m^{(1)}_{01}\Psi_2\Psi_3|^2 - |m^{(1)}_{02}\Psi_1\Psi_3|^2\phi_1\phi_2] - r^{2(1+k-1/n)}|m^{(1)}_{03}\Psi_1\Psi_2|^2(\phi_1^3/\phi_2)\}, \tag{4.6}$$

where

$$\Delta' \equiv |m^{(1)}_{01}\Psi_2\Psi_3|^2 - |m^{(1)}_{02}\Psi_1\Psi_3|^2\phi_1\phi_2 + |m^{(1)}_{03}\Psi_2\Psi_1|^2(\phi_1^2/\phi_2), \tag{4.7}$$

$$\mu = a - \omega + \sqrt{(a-\omega)^2 \pm r^2}, \tag{4.8}$$

$$\Psi_1\Psi_2\Psi_3 = -(-2\mu a)^{2/n + k/2}, \tag{4.9}$$

with

$$\phi_1 = r^{2b/n}e^{-X_1}, \quad \phi_2 = r^{2b/n}e^{X_2}. \tag{4.10}$$

The "physical" g is given by the expression

$$g^{\mathrm{ph}} = (|\mu|)/r)g. \tag{4.11}$$

V. TWO-SOLITON SOLUTION IN SU(3)

In the case where the matrix χ has two poles we have

$$\chi \equiv I + R_1/(\lambda - \mu_1) + R_2/(\lambda - \mu_2) \tag{5.1}$$

and

$$\begin{aligned} g_{ab} &= (g_0)_{ab} - \sum_{k,l}\frac{\bar N^{(k)}_a(\bar D_{kl})N^{(l)}_b}{\mu_k\bar\mu_l} \\ &= (g_0)_{ab} - \frac{\bar N^{(1)}_a N^{(1)}_b}{\mu_1\bar\mu_1}\frac{\bar\Gamma_{22}}{\Delta} + \frac{\bar N^{(1)}_a N^{(1)}_b}{\mu_1\bar\mu_2}\frac{\bar\Gamma_{21}}{\Delta} \\ &\quad + \frac{\bar N^{(2)}_a N^{(1)}_b}{\mu_2\bar\mu_1}\frac{\bar\Gamma_{12}}{\Delta} - \frac{\bar N^{(2)}_a N^{(2)}_b}{\mu_2\bar\mu_2}\frac{\bar\Gamma_{22}}{\Delta}, \end{aligned} \tag{5.2}$$

where

$$\bar\Delta = \Delta \equiv \frac{m^{(1)}\cdot\bar m^{(1)}}{r^2+|\mu|^2}\cdot\frac{m^{(2)}\cdot\bar m^{(2)}}{r^2+|\mu|^2} - \left|\frac{m^{(1)}\cdot\bar m^{(2)}}{r^2+|\mu|^2}\right|^2, \tag{5.3}$$

$$\begin{aligned} m^{(k)}\cdot\bar m^{(l)} &\equiv m^{(k)}_a(g_0)_{ab}\bar m^{(l)}_b = m^{(k)}_{01}\bar m^{(l)}_{01}(4a^2\mu_k\bar\mu_l)^{(b-1)/n} \\ &\quad\times r^{2(1-b)/n}\exp\{a_1[\omega + \tfrac12(\mu_k + \bar\mu_l)] \\ &\quad + c_1[r^2/2 - \omega^2 - \omega(\mu_k + \bar\mu_l) - \tfrac14(\mu_k^2 + \bar\mu_l^2]\} \\ &\quad + m^{(k)}_{02}\bar m^{(l)}_{02}(4a^2\mu_k\bar\mu_l)^{-(b+1)/n} \\ &\quad\times r^{2(1+b)/n}\exp\{a_2[\omega + \tfrac12(\mu_k + \bar\mu_l)] \\ &\quad + c_2[r^2/2 - \omega^2 - \omega(\mu_k + \bar\mu_l) - \tfrac14(\mu_k^2 + \bar\mu_l^2)]\} \\ &\quad + m^{(k)}_{03}\bar m^{(l)}_{03}(4a^2\mu_k\bar\mu_l)^{-k/2} \\ &\quad\times r^k\exp\{a_3[\omega + \tfrac12(\mu_k + \bar\mu_l)] \\ &\quad + c_3[r^2/2 - \omega^2 - \omega(\mu_k + \bar\mu_l) - \tfrac14(\mu_k^2 + \bar\mu_l^2)]\}, \end{aligned} \tag{5.4}$$

with

$$N^{(k)}_1 \equiv m^{(k)}_{01}(-2a\mu_k)^{(b-1)/n} r^{(1-b)/n} \times\exp\{\tfrac12 a_1\mu_k - c_1(\mu_k^2/2 + \omega\mu_k)\}, \tag{5.5a}$$

$$N^{(k)}_2 \equiv m^{(k)}_{02}(-2a\mu_k)^{-(b+1)/n} r^{(1+b)/n} \times\exp\{\tfrac12 a_2\mu_k - c_2(\mu_k^2/2 + \omega\mu_k)\}, \tag{5.5b}$$

$$N^{(k)}_3 \equiv m^{(k)}_{03}(-2a\mu_k)^{-k/2} r^{k/2} \times\exp\{\tfrac12 a_3\mu_k - c_3(\mu_k^2/2 + \omega\mu_k)\}, \tag{5.5c}$$

The physical g is obtained from the relation

$$g^{\mathrm{ph}} = (|\mu_1\mu_2|/r^2)g. \tag{5.6}$$

The procedure can be repeated n times to give the n-soliton solution in the SU(3) Lie group.

Note that for $n = 2$, $\phi_1 = \phi_2 = \phi$, $\rho_1 = \rho_2 = 0$, and $\rho_3 = \rho$ Eq. (2.4) gives the g of the SU(2) Lie group.[5,6] Furthermore, by setting $n = 2$ and $k = 0$ in Eqs. (3.8), (3.12), and (3.13), we find the corresponding g_0 and Ψ_0 of the SU(2) presented in Ref. 6.

ACKNOWLEDGMENTS

The author expresses his gratitude to Dr. P. S. Letelier for his comments for this article.

[1]C. N. Yang, Phys. Rev. Lett. 38, 1377 (1977).
[2]M. K. Prasad, Phys. Rev. D 17, 3243 (1978).

[3]Y. Brihaye, D. B. Fairlie, J. Nuyts, and R. G. Yates, J. Math. Phys. **19**, 2528 (1978).
[4]L. Witten, Phys. Rev. D **18**, 3877 (1978).
[5]D. Papadopoulos and L. Witten, Phys. Rev. D **24**, 3161 (1981).
[6]P. S. Letelier, J. Math. Phys. **23**, 1175 (1982).
[7]V. A. Belinski and V. E. Zakharov, Zh. Eksp. Teor. Fiz. **75**, 1955 (1978) [Sov. Phys. JETP **48**, 985 (1978)]; *ibid.* **77**, 3 (1979) [**50**, 1 (1978)].
[8]P. S. Letelier, J. Math. Phys. **25**, 2675 (1984).

Collectivity and geometry. IV. Sp(6) $\supset$ Sp(2)$\times$O(3) basis states for open shells

M. Moshinsky[a)]
Instituto de Física, Universidad Nacional Autonoma de México, Apartado Postal 20364, México, D. F. 01000, Mexico

M. Nicolescu and R. T. Sharp
Physics Department, McGill University, Montréal, PQ H3A 2T8, Canada

(Received 27 March 1985; accepted for publication 18 June 1985)

Using generating function methods, branching rules for Sp(6) $\supset$ Sp(2)$\times$O(3) are derived. The branching rules suggest an integrity basis, or set of elementary permissible diagrams, in terms of which the subgroup basis states are defined; they correspond to vibrational, or Bohr–Mottelson type, states in the nuclear symplectic model.

I. INTRODUCTION

We refer the reader to the earlier papers in this series[1–3] for general literature references and for a historical and physical introduction to the subject. The nuclear symplectic model combines the features of the Bohr–Mottelson and Elliott models.

The basis states of a nucleus of $n+1$ nucleons are taken to be the energy eigenstates of an isotropic $3n$-dimensional harmonic oscillator. The symmetry group is then SU($3n$) and the states are those of symmetric representations (all representation labels zero except the first one). The metaplectic irreducible representations (IR's) $[(1/2)^{3n}]$, $[(1/2)^{3n-1}, (3/2)]$ of Sp($6n$) are spanned by the SU($3n$) states of even, odd representation labels, respectively.

The physically significant subgroup of Sp($6n$) is Sp(6)$\times$O(n), and for the IR's of Sp($6n$) under consideration the Sp(6) and O(n) IR's are correlated; see Eq. (3.1) below. The Hamiltonian for nuclear collective motion is assumed to be in the enveloping algebra of Sp(6).

The Sp(6) basis states may be classified according to the U(3) subgroup, yielding Elliott or rotational type states, or according to the subgroup Sp(2)$\times$O(3), yielding Bohr–Mottelson, or vibrational type states. It is our purpose in this paper to derive the integrity basis, or elementary permissible diagrams (epd's) with their syzygies (incompatible products); they define vibrational, or Sp(2)$\times$O(3), type basis states for general IR's of Sp(6), corresponding to open shells.

In Sec. II we derive the generating function for Sp(6) $\supset$ Sp(2)$\times$O(3) branching rules and interpret it in terms of a finite set of epd's. The basis states are defined in terms of products of powers of the epd's, with certain combinations forbidden because of syzygies (polynomial identities). The basis states obtained are not orthonormal, but are complete, nonredundant, and analytic.

In Sec. III we discuss briefly the problem of computing generator matrix elements of Sp(6) between our states; it is in terms of them that the Hamiltonian operator for collective nuclear motions is defined.

Section IV shows how to convert the Sp(6) $\supset$ U(3) generating function to that for Su(n) $\supset$ O(n), all but the first three labels of SU(n) zero, and how to convert the Sp(6) $\supset$ Sp(2)$\times$O(3) to O($3n$) $\supset$ O(3)$\times$O(n), all but the first label of O($3n$) zero; the respective generating functions (and the branching rules) are related because of complementarity conditions.

We use Dynkin representation labels λ_i for the compact groups O(n), SU(n):

$$\lambda_i = 2\langle M_\lambda|\alpha_i\rangle/\langle\alpha_i|\alpha_i\rangle,$$

where M_λ is the highest weight of the IR (λ), and α_i are the simple roots; the exception is SO(3), where we use $\lambda/2 = l$ as the IR label. For (noncompact) Sp(6) we use the labels (p,q,d) of the "bottom" U(3) IR; (p,q) are its SU(3) labels and (d) its "vertical" weight component. The O(n) labels $(\omega_1,\omega_2,\omega_3)$ of Ref. 3 are related to the Dynkin labels used here by

$$\omega_1=\lambda_1+\lambda_2+\lambda_3,\quad \omega_2=\lambda_2+\lambda_3,\quad \omega_3=\lambda_3 \quad \text{for} \quad n\geqslant 9;$$

for O(8) we have $\omega_3=\lambda_3=\lambda_4$; and for O(7), $\omega_3=\lambda_3/2$.

II. Sp(6) $\supset$ Sp(2)$\times$O(3) BRANCHING RULES

We begin our derivation of Sp(6) $\supset$ Sp(2)$\times$O(3) branching rules with the known[4] generating function for Sp(6) $\supset$ U(3) branching rules

$$\begin{aligned} F(P,Q,D;A_1,B_1;A,B,Z) = {} & [(1-\alpha)(1-\beta)(1-\gamma)(1-\gamma^*)(1-\delta)(1-\delta^*)(1-\epsilon)(1-\epsilon^*)]^{-1} \\ & \times[(1+\eta+\theta+\kappa+\zeta\eta^*+\eta\theta+\eta\kappa+\eta\eta^*)(1-\zeta)^{-1} \\ & +(\zeta^*+\eta^*+\theta^*+\kappa^*+\zeta^*\eta+\eta^*\theta^*+\eta^*\kappa^*+\zeta^*\eta\eta^*)(1-\zeta^*)^{-1}]. \end{aligned} \tag{2.1}$$

[a)] Member of Instituto Nacional de Investigaciones Nucleares and El Colegio Nacional.

The letters on the right stand for the epd's:

$$\alpha = Z^3,\quad \beta = DZ,\quad \gamma = PA,\quad \delta = Q^2A_1{}^2Z,$$
$$\epsilon = A_1{}^2A^2Z,\quad \zeta = P^2A_1{}^2B^2Z,\quad \eta = PB_1{}^2BZ^2,$$
$$\theta = PA_1{}^2ABZ,\quad \kappa = PQA_1{}^2BZ.$$

The "conjugate" of an epd, denoted above by an asterisk, is obtained by the replacements $P \leftrightarrow Q$, $A_1{}^2Z \leftrightarrow B_1{}^2Z^2$, $A \leftrightarrow B$; the generating function (2.1) is conjugation symmetric, i.e., is unaffected by these interchanges. Equation (2.1) with A_1 and B_1 set equal to unity is just Eq. (3.6) of Ref. 4. Strictly, the labels a_1,b_1, carried as exponents by the dummies A_1,B_1, are not necessary; we comment below on their usefulness. When (2.1) is expanded in a power series

$$F = \sum P^pQ^qD^dA_1{}^{a_1}B_1{}^{b_1}A^aB^bZ^zC_{pqd,a_1b_1,abz}, \tag{2.2}$$

the coefficient C, summed over a_1,b_1, gives the multiplicity of the U(3) multiplet (a,b,z) in the Sp(6) IR (p,q,d).

Throughout this paper we follow the convention that representation labels, denoted by lowercase letters, are carried as exponents by the corresponding uppercase letters.

The exponents in (2.2) (or in the epd's) provide instructions for constructing the basis states (or the epd's): couple the U(3) multiplet $(a_1,b_1,a_1/2+b_1)$, whose components are polynomials of degree $a_1/2+b_1$ in the Sp(6) raising generators [they form the U(3) multiplet (2,0,1) and are the B^+_{ij} of Eq. (3.2a) or B^+_{lm} of Eq. (3.6) of Ref. 3], to the bottom U(3) multiplet (p,q,d) of the Sp(6) IR to obtain the U(3) multiplet $(a,b,a_1/2+b_1+d)$. The U(1) label is greater than $a_1/2+b_1+d$ by three times the degree in α, the SU(3) scalar of third degree in the raising generators.

The labels a_1,b_1 help in the interpretation of the epd's and of the basis states. For example, without them, one might, erroneously, think that ζ^* is the square of η^*. When (2.1) has been converted to give Sp(6) $\supset$ Sp(2)$\times$O(3) branching rules, the difficulties and ambiguities in interpreting it in terms of epd's are greatly increased. It is important to keep labels like a_1,b_1.

The subgroup SU(3) of Sp(6) is converted to SO(3) by substituting into Eq. (2.1) the SU(3) $\supset$ O(3) branching rules generating function

$$G(A,B,L) = [(1-A^2)(1-B^2)(1-AL)(1-BL)]^{-1}(1+ABL). \tag{2.3}$$

The substitution is accomplished[5] by evaluating

$$F(P,Q,D;A_1B_1;A',B',Z)G(A'^{-1}A,B'^{-1}B,L)\big|_{A'^0B'^0}.$$

The subscript $A'^0B'^0$ is an instruction to retain only the term in A' and B' of degree zero. The variables A,B are inserted to retain the SU(3) representation labels; as noted above, we will need all the labels we can get. The U(1) label z now becomes the weight label of the Sp(2) subgroup. The U(1) group is converted to (noncompact) Sp(2) simply by multiplying by 1-Z, or, more precisely, 1-M^*, where M^* is the epd $A_1{}^2A^2Z$ defined below. Then z, the exponent of Z, is the Sp(2) representation label, the lowest weight of the Sp(2) multiplet.

The result of the above operation is the desired Sp(6) $\supset$ Sp(2)$\times$O(3) generating function. It is given as follows:

$$\begin{aligned}
&H(P,Q,D;\ A_1,B_1;\ A,B;\ Z,L)\\
&= [(1-\alpha)(1-\beta)(1-J)(1-J^*)(1-K)(1-K^*)(1-M)]^{-1}\\
&\quad\times\{[(1-a)(1-a^*)(1-c)]^{-1}[a^*+e+r+a^*h+i^*+a^*k+l+a^*n+ci+ck^*\\
&\quad+u+ch^*+q+eq+ii^*+a^*s]+[(1-a)(1-a^*)(1-c^*)]^{-1}[c^*+c^*e+r^*+ah^*+i+ak^*\\
&\quad+l^*+an^*+c^*i^*+c^*k+u^*+c^*h+q^*+eq^*+c^*ii^*+c^*s]+[(1-a)(1-c)(1-d)]^{-1}[c+v+cp+ch\\
&\quad+cf+ck+cg+cn+cdi+cgh+du+cfh+cfg+cfn+cfi+cs]+[(1-a^*)(1-c^*)(1-d^*)]^{-1}[c^*d^*\\
&\quad+v^*+c^*p^*+c^*h^*+c^*f^*+c^*k^*+c^*g^*+c^*n^*+c^*d^*i^*+c^*g^*h^*+d^*u^*+c^*f^*h^*\\
&\quad+c^*f^*g^*+c^*f^*n^*+c^*f^*i^*+c^*d^*s]+[(1-a)(1-b^*)(1-d)]^{-1}[1+j+p+h+f+k+g+n+di\\
&\quad+gh+fp+fh+fg+fn+fi+s]+[(1-a^*)(1-b)(1-d^*)]^{-1}[d^*+d^*j^*+p^*+h^*+f^*+k^*\\
&\quad+g^*+n^*+d^*i^*+g^*h^*+f^*p^*+f^*h^*+f^*g^*+f^*n^*+d^*f^*i^*+d^*s]\\
&\quad+[(1-a^*)(1-b)(1-c)]^{-1}[a^*b+j^*+t+a^*bh+bi^*+a^*x+bl+a^*f^*h+cf^*+bck^*+w\\
&\quad+bch^*+hs+k^*w+f^*i^*+a^*bs]+[(1-a)(1-b^*)(1-c^*)]^{-1}[b^*c^*+c^*j+t^*+ab^*h^*+b^*i\\
&\quad+ax^*+b^*l^*+afh^*+c^*f+b^*c^*k+w^*+b^*c^*h+h^*s+kw^*+c^*fi+b^*c^*s]\\
&\quad+[(1-b)(1-c)(1-d)]^{-1}[bc+hm+cm+bch+bcf+cx+bcg+cf^*h+cdf^*\\
&\quad+bcgh+dw+bcfh+dhs+cms+cff^*+bcs]+[(1-b^*)(1-c^*)(1-d^*)]^{-1}[b^*c^*d^*\\
&\quad+h^*m^*+c^*m^*+b^*c^*h^*+b^*c^*f^*+c^*x^*+b^*c^*g^*+c^*fh^*+c^*d^*f+b^*c^*g^*h^*+d^*w^*\\
&\quad+b^*c^*f^*h^*+d^*h^*s+c^*m^*s+c^*ff^*+b^*c^*d^*s]+[(1-b)(1-b^*)(1-d)]^{-1}[b+y\\
&\quad+m+bh+bf+x+bg+f^*h+df^*+bfg+fm+bfh+bfg+ms+ff^*+bs]\\
&\quad+[(1-b)(1-b^*)(1-d^*)]^{-1}[b^*d^*+d^*y+m^*+b^*h^*+b^*f^*+x^*+b^*g^*+fh^*+d^*f+b^*g^*h^*\\
&\quad+f^*m^*+b^*f^*h^*+b^*f^*g^*+m^*s+d^*ff^*+b^*d^*s]\}.
\end{aligned} \tag{2.4}$$

The epd's α,β in Eq. (2.4) are the same as in (2.1); the others are as follows [the notation is (pq,a_1b_1,ab,l) which stands for $P^pQ^qA_1{}^{a_1}B_1{}^{b_1}A^aB^bZ^{(1/2)a_1+b_1}L^l$]:

$$
\begin{array}{lll}
J=(02,20,00,0), & K=(10,00,10,1), & M=(00,02,02,0),\\
a=(20,00,20,0), & b=(00,20,20,2), & c=(20,20,02,0),\\
d=(20,20,02,2), & e=(11,00,11,1), & f=(10,02,01,1),\\
g=(11,20,01,1), & h=(10,20,11,1), & i=(11,20,20,0),\\
j=(10,02,12,2), & k=(20,02,11,1), & l=(12,20,02,0),\\
m=(10,20,11,2), & n=(21,20,11,1), & p=(20,20,21,1),\\
q=(21,22,02,0), & r=(21,20,22,0), & s=(11,22,11,1),\\
t=(11,20,12,1), & u=(30,22,22,0), & v=(30,20,12,2),\\
w=(20,22,12,1), & x=(10,22,21,2), & y=(00,22,22,3).
\end{array}
$$

The conjugate of an epd, denoted by an asterisk in Eq. (2.4) obtained by interchanging the SU(3) labels in each of the three pairs: $(pq,a_1b_1,ab,)^* = (qp,b_1a_1,ba,)$; the epd's α,β,e,s,y are self-conjugate. The generating function in (2.4) apart from the missing denominator factor $1-M^*$, removed in converting from U(1) to Sp(2), is conjugation symmetric, a fact which was helpful in writing it in terms of the epd's. The generating function was also subjected to what may be called consistency checks. For example, the coefficient of i in Eq. (2.4)

$$
\begin{aligned}
&[(1-a)(1-a^*)(1-c^*)]^{-1} + c[(1-a)(1-a^*)(1-c)]^{-1}\\
&\quad + d\,[(1-a)(1-b^*)(1-d)]^{-1} + b^*[(1-a)\\
&\quad \times(1-b^*)(1-c^*)]^{-1}\\
&\quad + cd\,[(1-a)(1-c)(1-d)]^{-1}.
\end{aligned}
$$

It may be verified that each product of powers of three denominator epd's which appear in the same fraction (including those in which one or more exponents are zero) appears just once in the above expression; this check was made separately for each numerator epd and for each product of numerator epd's. As a final check we converted the expression in Eq. (2.4), by appropriate substitutions, into a generating function for SO(3) weights instead of SO(3) multiplets; it was then compared with the corresponding weight generating function obtained by converting (2.1) directly; since an analytic comparison would be prohibitively laborious, the necessary substitutions were made by a computer program and the two generating functions compared for random values of their arguments.

III. CONSTRUCTING THE BASIS STATES

The $\mathrm{Sp}(6)\supset \mathrm{Sp}(2)\times \mathrm{O}(3)$ generating function, as given in Eq. (2.4) defines the epd's (integrity basis) in terms of which all subgroup representations are given as stretched (all representation labels additive) products; epd's which are incompatible because of syzygies may be read from the generating function: they never appear multiplied together.

It is straightforward to construct the epd's, using their labels $(p,q;a_1,b_1;a,b;l)$: couple the SU(3) IR(a_1,b_1), of degree $a_1/2+b_1$ in the raising Sp(6) generators, to the bottom SU(3) multiplet (p,q) to obtain the IR (a,b); in every case the coupling is nondegenerate, i.e., unique. Next, choose the SO(3) multiplet contained in the SU(3) multiplet (a,b); again, the multiplet l is always nondegenerate. Apart from their usefulness in constructing the epd's, the labels a_1,b_1,a,b were invaluable in sorting out the epd's and their syzygies. Finally, to ensure that the states we are constructing are bottom states of Sp(2) multiplets, they must be rendered traceless (harmonic) by the use of Eq. (4.7b) of Ref. 3.

The bottom [lowest U(1)] multiplet of the Sp(6) IR (pqd) is best visualized in terms of the epd's for $\mathrm{Sp}(6n)\supset \mathrm{Sp}(6)\times \mathrm{O}(n)$.

The branching rules generating function

$$
\begin{aligned}
&[(1-PD^{1/2}H)(1-QDJ)(1-D^{3/2}K)]^{-1}\\
&\quad =\sum_{hjk} H^hJ^jK^kP^hQ^jD^{(1/2)(h+2j+3k)} \qquad (3.1)
\end{aligned}
$$

shows that the O(n) IR (hjk) is correlated with the Sp(6) IR (pqd) with $p=h$, $q=j$, $d=(h+2j+3k)/2$; integer values of d belong to even metaplectic Sp($6n$), $[(\frac{1}{2})^{3n}]$, half-odd values of d to odd metaplectic Sp($6n$), $[(\frac{1}{2})^{3n-1},\frac{3}{2}]$. Each of the three epd's stands for an elementary $\mathrm{Sp}(6)\times \mathrm{O}(n)$ multiplet, and is conveniently represented by the state of the multiplet which has the highest O(n) weight and, for Sp(6), the lowest U(1) and highest SU(3) [or SO(3)] weight. The exponent of D is one half the number of quanta in the state in question. Then $PD^{1/2}H$ is represented by η_{11}, where the first subscript denotes the highest state of the SU(3) [or SO(3)] triplet while the second one implies the highest state of the O(n) multiplet $(100\cdots 0)$. The epd QDJ, represented by

$$
\begin{vmatrix} \eta_{11} & \eta_{12} \\ \eta_{21} & \eta_{22} \end{vmatrix},
$$

is the highest state of an SU(3) antitriplet [or O(3) triplet] and the highest state of the O(n) multiplet $(010\cdots 0)$. The third epd $D^{3/2}K$ is represented by

$$
\begin{vmatrix} \eta_{11} & \eta_{12} & \eta_{13} \\ \eta_{21} & \eta_{22} & \eta_{23} \\ \eta_{31} & \eta_{32} & \eta_{33} \end{vmatrix};
$$

it is an SU(3) [or O(3)] scalar and is the highest state of the O(n) multiplet $(0010\cdots 0)$. Thus the bottom states of the Sp(6) IR(pqd) are defined by the product of powers of the three epd's with respective exponents p, q, $(2d\text{-}p\text{-}2q)/3$. These states are annihilated by the Sp(6) lowering (annihilation) generators, and no steps are needed to render them traceless.

The 21 generators of Sp(6) decompose under the subgroup $\mathrm{Sp}(2)\times \mathrm{O}(3)$ into three irreducible tensors which can be denoted by (1,0), (0,1), (1,2). The first two triplets are just the generators of Sp(2) and O(3); their matrix elements are well known. The matrix elements of only the (1,2) 15-plet need to be computed between our basis states. For that purpose it is necessary to compute only its reduced matrix elements between pairs of subgroup multiplets; although straightforward, that task is made laborious by the size of the integrity basis and the consequent large number of types of subgroup multiplet. We hope to complete it in a future publication.

IV. RELATED BRANCHING RULES

Complementarity relations in group–subgroup chains imply connections between apparently unrelated branching rules. Thus the generating functions of Eqs. (2.1) and (2.4), for Sp(6) $\supset$ U(3) and Sp(6) $\supset$ Sp(2)$\times$O(3), respectively, imply branching rules generating functions for SU(n) $\supset$ SO(n), all but the first three SU(n) labels zero, and for O($3n$) $\supset$ O(3)$\times$O(n), all but the first O($3n$) label zero, respectively.

Although not needed for the theory of nuclear collective motions, we present the results here since we get them at no extra cost.

For the chains of subgroups

$$\mathrm{Sp}(6n) \supset \mathrm{Sp}(6)\times \mathrm{O}(n), \tag{4.1a}$$

$$\mathrm{Sp}(6n) \supset \mathrm{Sp}(2)\times \mathrm{O}(3n), \tag{4.1b}$$

complementarity relations hold (see Ref. 3, Sec. II). This means that, since the representation of Sp($6n$) is $[(\frac{1}{2})^{3n}]$ or $[(\frac{1}{2})^{3n-1},(\frac{3}{2})]$ (the metaplectic ones), the IR of Sp(6) determines the IR of O(n) and vice versa in (4.1a) and the same holds for Sp(2) and O($3n$) in (4.1b).

Because of the complementarity in the chain in Eq. (4.1a) we can convert the generating function (2.1) giving the branching rules for Sp(6) $\supset$ U(3) to a generating function for the chain SU(n) $\supset$ O(n) [all but the first three labels of SU(n) zero] by the substitutions

$$P \to G^{1/3}H, \quad Q \to G^{2/3}J, \quad D \to G^{1/3}K, \quad A \to EG^{-1/3},$$

$$B \to FG^{-2/3}, \quad Z \to G^{2/3}, \quad A_1 \to 1, \quad B_1 \to 1;$$

the SU(n) nonzero labels are denoted by (e,f,g) and the O(n) ones by (h,j,k).

The above substitutions are valid when $n \geqslant 9$. For $n = 8$ the substitution for D changes to $D \to G^{1/3}KK'$, where the nonzero O(8) labels are (h,j,k,k') with $k = k'$, and for $n = 7$ the substitution for D is $D \to G^{1/3}K^2$ [here the three O(7) labels are (h,j,k)]. We do not consider the case $n \leqslant 6$.

Similarly, starting with the generating function for Sp(6) $\supset$ Sp(2)$\times$O(3) given in Eq. (2.4) we get the branching rules generating function for the chain O($3n$) $\supset$ O(3)$\times$O(n), all but the first label of O($3n$) zero, by the substitutions

$$P \to UH, \quad Q \to U^2J, \quad D \to UK, \quad Z \to U^2, \quad L \to L,$$

$$A_1 \to 1, \quad B_1 \to 1, \quad A \to 1, \quad B \to 1;$$

(u) labels O($3n$) IR's, (l) labels O(3) IR's, and (h,j,k) are the O(n) labels (all but the first three zero). These substitutions hold for $n \geqslant 9$. The substitutions for D become $D \to UKK'$ ($n = 8$) and $D \to UK^2$ ($n = 7$).

ACKNOWLEDGMENTS

We are grateful to E. Chacón and O. Castaños for helpful discussions.

This work was supported by the National Sciences and Engineering Research Council of Canada and by the Ministère de l'Education du Québec.

[1]M. Moshinsky, J. Math. Phys. 25, 1555 (1984).
[2]E. Chacón, P. Hess, and M. Moshinsky, J. Math. Phys. 25, 1565 (1984).
[3]O. Castaños, E. Chacón, and M. Moshinsky, J. Math. Phys. 25, 2815 (1984).
[4]R. Gaskell, G. Rosensteel, and R. T. Sharp, J. Math. Phys. 22, 2732 (1981).
[5]J. Patera and R. T. Sharp, J. Phys. A: Math. Gen. 13, 397 (1980).

Convergence of lattice sums and Madelung's constant

David Borwein
Department of Mathematics, University of Western Ontario, London, Ontario, N6A 5B9, Canada
Jonathan M. Borwein
Department of Mathematics, Statistics and Computing Science, Dalhousie University, Halifax, Nova Scotia, B3H 4H8, Canada
Keith F. Taylor
Department of Mathematics, University of Saskatchewan, Saskatoon, Saskatchewan, S7N 0W0, Canada

(Received 13 September 1984; accepted for publication 3 June 1985)

The lattice sums involved in the definition of Madelung's constant of an NaCl-type crystal lattice in two or three dimensions are investigated. The fundamental mathematical questions of convergence and uniqueness of the sum of these, not absolutely convergent, series are considered. It is shown that some of the simplest direct sum methods converge and some do not converge. In particular, the very common method of expressing Madelung's constant by a series obtained from expanding spheres does not converge. The concept of analytic continuation of a complex function to provide a basis for an unambiguous mathematical definition of Madelung's constant is introduced. By these means, the simple intuitive direct sum methods and the powerful integral transformation methods, which are based on theta function identities and the Mellin transform, are brought together. A brief analysis of a hexagonal lattice is also given.

I. INTRODUCTION

Lattice sums have played a role in physics for many years and have received a great deal of attention on both practical and abstract levels. The term lattice sum is not a precisely defined concept: it refers generally to the addition of the elements of an infinite set of real numbers, which are indexed by the points of some lattice in N-dimensional space. A method of performing a lattice sum involves accumulating the contributions of all these elements in some sequential order. Unfortunately, the elements of the set are not, in general, absolutely summable so the sequential order chosen can affect the answer. In this paper we are concerned with the particular lattice sums involved in Madelung's constant. Indeed, attaining specificity in the definition of Madelung's constant is our primary motive. Although we are dealing with purely mathematical questions it is our belief that the results presented here may shed some light on the physics of crystals. Other researchers[1–3] have expressed concern about the ambiguities involved in summing a nonabsolutely convergent series in a different manner, but it appears that no one has confronted it fully.

Let L be a lattice in N-dimensional space and let $A_L = \{a_l : l \in L\}$ be a set of real numbers indexed by L. There are two basic approaches to summing the elements of A_L: by direct summation or by integral transformations. The major factors involved in choosing a method are physical meaningfulness and speed of convergence.

The direct summation methods involve an orderly grouping of the elements of A_L into sequentially indexed finite subsets increasing in size to eventually include any element of A_L. Sometimes fractions of elements are included in the subsets to maintain a physical principal such as electrical neutrality. Two commonly used direct summation methods are due to Evjen[4] and Højendahl.[5]

The most commonly used integral transformation method is known as the Ewald method.[6] More recently the Mellin transformation applied to theta functions has been used to put the integral transformation methods in a general context. An excellent review article by Glasser and Zucker[1] gives a development of these methods and an extensive bibliography.

In this paper we deal primarily with NaCl-type ionic crystals in two or three dimensions. This is for two main reasons: the ease of notation and the fact that almost every textbook introduces Madelung's constant on this crystal first. From a mathematical and physical point of view there are two very reasonable simple direct summation methods that could be applied to give Madelung's constant for an NaCl-type ionic crystal. One could take a basic cube centered at the referenced ion with sides parallel to the basic vectors and let the cube expand as the contributions from all lattice points within the cube are accumulated. Alternately one could use expanding spheres centered at the reference ion. This latter method is intuitively appealing since all ions an equal distance from the reference ion are given equal treatment. Thus, many textbooks[7,8] (and some research articles) write down the resulting infinite series $(6 - 12/\sqrt{2} + 8/\sqrt{3} - 6/\sqrt{4} + \cdots)$ as giving Madelung's constant for an NaCl-type ionic crystal. Unfortunately, this infinite series does *not* converge. This was proven by Emersleben[9] and, in light of the fact that most people are unaware of this divergence, we include a short elementary proof in Theorem 3.

Section II is devoted to the two-dimensional square lattice while Sec. III contains the above-mentioned result on expanding spheres. In Theorem 4 we prove that the method of expanding cubes converges. In Sec. IV, the mathematical tools become more sophisticated as we consider integral transformation methods and their relation to the direct summation methods dealt with in Sec. III. We have included, in Sec. V, a careful analysis of some direct summation methods in two dimensions in light of the property of being analytic in the inverse power exponent. This analysis is quite illustrative of the relations between the various summation methods. In Sec. VI we do a brief analysis of a two-dimensional hexagonal lattice. Section VII gives our conclusions.

II. TWO DIMENSIONS

It is convenient to introduce the notation in the two-dimensional case of a simple lattice in the plane with unit charges located at integer lattice points (j,k) and of sign $(-1)^{j+k}$. The potential energy at the origin due to the charge at (j,k) is $-(-1)^{j+k}/(j^2+k^2)^{1/2}$. If we want the total potential energy at the origin due to all other charges, then we must sum all the numbers in the following set:

$$A = \{(-1)^{j+k}/(j^2+k^2)^{1/2} : (j,k) \in \mathbf{Z}/(0,0)\},$$

where $\mathbf{Z}$ denotes the integers. Because the elements of the subset of A with $j=k$ form a set of positive numbers with divergent sum, it is clear that the value of the sum is highly dependent on the order in which the elements of A are added. It is not immediately clear that any reasonable method will produce a convergent series. In addition, for the model to be physically relevant, all "reasonable" methods should converge to the same number. Here are two very reasonable methods.

First, consider the total potential due to all the charges within a circle of radius r about the origin and let $r \to \infty$. This leads to the series

$$\sum_{n=1}^{\infty} \frac{(-1)^n C_2(n)}{n^{1/2}}, \tag{1}$$

where $C_2(n)$ is the number of ways of writing n as a sum of two squares of integers (positive, negative, or zero). In deriving (1), use the fact that $(-1)^{j+k} = (-1)^{j^2+k^2} = (-1)^n$, for any $j,k \in \mathbf{Z}$ with $j^2+k^2=n$. We will refer to (1) as *the method of expanding circles.*

Second, there is *the method of expanding squares.* This is intuitively appealing, as a perfect crystal grows by expansion of the shape of a basic unit cell. For each natural number n, let

$$S_2(n) = \sum \left\{ \frac{(-1)^{j+k}}{(j^2+k^2)^{1/2}} : \ -n < j,k < n \ \text{and} \ (j,k) \neq (0,0) \right\}.$$

Then

$$\lim_{n\to\infty} S_2(n) \tag{2}$$

is a way of expressing the series obtained by expanding squares.

It turns out that both these methods converge as we will now show.

Theorem 1: The series in (1) converges.

Proof: To carry out the proof that the series in (1) converges we introduce some notation and standard facts from number theory. For any sequence of real numbers $\{a_n\}_{n=1}^{\infty}$ and β real we write $a_n = O(n^{\beta})$ if the sequence $\{n^{-\beta} a_n\}$ is bounded. Let $A_n = \sum_{k=1}^{n} C_2(k)$, for each natural number n. Then A_n denotes the number of nonorigin lattice points inside or on a circle of radius $n^{1/2}$. It is fairly easy to see that A_n should be approximately πn; in fact, the reader can easily show that $A_n - \pi n = O(n^{1/2})$. However, this is not quite good enough for us here so we quote a stronger result, which can be found in Dickson[10]:

$$A_n - \pi n = O(n^{\alpha}), \quad \text{for some } \alpha, \quad \tfrac{1}{4} < \alpha < \tfrac{1}{3}. \tag{3}$$

For a natural number n, a divisor of n is a natural number d such that d divides n. Let $d(n)$ denote the number of divisors of n and let $d_k(n)$ denote the number of divisors d of n with d congruent to k modulo 4, for $k=1$ or 3. With this notation, Theorem 278 of Hardy and Wright[11] implies

$$C_2(n) = 4(d_1(n) - d_3(n)). \tag{4}$$

This together with Theorem 315 of Hardy and Wright[11] implies that

$$C_2(n) = O(n^{\delta}), \quad \text{for any } \delta > 0. \tag{5}$$

Note that, $d_k(2n) = d_k(n)$, for $k=1$ or 3 and any n. So

$$C_2(2n) = C_2(n), \quad \text{for all natural numbers } n. \tag{6}$$

Let $B_n = \sum_{k=1}^{n} (-1)^k C_2(k)$. Then

$$\begin{aligned} B_{2n} &= \sum_{k=1}^{2n} (-1)^k C_2(k) \\ &= \sum_{k=1}^{n} C_2(2k) - \sum_{k=1}^{n} C_2(2k-1) \\ &= 2\sum_{k=1}^{n} C_2(2k) - \sum_{k=1}^{2n} C_2(k). \end{aligned}$$

Using (6),

$$B_{2n} = 2A_n - A_{2n}. \tag{7}$$

From (3) and (7), we have that, with α as in (3),

$$B_{2n} = O(n^{\alpha}).$$

Furthermore, this along with (5) implies that

$$B_{2n+1} = B_{2n} - C_2(2n+1) = O(n^{\alpha})$$

Therefore,

$$B_n = O(n^{\alpha}). \tag{8}$$

Now consider the partial sums of the series in (1):

$$\begin{aligned} T_n &= \sum_{k=1}^{n} \frac{(-1)^k C_2(k)}{k^{1/2}} = \frac{B_n}{(n+1)^{1/2}} \\ &\quad + \sum_{k=1}^{n} B_k [k^{-1/2} - (k+1)^{-1/2}] \\ &= O(n^{\alpha - 1/2}) - \sum_{k=1}^{n} B_k [(k+1)^{-1/2} - k^{-1/2}]. \end{aligned} \tag{9}$$

By the mean value theorem, $|(k+1)^{-1/2} - k^{-1/2}| < \tfrac{1}{2} k^{-3/2}$ and therefore

$$|B_k[(k+1)^{-1/2} - k^{-1/2}]| = O(k^{\alpha - 3/2}).$$

Since $\alpha - 3/2 < -1$, $\sum_{k=1}^{\infty} B_k[(k+1)^{-1/2} - k^{-1/2}]$ converges absolutely. Since $\alpha - \tfrac{1}{2} < 0$, $\lim_{n\to\infty} T_n = -\sum_{k=1}^{\infty} B_k[(k+1)^{-1/2} - k^{-1/2}]$ exists. That is, the series in (1) converges. Q.E.D.

We now turn to the limit in (2). We need an easy lemma from calculus that will be left to the reader to verify. This lemma will also be used in the proof of Theorem 4.

Lemma: For any positive real numbers, a,b, and s, each of the following functions are strictly decreasing in t, for $0<t<\infty$:

$$f_{1,s}(t)=t^{-s},$$
$$f_{2,s}(t)=t^{-s}-(t+a)^{-s},$$
$$f_{3,s}(t)=t^{-s}-(t+a)^{-s}-(t+b)^{-s}+(t+a+b)^{-s}.$$

Theorem 2: The limit in (2) exists.

Proof: We apply the lemma to $f_{2,s}$ with $a=(k+1)^2-k^2$ and $s=\frac{1}{2}$. Then, if $j\geqslant 0$ and $k\geqslant 0$ with $j+k\geqslant 1$, we have

$$f_{2,1/2}(j^2+k^2)>f_{2,1/2}((j+1)^2+k^2).$$

Explicitly this is

$$(j^2+k^2)^{-1/2}-(j^2+(k+1)^2)^{-1/2}-((j+1)^2+k^2)^{-1/2}+((j+1)^2+(k+1)^2)^{-1/2}>0. \tag{10}$$

Let $g(j,k)$ denote the left-hand side of (10). Then $(-1)^{j+k}\times g(j,k)$ is the contribution to the potential at the origin due to a basic cell of four adjacent ions with the closest ion at (j,k). Inequality (10) says that the contribution always has the same sign as that of the nearest ion.

Rewrite $S_2(n)$ using the symmetries to get

$$S_2(n)=4Q(n)+4X(n),$$

where

$$Q(n)=\sum_{j,k=1}^{n}\frac{(-1)^{j+k}}{(j^2+k^2)^{1/2}},$$
$$X(n)=\sum_{k=1}^{n}\frac{(-1)^k}{k}.$$

Since $\lim_{n\to\infty}X(n)=-\ln 2$, if we prove that $\lim_{n\to\infty}Q(n)$ exists, then the limit in (2) will exist. We will establish a number of properties of the sequence $\{Q(n)\}_{n=1}^{\infty}$, which will be used to prove its convergence.

Property 1:

$$Q(2n)-Q(2n-2)>0,\quad \text{for all } n\geqslant 2.$$

That is, the even indexed elements increase. To see this we group the terms of $Q(2n)-Q(2n-2)$ into basic cells of 4, as is illustrated in Fig. 1(a) for $Q(6)-Q(4)$. Thus,

$$Q(2n)-Q(2n-2)=\sum_{l=1}^{n}g(2l-1,2n-1)+\sum_{m=1}^{n-1}g(2n-1,2m-1),$$

where as before, $g(j,k)$ denotes the left-hand side of (10). So property 1 holds.

Property 2:

$$Q(2n+1)-Q(2n-1)<0,\quad \text{for all } n\geqslant 1.$$

That is, the odd indexed elements of the sequence decreases. Referring to Fig. 1(b) and correcting for the overlap at the $(2n,2n)$ point we are led to the following grouping:

$$Q(2n+1)-Q(2n-1)=\sum_{l=1}^{n}[-g(2l-1,2n)]+\sum_{l=1}^{n}[-g(2n,2l-1)]-[1/n\sqrt{2}-1/(n+1)\sqrt{2}]<0.$$

Property 3:

$$Q(2n+1)-Q(2n)>0,\quad \text{for all } n\geqslant 1.$$

Thus, the odd indexed elements are all greater than any even indexed element. This is clear again from a simple grouping of terms and using the monotonicity of $f_{1,1/2}$ of the lemma:

$$Q(2n+1)-Q(2n)=2\sum_{l=1}^{n}\left[\frac{1}{((2l-1)^2+(2n+1)^2)^{1/2}}-\frac{1}{(4l^2+(2n+1)^2)^{1/2}}\right]+\frac{1}{(2n+1)\sqrt{2}}>0.$$

Property 4:

$$\lim_{n\to\infty}Q(2n-1)-Q(2n)=0.$$

Thus, the difference between successive elements goes to zero. To see this, simply note that

$$0<Q(2n+1)-Q(2n)<2/[1+(2n+1)^2]^{1/2}+1/(2n+1)\sqrt{2}\to 0,\quad \text{as } n\to\infty.$$

It is now easy to see that properties 1–4 imply that $\lim_{n\to\infty}Q(n)$ exists. Q.E.D.

This completes the proof of Theorem 2. Thus, we have shown that two of the most obvious methods of summing for a Madelung constant in two dimensions converge. At this point, no indication has been given that the two methods yield the same number. That this is indeed so will be shown in Sec. V.

III. THREE DIMENSIONS

In this section, the three-dimensional case will be considered. For Madelung's constant of an NaCl-type crystal one must investigate ways of summing the elements of the following set:

$$B=\{(-1)^{j+k+l}/(j^2+k^2+l^2)^{1/2}:(j,k,l)\in\mathbb{Z}^3/(0,0,0)\}.$$

In analogy with the two-dimensional case we will consider the method of expanding spheres about the origin and the method of expanding cubes.

Our next theorem is a negative result, which is quite startling. Many textbooks in physical chemistry and solid state physics give the series dealt with in Theorem 3 as Madelung's constant for a NaCl-type crystal.[7,8] It also appears

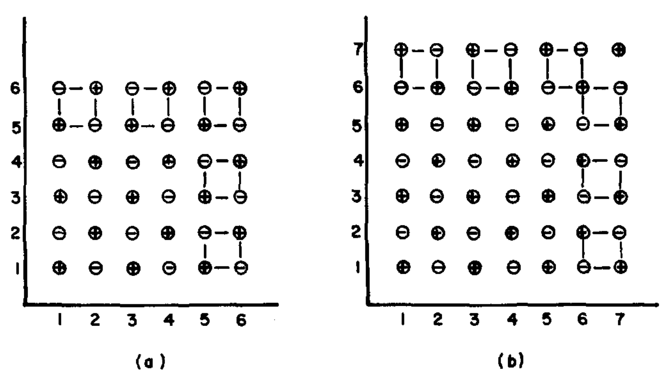


FIG. 1. Illustrations of (a) property 1 and (b) property 2.

in research articles. Although no one sums this series directly, it is physically misleading to believe that it converges to anything.

Let $C_3(n)$ denote the number of ways of writing n as a sum of three squares. If we consider a sphere centered at the origin in three-space, add all the elements of B that correspond to lattice points within the sphere, and then let the radius go to infinity, we are led to the infinite series

$$\sum_{n=1}^{\infty} \frac{(-1)^n C_3(n)}{\sqrt{n}}. \tag{11}$$

Theorem 3 (Emersleben[9]): The series in (11) diverges.

Proof: It is interesting that the proof that the series in (11) diverges is much less sophisticated than the proof in Theorem 1 that the series in (1) converges. Our main tool is a simple estimate of the number of nonzero lattice points on or inside a sphere of radius r. Call this number L_r. Notice that, for $\sqrt{n} < r < \sqrt{n+1}$,

$$L_r = \sum_{k=1}^{n} C_3(k).$$

We leave to the reader the easy task of verifying that

$$L_r - \tfrac{4}{3}\pi r^3 = O(r^2).$$

This implies that

$$\lim_{r\to\infty} L_r / r^3 = \tfrac{4}{3}\pi. \tag{12}$$

Proceeding with a proof by contradiction we assume that

$$\sum_{n=1}^{\infty} \frac{(-1)^n C_3(n)}{\sqrt{n}} \quad \text{converges.}$$

This implies that $\epsilon_n = C_3(n)/\sqrt{n} \to 0$, as $n\to\infty$. For a natural number N, let $M_N = \max\{\epsilon_n : n \geqslant N\}$. Then $M_N \to 0$, as $N\to\infty$. Fix N for the moment and consider, for $n > N$,

$$\begin{aligned}\frac{L_{\sqrt{n}}}{(\sqrt{n})^3} &= n^{-3/2}\left[\sum_{k=1}^{n} \epsilon_k \sqrt{k}\right] \\ &< n^{-3/2}\left[\sum_{k=1}^{N} \epsilon_k \sqrt{k}\right] + M_N n^{-3/2}\left[\sum_{k=N+1}^{n} \sqrt{k}\right].\end{aligned} \tag{13}$$

Now,

$$\begin{aligned}\sum_{k=N+1}^{n} \sqrt{k} &< \int_{N+1}^{n+1} t^{1/2}\,dt \\ &= \tfrac{2}{3}[(n+1)^{3/2} - (N+1)^{3/2}].\end{aligned}$$

Inserting this in (13) implies that

$$\begin{aligned}\frac{L_{\sqrt{n}}}{(\sqrt{n})^3} &< n^{-3/2}\left[\sum_{k=1}^{N} \epsilon_k \sqrt{k}\right] \\ &\quad + \frac{2}{3} M_n \left[\left(\frac{n+1}{n}\right)^{3/2} - \left(\frac{N+1}{n}\right)^{3/2}\right].\end{aligned}$$

Letting $n\to\infty$, we see that $\limsup_{n\to\infty} L_{\sqrt{n}}/(\sqrt{n})^3 \leqslant \tfrac{2}{3} M_N$, for any N. Since $M_N \to 0$ as $N\to\infty$, we have that

$$\lim_{n\to\infty} \frac{L_{\sqrt{n}}}{(\sqrt{n})^3} = 0.$$

This is a contradiction of (12). Therefore,

$$\sum_{n=1}^{\infty} \frac{(-1)^n C_3(n)}{\sqrt{n}} \quad \text{diverges.}$$

In fact, the contributions of individual spherical shells do not tend to zero. Q.E.D.

So it is not at all appropriate to define Madelung's constant via the method of expanding spheres. We turn to the method of expanding cubes. Let

$$S_3(n) = \sum\left\{\frac{(-1)^{j+k+l}}{(j^2+k^2+l^2)^{1/2}} : -n < j,k,l < n,\ (j,k,l) \neq (0,0,0)\right\}.$$

Theorem 4: $\lim_{n\to\infty} S_3(n)$ exists.

Proof: We proceed as in the proof of Theorem 2 in Sec. II. For $j,k,l \geqslant 1$ let

$$\begin{aligned}g(j,k,l) &= (j^2+k^2+l^2)^{-1/2} - (j^2+(k+1)^2+l^2)^{-1/2} \\ &\quad - (j^2+k^2+(l+1)^2)^{-1/2} \\ &\quad + (j^2+(k+1)^2+(l+1)^2)^{-1/2} \\ &\quad - ((j+1)^2+k^2+l^2)^{-1/2} \\ &\quad + ((j+1)^2+(k+1)^2+l^2)^{-1/2} \\ &\quad + ((j+1)^2+k^2+(l+1)^2)^{-1/2} \\ &\quad - ((j+1)^2+(k+1)^2 \\ &\quad + (l+1)^2)^{-1/2}.\end{aligned}$$

Then, $(-1)^{j+k+l} g(j,k,l)$ represents the contribution to the potential at the origin of a basic unit cell whose closest corner is at (j,k,l). An appropriate use of the monotonicity of $f_{3,1/2}$ from the lemma in Sec. II shows that $g(j,k,l) > 0$, for all $j,k,l \geqslant 1$.

Let

$$\begin{aligned}h(k,l) &= (2k^2+l^2)^{-1/2} - (2k^2+(l+1)^2)^{-1/2} \\ &\quad + (2(k+1)^2+l^2)^{-1/2} \\ &\quad - (2(k+1)^2+(l+1)^2)^{-1/2}.\end{aligned}$$

Using $f_{2,1/2}$, we get that $h(k,l) > 0$, for all $k,l \geqslant 1$.

Let $P(n)$ denote that part of $S_3(n)$ that comes from the positive octant. That is, for $n \geqslant 1$,

$$P(n) = \sum_{j,k,l=1}^{n} \frac{(-1)^{j+k+l}}{(j^2+k^2+l^2)^{1/2}}.$$

Then in a manner similar that used for the $Q(n)$'s, it can be shown that $\lim_{n\to\infty} P(n)$ exists. We proceed with the details of this demonstration.

The following identities are most easily seen by drawing a three-dimensional version of Fig. 1, but they can be verified directly:

$$\begin{aligned}P(2n+1) - P(2n-1) &= 3\sum_{k,l=1}^{n} g(2n,2k-1,2l-1) \\ &\quad + 3\sum_{j=1}^{n} h(2n,2j-1) \\ &\quad + (1/\sqrt{3})(1/2n - 1/(2n+1)),\end{aligned} \tag{14}$$

$$
\begin{aligned}
P(2n+2)-P(2n) = &-3\sum_{k,l=1}^{n} g(2n+1,2k-1,2l-1)\\
&-3\sum_{j=1}^{n} g(2n+1,2n+1,2j-1)\\
&-g(2n+1,2n+1,2n+1). \qquad (15)
\end{aligned}
$$

Both (14) and (15) hold for all $n \geqslant 1$. From (14) and (15) we get the properties of odd or even element monotonicity of the sequence of $P(n)$'s.

Property 1':

$P(2n)-P(2n-2)<0, \quad$ for all $n \geqslant 2$.

Property 2':

$P(2n+1)-P(2n-1)>0, \quad$ for all $n \geqslant 1$.

Notice that the inequalities are reversed from those of properties 1 and 2 in the two-dimensional case. To get the analogs of properties 3 and 4 for the $P(n)$'s we need to refer to the lemma one final time. For $n,j,k \geqslant 1$, let

$$
\begin{aligned}
h_0(n,j,k) = &(n^2+j^2+k^2)^{-1/2}-(n^2+(j+1)^2+k^2)^{-1/2}\\
&-(n^2+j^2+(k+1)^2)^{-1/2}\\
&+(n^2+(j+1)^2+(k+1)^2)^{-1/2}.
\end{aligned}
$$

With $a=(k+1)^2-k^2$,

$h_0(n,j,k)=f_{2,1/2}(n^2+j^2+k^2)-f_{2,1/2}(n^2+(j+1)^2+k^2)$,

which is positive for all $n,j,k \geqslant 1$.

With this notation

$$
\begin{aligned}
P(2n+1)-P(2n) = &-3\sum_{j,k=1}^{n} h_0(2n+1,2j-1,2k-1)\\
&-3\sum_{l=1}^{n}\left[(2(2n+1)^2+(2l-1)^2)^{-1/2}\right.\\
&\left.-(2(2n+1)^2+(2l)^2)^{-1/2}\right],\\
&-1/((2n+1)\sqrt{3}). \qquad (16)
\end{aligned}
$$

This leads to the following property.

Property 3':

$P(2n+1)-P(2n)<0, \quad$ for all $n \geqslant 1$.

Therefore, the decreasing even indexed elements are all greater than the increasing odd indexed elements. To see that there is a unique limit to the sequence of the $P(n)$'s, we only need the last property, which implies that the distance between successive terms approaches zero. This follows from (16) and

$$
\begin{aligned}
P(2n+1)-P(2n) > &-3/((2n+1)^2+2)^{1/2}\\
&-1/((2n+1)\sqrt{3}). \qquad (17)
\end{aligned}
$$

To verify (17) let

$$
x_j=\sum_{k=1}^{2n}\frac{(-1)^{1+j+k}}{((2n+1)^2+j^2+k^2)^{1/2}}, \quad \text{for } 1 \leqslant j \leqslant 2n+1.
$$

Using the function h_0, defined above, write

$$
|x_j|-|x_{j+1}|=\sum_{k=1}^{n} h_0(2n+1,j,2k-1)>0. \qquad (18)
$$

Note that x_j itself is an alternating sum of decreasing terms, so the sign of x_j is $(-1)^j$. With (18), this implies that

$$
0>\sum_{j=1}^{2n+1} x_j > x_1 > \frac{-1}{((2n+1)^2+2)^{-1/2}}.
$$

Then,

$$
\begin{aligned}
P(2n+1)-P(2n) &= 3\sum_{j=1}^{2n+1} x_j - \frac{1}{(2n+1)\sqrt{3}}\\
&> -\frac{3}{((2n+1)^2+2)^{1/2}}\\
&\quad -\frac{1}{((2n+1)\sqrt{3})}.
\end{aligned}
$$

So (17) holds. Thus, the following property has been established.

Property 4':

$\lim_{n\to\infty} P(2n+1)-P(2n)=0.$

Properties 1'–4' imply that $\lim_{n\to\infty} P(n)$ exists.

Finally,

$$
S_3(n)=8P(n)+8Q(n)+6X(n), \qquad (19)
$$

where, as before,

$$
Q(n)=\sum_{j,k=1}^{n}\frac{(-1)^{j+k}}{(j^2+k^2)^{1/2}}
$$

and

$$
X(n)=\sum_{k=1}^{n}\frac{(-1)^k}{k}.
$$

Since each of the terms on the right-hand side of (19) approach a limit as $n\to\infty$, we have that $\lim_{n\to\infty} S_3(n)$ exists. Q.E.D.

Remark 1: Although this method of summing over expanding cubes is not rapidly convergent, it is extremely well behaved. The alternation of the $P(n)$ and $Q(n)$ above and below their limiting values provide precise error bounds, which may be useful in theoretical considerations.

Remark 2: The work of Campbell[3] must be mentioned at this point. He states general conditions on a doubly indexed series and concludes a convergence result, which is stronger than Theorem 2 above. However, there is a serious error in his proof and his general theorem is false. A simplified version of Campbell's claimed result would be the following: Let $\{a_{ij}\}_{i=1,j=1}^{\infty,\infty}$ be a doubly indexed "sequence" of reals satisfying (A) for all i, $\{|a_{i1}|, |a_{i2}|, |a_{i3}|,...\}$ is a monotonically decreasing sequence with $\lim_{j\to\infty} a_{ij}=0$, and for all j, $\{|a_{1j}|, |a_{2j}|,...\}$ is a montonically decreasing sequence with $\lim_{i\to\infty} a_{ij}=0$; and (B) the sign of a_{ij} is $(-1)^{i+j+1}$. Then $\Sigma_{i=1}^{\infty}(\Sigma_{j=1}^{\infty} a_{ij})$ exists.

Here is a counterexample to this claim. Let the a_{ij}'s be defined as in the array in Table I. Let

TABLE I. The *ij* th entry in the array is denoted a_{ij}, to form a counterexample to the general convergence result claimed by Campbell.

$i \backslash j$	1	2	3	4	5	
1	1	$-\frac{1}{2}$	$\frac{1}{3}$	$-\frac{1}{4}$	$\frac{1}{5}$	...
2	$-\frac{1}{2}-10^{-2}$	$\frac{1}{2}-10^{-3}$	$-\frac{1}{4}-10^{-4}$	$\frac{1}{4}-10^{-5}$	$-\frac{1}{6}-10^{-6}$	...
3	$\frac{1}{2}$	$-\frac{1}{3}$	$\frac{1}{4}$	$-\frac{1}{5}$	$\frac{1}{6}$	...
4	$-\frac{1}{3}-10^{-3}$	$\frac{1}{3}-10^{-4}$	$-\frac{1}{5}-10^{-5}$	$\frac{1}{5}-10^{-6}$	$-\frac{1}{7}-10^{-7}$	...
5	$\frac{1}{3}$	$-\frac{1}{4}$	$\frac{1}{5}$	$-\frac{1}{6}$	$\frac{1}{7}$	...
6	$-\frac{1}{4}-10^{-4}$	$\frac{1}{4}-10^{-5}$	$-\frac{1}{6}-10^{-6}$	$\frac{1}{6}-10^{-7}$	$-\frac{1}{8}-10^{-8}$	...
⋮	⋮	⋮	⋮	⋮	⋮	

$$U_i = \sum_{j=1}^{\infty} a_{ij}, \quad \text{for } i = 1,2,3,\ldots .$$

Clearly, each U_i exists and the sign of U_i is $(-1)^{i+1}$. The odd indexed U_i are all positive and easily calculated:

$$U_1 = \ln 2 = 1/1\times 2 + 1/3\times 4 + 1/5\times 6 + \cdots,$$

$$U_3 = \ln 2 + 1 = 1/2\times 3 + 1/4\times 5 + 1/6\times 7 + \cdots .$$

In general,

$$U_{2k-1} = (-1)^{k+1}\left[\ln 2 - (1 - \tfrac{1}{2} + \cdots + (-1)^k/(k-1))\right]$$

$$= \sum_{j=0}^{\infty} \frac{1}{(k+2j)(k+2j+1)}, \quad k \geqslant 2.$$

Any U_{2k} is negative and given by

$$U_{2k} = \sum_{j=k+1}^{\infty} -\frac{1}{10^j} = -\frac{10^{-k}}{9}.$$

Note, that

$$\sum_{k=1}^{\infty} U_{2k} = -\frac{1}{9}\left(\sum_{k=1}^{\infty} 10^{-k}\right) = -\frac{1}{81}.$$

On the other hand

$$\sum_{k=1}^{\infty} U_{2k-1} = \sum_{k=1}^{\infty}\left(\frac{1}{k(k+1)} + \frac{1}{(k+2)(k+3)} + \cdots\right)$$

$$= \sum_{n=1}^{\infty}\left(\frac{1}{n(n+1)} + \frac{1}{(n+1)(n+2)} + \cdots\right)$$

$$> \sum_{n=1}^{\infty}\sum_{j=n}^{\infty}\frac{1}{2j^2} > \frac{1}{2}\sum_{n=1}^{\infty}\int_n^{\infty}\frac{1}{t^2}\,dt$$

$$= \frac{1}{2}\sum_{n=1}^{\infty}\frac{1}{n} = \infty .$$

Since the sum of the positive terms diverge and the sum of the negative terms converge, it follows that $\lim_{n\to\infty}\sum_{i=1}^{n} U_i$ does not exist.

Thus,

$$\sum_{i=1}^{\infty}\left(\sum_{j=1}^{\infty} a_{ij}\right) \text{ diverges,}$$

even though it satisfies Campbell's conditions for convergence. Campbell goes on to claim that the analogous result holds for any dimension and that one could also prove convergence if one summed by expanding rectangles. Both these statements are unfounded.

Remark 3: In light of the above, it appears that there is no simple proof in the literature of the convergence of any of the most elementary direct summation methods. That is why the detailed proofs of Theorems 2 and 4 are given. Emersleben's result (Theorem 3) indicates that a noncasual approach is justified.

Remark 4: The proofs given to Theorems 2 and 4 are simple and intuitive, based as they are on the fact that the contribution of a basic unit cell to the sum is always of the sign of the nearest point in the cell to the origin. We have abstracted this property and have obtained quite general convergence results for multidimensional alternating series. These results will be published elsewhere. We will point out later that Theorems 2 and 4 also follow from the deeper considerations of the next section.

IV. INTEGRAL TRANSFORMATIONS AND ANALYTICITY

Our purpose in this section is to establish a firm connection between the elementary direct summation methods discussed above and the integral transformation methods, which are described by Glasser and Zucker in their survey article.[1] One major consequence of this connection is that we can give a definition of Madelung's constant, which has a firm mathematical foundation and is unique in a strong enough sense to indicate why diverse methods of performing the lattice sums lead to the same number. We begin with a general discussion of analyticity of certain lattice sums in N-dimensional space. Of course, $N=2$ and 3 are the most interesting cases, but the general notation is just as convenient.

For a complex number s, let Re s denote the real part of s and let

$$A^N(s) = \{(-1)^{\mathbf{n}}/\|\mathbf{n}\|^{2s}: \mathbf{n}\in\mathbb{Z}^N/\{\mathbf{0}\}\},$$

where for $\mathbf{n} = (n_1,n_2,\ldots,n_N)\in\mathbb{Z}^N$, $(-1)^{\mathbf{n}} = (-1)^{n_1+\cdots+n_N}$ and $\|\mathbf{n}\| = (n_1^2 + n_2^2 + \cdots + n_N^2)^{1/2}$. We also use the notations

$$|\mathbf{n}| = (|n_1|,|n_2|,\ldots,|n_N|)\in\mathbb{Z}^N,$$

and for $\mathbf{m}\in\mathbb{Z}^N$,

$$\mathbf{n}\geqslant\mathbf{m}, \quad \text{if } n_j \geqslant m_j, \quad \text{for } 1\leqslant j\leqslant N.$$

If $\operatorname{Re} s > N/2$, then a simple comparison test shows that $\sum_{\mathbf{n}\neq 0} 1/\|\mathbf{n}\|^{2s} < \infty$. So the elements of $A^N(s)$ are absolutely summable if $\operatorname{Re} s > N/2$. Let

$$d_N(2s) = \sum\left\{\frac{(-1)^{\mathbf{n}}}{\|\mathbf{n}\|^{2s}}: \mathbf{n}\in\mathbb{Z}^N/\{\mathbf{0}\}\right\}.$$

Then $d_N(z)$ is a function of the complex variable z for $\operatorname{Re} z > N$. In fact, it is a multidimensional zeta function, analytic on this domain. To see this, define for $\mathbf{m}\in\mathbb{Z}^n$, with $\mathbf{m} > \mathbf{0}$,

$$d^{\mathbf{m}}(z) = \sum\left\{\frac{(-1)^{\mathbf{n}}}{\|n\|^{z}}: \mathbf{n}\in\mathbb{Z}^N/\{0\} \text{ and } |\mathbf{n}|<\mathbf{m}\right\}. \quad (20)$$

The $d^{\mathbf{m}}(z)$ is analytic for $\operatorname{Re} z > 0$. For fixed $\delta > 0$, if $\operatorname{Re} z > N+\delta$,

$$|d_N(z) - d^{\mathbf{m}}(z)| \leqslant \sum\left\{\frac{1}{\|\mathbf{n}\|^{N+\delta}}: \mathbf{n}\in\mathbb{Z}^N\setminus\{\mathbf{l}\in\mathbb{Z}^N: |\mathbf{l}|<\mathbf{m}\}\right\}. \quad (21)$$

The right-hand side of (21) can be made arbitrarily small by letting the minimal coefficient of $\mathbf{m}$ get large. Thus, on the region ($\operatorname{Re} z > N+\delta$), $d_N(z)$ is the uniform limit of analytic functions and is therefore analytic. Since $\delta > 0$ was arbitrary we have established the following proposition.

Proposition 1: The function $d_N(2s)$ is analytic in s for $\operatorname{Re} s > N/2$.

Now comes the crucial step for the definition of Madelung's constant. The functions $d_N(2s)$ can be analytically continued to the region ($\operatorname{Re} s > 0$). To accomplish this we follow the ideas of Glasser and Zucker[1] and introduce θ functions and the Mellin transform. We need, in particular,

$$\theta_4(q) = \sum_{n=-\infty}^{\infty}(-1)^n q^{n^2}, \quad 0 < q < 1.$$

So

$$\theta_4(e^{-t}) = \sum_{n=-\infty}^{\infty} (-1)^n e^{-n^2 t}, \quad 0<t<\infty. \tag{22}$$

For a continuous function $f(t)$ defined for $0<t<\infty$, bounded as $t \to 0$ and decaying sufficiently fast as $t \to \infty$, one can define a normalized Mellin transform $M_s(f)$ for $\operatorname{Re} s > 0$ by

$$M_s(f) = \Gamma^{-1}(s) \int_0^\infty f(t) t^{s-1}\, dt,$$

where Γ is the usual gamma function given by

$$\Gamma(s) = \int_0^\infty e^{-t} t^{s-1}\, dt, \quad \text{for } \operatorname{Re} s > 0.$$

Of course, Γ and Γ^{-1} are analytic functions on $(\operatorname{Re} s > 0)$. A useful property of the Mellin transform is that for $a>0$ and f such that its Mellin transform exists, $M_s(\tau_a f) = M_s(f)/a^s$, where $\tau_a f(t) = f(at)$ for all $t>0$. In particular,

$$M_s(e^{-at}) = 1/a^s, \quad \text{for } a>0. \tag{23}$$

Consider now a truncation of the series for θ_4. For some positive integer m, let $\phi_m(q) = \Sigma_{n=-m}^{m} (-1)^n q^{n^2}$. If $\mathbf{m} \in \mathrm{N}^N$, say $\mathbf{m} = (m_1,\ldots,m_N)$, then let $m = \min\{m_1,\ldots m_N\}$. We wish to approximate the Nth power of θ_4 with products of ϕ_{m_i}. For $0<t<\infty$,

$$\begin{aligned}
&\left|\theta_4^N(e^{-t}) - \prod_{i=1}^N \phi_{m_i}(e^{-t})\right| \\
&\leqslant \left|\prod_{i=1}^N [\,|\theta_4(e^{-t}) - \phi_{m_i}(e^{-t})| + \phi_{m_i}(e^{-t})] - \prod_{i=1}^N \phi_{m_i}(e^{-t})\right| \\
&= \prod_{i=1}^N (b_i + a_i) - \prod_{i=1}^N a_i,
\end{aligned} \tag{24}$$

where $a_i = \phi_{m_i}(e^{-t})$ and $b_i = |\theta_4(e^{-t}) - \phi_{m_i}(e^{-t})|$. Note that $0<a_i,\ b_i<1$, for $i=1,\ldots,N$ and the last expression in (24), $\Pi_{i=1}^N (b_i + a_i) - \Pi_{i=1}^N a_i$ represents the difference in volume between an N box of side lengths $b_i + a_i$, $i=1,\ldots,N$ and one of side lengths $a_i, i=1,\ldots,N$. Clearly $\Pi_{i=1}^N (b_i + a_i) - \Pi_{i=1}^N a_i < \max\{b_i : 1 \leqslant i \leqslant N\} N 2^{N-1}$. Now $|\theta_4(e^{-t}) - \phi_m(e^{-t})|$ is the maximum b_i and $|\theta_4(e^{-t}) - \phi_m(e^{t})| < 2e^{-m^2 t}$. So (24) becomes

$$\left|\theta_4^N(e^{-t}) - \prod_{i=1}^N \phi_{m_i}(e^{-t})\right| < N 2^N e^{-m^2 t}. \tag{25}$$

We also need the Mellin transform of $\Pi_{i=1}^N \phi_{m_i}(e^{-t}) - 1$, which is easily found using linearity and (23):

$$\begin{aligned}
M_s\left[\prod_{i=1}^N \phi_{m_i}(e^{-t}) - 1\right] &= M_s\left[\sum\nolimits'_{|\mathbf{n}| \leqslant \mathbf{m}} (-1)^{\mathbf{n}} e^{-\|\mathbf{n}\|^2 t}\right] \\
&= \sum\nolimits'_{|\mathbf{n}| \leqslant \mathbf{m}} \frac{(-1)^{\mathbf{n}}}{\|\mathbf{n}\|^{2s}} \\
&= d^{\mathbf{m}}(2s).
\end{aligned} \tag{26}$$

The prime on the summation sign indicates that the $\mathbf{n} = 0$ term is omitted.

We are now ready for the main theorems of this section. Define $F(s) = M_s\ [\theta_4^N(e^{-t}) - 1]$ wherever it exists.

Theorem 5: The Mellin transform $F(s)$ of $\theta_4^N(e^{-t}) - 1$ exists and is analytic for all s with $\operatorname{Re} s > 0$. Furthermore F provides an analytic continuation of $d_N(2s)$ to the region $(\operatorname{Re} s > 0)$.

Theorem 6: For any $\mathbf{m} \in \mathbf{Z}^N$, $\mathbf{m} \gtrsim \mathbf{0}$, $m = \min\{m_i : 1 \leqslant i \leqslant N\}$ and $\operatorname{Re} s > 0$

$$|F(s) - d^{\mathbf{m}}(2s)| < N 2^N \Gamma(\operatorname{Re} s)/(m^{2 \operatorname{Re} s} |\Gamma(s)|). \tag{27}$$

Proof (of Theorems 5 and 6 combined): Let s be a complex number such that $\operatorname{Re} s > 0$. Since

$$0 \leqslant 1 - \theta_4^N(e^{-t}) \leqslant N[1 - \theta_4(e^{-t})] \leqslant N e^{-t}, \quad \text{for all } 0<t<\infty,$$

then

$$\int_0^\infty |\theta_4^N(e^{-t}) - 1|\, |t^{s-1}| dt \leqslant N \int_0^\infty e^{-t} t^{\operatorname{Re} s - 1}\, dt = N\Gamma(\operatorname{Re} s).$$

Therefore, if $\operatorname{Re} s > 0$, then

$$\int_0^\infty [\theta_4^N(e^{-t}) - 1] t^{s-1}\, dt = F(s)$$

exists. Using (25) and (26), with $\mathbf{m}$ as in Theorem 6,

$$\begin{aligned}
|\Gamma(s)|\, |F(s) - d^{\mathbf{m}}(2s)| &\leqslant \int_0^\infty \left|\theta_4^N(e^{-t}) - \prod_{i=1}^N \phi_{m_i}(e^{-t})\right| |t^{s-1}| dt \\
&\leqslant N 2^N \int_0^\infty e^{-m^2 t} t^{\operatorname{Re} s - 1}\, dt \\
&= N 2^N \Gamma(\operatorname{Re} s)/m^{2 \operatorname{Re} s}.
\end{aligned}$$

Thus (27) holds. In turn, (27) implies that $F(s)$ can be uniformly approximated by the $d^{\mathbf{m}}(2s)$ on any region of the form $R_{\delta,M} = \{s : |s| < M \text{ and } \operatorname{Re} s > \delta\}$. To see this let K be an upper bound for the continuous function $N 2^N (\operatorname{Re} s)/|\Gamma(s)|$ on the closure of $R_{\delta,M}$. Then, for any $\epsilon > 0$ and any $\mathbf{m}$ such that $m = \min\{m_1,\ldots,m_N\} > (K/\epsilon)^{1/2\delta}$,

$$|F(s) - d^{\mathbf{m}}(2s)| < \epsilon, \quad \text{for all } s \in R_{\delta,M}.$$

Since ϵ is arbitrary and $\mathrm{d}^{\mathbf{m}}$ is analytic, F is analytic on $R_{\delta,M}$ for any $\delta > 0$ and $0 < M < \infty$. Therefore F is analytic on $(\operatorname{Re} s > 0)$. Finally, it is now clear that $F(s)$ agrees with $d_N(2s)$ if $\operatorname{Re} s > N/2$. Thus F is an analytic continuation of d_N. Q.E.D.

In light of Theorem 5, we will drop the use of F and write

$$d_N(2s) = M_s[\theta_4^N(e^{-t}) - 1], \quad \text{for } \operatorname{Re} s > 0. \tag{28}$$

A rigorous mathematical definition can now be given for Madelung's constant.

Definition: For a three-dimensional NaCl-type ionic crystal, Madelung's constant is the number

$$d_3(1) = M_{1/2}[\theta_4^3(e^{-t}) - 1].$$

Of course, this is the very number that has been approximated by many different methods over the years. We have just given a definition that avoids all the ambiguities of meaning that have existed. The uniqueness of analytic continuation explains the special significance of this particular sum of elements of

$$\{(-1)^{i+j+k}/(i^2+j^2+k^2)^{1/2}:\ (i,j,k)\in\mathbb{Z}^3/(0,0,0)\}.$$

Formula (27) emphasizes the strong connection between the integral transformation methods and the direct summation methods. In fact it is worthwhile to formulate a corollary to Theorem 6, which gives explicit error bounds for a finite sum approximation to Madelung's constant.

Corollary 1: Let $m_i > 0$, for $i = 1,2,3$ and $m = \min\{m_1,m_2,m_3\}$. Then

$$\left| d_3(1) - \sum_{(|i|,|j|,|k|)<m}{}' \frac{(-1)^{i+j+k}}{(i^2+j^2+k^2)^{1/2}} \right| < \frac{12}{m}.$$

Remark 5: The above corollary says the Madelung's constant for NaCl can be obtained, not only by expanding cubes, but by expanding any rectilinear shape and the order of convergence is the inverse of the minimum dimension. In fact, it is permissible to let some coordinates go to infinity before others.

Remark 6: Of course Theorems 2 and 4 follow immediately from Theorems 5 and 6 but we preferred to present the simple direct proofs of Secs. II and III for the reasons given in remark 4.

V. BACK TO TWO DIMENSIONS

In this section we consider the analyticity of various methods of summing the elements of the set

$$A_s = \{(-1)^{j+k}/(j^2+k^2)^s:\ (j,k)\in\mathbb{Z}/(0,0)\}.$$

From Theorem 6, it follows that the method of expanding squares leads to $d_2(2s)$, which is analytic for $\operatorname{Re} s > 0$. In fact, expanding rectangles of any shape with sides parallel to the axes lead to $d_2(2s)$. In Theorem 1, we showed that the method of expanding circles converged when $s = \frac{1}{2}$, but there is no reason to believe that $d_2(1)$ is obtained unless one shows that the appropriate function is analytic. Using the notation of Sec. II, let

$$G(s) = \sum_{n=1}^{\infty} \frac{(-1)^n C_2(n)}{n^s}, \tag{29}$$

whenever the right-hand side converges. Then $G(s)$ is the sum of the elements of A_s obtained by expanding circles.

Theorem 7: The function $G(s)$ exists and is analytic for $\operatorname{Re} s > \frac{1}{3}$. Thus, $G(s) = d_2(2s)$ if $\operatorname{Re} s > \frac{1}{3}$; in particular, $d_2(1) = \sum_{n=1}^{\infty}(-1)^n C_2(n)/n^{1/2}$.

Proof: As in the proof of Theorem 1, let $B_n = \sum_{k=1}^{n}(-1)^k C_2(k)$. By (8),

$$B_n = O(n^{1/3-\epsilon}), \quad \text{for some } \epsilon > 0. \tag{30}$$

Define $G_n(s)$ for all s with $\operatorname{Re} s > 0$ by

$$G_n(s) = \sum_{k=1}^{n} \frac{(-1)^k C_2(k)}{k^s}.$$

As in (9)

$$G_n(s) = \frac{B_n}{(n+1)^s} + \sum_{k=1}^{n} B_k[k^{-s} - (k+1)^{-s}]. \tag{31}$$

By (30), if $\operatorname{Re} s > \frac{1}{3}$, then $|B_n/(n+1)^s| \to 0$, uniformly in s. Note that

$$\begin{aligned} &|k^{-s} - (k+1)^s| \\ &= \left| (-s)\int_k^{k+1} t^{-(s+1)}\,dt \right| < |s| \int_k^{k+1} t^{-(\operatorname{Re} s+1)}\,dt \\ &< |s|\, k^{-(\operatorname{Re} s+1)}. \end{aligned} \tag{32}$$

So for $s \in R_M = \{z: \operatorname{Re} z > \frac{1}{3},\ |z| < M\}$ with M a fixed positive number and $1 \leqslant N \leqslant N'$,

$$\begin{aligned} &\left| \sum_{k=N}^{N'} B_k[k^{-s} - (k+1)^{-s}] \right| \\ &< \sum_{k=N}^{N'} |B_k|\, |k^{-s} - (k+1)^{-s}| \\ &< K \sum_{k=N}^{N'} k^{1/3-\epsilon} |k^{-s} - (k+1)^{-s}| \quad \text{by (30)} \\ &< KM \sum_{k=N}^{N'} k^{1/3-\epsilon-\operatorname{Re} s-1} \quad \text{by (32)} \\ &< KMN^{1/3-\epsilon-\operatorname{Re} s} \\ &< KMN^{-\epsilon} \\ &\to 0 \quad \text{as } N,N' \to \infty, \end{aligned}$$

uniformly for $s \in R_m$. Thus, the sequence of functions $\{G_n(s)\}_{n=1}^{\infty}$ is uniformly Cauchy on R_M and it converges uniformly to a limit function $G(s)$. Furthermore, each $G_n(s)$ is analytic, so $G(s)$ is analytic for $s \in R_M$. Since M is arbitrary, $G(s)$ exists and is analytic for all s with $\operatorname{Re} s > \frac{1}{3}$. Q.E.D.

Remark 7: It is not know what the minimum non-negative β is, such that $G(s)$ exists for all s with $\operatorname{Re} s > \beta$. However, if we consider another method of summing the elements of A_s, we can get a very complete and illuminating analysis. This is the *method of expanding diamonds.*

For each $k = 1,2,3,\ldots$, and complex s with $\operatorname{Re} s > 0$, let

$$\delta_k(s) = \sum_{j=0}^{k} \{(k-j)^2 + j^2\}^{-s}.$$

For each $n = 1,2,3,\ldots$, let

$$\Delta_n(s) = 4\sum_{k=1}^{n}(-1)^k \delta_k(s) - 4X_n(s), \tag{33}$$

where $X_n(s) = \sum_{k=1}^{n}(-1)^k/k^s$. Note that $\Delta_n(s)$ counts the contributions within the diamond $|k| + |j| \leqslant n$. Now, $\lim_{n\to\infty} X_n(s) = \sum_{k=1}^{\infty}(-1)^k/k^s = -\eta(s)$ and η is known to be analytic for $\operatorname{Re} s > 0$. Therefore, in order to determine for which s the limit of the $\Delta_n(s)$ exists and is analytic, it is sufficient to analyze $\sum_{k=1}^{\infty}(-1)^k \delta_k(s)$. We begin by establishing a number of facts about the sequence of δ_k.

Proposition 2: (a) $\lim_{k\to\infty} \delta_k(\frac{1}{2}) = \sqrt{2}\ln(\sqrt{2}+1)$. Thus $\sum_{k=1}^{\infty}(-1)^k \delta_k(\frac{1}{2})$ diverges.

(b) For real $r > \frac{1}{2}$, $\delta_{k-1}(r) > \delta_k(r)$, $k = 2,3,4,\ldots$.

(c) $\sum_{k=1}^{\infty}(-1)^k \delta_k(s)$ exists and is analytic for $\operatorname{Re} s > \frac{1}{2}$.

Proof: (a)

$$\begin{aligned} \delta_k\left(\frac{1}{2}\right) &= \sum_{j=0}^{k} [(k-j)^2 + j^2]^{-1/2} \\ &= \sum_{j=0}^{k} \left(\frac{1}{k}\right) \left[\left(1 - \frac{j}{k}\right)^2 + \left(\frac{j}{k}\right)^2 \right]^{-1/2} \\ &\to \int_0^1 [(1-t)^2 + t^2]^{-1/2}\,dt, \quad \text{as } k \to \infty \\ &= \sqrt{2}\ln(\sqrt{2}+1). \end{aligned}$$

For the proofs of (b) and (c) it is very convenient to introduce the following function. For $\operatorname{Re} s \geqslant \frac{1}{2}$, let

$$V(s) = 2s(2^s)\int_0^{\pi/4} \cos^{2s}\theta\, d\theta - 1.$$

Then V is continuous and $V(\frac{1}{2}) = 0$. If $r \geqslant \frac{1}{2}$, then

$$V(r) = 2r\int_0^{\pi/4} (2\cos^2\theta)^r\, d\theta - 1$$
$$\geqslant \int_0^{\pi/4} \sqrt{2}\cos\theta\, d\theta - 1 = 0. \tag{34}$$

We proceed now to the proof of (b).

Let $r \geqslant \frac{1}{2}$ and $k \geqslant 2$,

$$\begin{aligned}
&\delta_{k-1}(r) - \delta_k(r)\\
&= \sum_{j=1}^{k} \{[(k-j)^2 + (j-1)^2]^{-r} - [(k-j)^2 + j^2]^{-r}\} - k^{-2r}\\
&= \sum_{j=1}^{k}\int_{j-1}^{j} \frac{2rt\, dt}{[(k-j)^2 + t^2]^{r+1}} - k^{-2r} \geqslant \sum_{j=1}^{k}\int_{j-1}^{j} \frac{2rt\, dt}{[(k-t)^2 + t^2]^{r+1}} - k^{-2r}\\
&= \int_0^k \frac{2rt\, dt}{[(k-t)^2 + t^2]^{r+1}} - k^{-2r} = k^{-2r}\left\{\int_0^1 \frac{2ru\, du}{[(1-u)^2 + u^2]^{r+1}} - 1\right\}\\
&= k^{-2r}\left\{2r\int_{-1/2}^{1/2} \frac{(v+1/2)dv}{2^{r+1}(v^2 + 1/4)^{r+1}} - 1\right\} \quad \left(v = u - \frac{1}{2}\right)\\
&= k^{-2r}\left\{\frac{2r}{2^{r+1}}\int_0^{1/2} \frac{dv}{(v^2 + 1/4)^{r+1}} - 1\right\} = k^{-2r}\left\{2r\int_0^{\pi/4} (2\cos^2\theta)^r\, d\theta - 1\right\} \quad (\tan\theta = 2v)\\
&= k^{-2r}V(r) \geqslant 0.
\end{aligned}$$

That is, $\delta_{k-1}(r) \geqslant \delta_k(r)$, for $r \geqslant \frac{1}{2}$, $k = 2,3,4,\ldots$.

To prove (c), let $\epsilon > 0$ and $M < \infty$ be arbitrary. Let

$R = \{z: \operatorname{Re} z \geqslant \frac{1}{2} + \epsilon \text{ and } |z| < M\}.$

For $s \in R$, let $r = \operatorname{Re} s$. For $k \geqslant 2$, we can write

$$\begin{aligned}
\delta_{k-1}(s) - \delta_k(s) &= \sum_{j=1}^{k-1} \{[(k-j)^2 + (j-1)^2]^{-s} - [(k-j)^2 + j^2]^{-s}\} + (k-1)^{-2s} - 2k^{-2s}\\
&= \sum_{j=1}^{k-1}\int_{j-1}^{j} \frac{2st\, dt}{[(k-j)^2 + t^2]^{s+1}} + (k-1)^{-2s} - 2k^{-2s}.
\end{aligned}$$

Thus,

$$\begin{aligned}
|\delta_{k-1}(s) - \delta_k(s)| &\leqslant \sum_{j=1}^{k-1}\int_{j-1}^{j} \frac{2|s|t\, dt}{[(k-j)^2 + t^2]^{r+1}} + 3(k-1)^{-2r}\\
&\leqslant 2M\sum_{j=1}^{k-1}\int_{j-1}^{j} \frac{t\, dt}{[(k-1-t)^2 + t^2]^{r+1}} + 3(k-1)^{-2r}\\
&= 2M\int_0^{k-1} \frac{t\, dt}{[(k-1-t)^2 + t^2]^{r+1}} + 3(k-1)^{-2r}\\
&= (k-1)^{-2r}\left\{2M\int_0^1 \frac{u\, du}{[(1-u)^2 + u^2]^{r+1}} + 3\right\} = (k-1)^{-2r}\{(M/r)V(r) + (M/r) + 3\} \leqslant (k-1)^{-2r}C,
\end{aligned} \tag{35}$$

where C is the maximum of the continuous function $(M/r)V(r) + M/r + 3$ for $\frac{1}{2} + \epsilon \leqslant r \leqslant M$. Now, for each n, $\sum_{k=1}^{n}(\delta_{2k-1}(s) - \delta_{2k}(s))$ is an analytic function of s for $\operatorname{Re} s > \frac{1}{2}$ and

$$\sum_{k=1}^{\infty}(-1)^k\delta_k(s) = \lim_{n\to\infty}\left[-\sum_{k=1}^{n}(\delta_{2k-1}(s) - \delta_{2k}(s))\right]$$

exists uniformly on R, by (35) and the Weierstrass M test. Since $\epsilon > 0$ and $M < \infty$ are arbitrary, (c) has been established.

Q.E.D.

We can now describe the behavior of the diamond sums.

Theorem 8: For each complex number s with $\operatorname{Re} s > 0$ and each $n = 1,2,\ldots$, let

$$\varDelta_n(s) = \sum_{l=1}^{n}(-1)^l\left\{\sum_{|j|+|k|=l}[j^2 + k^2]^{-s}\right\}.$$

Then $\lim_{n\to\infty}\varDelta_n(s)$ exists and is analytic for $\operatorname{Re} s > \frac{1}{2}$. Although $\{\varDelta_n(\frac{1}{2})\}_{n=1}^{\infty}$ fails to converge,

$$d_2(1) = \lim_{r\to 1/2+}\left(\lim_{n\to\infty}\varDelta_n(r)\right). \tag{36}$$

Proof: These claims all follow immediately from Proposition 2.

Remark 8: Further analysis along the lines of Proposition 2 shows that although

$$\sum_{l=1}^{\infty}(-1)^l\left\{\sum_{|j|+|k|=l}[j^2 + k^2]^{-1/2}\right\}$$

is divergent, it is Cesáro summable or Abel summable to $d_2(1)$.

The diamond sums provide a nice illustration of how a method of summing the elements of A_s can be analytic in s for Re s large, then with decreasing Re s, this analyticity fails at a specific point. With the diamond sums it happens to be at $\frac{1}{2}$, with expanding squares or rectangles it is at 0. It is not clear where the expanding circles fails; it is at some point less than $\frac{1}{2}$. In three dimensions the method of expanding spheres fails at some point greater then $\frac{1}{2}$.

VI. THE HEXAGONAL LATTICE

As an illustration of what is obtained when one studies other crystal lattices in the above manner, we include a brief summary of results on Madelung's constant of a two-dimensional regular hexagonal lattice with ions of alternating unit charge.

In order to obtain a tractable expression for the terms appearing in the lattice sum, choose a coordinate system with an angle of $\phi = \pi/3$ between the positive axes. Then an arbitrary site in the lattice has coordinates $(\overline{n,m})$ with n and m integers. A charge of $+1$, -1, or 0 is attached to that site in a regular fashion (see Fig. 2). By considering the two parallelograms indicated in Fig. 2, one can see that this charge may be expressed by

$$q(n,m) = \tfrac{4}{3}[-\sin(n\theta)\sin((m-1)\theta) + \sin((m+1)\theta)\sin((n+1)\theta)], \quad \theta = 2\pi/3.$$

The distance of the point $(\overline{n,m})$ from the origin is given by

$$|(\overline{n,m})| = [(n+m/2)^2 + 3(m/2)^2]^{1/2}.$$

The set of numbers to be summed is then

$$C_s = \{q(n,m)/|(\overline{n,m})|^{2s}: \ (n,m)\in \mathbb{Z}^2/(0,0)\}, \quad \text{for } \operatorname{Re} s > 0.$$

As before, the elements of C_s are absolutely summable for Re $s > 1$ and we wish an analytic continuation of their sum to a region which includes $s = \frac{1}{2}$. Arguments, like those used for the diamond sums, will show that direct summation by expanding shells of hexagons will converge analytically for Re $s > \frac{1}{2}$ and even have a limit as s approaches $\frac{1}{2}$ from the right. However, for precise calculation purposes an analytic continuation via the integral transform methods is far superior. Let

$$H_2(2s) = \sum\left\{\frac{q(n,m)}{|(\overline{n,m})|^{2s}}: \ (n,m)\in \mathbb{Z}^2/(0,0)\right\}, \tag{37}$$

for Re $s > 1$. Then H_2 is an analytic function of s and the series converges absolutely. Substituting the expression for $q(n,m)$ and using elementary trignometric identities yields

$$H_2(2s) = \frac{2}{3}\sum{}'\frac{\cos((m-n)\theta)}{|(\overline{n,m})|^{2s}} - \left[\frac{1}{3}\sum{}'\frac{\sin((m-n)\theta)}{|(\overline{n,m})|^{2s}}\right], \tag{38}$$

where Σ' indicates the sum is over $(n,m)\in \mathbb{Z}^2/(0,0)$. By symmetry considerations, the second term in the right-hand side of (38) is zero. Further manipulation of theta functions (using the modular equation of order 3) produces a rectangular sum

$$H_2(2s) = (1 - 3^{1-s})\left[\sum{}'\frac{1}{(n^2+3m^2)^s} - \frac{1}{2}\sum{}'\frac{(-1)^{n+m}}{(n^2+3m^2)^s}\right]. \tag{39}$$

A theta function identity due to Cauchy[10] and a Mellin transform yields

$$H_2(2s) = 3(1 - 3^{1-s})\xi(s)L_{-3}(s), \tag{40}$$

where $\xi(s)$ is the standard zeta function $(\Sigma_{n=1}^{\infty} n^{-s})$ and

$$L_{-3}(s) = 1 - 2^{-s} + 4^{-s} - 5^{-s} + 7^{-s} - 8^{-s} + \cdots.$$

The formula (40) can also be deduced directly from (38) by using results in Sec. IV of Glasser and Zucker.[1] While the intermediate sums (39) and (40) are only analytic for Re $s > 1$, the standard continuation of the zeta function,

$$(1 - 2^{1-s})\xi(s) = \sum_{n=1}^{\infty}(-1)^{n+1}n^{-s} = \alpha(s),$$

gives

$$H_2(2s) = 3(1 - 3^{1-s})(1 - 2^{1-s})^{-1}\alpha(s)L_{-3}(s). \tag{41}$$

The right-hand side of (41) is an analytic function of s for Re $s > 0$ and therefore (41) provides the required analytic continuation, which is necessary for Madelung's constant of this hexagonal crystal lattice:

$$H_2(1) = 3(\sqrt{3} - 1)(\sqrt{2} + 1)\alpha(1/2)L_{-3}(1/2). \tag{42}$$

This can be considered as a solution to this lattice sum problem, as both $\alpha(\frac{1}{2})$ and $L_{-3}(\frac{1}{2})$ can be rapidly calculated by known techniques. At $s = 1$, we have an exact result:

$$H_2(2) = \sqrt{3}\pi \log 3. \tag{43}$$

VII. CONCLUDING REMARKS

We have investigated some fundamental properties of the multiply indexed series involved in the definition of Madelung's constant for an NaCl-type ionic crystal in two and three dimensions. We have provided elementary proofs that convergent series are obtained if the series is summed by letting the shape of a basic unit cell expand. The natural method of summing the effects of all ions within a fixed distance and letting the distance go to infinity leads to a convergent series in two dimensions but not so in three dimensions.

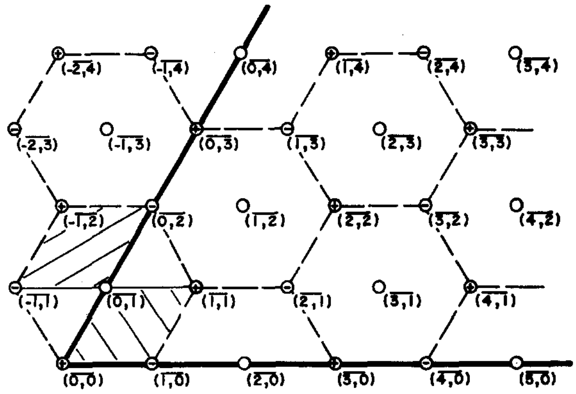


FIG. 2. The hexagonal lattice.

We have provided a unity to the concept of Madelung's constant by the use of analytic continuation of a complex function. Thus, although conditionally convergent when summed by expanding squares (or cubes), other methods of summing will provide the same answer provided that they are "analytic" in the correct sense. We have provided an analysis of the expanding circles and expanding diamonds methods in two dimensions to illustrate this point.

Perhaps the most important results are those in Sec. IV, rationalizing the integral transformation methods with the direct summation methods. These integral transform methods are the most useful in practice as they lead to very rapidly convergent series.

In the course of these investigations we have encountered many curious facts, most of which are probably known to experts in the area. However, the formulas (42) and (43) seem to be unknown and may be of sufficient interest to have been included; at least, as an illustration that the techniques of analytic continuation are applicable to other lattices.

ACKNOWLEDGMENTS

The authors would like to thank Professor O. Knop for originally bringing our attention to the problems involved in lattice sums.

The Natural Sciences and Engineering Research Council of Canada partially supported this work.

[1]M. L. Glasser and I. J. Zucker, in *Theoretical Chemistry: Advances and Perspectives*, edited by D. Henderson (Academic, New York, 1980), Vol. 5, pp. 67–139.

[2]E. L. Burrows and S. F. A. Kettle, J. Chem. Ed. **52**, 58 (1975).

[3]E. S. Campbell, J. Phys. Chem. Solids **24**, 197 (1963).

[4]H. Evjen, Phys. Rev. **39**, 675 (1932).

[5]K. Højendahl, K. Dan. Vidensk. Selsk. Mat. Fys. Medd. **16**, 135 (1938).

[6]P. Ewald, Ann. Phys. **64**, 253 (1921).

[7]C. Kittel, *Introduction To Solid State Physics* (Wiley, New York, 1953).

[8]J. S. Blakemore, *Solid State Physics* (Saunders, Philadelphia, 1969).

[9]O. Emersleben, Math. Nachr. **4**, 468 (1951).

[10]L. F. Dickson, *Theory of Numbers* (Chelsea, New York, 1920), Vol. II.

[11]G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, *4*[e] (Clarendon, Oxford, 1960).

Bianchi type I cosmological model with a viscous fluid

A. Banerjee, S. B. Duttachoudhury, and Abhik Kumar Sanyal
Department of Physics, Jadavpur University, Calcutta 700 032, India

(Received 25 February 1985; accepted for publication 26 April 1985)

Bianchi I cosmological models consisting of a fluid with both bulk and shear viscosity are studied. It is shown how the dynamical importance of the shear and the fluid density change in the course of evolution. Exact solutions with an equation of state $p = \rho$ for a stiff fluid are also obtained in several special cases, assuming the viscosity coefficients to be the power functions of the density. The results are, in some relevant cases, compared with those of Belinskiĭ and Khalatnikov (1976) in the asymptotic limits and are seen to agree with them in that the models start with $\rho = 0$ at the beginning and evolve with the creation of matter by the gravitational field, finally approaching the Friedmann universe.

I. INTRODUCTION

The presence of viscosity in the fluid content introduces many interesting features in the dynamics of homogeneous cosmological models.[1–5] The dissipative mechanisms not only modify the nature of the singularity usually occurring for a perfect fluid, but also can successfully account for the large entropy per baryon in the present universe. Misner[6,7] suggested that any anisotropy in an expanding universe would be reduced to a rather insignificant level today by neutrino viscosity. Murphy[8] presented an exactly soluble cosmological model of Friedmann type in the presence of bulk viscosity alone. Later Banerjee and Santos[9] extended the calculations to more general cases such as $k = \pm 1$. It was shown, however, that in all the cases where the singularity is said to appear at infinite past, the fluid did not satisfy Hawking–Penrose energy conditions.

Exact solutions for homogeneous anisotropic models are not much known in the literature. There are, however, a few[10,11] which utilize certain simplifying assumptions to get exact solutions at the cost of a physically reasonable equation of state. Belinskiĭ and Khalatnikov[4] assumed an equation of state of the form $\rho \propto p$, but did not give any exact solution. They have, however, investigated some general features of the isotropic and anisotropic homogeneous cosmological models in the presence of bulk as well as shear viscosity in asymptotic limits. We consider in this paper the Bianchi I model with a fluid characterized by both bulk and shear viscosity and having an equation of state $\rho \propto p$. The viscosity coefficients are further assumed to be power functions of the matter density as suggested by Belinskiĭ and Khalatnikov.[4] Exact solutions in several particular cases for stiff fluid $\rho = p$ are worked out.

In Sec. II we have considered Einstein's field equation for a Bianchi I cosmological model and discussed how the dynamical importance of the shear and matter change in the course of the cosmological evolution. The entropy variation is also explicitly stated.

In Sec. III exact solutions are obtained and their asymptotic characteristics are studied. The general behavior in the limits is compared with that of Belinskiĭ and Khalatnikov[4] at some places. Explicit solutions could be obtained considering the viscous coefficients only for a few restricted power functions of the mass density. These include the special cases with constant viscosity coefficients. In many of these cases the matter density vanishes at the initial instant, then increases in the course of evolution and finally reduces to zero. In such cases, therefore, at the initial singularity the metric is determined by the free space Einstein equations. In this context Belinskiĭ and Khalatnikov[4] remarked that the gravitational field creates the matter in the course of evolution.

II. EINSTEIN'S FIELD EQUATIONS AND SOME GENERAL RESULTS

The metric of the homogeneous Bianchi type I model is

$$ds^2 = -dt^2 + e^{2\alpha} dx^2 + e^{2\beta} dy^2 + e^{2\gamma} dz^2, \tag{2.1}$$

where α, β, and γ are functions of time alone. The energy momentum tensor of the viscous fluid[12] is given by

$$T_{ij} = (\rho + \bar{p}) v_i v_j + \bar{p} g_{ij} - \eta \mu_{ij}, \tag{2.2}$$

with

$$\bar{p} = p - (\zeta - \tfrac{2}{3}\eta)\theta \tag{2.3}$$

and

$$\mu_{ij} = v_{i;j} + v_{j;i} + v_i v^a v_{j;a} + v_j v^a v_{i;a}, \tag{2.4}$$

where

$$\theta = v^a;a.$$

In the above equations ζ and η stand for the bulk and shear viscosity coefficients, ρ and p are the mass density and pressure, respectively, $\bar{p}$ is the effective pressure, and v_i represents the four-velocity, so that

$$v_i v^i = -1. \tag{2.5}$$

Choosing units $8\pi G = C = 1$, Einstein's field equations can be written as

$$R^i_j - \tfrac{1}{2}\delta^i_j R = -T^i_j. \tag{2.6}$$

Using comoving coordinates, so that $v^i = \delta^i_0$, the explicit forms of Eq. (2.6) are

$$\tfrac{2}{3}(\dot{R}/R)^2 - \tfrac{1}{3}(\dot{\alpha}^2 + \dot{\beta}^2 + \dot{\gamma}^2) = \rho, \tag{2.7}$$

$$\ddot{\beta} + \ddot{\gamma} + \tfrac{3}{2}(\dot{R}/R)(\dot{\beta} + \dot{\gamma} - \dot{\alpha}) + \tfrac{1}{2}(\dot{\alpha}^2 + \dot{\beta}^2 + \dot{\gamma}^2) = -(\bar{p} - 2\eta\dot{\alpha}), \tag{2.8}$$

$$\ddot{\gamma} + \ddot{\alpha} + \tfrac{3}{2}(\dot{R}/R)(\dot{\gamma} + \dot{\alpha} - \dot{\beta}) + \tfrac{1}{2}(\dot{\alpha}^2 + \dot{\beta}^2 + \dot{\gamma}^2) = -(\bar{p} - 2\eta\dot{\beta}), \tag{2.9}$$

 0022-2488/85/113010-06$02.50

$$\ddot{\alpha} + \ddot{\beta} + \tfrac{3}{2}(\dot{R}/R)(\dot{\alpha} + \dot{\beta} - \dot{\gamma}) + (\dot{\alpha}^2 + \dot{\beta}^2 + \dot{\gamma}^2) = -(\bar{p} - 2\eta\dot{\gamma}), \tag{2.10}$$

where the dot indicates time differentiation and

$$R^3 = \exp(\alpha + \beta + \gamma). \tag{2.11}$$

The usual definitions of the dynamical scalars such as the expansion scalar θ and the shear scalar σ are considered to be

$$\theta = v^i_{;i} \quad \text{and} \quad \sigma^2 = \tfrac{1}{2}\sigma_{ij}\sigma^{ij}, \tag{2.12}$$

where

$$\sigma_{ij} = v_{(i;j)} + \tfrac{1}{2}(v_{i;k}v^k v_j + v_{j;k}v^k v_i) + \tfrac{1}{3}\theta(g_{ij} + V_i v_j). \tag{2.13}$$

For the Bianchi type I metric with comoving coordinates we have

$$\theta = 3(\dot{R}/R) \tag{2.14}$$

and

$$2\sigma^2 = (\dot{\alpha}^2 + \dot{\beta}^2 + \dot{\gamma}^2) - \tfrac{1}{3}\theta^2. \tag{2.15}$$

The field equations (2.7)–(2.10) now yield

$$T^4_4 = \tfrac{1}{3}\theta^2 - \sigma^2 = \rho \tag{2.16}$$

and

$$g_{ij}G^{ij} = 2\dot{\theta} + \tfrac{4}{3}\theta^2 + 2\sigma^2 = \rho - 3(\bar{p} - \zeta\theta). \tag{2.17}$$

One can further obtain from the Bianchi identity

$$\dot{\rho} = -(\rho + p)\theta + \zeta\theta^2 + 4\eta\sigma^2. \tag{2.18}$$

It follows directly from Eq. (2.18) that for contraction, that is, $\theta < 0$ we have $\dot{\rho} > 0$ so that the matter density increases or decreases depending on whether the viscous heating is greater or less than the cooling due to expansion. It may be mentioned here that for an ultrarelativistic fluid $p = \tfrac{1}{3}\rho$ and $\zeta = 0$, Stewart[13] has shown that the rate of viscous heating does not exceed one-half the rate of adiabatic cooling due to expansion.

Now, eliminating σ^2 from Eqs. (2.17) and (2.18), one readily obtains

$$\dot{\theta} = \tfrac{3}{2}(\rho - p + \zeta\theta) - \theta^2 \tag{2.19}$$

and

$$\dot{\rho} = -(\rho + p)\theta + \zeta\theta^2 + 4\eta(\tfrac{1}{3}\theta^2 - \rho). \tag{2.20}$$

The relation (2.19) may be written as

$$\frac{d}{dt}[\ln(\theta^2 R^6)] = 3[(\rho - p)\theta^{-1} + \zeta]. \tag{2.21}$$

Using Eq. (2.16), Eq. (2.19) can also be expressed in a different form such as

$$\dot{\theta} = -2\sigma^2 - \tfrac{1}{3}\theta^2 - \tfrac{1}{2}[\rho + 3(p - \zeta\theta)]. \tag{2.22}$$

This is exactly the Raychaudhuri equation.[14] Further, we have

$$R_{ij}v^i v^j = -\tfrac{1}{2}[\rho + 3(p - \zeta\theta)]. \tag{2.23}$$

The Hawking–Penrose energy condition is satisfied when $R_{ij}v^i v^j \leqslant 0$. Thus, in a contracting model, so long as the fluid density and pressure remain positive, the energy conditions are satisfied. When the bulk viscosity is insignificant the energy condition is independent of the viscosity of the fluid. Again from the Raychaudhuri equation (2.22) it is evident that with the energy condition being satisfied, $\dot{\theta} < 0$, so that there is no bounce from a minimum volume.

Now for $\theta \neq 0$, that is, for a nonstatic model, one has, in view of Eq. (2.16),

$$(\sigma^2/\theta^2)^{\cdot} = -(\rho/\theta^2)^{\cdot}. \tag{2.24}$$

Using the expressions for $\dot{\rho}$ and $\dot{\theta}$ from Eqs. (2.19) and (2.20) in Eq. (2.24), one obtains after simplification the following result:

$$(\sigma^2/\theta^2)^{\cdot} = -(\rho/\theta^2)^{\cdot} = -(\sigma^2/\theta^2)[3(\rho - p)\theta^{-1} + 3\zeta + 4\eta]. \tag{2.25}$$

The reasonable physical properties of the fluid demand $\rho \geqslant p > 0$, $\zeta > 0$, and $\eta > 0$, so that for an expanding model $(\theta > 0)$ we have $(\rho/\theta^2)^{\cdot} > 0$ and $(\sigma^2/\theta^2)^{\cdot} < 0$. It is evident, therefore, that (ρ/θ^2) increases with time, while (σ^2/θ^2) decreases. The dynamical influence of matter, therefore, increases with expansion, whereas that of shear decreases. For contraction, however, $\theta < 0$ and nothing can be said with certainty. It is interesting to note that for a stiff fluid, that is, $\rho = p$, (ρ/θ^2) is greater than zero and (σ^2/θ^2) is less than zero so long as viscosity coefficients are positive and this behavior holds irrespective of whether the model expands or contracts. Combining Eqs. (2.21) and (2.25), one obtains

$$[\ln(\sigma^2 R^6)]^{\cdot} = -4\eta, \tag{2.26}$$

which in turn can also be written as

$$(\sigma^2)^{\cdot} = -2(2\eta + \theta)\sigma^2. \tag{2.27}$$

This is the shear propagation equation. It follows from Eq. (2.27) that for the expanding model $\theta > 0$ the shear decreases with time. The rate of work done by anisotropic stresses augments the shear dissipation. It is also evident from Eq. (2.27) that the shear dissipation depends on the expansion rate θ, which is again affected by the presence of bulk viscosity as is evident from Eq. (2.22). The bulk viscosity has thus a significant role in the process of the shear dissipation mechanism.

When η is assumed to be a constant the relation (2.27) can be directly integrated to yield

$$\sigma^2 = (\sigma_0^2/R^6)\,e^{-4\eta t}, \tag{2.28}$$

σ_0 being the integration constant. The effect of shear viscosity is to reduce the anisotropy in the course of time in the form of an exponential factor. This purely relativistic result is due to Misner.[6,7]

Following Belinskiĭ and Khalatnikov[4] the time derivative of the entropy density in the model is given by

$$\dot{\Sigma}/\Sigma = \dot{\rho}/(\rho + p), \tag{2.29}$$

where Σ is the entropy density. Defining the total entropy by $S = R^3\Sigma$, one gets from Eq. (2.20) using Eq. (2.29) the relation

$$\dot{S}/S = (\zeta\theta^2 + 4\eta\sigma^2)/(\rho + p). \tag{2.30}$$

We now restrict ourselves to an equation of state of the form

$$p = (\gamma - 1)\rho, \quad 1 \leqslant \gamma \leqslant 2, \tag{2.31}$$

and assume that the viscosity coefficients are constants so that $\zeta = \zeta_0$ and $\eta = \eta_0$. The qualitative aspects of the presence of viscosity will, however, be present in such a restricted case also (see Misner[6,7] and Treciokas and Ellis[15]). In this case Eq. (2.30) can be written as

$$\dot{S}/S = [\zeta_0 + 4\eta_0(\sigma^2/\theta^2)]/\gamma(\rho/\theta^2). \tag{2.32}$$

By further differentiation with respect to time and using Eqs. (2.25) and (2.32), we find

$$\begin{aligned} S^{-1}\ddot{S} &= \gamma^{-2}(\rho/\theta^2)^{-2}[\zeta_0^2 - (\sigma^2/\theta^2)(\rho/\theta^2) \\ &\quad \times \{16\eta_0^2 + \gamma(2-\gamma)(3\zeta_0 + 4\eta_0)\theta\} \\ &\quad + 8\eta_0(\gamma - 1)(\zeta_0 + \tfrac{2}{3}\eta_0) + 3\gamma\zeta_0^2]. \end{aligned} \tag{2.33}$$

For an expanding model, (σ^2/θ^2) decreases with time, whereas (ρ/θ^2) increases, as was already discussed. The minimum of (σ^2/θ^2) is zero, when $\rho/\theta^2 = \frac{1}{3}$ and $S^{-1}\ddot{S} = \zeta_0^2\gamma^{-2}(\rho/\theta^2)^{-2}$, which is greater than zero. In the course of time (ρ/θ^2) decreases and we consider the extreme case when $\rho/\theta^2 = 0$. At this instant $\sigma^2/\theta^2 = \frac{1}{3}$ and from Eq. (2.33) it is evident that $\ddot{S} < 0$. The relation (2.32) indicates that $\dot{S} > 0$, that is, the total entropy always increases for nonnegative values of matter density, whereas for expansion, $\ddot{S}$ is initially negative and later becomes positive in the course of time. Since $\dot{S}/S \to \infty$ as $\rho/\theta^2 \to 0$, we have the $S - t$ curve intersecting the time axis. It means that S reduces to zero at some finite time. The picture is more clear for a stiff fluid when $\rho = p$. We have then from Eq. (2.25)

$$(\sigma^2/\theta^2)^{\cdot}/(\sigma^2/\theta^2) = (3\zeta_0 + 4\eta_0),$$

which yields on integration

$$\sigma^2/\theta^2 = A^2 e^{-(3\zeta_0 + 4\eta_0)t}, \tag{2.34}$$

where A^2 is the magnitude of the ratio σ^2/θ^2 at $t = 0$. It is evident from Eq. (2.34) that for both ζ and η as constants the ratio of shear to expansion decays exponentially and the rate falls in the absence of either bulk or shear viscosity. From Eqs. (2.16) and (2.34) one gets the expression for (ρ/θ^2) in the form

$$\rho/\theta^2 = \tfrac{1}{3} - A^2 e^{-(3\zeta_0 + 4\eta_0)t}. \tag{2.35}$$

So, initially, if one starts with zero mass density at some finite time one must have $(\rho/\theta^2)_{\max} = \frac{1}{3}$ at $t \to \infty$. In view of Eqs. (2.34) and (2.35), Eq. (2.32) can now be written as

$$\frac{\dot{S}}{S} = \frac{\zeta_0 + 4\eta_0 A^2 e^{-(3\zeta_0 + 4\eta_0)t}}{2(\frac{1}{3} - A^2 e^{-(3\zeta_0 + 4\eta_0)t})}, \tag{2.36}$$

which in turn yields on integration

$$S = S_0[\tfrac{1}{3}e^{3\zeta_0 t} - A^2 e^{-4\eta_0 t}]^{1/2}, \tag{2.37}$$

with S_0 being the constant of integration. It is evident that at

$$t = [\ln(3A^2)](3\zeta_0 + 4\eta_0)^{-1}, \tag{2.38}$$

the total entropy $S = 0$, when we also have $\rho/\theta^2 = 0$. Again as $t \to \infty$, $S \to \infty$ and ρ/θ^2 approaches its maximum value. Combining Eqs. (2.35) and (2.37) together one can also write

$$S^2 = S_0^2(\rho/\theta^2)e^{3\zeta_0 t}, \tag{2.39}$$

which in turn demands that the matter density must be nonnegative in this case.

III. SPECIAL SOLUTIONS FOR A BIANCHI I MODEL WITH A VISCOUS FLUID

In what follows we consider some special cases with restrictions on the behavior of the bulk and shear viscosity coefficients. It is true that the assumptions regarding these viscosity coefficients may not always be valid in an actual fluid throughout the entire evolution history of the cosmological models hitherto discussed: the solutions are nevertheless interesting in indicating the role of viscosity in cosmological evolution.

In view of Eqs. (2.18) and (2.26) we now have the relation

$$\frac{d}{dt}[(\rho + \sigma^2)R^6] = \zeta\theta^2 R^6. \tag{3.1}$$

Using Eq. (2.16) in Eq. (3.1) we further obtain

$$\frac{d}{dt}[\ln(\theta^2 R^6)] = 3\zeta. \tag{3.2}$$

A. Case I

In this case,

$$\zeta = 0, \quad \eta = \eta_0 \rho^n,$$

where η_0 and n are constants. Equation (3.2) can immediately be integrated in this case to yield

$$R^3 = R_0^3 t, \tag{3.3}$$

where R_0 is an arbitrary constant and the time coordinate is chosen such that the proper volume vanishes at $t = 0$. It is, however, not difficult to show that one will get the same solution (3.3) if one assumes the shear and bulk viscosity to be absent for a perfect fluid. The expansion scalar θ is given by

$$\theta = t^{-1}. \tag{3.4}$$

Equation (2.26) now yields in view of Eq. (2.16)

$$(\sigma^2/\theta^2)^{-1}(\sigma^2/\theta^2)^{\cdot} = -4\eta_0\theta^{2n}(\tfrac{1}{3} - \sigma^2/\theta^2)^n. \tag{3.5}$$

Writing $\sigma^2/\theta^2 = y$ and using the relation (3.4), Eq. (3.5) can be written as a first-order differential equation

$$\dot{y}/y = -4\eta_0 t^{-2n}(\tfrac{1}{3} - y)^n. \tag{3.6}$$

In view of Eq. (2.16) ρ is positive when $\frac{1}{3} - y > 0$. For $n = 1$ the solution for y is obtained by integrating Eq. (3.6) as

$$y = \sigma^2/\theta^2 = \tfrac{1}{3}(1 + a^2 e^{-(4/3)(\eta_0/t)})^{-1}, \tag{3.7}$$

where a^2 is a positive constant. It follows from Eq. (3.7), in view of Eq. (3.4), that

$$\sigma^2 = (1/3t^2)(1 + a^2 e^{-(4/3)(\eta_0/t)})^{-1}. \tag{3.8}$$

Using Eq. (2.16) we, therefore, obtain

$$\rho = (a^2/3t^2)e^{-(4/3)(\eta_0/t)}(1 + a^2 e^{-(4/3)(\eta_0/t)})^{-1}. \tag{3.9}$$

Now consider an expanding model, for which $\theta > 0$. In this case at $t \to 0$, $R^3 \to 0$ and both the shear and expansion scalars attain infinitely large magnitudes, while the density reduces to zero. It is an interesting behavior as noted previously by Belinskiĭ and Khalatnikov.[4] The singularity at $t = 0$ in this case corresponds to vanishing proper volume $R^3 = 0$. But, unlike in the usual case of cosmological singularity, the density vanishes in the limit instead of increasing to infinity. The matter density subsequently increases and again decreases to approach zero magnitude at the final stage at $t \to \infty$, as is evident from Eq. (3.9). In this limit, however, both the expansion (θ) and shear (σ^2) scalars vanish and $R^3 \to \infty$. In other words the model may be said to approach asymptotically the isotropic Friedmann universe (cf. Belinskiĭ and Khalatnikov[4]). On the other hand, collapse may be discussed for a negative value of t when $t<0$, $\theta<0$, which represents a contracting model. As $t \to -\infty$, all the quantities such as θ, σ^2, and ρ vanish with infinitely large proper volume representing a Friedmann model. In the course of time as $t \to 0$, the proper volume reduces to zero whereas $|\theta|$, σ^2, and ρ all increase indefinitely. These properties in asymptotic limits only are discussed by Belinskiĭ and Khalatnikov.[4]

For $n = \frac{3}{2}$ the relation (3.2) takes the following form:

$$\dot{y}/y = -(4\eta_0/t^3)(\tfrac{1}{3}-y)^{3/2}. \tag{3.10}$$

Since $(\sigma^2/\theta^2)^{\cdot} < 0$ for positive ζ and η, that is, $\dot{y}<0$, t should assume only positive values and thus one can have only expansion allowed in this case ($\theta>0$). Integrating Eq. (3.10), we have

$$\ln\left\{\frac{1-(1-3y)^{1/2}}{1+(1-3y)^{1/2}}\right\} + \frac{2}{(1-3y)^{1/2}} = \frac{2\eta_0}{3\sqrt{3}}\, t^{-2} + \text{const.} \tag{3.11}$$

For $n=2$ in Eq. (3.6) we find that the solution of y is given by

$$3\ln[y/(\tfrac{1}{3}-y)] + 1/(\tfrac{1}{3}-y) = \tfrac{4}{3}\eta_0 t^{-3} + \text{const.} \tag{3.12}$$

Though the solutions (3.11) and (3.12) are not in closed form, it is not very difficult to investigate the properties of these models at limits. The analysis is done in an identical manner as for $n=1$. The behavior can be seen to be almost identical in the limits $t \to 0$ or $t \to \infty$.

For $n = \frac{1}{2}$, Eq. (3.6) can be expressed as

$$\dot{y}/y = -4\eta_0 t^{-1}(\tfrac{1}{3}-y)^{1/2}, \tag{3.13}$$

which on integration yields

$$\frac{1-(1-3y)^{1/2}}{1+(1-3y)^{1/2}} = \left(\frac{t}{t_0}\right)^{-(4/\sqrt{3})\eta_0},$$

where t_0 is the constant of integration and is less than t as is clear from the above equation, which on further simplification gives

$$y = \tfrac{4}{3}\frac{(t/t_0)^{-(4/\sqrt{3})\eta_0}}{(1+(t/t_0)^{-(4/\sqrt{3})\eta_0})^2}. \tag{3.14}$$

One can now use Eq. (3.4) in Eq. (3.14) to obtain an expression for $\sigma^2(\because y = \sigma^2/\theta^2)$ as

$$\sigma^2 = (4/3t^2)(t/t_0)^{-(4/\sqrt{3})\eta_0}(1+(t/t_0)^{-4/\sqrt{3})\eta_0})^{-2} \tag{3.15}$$

and also

$$\rho = (1/3t^2)(1-(t/t_0)^{-(4/\sqrt{3})\eta_0})^2(1+(t/t_0)^{-(4/\sqrt{3})\eta_0})^{-2}. \tag{3.16}$$

For an expanding model $\theta>0$. In this case as $t \to t_0$, θ, R^3, and σ^2 all remain finite, while the matter density vanishes. On the other hand, as $t \to \infty$ the expansion (θ) and the shear (σ^2) scalars vanish although the proper volume increases indefinitely. The matter density ρ approaches zero, thus representive asymptotically of an isotropic Friedmann universe.

The simplest case is for $n=0$, that is, $\eta=\eta_0$. We now have from Eq. (3.6)

$$\dot{y}/y = -4\eta_0, \tag{3.17}$$

so that $y = \sigma^2/\theta^2 \propto e^{-4\eta_0 t}$, yielding the relation

$$\sigma^2 = (\sigma_0^2/t^2)e^{-4\eta_0 t}. \tag{3.18}$$

Further,

$$\rho = \tfrac{1}{3}\theta^2 - \sigma^2 = (1/t^2)[\tfrac{1}{3} - \sigma_0^2 e^{-4\eta_0 t}]. \tag{3.19}$$

The behavior in this case is quite different from the previous cases at least in the initial phase of expansion. Here at $t \to 0$, the physical and geometrical quantities such as ρ,θ,σ^2 become infinitely large. The magnitude of the constant σ_0^2, however, cannot be greater than $\frac{1}{3}$ for positive values of matter density ρ. On the other hand, as $t \to \infty$, all the quantities ρ, θ, σ^2 reduce to vanishingly small quantities.

It should be mentioned at this point that in all the models discussed so far, one can conclude that the Hawking–Penrose energy condition ($R_{ij}v^iv^j \leqslant 0$) is satisfied so long as $\rho \geqslant 0$. This is clear from the relation (2.23) and the fact that in the above models we have assumed $\zeta = 0$.

B. Case II

Let us now turn our attention to the nonvanishing values of bulk viscosity coefficients, that is, the situation where the bulk viscosity of the fluid cannot be completely ignored.

We assume

$$\zeta = \zeta_0 \quad \text{and} \quad \eta = \eta_0\rho^q, \tag{3.20}$$

where ζ_0 and η_0 are constants. Now, integrating Eq. (3.2) one gets

$$\theta = (\theta_0/R^6)e^{(3/2)\zeta_0 t}, \tag{3.21}$$

θ_0 being the constant of integration. Remembering that for the Bianchi I metric $\theta = 3(\dot{R}/R)$ and so integrating further Eq. (3.21), we have the solution for R, which is given by

$$R^3 = (2\theta_0/\zeta_0)(e^{(3/2)\zeta_0 t} + D). \tag{3.22}$$

Equation (3.20) therefore yields

$$\theta = \tfrac{3}{2}\zeta_0(e^{(3/2)\zeta_0 t}/[e^{(3/2)\zeta_0 t} + D]). \tag{3.23}$$

Now from Eqs. (2.25) and (3.20) one gets

$$\dot{y}/y = -(3\zeta_0 + 4\eta_0\rho^q).$$

which in turn can be written in view of Eq. (2.16) as

$$\dot{y}/y = -3\zeta_0 - 4\eta_0\theta^{2q}(\tfrac{1}{3}-y)^q. \tag{3.24}$$

The special case for $D=0$ is particularly simple and we discuss only this case. We therefore have $\theta = \frac{3}{2}\zeta_0$ so that $\dot{\theta}=0$. Here the expansion is steady and the rate is constant. One of the relatively simple cases is $q=1$. In this case $\eta = \eta_0\rho$ and hence we obtain from Eq. (3.24)

$$\dot{y}/y = -3\zeta_0 - 4\eta_0\theta^2(\tfrac{1}{3}-y). \tag{3.25}$$

Writing $3\zeta_0 = a_0$ and $9\eta_0^2\zeta_0^2 = b_0$, the relation (3.25) may be written as

$$\dot{y}/y = -a_0 - b_0(\tfrac{1}{3}-y).$$

where both a_0 and b_0 are greater than zero. Integrating Eq. (3.25) we further obtain

$$y/(c_0-y) = e^{b_0c_0(t_0-t)}, \tag{3.26}$$

So that one can write explicitly

$$\sigma^2/\theta^2 = c_0/(1+e^{b_0c_0(t-t_0)}), \tag{3.27}$$

where $c_0 = a_0/b_0 + \frac{1}{3}$ and t_0 is the constant of integration. Since here $\theta = \frac{3}{2}\zeta_0$, the expansion scalar θ is positive for the physical requirement $\zeta_0 > 0$. The above solution, therefore, represents an expanding model only. The density ρ vanishes at a finite time, when $\sigma^2/\theta^2 = \frac{1}{3}$, so that σ^2 remains finite. The proper volume represented by R^3 also has finite magnitude. But for $t \to \infty$, $R^3 \to \infty$, $\sigma^2 \to 0$, and $\rho \to \frac{3}{4}\zeta_0^2$. The singularity of vanishing volume $R^3 = 0$ exists at $t \to -\infty$, where the density is negative infinity. In fact prior to the instant corresponding to $\sigma^2/\theta^2 = \frac{1}{3}$ the density assumes only negative values. If one calculates $R_{ij}v^iv^j$ in this model, one finds it to be positive so that the energy condition is violated throughout.

Particularly, simple models can be constructed in this case, taking $q = 0$ in Eq. (3.20), so that

$$\zeta = \zeta_0, \quad \eta = \eta_0. \tag{3.28}$$

The expressions for R^3 and θ remain unaltered from those given in Eqs. (3.22) and (3.23), respectively. Equation (3.24) then reduces to

$$\dot{y}/y = -(3\zeta_0 + 4\eta_0). \tag{3.29}$$

This case is already mentioned in Eq. (2.34). Integrating Eq. (3.29), the solution can be obtained in the form

$$y = \sigma^2/\theta^2 = A^2 e^{-(3\zeta_0 + 4\eta_0)t},$$

so that

$$\sigma^2 = \tfrac{9}{4}A^2\zeta_0^2 e^{-4\eta_0 t}\left[e^{(3/2)\zeta_0 t} - D\right]^{-2}, \tag{3.30}$$

and the matter density ρ is expressed as

$$\rho = \theta^2\left(\frac{1}{3} - \frac{\sigma^2}{\theta^2}\right)$$

$$= \frac{9}{4}\frac{\zeta_0^2 e^{3\zeta_0 t}}{\left[e^{(3/2)\zeta_0 t} + D\right]^2}\left[\tfrac{1}{3} - A^2 e^{-(3\zeta_0 + 4\eta_0)t}\right]. \tag{3.31}$$

The maximum of σ^2/θ^2 is $\frac{1}{3}$ and this occurs at some finite time t given by Eq. (2.38). In this limit $\rho = 0$ and σ^2, θ^2 are both finite. For $D \geqslant 0$ the proper volume never reduces to zero magnitude as is evident from Eq. (3.22). On the other hand, as $t \to \infty$, $\theta \to \frac{3}{2}\zeta_0$, which is finite, $\sigma^2 \to 0$, and $\rho \to \frac{3}{4}\zeta_0^2$, but $R^3 \to \infty$. We note that though the proper volume increases to large dimension and the anisotropy reduces to zero the fluid density ρ does not vanish, unlike the Friedmann universe. For $D < 0$ we note that at a finite time $R^3 = 0$ and θ, σ^2, ρ all become infinitely large. But $t \to \infty$, $R^3 \to \infty$, $\sigma^2 \to 0$ but θ and ρ both remain finite. In the limit $t \to \infty$ the behavior of the model is independent of the sign of the constant D.

For $D \geqslant 0$ it can be shown that the Hawking–Penrose energy conditions are violated throughout and for $D < 0$ this happens for large time t.

IV. CONCLUSION

In summary, we have considered the Bianchi type I cosmological model with a viscous fluid, assuming the coefficients of viscosity as power functions of the matter density and considering an equation of state for a stiff fluid ($p = \rho$). In most of the cases it has been observed that the matter density is zero at the initial singularity but then increases in the course of evolution, finally vanishing again in the asymtotic limit, which implies that the gravitational field creates matter. In the case of expanding models it is found that the dynamical importance of matter increases while that of shear decreases in the course of evolution. For a stiff fluid, in particular, this result is shown to hold also for contracting models. In addition to the role of shear viscosity in the dissipation of shear it is pointed out that the bulk viscosity can also augment the shear dissipation. For the stiff fluid with constant viscosity coefficients it is observed that the bulk viscosity can be an effective mechanism for large entropy in the asymptotic limit when the model approaches the isotropic Friedmann universe. Although the role of shear viscosity also is to increase the entropy in the course of the expansion, its effect becomes gradually less compared to that of the bulk viscosity in the asymptotic limit. The magnitude of the entropy is low at the highly anisotropic initial phase of evolutin, as is observed from Eq. (2.37), and then increases subsequently.

In the preceding section we have considered two different cases. The first case is $\eta = \eta_0\rho^n$ and $\zeta = 0$. Solutions for particular values of n such as $n = \frac{1}{2}, 1, \frac{3}{2}, 2$ are explicitly given. It is found that in all these cases except for $n = 1$, only expanding models are allowed. For $n = \frac{1}{2}$, 1, and 2, the models have been found to approach the isotropic Friedmann universe asymptotically. The matter density vanishes at the initial phase of singularity, increasing subsequently during the evolution, ultimately reducing to zero in the asymptotic limit. For $n = \frac{3}{2}$, the solution is not in the closed form and as such the behavior cannot be studied. For $n = 1$ as $t \to -\infty$ the proper volume R^3 tends to infinity whereas θ, σ^2, and ρ vanish, representing a Friedmann model. In the course of time as $t \to 0$ the volume contracts and reduces to zero and the other physical scalars become infinitely large. In the case where the shear viscosity coefficient is assumed to be constant, i.e., $n = 0$, the behavior is different. Here at the initial phase of expansion ρ, θ, and σ^2 are infinitely large, but asymptotically, however, all of them reduce to vanishingly small quantities.

In the second case we have considered, $\eta = \eta_0\rho^q$, $\zeta = \zeta_0$. For $q = 0$ and 1 only expansion is found to be allowed. For $q = 1$, ρ vanishes at a finite time, keeping the proper volume finite. In the limit $t \to \infty$, the model isotropizes with infinite proper volume but the matter density is nonvanishing, unlike the previous cases. For $q = 0$, either of the two cases is observed, depending on the sign of an integration constant D. For $D \geqslant 0$, the asymptotic behavior is the same as for $q = 1$. But for $D < 0$, the proper volume vanishes at a finite time while θ, σ^2, and ρ take infinitely large magnitude at this instant. In the asymtotic limit, $t \to \infty$, the model is again identical to the case $q = 1$.

[1]C.B. Collins and J. M. Stewart, Mon. Not. R. Astron. Soc. **153**, 419 (1971).
[2]S. Weinberg, Astrophys. J. **168**, 175 (1971).

[3]J. D. Nightingale, Astrophys. J. **185**, 105 (1973).
[4]V. A. Belinskiĭ and I. M. Khalatnikov, Sov. Phys. JETP **42**, 205 (1976).
[5]V. A. Belinskiĭ and I. M. Khalatnikov, Sov. Phys. JETP **45**, 1 (1977).
[6]C. W. Misner, Nature **214**, 40 (1967).
[7]C. W. Misner, Astrophys. J. **151**, 431 (1968).
[8]G. L. Murphy, Phys. Rev. D **8**, 4231 (1973).
[9]A. Banerjee and N. O. Santos, J. Math. Phys. **26**, 878 (1985).
[10]A. Banerjee and N. O. Santos, Gen. Relativ. Gravit. **16**, 217 (1984).
[11]A. Banerjee and N. O. Santos, J. Math Phys. **24**, 2689 (1983).
[12]L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* (Pergamon, New York, 1959), pp. 47, 304, 499.
[13]J. M. Stewart, Mon. Not. R. Astron. Soc. **145**, 347 (1969).
[14]A. K. Raychaudhuri, Phys. Rev. **98**, 1123 (1955).
[15]R. Treciokas and G. F. R. Ellis, Commun. Math. Phys. **23**, 1 (1971).

ERRATUM

Erratum: New inequalities for the Coulomb T matrix in momentum space [J. Math. Phys. 25, 3033 (1984)]

H. van Haeringen
Department of Mathematics and Informatics, Delft University of Technology, Delft, The Netherlands

L. P. Kok
Institute for Theoretical Physics, P. O. Box 800, Groningen, The Netherlands

(Received 4 June 1985; accepted for publication 12 June 1985)

The four-line paragraph just before Eq. (1.3) was mistakenly replaced by a duplicate of the four-line paragraph at the end of Sec. I. Instead, it should read as follows:

In Ref. 1, the ratios R_c and R_{cl} have been studied extensively, by analytical means and by numerical calculations, and a number of inequalities have been derived. One particularly interesting inequality, viz.

