The construction of $B_l^{(1)}$ in terms of vertex operators and a Lie-admissible not associative algebra

M. V. Cougo-Pinto
Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627

(Received 16 July 1987; accepted for publication 23 September 1987)

The affine nontwisted Kac–Moody algebra $B_l^{(1)}$ is constructed in terms of an underlying not associative algebra obtained by tensoring an associative algebra of vertex operators with a given Lie-admissible algebra. The former has its origin in bosonic string theory and the latter is found to be essentially an algebra of contractions used in canonical field theory.

I. INTRODUCTION

An algebra $\mathfrak{a}$ is a vector space in which a bilinear closed operation,

$$\cdot:\ \mathfrak{a}\times\mathfrak{a}\to\mathfrak{a},\quad (x,y)\mapsto x\cdot y, \tag{1.1}$$

is defined. If the associative law is not assumed for this operation the algebra $\mathfrak{a}$ is referred to as nonassociative; if the associative law is actually violated in $\mathfrak{a}$ we call the algebra not associative.[1] A nonassociative algebra $\mathfrak{a}$ is completely defined by a multiplication table; by this we mean the definition of how the bilinear operation acts on some basis of $\mathfrak{a}$. For example, if we have

$$\mathfrak{a}=\bigoplus_{i\in Z}\mathbb{C}\,x_i, \tag{1.2a}$$

a multiplication table is given by

$$x_i\cdot x_j=\sum_{k\in Z}c_{ijk}x_k, \tag{1.2b}$$

where only a finite number of terms should be different from zero. Among the not associative algebras the ones that have a richer structure and are most well known are the Lie algebras.

It is possible to construct new algebras out of some given algebra $\mathfrak{a}$. For example, if $\mathfrak{a}$ is defined by (1.2) we can define $\mathfrak{a}^-$ as (1.2a) endowed with the operation

$$[x,y]:=x\cdot y-y\cdot x\,. \tag{1.3}$$

If $\mathfrak{a}^-$ is a Lie algebra then $\mathfrak{a}$ is called a Lie-admissible algebra.[2] Obviously, every associative algebra is Lie admissible. If $\mathfrak{a}$ is an associative algebra of endomorphisms on a vector space V then the construction of $\mathfrak{a}^-$ is a problem in the theory of representations of Lie algebras; the purpose of the construction is to use the much simpler properties of the associative algebra $\mathfrak{a}$ to study $\mathfrak{a}^-$. An antipodal problem consists in constructing a Lie algebra $\mathfrak{a}^-$ from a not associative algebra $\mathfrak{a}$ to use the results from the theory of Lie algebras to understand better the underlying $\mathfrak{a}$; with this purpose Albert introduced in 1948 the concept of Lie-admissible algebras.[3] In this case, the usefulness of the construction comes from the fact that not associative algebras, which are not Lie algebras, are in general quite untractable.

In what follows we present a construction of the affine Kac–Moody algebra[4] $B_l^{(1)}$ from a not associative algebra $\mathfrak{b}$. In Sec. II we present a not associative algebra $\mathfrak{a}$, whose origin is explained in Sec. V. Section III is a summary of the vertex operator formalism for constructing representations of simply laced algebras.[5–8] In Sec. IV we define $\mathfrak{b}$ in terms of the algebras of Secs. II and III and construct $B_l^{(1)}$ in terms of it. In Sec. V the relation between our construction and some representations of $B_l^{(1)}$ is established; from this relation we can obtain a method of constructing representations for $\mathfrak{a}$. Although in our case $\mathfrak{a}$ may be simple enough to allow the solution (5.5) to be guessed, the method presented can be useful in dealing with more complicated examples of not associative algebras.

The fact that $\mathfrak{a}$ is essentially an algebra of contractions used in quantum field theory shows that it can be of some relevance in renormalization theory (the most naive renormalization consists in subtracting contractions from a divergent free-field Hamiltonian). Although not associative algebras, other than Lie algebras and special Jordan algebras, have not been widely used in physics yet, it has been remarked recently that they may play an important role in the study of anomalies in quantum field theory.[9–12] In Witten's string field theory, for example, the question of associativity of the basic string operations is a fundamental one from the very beginning,[10] and the key to the construction of the closed string field theory of Strominger[11] is an associativity anomaly.[12] The discussions and results of Sec. IV show in a much simpler context a mechanism by which an algebra of operators may or may not exhibit associativity anomaly. The construction of the representations (5.5) and (5.6) shows then a process to eliminate the anomaly; the representations relate the two classes of algebras [the not associative like $\mathfrak{a}$ and the "anomaly-free" like $\mathbb{C}(\Psi)$ and $\mathbb{C}(\tilde{\Psi})$] that can be used in a given theory (in our case a theory possessing $B_l^{(1)}$ symmetry).

II. THE ALGEBRA OF CONTRACTIONS

Let us consider the vector space

$$\mathfrak{a}=\bigoplus_{m\in Z}\mathbb{C}a_m\oplus\mathbb{C}1 \tag{2.1a}$$

endowed with a bilinear operation defined by

$$a_m\cdot a_n=\theta(m)\delta_{m+n,1}\,1,\quad a_m\cdot 1=a_m,$$
$$1\cdot a_m=a_m,\quad 1\cdot 1=1\quad (m,n\in\mathbb{Z}), \tag{2.1b}$$

where $\theta:\mathbb{Z}\to\mathbb{C}$ is defined to be unity if $m>0$ and zero if $m\leqslant 0$. The symbol 1 as well as the dot indicating operation in $\mathfrak{a}$ will in general be omitted. It is trivial to check that $\mathfrak{a}$ is an algebra which is not commutative and not associative; this last prop-

 0022-2488/88/020275-06$02.50

erty is crucial for what follows. We call $\mathfrak{a}$ the algebra of contractions. For any algebra $\mathfrak{a}$ the commutator of two elements $x,y\in\mathfrak{a}$ is defined by

$$[x,y]:=xy-yx. \tag{2.2}$$

As an expression of nonassociativity in $\mathfrak{a}$ it is convenient to define also the associator of three elements $x,y,z\in\mathfrak{a}$:

$$(x,y,z):=(xy)z-x(yz). \tag{2.3}$$

In the algebra defined by (2.1) we have, e.g., $[a_1,a_0]=1$ and $(a_1,a_0,a_1)=a_1$.

Then $\mathfrak{a}$ has the obvious grading $\mathfrak{a}=\mathfrak{a}_{\bar{0}}\oplus\mathfrak{a}_{\bar{1}}$, where $\mathfrak{a}_{\bar{0}}=\mathbb{C}1$ and $\mathfrak{a}_{\bar{1}}=\oplus_{m\in\mathbb{Z}}\mathbb{C}a_m$ $(\bar{0},\bar{1}\in\mathbb{Z}_2)$, and is simple although $\oplus_{m>0}\mathbb{C}a_m$ and $\oplus_{m<0}\mathbb{C}a_m$ are right and left ideals, respectively. Thus $\mathfrak{a}$ gives an example of a Lie-admissible algebra which is not associative. The multiplication table of $\mathfrak{a}^-$ is given by

$$[a_m,a_n]=(\theta(m)-\theta(1-m))\delta_{m+n,1},\quad [a_m,1]=0. \tag{2.4}$$

Although $\mathfrak{a}$ does not seem to provide an interesting example of a Lie-admissible algebra, its "affinization," when tensored with some associative algebra, to be defined in the next paragraphs, will provide us with an algebra isomorphic to $B_l^{(1)}$. The "affinization" of $\mathfrak{a}$ is done by defining the "generating function"

$$a(z):=\sum_{m\in\mathbb{Z}}a_m z^{-m},\quad a(z)\in\mathfrak{a}\{z\}, \tag{2.5}$$

where $\mathfrak{a}\{z\}$ is the vector space of formal Laurent series in the indeterminate z with coefficients in $\mathfrak{a}$.[13] In terms of $a(z)$ the multiplication table (2.1b) can be written as

$$a(z)a(w)=(z-w)^{-1},\quad |z|>|w|. \tag{2.6}$$

The nonassociativity in $\mathfrak{a}$ can be expressed by

$$(a(u),a(v),a(w))=(u-v)^{-1}a(w)+(w-v)^{-1}a(u)\qquad(|u|>|v|>|w|). \tag{2.7}$$

III. THE VERTEX OPERATOR REPRESENTATIONS

Let $\mathfrak{h}=\mathbb{C}p^1\oplus\cdots\oplus\mathbb{C}p^l$ be an l-dimensional complex vector space and $\mathfrak{h}^*=\mathbb{C}q^1\oplus\cdots\oplus\mathbb{C}q^l$ its dual, where $q^1,\ldots,q^l$ is the basis dual to $p^1,\ldots,p^l$. Let Λ be a lattice in $\mathfrak{h}_{\mathbb{R}}^*=\mathbb{R}q^1\oplus\cdots\oplus\mathbb{R}q^l$, and $\mathbb{C}(\Lambda)$ its group algebra. We represent a vector in any vector space as a pairing, denoted by a dot, between its components and corresponding basis elements, e.g., $\alpha^1q^1+\cdots+\alpha^lq^l=:\alpha\cdot q\in\mathfrak{h}^*$. Let us consider in $\mathfrak{h}$ an inner product, inherited by $\mathfrak{h}^*$ in the usual way, and denote it also by a dot. For convenience we take

$$p^i\cdot p^j=q^i\cdot q^j=\delta^{ij}\quad(i,j=1,\ldots,l).$$

We make $\mathfrak{h}$ an l-dimensional (Abelian) algebra of derivations on $\mathbb{C}(\Lambda)$ by defining

$$p^j|\gamma\rangle=\gamma^j|\gamma\rangle\quad[j=1,\ldots,l,\quad|\gamma\rangle\in\mathbb{C}(\Lambda)]. \tag{3.1}$$

We define also an action of the group Λ on $\mathbb{C}(\Lambda)$ by means of

$$e^{i\alpha\cdot q}|\gamma\rangle=|\alpha+\gamma\rangle\quad(\alpha,\gamma\in\Lambda), \tag{3.2a}$$

and so

$$e^{i\alpha\cdot q}e^{i\beta\cdot q}=e^{i(\alpha+\beta)\cdot q},\quad e^{i0\cdot q}=1\quad(\alpha,\beta\in\Lambda). \tag{3.2b}$$

Notice that $\mathbb{C}(\Lambda)$ can be generated from its vacuum by the action of Λ.

From (3.1) and (3.2) we get immediately

$$[p^i,e^{i\alpha\cdot q}]=\alpha^i e^{i\alpha\cdot q}. \tag{3.3}$$

It is sometimes convenient to use the (ill-defined) operators $q^1,\ldots,q^l$ and write

$$[q^i,q^j]=0=[p^i,p^j],\quad [q^i,p^j]=i\delta^{ij}. \tag{3.4}$$

The first two equations express commutativity in the group Λ and in the algebra $\mathfrak{h}$, respectively, and the third one implies (3.3).

Let us now consider a Heisenberg algebra

$$\mathrm{Heis}=\bigoplus_{m\in\mathbb{Z}\setminus\{0\}}\bigoplus_{i=1}^{l}\mathbb{C}H^i_m\oplus\mathbb{C}1,$$
$$[H^i_m,H^j_n]=\delta^{ij}m\delta_{m+n,0}\quad(i,j=1,\ldots,l,\quad m,n\in\mathbb{Z}\setminus\{0\})$$

commuting with q^i, p^i $(i=1,\ldots,l)$ and acting on $S(\mathrm{Heis}_<)$, the symmetric algebra of

$$\mathrm{Heis}_<:=\bigoplus_{m<0}\bigoplus_{i=1}^{l}\mathbb{C}H^i_m.$$

We write $H^i_0:=p^i$ and define

$$\hat{\mathfrak{h}}:=\mathfrak{h}\oplus\mathrm{Heis}=\bigoplus_{m\in\mathbb{Z}}\bigoplus_{i=1}^{l}\mathbb{C}H^i_m\oplus\mathbb{C}1,$$
$$[H^i_m,H^j_n]=\delta^{ij}m\delta_{m+n,0}\quad(i,j=1,\ldots,l,\quad m,n\in\mathbb{Z}), \tag{3.5}$$

with $\hat{\mathfrak{h}}$ acting on

$$F(\Lambda):=S(\mathrm{Heis}_<)\otimes\mathbb{C}(\Lambda). \tag{3.6}$$

We write $1\otimes|0\rangle\in F(\Lambda)$ simply as $|0\rangle$ and consider

$$\mathrm{Heis}_>:=\bigoplus_{m>0}\bigoplus_{i=1}^{l}\mathbb{C}H^i_m$$

killing all vectors in $\mathbb{C}(\Lambda)$.

Now we define the Fubini–Veneziano–Gervais fields from the theory of bosonic strings[14]:

$$Q^j(z):=q^j-ip^j\ln z+i\sum_{m\in\mathbb{Z}\setminus\{0\}}\frac{1}{m}H^j_m z^{-m}\qquad(z\in\mathbb{C}\setminus(-\infty;0]), \tag{3.7a}$$

$$P^j(z):=iz\frac{dQ^j(z)}{dz}=\sum_{m\in\mathbb{Z}}H^j_m z^{-m}\quad(z\in\mathbb{C}\setminus\{0\}). \tag{3.7b}$$

Here $Q^j(z)-q^j+ip^j\ln z$ and $P^j(z)$ are in the vector space $\hat{\mathfrak{h}}\{z\}$ of formal Laurent series in the indeterminate z with coefficients in $\hat{\mathfrak{h}}$, and $P^j(z)$ is just the "generating function" for $\hat{\mathfrak{h}}$. Finally we define the vertex operator associated to a point α of Λ:

$$U(\alpha,z):=z^{\alpha^2/2}:e^{i\alpha\cdot Q(z)}:\quad(\alpha\in\Lambda,\quad z\in\mathbb{C}\setminus(-\infty;0]), \tag{3.8}$$

where we have used the normal ordering

$$:q^jp^j:=:p^jq^j:=q^jp^j,$$
$$:H^j_mH^j_n:=\begin{cases}H^j_mH^j_n, & \text{if } n>0,\\ H^j_nH^j_m, & \text{if } n<0.\end{cases} \tag{3.9}$$

From (3.8) and (3.9) we obtain

$$U(\alpha,z)U(\beta,w) = (z-w)^{\alpha\cdot\beta} :U(\alpha,z)U(\beta,w):$$
$$(|z|>|w|). \tag{3.10}$$

To construct a representation of an affine Lie algebra $\hat{\mathfrak{g}}$ out of vertex operators[6] we take α in (3.8) to be in the root system Φ of the associated finite algebra $\mathfrak{g}$ ($\Phi\subset\Lambda$), and restrict ourselves to some $M\subset\Lambda$ in (3.6) for which $U(\alpha,z)$ can be extended analytically to $\mathbb{C}\setminus\{0\}$. From (3.8) it is easy to see that this requires

$$\tfrac{1}{2}\alpha^2 + \alpha\cdot\gamma\in\mathbb{Z}, \quad \text{if } \alpha\in\Phi \text{ and } \gamma\in M. \tag{3.11a}$$

For $F(M)$ to be a $\hat{\mathfrak{g}}$-module we need also that

$$\Phi + M\subset M. \tag{3.11b}$$

Since we have to avoid cuts in the z and w planes we must require also that [cf. Eq. (3.10)]

$$\alpha\cdot\beta\in\mathbb{Z} \quad (\alpha,\beta\in\Phi). \tag{3.11c}$$

A solution of (3.11) is obtained if Φ is the root system of a simply laced algebra $\mathfrak{g}$. By choosing $\alpha^2=2$ ($\alpha\in\Phi$), (3.11c) is automatically satisfied and we can take $M=\Lambda_w(\mathfrak{g})$, the weight lattice of $\mathfrak{g}$. With this solution at hand we can define the Del Giudice–Di Vechia–Fubini excitation operators from bosonic string theory[14]:

$$U^\alpha_m := \frac{1}{2\pi i}\oint_{(0)}\frac{dz}{z} z^m U(\alpha,z) \quad (\alpha\in\Phi,\ m\in\mathbb{Z}), \tag{3.12}$$

where the integration is along a simple loop around $z=0$. Using (3.10) we get

$$U^\alpha_m U^\beta_n - (-1)^{\alpha\cdot\beta} U^\beta_n U^\alpha_m = \frac{1}{2\pi i}\oint_{(0)}\frac{dw}{w} w^n \frac{1}{2\pi i}\oint_{(w)}\frac{dz}{z} z^m \times (z-w)^{\alpha\cdot\beta}:U(\alpha,z)U(\beta,w):, \tag{3.13}$$

where the w integration is around the origin and the z integration around $z=w$.

From the Schwarz inequality we get that $\alpha\cdot\beta = 2,1,0,-1,-2$ and consequently, that $(\alpha+\beta)^2 = 8,6,4,2,0$. As a result $\alpha+\beta\notin\Phi\cup\{0\}\Rightarrow\alpha\cdot\beta = 0,1,2$, $\alpha+\beta\in\Phi\Rightarrow\alpha\cdot\beta=-1$, and $\alpha+\beta=0\Rightarrow\alpha\cdot\beta=-2$. Using those cases in (3.13) we get

$$U^\alpha_m U^\beta_n - (-1)^{\alpha\cdot\beta} U^\beta_n U^\alpha_m = \begin{cases} 0, & \text{if } \alpha+\beta\notin\Phi\cup\{0\},\\ U^{\alpha+\beta}_{m+n}, & \text{if } \alpha+\beta\in\Phi,\\ \alpha\cdot H_{m+n} + m\delta_{m+n,0}, & \text{if } \alpha+\beta=0. \end{cases} \tag{3.14}$$

By eliminating the factor $(-1)^{\alpha\cdot\beta}$ it is possible to turn (3.14) into the commutation relations for the root vectors of an affine algebra. This method was discovered by Frenkel and Kac.[6] It is possible to define a projective representation of the root lattice of $\mathfrak{g}$, $\Lambda_r(\mathfrak{g})$, in some Hilbert space V_c:

$$c:\alpha\in\Lambda_r(\mathfrak{g})\mapsto c(\alpha)\in\mathrm{End}(V_c),$$
$$c(\alpha)c(\beta) = \epsilon(\alpha,\beta)c(\alpha+\beta), \quad c(0)=1, \tag{3.15a}$$

where the two-cocycle is given by

$$\epsilon:\Lambda_r(\mathfrak{g})\times\Lambda_r(\mathfrak{g})\to\{\pm 1\},$$
$$\epsilon(\alpha,\beta) = (-1)^{S(\alpha,\beta)}\epsilon(\beta,\alpha), \tag{3.15b}$$

and

$$S(\alpha,\beta)\equiv\alpha\cdot\beta\equiv -(\alpha\cdot\beta)^2 \pmod 2. \tag{3.15c}$$

Using the associativity in the group $\Lambda_r(\mathfrak{g})$ we obtain immediately

$$\epsilon(\alpha,\beta)\epsilon(\alpha+\beta,\gamma) = \epsilon(\alpha,\beta+\gamma)\epsilon(\beta,\gamma), \tag{3.15d}$$

for α,β,γ in $\Lambda_r(\mathfrak{g})$. It is possible to choose

$$\epsilon(\alpha,0)=\epsilon(0,\alpha)=1, \quad \epsilon(\alpha,-\alpha)=1 \quad [\alpha\in\Lambda_r(\mathfrak{g})]. \tag{3.15e}$$

For the simply laced case it is obvious that

$$\epsilon(\alpha,\beta)\epsilon(\beta,\alpha) = -1, \quad \text{if } \alpha+\beta\in\Phi. \tag{3.15f}$$

It is possible to take $V_c=\mathbb{C}(\Lambda)$ and then obtain the commutation relations between the cocycle c and the vertex operators; another possibility consists in taking V_c to be an independent space to be tensored with $F(M)$ and then getting a cocycle c which commutes with the vertex operators.[15] We are going to do the latter and define

$$E^\alpha(z) := U(\alpha,z)c(\alpha),$$
$$E^\alpha_m := U^\alpha_m c(\alpha) \quad (\alpha\in\Phi,\ m\in\mathbb{Z},\ z\in\mathbb{C}\setminus\{0\}) \tag{3.16}$$

acting on $F_c(M)=F(M)\otimes V_c$. From (3.14) and (3.16) we get the commutation relations for the operators E^α_m ($\alpha\in\Phi$; $m\in\mathbb{Z}$). The commutation relations for the operators H^i_m ($i=1,\ldots,l$, $m\in\mathbb{Z}$) are known from the beginning [Eq. (3.5)] and those for H^i_m and E^α_n ($i=1,\ldots,l$, $\alpha\in\Phi$, $m,n\in\mathbb{Z}$) are trivial to obtain [by using Eq. (3.3)]. The complete algebra is given by

$$\hat{\mathfrak{g}} := \hat{\mathfrak{h}}\oplus\bigoplus_{m\in\mathbb{Z}}\bigoplus_{\alpha\in\Phi}\mathbb{C}E^\alpha_m, \tag{3.17a}$$

with

$$[H^i_m,H^j_n] = \delta^{ij}m\delta_{m+n,0},$$
$$[H^i_m,E^\alpha_n] = \alpha^i E^\alpha_{m+n}, \tag{3.17b}$$
$$[E^\alpha_m,E^\beta_n] = \begin{cases} 0, & \text{if } \alpha+\beta\notin\Phi\cup\{0\},\\ \epsilon(\alpha,\beta)E^{\alpha+\beta}_{m+n}, & \text{if } \alpha+\beta\in\Phi,\\ \alpha\cdot H_{m+n} + m\delta_{m+n,0}, & \text{if } \alpha+\beta=0. \end{cases}$$

This should be recognized as a nontwisted affine simply laced Kac–Moody algebra (without the degree derivation, to be precise).

Now, if we consider the root system Φ to be $\Phi(B_l)$, the root system of the non-simply-laced Lie algebra B_l (with long roots having square length 2), we still have (3.11c) satisfied, and (3.11b) can be trivially satisfied taking $M=\Lambda_r(B_l) := \mathrm{Span}_{\mathbb{Z}}\Phi$. The condition (3.11a), however, will be satisfied only if α is a long root of B_l. Despite that we can still associate excitation operators to the short roots of B_l by defining them as the Laurent coefficients of $z^{1/2}U(\alpha,z)$ [$\alpha\in\Phi(B_l)$, $\alpha^2=1$]. Let us call Φ_L and Φ_S the set of long and short roots of B_l, respectively; we have

$$\Phi(B_l) = \Phi_S\cup\Phi_L. \tag{3.18}$$

We define the excitation operators for $\Phi(B_l)$ as

$$U^\alpha_m := \frac{1}{2\pi i}\oint_{(0)}\frac{dz}{z} z^m U(\alpha,z) \quad (m\in\mathbb{Z}), \quad \text{if } \alpha\in\Phi_L, \tag{3.19a}$$

and

$$V_s^\rho = \frac{1}{2\pi i}\oint_{(0)} \frac{dz}{z} z^s U(\rho,z) \quad (s\in\mathbf{Z}+\tfrac{1}{2}), \quad \text{if } \rho\in\Phi_S . \tag{3.19b}$$

Since Φ_L is the root system of a simply laced Lie algebra we can use (3.19a) to get again (3.17b) with the understanding that in (3.17b) and (3.15f) Φ is replaced by Φ_L, and in (3.15a)–(3.15e) $\Lambda_r(\mathfrak{g})$ is replaced by $\mathrm{Span}_{\mathbf{Z}}\,\Phi_L$. The equation analogous to (3.13) involving excitation operators for short roots will result in the following relations, to be compared with Eq. (3.14):

$$V_s^\rho U_m^\alpha - (-1)^{\rho\cdot\alpha} U_m^\alpha V_s^\rho = \begin{cases} 0, & \text{if } \rho+\alpha\notin\Phi(B_l)\cup\{0\}, \\ V_{s+m}^{\rho+\alpha}, & \text{if } \rho+\alpha\in\Phi(B_l); \end{cases} \tag{3.20}$$

$$V_s^\rho V_{s'}^\sigma - (-1)^{\rho\cdot\sigma} V_{s'}^\sigma V_s^\rho = \begin{cases} 0, & \text{if } \rho\cdot\sigma = 1,0, \\ \delta_{s+s',0}, & \text{if } \rho\cdot\sigma = -1; \end{cases} \tag{3.21}$$

where $m\in\mathbf{Z}$ and $s,s'\in\mathbf{Z}+\frac{1}{2}$.

To obtain (3.20) we notice that $|\rho\cdot\alpha|\leqslant|\rho|\cdot|\alpha| = \sqrt{2}$ and that $\Lambda_r(\Phi(B_l))$ is an integral lattice; then it follows that $\rho\cdot\alpha = 1,0,-1$ and that $(\rho+\alpha)^2 = 5,3,1$. Consequently

$$(\rho+\alpha)\notin\Phi(B_l)\cup\{0\} \Rightarrow (\rho+\alpha)^2 = 5,3 \Rightarrow \rho\cdot\alpha = 1,0$$

and

$$(\rho+\alpha)\in\Phi(B_l) \Rightarrow (\rho+\alpha)^2 = 1 \Rightarrow \rho\cdot\alpha = -1.$$

Concerning (3.21) it is obvious that no correction of sign can turn it into the commutation relations among the short root vectors of $B_l^{(1)}$. However, (3.15a)–(3.15e) can be extended[15] from the lattice $\mathrm{Span}_{\mathbf{Z}}\,\Phi_L$ to the whole $\Lambda_r(\Phi(B_l))$ in such a way that all the equations (3.15a)–(3.15f) remain valid, except for (3.15c), which is modified for $\alpha,\beta\in\Phi_S$:

$$S(\alpha,\beta)\equiv\alpha^2\beta^2 - (\alpha\cdot\beta)^2 \pmod 2. \tag{3.22}$$

Using this fact and the definition

$$\psi_s^\rho := V_s^\rho c(\rho) \quad (\rho\in\Phi_S, \quad s\in\mathbf{Z}+\tfrac{1}{2}), \tag{3.23}$$

(3.21) can be made a Clifford algebra $\mathbb{C}(\psi)$ written in terms of generators ψ_s^ρ ($\rho\in\Phi_S$, $s\in\mathbf{Z}+\frac{1}{2}$). The whole algebra obtained from the vertex operators is given by

$$[H_m^i, H_n^j] = \delta^{ij} m\delta_{m+n,0},$$

$$[H_m^i, E_n^\alpha] = \alpha^i E_{m+n}^\alpha, \quad [H_m^i, \psi_S^\rho] = \rho^i \psi_{m+s}^\rho,$$

$$[E_m^\alpha, E_n^\beta] = \begin{cases} 0, & \text{if } \alpha+\beta\notin\Phi(B_l)\cup\{0\}, \\ \epsilon(\alpha,\beta)E_{m+n}^{\alpha+\beta}, & \text{if } \alpha+\beta\in\Phi(B_l), \\ \alpha\cdot H_{m+n} + m\delta_{m+n,0}, & \text{if } \alpha+\beta=0, \end{cases} \tag{3.24}$$

$$\{\psi_s^\rho, \psi_{s'}^\sigma\} = \delta^{\rho+\sigma,0}\delta_{s+s',0},$$

$$[\psi_s^\rho, E_n^\alpha] = \begin{cases} 0, & \text{if } \rho+\alpha\notin\Phi(B_l)\cup\{0\}, \\ \epsilon(\rho,\alpha)\psi_{s+n}^{\rho+\alpha}, & \text{if } \rho+\alpha\in\Phi(B_l), \end{cases}$$

where $i,j = 1,\ldots,l$, $m,n\in\mathbf{Z}$, $s,s'\in\mathbf{Z}+\frac{1}{2}$, $\alpha,\beta\in\Phi_L$, $\rho,\sigma\in\Phi_S$. Therefore, in this case, instead of obtaining a representation of $B_l^{(1)}$ we obtain a representation of a superalgebra.[15]

IV. A NONASSOCIATIVE REALIZATION OF $B_l^{(1)}$

The operators E_m^α and $\Psi_m^\rho := \psi_{m+1/2}^\rho$ ($\alpha\in\Phi_L$, $\rho\in\Phi_S$, $m\in\mathbf{Z}$) were defined in terms of the operators H_m^i ($i = 1,\ldots,l$, $m\in\mathbf{Z}$) in the Heisenberg system (3.5) and the operators $e^{i\alpha\cdot q}$ ($\alpha\in\Phi$) in the representation (3.2) of the group Λ. We can think of all of them as generators of an algebra A of endomorphisms of $F_c(M)$. This algebra can be viewed more abstractly as the free associative algebra generated by $e^{i\alpha\cdot q}, H_m^j, E_m^\alpha, \Psi_m^\rho$ in which some elements are then identified according to (3.2b), (3.10), and (3.24), i.e., the free associative algebra modulo the ideal generated by the elements $e^{i\alpha\cdot q}e^{i\beta\cdot q} - e^{i(\alpha+\beta)\cdot q}$, $[H_m^i, H_n^j] - \delta^{ij}\delta_{m+n,0}$, etc. We take the affinization of A and in the direct product of $A\{z\}$ by $\mathfrak{a}\{z\}$ we take the elements $H^i(z)\otimes 1, E^\alpha(z)\otimes 1$, and $\Psi^\rho(z)\otimes a(z)$ ($i = 1,\ldots,l, \alpha\in\Phi_L, \rho\in\Phi_S$), in terms of which we define

$$h_m^i = \frac{1}{2\pi i}\oint_{(0)}\frac{dz}{z} z^m H^i(z)\otimes 1,$$

$$e_m^\alpha = \frac{1}{2\pi i}\oint_{(0)}\frac{dz}{z} z^m E^\alpha(z)\otimes 1, \tag{4.1}$$

$$e_m^\rho = \frac{1}{2\pi i}\oint_{(0)}\frac{dz}{z} z^m \Psi^\rho(z)\otimes a(z),$$

and write 1 instead of $1\otimes 1$ whenever it is convenient. Let us call $\mathfrak{b}$ the resulting algebra with generators (4.1); we have that $\mathfrak{b}\{z\}\subset A\{z\}\otimes\mathfrak{a}\{z\}$. Now we want to show that $B_l^{(1)}\subset\mathfrak{b}^-$. This will provide us with an example of a Lie algebra $B_l^{(1)}$ constructed from elements of the underlying not associative algebra $\mathfrak{b}$.

From the definition (4.1) and the basic identity (3.10) we obtain

$$[e_m^\rho, e_n^\sigma] = \epsilon(\rho,\sigma)\frac{1}{2\pi i}\oint_{(0)}\frac{dw}{w} w^n \frac{1}{2\pi i}\oint_{(w)}\frac{dz}{z} z^m \times (z-w)^{\rho\cdot\sigma} :ze^{i\rho\cdot Q(z)} we^{i\sigma\cdot Q(w)}: c(\rho+\sigma)\otimes 1 \quad (\rho,\sigma\in\Phi_S, \quad m,n\in\mathbf{Z}). \tag{4.2}$$

For $\rho+\sigma\notin\Phi\cup\{0\}$ we have $\rho\cdot\sigma = 4$; for $\rho+\sigma\in\Phi$ ($\Leftrightarrow\rho+\sigma\in\Phi_L$) we have $\rho\cdot\sigma = 0$, and for $\rho+\sigma = 0$ we have $\rho\cdot\sigma = -1$. In those cases (4.2) reduces to

$$[e_m^\rho, e_n^\sigma] = \begin{cases} 0, & \text{if } \rho+\sigma\notin\Phi\cup\{0\}, \\ \epsilon(\rho,\sigma)e_{m+n}^{\rho+\sigma}, & \text{if } \rho+\sigma\in\Phi, \\ \rho\cdot h_{m+n} + m\delta_{m+n,0}, & \text{if } \rho+\sigma=0 \end{cases} \quad (\rho,\sigma\in\Phi_S, \quad m,n\in\mathbf{Z}). \tag{4.3}$$

It is easy to get from the rest of the equations in (3.24) all the other commutation relations necessary to obtain the algebra

$$\hat{\mathfrak{g}} = \bigoplus_{m\in\mathbf{Z}}\left(\bigoplus_{i=1}^{l}\mathbb{C}h_m^i \oplus \bigoplus_{\alpha\in\Phi}\mathbb{C}e_m^\alpha\right)\oplus\mathbb{C}1, \tag{4.4a}$$

with

$$[h^i_m,h^j_n]=\delta^{ij}m\delta_{m+n,0},$$
$$[h^i_m,e^\alpha_n]=\alpha^i e^\alpha_{m+n},$$
$$[e^\alpha_m,e^\beta_n]=\begin{cases}0, & \text{if } \alpha+\beta\notin\Phi\cup\{0\},\\ \epsilon(\alpha,\beta)\,e^{\alpha+\beta}_{m+n}, & \text{if } \alpha+\beta\in\Phi,\\ \alpha\cdot h_{m+n}+m\delta_{m+n,0}, & \text{if } \alpha+\beta=0\end{cases} \tag{4.4b}$$
$$[i,j=1,\ldots,l,\quad \alpha,\beta\in\Phi,\quad m,n\in\mathbb{Z},\quad \Phi=\Phi(B_l)]\,.$$

Those equations define a Lie algebra $[B_l^{(1)}$ because $\Phi=\Phi(B_l)]$ if we can obtain from them the Jacobi identity for any three elements of $\hat{\mathfrak{g}}$. For algebras like (3.17) the Jacobi identity is an immediate consequence of the associativity of the underlying algebra of operators. However, the underlying algebra that gives (4.4) is not associative:

$$(e^\rho_m,e^\sigma_n,e^\tau_p)=\frac{1}{2\pi i}\oint\frac{du}{u}u^m\oint\frac{dv}{v}v^n\oint\frac{dw}{w}w^p \times\Psi^\rho(z)\Psi^\sigma(v)\Psi^\tau(w)\otimes(a(u),a(v),a(w)),$$

which is not identically zero. But once the underlying algebra is not associative the Jacobi identity may be violated and its possible validity depends crucially on the properties of the structure constants of $\hat{\mathfrak{g}}$. Essentially, what is required to prove the Jacobi identity is that the two-cocycle ϵ in (4.4b) satisfy the properties given in (3.15). Since that is the case we have that $\hat{\mathfrak{g}}$ in (4.4) is indeed a Lie algebra: the affine nontwisted Kac–Moody algebra $B_l^{(1)}$. The proof of the validity of the Jacobi identity is by enumeration of cases and therefore not very illuminating; a concise and elegant presentation is given by Mitzman.[16]

V. THE RELATION BETWEEN $\mathfrak{a}$ AND THE REPRESENTATIONS OF $B_l^{(1)}$

A method of generalizing the vertex operator formalism, to obtain also a representation for the non-simply-laced Lie algebra $B_l^{(1)}$, consists in replacing (3.19b) by

$$U^\rho_m:=\frac{1}{2\pi i}\oint\frac{dz}{z}z^m z^{1/2}U(\rho,z)U(z), \tag{5.1}$$

where $U(z)$ is an auxiliary field, to be regarded as an endomorphism of some vector space, and whose basic properties are to commute with $U(\alpha,z)$ $[\alpha\in\Phi(B_l)]$, to be analytic in $\mathbb{C}\setminus\{0\}$, and to satisfy the relation

$$U(z)U(w)=(z-w)^{-1}f(z,w)+F(z,w)\quad(|z|>|w|), \tag{5.2}$$

where f is a complex function, symmetric, and equal to unity for $z=w$, and F is an operator function, antisymmetric, zero at $z=w$, and analytic in $\mathbb{C}\setminus\{0\}\times\mathbb{C}\setminus\{0\}$. With this ansatz the problem of finding a representation for $B_l^{(1)}$ reduces to the problem of finding the operator $U(z)$. Solutions for this problem, and its generalization to other non-simply-laced algebras, were given before by several authors, with different approaches.[17–20] A possible solution consists in taking $U(z)=z^{1/2}\Psi(z)$, where

$$\Psi(z)=\sum_{m\in\mathbb{Z}+1/2}\Psi_m z^{-m} \tag{5.3a}$$

is the "generating function" for a Clifford algebra $\mathbb{C}(\Psi)$ with

$$\Psi_m\Psi_n+\Psi_n\Psi_m=\delta_{m+n,0}\quad(m,n\in\mathbb{Z}+\tfrac{1}{2})\,. \tag{5.3b}$$

Defining the normal ordering

$$:\Psi_m\Psi_n:=\begin{cases}\Psi_m\Psi_n, & \text{if } n>0,\\ -\Psi_n\Psi_m, & \text{if } n<0,\end{cases} \tag{5.3c}$$

it is easy to obtain the following operator product expansion:

$$\Psi(z)\Psi(w)=\frac{z^{1/2}w^{1/2}}{z-w}+:\Psi(z)\Psi(w):\quad(|z|>|w|). \tag{5.4}$$

Comparing (5.4) with (5.2) we see that $U(z)=z^{-1/2}\Psi(z)$ is indeed a solution of (5.2). Since any Clifford algebra can be realized as endomorphisms of a vector space, this solution provides a representation of $B_l^{(1)}$. A choice that cannot be made in (5.2) is $F(z,w)=0$, because in this case $U(z)$ will inevitably be an element of a not associative algebra, as was shown in Sec. II, and that makes it impossible for U^ρ_m to be used in the vertex operator construction of a representation of a Lie algebra. Of course, it was precisely this choice that provided the example of $B_l^{(1)}$ being constructed from the not associative algebra $\mathfrak{b}$. By comparing the choice (2.6) with (5.4) we want to obtain more information about the not associative algebra treated in Sec. II.

A representation π of a nonassociative algebra A by means of an algebra B is defined by a linear transformation $\pi: A\to B$, such that

$$\pi(a)*\pi(a')=\pi(a\cdot a')\quad(a,a'\in A),$$

where the operation in A is denoted by a dot and $*: B\times B\to B$ is a bilinear function. In case A is a Lie algebra then B is $\mathrm{End}(V)$ for some vector space V and $*$ is fixed to be

$$\pi(a)*\pi(a'):=\pi(a)\times\pi(a')-\pi(a')\times\pi(a),$$

where the composition of endomorphisms is denoted by a $\times$ for clarity. With this definition in mind it is quite straightforward to define a representation of $\mathfrak{a}$ by means of $\mathbb{C}(\Psi)$. We define

$$\pi:\mathfrak{a}\to\mathbb{C}(\Psi) \tag{5.5a}$$

to be the linear extension of

$$\pi(a_m)=\Psi_{m-1/2}\quad(m\in\mathbb{Z}),\quad \pi(1)=1, \tag{5.5b}$$

and require $*$ to be such that $\pi(a(z))*\pi(a(w))=(z-w)^{-1}$. By using (5.4) it is straightforward to see that $*$ should be defined as the linear extension of

$$1*1:=1,\quad 1*\Psi_m=\Psi_m*1=\Psi_m,$$
$$\Psi_m*\Psi_n:=\overbrace{\Psi_m\Psi_n}\quad(m,n\in\mathbb{Z}+\tfrac{1}{2}), \tag{5.5c}$$

where we have used the symbol of contraction from quantum field theory:

$$\overbrace{\Psi_m\Psi_n}=\Psi_m\Psi_n-:\Psi_m\Psi_n:. \tag{5.5d}$$

The formulas (5.5) justify the name given to $\mathfrak{a}$ as the algebra of contractions.

Another Clifford algebra $\mathbb{C}(\tilde{\Psi})$, which is used to construct representations of $B_l^{(1)}$, is given by

$$\tilde{\Psi}_m\tilde{\Psi}_n+\tilde{\Psi}_n\tilde{\Psi}_m=2\delta_{m+n,0}\quad(m,n\in\mathbb{Z}).$$

Defining

$$\tilde{\Psi}(z) = \sum_{m\in\mathbb{Z}} \tilde{\Psi}_m z^{-m}$$

and the normal ordering

$$:\tilde{\Psi}_m\tilde{\Psi}_n: = \begin{cases} \tilde{\Psi}_m\tilde{\Psi}_n, & \text{if } n>0, \\ \frac{1}{2}(\tilde{\Psi}_m\tilde{\Psi}_0 - \tilde{\Psi}_0\tilde{\Psi}_m), & \text{if } n=0, \\ -\tilde{\Psi}_n\tilde{\Psi}_m, & \text{if } n<0, \end{cases}$$

we obtain

$$\tilde{\Psi}(z)\tilde{\Psi}(w) = (z+w)(z-w)^{-1} + :\tilde{\Psi}(z)\tilde{\Psi}(w): \quad (|z|>|w|).$$

Using the same method followed to obtain (5.5) we find the representation of $\mathfrak{a}$ by means of $\mathbb{C}(\tilde{\Psi})$:

$$\tilde{\pi}: \mathfrak{a}\to\mathbb{C}(\tilde{\Psi}); \tag{5.6a}$$

$$\tilde{\pi}(a_m) = \tilde{\Psi}_m \quad (m\in\mathbb{Z}), \quad \tilde{\pi}(1) = 1; \tag{5.6b}$$

$$1*1 := 1, \quad 1*\tilde{\Psi}_m = \tilde{\Psi}_m*1 = \tilde{\Psi}_m,$$
$$\tilde{\Psi}_m*\tilde{\Psi}_n := \sum_{0<k\in\mathbb{Z}} (-1)^k \overbrace{\tilde{\Psi}_{m-k-1}\tilde{\Psi}_{n+k}}, \tag{5.6c}$$

where the series in the last expression contains only a finite number of nonzero terms.

From those two examples we can see that $\mathfrak{a}$ can be related to all the representations of $B_l^{(1)}$ obtained from the ansatz (5.1). This $\mathfrak{a}$ expresses algebraically the essential common features of all the auxiliary fields used in (5.1) to extend the vertex operator formalism to the non-simply-laced algebras.

It seems that the construction of Sec. IV can be generalized to other non-simply-laced algebras, by attaching a not associative algebra to each distinct orbit of the Weyl group generated by the long roots reflections acting on each short root.[19]

ACKNOWLEDGMENTS

I wish to thank Professor S. Okubo for his guiding remarks that were essential for the completion of this work. I also wish to thank Professor A. Feingold for explaining to me the correct meaning of some of the mathematical concepts used above.

This work was supported through a scholarship from CNPq (National Research Council of Brazil) and by the U. S. Department of Energy Contract No. DE-AC02-76ER13065.

[1] R. D. Schafer, *An Introduction to Nonassociative Algebras* (Academic, New York, 1966).
[2] For a short review, see H. C. Myung, Hadronic J. **1**, 169 (1978).
[3] A. A. Albert, Trans. Am. Math. Soc. **64**, 552 (1948).
[4] V. G. Kac, *Infinite Dimensional Lie Algebras* (Cambridge U. P., Cambridge, 1985).
[5] J. Lepowsky and R. L. Wilson, Commun. Math. Phys. **62**, 43 (1978).
[6] B. Frenkel, II and V. G. Kac, Invent. Math. **62**, 23 (1980).
[7] B. Frenkel, II, J. Funct. Anal. **44**, 259 (1981).
[8] G. Segal, Commun. Math. Phys. **80**, 301 (1981).
[9] R. Jackiw, "Quantum mechanical symmetry breaking," in *Recent Developments in Quantum Field Theory*, edited by J. Ambjorn, B. Durhuus, and J. Petersen (North-Holland, Amsterdam, 1985).
[10] E. Witten, Nucl. Phys. B **268**, 253 (1986).
[11] A. Strominger, The Institute for Advanced Study, Preprint HEP 87/16.
[12] G. T. Horowitz and A. Strominger, Phys. Lett. B **185**, 45 (1987).
[13] For the formalism of $\mathfrak{a}\{z\}$ see J. Lepowsky and R. L. Wilson, Invent. Math. **77**, 199 (1984).
[14] *Dual Theory*, Physics Reports Reprint Vol. I, edited by M. Jacob (North-Holland, Amsterdam, 1974).
[15] P. Goddard and D. Olive, "Algebras, lattices and strings," in *Vertex Operators in Mathematics and Physics*, Mathematical Sciences Research Institute Publications, Vol. 3 (Springer, New York, 1984), pp. 51–96.
[16] D. Mitzman, *Contemporary Mathematics*, Vol. 40 (Am. Math. Soc., Providence, RI, 1985).
[17] J. Lepowsky and M. Primc, "Structure of the standard modules for the affine Lie algebra $A_1^{(1)}$ in the homogeneous picture," in *Vertex Operators in Mathematics and Physics*, Mathematical Sciences Research Institute Publications, Vol. 3 (Springer, New York, 1984), pp. 143–162.
[18] A. J. Feingold and B. Frenkel, II, Adv. Math. **56**, 117 (1985).
[19] P. Goddard, W. Nahm, D. Olive, and A. Schwimmer, Commun. Math. Phys. **107**, 179 (1986).
[20] D. Bernard and J. Thierry-Mieg, Preprint GAR Observatoire de Paris-Meudon, August 1986.

Loop algebras and superalgebras based on S^7

F. Englert
Université Libre de Bruxelles, Pool de Physique, CP 225, B-1050 Brussels, Belgium

A. Sevrin,[a] W. Troost,[b] and A. Van Proeyen[b]
Instituut voor Theoretische Fysica, Universiteit Leuven, B-3030 Leuven, Belgium

Ph. Spindel
Faculté des Sciences, Université de l'Etat à Mons, B-7000 Mons, Belgium

(Received 3 June 1987; accepted for publication 15 July 1987)

Loop algebras of parallel transformations are defined on seven spheres. These algebras can be extended with eight fermionic charges and the Virasoro algebra to a soft superalgebra using the geometry of cosets of supergroups, or to a nonassociative infinite-dimensional $N = 8$, $d = 2$ superconformal algebra.

I. INTRODUCTION

In recent years infinite-dimensional algebras have played an ever increasing role in physics, especially in statistical physics and in string theories (a review is given in Ref. 1). The simplest infinite-dimensional algebras are the algebras of loop groups defined by maps from the circle into a group manifold. Composition of these maps is done by using the group multiplication. Since the group multiplication corresponds to an absolute parallelism (see Appendix B), this composition can also be interpreted in the following way: one composes two loops in group space by means of an absolute parallelism. This notion of parallel transport allows an extension of the loop composition on a different space, the seven-sphere S^7. Indeed, S^7 is the only compact Riemannian manifold which shares with Lie groups the property of absolute parallelism.[2] We will use this property to define two new loop algebras. One of these will be a nonassociative algebra. The other will be the "soft" algebra of covariant derivatives. By soft algebra[3] we mean that it has structure functions depending on a point of the seven-sphere rather than structure constants. If one takes the values of the structure functions at one point of the sphere, one reobtains the nonassociative algebras.

These considerations can also be applied to superalgebras. This will allow us to obtain an infinite-dimensional superalgebra for $N = 8$ in 2 dimensions, generalizing the known infinite superalgebras for $N = 1$, 2, and 4.[4]

II. SOFT ALGEBRAS ON S^7

In this section we show how algebras defined on group manifolds can be generalized to the seven-sphere, using its parallelizability. We obtain thereby a soft algebra, as has been remarked previously in Ref. 5. Then we will get the same algebra by using more specifically the coset space structure of the seven-sphere: SO(8)/SO(7), and we show how a loop algebra can be defined.

The seven-sphere can be parametrized by unit octonions. There are two seven-parameter families of parallelisms depending on a unit octonion A. We will use here only "left parallelisms."[6] The parallelisms and corresponding composition laws (see Appendix A for notations) satisfying the criteria of Appendix B are

$$\begin{aligned}&(X,Y)\|_A(X',Y') \quad \text{if } Y(\bar{X}A) = Y'(\bar{X}'A),\\ &X *_A Y = (XA)(\bar{A}Y) = XY + A[X,A,Y].\end{aligned} \tag{1}$$

We write the algebra for the parallelism defined by $A = 1$. We parametrize all points of the seven-sphere as $e^{t\eta}$, where t is a real number and η a unit purely imaginary octonion ($\eta = -\bar{\eta}$). The couple starting from X, and parallel to $(1,e^{t\eta})$, is given by $(X,e^{t\eta}X)$. Henceforth, the end point obtained in this way we will call the transport of X by $e^{t\eta}$,

$$e^{t\eta}X = (\cos t)X + (\sin t)\eta X. \tag{2}$$

We define therefore the generator K_η of transport in η direction acting on the point X as

$$K_\eta X = \eta X. \tag{3}$$

We define the composition of parallel transformations by performing the two parallel transports successively. So this implies

$$\begin{aligned}[K_\eta,K_\lambda]X &= \eta(\lambda X) - \lambda(\eta X)\\ &= ([\eta,\lambda])X - 2[\eta,\lambda,X].\end{aligned} \tag{4}$$

The associator in Eq. (4) states that the result of the operation is not a uniform parallel transport. It can be viewed as a point-dependent parallel transport, by rewriting the commutator as

$$[K_\eta,K_\lambda]X = K_{\rho(X)}X \tag{5}$$

with

$$\rho(X) = [\eta,\lambda] - 2X[\eta,\lambda,X]. \tag{6}$$

Define the imaginary units e^a and $e^0 = 1$, then

$$e^i e^j = \Delta^{ijk}e^k. \tag{7}$$

We denote by K^a the generator of transports in the $(-e^a)$ direction. In this basis, the commutator is

[a] Onderzoeker I.I.K.W., Belgium.
[b] Bevoegdverklaard Navorser N.F.W.O., Belgium.

 0022-2488/88/020281-06$02.50

$$[K^a,K^b]X = e^a(e^bX) - e^b(e^aX)$$
$$= -(2\Delta^{abc} - (2X[e^a,e^b,X])^c)K^cX$$
$$= 2S^{abc}(X)K^cX. \quad (8)$$

As we will see in a moment S is the torsion of the seven-sphere corresponding to the parallel transport.[6] The Jacobi identity is trivially satisfied as the composition is realized by successive transformations. Nevertheless it is instructive to see what happens explicitly. In fact, what we have here is a *soft algebra*.[3] Equation (5) defines an algebra where the action of the generators depends on extra variables, which are in our case points of the seven-sphere. The structure constants depend on these points, too ("structure functions"). The action of the generators also transforms these structure functions such that the Jacobi identities are preserved. Consider what happens if we calculate

$$[\exp(te^c),[K^a,K^b]]X, \quad (9)$$

where t is small. We obtain then

$$\exp(te^c)(-2S^{abd}(X)K^dX)$$
$$-(-2S^{abd}(\exp(te^c)X))K^d(\exp(te^c)X), \quad (10)$$

which to first order in t is equal to

$$t(4S^{abd}(X)S^{cde}(X)K^eX + 2\partial^cS^{abd}(X)K^dX). \quad (11)$$

The Jacobi identity states that the antisymmetric part in $[abc]$ is zero, recovering the known differential equation for the torsion.[2] This mechanism can also be seen in the *algebra of covariant derivatives* on the seven-sphere. Choosing a sevenbein $u^{a\alpha}$, where α indices label the coordinates on the sphere, and a labels the coordinates in the tangent space, we have on world-scalar functions (no α-type indices)

$$\nabla^a = u^{a\alpha}\partial_\alpha. \quad (12)$$

We obtain then for the commutator

$$[\nabla^a,\nabla^b] = 2S^{abc}(X)\nabla_c. \quad (13)$$

This equation remains valid on arbitrary tensors, if the sevenbein u_a is covariantly constant. Then $S^{abc}(X)$ are the components of the torsion tensor. Choosing for the sevenbein u^a at the point X the vectors e^aX, we have

$$S^{abc}(X)u^{c\alpha} = u^{[a\beta}\partial_\beta u^{b]\alpha}$$
$$= (e^{[a}X)^\beta\partial_\beta(e^{b]}X)^\alpha = (e^{[b}(e^{a]}X))^\alpha. \quad (14)$$

Comparing with Eq. (8) justifies the use of the name "torsion" for the quantity found there. As calculated in Ref. 6 the covariant derivative of the torsion is

$$\nabla^dS^{abc}(X) = \langle[e^a,e^bX,e^dX] + [e^a,e^b(e^dX),X],e^c\rangle, \quad (15)$$

where $\langle\ ,\ \rangle$ denotes the positive definite scalar product. At the points 1 and -1,

$$S^{abc} = -\Delta^{abc},$$
$$\nabla^dS^{abc} = -2\phi^{abcd}. \quad (16)$$

For the Jacobi identity we act with ∇^c on Eq. (13), and the second term of Eq. (11) now corresponds to Eq. (15). The Jacobi identity then corresponds to Eq. (A4).

In Eq. (8), we could interpret the action $S^{abc}(X)K^c$ as 21 new generators. Then together with the original seven K^a they combine to the SO(8) group. We will explain how the soft algebra can be seen to originate from this *coset structure*, and this we will apply in Sec. IV to the supersymmetric case.

Let $K^{A\mu}\partial_\mu$ represent the generators of a group G in a coset space G/H. Now we suppose that the subset $K^{a\mu}\partial_\mu$, where a runs over $\dim(G/H)$ values, forms a basis for the action of the generators on the coset space. As a consequence, they generate regular motions without fixed points. This implies that the manifold is parallelizable: parallel vectors can then be defined as those having the same components in the frames defined by $K^{a\mu}$. With respect to the metric

$$g_{\mu\nu} = K^{A\mu}K^{B\nu}g_{AB}, \quad (17)$$

where g_{AB} is the Killing metric, this becomes an absolute parallelism. Therefore—in the framework of positive definite metrics—we are led to coset spaces isomorphic with group manifolds, and to the seven-sphere.[2]

For use in Sec. IV we will construct these generators explicitly for SO(8)/SO(7). First define the SO(8) generators ($I = 0,1,...,7$)

$$M_{IJ} = x_I\partial_J - x_J\partial_I,$$
$$[M_{IJ},M_{KL}] = \delta_{IL}M_{JK} + \delta_{JK}M_{IL} \quad (18)$$
$$-\delta_{JL}M_{IK} - \delta_{IK}M_{JL}.$$

One would be tempted to take here the straightforward choice M_{I0} for the generators K^a, but these become zero at some points of the sphere. Actually we need nonvanishing vector fields, e.g., the seven vector fields associated to the generators K^a defined previously. To this aim we first define the Γ matrices by identifying the octonionic multiplication with

$$X \mapsto e_IX \quad (19)$$

with a linear transformation

$$X = x_ie^i \mapsto e_IX = (\Gamma_I^{ij}x_i)e_j, \quad \Gamma_I^{ij} = \Delta^{Iij}. \quad (20)$$

Obviously Γ_0 is the identity matrix, while the seven Γ_A obey the Clifford algebra

$$\Gamma_A\Gamma_B + \Gamma_B\Gamma_A = -2\delta_{AB}, \quad A,B = 1,...,7, \quad (21)$$

thanks to the complete antisymmetry of the associator. From them we define Γ_{IJ}

$$\Gamma_{A0} = -\Gamma_{0A} = \Gamma_A, \quad \Gamma_{AB} = \Gamma_{[A}\Gamma_{B]}. \quad (22)$$

We indicate further the spinor indices of the Γ matrices by $i,j,...$. These Γ matrices are antisymmetric in these indices which run also over eight values. Due to triality of SO(8), the set Γ_{IJ}^{a0} with $a = 1,...,7$ forms again a seven-dimensional Clifford algebra where now the $[IJ]$ are spinor indices. We define

$$M^{ij} \equiv \tfrac{1}{4}\Gamma_{IJ}^{ij}M_{JI} \rightsquigarrow M_{IJ} = \tfrac{1}{4}\Gamma_{IJ}^{ij}M^{ji}, \quad K^a \equiv -2M^{a0}. \quad (23)$$

The M^{ij} satisfy in any point the SO(8) algebra as the M_{IJ} equation (18). In contrast to the basis M_{A0}, the seven generators K^a vanish nowhere on S^7. The M^{ab} can be expressed in terms of the K^a:

$$M^{ab} = -S^{abc}(X)K^c/2. \tag{24}$$

Therefore the SO(8) algebra implies

$$\begin{aligned}[K^a,K^b] &= 4[M^{a0},M^{b0}] \\ &= -4M^{ab} = 2S^{abc}(X)K^c.\end{aligned} \tag{25}$$

So this gives the same result as what we obtained before with the covariant derivatives, Eq. (13), or with the parallel transport of points, Eq. (8).

Now the definition of *loop algebras* is straightforward. As on group manifolds we define a map from the circle, parametrized by a variable z, to parallel transformations of the seven-sphere. The multiplication is defined pointwise. The resulting infinite-dimensional "Lie" algebra is a direct product of the previous one, and a multiplication by a function of z. So we obtain

$$[z^m K^a, z^n K^b]X = 2S^{abc}(X)z^{m+n}K^c X. \tag{26}$$

We have not been able to define a central extension to this soft Lie algebra.

III. NONASSOCIATIVE ALGEBRAS ON S^7

We come now to another way of defining the composition of two parallel transports. It is again a parallel transport, namely the one corresponding to the multiplied octonions. This yields a nonassociative loop algebra. In the vicinity of the unit element we have

$$[K_\eta,K_\lambda]' = K_{\eta\lambda-\lambda\eta}. \tag{27}$$

The superscript $'$ is added to draw attention to the fact that a new composition law for parallel transports has to be used. In the basis defined before, we have

$$[K^a,K^b] = -2\Delta^{abc}K^c. \tag{28}$$

This yields the nonassociative Lie algebra. In fact, we obtain a violation of the Jacobi identity

$$\begin{aligned}&[[K^a,K^b]',K^c]' + \text{two cyclic permutations} \\ &\quad = 12\phi^{abcd}K_d,\end{aligned} \tag{29}$$

where ϕ has been defined in Appendix A, Eq. (A4). This violation of associativity is due to the fact that we apply the successive transformations at different points.

At the antipodal points $X = \pm 1$, the commutator according to the composition of the previous section, Eq. (5), coincides with the commutator [Eq. (27)].

Now we can define an infinite-dimensional algebra with a central extension,

$$[K^a_n,K^b_m] = -2\Delta^{abc}K^c_{n+m} - \tfrac{1}{3}cn\delta^{ab}\delta_{n+m}. \tag{30}$$

Indeed, the central extension does not modify Eq. (29).

IV. SOFT $N=8$ SUPERALGEBRAS

The supergroup OSp(8|2) contains an SO(8) subgroup. We will extend the construction of the soft algebra, outlined in Sec. II to this case. The generators of the supergroup OSp(8|2) act on eight bosonic real coordinates x^I and a complex Grassman coordinate ξ. The group is defined as the operations which leave invariant the quadratic form

$$X^2 \equiv x^I x^I + \bar{\xi}\xi. \tag{31}$$

The "super-seven-sphere" is defined by the space $X^2 = 1$. Apart from the SO(8) generators [Eqs. (18)], the supergroup contains the supersymmetries

$$\begin{aligned}G_I^+ &= \sqrt{2}x_I \frac{\partial}{\partial\xi} + \frac{1}{\sqrt{2}}\bar{\xi}\frac{\partial}{\partial x^I}, \\ G_I^- &= \sqrt{2}x_I \frac{\partial}{\partial\bar{\xi}} - \frac{1}{\sqrt{2}}\xi\frac{\partial}{\partial x^I}.\end{aligned} \tag{32}$$

Further there is the Sp(2) subgroup of the Virasoro algebra,

$$\begin{aligned}L_0 &= \frac{1}{2}\left[\xi\frac{\partial}{\partial\xi} - \bar{\xi}\frac{\partial}{\partial\bar{\xi}}\right], \\ L_1 &= -\bar{\xi}\frac{\partial}{\partial\xi}, \\ L_{-1} &= \xi\frac{\partial}{\partial\bar{\xi}}.\end{aligned} \tag{33}$$

The coset space which we describe in this way is OSp(8|2)/OSp(7|2). To see this, first remark that by definition the supergroup OSp(8|2) connects all points on the supersphere. Therefore we should just consider the stability group of one point. This is easily done for the point

$$\xi = 0, \quad x^I = \delta^I_0, \tag{34}$$

which is left invariant by all the generators of a OSp(7|2) supergroup. The algebra of OSp(8|2) is

$$\begin{aligned}&[L_0,L_{\pm1}] = \mp L_{\pm1}, \\ &[L_1,L_{-1}] = 2L_0, \\ &[L_0,G_I^\pm] = \mp\tfrac{1}{2}G_I^\pm, \\ &[L_{\pm1},G_I^\pm] = 0, \\ &[L_{\pm1},G_I^\mp] = \pm G_I^\pm, \\ &\{G_I^\pm,G_J^\pm\} = -2L_{\pm1}\delta_{IJ}, \\ &\{G_I^+,G_J^-\} = -2L_0\delta_{IJ} - M_{IJ}, \\ &[M_{IJ},M_{KL}] = \delta_{IL}M_{JK} + \delta_{JK}M_{IL} \\ &\qquad\qquad - \delta_{JL}M_{IK} - \delta_{IK}M_{JL}, \\ &[M_{IJ},G_K^\pm] = \delta_{JK}G_I^\pm - \delta_{IK}G_J^\pm\end{aligned} \tag{35}$$

(M commutes with L). As in Sec. II, the M_{IJ} do not form an independent set on the seven-sphere, and can be written in terms of the K^a defined there, Eq. (23). If the supercoset is OSp(8|2)/OSp(7|2) one can express the 28 bosonic generators of SO(8) in terms of seven generators which induce motions without fixed points: K^a. A similar reduction is not possible for the other generators. If this is done, we get new

commutation relations, as Eq. (25) and

$$\begin{aligned}[K^a,G_K^{\pm}] &= -\Gamma_{KI}^{a0}G_I^{\pm},\\ \{G_I^{+},G_J^{-}\} &= -2L_0\delta_{IJ}\\ &\quad -\tfrac{1}{4}[\Gamma_{IJ}^{a0}+\tfrac{1}{2}S^{abc}(X)\Gamma_{IJ}^{bc}]K^a.\end{aligned} \tag{36}$$

So this algebra is a soft $N=8$ conformal superalgebra.

Clearly the same soft algebra is obtained from any coset space of OSp(8|2) containing SO(8)/SO(7). As

$$\frac{\mathrm{SO}(8)}{\mathrm{SO}(7)}=\frac{\mathrm{SU}(4)}{\mathrm{SU}(3)}=\frac{\mathrm{Sp}(4)}{\mathrm{Sp}(2)}=\frac{\mathrm{Spin}(7)}{\mathrm{G}_2} \tag{37}$$

we can look for supergroups containing the subgroups SU(4), Sp(4), or Spin(7) of SO(8). It seems interesting that the last coset is contained in the coset of two exceptional supergroups: F(4) and G(3). Here F(4) contains the right number of fermionic generators and the Sp(2) group. However, we leave for further study whether a soft algebra for these supergroups can be constructed.

In this way we have constructed a class of finite-dimensional superalgebras (for other superalgebras using octonions, see Ref. 5). We did not obtain soft infinite-dimensional superalgebras, but the superalgebras we constructed correspond to the finite-dimensional subalgebras of an interesting class of infinite-dimensional superalgebras, which are associative for $N=1$, 2, and 4 and nonassociative for $N=8$.

V. NONASSOCIATIVE $N=8$ SUPERALGEBRAS

There is another motivation to consider algebras related to the seven-sphere. The infinite-dimensional extended superalgebras which are useful for string theories seem to be those which contain the following operators: the Virasoro generators L_n, N fermionic dimension $\frac{3}{2}$ operators G_r^i, and generators of a loop algebra K_n^a.[4] Using Jacobi identities one proves[7] that there are only solutions for $N=1$, 2, and 4 and that the loop algebra is not necessary for $N=1$, is $\hat{\mathrm{U}}(1)$ for $N=2$ and $\hat{\mathrm{SU}}(2)$ for $N=4$. These algebras are $(i,j,\ldots=0,\ldots,N-1;\ a,b,c=1,\ldots,N-1)$

$$\begin{aligned}[L_m,L_n] &= (m-n)L_{m+n}+(c/12)(m^3-m)\delta_{m+n},\\ [L_n,G_r^i] &= (\tfrac{1}{2}n-r)G_{n+r}^i,\\ \{G_r^i,G_s^j\} &= -2\delta^{ij}(L_{r+s}+(c/6)(r^2-\tfrac{1}{4})\delta_{r+s})\\ &\quad -(r-s)\Delta^{aij}K_{r+s}^a,\\ [L_n,K_m^a] &= -mK_{n+m}^a,\\ [K_n^a,G_r^i] &= -\Delta^{aij}G_{n+r}^j,\\ [K_n^a,K_m^b] &= -2\Delta^{abc}K_{n+m}^c-\tfrac{1}{3}cn\delta^{ab}\delta_{n+m},\\ \Delta^{aob} &= -\Delta^{abo}=\delta^{ab},\\ \Delta^{abc} &= \epsilon^{abc}\quad \text{for } N=4, \text{ and else zero.}\end{aligned} \tag{38, 39}$$

In other words, these algebras are connected to the division algebras $\mathbb{R}(N=1)$, $\mathbb{C}(N=2)$, and $\mathbb{H}(N=4)$. The Δ^{ijk} are the structure constants of these algebras [see Eq. (7)]. When we want to extend this to $N=8$ we get to the octonion algebra. This division algebra is nonassociative, and therefore the Jacobi identities are not satisfied.

The generators L_0, $L_{\pm 1}G_{\pm 1/2}$, and K_0 define a finite-dimensional superalgebra. The soft superalgebras from the previous section have exactly these generators. They do not satisfy the Jacobi identities when we consider the values of the structure constants at one point of the manifold, but they define the nonassociative algebras.

If we compare the result Eqs. (35) and (36) at the point "1" with the finite-dimensional part of Eq. (38), using $G_{+1/2}\equiv G^+$ and $G_{-1/2}\equiv G^-$, only the last commutator of Eq. (36) gives a different result because of

$$\Gamma_{BC}^{a0}=\Delta^{aBC}. \tag{40}$$

At the point "1" we get

$$\{G_I^{+},G_J^{-}\}' = -2L_0\delta_{IJ}+[\delta_I^a\delta_J^0-\delta_I^0\delta_J^a]K^a. \tag{41}$$

With Eq. (38), the Jacobi identities are only violated for three generators with an a-type index,

$$\begin{aligned}[A,B,C] &\equiv [[A,B\},C\}-[A,[B,C\}\}\\ &\quad +(-1)^{ab}[B,[A,C\}\},\end{aligned} \tag{42}$$

$$\begin{aligned}&[G_r^a,G_s^b,K_m^c]=4(r-s)\phi^{abcd}K_{r+s+m}^d,\\ &[K_m^a,K_n^b,G_r^c]=4\phi^{abcd}G_{m+n+r}^d,\\ &[K_m^a,K_n^b,K_p^c]=12\phi^{abcd}K_{m+n+p}^d,\\ &[G_r^a,G_s^b,G_t^c]=0.\end{aligned} \tag{43}$$

So all violations of Jacobi identities are proportional to ϕ, are therefore completely antisymmetric in four indices and satisfy

$$\Delta^{abc}[X^a,X^b,X^c]=0, \tag{44}$$

where any X can be G or K.

Taking the structure constants at the point 1 of the algebra of the previous section [Eqs. (36)], we get a more complicated result for this violation,

$$\begin{aligned}&[K^a,K^b,K^c]'=12\phi^{abcd}K^d,\\ &[K^a,K^b,G^c]'=4\phi^{abcd}G^d,\\ &[G_+^a,G_-^b,K^c]'=K^a\delta^{bc}-K^b\delta^{ac},\\ &[G_{\pm}^0,G_{\mp}^a,K^b]'=\pm\Delta^{abc}K^c,\\ &[G_{\pm}^a,G_{\pm}^b,G_{\mp}^0]'=\pm 2\Delta^{abc}G_{\pm}^c,\\ &[G_+^a,G_-^b,G_{\pm}^0]'=\mp\Delta^{abc}G_{\pm}^c,\\ &[G_+^a,G_+^b,G_-^c]'=G_+^a\delta^{bc}+G_+^b\delta^{ac}-2G_+^c\delta^{ab},\end{aligned} \tag{45}$$

where' again indicates that we take the structure constants rather than the structure functions, in which case we would of course get zero. These results are not completely antisymmetric in four indices, or do not satisfy Eq. (44). It would be interesting to check whether in the supercoset F(4)/G(3)

the restriction of the structure functions of the corresponding soft superalgebra to one point would give those in Eq. (38).

VI. DISCUSSION

In this paper, algebras usually associated with group manifolds were extended to the seven-sphere, having in mind the structure of Eq. (38) for $N = 8$.

On the one hand we have used the parallelizability properties of the seven-sphere to construct a soft algebra. At one point this algebra corresponds to the nonassociative algebra of octonion multiplication. We used this structure to build a loop algebra. Again this defines a soft algebra, while specifying the structure functions to one point, a nonassociative infinite-dimensional algebra is obtained.

On the other hand the algebra of parallel transformations on the seven-sphere can also be generalized to a soft superalgebra of transformations on a super-seven-sphere. We have illustrated this with a construction of the coset $OSp(8|2)/OSp(7|2)$.

We do not know a generalization of soft superalgebras to an infinite-dimensional superalgebra. However, they belong to a class containing the infinite-dimensional conformal algebras in two dimensions, where one has only dimension 2, $\frac{3}{2}$, and 1 generators. Such algebras have been constructed with $N = 1, 2$, and 4 fermionic dimension $\frac{3}{2}$ operators.[4] These algebras have always a finite-dimensional subalgebra, which for $N = 2$ and 4 are connected to a construction on the other parallelizable spheres S^1 and S^3. Here we have constructed a soft finite-dimensional algebra which continues this series, and a nonassociative infinite-dimensional algebra where the violation of the Jacobi identities [Eqs. (43)] is proportional to the completely antisymmetric associator of the octonions.

We did not look to field representations of these $N = 8$ algebras. Recently two-dimensional actions with $N = 8$ have been proposed,[8] but we do not yet know whether they are related to these algebras.

APPENDIX A: NOTATION AND USEFUL FORMULA

We give here some notation concerning the associator that gives the nonassociativity of the octonion algebra. More detail can be found in Refs. 6 and 9. The associator is defined by

$$[A,B,C] \equiv (AB)C - A(BC). \tag{A1}$$

It is antisymmetric, pure imaginary and

$$[A,B,C] = -[\bar{A},B,C] \tag{A2}$$

$$A[A,B,C] = [A,\bar{A}B,C] = [A,BA,C] = [A,B,C]\bar{A}. \tag{A3}$$

For the basis imaginary units this defines a tensor antisymmetric in four indices:

$$[e^a,e^b,e^c] = 2\phi^{abcd}e^d,$$
$$\phi^{abcd} = -(1/3!)\epsilon^{abcdefg}\Delta^{cfg} = \Delta^{e[ab}\Delta^{cd]}{}_e, \tag{A4}$$

where Δ has been defined in Eq. (7).

For the Γ matrices defined in Sec. II we have the useful property

$$\Gamma^{ik}_{IJ}\Gamma^{jl}_{IJ} = 8(\delta^{ij}\delta^{kl} - \delta^{il}\delta^{jk}),$$
$$\Gamma^{ij}_{IK}\Gamma^{ij}_{JL} = 8(\delta^{IJ}\delta^{KL} - \delta^{IL}\delta^{JK}). \tag{A5}$$

APPENDIX B: RELATION BETWEEN A PARALLELISM AND A COMPOSITION LAW

We define a parallelism on a set of points by an equivalence relation—denoted by $\|$—on couples of points such that

(A) $(A,B)\|(C,D) \Rightarrow (B,A)\|(D,C)$,

(B) $(A,B)\|(A,C) \Rightarrow B = C$,

(C) for all A, B, and C there is a D such that $(A,B)\|(C,D)$, (B1)

(D) for all A and B $(A,A)\|(B,B)$.

From such a parallelism we can define a composition law with an identity and unique inverse elements by choosing a base point O and defining

$$A * B = C \quad \text{if } (O,A)\|(B,C). \tag{B2}$$

The typical illustration of this concept is given by the usual parallelism of the plane, directly related to its vectorial structure. Then obviously the composition law defines a commutative group. However, in general we cannot reconstruct a parallelism from a composition law without extra assumptions. But if we have a composition law with

(a) a unit element 1,

(b) for each A there is a unique right and left inverse A^{-1}, (B3)

(c) $(A * B) * B^{-1} = A$,

(d) $(A * B)^{-1} = B^{-1} * A^{-1}$,

then a parallelism can be defined by

$$(A,B)\|(C,D) \quad \text{if} \quad B * A^{-1} = D * C^{-1}. \tag{B4}$$

This parallelism then also satisfies an extra property for the point O corresponding to the unit element 1

(E) if $(B,O)\|(O,C)$ and $(O,D)\|(C,E)$
$\Rightarrow (O,E)\|(B,D)$. (B5)

There is now an equivalence between a parallelism satisfying (A–E) and a composition law satisfying (a–d), where the special point O is mapped on the unit element.

Group composition laws are associative and therefore conditions (B3) are satisfied. We can define * as the left or right multiplication in the group, and this defines then left

and right parallelisms on group manifolds, where we have chosen the base point O as the identity of the group. Conversely, from the parallelism one recovers the group composition law by Eq. (B2).

A group composition law is not really necessary to define a parallelism. As Cartan and Schouten[2] showed, we can define parallelisms on the seven-sphere, which we discussed in the main text. However, we implicitly used there a metric preserved by the parallelisms, such that these parallelisms are called absolute parallelisms." In this way we are restricted there to the round seven-sphere.

[1]P. Goddard and D. Olive, IJMPA **1**, 303 (1986).

[2]E. Cartan and J. A. Schouten, Proc. Kon. Akad. Wet. Amsterdam **29**, 803 (1926); J. A. Wolf, J. Diff. Geom. **6**, 317 (1972); **7**, 19 (1972).

[3]M. F. Sohnius, Z. Phys. C **18**, 229 (1983).

[4]M. Ademollo, L. Brink, A. D'Adda, R. D'Auria, E. Napolitano, S. Sciuto, E. Del Giudice, P. di Vecchia, S. Ferrara, F. Gliossi, R. Musto, and R. Peltorino, Phys. Lett. B **62**, 105 (1976); Nucl. Phys. B **114**, 297 (1976); M. Ademollo, L. Brink, A. D'Adda, R. D'Auria, E. Napolitano, S. Sciuto, E. Del Giudice, P. di Vecchia, S. Ferrara, F. Gliossi, R. Musto, R. Peltorino, and J. Schwarz, Nucl. Phys. B **111**, 77 (1976).

[5]Z. Hasiewicz and J. Lukierski, Phys. Lett. B **145**, 65 (1984).

[6]M. Rooman, Nucl. Phys. B **236**, 501 (1984); Ph.D. thesis, ULB, Brussels, 1984.

[7]P. Ramond and J. H. Schwarz, Phys. Lett. B **64**, 75 (1976); R. Gastmans, A. Sevrin, W. Troost, and A. Van Proeyen, IJMPA **2**, 195 (1987).

[8]E. Bergshoeff, E. Sezgin, and H. Nishino, Phys. Lett. B **186**, 167 (1987).

[9]M. Günaydin and F. Gürsey, J. Math. Phys. **14**, 1651 (1973); R. Dündarer, F. Gürsey, and C.-H. Tze, *ibid.* **25**, 1496 (1984).

Vector coherent state theory and its application to the orthogonal groups

D. J. Rowe
Physics Department, University of Toronto, Toronto, Ontario, Canada M5S 1A7

R. Le Blanc
A. W. Wright Nuclear Structure Laboratory, Yale University, New Haven, Connecticut 06511

K. T. Hecht[a)]
Max Planck Institut für Kernphysik, Postfach 103980, D-6900 Heidelberg, Federal Republic of Germany

(Received 19 May 1987; accepted for publication 2 September 1987)

Vector coherent state theory is developed and presented in a form that explicitly exhibits its general applicability to the ladder representations of all semisimple Lie groups and their Lie algebras. It is shown that, in a suitable basis, the vector coherent state inner product can be inferred algebraically, by K-matrix theory, and changed to a simpler Bargmann inner product thereby facilitating the explicit calculation of the matrix representaions of Lie algebras. Applications are made to the even and odd orthogonal Lie algebras.

I. INTRODUCTION

Two recent advances have added enormously to the power of coherent state theory as a tool in Lie group and Lie algebra representation theory. The first is the extension of standard coherent state theory to admit vector-valued functions.[1-3] This apparent complication effects major simplification because it allows one to package much of the complexity of a coherent state wave function into a simple and well understood intrinsic structure, thereby greatly simplifying the remaining aspects of the wave function. It also allows one to use tensorial methods and hence the Wigner–Eckart theorem to effect further simplifications. The second advance is a simple K-matrix technique for determining inner products[2] without which coherent state representation theory would not be viable as a computational tool. Thus K-matrix theory obviates the need for invoking the cumbersome, and often unknown, integral form of the coherent state identity resolution. It solves the long standing problem of how to map the so-called Dyson representations[4] into Holstein–Primakoff representations[5] and hence allows one to determine explicitly the matrices of Lie algebra ladder representations.

The combination of these two advances is a powerful new construction for inducing explicit matrix representations of semisimple Lie groups and their Lie algebras from ladder representations of specific subgroups. Many new insights are gained and many hitherto unsolved problems can be solved in a straightforward and often routine way. In particular, Rowe[3] and Rowe, Rosensteel, and Carr[6] were able to derive analytic expressions for many matrix elements of the $sp(6,\Re)$ algebra [also called $sp(3,\Re)$], in a $u(3)$ basis, and to find a simple numerical algorithm to compute those matrix elements for which exact analytic expressions, in principle, do not exist. Thus they were able to extend to arbitrary ladder representations of $sp(6,\Re)$ the results of Castaños, Chacón, and Moshinsky[7] and Deenen and Quesne[8] for representations induced from one-dimensional representations of $u(3)$.

In other applications, Hecht and Elliott[9] were able to determine the matrices of the compact $sp(4) \supset u(2)$ algebra and Hecht[10] those of $so(8) \supset u(4)$. Applications to $sp(4,\Re)$ were made by Castaños *et al.*[11] Vector coherent state (VCS) theory was combined with the powerful complementarity principle[12] of Moshinsky and Quesne and Kashiwara and Vergne to determine the $sp(2n,\Re) \supset u(n)$ branching rules.[13] It was used to determine the matrices for irreducible representations of $su(3)$ in a canonical $so(3)$ basis[14] and to give a canonical reduction of the $u(4) \supset so(4)$ Wigner supermultiplet structure.[15] The fundamental spinor representations of the $so(2n)$ Lie algebras of major interest in the theory of many-fermion systems were constructed by Rowe and Carvalho.[16] The $so(n,2)$ and $so^*(2n)$ Lie algebras were analyzed by Le Blanc and Rowe.[17] Finally, Hecht, Le Blanc, and Rowe[18] showed that the above-mentioned techniques can be used to give the matrix elements of the $u(n)$ Lie algebras in terms of $u(n-1)$ Wigner and Racah coefficients. The latter were then given, using the VCS framework, by Le Blanc and Hecht.[19] Hence, by proceeding recursively, one is able to obtain the matrix representations of $u(n)$ in the familiar Gel'fand bases.

An extension of K-matrix theory to the tensor algebras of Lie groups has been proposed by Le Blanc and Rowe and used to determine elementary Wigner and Racah coefficients.[14,15,19]

Thus the VCS and K-matrix theories have been shown to play a central role in the representation theory of Lie algebras and their associated Wigner–Racah calculus and, with each new application, their power and versatility have become more evident. For this reason we herein analyze their structure in some depth and extend their domains of applicability. Although it has not been stressed previously, it is important to note that VCS theory and the K-matrix techniques for calculating matrix elements work with the complex extensions of Lie algebras and thus simultaneously treat all real forms of a given complex Lie algebra. For example, the theories of the noncompact $sp(2n,\Re)$ $[so^*(2n)]$ and compact $sp(2n)$ $[so(2n)]$ Lie algebras are the same. Thus, among the semisimple Lie algebras, the only ones which have not been studied using the new theory are the odd or-

[a)] On leave of absence from Physics Department, University of Michigan, Ann Arbor, Michigan 48109.

 0022-2488/88/020287-18$02.50

thogonal and exceptional Lie algebras. This is because, although the VCS theory applies equally well to all semisimple Lie groups and their Lie algebras, the K-matrix technique for finding inner products for the $so(2n+1)$ representations presents some additional features. In this paper, we present the VCS and K-matrix theories in a general context and give their properties and relationships. We show that they are applicable to all the orthogonal groups, both even and odd, thus completing the demonstration that they provide a simple systematic way of determining the explicit matrices for the ladder representations of any classical Lie algebra in bases that reduce its $u(n)$ or $gl(n)$ subalgebras [as opposed to the Gel'fand and Tsetlin[20] bases that reduce the canonical chain $so(m) \supset so(m-1) \supset so(m-2) \supset \cdots$; cf. also Baird and Biedenharn[21]]. We make application to the exceptional Lie algebra g_2 in a following paper.[22]

VCS theory is fundamentally a systematic prescription for representing any Lie group G, which admits a ladder representation, as a group of linear transformations of a Hilbert space of holomorphic vector-valued functions of a set of complex variables. In coordinate free terms, the Hilbert space is a space of holomorphic sections of a fiber bundle associated to a principle H bundle over a base manifold G/H, where H is a suitably defined subgroup of G. Thus VCS theory is a generalization of standard (scalar) coherent state theory for which the vectors are one dimensional.

The idea of holomorphic vector-valued representations of Lie groups originated with Harish-Chandra.[23] The discrete series representations of the $Sp(2n,\Re)$ groups were expressed by Godement[24] on a Hilbert space of vector-valued functions on the Siegel half-plane,[25] taking vector values in the carrier space of an irreducible representation of the subgroup $U(n)$. Since $U(n)$ is the symmetry group of the n-dimensional harmonic oscillator, these representations are particularly appropriate to the solution of the many-body problem in a harmonic-oscillator basis, as shown by Rosensteel and Rowe[26] who applied them to the microscopic nuclear collective model.[26,27]

Its vector holomorphic representations give $sp(2n,\Re)$ an obvious realization as a subalgebra of the universal enveloping algebra of what Rosensteel and Rowe[28] called a $u(n)$-boson, or $u(n)$-Weyl, algebra. The latter algebra is a direct sum of $u(n)$ and the mth Heisenberg–Weyl algebra $hw(m)$, where $m = n(n+1)/2$. Such realizations are called by physicists *boson expansions*. Rosensteel and Rowe[28] and subsequently many others[11,29] showed that, in a certain limit, the $sp(2n,\Re)$ algebra contracts to the much simpler (semidirect sum) $u(n)$-Weyl algebra. The contraction yields valuable physical insight into the macroscopic interpretation of the microscopic symplectic model of the nucleus, which is a collective model having an $sp(6,\Re)$ dynamical structure. It also gives highly accurate analytical approximations to $sp(6,\Re)$ matrix elements[6,30] which can be used in model calculations.[31]

In a similar way, its vector holomorphic representations give $u(n+1)$ a realization as a subalgebra of the $u(n) \oplus hw(n)$ enveloping algebra. We shall show here, that the even orthogonal algebra $so(2n)$ is realized in terms of $u(n) \oplus hw(n(n-1)/2)$ and that the odd $so(2n+1)$ is realized in terms of $u(n) \oplus hw(n) \oplus hw(n(n-1)/2)$.

Bosons of $hw(n(n \pm 1)/2)$ Lie algebras appear in physics as composites of more elementary particles. The Cooper pairs of superconductivity theory[32] and the s and d bosons of the nuclear interacting boson approximation[33] (IBA) are just two of many examples. Their algebraic substructure is seen from the well-known fact that while the $hw(n(n+1)/2)$ bosons carry the fundamental $\{1\}$ representations of $u(n(n+1)/2)$, they also carry the symmetric $\{2\}$ representation of $u(n)$. Similarly, while the $hw(n(n-1)/2)$ bosons carry the fundamental $\{1\}$ representations of $u(n(n-1)/2)$, they carry the antisymmetric $\{11\}$ representation of $u(n)$. The $hw(n(n \pm 1)/2)$ Lie algebras have therefore been referred to by Le Blanc and Rowe[34] as Heisenberg–Weyl algebras of symmetric and antisymmetric bosons, respectively. For $n = 3$, the symmetric bosons have been well studied in the context of the IBA and their matrix elements are known.[30,35] Matrix elements for both the symmetric and antisymmetric bosons have been given for all n by Le Blanc and Rowe.[34]

Now, it is well known that (scalar) holomorphic functional representations of Lie groups are obtainable by coherent state methods. The prototype is the Bargmann representation[36] of the Heisenberg–Weyl group and its Lie algebra $hw(n)$. At the Lie algebra level, the $hw(n)$ raising and lowering operators, which satisfy the commutation relations

$$[c_i^\dagger, c_j^\dagger] = [c_i, c_j] = 0\,, \quad [c_i, c_j^\dagger] = \delta_{ij}\,,$$

are simply represented in terms of n complex variables by

$$c_i^\dagger \to z_i\,, \quad c_i \to \frac{\partial}{\partial z_i}\,.$$

The extension of scalar coherent state representation theory to other Lie groups, which admit ladder representations, was made by Perelomov[37] and Onofri.[38]

Coherent state theory has been enormously important in physics and in mathematical physics.[39] In particular, it has led to an understanding of the coherence properties of electromagnetic radiation[40] and to predictions of new kinds of coherence that could be realized physically in, for example, a two-photon laser.[41] It has provided an understanding of minimal wave packets (group orbits of lowest weight states) and their importance for the classical limit of quantum mechanics. It underlies the Hartree–Fock and time-dependent Hartree–Fock variational equations, which are fundamental to many-body quantum theory, and it has suggested extensions to these theories.[42] Finally it has been used extensively in the realization of Lie algebras in terms of Bargmann variables and the construction of approximate truncated expansions (boson expansion theory). Its use in this latter context has been reviewed by Dobaczewski.[43]

From its successful applications to date, it is clear that the versatility of coherent state theory is enhanced substantially by its extension to admit vector-valued functions and by the simple K-matrix techniques for determining its inner products. We have discussed above its value in Lie algebra representation theory. Another important recent application is to the extension of the ordinary random-phase approximation (RPA) to what has been called a *partial*

RPA.[44] In contrast to the standard theory, the partial RPA admits the possibility of many-body particle-hole vacuum (i.e., lowest weight) states having an intrinsic structure and opens up new ways of thinking of the elementary excitations of composite systems and their phase transitions.

II. VECTOR COHERENT STATE THEORY

A. General theory

Let G be a semisimple Lie group with Lie algebra $\mathbf{g}$ and let $\mathbf{g}^c$ denote its complex extension. To determine a VCS representation of $\mathbf{g}$, one must select a subalgebra $\mathbf{h}\subset\mathbf{g}$ that contains the Cartan subalgebra and is such that $\mathbf{g}^c$ can be decomposed

$$\mathbf{g}^c = \mathbf{n}_- + \mathbf{h}^c + \mathbf{n}_+ , \tag{2.1}$$

where $\mathbf{n}_\pm$ are, respectively, nilpotent subalgebras of raising and lowering operators.

We require that, like the familiar Cartan decomposition, the above decomposition is reductive in the sense that, for $C\in\mathbf{h}^c$,

$$A\in\mathbf{n}_+ \Rightarrow [C,A]\in\mathbf{n}_+ , \quad B\in\mathbf{n}_- \Rightarrow [C,B]\in\mathbf{n}_- ,$$
$$C'\in\mathbf{h}^c \Rightarrow [C,C']\in\mathbf{h}^c . \tag{2.2}$$

This ensures that elements of $\mathbf{n}_\pm$ transform as components of (possibly reducible) tensors under the subalgebra $\mathbf{h}$. Consequently, the reduction (2.1) facilitates the decomposition of ladder representations of $\mathbf{g}$ into irreducible representations of $\mathbf{h}$. In particular, it follows that the subspace of states of an irreducible ladder representation of $\mathbf{g}$ that satisfy

$$A\,|\psi\rangle = 0 , \quad \forall A\in\mathbf{n}_+ , \tag{2.3a}$$

carries an irreducible representation of $\mathbf{h}$. We refer to this subspace as the highest-weight $\mathbf{h}$ space. Similarly, the subspace of states that satisfy

$$B\,|\psi\rangle = 0 , \quad \forall B\in\mathbf{n}_- , \tag{2.3b}$$

is an irreducible lowest-weight $\mathbf{h}$ space. Application of the lowering tensors to the highest-weight states, or the raising tensors to the lowest-weight states, then generates intermediate-weight $\mathbf{h}$ spaces. This generalizes the familiar Cartan construction of lowering and raising from one-dimensional highest- and lowest-weight spaces.

Note that the group H generated by the subalgebra $\mathbf{h}$ is the stability subgroup of all elements in G that leave the highest- (and lowest-) weight spaces invariant. We therefore naturally refer to $\mathbf{h}$ as the *stability subalgebra* of $\mathbf{g}$.

Let $[\sigma]$ label an irreducible ladder representation of $\mathbf{g}$ and let $\{|\sigma\alpha\rangle\}$ be a basis for the corresponding highest-weight $\mathbf{h}$ space. (We could equally well choose the lowest-weight $\mathbf{h}$ space and, for infinite-dimensional representations with lowest- but without highest-weight $\mathbf{h}$ states, it is necessary to do so.) Let $\{A_\nu\}$ be a basis of raising operators for $\mathbf{n}_+$. An arbitrary vector $\mathbf{z}\in\mathbf{n}_+$ can then be expanded as

$$\mathbf{z} = \sum_\nu z_\nu A_\nu . \tag{2.4}$$

[The components (z_ν) of $\mathbf{z}$ can be regarded as coordinates for $\mathbf{n}_+$ and hence for the factor space G/H in the standard way.[45]] The VCS representation of an arbitrary state $|\Psi\rangle$ in the irrep is then defined by

$$\Psi(z) = (z|\Psi\rangle = \sum_\alpha |\sigma\alpha\rangle\Psi_\alpha(z) , \tag{2.5a}$$

where $\Psi_\alpha(z)$ is the holomorphic function

$$\Psi_\alpha(z) = \langle\sigma\alpha|e^{\mathbf{z}}|\Psi\rangle . \tag{2.5b}$$

The function $\Psi(z)$ is therefore holomorphic in the z coordinates and takes vector values in the highest-weight $\mathbf{h}$ space, i.e., the carrier space for an irrep $\{\sigma\}$ of the stability subgroup H.

The carrier space for the VCS representation is thus defined as the space of all such holomorphic vector-valued functions. It can be defined more satisfactorily, independently of the abstract space for the ladder representation, as the Hilbert space of all holomorphic vector-valued functions that are square integrable with respect to the VCS inner product (cf. next section and Ref. 3).

The VCS representation of a group element $g\in G$ is defined by

$$[\Gamma(g)\Psi](z) = (z|g|\Psi\rangle = \sum_\alpha |\sigma\alpha\rangle\langle\sigma\alpha|e^{\mathbf{z}}g|\Psi\rangle , \tag{2.6}$$

and the corresponding representation of an infinitesimal generator by

$$[\Gamma(X)\Psi](z) = \sum_\alpha |\sigma\alpha\rangle\langle\sigma\alpha|e^{\mathbf{z}}X\,|\Psi\rangle . \tag{2.7}$$

Consider, for example, the VCS action of an element $h\in H$. First observe that the subalgebra $\mathbf{n}_+$ carries a (possibly reducible) representation ρ under the adjoint action, i.e.,

$$A_\nu \to hA_\nu h^{-1} = A_\mu\rho_{\mu\nu}(h) , \tag{2.8}$$

for $A_\nu\in\mathbf{n}_+$, $h\in H$, with summation over repeated indices. It follows that

$$[\Gamma(h)\Psi](z) = \sum_\alpha |\sigma\alpha\rangle\langle\sigma\alpha|h\exp[h^{-1}\mathbf{z}h\,]|\Psi\rangle$$
$$= \sum_\alpha h\,|\sigma\alpha\rangle\Psi_\alpha(h^{-1}\cdot z) , \tag{2.9}$$

where $h^{-1}\cdot z$ is defined by

$$(h^{-1}\cdot z)_\mu = \rho_{\mu\nu}(h^{-1})z_\nu . \tag{2.10}$$

The VCS representation of any element $X\in\mathbf{g}^c$ is given explicitly as a linear differential operator as follows. First expand

$$e^{\mathbf{z}}X = (e^{\mathbf{z}}Xe^{-\mathbf{z}})e^{\mathbf{z}} = (X + [\mathbf{z},X\,] + \tfrac{1}{2}[\mathbf{z}[\mathbf{z},X\,]\,] + \cdots)e^{\mathbf{z}} ,$$

and note that, since $\mathbf{n}_+$ is nilpotent, the series terminates. Let $\{A_\nu\}$, $\{B_\nu\}$, and $\{C_i\}$ denote bases for the subalgebras $\mathbf{n}_+$, $\mathbf{n}_-$, and $\mathbf{h}^c$, respectively. Now, again since $\mathbf{n}_+$ is nilpotent, $A_\nu e^{\mathbf{z}}$ is expressible as a differential operator on $e^{\mathbf{z}}$. Thus, for example, in the special case that $\mathbf{n}_+$ is Abelian, we have

$$\sum_\alpha |\sigma\alpha\rangle\langle\sigma\alpha|A_\nu e^{\mathbf{z}}|\Psi\rangle = \partial_\nu\Psi(z) ,$$

where $\partial_\nu = \partial/\partial z_\nu$. More generally, from

$$\partial_\nu e^{\mathbf{z}} = e^{\mathbf{z}}(e^{-\mathbf{z}}\partial_\nu e^{\mathbf{z}}) = e^{\mathbf{z}}(A_\nu + \tfrac{1}{2}z_\mu\,[A_\nu,A_\mu\,] + \cdots) ,$$
$$A_\nu e^{\mathbf{z}} = e^{\mathbf{z}}(e^{-\mathbf{z}}A_\nu e^{\mathbf{z}}) = e^{\mathbf{z}}(A_\nu + z_\mu\,[A_\nu,A_\mu\,] + \cdots) ,$$

we have

$$\sum|\sigma\alpha\rangle\langle\sigma\alpha|A_\nu e^z|\Psi\rangle = \left(\partial_\nu - \frac{1}{2} z_\mu c^{\rho}_{\mu\nu}\,\partial_\rho + \cdots\right)\Psi(z)\,, \tag{2.11}$$

where $c^{\rho}_{\mu\nu}$ is the Lie algebra structure constant appearing in the equation

$$[A_\mu, A_\nu] = c^{\rho}_{\mu\nu} A_\rho\,. \tag{2.12}$$

We also have, since $|\sigma\alpha\rangle$ belongs to the highest-weight subspace, that

$$\sum_\alpha |\sigma\alpha\rangle\langle\sigma\alpha|B_\nu e^z|\Psi\rangle = 0\,. \tag{2.13}$$

Finally,

$$\sum|\sigma\alpha\rangle\langle\sigma\alpha|C_i e^z|\Psi\rangle = \mathbb{C}_i\Psi(z)\,, \tag{2.14a}$$

where $\mathbb{C}_i$ is the representation of C_i carried by the highest-weight invariant subspace; i.e.,

$$\mathbb{C}_i\Psi(z) = \sum_\alpha (C_i|\sigma\alpha\rangle)\Psi_\alpha(z). \tag{2.14b}$$

We sometimes refer to $\mathbb{C}_i$ as the *intrinsic* representation of $C_i \in \mathbf{h}^c$.

Proposition 1: The VCS representation of an element C_i of the stability subalgebra $\mathbf{h}^c$ is given by

$$\Gamma(C_i) = \mathbb{C}_i + \mathscr{C}_i\,, \tag{2.15a}$$

where

$$\mathscr{C}_i = c^{\mu}_{\nu i} z_\nu\,\partial_\mu\,. \tag{2.15b}$$

Proof: Observe that $\mathscr{C}_i$ is defined such that

$$[C_i + \mathscr{C}_i, \mathbf{z}] = 0\,. \tag{2.16}$$

Next observe that

$$e^z C_i|\Psi\rangle = e^z(C_i + \mathscr{C}_i)|\Psi\rangle = (C_i + \mathscr{C}_i)e^z|\Psi\rangle\,.$$

Use of Eq. (2.14) now gives the desired result. Q.E.D.

Note that, if the element $A_\nu \in \mathbf{n}_+$ has root vector ν, it is convenient to denote by $B_\nu \in \mathbf{n}_-$ the element having root vector $-\nu$. It is also convenient to choose basis vectors for $\mathbf{n}_\pm$ in the canonical way such that the Killing form $(K_{\alpha\beta})$ satisfies

$$K_{\mu,-\nu} = k\delta_{\mu\nu}\,,\quad k = \text{const}\,,\quad K_{i,\pm\nu} = 0\,, \tag{2.17}$$

for μ,ν labeling positive roots and i labeling a basis for $\mathbf{h}^c$. With such a choice, it follows that the scalar product

$$B\cdot A = K^{\mu,-\nu}B_\nu A_\mu = k^{-1}B_\nu A_\nu$$

is the difference of the quadratic Casimir invariants for $\mathbf{g}$ and $\mathbf{h}$. Hence it is invariant under the adjoint action of the stability algebra, i.e.,

$$[C_i, B_\nu A_\nu] = 0\,,\quad C_i \in \mathbf{h}^c\,. \tag{2.18}$$

The constant k is given from the general expression for the Killing form by

$$k = c^{\tau}_{\nu\sigma}\,c^{\sigma}_{-\nu\tau}\,,\quad \text{no sum on } \nu\,, \tag{2.19}$$

where σ and τ run over all μ, $-\mu$, and i indices.

Proposition 2: With a canonical choice of basis vectors for $\mathbf{n}_\pm$, the complex variables (z_ν) transform under the stability subalgebra in precisely the same way as the basis elements (B_ν) of $\mathbf{n}_-$.

Proof: The equation

$$[C_i, A_\nu] = c^{\mu}_{i\nu} A_\mu$$

together with Eq. (2.18) implies that

$$[C_i, B_\mu] = c^{\mu}_{\nu i} B_\nu\,,$$

which may be compared to

$$[\mathscr{C}_i, z_\mu] = c^{\mu}_{\nu i} z_\nu\,. \qquad \text{Q.E.D.}$$

B. VCS representation of so(2n)

The fundamental realization of the SO($2n$) Lie group is as a group of orthogonal transformations of $\mathfrak{R}^{2n}$. A basis for the corresponding fundamental realization of the so($2n$) Lie algebra is given by the generalized angular momentum operators

$$L_{\alpha\beta} = -i\left(x_\alpha\frac{\partial}{\partial x_\beta} - x_\beta\frac{\partial}{\partial x_\alpha}\right),\quad \alpha,\beta = 1,\ldots,2n\,, \tag{2.20}$$

which satisfy the commutation relations

$$[L_{\alpha\beta}, L_{\gamma\delta}] = -i(\delta_{\beta\gamma}L_{\alpha\delta} - \delta_{\beta\delta}L_{\alpha\gamma} + \delta_{\alpha\delta}L_{\beta\gamma} - \delta_{\alpha\gamma}L_{\beta\delta})\,. \tag{2.21}$$

We choose as stability subalgebra the u(n) subalgebra of so($2n$). We then determine a basis of raising operators,

$$A_{ij} = -A_{ji} = \tfrac{1}{2}(L_{2i-1,2j} + L_{2i,2j-1} + iL_{2i,2j} - iL_{2i-1,2j-1})\,,\quad 1\leqslant i,j\leqslant n\,, \tag{2.22a}$$

a basis of lowering operators,

$$B_{ij} = -B_{ji} = \tfrac{1}{2}(L_{2i-1,2j} + L_{2i,2j-1} - iL_{2i,2j} + iL_{2i-1,2j-1})\,,\quad 1\leqslant i,j\leqslant n\,, \tag{2.22b}$$

and a basis for u(n),

$$C_{ij} = \tfrac{1}{2}(L_{2i-1,2j} - L_{2i,2j-1} + iL_{2i,2j} + iL_{2i-1,2j-1})\,,\quad 1\leqslant i,j\leqslant n\,. \tag{2.22c}$$

These operators satisfy the commutation relations

$$\begin{aligned}
&[B_{ij}, A_{kl}] = \delta_{il}C_{kj} - \delta_{ik}C_{lj} + \delta_{jk}C_{li} - \delta_{jl}C_{ki}\,,\\
&[C_{ij}, A_{kl}] = \delta_{jk}A_{il} + \delta_{jl}A_{ki}\,,\\
&[C_{ij}, B_{kl}] = -\delta_{il}B_{kj} - \delta_{ik}B_{jl}\,,\\
&[C_{ij}, C_{kl}] = \delta_{jk}C_{il} - \delta_{il}C_{kj}\,.
\end{aligned} \tag{2.23}$$

Observe that for this choice of stability subalgebra the raising operators A_{ij} span an Abelian algebra. Note also that they transform under U(n) as components of a u(n) $\{11\}$ tensor whereas the lowering operators B_{ij} transform as components of a $\{-1-1\}$ tensor. [We omit all zeros in labeling a u(n) irrep. Thus $\{11\}\sim\{110\cdots0\}$ and $\{-1-1\}\sim\{0\cdots0-1-1\}$.]

The so($2n$) operators also have a well-known realization in terms of fermion creation and annihilation operators given by the isomorphism

$$A_{ij}\sim\alpha_i^\dagger\alpha_j^\dagger\,,\quad B_{ij}\sim\alpha_j\alpha_i\,,\quad C_{ij}\sim\tfrac{1}{2}(\alpha_i^\dagger\alpha_j - \alpha_j\alpha_i^\dagger)\,,$$

where the fermion operators satisfy the anticommutation relations

$$\{\alpha_i,\alpha_j\} = \{\alpha_i^\dagger,\alpha_j^\dagger\} = 0\,,\quad \{\alpha_i,\alpha_j^\dagger\} = \delta_{ij}\,.$$

For this reason the so($2n$) algebra is often referred to in physics as the fermion-pair algebra and its representations are of major importance in the quantum theory of many-fermion systems.

A highest-weight state $|\mathrm{hw}\rangle$ for a unitary irreducible representation (UIR) of so($2n$) is defined by

$$\begin{aligned} &A_{ij}|\mathrm{hw}\rangle = 0\,, \quad i,j = 1,\ldots,n\,, \\ &C_{ij}|\mathrm{hw}\rangle = 0\,, \quad i<j\,, \\ &C_{ii}|\mathrm{hw}\rangle = \lambda_i|\mathrm{hw}\rangle\,, \quad \text{no sum on } i. \end{aligned} \tag{2.24}$$

The integers (or half-odd integers for a spinor representation) $[\lambda]\sim[\lambda_1\cdots\lambda_n]$ serve to label the representation. Now observe that $|\mathrm{hw}\rangle$ is also a highest-weight state for a highest-weight UIR $\{\lambda\}$ of the $u(n)$ stability subalgebra. Let $\{|\lambda\alpha\rangle\}$ denote a basis for this highest-weight u(n) subrepresentation. Define

$$\mathbf{z} = \frac{1}{2}\sum_{ij} y_{ij}\,A_{ij}\,, \quad y_{ij} = -y_{ji}\,. \tag{2.25}$$

The VCS representation of an arbitrary state $|\Psi\rangle$ in the so($2n$) representation space is then defined by Eq. (2.7). We easily derive the VCS representation of the so($2n$) Lie algebra

$$\begin{aligned} &\Gamma(A_{ij}) = \nabla_{ij} = \frac{\partial}{\partial y_{ij}}\,, \quad \Gamma(C_{ij}) = \mathbb{C}_{ij} + \mathscr{C}_{ij}\,, \\ &\Gamma(B_{ij}) = y_{il}\mathbb{C}_{lj} - y_{jl}\mathbb{C}_{li} + y_{ik}\,y_{lj}\nabla_{kl}\,, \end{aligned} \tag{2.26}$$

where $\mathbb{C}_{ij}$ is the highest-weight u(n) UIR of C_i and

$$\mathscr{C}_{ij} = y_{jl}\nabla_{li}\,, \tag{2.27}$$

and where summation over repeated indices is implied.

C. VCS representation of so($2n$+1)

For so($2n+1$), we must augment the angular momentum operators of so($2n$) by

$$L_{2n+1,2i-1}\,, \quad L_{2n+1,2i}\,, \quad i = 1,\ldots,n\,.$$

Thus we obtain additional raising and lowering operators given by

$$\begin{aligned} \mathscr{A}_i &= (1/\sqrt{2})(L_{2n+1,2i} - iL_{2n+1,2i-1})\,, \\ \mathscr{B}_i &= (1/\sqrt{2})(L_{2n+1,2i} + iL_{2n+1,2i-1})\,, \end{aligned} \tag{2.28}$$

respectively. These operators satisfy the commutation relations

$$\begin{aligned} &[\mathscr{A}_i,\mathscr{A}_j] = A_{ij}\,, \quad [\mathscr{B}_i,\mathscr{B}_j] = \mathscr{B}_{ji}\,, \quad [\mathscr{A}_i,\mathscr{B}_j] = C_{ij}\,, \\ &[C_{ij},\mathscr{A}_k] = \delta_{jk}\mathscr{A}_i\,, \quad [C_{ij},\mathscr{B}_k] = -\delta_{ik}\mathscr{B}_j\,, \\ &[\mathscr{A}_i,A_{kl}] = [\mathscr{B}_i,B_{kl}] = 0\,, \\ &[\mathscr{A}_i,B_{kl}] = \delta_{il}\mathscr{B}_k - \delta_{ik}\mathscr{B}_l\,, \\ &[\mathscr{B}_i,A_{kl}] = \delta_{ik}\mathscr{A}_l - \delta_{il}\mathscr{A}_k\,. \end{aligned} \tag{2.29}$$

One now observes that the complete set of raising operators $(\mathscr{A}_i,A_{ij})$ no longer spans an Abelian algebra. One also notes that the $(\mathscr{A}_i)$ and $(\mathscr{B}_i)$ operators transform as components of $\{1\}$ and $\{-1\}$ tensors, respectively, under u(n).

It is of interest to note that, like so($2n$), the so($2n+1$) Lie algebra has a realization in terms of fermion operators given by the isomorphism[46]

$$\mathscr{A}_i \sim (1/\sqrt{2})\alpha_i^\dagger\,, \quad \mathscr{B}_i \sim (1/\sqrt{2})\alpha_i\,.$$

If $\{|\sigma\alpha\rangle\}$ denotes a basis for a highest-weight u(n) subrepresentation of an so($2n+1$) ladder representation, then the so($2n+1$) VCS wave functions are defined by

$$\Psi(y,z) = \sum_\alpha |\sigma\alpha\rangle\langle\sigma\alpha|e^{\mathbf{z}}|\Psi\rangle\,, \tag{2.30}$$

where

$$\mathbf{z} = z_i\,\mathscr{A}_i + \tfrac{1}{2}y_{ij}\,A_{ij}\,. \tag{2.31}$$

We readily derive the VCS representation of the so($2n+1$) Lie algebra

$$\begin{aligned} &\Gamma(A_{ij}) = \nabla_{ij} = \frac{\partial}{\partial_{ij}}\,, \\ &\Gamma(\mathscr{A}_i) = \partial_i + \frac{1}{2}z_l\nabla_{li}\,, \quad \left(\partial_i = \frac{\partial}{\partial z_i}\right), \\ &\Gamma(C_{ij}) = \mathbb{C}_{ij} - z_j\,\partial_i + y_{jl}\nabla_{li}\,, \\ &\Gamma(\mathscr{B}_i) = z_l\mathbb{C}_{li} - \tfrac{1}{2}z_i\,z_l\,\partial_l - y_{il}\,\partial_l + \tfrac{1}{2}z_k\,y_{il}\nabla_{lk}\,, \\ &\Gamma(B_{ij}) = y_{il}\mathbb{C}_{lj} - y_{jl}\mathbb{C}_{li} + y_{il}\,y_{kj}\nabla_{lk} \\ &\qquad - \tfrac{1}{2}z_i\,z_l\mathbb{C}_{lj} + \tfrac{1}{2}z_j\,z_l\mathbb{C}_{li} - y_{il}z_j\,\partial_l + y_{jl}z_i\,\partial_l\,. \end{aligned} \tag{2.32}$$

III. THE INNER PRODUCT AND THE CALCULATION OF MATRIX ELEMENTS

The VCS inner product of two states is given by an integral of the form

$$\langle\Psi|\Psi'\rangle_{\mathrm{VCS}} = \int \langle\Psi|z\rangle(z|\Psi'\rangle\,d\mu(z)\,,$$

where $d\mu(z)$ is the VCS measure.[3] However, while important for the completeness of VCS theory, this integral form of the inner product is cumbersome and difficult to use in practice.

The problem of deducing the inner product implicit in a ladder representation is not unique to VCS theory. Conversely, techniques for calculating VCS inner products are more generally applicable. A standard approach, which is particularly useful for calculating the overlaps of states in a nonorthonormal basis, is by means of generator functions.[47] This approach has been successfully employed by Castaños *et al.*[11] in the construction of orthonormal bases for sp(4,$\Re$) representations. However, to our knowledge, no systematic algorithm is known at present to construct orthonormal bases in general by generator function methods.

We therefore give a generalization of the K-matrix approach of Rowe[2] which derives the inner product directly from the Lie algebra structure. This approach has proved to be remarkably simple and effective in application to those cases where the algebra of raising (or lowering) operators is Abelian. We show here that it applies more generally.

A. Generalized K-matrix theory

An inner product on a vector space is conveniently defined by specification of an orthonormal basis. We therefore seek a basis $\{|\phi(\alpha)\rangle\}$ for the space of vector-valued functions that is orthonormal with respect to the VCS inner product and indexed by a suitable parameter set (α). We shall refer to such a basis simply as an *orthonormal VCS basis.*

We start by considering a basis $\{|\alpha\rangle\}$ that is orthonor-

mal with respect to a Bargmann, as opposed to the VCS, inner product and then seek a transformation K which will map it onto an orthonormal VCS basis;

$$K\colon |\alpha\rangle \to |\phi(\alpha)\rangle = K|\alpha\rangle .$$

If $\{|i\rangle\}$ denotes an orthonormal basis for the highest-weight irrep of the stability subalgebra, then an orthonormal Bargmann basis can be defined by the vector-valued polynomials

$$(z|ni\rangle = \prod_{\nu} \frac{(z_\nu)^{n_\nu}}{\sqrt{n_\nu!}} |i\rangle . \tag{3.1}$$

In practice, it will be convenient to select a basis having useful reducibility properties with respect to the subalgebra chains involved. But for the moment, let $\{|\alpha\rangle\}$ simply denote an arbitrary Bargmann basis *indexed, for convenience, by the same parameter set as the VCS basis.*

It follows that a VCS basis state can be expanded on the Bargmann basis

$$|\phi(\alpha)\rangle = |\beta\rangle K_{\beta\alpha} . \tag{3.2a}$$

We therefore define the operator K that maps the Bargmann basis state $|\alpha\rangle$ to the VCS basis state $|\phi(\alpha)\rangle$ by

$$|\alpha\rangle \to |\phi(\alpha)\rangle = K|\alpha\rangle = |\beta\rangle K_{\beta\alpha} . \tag{3.2b}$$

The operator K must clearly satisfy the equation

$$\langle\alpha|K^\dagger K|\beta\rangle = \langle\phi(\alpha)|\phi(\beta)\rangle \tag{3.3a}$$

in which here and throughout this paper *it is understood that all inner products of holomorphic vector-valued wave functions, even for VCS basis wave functions, are Bargmann inner products* unless explicitly stated otherwise, e.g., by a VCS subscript.

It should be emphasized that a fundamental assumption of K-matrix theory is that the space of vector-valued wave functions that are square integrable with respect to the VCS inner product is a vector subspace of the Bargmann space (the space of states that are square integrable with respect to the Bargmann inner product). Such a subspace is usually called a *physical* subspace by physicists. Thus, in general, K is a map from Bargmann space onto its physical subspace. Conversely, the *pullback* $P = K^{-1}$ is a well-defined map from the physical subspace into Bargmann space. Note, however, that care must be exercised in the use of P since it is undefined outside of the physical subspace.

Proposition 3: The operator

$$K^\dagger K = \sum_{\alpha\beta} |\alpha\rangle\langle\alpha|K^\dagger K|\beta\rangle\langle\beta| \tag{3.3b}$$

is uniquely defined independently of the choice of either the Bargmann or the VCS basis.

Proof: First observe that the physical subspace of the Bargmann space is defined without reference to a basis. Then, since K annihilates any state belonging to the orthogonal complement of the physical subspace, the sum on α and β in Eq. (3.3b) can be extended to a sum over a complete orthonormal basis for all the Bargmann space. It follows that $K^\dagger K$ is independent of this basis. Finally, observe that the overlap $\langle\phi(\alpha)|\phi(\beta)\rangle$ is invariant under a unitary transformation

$$|\phi(\alpha)\rangle \to |\phi(\alpha')\rangle U_{\alpha'\alpha} ,$$

demonstrating that $K^\dagger K$ is independent of the VCS basis. Q.E.D.

Once the map K and hence its pullback P have been determined, the inner product of any two state vectors $|\Psi\rangle$ and $|\Psi'\rangle$ in the carrier space of a ladder representation of a Lie group can be determined. If $(z|\Psi\rangle$ denotes the VCS wave function for the state $|\Psi\rangle$ then the corresponding Bargmann wave function is given by $(z|\psi\rangle = (z|P|\Psi\rangle$ and the inner product by

$$\langle\Psi|\Psi'\rangle_{\rm VCS} = \langle\psi|\psi'\rangle = \langle\Psi|P^\dagger P|\Psi'\rangle . \tag{3.4}$$

Thus K-matrix theory allows one to combine the simplicity of the Bargmann product with the VCS construction of wave functions.

Now, if $\Gamma(X)$ is the VCS representation of an element $X\in\mathbf{g}^c$, its matrix elements $\gamma_{\beta\alpha}(X)$ are defined by

$$\Gamma(X)|\phi(\alpha)\rangle = |\phi(\beta)\rangle\gamma_{\beta\alpha}(X) .$$

We then have the relationship

$$\Gamma(X)K|\alpha\rangle = K\gamma(X)|\alpha\rangle ,$$

where the operator $\gamma(X)$ is defined by

$$\gamma(X)|\alpha\rangle = |\beta\rangle\gamma_{\beta\alpha}(X) . \tag{3.5}$$

Hence we infer that

$$\Gamma(X)K = K\gamma(X) . \tag{3.6}$$

Thus K is an intertwining operator that relates the VCS representation Γ to an equivalent representation γ. Once K has been determined, the desired $\gamma_{\beta\alpha}(X)$ matrix elements can be evaluated from

$$\gamma_{\beta\alpha}(X) = \langle\beta|\gamma(X)|\alpha\rangle = \langle\beta|K^{-1}\Gamma(X)K|\alpha\rangle . \tag{3.7}$$

Of particular importance is the fact that, if Γ is unitary with respect to the VCS inner product, then γ will be unitary with respect to the much simpler Bargmann inner product. Consequently, Γ will generally be nonunitary with respect to the Bargmann measure. The representation Γ is therefore sometimes referred to as a nonunitary Dyson representation[4] and the associated unitary representation γ is referred to as a Holstein–Primakoff[5] representation. More generally, we can say (definition below) that γ is compatible with the Bargmann inner product but Γ is not.

In determining K by algebraic methods, we make use of Hermitian adjoint relationships. Let $\Gamma^\dagger(X)$ and $\gamma^\dagger(X)$ denote the Hermitian adjoints of $\Gamma(X)$ and $\gamma(X)$, respectively, with respect to the Bargmann inner product, and let $\Gamma(X^\dagger)$ be the Hermitian adjoint of $\Gamma(X)$ with respect to the VCS inner product, i.e.,

$$\langle\phi(\alpha)|\Gamma(X^\dagger)|\phi(\beta)\rangle_{\rm VCS} = \langle\phi(\beta)|\Gamma(X)|\phi(\alpha)\rangle^*_{\rm VCS} .$$

It follows from

$$\langle\phi(\beta)|\Gamma(X)|\phi(\alpha)\rangle_{\rm VCS} = \langle\beta|\gamma(X)|\alpha\rangle$$

that

$$\gamma^\dagger(X) = \gamma(X^\dagger), \tag{3.8}$$

but that, in general,

$$\Gamma^\dagger(X) \neq \Gamma(X^\dagger) . \tag{3.9}$$

Definition: Let $\mathbf{k}$ be a subalgebra of $\mathbf{g}^c$ and let R be a representation of $\mathbf{k}$. If, for all $X\in\mathbf{k}$, $R(X)$ and $R(X^\dagger)$ are

Hermitian adjoints with respect to a given inner product then we say that the representation R of $\mathbf{k}$ is compatible with the given inner product. Thus, for example, the Bargmann (generalized Holstein–Primakoff) representation γ of $\mathbf{g}^c$ is compatible with the Bargmann inner product, and the VCS (generalized Dyson) representation Γ is compatible with the VCS inner product. But Γ is not compatible with the Bargmann inner product. This gives a generalization of the concepts of Dyson and Holstein–Primakoff representation to situations where the underlying representations may or may not be unitary.

Proposition 4: The VCS representation of the stability subalgebra $\mathbf{h}^c \subset \mathbf{g}^c$ is compatible with the Bargmann inner product.

Proof: Define $B_\nu = A_\nu^\dagger$ (compatible with the requirement that the quantity $B_\nu A_\nu$ should be invariant under H). Now, for $C_i \in \mathbf{h}^c$,

$$[C_i, A_\nu] = c_{i\nu}^{\mu} A_\mu$$

implies

$$[C_i^\dagger, B_\nu] = c_{\nu i}^{\mu *} B_\mu .$$

Thus $C_i^\dagger$ is identified as the unique element of $\mathbf{h}^c$ that satisfies the last equation and

$$[C_i^\dagger, C_j^\dagger] = c_{ji}^{k *} C_k^\dagger .$$

It follows from these equations and Proposition 2 that $\Gamma(C_i^\dagger)$ is the unique operator that satisfies the equations

$$[\Gamma(C_i^\dagger), z_\nu] = c_{\nu i}^{\mu *} z_\mu , \quad [\Gamma(C_i^\dagger), \Gamma(C_j^\dagger)] = c_{ji}^{k *} \Gamma(C_k^\dagger) .$$

These equations are seen to be satisfied by

$$\Gamma(C_i^\dagger) = \Gamma^\dagger(C_i) = \mathbb{C}_i^\dagger + c_{\nu i}^{\mu *} z_\mu \partial_\nu . \tag{3.10}$$

Q.E.D.

As a consequence of Proposition 4, it is possible to construct an orthonormal Bargmann basis $\{|i\{\mu\}\nu\rangle\}$ of vector-valued wave functions that reduces the stability subalgebra $\mathbf{h}$. We use the notation that the set of wave functions with $\imath$ and $\{\mu\}$ fixed carry an irrep $\{\mu\}$ of $\mathbf{h}^c$. Thus ν labels a basis for the irrep $\{\mu\}$ and i is a set of multiplicity indices. Such a basis is easily constructed by first constructing orthonormal polynomials $Z_\nu^{\tau\{\zeta\}}(z)$ in (z_ν), which transform as the components of an irreducible tensor $\{\zeta\}$ under h, and subsequently coupling them to the vectors of the (intrinsic) highest-weight irrep $\{\sigma\}$ of $\mathbf{h}$; i.e.,

$$\begin{aligned} (z|i\{\mu\}\nu\rangle &= (z|[\sigma]\tau\{\zeta\}\rho\{\mu\}\nu\rangle \\ &= [Z^{\tau\{\zeta\}}(z) \times |\{\sigma\}\rangle]_\nu^{\rho\{\mu\}} . \end{aligned} \tag{3.11}$$

Note that the three labels $i \sim (\tau\{\zeta\}\rho)$ indicate the three kinds of $\mathbf{h}$ multiplicity that can arise in the decomposition $\mathbf{g}\downarrow\mathbf{h}$. The label $\rho = 1,\ldots,m_{\zeta\sigma}^{\mu}$ resolves the coupling

$$\{\zeta\} \otimes \{\sigma\} = \sum_\mu m_{\zeta\sigma}^{\mu} \{\mu\} ,$$

where $m_{\zeta\sigma}^{\mu}\{\mu\}$ denotes the multiplicity of a given irrep $\{\mu\}$ in the outer (Kronecker) product. The label $\{\zeta\}$ refers to the tensorial properties of the polynomial $Z^{\tau\{\zeta\}}$ under $\mathbf{h}$ while the multiplicity label τ resolves any ambiguities in its construction.

Proposition 5: The VCS and the Bargmann representation of $\mathbf{g}^c$ become identical on restriction to $\mathbf{h}^c \subset \mathbf{g}^c$; i.e.,

$$\gamma(C) = \Gamma(C) = \mathbb{C} + \mathscr{C} , \quad \forall C \in \mathbf{h}^c . \tag{3.12}$$

Proof: Since the VCS irrep Γ is compatible with the Bargmann inner product when restricted to $\mathbf{h}^c \subset \mathbf{g}^c$, it follows that a corresponding VCS orthonormal basis exists of the form

$$\begin{aligned} |\phi(i\{\mu\}\nu)\rangle &= K\,|i\{\mu\}\nu\rangle \\ &= |j\{\mu\}\nu\rangle K_{ji}(\{\mu\}) \quad \text{(no sum on } \mu\text{)}. \end{aligned} \tag{3.13}$$

Now, by definition,

$$\Gamma(C)|\phi(i\{\mu\}\nu)\rangle = K\gamma(C)|i\{\mu\}\nu\rangle .$$

But

$$\begin{aligned} K\gamma(C)|i\{\mu\}\nu\rangle &= |j\{\mu\}\nu'\rangle K_{ji}\gamma_{\nu'\nu}^{\{\mu\}}(C) \\ &= \gamma(C)|\phi(i\{\mu\}\nu)] . \end{aligned}$$

Q.E.D.

It follows from Eq. (3.13) that, by choosing basis states in both the VCS and Bargmann Hilbert spaces to reduce the stability subalgebra $\mathbf{h}^c$, we automatically block-diagonalize the K matrix. Furthermore, since the Bargmann and VCS inner products are identical on the highest-weight $\mathbf{h}$ subspace, we require, without loss of generality, the highest-weight $K(\{\sigma\})$ submatrix to be the identity matrix. Thus we are naturally led to seek a recursion relation for the matrices starting from the highest-weight matrix.

From the two equations

$$\Gamma(X)KK^\dagger = K\gamma(X)K^\dagger , \quad \Gamma(X^\dagger)KK^\dagger = K\gamma(X^\dagger)K^\dagger , \tag{3.14}$$

we easily derive

$$KK^\dagger\Gamma^\dagger(X) = \Gamma(X^\dagger)KK^\dagger , \quad \forall X \in \mathbf{g}^c . \tag{3.15a}$$

Setting $X = A_\nu$ and $X^\dagger = B_\nu$ then gives

$$KK^\dagger\Gamma^\dagger(A_\nu) = \Gamma(B_\nu)KK^\dagger . \tag{3.15b}$$

If we choose the solution to Eq. (3.3) for which K is the positive Hermitian square root of $K^\dagger K$, then

$$K = K^\dagger , \quad KK^\dagger = K^\dagger K = K^2 ,$$

and Eq. (3.15) leads to a recursion relation in K^2.

Equation (3.15) is particularly useful when the raising operators commute and span an invariant subspace under the adjoint action of the stability algebra. Oddly enough, the corresponding equation

$$\Gamma^\dagger(X)P^\dagger P = P^\dagger P\Gamma(X^\dagger) \tag{3.16}$$

for the inverse operator $P^\dagger P$ is sometimes more useful when the raising operator algebra is non-Abelian.

For $\mathrm{u}(n+1)$, $\mathrm{sp}(2n,\Re)$, and $\mathrm{so}(2n)$ with $\mathrm{u}(n)$ as stability algebra, the raising operator algebras are all Abelian. Furthermore, in each of these cases, the raising operators are all components of a single irreducible tensor under $\mathrm{u}(n)$. However, in contrast $\mathrm{so}(2n+1)$ has a non-Abelian raising operator algebra that carries a reducible representation of the stability algebra $\mathrm{u}(n)$. (The same is true of the exceptional Lie algebras.) We therefore consider the two situations separately.

B. Case (i): Abelian raising operator algebras

It follows immediately from the VCS expansion that

$$\Gamma(A_\nu) = \partial_\nu , \tag{3.17a}$$

$$\Gamma^\dagger(A_\nu) = z_\nu . \tag{3.17b}$$

Furthermore, one easily shows (Appendix A) that $\Gamma(B_\nu)$ can be expressed in the form

$$\Gamma(B_\nu) = [\hat{\Omega}, z_\nu] , \tag{3.18}$$

where $\hat{\Omega}$ is the combination of Casimir invariants

$$\hat{\Omega} = \tfrac{1}{4}k\,[\mathscr{C}\cdot\mathscr{C} - 2\gamma(C\cdot C) + \tfrac{1}{2}\hat{N}_z] \tag{3.19}$$

in the $\mathscr{C}$ and $\gamma(C)$ realizations of the stability algebra $\mathbf{h}^c$; k is defined by Eq. (2.19) and the operator $\hat{N}_z = z_\nu\,\partial_\nu$ simply measures the degree of a polynomial wave function in the (z_ν).

We therefore set $K = K^\dagger$ in Eq. (3.15) and obtain the simple equation of Rowe,[2]

$$K^2 z_\nu = [\hat{\Omega}, z_\nu] K^2 , \tag{3.20a}$$

which can also be expressed in the form

$$K^2 \hat{N}_z = K^2 z_\nu\,\partial_\nu = [\hat{\Omega}, z_\nu] K^2\,\partial_\nu . \tag{3.20b}$$

Since both $\hat{N}_z$ and $\hat{\Omega}$ are diagonal in the chosen basis, i.e.,

$$\hat{\Omega}|i\{\mu\}\nu\rangle = \Omega(i\{\mu\})\,|i\{\mu\}\nu\rangle , \tag{3.19'}$$

Eq. (3.20) immediately gives a recursion relation for the $K(\{\mu\})$ submatrices [see, e.g., Eqs. (3.22), (3.25), and (4.8)] upon taking matrix elements of both sides between states of successively lower weight starting from the highest, for which we have already noted the $K(\{\sigma\})$ matrix to be a multiple of the identity.

An especially simple case arises for matrix elements between states that are multiplicity-free as they are, for example, for the completely multiplicity-free case of $\mathrm{u}(n+1)$ in a $\mathrm{u}(n)$ Gel'fand basis reviewed by Hecht, Le Blanc, and Rowe.[18] The $K(\{\mu\})$ submatrices are then one dimensional and their ratios are given immediately from Eq. (3.20) by

$$K^2(\{\mu\})\langle\{\mu\}\|z\|\{\nu\}\rangle = (\Omega(\{\mu\}) - \Omega(\{\nu\}))\langle\{\mu\}\|z\|\{\nu\}\rangle K^2(\{\nu\}) . \tag{3.21}$$

Thus, when $\langle\{\mu\}\|z\|\{\nu\}\rangle \neq 0$, we obtain, choosing the positive square root, the solution

$$K(\{\mu\})/K(\{\nu\}) = (\Omega(\mu) - \Omega(\nu))^{1/2} . \tag{3.22}$$

For the $\mathrm{u}(n)$ algebras, this difference was shown[18] to be simply expressible in terms of the well-known *hooks*[48] of the theory of the symmetric group. We find, in Sec. IV, a similar expression for the orthogonal groups. These results reflect the pervasive role of the symmetric group in the theory of the tensor representations of Lie algebras, in general.

Finally, matrix elements of the Lie algebra are obtained from the generalized Holstein–Primakoff representation

$$\gamma(A_\nu) = K^{-1}\Gamma(A_\nu)K = K^{-1}\,\partial_\nu K ,$$
$$\gamma(B_\nu) = \gamma^\dagger(A_\nu) = K z_\nu K^{-1} , \tag{3.23}$$
$$\gamma(C_i) = \Gamma(C_i) = \mathbb{C}_i + c^{\mu}_{\nu i}\, z_\nu\,\partial_\mu .$$

One sees that to evaluate matrix elements, one only needs the ratios of the $K(\{\mu\})$ submatrices and not their absolute values. Thus, for a multiplicity-free case, for example, we immediately obtain from Eqs. (3.20a) and (3.22) the analytic expression

$$(\{\mu\}\|\gamma(B)\|\{\nu\}\rangle = (\Omega(\{\mu\}) - \Omega(\{\nu\}))^{1/2}\langle\{\mu\}\|z\|\{\nu\}\rangle . \tag{3.24}$$

For the general situation, the recursion relation for the $K(\{\mu\})$ matrices is found from Eq. (3.20b) to be

$$K^2_{ij}(\{\mu\}) = \frac{1}{N_z(j)} \sum_{k\{\mu'\}} (\Omega(i\{\mu\}) - \Omega(k\{\mu'\})) \times \langle i\{\mu\}\|z\|k\{\mu'\}\rangle\langle j\{\mu\}\|z\|l\{\mu'\}\rangle^* K^2_{kl}(\{\mu'\}) , \tag{3.25}$$

where

$$N_z(j) = \langle j\{\mu\}\|\hat{N}_z\|j\{\mu\}\rangle .$$

It should be emphasized, however, that Eq. (3.24) applies not only to multiplicity-free representations, but also to the more common situation of matrix elements between particular states of an arbitrary representation that happen to be multiplicity-free. Thus, even for representations of physical interest with multiplicities, one often needs to solve the general Eq. (3.25) for only a relatively small number of cases.

We conclude this section with the observation that the simple form (3.23) of $\gamma(B_\nu)$ together with the isomorphism

$$[Z^{\tau\{\zeta\}}(B)\times|\{\sigma\}\rangle]^{\rho\{\mu\}}_\nu \sim [Z^{\tau\{\zeta\}}(KzK^{-1})\times|\{\sigma\}\rangle]^{\rho\{\mu\}}_\nu$$

implies that the abstract state vectors corresponding to the Bargmann basis (3.13) are given by

$$|[\sigma]\tau\{\zeta\}\rho\{\mu\}\nu\rangle = K^{-1}[Z^{\tau\{\zeta\}}(B)\times|\{\sigma\}\rangle]^{\rho\{\mu\}}_\nu . \tag{3.26}$$

C. Case (ii): Non-Abelian raising operator algebras

Proposition 6: The Lie algebra $\mathbf{n}_+$ (and likewise $\mathbf{n}_-$) contains an Abelian subalgebra invariant under the adjoint action of $\mathbf{h}$.

Proof: The Lie algebra $\mathbf{n}_+$ is invariant under the adjoint action of $\mathbf{h}$. Decompose it into a direct sum of vector subspaces each of which is invariant and irreducible under $\mathbf{h}$. Let $\mathbf{n}^1_+$ be the highest-weight subspace. The commutator of two elements of $\mathbf{n}^1_+$ must vanish or belong to a higher-weight subspace. Therefore since $\mathbf{n}^1_+$ is the highest-weight subspace, it must be an Abelian subalgebra. A parallel argument shows the existence of a lowest-weight Abelian subalgebra of lowering operators. Q.E.D.

Proposition 7: The direct sum

$$\mathbf{l}^c = \mathbf{n}^1_- + \mathbf{h}^c + \mathbf{n}^1_+$$

is a subalgebra of $\mathbf{g}^c$.

Proof: Simply observe that $[\mathbf{n}^1_-, \mathbf{n}^1_+] \subset \mathbf{h}^c$ by addition of weights. Q.E.D.

Let us set

$$\mathbf{n}_\pm = \mathbf{n}^1_\pm + \mathbf{n}^2_\pm , \tag{3.27}$$

where $\mathbf{n}^2_\pm$ is the $\mathbf{h}$-invariant complement of $\mathbf{n}^1_\pm$. The Abelian subalgebra $\mathbf{n}^1_\pm$ is, by construction, irreducible under $\mathbf{h}$. *For simplicity, we shall restrict consideration to the situation in*

which $\mathbf{n}^2_\pm$ *is also irreducible,* although many of the following results are more general. The more general case is considered in Ref. 22. We then have, for example,

$$[\mathbf{n}^2_+ ,\mathbf{n}^2_+] \subset \mathbf{n}^1_+ . \tag{3.28}$$

Proposition 8: The direct sum

$$\mathbf{n}^2 = \mathbf{n}^2_- + \mathbf{n}^2_+ \tag{3.29}$$

is invariant under the adjoint action of $\mathbf{l}$.

Proof: By construction,

$$[\mathbf{h}^c,\mathbf{n}^2] \subset \mathbf{n}^2 ,$$

and by addition of weights,

$$[\mathbf{n}^1_+ ,\mathbf{n}^2_-] \subset \mathbf{n}^2_+ , \quad [\mathbf{n}^1_- ,\mathbf{n}^2_+] \subset \mathbf{n}^2_- . \qquad \text{Q.E.D.}$$

By Proposition 7, we find that we always have a Lie algebra $\mathbf{l}^c$ with Abelian raising and lowering operator subalgebras sitting between $\mathbf{g}^c$ and the chosen stability subalgebra,

$$\mathbf{g}^c \supset \mathbf{l}^c \supset \mathbf{h}^c . \tag{3.30}$$

And, by Proposition 8, we have the decomposition

$$\mathbf{g}^c = \mathbf{l}^c + \mathbf{n}^2, \tag{3.31}$$

with the raising and lowering operators of $\mathbf{n}^2$ transforming together as the components of a (possibly reducible) tensor under $\mathbf{l}$.

Let us denote by $\{A_\nu\}$, $\{B_\nu\}$, and $\{C_i\}$ basis operators for the subalgebras $\mathbf{n}^1_\pm$ and $\mathbf{h}^c$, respectively, and by $\{\mathscr{A}_\alpha\}$ and $\{\mathscr{B}_\alpha\}$ basis operators for $\mathbf{n}^2_\pm$, respectively. The general raising operator of VCS theory then has the two term expansion

$$\mathbf{z} = z_\alpha \mathscr{A}_\alpha + y_\nu A_\nu . \tag{3.32}$$

Proposition 1 gives

$$\mathscr{C}_i = \mathscr{C}_i^{(z)} + \mathscr{C}_i^{(y)} , \tag{3.33a}$$

where

$$\mathscr{C}_i^{(z)} = c^{\beta}_{\alpha i} z_\alpha \partial_\beta , \quad \mathscr{C}_i^{(y)} = c^{\mu}_{\nu i} y_\nu \nabla_\mu , \tag{3.33b}$$

with

$$\partial_\beta = \frac{\partial}{\partial z_\beta} , \quad \nabla_\nu = \frac{\partial}{\partial y_\nu} . \tag{3.34}$$

Since $\mathbf{n}^1_+$ is Abelian, we have

$$\Gamma(A_\nu) = \nabla_\nu , \quad A_\nu \in \mathbf{n}^1_+ , \tag{3.35}$$

and, with $\mathbf{n}^2_+$ irreducible under $\mathbf{h}$,

$$\Gamma(\mathscr{A}_\alpha) = \partial_\alpha - \tfrac{1}{2} c^{\nu}_{\alpha\beta} z_\beta \nabla_\nu , \quad \mathscr{A}_\alpha \in \mathbf{n}^2_+ . \tag{3.36}$$

With a canonical basis, meaning that $K_{\mu,-\nu} = \delta_{\mu\nu} \times \text{const}$, $K_{\beta,-\alpha} = \delta_{\beta\alpha} \times \text{const}$, we have (cf. Proposition 2)

$$[C_i,B_\mu] = c^{\mu}_{\nu i} B_\nu , \quad [C_i,\mathscr{B}_\beta] = c^{\beta}_{\alpha i} \mathscr{B}_\alpha . \tag{3.37}$$

Comparison with

$$[\mathscr{C}_i, y_\mu] = c^{\mu}_{\nu i} y_\nu , \quad [\mathscr{C}_i, z_\beta] = c^{\beta}_{\alpha i} z_\alpha \tag{3.38}$$

then shows that (z_α) and (y_ν) transform as components of tensors under $\mathbf{h}$ in exactly the same ways as $(\mathscr{B}_\alpha)$ and (B_ν).

We can now easily construct an orthonormal Bargmann basis of states, which reduces the $\mathbf{g} \supset \mathbf{h}$ subalgebra chain, by

$$|i[\lambda]j\{\mu\}\nu\rangle = [Y^{\tau\{\xi\}}(y) \times [Z^{\tau\{\xi\}}(z)|\{\sigma\})]^{\rho\{\lambda\}}]^{\rho'\{\mu\}}_{\nu} , \tag{3.39}$$

where $i \sim (\tau\{\xi\}\rho)$ and $j \sim (\tau'\{\zeta\}\rho')$ label the multiplicities arising in the $\mathbf{h}$ couplings. (N.B. Throughout this paper, we assume all couplings to be ordered sequentially from right to left.)

We next seek a K operator that will map this basis into an orthonormal VCS basis $|\phi(i[\lambda]j\{\mu\}\nu)\rangle$. We find that considerable simplification arises if we require the K operator to give VCS basis states that reduce the subalgebra chain

$$\begin{array}{ccccccc} \mathbf{g} & \supset & \mathbf{l} & \supset & \mathbf{h} \\ \langle\sigma\rangle & i & [\lambda] & j & \{\mu\} \end{array}$$

with $[\lambda]$, in particular, labeling an irrep of $\mathbf{l}$. We know that the desired K operator is block diagonal in $\{\mu\}$ as a result of Proposition 4. However, there is no reason to suppose that it should be block diagonal in $[\lambda]$. Nevertheless, when acting on (multiplicity-free) highest-weight $\{\mu = \lambda\}$ $\mathbf{h}$-irrep states, we have the following result.

Proposition 6:

$$\langle i'[\lambda']j'\{\mu'\}\nu'|K|i[\lambda]\{\lambda\}\nu\rangle = \delta_{\lambda'\lambda}\delta_{\mu'\lambda}\delta_{\nu'\nu}\mathscr{K}(\{\lambda\})_{i'i} . \tag{3.40}$$

[See discussion following (3.42); K is *not* Hermitian in general.]

Proof: It follows from Eq. (3.35) that states of a highest-weight $\mathbf{h}$ subirrep $\{\lambda\}$ of an $\mathbf{l}$-irrep $[\lambda]$ by definition satisfy

$$\nabla_\nu |\phi(i[\lambda]\{\lambda\}\nu')\rangle = 0 . \tag{3.41}$$

Thus the VCS wave function $|\phi(i[\lambda]\{\lambda\}\nu)\rangle$ is independent of the (y_ν) variables and its expansion on the Bargmann basis is of the form

$$|\phi(i[\lambda]\{\lambda\}\nu)\rangle = \sum_{i'} |i'[\lambda]\{\lambda\}\nu\rangle \mathscr{K}(\{\lambda\})_{i'i} . \tag{3.42}$$

Q.E.D.

It is most important to note that having chosen the Bargmann basis for convenience and having required the VCS basis to reduce the $\mathbf{g} \supset \mathbf{l} \supset \mathbf{h}$ subalgebra chain, we have given up any freedom we had to choose the whole K operator to be Hermitian. However, no constraint has been imposed on the combinations of multiplicities of states that transform in the same way under these subgroups and so we are free to choose them such that the $\mathscr{K}(\{\lambda\})$ submatrices, defined by Eq. (3.42), are Hermitian and it is convenient to do so.

Choosing a VCS basis to reduce the $\mathbf{g} \supset \mathbf{l} \supset \mathbf{h}$ subalgebra chain means that we can proceed in steps. The first step is to induce the ladder irreps of $\mathbf{l}$ from those of its $\mathbf{h}$ highest-weight irrep, which is easily done since $\mathbf{l}$ is a Lie algebra with Abelian raising and lowering operator subalgebras for which Sec. III B applies. Let us suppose that this has been done. The remaining step is then to induce the required irrep of $\mathbf{g}$ from its highest-weight $\mathbf{l}$ irrep. To do this we find that we do not need the whole K operator; we need only its restrictions to the highest-weight $\mathbf{h}$ subspaces given by the $\mathscr{K}(\{\lambda\})$ submatrices.

We start with the general intertwining relationship from Eq. (3.6),

$$\Gamma(\mathscr{A}_\alpha)K = K\gamma(\mathscr{A}_\alpha) , \tag{3.43}$$

and observe that, since the raising operators $\mathscr{A}_\alpha$ commute

with the A_ν raising operators, they can only raise states of a highest-weight **h** irrep to other highest-weight irreps; i.e.,

$$\gamma(\mathscr{A}_\alpha)|j[\lambda]\{\lambda\}\nu\rangle = \sum_{i\lambda'\nu'} |i[\lambda']\{\lambda'\}\nu'\rangle \times\langle i[\lambda']\{\lambda'\}\nu'|\gamma(\mathscr{A}_\alpha)|j[\lambda]\{\lambda\}\nu\rangle .$$

Thus, taking **h**-reduced matrix elements of Eq. (3.43), we obtain

$$\langle i[\lambda']\{\lambda'\}\|\Gamma(\mathscr{A})\mathscr{K}(\{\lambda\})\|j[\lambda]\{\lambda\}\rangle = \langle i[\lambda']\{\lambda'\}\|\mathscr{K}(\{\lambda'\}\gamma(\mathscr{A})\|j[\lambda]\{\lambda\}\rangle ,$$

which, with Eq. (3.36), gives, for $[\lambda'] > [\lambda]$,

$$\langle i[\lambda']\{\lambda'\}\|\gamma(\mathscr{A})\|j[\lambda]\{\lambda\}\rangle = \langle i[\lambda']\{\lambda'\}\|\mathscr{K}(\{\lambda'\})^{-1}\partial\mathscr{K}(\{\lambda\})\|j[\lambda]\{\lambda\}\rangle . \tag{3.44a}$$

Taking the complex conjugate of this equation and recalling that $\mathscr{K}(\{\lambda\})$ and $\mathscr{K}(\{\lambda'\})$ are Hermitian, we then obtain for $[\lambda'] < [\lambda]$,

$$\langle j[\lambda']\{\lambda'\}\|\gamma(\mathscr{B})\|i[\lambda]\{\lambda\}\rangle = \langle j[\lambda']\{\lambda'\}\|\mathscr{K}(\{\lambda'\})z\mathscr{K}(\{\lambda\})^{-1}\|i[\lambda]\{\lambda\}\rangle . \tag{3.44b}$$

We know by Proposition 8 that $\mathscr{A}_\alpha$ and $\mathscr{B}_\alpha$ are components of a tensor under **l**; call it $\mathscr{F}$. For simplicity, let us suppose that this tensor is irreducible and of rank $[\mathscr{F}]$. Let the tensor ranks of its $\mathscr{A}$ and $\mathscr{B}$ components be denoted $\{\mathscr{A}\}$ and $\{\mathscr{B}\}$, respectively, under **h**. Then, once we have determined the $\mathscr{K}(\{\lambda\})$ matrices, we obtain from the above equations the **l**-reduced (triple bar) matrix elements of $\mathscr{F}$,

$$\langle i[\lambda']|\,\|\gamma(\mathscr{F})\|\,|j[\lambda]\rangle = \frac{\langle i[\lambda']\{\lambda'\}\|\mathscr{K}(\{\lambda'\})^{-1}\partial\mathscr{K}(\{\lambda\})\|j[\lambda]\{\lambda\}\rangle}{\langle[\lambda]\{\lambda\};[\mathscr{F}]\{\mathscr{A}\}\|[\lambda']\{\lambda'\}\rangle} \tag{3.45a}$$

for $[\lambda'] > [\lambda]$ and

$$\langle i[\lambda']|\,\|\gamma(\mathscr{F})\|\,|j[\lambda]\rangle = \frac{\langle i[\lambda']\{\lambda'\}\|\mathscr{K}(\{\lambda'\}z\mathscr{K}(\{\lambda\})^{-1}\|j[\lambda]\{\lambda\}\rangle}{\langle[\lambda]\{\lambda\};[\mathscr{F}]\{\mathscr{B}\}\|[\lambda']\{\lambda'\}\rangle} \tag{3.45b}$$

for $[\lambda'] < [\lambda]$, where $\langle[\lambda]\{\lambda\};[\mathscr{F}]\{\mathscr{A}\}\|[\lambda']\{\lambda'\}\rangle$ and $\langle[\lambda]\{\lambda\};[\mathscr{F}]\{\mathscr{B}\}\|[\lambda']\{\lambda'\}\rangle$ are $\mathbf{l}\supset\mathbf{h}$ reduced Wigner coefficients. In the event of a multiplicity in the $[\lambda]\times[\mathscr{F}]\to[\lambda']$ coupling, the specific couplings are the ones in which states transforming according to $\gamma(F)|j[\lambda]\{\lambda\}\nu\rangle$ are chosen.[49]

Thus it remains to determine the $\mathscr{K}(\{\lambda\})$ matrices. We recall that, if P is any operator (not necessarily equal to the pullback K^{-1} of the chosen K operator) that maps an orthonormal VCS basis into an orthonormal Bargmann basis for the physical subspace, then, in particular, it must map the orthonormal VCS states

$$\{|\phi(i[\lambda]\{\lambda\}\nu)\rangle = \mathscr{K}(\{\lambda\})|i[\lambda]\{\lambda\}\nu\rangle\}$$

into orthonormal Bargmann states

$$\{P\mathscr{K}(\{\lambda\})|i[\lambda]\{\lambda\}\nu\rangle\} .$$

Hence it must satisfy

$$\langle i'[\lambda']\{\lambda'\}\nu'|\mathscr{K}(\{\lambda'\})P^\dagger P\mathscr{K}(\{\lambda\})|i[\lambda]\{\lambda\}\nu\rangle = \delta_{i'i}\delta_{\lambda'\lambda}\delta_{\nu'\nu} . \tag{3.46}$$

Let us choose an operator P that is Hermitian; $P = P^\dagger$. Then Eq. (3.46) implies that

$$P^2_{ij}(\{\lambda\})\mathscr{K}(\{\lambda\})^2_{jk} = \delta_{ik} , \tag{3.47}$$

where $P^2(\{\lambda\})$ is the projection of $P^2 = P^\dagger P$ to the subspace of $\{\mu = \lambda\}$ highest-weight states; i.e.,

$$P^2_{ij}(\{\lambda\}) = \langle i[\lambda]\{\lambda\}\nu|P^2|j[\lambda]\{\lambda\}\nu\rangle . \tag{3.48}$$

It is important to note that $P \neq K^{-1}$ since we have chosen $P = P^\dagger$ whereas $K \neq K^\dagger$. Nevertheless $P^2(\{\lambda\})$, which is the projection of P^2, is the inverse of $\mathscr{K}(\{\lambda\})^2$, the square of a projection. Equation (3.47) is therefore a nontrivial result.

We now use Eq. (3.16) to obtain a recursion relation for $P^2(\{\lambda\})$. Setting $X = \mathscr{A}$ in Eq. (3.16) and taking matrix elements between $[\lambda'] < [\lambda]$ highest-weight states, gives

$$\sum_k \langle i[\lambda']\{\lambda'\}\|z\|k[\lambda]\{\lambda\}\rangle P^2_{kj}(\{\lambda\}) = \sum_l P^2_{il}(\{\lambda'\})\langle l[\lambda']\{\lambda'\}\|\Gamma(\mathscr{B})\|j[\lambda]\{\lambda\}\rangle + \sum_{k,\lambda''>\lambda}\langle i[\lambda']\{\lambda'\}|P^2|k[\lambda'']\{\lambda'\}\rangle \times\langle k[\lambda'']\{\lambda'\}\|\Gamma(\mathscr{B})\|j[\lambda]\{\lambda\}\rangle . \tag{3.49}$$

Setting $X = A$ in (3.16) and noting that

$$\langle i[\lambda']\{\lambda'\}\|\Gamma^\dagger(A)P^2\|k[\lambda'']\{\lambda''\}\rangle = \langle i[\lambda']\{\lambda'\}\|yP^2\|k[\lambda'']\{\lambda''\}\rangle = 0 ,$$

we find, for $[\lambda'] < [\lambda'']$,

$$\langle i[\lambda']\{\lambda'\}\|P^2\Gamma(B)\|k[\lambda'']\{\lambda''\}\rangle = \sum_l P^2_{il}(\{\lambda'\})\langle l[\lambda']\{\lambda'\}\|\Gamma(B)\|k[\lambda'']\{\lambda''\}\rangle + \sum_{l,\lambda'''>\lambda'}\langle i[\lambda']\{\lambda'\}|P^2|l[\lambda''']\{\lambda'\}\rangle \times\langle l[\lambda''']\{\lambda'\}\|\Gamma(B)\|k[\lambda'']\{\lambda''\}\rangle = 0 . \tag{3.50}$$

Now, making an expansion of $\Gamma(B)$ in powers of (y_ν) [cf. Eq. (2.32)], one easily determines that the only component that contributes in this equation is the term independent of (∇_ν) given by

$$\Gamma^{(1)}(B_\nu) = y_\mu c^{\,i}_{\mu-\nu}[\mathbb{C}_i + \mathscr{C}_i^{(z)}] , \tag{3.51}$$

where $\mathscr{C}^{(z)}$ is defined by Eq. (3.33). It follows immediately from the structure (3.39) of the Bargmann wave functions that the only component surviving in the second term of Eq. (3.50) is the term with $l = k$ and $\lambda''' = \lambda''$. Thus Eq. (3.50) gives an explicit expression for the matrix element $\langle i[\lambda']\{\lambda'\}|P^2|k[\lambda'']\{\lambda'\}\rangle$ that can be substituted into Eq. (3.49) to give

$$\sum_k \langle i[\lambda']\{\lambda'\}\|z\|k[\lambda]\{\lambda\}\rangle P^2_{kj}(\{\lambda\}) = \sum_l P^2_{il}(\{\lambda'\})\Big[\langle l[\lambda']\{\lambda'\}\|\Gamma^{(0)}(\mathscr{B})\|j[\lambda]\{\lambda\}\rangle - \sum_{k,\lambda''>\lambda}\frac{\langle l[\lambda']\{\lambda'\}\|\Gamma^{(0)}(B)\|k[\lambda'']\{\lambda''\}\rangle}{\langle l[\lambda'']\{\lambda'\}\|\Gamma^{(1)}(B)\|k[\lambda'']\{\lambda''\}\rangle} \times\langle k[\lambda'']\{\lambda'\}\|\Gamma^{(1)}(\mathscr{B})\|j[\lambda]\{\lambda\}\rangle\Big] . \tag{3.52}$$

One observes that the only components of $\Gamma(\mathscr{B})$ and $\Gamma(B)$ contributing to matrix elements in Eq. (3.52) are the (y_ν)- and (∇_ν)-independent components $\Gamma^{(0)}(\mathscr{B})$ and $\Gamma^{(0)}(B)$, respectively, and the components $\Gamma^{(1)}(\mathscr{B})$ and $\Gamma^{(1)}(B)$, which are linear in (y_ν) and (∇_ν) independent.

Now, as shown in Appendix B, $\Gamma^{(0)}(\mathscr{B})$ can be expressed

$$\Gamma^{(0)}(\mathscr{B}_\alpha) = [\hat{\Omega}, z_\alpha], \tag{3.53}$$

where

$$\hat{\Omega} = \tfrac{1}{4}k(\mathscr{C}^{(z)}\cdot\mathscr{C}^{(z)} - 2\gamma(C\cdot C) + \tfrac{1}{2}\hat{N}_z) \tag{3.54}$$

with

$$k = K_{\alpha,-\alpha}, \quad \hat{N}_z = z_\alpha \partial_\alpha .$$

One easily determines that

$$\Gamma^{(1)}(\mathscr{B}_\alpha) = y_\nu c^{\beta}_{\nu-\alpha}\partial_\beta, \tag{3.55a}$$

which, since $\mathscr{B}$, y, and ∂ are irreducible tensors under $\mathbf{h}$, must be proportional to the coupled tensor product $[y\times\partial]^{\{\mathscr{B}\}}$, where the superscript $\{\mathscr{B}\}$ denotes the tensor rank of $\mathscr{B}$ under $\mathbf{h}$. It follows that the structure factor $c^{\beta}_{\nu-\alpha}$ is proportional to a Wigner coefficient for the stability subalgebra $\mathbf{h}$. Hence it is convenient to write

$$\Gamma^{(1)}(\mathscr{B}_\alpha) = [y\times\partial]^{\{\mathscr{B}\}}, \tag{3.55b}$$

where

$$[y\times\partial]^{\{\mathscr{B}\}} = y_\nu c^{\beta}_{\nu-\alpha}\partial_b$$

is proportional to $[y\times\partial]^{\{\mathscr{B}\}}$. We likewise find that

$$\Gamma^{(0)}(B) = \tfrac{1}{2}[z\times z\times(\gamma(C) - \tfrac{2}{3}\mathscr{C}^{(z)})]^{\{B\}}, \tag{3.56}$$

where

$$[z\times z\times(\gamma(C) - \tfrac{2}{3}\mathscr{C}^{(z)})]^{\{B\}}_\nu = z_\alpha c^{\beta}_{\alpha-\nu} z_\gamma c^{i}_{\gamma-\beta}(\gamma(C_i) - \tfrac{2}{3}\mathscr{C}^{(z)}_i) \tag{3.57}$$

is proportional to the tensor coupled product $[z\times z\times(\gamma(C) - \tfrac{2}{3}\mathscr{C}^{(z)})]^{\{B\}}$ and the superscript $\{B\}$ denotes the tensor rank of B under $\mathbf{h}$. Finally, as shown in Appendix B, $\Gamma^{(1)}(B_\nu)$ is expressed as

$$\Gamma^{(1)}(B_\nu) = [\hat{\Omega}', y_\nu], \tag{3.58}$$

where

$$\hat{\Omega}' = \tfrac{1}{2}k'[\mathscr{C}^{(y)}\cdot\mathscr{C}^{(y)} - \gamma(C\cdot C)], \tag{3.59}$$

with $k' = K_{\nu,-\nu}$.

Substituting these expressions together with Eq. (3.47) into Eq. (3.52) gives the required relations for the $\mathscr{K}(\{\lambda\})$ matrices:

$$\sum_k \mathscr{K}(\{\lambda'\})^2_{ik}\langle k[\lambda']\{\lambda'\}\|z\|j[\lambda]\{\lambda\}\rangle = \sum_l\Big[\langle i[\lambda']\{\lambda'\}\|[\hat{\Omega},z]\|l[\lambda]\{\lambda\}\rangle - \frac{1}{2}\sum_{k,\lambda''>\lambda}\frac{\langle i[\lambda']\{\lambda'\}\|[z\times z\times(\gamma(C) - \tfrac{2}{3}\mathscr{C}^{(z)})]^{\{B\}}\|k[\lambda'']\{\lambda''\}\rangle}{\langle k[\lambda'']\{\lambda'\}\|[\hat{\Omega}', y]\|k[\lambda'']\{\lambda''\}\rangle} \times\langle k[\lambda'']\{\lambda'\}\|[y\times\partial]^{\{b\}}\|l[\lambda]\{\lambda\}\rangle\Big]\mathscr{K}(\{\lambda\})^2_{lj}. \tag{3.60}$$

Note that, if the basis for the Lie algebra is chosen such that $k = k'$, then $\hat{\Omega}$ and $\hat{\Omega}'$ in Eq. (3.60) can be replaced by the single operator

$$\hat{\Lambda} = \tfrac{1}{4}k(\mathscr{C}^{(z)}\cdot\mathscr{C}^{(z)} + 2\mathscr{C}^{(y)}\cdot\mathscr{C}^{(y)} - 2\gamma(C\cdot C) + \tfrac{1}{2}\hat{N}_z). \tag{3.61}$$

IV. APPLICATIONS TO THE ORTHOGONAL GROUPS

A. The matrix representations of so(2n)

With $\mathrm{u}(n)\subset\mathrm{so}(2n)$ chosen as stability subalgebra, the remaining raising operators of $\mathrm{so}(2n)$ span an Abelian algebra and the methods of Sec. II B apply.

The lowering operators (B_{ij}) and hence, by Proposition 2, the (y_{ij}) variables transform under $\mathrm{u}(n)$ as the components of $\{-1-1\}$ tensors. We may therefore construct a basis $\{Y^{\{-b\}}_\nu(y)\}$ of Bargmann polynomials in the (y_{ij}) that transform as the components of tensors of rank $\{-b\}$ under $\mathrm{u}(n)$, where $\{-b\} = \{-b_n, -b_{n-1},\ldots,-b_1\}$ is a partition of negative integers having the property

$$b_1 = b_2 \geqslant b_3 = b_4 \geqslant \cdots \geqslant b_{2n-1} = b_{2n}. \tag{4.1}$$

These polynomials may be chosen orthonormal with respect to the Bargmann measure and to span the space of polynomials in (y_{ij}). [Note that no multiplicity index is needed since the highest-weight polynomial $Y^{\{-b\}}_{\mathrm{hw}}(y)$ is unambiguously identified[34] by the partition $\{-b\}$.] An orthonormal Bargmann basis of vector-valued wave functions for an $\mathrm{so}(2n)$ irrep $[\lambda]$ is now given by the $\mathrm{u}(n)$-coupled wave functions

$$(y|[\lambda]\{-b\}\rho\{\mu\}\nu\rangle = [Y^{\{-b\}}(y)\times|\{\lambda\}\rangle]^{\rho\{\mu\}}_\nu. \tag{4.2}$$

The VCS representation of the $\mathrm{so}(2n)$ Lie algebra is given by Eq. (2.26). We easily determine that $\Gamma(B_{ij})$ can be expressed as

$$\Gamma(B_{ij}) = [\hat{\Omega}, y_{ij}], \tag{4.3a}$$

where

$$\hat{\Omega} = \tfrac{1}{4}\hat{I}(\mathscr{C}) - \tfrac{1}{2}\hat{I}(\gamma(C)) + \tfrac{1}{2}(n-1)\hat{N}_y, \tag{4.3b}$$

with

$$\hat{I}(\mathscr{C}) = k\mathscr{C}\cdot\mathscr{C} = \mathscr{C}_{ij}\mathscr{C}_{ji}, \tag{4.4}$$

where $\mathscr{C}_{ij}$ is given by Eq. (2.27) and

$$\gamma(C_{ij}) = \Gamma(C_{ij}) = \mathbb{C}_{ij} + \mathscr{C}_{ij}.$$

The operator $\hat{I}(\gamma(C))$ is proportional to the quadratic

Casimir invariant of u(n) and has eigenvalues given by

$$\hat{I}(\gamma(C))|[\lambda]\{-b\}\rho\{\mu\}\nu\rangle = I(\{\mu\})|[\lambda]\{-b\}\rho\{\mu\}\nu\rangle\,, \tag{4.5a}$$

where

$$I(\{\mu\}) = \sum_{i=1}^{n} \mu_i(\mu_i + n + 1 - 2i)\,. \tag{4.5b}$$

The operator $\hat{N}_y$ measures the "bosonic" degree of the Bargmann wave function. It is defined

$$\hat{N}_y = \tfrac{1}{2} y_{ij}\nabla_{ij}, \tag{4.6a}$$

and its eigenvalues are given by

$$\begin{aligned}\hat{N}_y|[\lambda]\{-b\}\rho\{\mu\}\nu\rangle &= N_y(\{-b\})|[\lambda]\{-b\}\rho\{\mu\}\nu\rangle \\ &= \frac{1}{2}\left(\sum_i b_i\right)|[\lambda]\{-b\}\rho\{\mu\}\nu\rangle\,.\end{aligned} \tag{4.6b}$$

Thus $\hat{\Omega}$ is diagonal in the above Bargmann basis and has eigenvalues given by

$$\begin{aligned}\hat{\Omega}|[\lambda]\{-b\}\rho\{\mu\}\nu\rangle &= \Omega(\{-b\}\{\mu\})|[\lambda]\{-b\}\rho\{\mu\}\nu\rangle\,, \\ \Omega(\{-b\}\{\mu\}) &= \tfrac{1}{4}I(\{-b\}) - \tfrac{1}{2}I(\{\mu\}) \\ &\quad + \tfrac{1}{2}(n-1)N_y(\{-b\})\,.\end{aligned} \tag{4.7a, 4.7b}$$

The so($2n$)$\supset$u(n) reduction of an arbitrary irrep $[\lambda]$ of so($2n$) is not, in general, multiplicity-free for $n \geqslant 4$. To simplify the notation, let us index the multiplicities by $i \sim \{-b\}\rho$ and write

$$|i\{\mu\}\nu\rangle \sim |[\lambda]\{-b\}\rho\{\mu\}\nu\rangle\,.$$

Since the K matrix is diagonal in $\{\mu\}$ and independent of ν, we denote the matrix elements of K as

$$K_{ij}(\{\mu\}) = \langle i\{\mu\}\nu|K\,|\,j\{\mu\}\nu\rangle\,.$$

From Eq. (3.25), we then immediately obtain the recursion relation for the $K(\{\mu\})$ submatrices:

$$\begin{aligned}K^2_{ij}(\{\mu\}) &= \frac{1}{N_y(j)}\sum_{kl\{\mu'\}} \Omega(i\{\mu\}) - \Omega(k\{\mu'\})\langle i\{\mu\}\|\,y\,\|k\{\mu'\}\rangle \\ &\quad\times\langle j\{\mu\}\|\,y\,\|l\{\mu'\}\rangle^* K^2_{kl}(\{\mu'\})\,,\end{aligned} \tag{4.8a}$$

where

$$N_y(j) = \langle j\{\mu\}|\hat{N}_y|\,j\{\mu\}\rangle \tag{4.8b}$$

has the value given by Eq. (4.6b).

The reduced matrix elements appearing in this equation can be expressed as

$$\begin{aligned}\langle[\lambda]\{-b'\}\rho'\{\mu'\}\|\,y\,\|[\lambda]\{-b\}\rho\{\mu\}\rangle &= U(\{\lambda\}\{-b\}\{\mu'\}\{-1-1\};\{\mu\}\rho\_\{-b'\}\_\rho') \\ &\quad\times\langle\{-b'\}\|\,y\,\|\{-b\}\rangle\,,\end{aligned} \tag{4.9}$$

where $U(\{\lambda\}\{-b\}\{\mu'\}\{-1-1\};\{\mu\}\rho\_\{-b'\}\_\rho')$ is a u(n) Racah coefficient (with multiplicity indices in Hecht *et al.* notation[49]) and $\langle\{-b'\}\|\,y\,\|\{-b\}\rangle$ is a reduced matrix element between simple Bargmann polynomials. Its nonzero values are given[34] by

$$\begin{aligned}\langle\{-b-\Delta^{(11)}(k)\|\,y\,\|\{-b\}\rangle &= \Bigg[(b_{2k-1}+n-2k+1) \\ &\quad\times\prod_{\substack{j=1\\ j\neq k}}^{m}\left(\frac{b_{2k-1}-b_{2j-1}+2j-2k-1}{b_{2k-1}-b_{2j-1}+2j-2k+1}\right)\Bigg]^{1/2},\end{aligned} \tag{4.10}$$

where $\Delta^{(11)}(k)$ is the partition $(0\cdots0110\cdots0)$ with null entries everywhere except for unity in its $(2k-1)$th and $(2k)$th entries.

Starting from the $\{\mu=\lambda\}$ highest-weight u(n) subirrep, which is always multiplicity-free and for which $K(\{\lambda\}) = \mathbf{1}$, the recursion relation (4.8a) easily gives all the $K(\{\mu\})$ matrices. Finally, one evaluates the so($2n$) matrix elements

$$\begin{aligned}\langle[\lambda]\{-b'\}\rho'\{\mu'\}\|\gamma(B)\|[\lambda]\{-b\}\rho\{\mu\}\rangle &= \langle[\lambda]\{-b'\}\rho'\{\mu'\}\|K(\{\mu'\})yK(\{\mu\})^{-1} \\ &\quad\|[\lambda]\{-b\}\rho\{\mu\}\rangle\,.\end{aligned} \tag{4.11}$$

It is important to note that, although multiplicities occur in general in the so(n)$\supset$u(n) reduction, many and often most of the u(n) states are, in fact, multiplicity-free for $[\lambda]$ irreps of physical interest. In particular, the highest-weight and the $N_y = 1$ u(n) states are always multiplicity-free. For multiplicity-free states, the ratios of the one-dimensional K matrices required to evaluate Eq. (4.11) are given immediately by

$$[K(\{\mu'\})/K(\{\mu\})]^2 = \Omega(\{\mu'\}) - \Omega(\{\mu\})\,. \tag{4.12}$$

Reduced matrix elements of $\gamma(A)$ are given by complex conjugation and the u(n) matrix elements of $\gamma(C)$ are already known.[15,19,21] Thus all the matrix elements of so($2n$) are expressed in terms of the matrix elements (4.10) and u(n) Racah coefficients.

Consider, for example, so(6). An so(6) irrep is labeled by a triplet of integer or half-integer numbers $[\lambda_1\lambda_2\lambda_3]$. Polynomials of degree b in the (y_{ij}) variables are tensors of u(3) rank $\{-b-b\}$. The u(3) coupling

$$\{\lambda_1\lambda_2\lambda_3\}\times\{-b-b\}\to\{\mu_1\mu_2\mu_3\}$$

being multiplicity-free, the K submatrices are all one dimensional. The eigenvalues of $\hat{\Omega}$ are given by

$$\Omega(\{\mu\}) = -\frac{1}{2}\sum_{i=1}^{3}\mu_i(\mu_i+4-2i) + \frac{1}{2}b(b+3)\,. \tag{4.13}$$

Hence one determines, for $b' = b+1$, the ratios

$$\left[\frac{K(\{\mu'\})}{K(\{\mu\})}\right]^2 = \left[\sum_{i=1}^{3}(\mu_i - i)\Delta_i\right] + b + 5, \tag{4.14}$$

for $\{\mu'\} = \{\mu-\Delta\}$ with $\Delta = (110)$, (101), or (011). Although they are not needed for the calculation of matrix elements, one also easily derives the absolute value

$$K^2(\{\mu\}) = \frac{(\lambda_1+\lambda_2+2)!(\lambda_1+\lambda_3+1)!(\lambda_2+\lambda_3)!}{(\mu_1+\mu_2+b+2)!(\mu_1+\mu_3+b+1)!(\mu_2+\mu_3+b)!}\,. \tag{4.15}$$

Equations (4.11), (4.12), and (4.14) immediately give the u(3)-reduced matrix elements for so(6),

$$\langle [\lambda]\{-b'-b'\}\{\mu'\}\|\gamma(B)\|[\lambda]\{-b-b\}\{\mu\}\rangle$$

$$= \frac{K(\{\mu'\})}{K(\{\mu\})} \times U(\{\lambda\}\{-b-b\}\{\mu'\}\{-1-1\};\{\mu\}\{-b'-b'\})\sqrt{b+1} \tag{4.16a}$$

$$= \left[\left[\left(\sum_{\substack{k=1\\k\neq l}}^{3} p^{\mu}_{k3} + b - 1\right)\right] \frac{\Pi_{k=1}^{3}(p^{\mu}_{l3} - p^{\lambda}_{k3} + b + 1)}{\Pi^{3}_{\substack{k=1\\k\neq l}}(p^{\mu}_{l3} - p^{\mu}_{k3} + 1)}\right]^{1/2}, \quad \{\mu'\} = \{\mu - \Delta\}, \quad \Delta = (1-\delta_{1l}, 1-\delta_{2l}, 1-\delta_{3l}), \tag{4.16b}$$

in which we have inserted

$$\langle\{0-b'-b'\}\|y\|\{0-b-b\}\rangle = \sqrt{b+1},$$

where the hooks p are defined by

$$p^{\mu}_{in} = \mu_{in} + n - i, \quad \mu_{in} \equiv \mu_i. \tag{4.17}$$

Since

$$\sum_{i<j=1}^{3} A_{ij}B_{ij} = \frac{1}{2}\sum_{\alpha<\beta=1}^{6} L_{\alpha\beta} - \frac{1}{2}\sum_{i,j=1}^{3} C_{ij}C_{ji} + \sum_{k=1}^{3} C_{kk},$$

the matrix elements (4.16b) must obey the sum rule

$$\sum_{\{\mu'\}} \langle [\lambda]\{-b'-b'\}\{\mu'\}\|\gamma(B)\|[\lambda]\{-b-b\}\{\mu\}\rangle^2 \times \left[\frac{\dim\{\mu'\}}{\dim\{\mu\}}\right] = \frac{1}{2}\sum_{k=1}^{3}\lambda_k(\lambda_k + 6 - 2k) - \frac{1}{2}\sum_{k=1}^{3}\mu_k(\mu_k + 4 - 2k) + \frac{1}{2}\sum_{k=1}^{3}\mu_k,$$

where, from Weyl,[50]

$$\dim\{\mu\} = \frac{\Pi_{r<s}^{n}(p^{\mu}_{rn} - p^{\mu}_{sn})}{1!2!\cdots(n-1)!}.$$

Such a sum rule is easily verified using complex function residue theory.[34]

B. The matrix representations of so(2n+1)

With $u(n) \subset so(2n+1)$ chosen as the stability subalgebra, the remaining raising operators of $so(2n+1)$ now include both the $so(2n)$ raising operators (A_{ij}), which we have noted transform as the components of a $u(n)$ $\{11\}$ tensor, and, in addition, the set $(\mathscr{A}_i)$, which transform as the components of a $u(n)$ $\{1\}$ tensor. Whereas the former span an Abelian raising operator algebra, the latter are not commutative. The K-matrix theory of Sec. III C therefore applies.

The intermediate algebra I sitting between $so(2n+1)$ and its stability subalgebra is $so(2n)$ and we can classify a $so(2n+1)$ basis by the labels of the subalgebra chain

$$\begin{array}{ccccccc} so(2n+1) & \supset & so(2n) & \supset & u(n) \\ \langle\sigma\rangle & & \{-w\} & [\lambda] & \{-b\}\rho & \{\mu\} \end{array} \tag{4.18}$$

[see also Eqs. (4.19) and (4.21)].

One recalls the well-known fact[51] that the reduction of an $so(2n+1)$ irrep $\langle\sigma\rangle$ into a sum of $so(2n)$irreps is multiplicity-free and given by the branching rule

$$so(2n+1)\downarrow so(2n): \langle\sigma\rangle\downarrow\sum_{w}[\sigma/w], \tag{4.19}$$

where $[\sigma/w]$ denotes a partition obtained by factoring out the fully symmetric Young tableaux $w = \{w0\cdots0\}$ from the tableau $[\sigma]$.

From the perspective of VCS theory, we have found that the only $so(2n)$ subirreps that can appear in the $so(2n+1)$ irrep $\langle\sigma\rangle$ are those that can be obtained by $u(n)$ coupling a $u(n)$ highest-weight state $|\{\sigma\}\rangle$ to a polynomial in (z), i.e., $[X^{\{-w\}}(z)\times|\{\sigma\}\rangle]^{\{\lambda\}}$, where $\{-w\} = \{0\cdots0-w\}$ and w is the degree of the polynomial. This result is seen to be consistent with and to naturally exhibit the significance of the branching rule (4.19). One also finds, and this is a particular strength of VCS theory, that the $\{\lambda\}\in\{-w\}\otimes\{\sigma\}$ coupled vector-valued wave functions that do not occur in the factorization $[\sigma/w]$ are automatically annihilated by the K-matrix operator on mapping the Bargmann space of vector-valued wave functions onto the irreducible VCS space.

For the chosen $so(2n+1)$ basis, Eqs. (2.22) and (2.28), we easily determine that

$$\Gamma^{(0)}(\mathscr{B}_i) = [\hat{\Lambda}, z_i], \quad \Gamma^{(1)}(B_{ij}) = [\hat{\Lambda}, y_{ij}], \tag{4.20a}$$

with

$$\hat{\Lambda} = \tfrac{1}{4}\hat{I}(\mathscr{C}^{(z)}) + \tfrac{1}{2}\hat{I}(\mathscr{C}^{(y)}) - \tfrac{1}{2}\hat{I}(\gamma(C)) + \tfrac{1}{4}n\hat{N}_z, \tag{4.20b}$$

in accord with Eq. (3.61).

Following Eq. (3.39), we define an orthonormal Bargmann basis of states parametrized by a set of labels reducing the $so(2n+1)\supset so(2n)\supset u(n)$ subalgebra chain:

$$|\langle\sigma\rangle\{-w\}[\lambda]\{-b\}\rho\{\mu\}\nu\rangle = [Y^{\{-b\}}(y)\times[Z^{\{-w\}}(z)\times|\{\sigma\}\rangle]^{\{\lambda\}}]^{\{\lambda\}}_{\nu}. \tag{4.21}$$

The operator $\hat{\Lambda}$ is diagonal in this basis and has eigenvalues

$$\Lambda(\{-w\}\{-b\}\{\mu\}) = \tfrac{1}{4}I(\{-w\}) + \tfrac{1}{2}I(\{-b\}) - \tfrac{1}{2}I(\{\mu\}) + \tfrac{1}{4}nw. \tag{4.20'}$$

Since the $so(2n+1)\downarrow so(2n)$ reduction is multiplicity-free, the $K(\{\lambda\})$ submatrices are one dimensional. Equation (3.60) becomes

$$\frac{\mathscr{K}(\{\lambda'\})^2}{\mathscr{K}(\{\lambda\})^2}=\Big\{[\Lambda(\{-(w+1)\}\{0\}[\{\lambda'\}-\Lambda(\{-w\}\{0\}\{\lambda\})]$$
$$-\frac{1}{2}\sum_{\lambda''}\Big[\frac{\langle\langle\sigma\rangle\{-(w+1)\}[\lambda']\{0\}\{\lambda'\}\|[\gamma(C)\times z\times z]^{\{-1-1\}}\|\langle\sigma\rangle\{-(w-1)\}[\lambda'']\{0\}\{\lambda''\}\rangle}{[\Lambda(\{-(w-1)\}\{-1-1\}\{\lambda'\})-\Lambda(\{-(w-1)\}\{0\}\{\lambda''\})]}$$
$$\times\frac{\langle\sigma\rangle\{-(w-1)\}[\lambda'']\{-1-1\}\{\lambda'\}\|[y\times\partial]^{\{-1\}}\|\langle\sigma\rangle\{-w\}[\lambda]\{0\}\{\lambda\}\rangle}{\langle\langle\sigma\rangle\{-(w+1)\}[\lambda']\{0\}\{\lambda'\}\|z\|\langle\sigma\rangle\{-w\}[\lambda]\{0\}\{\lambda\}\rangle}\Big]\Big\}, \tag{4.22}$$

where we have used the fact that

$$[z\times z\times\mathscr{C}^{(z)}]^{\{-1-1\}}=[z\times z\times z\times\partial]^{\{-1-1\}}=0.$$

Using standard definitions and notation for the u(n) Racah coefficients,[49] the various terms in Eq. (4.22) can be compared as follows:

$$\langle\langle\sigma\rangle\{-(w+1)\}[\lambda']\{0\}\{\lambda'\}\|z\|\langle\sigma\rangle\{-w\}[\lambda]\{0\}\{\lambda\}\rangle=U(\{\sigma\}\{-w\}\{\lambda'\}\{-1\};\{\lambda\}\{-(w+1)\})\sqrt{w+1}, \tag{4.23a}$$

$$\langle\langle\sigma\rangle\{-(w'+2)\}[\lambda']\{0\}\{\lambda'\}\|[\gamma(C)\times z\times z]^{\{-1-1\}}\|\langle\sigma\rangle\{-w'\}[\lambda'']\{0\}\{\lambda''\}\rangle$$
$$=U(\{\lambda''\}\{-2\}\{\lambda'\}\{-1-1\};\{\lambda'\}\_(\rho=1)\{-1-1\}\_)$$
$$\times U(\{\sigma\}\{-w'\}\{\lambda'\}\{-2\};\{\lambda''\}\{-(w'+2)\})\langle\{\lambda''\}\|\gamma(C)\|\{\lambda''\}\rangle\times[(w'+2)!/w'!]^{1/2}, \tag{4.23b}$$

$$\langle\langle\sigma\rangle\{-(w-1)\}[\lambda'']\{-1-1\}\{\lambda'\}\|[y\times\partial]^{\{-1\}}\|\langle\sigma\rangle\{-w\}[\lambda]\{0\}\{\lambda\}\rangle$$
$$=U(\{\lambda\}\{1\}\{\lambda'\}\{-1-1\};\{\lambda''\}\{-1\})\times\langle\langle\sigma\rangle\{-(w-1)\}\{\lambda''\}\{0\}\{\lambda''\}\|\partial\|\langle\sigma\rangle\{-w\}[\lambda]\{0\}\{\lambda\}\rangle$$
$$=(-1)^{\phi(\{\lambda\})+\phi(\{1\})-\phi(\{\lambda''\})}[\dim\{\lambda\}/\dim\{\lambda''\}]^{1/2}\times U(\{\lambda\}\{1\}\{\lambda'\}\{-1-1\};\{\lambda''\}\{-1\})$$
$$\times\langle\langle\sigma\rangle\{-w\}[\lambda]\{0\}\{\lambda\}\|z\|\langle\sigma\rangle\{-(w-1)\}[\lambda'']\{0\}\{\lambda''\}\rangle, \tag{4.23c}$$

where[18]

$$\phi(\{h\})=\frac{1}{2}\sum_{i=1}^{n}(n+1-2i)h_i.$$

Most of the U(n) Racah coefficients in (4.23) are of the multiplicity-free type calculated by Le Blanc and Hecht[19] [see Eqs. (A9), (A10), (A13), (A14), and (A27) of Appendix A] or are of the stretched type in the 23 position, or can be put into one of these forms using symmetry properties discussed in the Appendix of Hecht *et al.*[18]). Such symmetries are illustrated for SU(3) in Appendix A of Hecht *et al.*[49] Only the Racah coefficient

$$U(\{\lambda''\}\{-2\}\{\lambda'\}\{-1-1\};\{\lambda'\}\_(\rho=1)\{-1-1\}\_\_)$$

cannot be found in the literature but it can easily be computed from its definition using known Wigner coefficients of the elementary type[19,52] and the matrix elements of the u(n) generators.[18,21] We find for U(3),

$$U(\{\lambda''\}\{-2\}\{\lambda'\}\{-1-1\};\{\lambda'\}\_(\rho=1)\{-1-1\}\_\_)$$
$$=-\tfrac{1}{4}[[3/g(\lambda')](p_{j_1 3}^{\lambda'}-p_{j_2 3}^{\lambda'}-1)\times(p_{j_1 3}^{\lambda'}-p_{j_2 3}^{\lambda'}+1)]^{1/2} \tag{4.24a}$$

for $\{\lambda''\}=\{\lambda'+\Delta^{(1_2)}(j_1,j_2)\}$, where

$$\Delta^{(1_2)}(1,2)=(110),\quad \Delta^{(1_2)}(1,3)=(101),\quad \Delta^{(1_2)}(2,3)=(011), \tag{4.24b}$$

and where

$$g(\lambda')=(\lambda_1')^2+(\lambda_2')^2+(\lambda_3')^2-\lambda_1'\lambda_2'-\lambda_1'\lambda_3'-\lambda_2'\lambda_3'+3\lambda_1'-3\lambda_3'. \tag{4.25}$$

Also for U(3), we find

$$\langle\{\lambda'\}\|\gamma(C)\|\{\lambda'\}\rangle=[g(\lambda')/3]^{1/2}. \tag{4.26}$$

We find

$$[\gamma(C)\times z\times z]^{\{-1-1\}}=\sqrt{2}[\gamma(C)\times Z^{\{-2\}}(z)]^{\{-1-1\}}$$
$$=(-1)^{n+\omega(\{10-1\},\mathrm{lw})}\,2\sqrt{(n+1)}\times[\gamma(C)\times z\times z]^{\{-1-1\}}, \tag{4.27}$$

where $\omega(\{10-1\},\mathrm{lw})$ is a conjugation phase defined in the Appendix of Hecht *et al.*[18] We also find

$$[z\times\partial]^{\{-1\}}=(-1)^n\sqrt{(n-1)}\,[z\times\partial]^{\{-1\}}. \tag{4.28}$$

When the various expressions in Eqs. (4.23)–(4.28) are inserted in the recursion formula (4.22) for $\mathscr{K}(\{\lambda\})$, we find, for $n=2,3$,

$$\frac{\mathscr{K}(\{\lambda-\Delta^{(1)}(i)\})^2}{\mathscr{K}(\{\lambda\})^2}=\frac{\Pi_{k=1}^{n}(p_{kn}^{\sigma}+p_{in}^{\lambda})}{\Pi_{k=1,\,k\neq i}^{n}(p_{kn}^{\lambda}+p_{in}^{\lambda})}, \tag{4.29}$$

where $\Delta^{(1)}(i)$ is a null n-vector except for unity in its ith entry. While Eq. (4.24) has been verified for so(5) ($n=2$) and so(7) ($n=3$), we surmise that it holds for all n; work is in progress to test its validity using various selection and sum

rules. It respects the Gel'fand betweenness condition

$$\sigma_n \geqslant \lambda_n \geqslant -\sigma_n ,$$

i.e., $\mathscr{K}(\{\lambda - \Delta^{(i)}(n)\})$ vanishes for $\lambda_n = -\sigma_n$, and this is an example of the above-mentioned property of the K operator for Bargmann states with null image on the VCS basis.

Equation (4.24) has also been verified independently for so(5) by one of us[53] using cumbersome but straightforward commutator algebra, a technique not easily applied to Lie algebras of higher rank. It also has been verified[54] again for so(5) by Hecht and Suzuki [one of us (K.T.H.) would like to point out that some of the expressions given in this paper were based on the assumption that the operator K is Hermitian and are therefore, strictly speaking, incorrect; as shown in Sec. III C, K is not Hermitian if one requires both the VCS and the Bargmann basis to reduce the $so(5) \supset so(4)$ subalgebras] using vector coherent state expansions, K-matrix theory, and shift operator techniques; although this approach yields the correct answer, its use is also limited to Lie algebras of low rank as the construction of the shift operators becomes rapidly involved when considering higher rank Lie algebras (Pang and Hecht[55]).

Having determined the $\mathscr{K}(\{\lambda'\})/\mathscr{K}(\{\lambda\})$ ratios, the $u(n)$-reduced matrix elements for $\gamma(\mathscr{B})$ [and, hence, through Hermitian conjugation, for $\gamma(\mathscr{A})$] are given for $[\lambda'] < [\lambda]$ highest-weight $u(n)$ states from Eq. (3.45) by

$$\langle\langle\sigma\rangle\{-(w+1)\}[\lambda']\{0\}\{\lambda'\}\|\gamma(\mathscr{B})\|\langle\sigma\rangle\{-w\}[\lambda]\{0\}\{\lambda\}\rangle$$
$$= [\mathscr{K}(\{\lambda'\})/\mathscr{K}(\{\lambda\})] \times U(\{\sigma\}\{-w\}\{\lambda'\}\{-1\};\{\lambda\}\{-(w+1)\}) \times \sqrt{w+1} . \tag{4.30}$$

If the $so(2n) \supset u(n)$ reduced Wigner coefficients are available, one can derive from this expression (or its Hermitian conjugate for $[\lambda'] > [\lambda]$) the $so(2n)$ (triple reduced) matrix elements

$$\langle[\lambda']\| \|\mathscr{F}\| \|[\lambda]\rangle ,$$

cf. Eq. (3.45). Hence other $u(n)$-reduced matrix elements are given by

$$\langle\langle\sigma\rangle\{-w'\}[\lambda']i'\{\mu'\}\|\gamma(\mathscr{B})\|\langle\sigma\rangle\{-w\}[\lambda]i\{\mu\}\rangle = \langle[\lambda]i\{\mu\};[1]\{-1\}\|[\lambda']i'\{\mu'\}\rangle\langle[\lambda']\| \|\mathscr{F}\| \|[\lambda]\rangle ,$$

where $i \sim (\{-b\}\rho)$ and

$$\langle[\lambda]i\{\mu\};[1]\{-1\}\|[\lambda']i'\{\mu'\}\rangle$$

is a $so(2n) \supset u(n)$ reduced Wigner coefficient.

If the $so(2n) \supset u(n)$ reduced Wigner coefficients are not available, one can use the following alternative expressions (see Le Blanc and Rowe[14] and Hecht *et al.*[15] for their derivation in parallel contexts; see, in particular, Sec. 4.1 of the latter but note that this reference uses a different coupling order); these expressions then offer a method for computing the unknown reduced Wigner coefficients by equating (4.30) and (4.31). The expression for $[\lambda'] < [\lambda]$ is

$$\langle\langle\sigma\rangle\{-(w+1\}[\lambda']\{-b'\}\rho'\{\mu'\}\|\gamma(\mathscr{B})\|\langle\sigma\rangle\{-w\}[\lambda]\{-b\}\rho\{\mu\}\rangle$$
$$= \sum_{\{-\bar{b}\}\rho_i\rho_j} (-1)^{\phi(\{\mu\})-\phi(\{-1\})-\phi(\{\mu'\})}(-1)^{\phi(\{\lambda\})-\phi(\{-1\})-\phi(\{\lambda'\})}$$
$$\times K_{\{-b'\}\rho';\{-\bar{b}\}\rho_j}([\lambda'];\{\mu'\}) \times K^{-1}_{\{-\bar{b}\}\rho_i;\{-b\}\rho}(\{\lambda\};\{\mu\})$$
$$\times U(\{-1\}\{\lambda\}\{\mu'\}\{-\bar{b}\};\{\lambda'\}_{-}\rho_j\{\mu\}\rho_{i-}) \times \langle\langle\sigma\rangle\{-(w+1)\}[\lambda']\{0\}\{\lambda'\}\|\gamma(\mathscr{B})\|\langle\sigma\rangle\{-w\}[\lambda]\{0\}\{\lambda\}\rangle , \tag{4.31a}$$

while, for $[\lambda'] > [\lambda]$,

$$\langle\langle\sigma\rangle\{-(w-1)\}[\lambda']\{-b'\}\rho'\{\mu'\}\|\gamma(\mathscr{B})\|\langle\sigma\rangle\{-w\}[\lambda]\{-b\}\rho\{\mu\}\rangle$$
$$= \sum_{\{-b_i\}\rho_i;\{-b_j\}\rho_j\{\bar{l}\}\bar{\rho}} (-1)^{\phi(\{\mu\})-\phi(\{-1\})-\phi(\{\mu'\})}(-1)^{\phi(\{\lambda\})-\phi(\{-1\})-\phi(\{\bar{l}\})} \times Q([\lambda']\{\bar{l}\})$$
$$\times K_{\{-b'\}\rho';\{-b_j\}\rho_j}([\lambda'];\{\mu'\}) \times K^{-1}_{\{-b_i\}\rho_i;\{-b\}\rho}([\lambda];\{\mu\} \times (\{b_j\}\| y \|\{b_i\})$$
$$\times U(\{\lambda'\}\{-1-1\}\{\mu'\}\{-b_i\};\{\bar{l}\}_{-}\bar{\rho},\{-b_j\}_{-}\rho_j) \times U(\{-1\}\{\lambda\}\{\mu'\}\{-b_i\};\{\bar{l}\}_{-}\bar{\rho},\{\mu\}\rho_{i-}) , \tag{4.31b}$$

where the various K matrices are $so(2n) \supset u(n)$ K matrices and where

$$Q([\lambda']\{\bar{l}\}) = (-1)^{(\phi\{\bar{l}\})+\phi(\{11\})-\phi(\{\lambda'\})}(-1)^{\phi(\{\lambda\})+\phi(\{1\})-\phi(\{\lambda'\})}\left[\frac{\dim\{\lambda\}}{\dim\{\bar{l}\}}\right]^{1/2}$$
$$\times \frac{(-1)^{n-1}\sqrt{n-1}}{K^2_{\{-1-1\};\{-1-1\}}([\lambda'];\{\bar{l}\})} \times U(\{\lambda\}\{-1\}\{\lambda'\}\{11\};\{\bar{l}\},\{1\})$$
$$\times \langle\langle\sigma\rangle\{-w\}[\lambda]\{0\}\{\lambda\}\|\gamma(\mathscr{B})\|\langle\sigma\rangle\{-(w-1)\}[\lambda']\{0\}\{\lambda'\}\rangle . \tag{4.31c}$$

With the help of (4.30) and (4.31), we compute, for example, the following matrix elements for the completely multiplicity-free ($n = 3$) case of so(7): we find, for $[\lambda'] < [\lambda]$,

$$\langle\langle\sigma\rangle\{-(w+1)\}[\lambda-\Delta^{(1)}(i)]\{0-b-b\}\{\mu-\Delta^{(1)}\}(j)\|\gamma(\mathscr{B})\|\langle\sigma\rangle\{-w\}[\lambda]\{0-b-b\}\{\mu\}\rangle$$

$$=(-1)^{i-j}S(i-j)\times\left[\frac{1}{2}\frac{(\Pi^3_{\substack{k=1\\k\neq i}}[p^{\mu}_{j3}-p^{\lambda}_{k3}+b])(\Pi^3_{\substack{k=1\\k\neq j}}[p^{\lambda}_{i3}-p^{\mu}_{k3}-b-1])}{(\Pi^3_{\substack{k=1\\k\neq i}}[p^{\lambda}_{i3}-p^{\lambda}_{k3}])(\Pi^3_{\substack{k=1\\k\neq i}}[p^{\lambda}_{i3}-p^{\lambda}_{k3}-1])}\right.$$

$$\left.\times\frac{(\Pi^3_{\substack{k=1\\k\neq j}}[p^{\mu}_{j3}+p^{\mu}_{k3}+b-1])(\Pi^3_{k=1}[p^{\sigma}_{k3}-p^{\lambda}_{i3}+1])(\Pi^3_{k=1}[p^{\lambda}_{i3}+p^{\sigma}_{k3}])}{(\Pi^3_{\substack{k=1\\k\neq i}}[p^{\lambda}_{i3}+p^{\lambda}_{k3}-1])(\Pi^3_{\substack{k=1\\k\neq j}}[p^{\mu}_{j3}-p^{\mu}_{k3}-1])(\Pi^3_{\substack{k=1\\k\neq i}}[p^{\lambda}_{i3}+p^{\lambda}_{k3}])}\right]^{1/2}, \tag{4.32a}$$

while, for $[\lambda']>[\lambda]$,

$$\langle\langle\sigma\rangle\{-(w-1)\}[\lambda+\Delta^{(1)}(i)]\{0-(b+1)-(b+1)\}\{\mu-\Delta^{(1)}(j)\}\|\gamma(\mathscr{B})\|\langle\sigma\rangle\{-w\}[\lambda]\{0-b-b\}\{\mu\}\rangle$$

$$=(-1)^{i(i-1)/2}S(i-j)\times\left[\frac{1}{2}\frac{(\Pi^3_{\substack{k=1\\k\neq i}}\Pi^3_{\substack{l=1\\l\neq i}}[p^{\lambda}_{i3}-p^{\mu}_{k3}-b-1])([p^{\lambda}_{i3}-p^{\mu}_{j3}-b])}{(\Pi^3_{\substack{k=1\\k\neq i}}[p^{\lambda}_{i3}-p^{\lambda}_{k3}+1])(\Pi^3_{\substack{k=1\\k\neq i}}[p^{\lambda}_{i3}-p^{\lambda}_{k3}])}\right.$$

$$\left.\times\frac{([\Sigma_{\substack{k=1\\k\neq j}}p^{\mu}_{k3}+b])(\Pi^3_{k=1}[p^{\sigma}_{k3}-p^{\lambda}_{i3}])(\Pi^3_{k=1}[p^{\lambda}_{i3}+p^{\sigma}_{k3}+1])}{(\Pi^3_{\substack{k=1\\k\neq i}}[p^{\lambda}_{i3}+p^{\lambda}_{k3}])(\Pi^3_{\substack{k=1\\k\neq i}}[p^{\mu}_{j3}-p^{\mu}_{k3}-1])(\Pi^3_{\substack{k=1\\k\neq i}}[p^{\lambda}_{i3}+p^{\lambda}_{k3}+1])}\right]^{1/2}, \tag{4.32b}$$

where $S(i-j)$ is the sign of $(i-j)$.

V. SUMMARY AND CONCLUDING REMARKS

VCS theory is based on the decomposition (2.1) of the complex extension of Lie algebras into subalgebras of stability, raising and lowering operators. If the lowest-weight state for a ladder representation of a Lie algebra carries a one-dimensional representation of its stability subalgebra, then standard (scalar) coherent state theory applies. Thus VCS theory is an extension of standard coherent state theory to representations induced from nontrivial finite-dimensional representations of stability subalgebras.

Coupled with a change from the coherent state measure to a simpler Bargmann measure, coherent state theory is often regarded as a nonunitary Dyson mapping.[43] In parallel, VCS theory gives nonunitary mappings to generalized vector Dyson representations. However, for computational purposes, one needs to preserve the inner product. While the coherent state and vector coherent state inner products are known in integral form,[3] in terms of the invariant measure of the stability subgroup, they are cumbersome and difficult to use in practice. Therefore K-matrix theory was developed as a practical alternative for the construction of inner products. It may be regarded as a second mapping that restores the unitarity of the coherent state mapping on change to the Bargmann measure. Thus K-matrix theory generalizes, to highly nontrivial situations, the familiar mappings from Dyson to Holstein–Primakoff representations.

Since the original formulation of K-matrix theory was restricted to situations in which the raising operator subalgebra is Abelian, it did not apply, for example, to the odd orthogonal $so(2n+1)$ Lie algebras with $u(n)$ as stability subalgebra. We have therefore analyzed the underlying structure of K-matrix theory and shown that it is not fundamentally limited in this way. In particular, we have extended its domain of applicability to situations in which the raising operator subalgebra is not Abelian and is the sum of two irreducible subspaces [cf. Eq. (3.27)] under the adjoint action of the stability subalgebra. With this extension, VCS and K-matrix theory together provide algorithms for the construction of the matrix representations of all four series of classical Lie algebras with $u(n)$ [equivalently $gl(n)$] as stability subalgebra. The extension also provides the framework for the completely general situation in which the raising operator algebra is a sum of more than two irreducible subspaces, i.e,

$$\mathbf{n}_{\pm}=\mathbf{n}^1_{\pm}+\mathbf{n}^2_{\pm}+\cdots,$$

which is considered in Ref. 22.

A remarkable property of K-matrix theory is that the square of the similarity transformation, which restores the unitarity of the Dyson mapping, obeys a linear recursion relation. Furthermore, one is able to make full use of the tensor structure of the Lie algebra and its representations in order to exploit the symmetry properties of the K matrix. In particular, the invariance of the K matrix under the stability subalgebra (consequence of Proposition 4) results in it being block diagonal. As a consequence, the algorithm for calculating matrix representations is remarkably simple even for high-dimensional representations. In the absence of multiplicity, one obtains analytical solutions in terms of the Wigner and Racah coefficients of the stability subalgebra. In the presence of multiplicities, one has to solve linear recursion relations for the submatrices of the block-diagonal K matrix, where the dimension of a given block is equal to the multiplicity of occurrences of the corresponding irreducible representation of the stability subalgebra.

Another remarkable property of the theory is that it gives a natural interpretation of the various kinds of multiplicity that can occur in the reduction of a Lie algebra ladder representation under restriction to its stability subalgebra [cf. discussion following Eq. (3.11)]. One finds that the

multiplicities of the algebra are naturally expressed in terms of the outer product multiplicities of the stability subalgebra. Thus one finds that all the properties of an irreducible ladder representation of a Lie algebra are related to those of the tensor algebra of the stability subalgebra.

In view of this observaton, it is particularly significant therefore that all the classical Lie algebras have VCS representations with $\mathrm{u}(n)$ or, equivalently, $\mathrm{gl}(n)$ as stability subalgebra since the $\mathrm{u}(n)$ tensor algebra and its Wigner–Racah calculus are the most studied and best known.[19,52,56] It is also significant, that K-matrix theory has been found to play an essential role in this calculus.[14,15,19] Evidently VCS and K-matrix theory is playing a pervasive role throughout Lie algebra representation theory and it is our belief that one has only begun to appreciate its full power and versatility.

APPENDIX A: THE $\hat{\Omega}$ OPERATOR FOR ABELIAN RAISING OPERATOR ALGEBRAS

We show that for $\mathfrak{n}_\pm$ Abelian and $\mathfrak{h}$-invariant, $\Gamma(B_\nu)$ can be expressed in the form

$$\Gamma(B_\nu) = [\hat{\Omega}, z_\nu] . \tag{A1}$$

From its definition (2.7),

$$\Gamma(B_\nu) = z_\mu c^i_{\mu\,-\nu}(\mathbb{C}_i + \tfrac{1}{2}\mathscr{C}_i) . \tag{A2}$$

Recall that

$$\mathscr{C}_i = c^\nu_{\mu i}\, z_\mu\, \partial_\nu = K^{\nu,-\rho} c_{-\rho\mu i}\, z_\mu\, \partial_\nu = k^{-1} c_{-\nu\mu i}\, z_\mu\, \partial_\nu ,$$

where $K_{\nu,-\rho} \equiv k\delta_{\nu\rho}$ is given by Eq. (2.19) and

$$c_{\alpha\beta\gamma} = K_{\alpha\lambda} c^\lambda_{\beta\gamma}$$

is the fully antisymmetric form of the structure constants. It follows that

$$\mathscr{C}\cdot\mathbb{C} = K^{ij}\mathscr{C}_j\mathbb{C}_i \tag{A3}$$

can be expressed

$$\mathscr{C}\cdot\mathbb{C} = -k^{-1}\mathbb{C}_i c^i_{\mu\,-\nu}\, z_\mu\, \partial_\nu .$$

We therefore find that

$$c^i_{\mu\,-\nu}\, z_\mu \mathbb{C}_i = -\tfrac{1}{2}k\,[\gamma(C\cdot C) - \mathscr{C}\cdot\mathscr{C}, z_\nu] .$$

For the second term of $\Gamma(B_\nu)$, we can write

$$\tfrac{1}{2}c^i_{\mu\,-\nu}\, z_\mu \mathscr{C}_i = \tfrac{1}{4}c^i_{\mu\,-\nu}\, z_\mu \mathscr{C}_i + \tfrac{1}{4}c^i_{\mu\,-\nu}\, \mathscr{C}_i z_\mu + \tfrac{1}{4}c^i_{\mu\,-\nu}[z_\mu, \mathscr{C}_i] .$$

Then, since

$$\tfrac{1}{4}c^i_{\mu\,-\nu} z_\mu \mathscr{C}_i = -\tfrac{1}{4}k\,[\mathscr{C}^i, z_\nu]\mathscr{C}_i ,$$

and

$$\tfrac{1}{4}c^i_{\mu\,-\nu}[z_\mu, \mathscr{C}_i] = -\tfrac{1}{4}c^i_{\mu\,-\nu}\, c^{\mu}_{\rho i}\, z_\rho = \tfrac{1}{8}k z_\nu ,$$

Eq. (A1) follows with

$$\hat{\Omega} = \tfrac{1}{4}k\,(\mathscr{C}\cdot\mathscr{C} - 2\gamma(C\cdot C) + \tfrac{1}{2}\hat{N}_z) , \tag{A4}$$

where

$$\hat{N}_z = z_\nu\, \partial_\nu \tag{A5}$$

measures the degree of a wave function in (z_ν).

The value of k is easily determined from

$$k = 2W^i_i ,$$

where

$$W^j_i = c^{\,j}_{-\nu\rho} c^{\,\rho}_{\nu i} \quad \text{(no sum on } \nu\text{)}$$

is defined by the identity

$$[B_\nu, [A_\nu, C_i]] = C_j W^j_i \quad \text{(no sum on } \nu\text{)}.$$

Thus for the basis (2.22) for $\mathrm{so}(2n)$, we easily determine, for example,

$$k = 4(n-1).$$

APPENDIX B: THE $\hat{\Omega}$ OPERATORS FOR NON-ABELIAN RAISING OPERATOR ALGEBRAS

We easily determine that the (y_ν)- and (∇_ν)-independent component $\Gamma^{(0)}(\mathscr{B}_\alpha)$ of $\Gamma(\mathscr{B}_\alpha)$ is given by

$$\Gamma^{(0)}(\mathscr{B}_\alpha) = z_\beta c^i_{\beta\,-\alpha}\,[\mathbb{C}_i + \tfrac{1}{2}\mathscr{C}^{(z)}_i] . \tag{B1}$$

Thus, in parallel with the derivation of Eq. (A1), we obtain

$$\Gamma^{(0)}(\mathscr{B}_\alpha) = [\hat{\Omega}, z_\alpha] \tag{B2}$$

with

$$\hat{\Omega} = \tfrac{1}{4}k\,(\mathscr{C}^{(z)}\cdot\mathscr{C}^{(z)} - 2\gamma(C\cdot C) + \tfrac{1}{2}\hat{N}_z) , \tag{B3}$$

where

$$\hat{N}_z = z_\alpha\, \partial_\alpha \tag{B4}$$

measures the degree of a polynomial in (z_α) and now $k = K_{\alpha,-\alpha}$.

The component $\Gamma^{(1)}(B_\nu)$ of $\Gamma(B_\nu)$ that is linear in (y_ν) and independent of (∇_ν) is given by

$$\Gamma^{(1)}(B_\nu) = y_\mu c^i_{\mu\,-\nu}[\mathbb{C}_i + \mathscr{C}^{(z)}_i] \tag{B5}$$

and is likewise expressed as

$$\Gamma^{(1)}(B_\nu) = [\hat{\Omega}', y_\nu] , \tag{B6}$$

where

$$\hat{\Omega}' = \tfrac{1}{2}k'(\mathscr{C}^{(y)}\cdot\mathscr{C}^{(y)} - \gamma(C\cdot C)) \tag{B7}$$

with $k' = K_{\nu,-\nu}$.

[1] J. Deenen and C. Quesne, J. Math. Phys. **25**, 1638, 2354 (1984); **26**, 2705 (1985); C. Quesne, *ibid.* **27**, 428, 869 (1986).
[2] D. J. Rowe, J. Math. Phys. **25**, 2662 (1984); "Coherent states, contractions and classical limits of the non-compact symplectic groups," in *Proceedings of the XIII International Colloquium on Group Theoretical Methods in Physics*, edited by W. W. Zachary (World Scientific, Singapore, 1984).
[3] D. J. Rowe, G. Rosensteel, and R. Gilmore, J. Math. Phys. **26**, 2787 (1985).
[4] F. J. Dyson, Phys. Rev. **102**, 1217 (1956).
[5] T. Holstein and H. Primakoff, Phys. Rev. **58**, 1098 (1940).
[6] D. J. Rowe, G. Rosensteel, and R. Carr, J. Phys. A: Math. Gen. **17**, L399 (1984).
[7] O. Castaños, E. Chacòn, and M. Moshinsky, J. Math. Phys. **25**, 1211 (1984).
[8] J. Deenen and C. Quesne, J. Phys. A: Math. Gen. **17**, L404 (1984).
[9] K. T. Hecht and J. P. Elliott, Nucl. Phys. A **438**, 29 (1985).
[10] K. T. Hecht, Nucl. Phys. A **444**, 189 (1985).
[11] O. Castaños, E. Chacòn, M. Moshinsky, and C. Quesne, J. Math. Phys. **26**, 2107 (1985); O. Castaños, P. Kramer, and M. Moshinsky, J. Math. Phys. **27**, 924 (1986).
[12] M. Kashiwara and M. Vergne, Invent. Math. **44**, 1 (1978); M. Moshinsky and C. Quesne, J. Math. Phys. **12**, 1772 (1971).
[13] D. J. Rowe, B. G. Wybourne, and P. Butler, J. Phys. A: Math. Gen. **18**, 939 (1985).
[14] R. Le Blanc and D. J. Rowe, J. Phys. A: Math. Gen. **18**, 1891, 1905 (1985).
[15] K. T. Hecht, R. Le Blanc, and D. J. Rowe, J. Phys. A: Math. Gen. **20**, 257 (1987).

[16]D. J. Rowe and J. Carvalho, Phys. Lett. **175**, 243 (1986).
[17]R. Le Blanc and D. J. Rowe, J. Phys. A: Math. Gen. **19**, 1111 (1986).
[18]K. T. Hecht, R. Le Blanc, and D. J. Rowe, J. Phys. A: Math. Gen. **20**, 2241 (1987).
[19]R. Le Blanc and K. T. Hecht, J. Phys. A: Math. Gen. **20**, 4613 (1987).
[20]I. M. Gel'fand and M. L. Tsetlin, Dokl. Akad. Nauk. **71**, 825, 1017 (1950).
[21]G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).
[22]R. Le Blanc and D. J. Rowe, "The matrix representations of g_2. I. Representations in an so(4) basis," J. Math. Phys. (in press); "The matrix representations of g_2. II. Representations in an su(3) basis," J. Math. Phys. (in press).
[23]Harish-Chandra, Am. J. Math. **78**, 1 (1956).
[24]R. Godement, *Seminaire Cartan* (Ecole Normale Superieure, Paris, 1958).
[25]C. L. Siegel, Am. J. Math. **65**, 1 (1943).
[26]G. Rosensteel and D. J. Rowe, Int. J. Theor. Phys. **16**, 63 (1977).
[27]D. J. Rowe, Rep. Prog. Phys. **48**, 1419 (1985).
[28]G. Rosensteel and D. J. Rowe, Phys. Rev. Lett. **47**, 223 (1981).
[29]D. J. Rowe and G. Rosensteel, Phys. Rev. C **25**, 3236 (1982); J. Deenen and C. Quesne, J. Math. Phys. **23**, 878, 2004 (1982); J. Deenen and C. Quesne, J. Phys. A: Math. Gen. **16**, 2095 (1983); P. Kramer, Ann. Phys. (NY) **141**, 254, 259 (1982); M. Moshinsky, J. Math. Phys. **25**, 1555 (1984).
[30]G. Rosensteel and D. J. Rowe, J. Math. Phys. **24**, 2461 (1983).
[31]R. Le Blanc, C. Carvalho, and D. J. Rowe, Nucl. Phys. A **414**, 155 (1984); Y. Suzuki, *ibid.* **448**, 395 (1986); Y. Suzuki and K. T. Hecht, Nucl. Phys. A **455**, 315 (1986).
[32]J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Phys. Rev. **108**, 1175 (1957).
[33]A. Arima and F. Iachello, Phys. Rev. Lett. **35**, 1069 (1975).
[34]R. Le Blanc and D. J. Rowe, J. Phys. A: Math. Gen. **20**, L681 (1987).
[35]C. Quesne, J. Math. Phys. **22**, 1482 (1981).
[36]V. Bargmann, Commun. Pure Appl. Math. **14**, 187 (1961); I. Segal, *Mathematical Problems of Relativistic Physics* (Am. Math. Soc., Providence, RI, 1963).
[37]A. M. Perelomov, Commun. Math. Phys. **26**, 222 (1972).
[38]E. Onofri, J. Math. Phys. **16**, 1087 (1975).
[39]J. R. Klauder and B. S. Skagerstam, *Coherent States* (World Scientific, Singapore, 1985).
[40]R. J. Glauber, Phys. Rev. **131**, 2766 (1963).
[41]D. J. Rowe, Can. J. Phys. **56**, 442 (1978).
[42]D. J. Rowe and R. Basserman, Can. J. Phys. **54**, 1941 (1976); D. J. Rowe, A. Ryman, and G. Rosensteel, Phys. Rev. A **22**, 2362 (1980); P. Kramer and M. Saraceno, *Lecture Notes in Physics*, Vol. 140 (Springer, Berlin, 1981).
[43]J. Dobaczewski, Nucl. Phys. A **369**, 213, 237 (1981); **380**, 1 (1982).
[44]D. J. Rowe, P. Rochford, and R. Le Blanc, Nucl. Phys. A **464**, 39 (1987).
[45]W. Schmid, Ann. Math. **93**, 1 (1971).
[46]H. Fukutome, M. Yamamura, and S. Nishiyama, Prog. Theor. Phys. **57**, 1554 (1977); B. R. Judd, in *Group Theory and its Applications*, edited by E. M. Loebl (Academic, New York, 1968).
[47]J. Broeckhove and E. Deumens, Z. Phys. A **292**, 243 (1979); V. S. Vasilevski, Y. F. Smirnov, and G. Filippov, Sov. J. Nucl. Phys. **32**, 510 (1980); M. G. Vassanji and D. J. Rowe, Phys. Lett. **115** B 77 (1982).
[48]G. de B. Robinson, *Representation Theory of the Symmetric Group* (University of Toronto, Toronto, 1961).
[49]K. T. Hecht, E. J. Reske, T. H. Seligman, and W. Zahn, Nucl. Phys. A **356**, 146 (1981).
[50]H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover, New York, 1931).
[51]R. C. King, J. Phys. A: Math. Gen. **18**, 429 (1975).
[52]L. C. Biedenharn and J. D. Louck, Comm. Math. Phys. **8**, 89 (1968).
[53]K. T. Hecht, Nucl. Phys. **63**, 177 (1965).
[54]K. T. Hecht and Y. Suzuki, J. Phys. A: Math. Gen. **20**, L117 (1987).
[55]S. C. Pang and K. T. Hecht, J. Math. Phys. **8**, 1233 (1967).
[56]J. D. Louck, Am. J. Phys. **38**, 3 (1970).

Lie algebraic solutions of linear Fokker–Planck equations

Fritz Wolf
Fachbereich Physik der Universität Kaiserslautern, D-6750 Kaiserslautern, West Germany

(Received 1 July 1987; accepted for publication 23 September 1987)

Nonstationary solutions of linear Fokker–Planck equations with arbitrary drift and diffusion coefficients are derived. Using Lie algebraic techniques the time-evolution operator is given as a product of exponentials of the differential operators appearing in the equations.

I. INTRODUCTION

In this paper we will be concerned with the Lie algebraic solution of the linear Fokker–Planck equation

$$\frac{\partial P(x,t)}{\partial t} = \mathcal{L}(x,t)P(x,t), \tag{1.1}$$

with

$$\mathcal{L}(x,t) := -\frac{\partial}{\partial x}(a(t)x + b(t)) + c(t)\frac{\partial^2}{\partial x^2}. \tag{1.2}$$

Conventionally, one evaluates the time-ordered exponential

$$U(t) := \exp_+\left(\int_0^t \mathcal{L}(x,t')dt'\right), \tag{1.3}$$

giving a solution of (1.1) in the form

$$P(x,t) = U(t)P(x,0). \tag{1.4}$$

Solutions of (1.1) in the form (1.4) can be found in Ref. 1 for constant coefficients and in Refs. 2–4 for nonconstant functions $a(t)$, $b(t)$, $c(t)$, respectively.

Two drawbacks of this approach are at hand.

(i) The treatment depends on whether the coefficients in (1.2) are constant or not. In the first case the time-ordered exponential reduces to an ordinary one, whereas in the second case one has to evaluate the time ordering explicitly.

(ii) The exponential is not given in a product representation, e.g., in the form

$$U(t) = \exp\left[\alpha(t)\frac{\partial}{\partial x}x\right]\exp\left[\beta(t)\frac{\partial}{\partial x}\right]\exp\left[\gamma(t)\frac{\partial^2}{\partial x^2}\right], \tag{1.5}$$

thus not allowing the easy evaluation of the action of U on $P(x,0)$ and the study of the influence of the single factors on the time evolution of $P(x,t)$.

A first step to (ii) was undertaken in Ref. 2, where U was given as

$$U(t) = \exp_+\left[-\int_0^t \frac{\partial}{\partial x}(a(t')x + b(t'))dt'\right] \times \exp\left[\sigma(t)\frac{\partial^2}{\partial x^2}\right], \tag{1.6}$$

by using decomposition formulas for infinite-dimensional Lie algebras; (1.5) can be derived from (1.6) by using Lemma 1 in Ref. 2 for the two-dimensional Lie algebra $\{(\partial/\partial x)x, \partial/\partial x\}$.

Contrary to the aforementioned approaches, we want to give a unified approach to the solution of (1.1) based on the work of Wei and Norman,[5,6] which is recapitulated in Sec. II. In Sec. III we will see that $U(x,t)$ can be represented in the form (1.5) by using only solvable finite-dimensional Lie algebras, irrespectively of whether the coefficients in (1.2) are explicitly time dependent or not. A short summary is given in Sec. IV.

II. THE ALGEBRAIC METHOD

In this section we will shortly review the algebraic method developed by Wei and Norman[5,6] of solving the equation

$$\frac{dU(t)}{dt} = A(t)U(t), \quad U(0) = 1, \tag{2.1}$$

where A and U are linear operators and A is given as a linear combination of time-independent operators T_i with scalar valued functions $a_i(t)$ as coefficients

$$A(t) = \sum_{i=1}^{M} a_i(t)T_i. \tag{2.2}$$

If the Lie algebra L generated by (2.2) is finite dimensional with basis $\{T_1,...,T_N\}$, $N \geqslant M$, the solution of (2.1) can be represented (at least locally) in the factorized form

$$U(t) = \exp(g_1T_1)\exp(g_2T_2)\cdots\exp(g_NT_N). \tag{2.3}$$

The functions g_i, $i = 1,...,N$, have to be determined by a set of nonlinear differential equations

$$\begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix} = \begin{pmatrix} \xi_{11} & \cdots & \xi_{1N} \\ \vdots & & \vdots \\ \xi_{N1} & & \xi_{NN} \end{pmatrix}\begin{pmatrix} \dot{g}_1 \\ \vdots \\ \dot{g}_N \end{pmatrix},$$
$$g_i(0) = 0, \quad i = 1,...,N, \tag{2.4}$$

where the ξ_{kl} are analytic functions of the g_i's.

Since we want to apply this method to linear Fokker–Planck equations, it is sufficient to restrict ourselves to solvable Lie algebras (compare Sec. III), i.e., Lie algebras for which the derived series $L^{(0)} = L$, $L^{(1)} = [L,L]$,...,$L^{(i)} = [L^{(i-1)},L^{(i-1)}]$ truncates for some integer n.

For solvable Lie algebras Eqs. (2.4) are easily solved.[6] Arranging the basis of the Lie algebra L generated by (2.2) such that

$$L = L_N \supset L_{N-1} \supset \cdots \supset L_1 \supset \{0\}, \tag{2.5}$$

where each ideal L_i is exactly of dimension i, Eqs. (2.4) read

$$\begin{pmatrix} a_1 \\ \vdots \\ \vdots \\ a_N \end{pmatrix} = \begin{pmatrix} \xi_{11} & 0 & \cdots & & 0 \\ \xi_{21} & \xi_{22} & 0 & \cdots & 0 \\ \vdots & & & & \\ \xi_{N1} & \cdots & & & \xi_{NN} \end{pmatrix}\begin{pmatrix} \dot{g}_1 \\ \vdots \\ \vdots \\ \dot{g}_N \end{pmatrix}, \tag{2.6}$$

i.e., the system is in triangular form and the element ξ_{ki} depends only on g_j with $j < k$. As shown in Ref. 6, the ξ matrix is in this case invertible for all t, so that a global solution of

 0022-2488/88/020305-03$02.50

(2.1) is given by (2.3), where the functions $g_i(t)$ can be obtained by quadratures.

III. THE LINEAR FOKKER–PLANCK EQUATION

We will discuss the one-dimensional linear Fokker–Planck equation with time-dependent coefficients

$$\frac{\partial P(x,t)}{\partial t} = \left\{ -\frac{\partial}{\partial x}(a(t)x + b(t)) + c(t)\frac{\partial^2}{\partial x^2}\right\} P(x,t) \tag{3.1}$$

and search for a solution in the form

$$P(x,t) = U(t)P(x,0), \tag{3.2}$$

where U satisfies (2.1) with

$$A(t) = -a(t)\frac{\partial}{\partial x}x - b(t)\frac{\partial}{\partial x} + c(t)\frac{\partial^2}{\partial x^2}. \tag{3.3}$$

The Lie algebra generated by (3.1) is given by

$$L := \left\{\frac{\partial}{\partial x}x,\ \frac{\partial}{\partial x},\ \frac{\partial^2}{\partial x^2}\right\}, \tag{3.4}$$

with commutation relations

$$\left[\frac{\partial}{\partial x}x,\ \frac{\partial}{\partial x}\right] = -\frac{\partial}{\partial x}, \tag{3.5a}$$

$$\left[\frac{\partial^2}{\partial x^2},\ \frac{\partial}{\partial x}\right] = 0, \tag{3.5b}$$

$$\left[\frac{\partial}{\partial x}x,\ \frac{\partial^2}{\partial x^2}\right] = -2\frac{\partial^2}{\partial x^2}. \tag{3.5c}$$

Since $[L,L] = \{\partial/\partial x, \partial^2/\partial x^2\}$, $L^{(2)} = [[L,L],\ [L,L]] = 0$, i.e., L is solvable.

Arranging the basis in the form

$$\{0\} \subset \left\{\frac{\partial^2}{\partial x^2}\right\} \subset \left\{\frac{\partial^2}{\partial x^2},\ \frac{\partial}{\partial x}\right\} \subset \left\{\frac{\partial^2}{\partial x^2},\ \frac{\partial}{\partial x},\ \frac{\partial}{\partial x}x\right\} = L, \tag{3.6}$$

we see that the chain condition (2.5) is fulfilled.

Searching for the propagator $U(t)$ in the form

$$U(t) = \exp\left[\alpha(t)\frac{\partial}{\partial x}x\right]\exp\left[\beta(t)\frac{\partial}{\partial x}\right]\exp\left[\gamma(t)\frac{\partial^2}{\partial x^2}\right], \tag{3.7}$$

we find for the unknown functions $\alpha(t)$, $\beta(t)$, $\gamma(t)$ the linear equations

$$\begin{pmatrix} -a(t) \\ -b(t) \\ c(t) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{-\alpha(t)} & 0 \\ 0 & 0 & e^{-2\alpha(t)} \end{pmatrix} \begin{pmatrix} \dot\alpha(t) \\ \dot\beta(t) \\ \dot\gamma(t) \end{pmatrix},$$

$$\alpha(0) = \beta(0) = \gamma(0) = 0. \tag{3.8}$$

The solutions of (3.6) are given as

$$\alpha(t) = -\int_0^t a(s)ds, \tag{3.9a}$$

$$\beta(t) = -\int_0^t b(s)e^{\alpha(s)}\,ds, \tag{3.9b}$$

$$\gamma(t) = \int_0^t c(s)e^{2\alpha(s)}\,ds. \tag{3.9c}$$

Using the easily derivable expressions[2]

$$\exp\left[-f(t)\frac{\partial}{\partial x}x\right]P(x) = e^{-f(t)}P(xe^{-f(t)}), \tag{3.10}$$

$$\exp\left[f(t)\frac{\partial^2}{\partial x^2}\right]P(x) = (4\pi f(t))^{-1/2}\int_{-\infty}^{+\infty}\exp\left[-\frac{(x-y)^2}{4f(t)}\right]P(y)dy, \tag{3.11}$$

$$\exp\left[f(t)\frac{\partial}{\partial x}\right]P(x) = P(x+f(t)), \tag{3.12}$$

we find for $P(x,t)$,

$$\begin{aligned} P(x,t) &= \exp\left[\alpha(t)\frac{\partial}{\partial x}x\right]\exp\left[\beta(t)\frac{\partial}{\partial x}\right] \\ &\quad\times\exp\left[\gamma(t)\frac{\partial^2}{\partial x^2}\right]P(x,0) \\ &= \exp\left[\alpha(t)\frac{\partial}{\partial x}x\right]\exp\left[\beta(t)\frac{\partial}{\partial x}\right](4\pi\gamma(t))^{-1/2} \\ &\quad\times\int dy\,P(y,0)\exp\left[-\frac{(y-x)^2}{4\gamma(t)}\right] \\ &= \exp\left[\alpha(t)\frac{\partial}{\partial x}x\right](4\pi\gamma(t))^{-1/2} \\ &\quad\times\int dy\,P(y,0)\exp\left[\frac{-(y-(x+\beta(t)))^2}{4\gamma(t)}\right] \\ &= \frac{e^{\alpha}(t)}{\sqrt{4\pi\gamma(t)}}\int dy\,P(y,0) \\ &\quad\times\exp\left[\frac{-(y-(xe^{\alpha(t)}+\beta(t)))^2}{4\gamma(t)}\right], \end{aligned} \tag{3.13}$$

a result obtained already in Ref. 2 by recurring to decomposition formulas for infinite-dimensional Lie algebras.

A subalgebra of L given in (3.4) is

$$L' := \left\{\frac{\partial^2}{\partial x^2}, \frac{\partial}{\partial x}x\right\}, \tag{3.14}$$

generated by Eq. (3.1) with $b(t) \equiv 0$,

$$\frac{\partial P(x,t)}{\partial t} = \left(-a(t)\frac{\partial}{\partial x}x + c(t)\frac{\partial^2}{\partial x^2}\right)P(x,t). \tag{3.15}$$

The solution of (3.15) is then given by (3.13) with $\beta(t) \equiv 0$. Equation (3.13) can of course be solved independently by imitating the above procedure using the solvable algebra L' instead of L. Writing the propagator in the form

$$U(t) = \exp\left[\delta(t)\frac{\partial}{\partial x}x\right]\exp\left[\epsilon(t)\frac{\partial}{\partial x}\right], \tag{3.16}$$

we obtain for the functions $\delta(t)$, $\epsilon(t)$ just the expressions (3.9a) and (3.9c), respectively.

IV. SUMMARY

In the present paper we recommended the method of Wei and Norman as a systematic way of treating evolution equations and applied it to linear Fokker–Planck equations in one variable. Two advantages are at hand. First, there is no principal distinction in treating equations with constant or time-dependent drift and diffusion functions, since they appear just as coefficients of operators of a finite-dimensional Lie algebra, therefore avoiding the use of infinite-dimensional algebras for time-dependent equations.[2] Second, the propagator U is represented in factorized form, which allows

the treatment of each factor separately and avoids the use of time-ordered exponentials.

ACKNOWLEDGMENT

This work has been supported by the Deutsche Forschungsgemeinschaft under SFB 91: "Energietransfer bei atomaren und molekularen Stoßprozessen."

[1]H. Risken, *The Fokker–Planck Equation* (Springer, Berlin, 1984), and references therein.
[2]M. Suzuki, J. Math. Phys. **26**, 601 (1985).
[3]S. Chandrasekhar, Rev. Mod. Phys. **15**, 1 (1943).
[4]P. Hänggi and H. Thomas, Z. Phys. B **22**, 295 (1975).
[5]J. Wei and E. Norman, J. Math. Phys. **4**, 575 (1963).
[6]J. Wei and E. Norman, Proc. Am. Math. Soc. **15**, 327 (1963).

Gauge transformation and the higher order Korteweg–de Vries equation

Yu-kun Zheng and W. L. Chan
Department of Mathematics, Science Centre, The Chinese University of Hong Kong, Shatin, N. T., Hong Kong

(Received 27 March 1987; accepted for publication 23 September 1987)

To the family of higher order Korteweg–de Vries (KdV) equations a new family of higher order mKdV equations depending on a parameter η is constructed. They are connected by an η^2-dependent Miura transformation. A Bäcklund transformation is also established. Furthermore, a new gauge transformation is associated to this Bäcklund transformation. It enables one to derive auto-Bäcklund transformations for the families of higher order η^2-mKdV and KdV equations, respectively. For illustration, four generations of explicit solutions of the second order KdV equation are presented.

I. INTRODUCTION

In the paper by Chern and Peng,[1] a family of higher order Korteweg–de Vries (ho-KdV) equations and an associated family of higher order modified Korteweg–de Vries (ho-mKdV) equations had been derived from a SL(2,R) formalism. The Miura transformation between these two classes of equations was also discussed. In the present study we will derive a family of η^2-dependent ho-mKdV (η^2-ho-mKdV) equations from the ho-KdV equations and an η^2-dependent Miura transformation connecting these two families. Furthermore, we will present a gauge transformation corresponding to the Bäcklund transformation of the solutions of ho-KdV equations and an auto-Bäcklund transformation of the η^2-ho-mKdV equations.

We first briefly state the derivation of the ho-KdV equations obtained by Chern and Peng.[1] For convenience, we state our problem with the following Ablowitz–Kaup–Newell–Segur (AKNS) system[2]:

$$\Phi_x = P\Phi, \tag{1.1}$$

$$\Phi_t = Q\Phi, \tag{1.2}$$

where

$$\Phi = \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix}, \tag{1.3}$$

$$P = \begin{pmatrix} \eta & u \\ r & -\eta \end{pmatrix}, \tag{1.4}$$

$$\eta, \text{ a parameter, independent of } x \text{ and } t, \tag{1.5}$$

$$r = -1, \tag{1.6}$$

$$u = u(x,t), \tag{1.7}$$

$$Q = \begin{pmatrix} A & B \\ C & -A \end{pmatrix}. \tag{1.8}$$

Here Q will be determined by the following integrability condition:

$$P_t - Q_x + PQ - QP = 0, \tag{1.9}$$

or in component form

$$-A_\alpha + uC + B = 0, \tag{1.10}$$

$$u_t - B_x + 2\eta B - 2uA = 0, \tag{1.11}$$

$$-C_x - 2A - 2\eta C = 0. \tag{1.12}$$

Solving (1.10)–(1.12) gives

$$A = -\tfrac{1}{2}C_x - \eta C, \tag{1.13}$$

$$B = \tfrac{1}{2}C_{xx} - \eta C_x - uC, \tag{1.14}$$

$$u_t = -\tfrac{1}{2}C_{xxx} + 2(\eta^2 - u)C_x - u_x C. \tag{1.15}$$

Following Ref. 1, and taking C to be an arbitrary polynomial in η^2,

$$C = C_{(n)} = \sum_{j=0}^{n} C_j(x,t)\eta^{2(n-j)}. \tag{1.16}$$

Here C depends on n; for convenience, we will still denote it by C. Substituting (1.16) into (1.15) and equating to zero the coefficients of the powers of η, we get

$$C_0 = \text{const (assumed to be 4)}, \tag{1.17}$$

$$C_{j+1,x} = \tfrac{1}{4}C_{j,xxx} + uC_{j,x} + \tfrac{1}{2}u_x C_j, \quad 0 \leqslant j < n, \tag{1.18}$$

$$u_t = K_n(u), \tag{1.19}$$

where

$$K_n(u) = -\tfrac{1}{2}C_{n,xxx} - 2uC_{n,x} - u_x C_n \overset{\text{def}}{=} -2C_{n+1,x}. \tag{1.20}$$

Equations (1.19) and (1.20) is the ho-KdV equation obtained by Chern and Peng[1] [there are some slight differences due to the choice of r in (1.4)]. The C_n in (1.20) is determined by the recursion formulas in (1.17) and (1.18). In fact, (1.18) can be rewritten in the following form:

$$C_{j+1} = EC_j, \quad 0 \leqslant j \leqslant n, \tag{1.21}$$

where

$$E = \tfrac{1}{4}D^2 + u - \tfrac{1}{2}D^{-1}u_x, \tag{1.22}$$

$$D = \frac{\partial}{\partial x}, \quad D^{-1} = \int dx. \tag{1.23}$$

For $j = 1,2,3$, the C_j's have the following expression:

$$C_1 = 2u, \tag{1.24}$$

$$C_2 = \tfrac{1}{2}u_{xx} + \tfrac{3}{2}u^2, \tag{1.25}$$

$$C_3 = \tfrac{1}{8}u_{xxxx} + \tfrac{5}{4}uu_{xx} + \tfrac{5}{8}u_x^2 + \tfrac{5}{4}u^3. \tag{1.26}$$

The K_n in (1.20) can also be expressed in a recursion for-

 0022-2488/88/020308-07$02.50

mula:

$$K_0 = -4u_x, \tag{1.27}$$

$$K_n = FK_{n-1}, \quad n \geqslant 1, \tag{1.28}$$

where

$$F = \tfrac{1}{4}D^2 + u + \tfrac{1}{2}u_x D^{-1}. \tag{1.29}$$

For $n = 0,1,2$, the K_n are given in the following:

$$K_0 = -4u_x, \tag{1.30}$$

$$K_1 = -u_{xxx} - 6uu_x, \tag{1.31}$$

$$K_2 = -\tfrac{1}{4}u_{xxxxx} - \tfrac{5}{2}uu_{xxx} - 5u_x u_{xx} - \tfrac{15}{2}u^2 u_x. \tag{1.32}$$

II. η^2-DEPENDENT HIGHER ORDER mKdV EQUATIONS

In this section we will derive a family of η^2-dependent higher order mKdV equations from the ho-KdV equations (1.19).

Writing (1.1) and (1.2) in component form, we have

$$\varphi_{1x} = \eta\varphi_1 + u\varphi_2, \tag{2.1}$$

$$\varphi_{2x} = -\varphi_1 - \eta\varphi_2, \tag{2.2}$$

$$\varphi_{1t} = A\varphi_1 + B\varphi_2, \tag{2.3}$$

$$\varphi_{2t} = C\varphi_1 - A\varphi_2. \tag{2.4}$$

Let

$$v = -\varphi_1/\varphi_2 - \eta. \tag{2.5}$$

Taking derivatives in (2.5) with respect to x and from (2.1) and (2.2) we get the following relationship between u and v:

$$u = \eta^2 - v_x - v^2. \tag{2.6}$$

Similarly, from (2.3)–(2.5) we have

$$v_t = C(v+\eta)^2 + 2A(v+\eta) - B. \tag{2.7}$$

Substituting (1.13) and (1.14) into (2.7) leads to

$$v_t = (\tfrac{1}{2}C_x - vC)_x = \tfrac{1}{2}C_{xx} - v_x C - vC_x. \tag{2.8}$$

From (1.16) and (1.21) we know that C is a polynomial in η^2 with degree n, and its coefficients C_j's are functionals of u. On the other hand, by the relation of u and v in (2.6), C_j's are functionals of v. Therefore (2.8) is a partial differential equation of v with η^2 as a parameter; this means that the function v defined by (2.5) is a solution of Eq. (2.8). Denote the expression of the right-hand side in (2.8) by M_n,

$$M_n = M_n(v) = (\tfrac{1}{2}C_x - vC)_x, \tag{2.9}$$

then Eq. (2.8) can be expressed in the following form:

$$v_t - M_n(v) = 0. \tag{2.10}$$

For $n = 0,1,2$, by (1.17), (1.21), and (1.16), the M_n are given in the following:

$$M_0 = -4v_x, \tag{2.11}$$

$$M_1 = -v_{xxx} - 6(\eta^2 - v^2)v_x, \tag{2.12}$$

$$M_2 = -\tfrac{1}{4}v_{xxxxx} - \tfrac{5}{2}(\eta^2 - v^2)v_{xxx} + 10vv_x v_{xx} + \tfrac{5}{2}v_x^3 - \tfrac{15}{2}(\eta^2 - v^2)^2 v_x. \tag{2.13}$$

For $n = 1$, by (2.12), Eq. (2.10) take the following form:

$$v_t + v_{xxx} + 6(\eta^2 - v^2)v_x = 0. \tag{2.14}$$

This equation is generally called the η^2-dependent modified KdV equation, therefore, we called (2.10), for $n > 1$, the η^2-dependent higher order modified KdV (η^2-ho-mKdV) equation. Denote

$$R = -D - 2v, \tag{2.15}$$

then by (2.6), (2.9), and (1.15), it is easy to verify that

$$u_t = Rv_t, \tag{2.16}$$

and

$$RM_n = K_n. \tag{2.17}$$

Hence, by combining (1.19), (2.16), (2.17), and (2.10), we find that the two families of equations of ho-KdV (1.19) and η^2-ho-mKdV (2.10) are connected by the following equality:

$$u_t - K_n(u) = R[v_t - M_n(v)]. \tag{2.18}$$

Summarizing the results obtained in this section, we have proved the following two theorems, respectively.

Theorem 1: Whenever $v = v(x,t)$ is a solution of the η^2-ho-mKdV equation (2.10), the function $u = u(x,t)$, determined by (2.6), is a solution of the ho-KdV equation (1.19).

Theorem 2: Whenever $u = u(x,t)$ is a solution of the ho-KdV equation (1.19), the function $v = v(x,t)$, defined by (2.5), is a solution of the η^2-ho-mKdV equation (2.10).

We now derive some further properties of M_n which will be used in the sequel. Let

$$H_n = \tfrac{1}{2}C_x - vC. \tag{2.19}$$

Then, by (2.9), we have

$$M_n = H_{n,x}. \tag{2.20}$$

Substituting (1.16) into (2.19) and using (1.20) and (2.17), we have, for $n > 0$,

$$\begin{aligned} H_n &= \frac{1}{2}C_{n,x} - vC_n + \eta^2 \sum_{j=0}^{n-1} \left(\frac{1}{2}C_{j,x} - vC_j\right)\eta^{2(n-1-j)} \\ &= (\tfrac{1}{2} - vD^{-1})C_{n,x} + \eta^2 H_{n-1} \\ &= (-\tfrac{1}{4} + \tfrac{1}{2}vD^{-1})K_{n-1} - vC_n(0) + \eta^2 H_{n-1} \\ &= (-\tfrac{1}{4} + \tfrac{1}{2}vD^{-1})RM_{n-1} + \eta^2 H_{n-1} - vC_n(0) \\ &= (-\tfrac{1}{4} + \tfrac{1}{2}vD^{-1})RDH_{n-1} + \eta^2 H_{n-1} - vC_n(0) \\ &= [(\tfrac{1}{4}D - vD^{-1}v)D + \eta^2]H_{n-1} - vC_n(0) \\ &= TH_{n-1} - vC_n(0), \end{aligned} \tag{2.21}$$

where

$$T = (\tfrac{1}{4}D - vD^{-1}v)D + \eta^2, \tag{2.22}$$

and $C_n(0)$ is obtained from C_n in (1.21) by taking v to be zero. Then $C_n(0)$ has the following explicit expression:

$$C_n(0) = [(2n-1)!!/2^{n-2}n!]\eta^{2n}, \quad \text{for } n > 0. \tag{2.23}$$

Thus, by (1.17), (2.19), and (2.21), we obtain a recursion formula for H_n:

$$H_0 = -4v, \tag{2.24}$$

$$H_n = TH_{n-1} - C_n(0)v, \quad \text{for } n > 0. \tag{2.25}$$

From these recursion formulas and (2.22), we see that H_n is odd with respect to v,

$$H_n(-v) = -H_n(v), \tag{2.26}$$

since H_0 is odd and T is even, and (2.26) follows by induction.

From (2.20) and (2.26) we see that M_n is also odd with respect to v,

$$M_n(-v) = -M_n(v). \tag{2.27}$$

Taking the derivative with respect to x in (2.24) and (2.25), respectively, and using (2.20), we obtain a recursion formula for M_n,

$$M_0 = -4v_x, \tag{2.28}$$

$$M_n = SM_{n-1} - C_n(0)v_x, \quad \text{for } n>0, \tag{2.29}$$

where

$$S = \tfrac{1}{4}D^2 + \eta^2 - v^2 - v_x D^{-1} v. \tag{2.30}$$

III. η^2-DEPENDENT MIURA TRANSFORMATION AND BÄCKLUND TRANSFORMATION

It is well known that the KdV and the mKdV equations are related by the Miura transformation. As for the ho-KdV and the η^2-ho-mKdV equations, we see that, from Theorems 1 and 2, relation (2.6) plays a similar role as the Miura transformation does. We call (2.6) the η^2-dependent Miura transformation. We use this transformation to derive a Bäcklund transformation (BT) of the ho-KdV equation (1.19).

We know from (2.27) that $M_n(v)$ is odd with respect to v. Thus the η^2-ho-mKdV equation (2.10) possesses with every solution v another solution $-v$, but then, by substituting $-v$ into (2.6) and by Theorem 2, we obtain another solution u' of the ho-KdV equation (1.19),

$$u' = \eta^2 + v_x - v^2. \tag{3.1}$$

Subtracting (2.6) from (3.1), we get

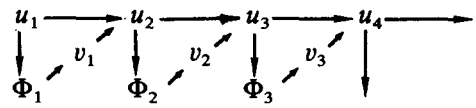
$$u' = u + 2v_x. \tag{3.2}$$

This is the Bäcklund transformation of the ho-KdV equation (1.19) that we want to establish (for $n=1$, it had been derived by Jager and Spannenberg[2]). The above results can be stated in the following two theorems.

Theorem 3: Whenever the function $v = v(x,t)$ is a solution of the η^2-ho-mKdV equation (2.10) and u a function determined by (2.6), then the function u' determined by (3.2), is also a solution of the ho-KdV equation (1.19).

Theorem 4: Whenever the function $u = u(x,t)$ is a solution of the ho-KdV equation (1.19), and $v = v(x,t)$ is a function determined by (2.5), then the function u' determined by (3.2) is also a solution of the ho-KdV equation (1.19).

The difference between these two theorems is that Theorem 3 indicates that (3.2) is a Bäcklund transformation between the η^2-ho-mKdV equation (2.10) and the ho-KdV equation (1.19), while Theorem 4 shows that (3.2) is an auto-Bäcklund transformation of the ho-KdV equation (1.19). Therefore, when applying these Bäcklund transformations to construct new solutions of the ho-KdV equation (1.19), it is evident that Theorem 3 tells us to find a solution v of the η^2-ho-mKdV equation (2.10) in advance, while Theorem 4 instructs us to take a known solution u of the ho-KdV equation (1.19) in advance and to solve the AKNS system (1.1)–(1.16) for the wave functions (1.3) corresponding to u, since the function v appearing in (3.2) is determined by the wave function Φ in (2.5). Obviously, Theorem 4 also indicates a method for obtaining a hierarchy of new solutions of the ho-KdV equation (1.19) starting from a known solution u_1 of this equation in the following manner:



(3.3)

Our current interest is in the application of Theorem 4 to obtain more solutions of the ho-KdV equation (1.19) from a known solution u_1 of this equation, and we will further study its ramifications in the next section.

IV. GAUGE TRANSFORMATION OF THE WAVE FUNCTION

In order to apply the Bäcklund transformation (3.2) in the sense of Theorem 4 to obtain a hierarchy of new solutions of the ho-KdV equation (1.19) from a known solution of the same equation, we see from (3.3) that it is necessary to seek a convenient method for obtaining the new wave function from a known wave function corresponding to a known solution of Eq. (1.19) so as to avoid the difficulty of solving the AKNS system (1.1) and (1.2). To this end, we will seek a gauge transformation G of the wave function (1.3) which takes the wave function Φ corresponding to the known solution u in (3.2) to a new wave function Φ', $\Phi' = G\Phi$ corresponding to the new solution u' (see the papers by Orfanidis[3] and Wadati and Sogo[4] for other gauge transformations). With this in mind, we proceed to construct such a G through the following three steps.

(i) In view of (1.19), (2.6), (2.10), and (2.18), we construct a gauge transformation S_1 which connects the AKNS system (1.1)–(1.9), denoted by $L_1(u)$, to a second linear system $L_2(v)$ with Eq. (2.10) as its integrability condition.

(ii) Exploiting the oddness of Eq. (2.10), we determine a second gauge transformation S_2 which maps $L_2(v)$ to a third linear system $L_3(-v)$ with the equation

$$(-v)_t - M_n(-v) = 0 \tag{4.1}$$

as its integrability condition. Obviously, Eq. (4.1) is the same as Eq. (2.10).

(iii) Finally, Eq. (3.1) leads us to construct a gauge transformation S_3 which transforms $L_3(-v)$ to $L_1(u')$ with u' in (3.1) as its potential function.

Then we combine the S_1, S_2, S_3 together to obtain the gauge transformation G,

$$G = S_3 S_2 S_1, \tag{4.2}$$

which amounts to taking $L_1(u)$ to $L_1(u')$. These steps can be

depicted by the following diagram:

$$
\begin{array}{ccc}
u_t - K_n(u) = 0 & & v_t - M_n(v) = 0 \\
L_1(u) & \xrightarrow{S_1} & L_2(v) \\
G \downarrow & & \downarrow S_2 \\
L_1(u') & \xleftarrow{S_3} & L_3(-v) \\
u'_t - K_n(u') = 0 & & (-v)_t - M_n(-v) = 0
\end{array}
\tag{4.3}
$$

Let u be a known solution of the ho-KdV equation (1.19) and Φ in (1.3) be the corresponding wave function, then u and Φ satisfy all the relations (1.1)–(1.20), and by (2.5) we have the function v which, as we know, is a solution of the η^2-ho-KdV equation (2.10). Moreover u and v satisfy (2.6). Now, let S_1 be a 2×2 matrix defined as

$$
S_1 = \frac{1}{\beta}\begin{pmatrix} -1 & -\eta - v \\ 0 & 1 \end{pmatrix}, \quad \beta = \exp\int v\,dx. \tag{4.4}
$$

Consider the following transformation:

$$
S_1: \Phi \to \Psi = S_1\Phi, \tag{4.5}
$$

where the components of Ψ are denoted by

$$
\Psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}. \tag{4.6}
$$

Then (4.5) will transform the linear system (1.1)–(1.8) into the following new linear system:

$$
\Psi_x = V\Psi, \tag{4.7}
$$

$$
\Psi_t = W\Psi, \tag{4.8}
$$

with

$$
V = S_{1x}S_1^{-1} + S_1PS_1^{-1} = \begin{pmatrix} -2v & 0 \\ 1 & 0 \end{pmatrix}, \tag{4.9}
$$

$$
W = S_{1t}S_1^{-1} + S_1QS_1^{-1} = \begin{pmatrix} -2H_n & 0 \\ -C & 0 \end{pmatrix}, \tag{4.10}
$$

where H_n is defined in (2.19). In the derivation of (4.10), we have made use of (1.13), (1.14), and (2.8). The system (4.7)–(4.10) is compatible, that is, the matrices V and W satisfy the following equation:

$$
V_t - W_x + VW - WV = 0. \tag{4.11}
$$

To check this, substituting (4.9) and (4.10) into (4.11) leads to the requirement that

$$
\begin{pmatrix} 2(M_n - v_t) & 0 \\ 0 & 0 \end{pmatrix} = 0. \tag{4.12}
$$

But (4.12) holds because of (2.10). This shows that (2.10) is the integrability condition of the system (4.7)–(4.10). Therefore the system (4.7)–(4.10) is the linear system $L_2(v)$ which we want to construct and the matrix (4.4) is the gauge which transforms the system $L_1(u)$ in (1.1)–(1.9) to the system $L_2(v)$ in (4.7)–(4.10). Referring to Theorem 2, we state this result as the following theorem.

Theorem 5: Let u be a solution of Eq. (1.19) and let v be related by u as (2.6). Then (4.5) is a gauge transformation which transforms the system (1.1)–(1.9) into the system (4.7)–(4.10) and Eq. (2.10) is the integrability condition of the system (4.7)–(4.10).

The inverse of this theorem is also true, since the gauge S_1 in (4.4) is invertible.

Theorem 6: Let v be a solution of Eq. (2.10), let u be determined by v in (2.6), and let C be determined by (1.16), (1.17), and (1.21). Then the gauge transformation

$$
S_1^{-1}: \Psi \to \Phi = S_1^{-1}\Psi \tag{4.13}
$$

transforms the linear system (4.7)–(4.10) into the linear system (1.1)–(1.9) and Eq. (1.19) is the integrability condition of the system (1.1)–(1.9).

Next we introduce two linear systems which are related to systems (1.1)–(1.9) and (4.7)–(4.10).

The first one is the following AKNS system:

$$
\Phi'_x = P'\Phi', \tag{4.14}
$$

$$
\Phi'_t = Q'\Phi', \tag{4.15}
$$

where

$$
\Phi' = \begin{pmatrix} \varphi'_1 \\ \varphi'_2 \end{pmatrix}, \tag{4.16}
$$

$$
P' = \begin{pmatrix} \eta & u' \\ -1 & -\eta \end{pmatrix}, \tag{4.17}
$$

$$
Q' = \begin{pmatrix} -\frac{1}{2}C'_x - \eta C' & -\frac{1}{2}C'_{xx} - \eta C'_x - u'C' \\ C' & \frac{1}{2}C'_x + \eta C' \end{pmatrix}. \tag{4.18}
$$

We take the function u' in (4.17) and (4.18) to be the function defined in (3.1), that is, it satisfies the ho-KdV equation (1.19); therefore the C' in (4.19) must satisfy a similar condition as the C in (1.8) does. Obviously, we have

$$
C' = C(-v), \tag{4.19}
$$

since u' in (3.1) is obtained from (2.6) by changing v into $-v$.

The second linear system is

$$
\Psi'_x = V'\Psi', \tag{4.20}
$$

$$
\Psi'_t = W'\Psi', \tag{4.21}
$$

where

$$
\Psi' = \begin{pmatrix} \psi'_1 \\ \psi'_2 \end{pmatrix}, \tag{4.22}
$$

$$
V' = \begin{pmatrix} 2v & 0 \\ 1 & 0 \end{pmatrix}, \tag{4.23}
$$

$$
W' = \begin{pmatrix} -2H_n(-v) & 0 \\ -C(-v) & 0 \end{pmatrix}. \tag{4.24}
$$

We take the function v in (4.23) and (4.24) to be the function defined in (2.5), that is, v together with u' satisfy (3.1). Obviously, the linear systems (4.14)–(4.18) and (4.20)–(4.24) are obtained from the linear systems (1.1)–(1.9) and (4.7)–(4.10), respectively, by changing all the signs in front of v, viz., v is replaced by $-v$. Therefore, by Theorem 6, taking all the v's there as $-v$, and by the oddness of M_n in (2.27) and (2.10), a gauge transformation S_3 which transforms the wave function Ψ' in (4.22) into the wave function Φ' in (4.16) exists, i.e.,

$$
S_3: \Psi' \to \Phi' = S_3\Psi', \tag{4.25}
$$

where

$$S_3=\frac{1}{\beta}\begin{pmatrix}-1 & -\eta+v\\ 0 & 1\end{pmatrix}, \tag{4.26}$$

and is obtained from S_1 in (4.4) by changing all v in S_1 into $-v$ and then taking the inverse.

Finally we show that there exists a gauge transformation which transforms the system (4.6)–(4.10) into the system (4.20)–(4.24).

Let S_2 be a 2×2 matrix,

$$S_2=\begin{pmatrix}a & b\\ c & d\end{pmatrix} \tag{4.27}$$

with

$$a=\beta^4\left(a_0-b_0\int\beta^{-2}\,dx\right), \tag{4.28}$$

$$b=b_0\beta^2, \tag{4.29}$$

$$c=\beta^2\left(a_0\int\beta^2\,dx-b_0\int\beta^2\,dx\times\int\beta^{-2}\,dx+c_0-d_0\int\beta^{-2}\,dx\right), \tag{4.30}$$

$$d=b_0\int\beta^2\,dx+d_0, \tag{4.31}$$

where β is defined in (4.4), and a_0, b_0, c_0, and d_0 are constants chosen such that

$$a_0d_0-b_0c_0=1. \tag{4.32}$$

We prove that S_3 is a gauge which transforms the wave function Ψ in (4.7) and (4.8) into the wave function Ψ' in (4.20) and (4.21),

$$S_2\colon\ \Psi\to\Psi'=S_2\Psi. \tag{4.33}$$

This means that we must check that the following two matrix equalities hold:

$$V'=S_{2x}S_2^{-1}+S_2VS_2^{-1}, \tag{4.34}$$

$$W'=S_{2t}S_2^{-1}+S_2WS_2^{-1}. \tag{4.35}$$

First we check (4.34). By (4.27)–(4.32), we have

$$\det S_2=ad-bc=\beta^4, \tag{4.36}$$

$$S_2^{-1}=\frac{1}{\beta^4}\begin{pmatrix}d & -b\\ -c & a\end{pmatrix}, \tag{4.37}$$

$$S_{2x}=\begin{pmatrix}4va-b & 2vb\\ a+2vc-d & b\end{pmatrix}. \tag{4.38}$$

Substituting (4.9), (4.27), (4.37), and (4.38) into the right-hand side of (4.34) and simplifying the expression leads to the matrix V' in (4.23), that is, (4.34) holds.

Next we check (4.35). By (4.4), (2.8), and (2.19), we have

$$\beta_t=\beta H_n, \tag{4.39}$$

$$\int\beta^{-2}H_n\,dx=\frac{1}{2}\beta^{-2}C. \tag{4.40}$$

Changing v into $-v$ in (4.40) and using (2.26), (4.40) leads to

$$\int\beta^2H_n\,dx=-\frac{1}{2}\beta^2C(-v). \tag{4.41}$$

Taking the derivative in (4.27)–(4.32) with respect to t and using (4.39)–(4.41), we get

$$S_{2t}=\begin{pmatrix}bC+4aH_n & 2bH_n\\ dC+2cH_n-aC(-v) & -bC(-v)\end{pmatrix}. \tag{4.42}$$

Substituting (4.10), (4.27), (4.37), and (4.42) into the right-hand side of (4.35) gives the matrix W' in (4.24), that is, equality (4.35) holds.

Thus (4.27)–(4.33) is the gauge transformation which we want to seek as mentioned in (ii), and (4.20)–(4.24) is the linear system $L_3(-v)$.

Now, we combine (4.5), (4.33), and (4.25) together so that

$$\Phi'=S_3S_2S_1\Phi. \tag{4.43}$$

Let

$$G=S_3S_2S_1, \tag{4.44}$$

then G is the gauge that we are looking for, it transforms a wave function Φ in the AKNS system (1.1)–(1.8) corresponding to a known solution u of the ho-KdV equation (1.19) to another wave function Φ' in the AKNS system (4.14)–(4.18) corresponding to a new solution u' of Eq. (1.19) defined by (3.1) or (3.2). Substituting (4.26), (4.27), and (4.4) into (4.44), we get an explicit expression for G,

$$G=\frac{1}{\beta}\begin{pmatrix}-1 & -\eta+v\\ 0 & 1\end{pmatrix}\cdot\begin{pmatrix}a & b\\ c & d\end{pmatrix}\cdot\frac{1}{\beta}\begin{pmatrix}-1 & -\eta-v\\ 0 & 1\end{pmatrix}$$
$$=\frac{1}{\beta^2}\begin{pmatrix}a+(\eta-v)c & (\eta+v)a-b+(\eta^2-v^2)c-(\eta-v)d\\ -c & -(\eta+v)c+d\end{pmatrix}, \tag{4.45}$$

where a, b, c, and d are defined in (4.28)–(4.31).

V. BÄCKLUND TRANSFORMATION FOR THE η^2-DEPENDENT HIGHER ORDER mKdV EQUATION

The original object of constructing a gauge transformation for the wave function is to seek a convenient method for obtaining the function v defined by (2.5). Now, by using the gauge transformation (4.45) we can derive a formula which connects the two adjacent v and v'. Suppose that v is known, obtained from a known wave function $\Phi=(\varphi_1,\varphi_2)^T$ by (2.5), then we construct the gauge G in (4.45) and obtain the successive wave function $\Phi'=(\varphi_1',\psi_2')^T$ by gauge

transformation,

$$\Phi' = G\Phi. \tag{5.1}$$

Substituting (4.45) into (5.1) and by (2.5) we have

$$v' = -\frac{[a+(\eta-v)c]\varphi_1 + [(\eta+v)a - b + (\eta^2-v^2)c - (\eta-v)d]\varphi_2}{-c\varphi_1 + [-(\eta+v)c+d]\varphi_2} - \eta. \tag{5.2}$$

Dividing simultaneously the numerator and denominator of the fraction (5.2) by φ_2 and using (2.5),

$$v' = -\frac{[a+(\eta-v)c](-v-\eta) + (\eta+v)a - b + (\eta^2-v^2)c - (\eta-v)d}{-c(-v-\eta) - (\eta+v)c + d} - \eta = -v + \frac{b}{d}. \tag{5.3}$$

Applying (4.29), (4.31), and (4.4), we finally obtain the required formula

$$v' = -v + \left\{\ln\left[\int\left(\exp 2\int v\,dx\right)dx + k_0\right]\right\}_x, \tag{5.4}$$

where

$$k_0 = d_0/b_0 \tag{5.5}$$

is a constant.

We recall that the functions v's are solutions of Eq. (2.10), therefore formula (5.4) is, in fact, an auto-Bäcklund transformation for the η^2-ho-mKdV equation (2.10).

VI. APPLICATIONS

Having formula (5.4), to obtain more new solutions of the ho-KdV equation (1.19) from a known solution by using the Bäcklund transformation (3.1) can be realized more easily. Suppose that u_1 is a known solution of Eq. (1.19) (we call it a seed solution), then, corresponding to this solution we have a wave function Φ_1 solved from the AKNS system (1.1)–(1.9), and by (2.5), we get a function v_1. Then, starting from this v_1 and using the Bäcklund transformation (5.4), we will obtain a hierarchy of v's,

$$v_1, v_2, v_3, \ldots. \tag{6.1}$$

By using (3.2) recursively, corresponding to these v's, we will obtain a hierarchy of solutions of the ho-KdV equation,

$$u_1, u_2, u_3, \ldots. \tag{6.2}$$

This process can be depicted by the following diagram:

$$\begin{array}{ccccccc} u_1 & \to & u_2 & \to & u_3 & \to & u_4 \to \\ \downarrow & & \uparrow & & \uparrow & & \uparrow \\ \Phi_1 & \to & v_1 & \to & v_2 & \to & v_3 \to \end{array}. \tag{6.3}$$

This diagram is an improvement of that in (3.3). By this diagram we see that in the application of the Bäcklund transformation (3.2) to find the hierarchy of solutions of the ho-KdV equation it is only necessary to solve one wave function in (1.1) and (1.2) corresponding to the seed solution of the ho-KdV equation (1.19), while the gauge transformation (4.45) only plays the role of a tool for deriving the auto-Bäcklund transformation (5.4). Perhaps (5.4) is a most meaningful result here.

Example: Solutions of the second order KdV equation.

For $n = 2$, by (1.19) and (1.32) we get the second order KdV equation as follows:

$$u_t + \tfrac{1}{4}u_{xxxxx} + \tfrac{5}{2}uu_{xxx} + 5u_xu_{xx} + \tfrac{15}{2}u^2u_x = 0. \tag{6.4}$$

Substituting (1.17), (1.24), and (1.25) into (1.16) gives

$$C = 4\eta^4 + 2\eta^2 u + \tfrac{1}{2}(u_{xx} + 3u^2), \tag{6.5}$$

while (6.5) into (1.13) and (1.14) gives

$$A = -4\eta^5 - 2\eta^3 u - \eta^2 u_x - \tfrac{1}{2}\eta(u_{xx} + 3u^2) - \tfrac{1}{4}u_{xxx} - \tfrac{3}{2}uu_x, \tag{6.6}$$

$$B = -4\eta^4 u - 2\eta^3 u_x - \eta^2(u_{xx} + 2u^2) - \eta(3uu_x + \tfrac{1}{2}u_{xxx}) - \tfrac{3}{2}u^3 - 2uu_{xx} - \tfrac{3}{2}u_x^2 - \tfrac{1}{4}u_{xxxx}. \tag{6.7}$$

Now we apply the method established in the above to find the solutions up to the fourth generation for Eq. (6.4). Obviously, Eq. (6.4) possesses a constant solution, denoted by u_1, so we take u_1 as the seed solution of Eq. (6.4). Substituting u_1 into (6.5)–(6.7), we get

$$A = -k\eta, \tag{6.8}$$

$$B = -ku_1, \tag{6.9}$$

$$C = k, \tag{6.10}$$

where

$$k = 4\eta^4 + 2\eta^2 u_1 + \tfrac{3}{2}u_1^2. \tag{6.11}$$

Now, (6.8)–(6.10) and (1.8) give

$$Q = -k\begin{pmatrix} \eta & u_1 \\ -1 & -\eta \end{pmatrix} = -kP. \tag{6.12}$$

Furthermore, from (1.1), (1.2), (1.4), (1.6), and (6.12), we find that

$$d\Phi = \Phi_x\,dx + \Phi_t\,dt = P\Phi\,dx - kP\Phi\,dt = P\Phi d(x-kt) = P\Phi\,d\rho, \tag{6.13}$$

where

$$\rho = x - kt. \tag{6.14}$$

The solution of Eq. (6.13) is

$$\Phi = \exp(P\rho)\Phi_0 = (I + \rho P + (\rho^2/2!)P^2 + (\rho^3/3!)P^3 + \cdots)\Phi_0, \tag{6.15}$$

where Φ_0 is a constant column vector. Denoting

$$\alpha^2 = \eta^2 - u_1, \tag{6.16}$$

then (6.15) leads to

$$\Phi = (I + \rho P + (\alpha^2\rho^2/2!)I + (\alpha^2\rho^3/3!)P + (\alpha^4\rho^4/4!)I + (\alpha^4\rho^5/5!)P + \cdots)\Phi_0$$
$$= I[1 + (\alpha\rho)^2/2! + (\alpha\rho)^4/4! + \cdots]\Phi_0 + (1/\alpha)P[(\alpha\rho) + (\alpha\rho)^3/3! + (\alpha\rho)^5/5! + \cdots]\Phi_0$$
$$= (I\cosh\alpha\rho + (1/\alpha)P\sinh\alpha\rho)\Phi_0$$
$$= \begin{pmatrix} \cosh\alpha\rho + (\eta/\alpha)\sinh\alpha\rho & (u_1/\alpha)\sinh\alpha\rho \\ -(1/\alpha)\sinh\alpha\rho & \cosh\alpha\rho - (\eta/\alpha)\sinh\alpha\rho \end{pmatrix}\Phi_0. \tag{6.17}$$

Taking $\Phi_0 = (1,0)^T$ in (6.17), then we deduce that

$$\begin{pmatrix}\varphi_1\\ \varphi_2\end{pmatrix} = \begin{pmatrix}\cosh\alpha\rho + (\eta/\alpha)\sinh\alpha\rho\\ -(1/\alpha)\sinh\alpha\rho\end{pmatrix}. \tag{6.18}$$

Hence from (2.5) we get

$$v_1 = \alpha\coth\alpha\rho. \tag{6.19}$$

Substituting u_1 and (6.19) into (3.2), we obtain the second solution of Eq. (6.4),

$$u_2 = u_1 - 2\alpha^2\operatorname{csch}^2\alpha\rho. \tag{6.20}$$

Next, (6.19) and (5.4) determine v_2 as

$$v_2 = -\alpha\coth\alpha\rho + \frac{4\alpha\sinh^2\alpha\rho}{\sinh 2\alpha\rho - 2\alpha\rho + 4\alpha k_1}. \tag{6.21}$$

Setting (6.21) into (3.2), we construct the third solution of Eq. (6.4),

$$u_3 = u_1 + 2\left(\frac{4\alpha\sinh\alpha\rho}{\sinh 2\alpha\rho - 2\alpha\rho + 4\alpha k_1}\right)_x$$
$$= u_1 + 2[\ln(\sinh 2\alpha\rho - 2\alpha\rho + 4\alpha k_1)]_{xx}. \tag{6.22}$$

Repeating the application of formulas (5.4) and (3.2) as done above, we obtain the v_3 as

$$v_3 = -v_2 + \left[\ln\left(\int\sigma\,dx + k_2\right)\right]_x, \tag{6.23}$$

with

$$\sigma = \exp 2\int v_2\,dx = \left(\frac{\sinh 2\alpha\rho - 2\alpha\rho + 4\alpha k_1}{\sinh\alpha\rho}\right)^2, \tag{6.24}$$

and the fourth solution of Eq. (6.4),

$$u_4 = u_1 + 2\left\{\ln\left[\left(\int\sigma\,dx + k_2\right)\sinh\alpha\rho\right]\right\}_{xx}. \tag{6.25}$$

Obviously, here we have also found three generations of solutions of the second order η^2-mKdV equation; they are the functions in (6.19), (6.21), and (6.23). By (2.13) and (2.10) for $n = 2$, we get the second order η^2-mKdV equation as follows:

$$v_t + \tfrac{1}{4}v_{xxxxx} + \tfrac{5}{2}(\eta^2 - v^2)v_{xxx} - 10vv_xv_{xx}$$
$$- \tfrac{5}{2}v_x^3 + \tfrac{15}{2}(\eta^2 - v^2)^2v_x = 0. \tag{6.26}$$

[1] S-s. Chern and C-k. Peng, Manuscr. Math. **28**, 207 (1979).
[2] E. M. Jager and S. Spannenburg, J. Phys. A: Math. Gen. **18**, 2177 (1985).
[3] S. J. Orfanidis, Phys. Rev. D **21**, 1513 (1980).
[4] M. Wadati and K. Sogo, J. Phys. Soc. Jpn. **52**, 394 (1983).

Multiplications of distributions in elasticity and hydrodynamics

J. F. Colombeau and A. Y. Le Roux
Mathematiques et Informatique, Universite de Bordeaux I, 351 Cours de la Liberation, 33405 Talence Cedex, France

(Received 12 February 1987; accepted for publication 31 August 1987)

It is shown how a mathematical theory of generalized functions, in which the multiplications of distributions appearing in nonlinear equations of physics make sense, gives new formulas and new numerical results. The new methods shown here are quite general but since each particular problem requires its own study, this paper is limited to elasticity and hydrodynamics. In elasticity Hooke's law gives systems in a nonconservative form; the study of shock waves for these systems gives nonclassical multiplications of distributions of the form $Y\cdot\delta$ (Y = Heaviside function, δ = Dirac mass at the origin). Using this new mathematical tool new formulas are obtained (more generally new numerical schemes): in a first step "ambiguous" results are obtained; then the ambiguity is removed. In hydrodynamics a formulation is obtained that has a nonconservative form and is at the basis of efficient new numerical schemes. Strictly speaking the reader is not assumed to know anything either on distributions or on elasticity and hydrodynamics, since the basic equations are recalled. All computations done in this paper are rigorous from the mathematical viewpoint.

I. INTRODUCTION

A one-dimensional model (obtained by splitting from the three-dimensional system) of the equations of elasticity is written in the form of the following system:

$$\rho_t + (\rho u)_x = 0, \quad \text{mass conservation},$$
$$(\rho u)_t + (\rho u^2)_x + (p - S)_x = 0,$$
$$\text{momentum conservation},$$
$$\mathscr{E}_t + (\mathscr{E} u)_x + [(p - S)u]_x = 0, \qquad (1)$$
$$\text{energy conservation},$$
$$S_t + uS_x - \tfrac{4}{3}\mu u_x = 0, \quad \text{Hooke's law},$$
$$p = \phi(\rho, I), \quad \text{state law},$$

where ρ is the density, u is the velocity, p is the pressure, S is the stress deviator, $\mathscr{E} = \rho e$ is the total energy per unit volume, and e is the total energy per unit mass; if I is the internal energy per unit mass, we have of course (sum of the internal energy and of the kinetic energy)

$$\mathscr{E} = \rho I + \rho u^2/2.$$

Here μ is an elasticity constant depending on the medium and $\sigma = S - p$ is the stress. For the applications it is very important to study the shock waves solutions of (1): these are the solutions of (1) that are discontinuous at the same point. The term uS_x in the fourth equation gives rise to a meaningless product of distributions of the form $Y\cdot\delta$ (Y is the Heaviside step function, δ is the Dirac mass at the origin). The shocks of the system (1) cannot be studied within distribution theory. Note that heuristic attempts to transform (1) into a conservative form have led to results that are unacceptable from the physical viewpoint.

For the purpose of mathematical simplification we shall study more particularly the two simpler systems that have a nonconservative form,

$$\rho_t + (\rho u)_x = 0, \quad (\rho u)_t + (\rho u^2)_x = \sigma_x,$$
$$\sigma_t + u\sigma_x = k^2 u_x, \qquad (2)$$

and a simplified form of (2) (when ρ is assumed to remain very close to a fixed value ρ_0),

$$\rho_0(u_t + uu_x) = \sigma_x, \quad \sigma_t + u\sigma_x = k^2 u_x, \qquad (3)$$

which corresponds to the coupling of a dynamical law and a Hooke law in a homogeneous medium in which the density varies very slightly in the neighborhood of the value ρ_0.

For the study of the shock waves solutions of these systems we use a nonlinear theory of generalized functions introduced in Refs. 1 and 2, in the survey[3] and in papers quoted there.

In the course of this study we were confronted with the particular case of (1) in which the stress deviator S is neglected: this is nothing else than classical hydrodynamics, which has a conservative form,

$$\rho_t + (\rho u)_x = 0, \quad (\rho u)_t + (\rho u^2 + p)_x = 0,$$
$$\mathscr{E}_t + (\mathscr{E} u + pu)_x = 0, \quad p = \phi(\rho, I). \qquad (4)$$

Our study will give, in the setting of the new theory of generalized functions, an equivalent system that has a nonconservative form. The interest of this system is that it leads very easily to efficient numerical schemes.

The numerical results thus obtained in elasticity are in agreement with the results of observations from physical experiments. Our new formulas and schemes in elasticity were at the basis of the LEA code developed in Ref. 4. They have been successfully compared with classical codes that are known to be in agreement with the experiments (see Ref. 4, Chap. 5). New schemes and convergence results for some of them may be found in Refs. 4–8.

II. THE NEW MATHEMATICAL TOOL

Although original, it is extremely elementary in the sense that the reader is only assumed to know classical differential calculus. If Ω is an open subset of $\mathbf{R}^n$ we define "new generalized functions" on Ω. Their set, denoted by $\mathscr{G}(\Omega)$, is obtained from the set $\mathscr{E}^\infty(\Omega)$ of all C^∞ functions on Ω by a

 0022-2488/88/020315-05$02.50

mathematical construction quite similar to the construction of the set $\mathbb{R}$ of all real numbers from the set $\mathbb{Q}$ of all rational numbers (by the classical method of Cauchy sequences). A precise (simplified) construction is given in Appendix A for the reader's curiosity but strictly speaking we do not need it to compute and get new physical results. It suffices to have in mind the following.

(A) One computes in $\mathscr{G}(\Omega)$ exactly as in $\mathscr{C}^\infty(\Omega)$: addition, multiplication, derivation, integration, point values, restrictions to open subsets and to subspaces, change of variables; in the same way as the basic operations in $\mathbb{R}$ are mere extensions of those in $\mathbb{Q}$, the operations in $\mathscr{G}(\Omega)$ are mere extensions of those in $\mathscr{C}^\infty(\Omega)$. Of course there are some natural differences: if $\delta\in\mathscr{G}(\mathbb{R})$ is such that its restriction to $(-\infty, 0[\cup]0, +\infty]$ is the zero function and such that $\int_{\mathbb{R}}\delta(x)dx = 1$ then clearly the point value $\delta(0)$ is not an element of $\mathbb{R}$; it is a "generalized number" in a sense made precise in our theory.

(B) Let $\mathscr{C}_0^\infty(\Omega)$ denote the set of all C^∞ functions on Ω with compact support. There are different elements $G_1, G_2\in\mathscr{G}(\Omega)$ such that for any $\Psi\in\mathscr{C}_0^\infty(\Omega)$ the integral

$$\int_\Omega (G_1(x) - G_2(x))\Psi(x)dx$$

gives—in a natural sense—the value 0. In this case we say that G_1 is associated with G_2 (notation $G_1\sim G_2$). For the reader acquainted with distribution theory *the association is the faithful generalization of the concept of equality of distributions.* It is not coherent with the multiplication, i.e., $G_1\sim G_2$ does not imply $GG_1\sim GG_2$, if $G\in\mathscr{G}(\Omega)$. Of course two C^∞ functions on Ω are associated iff they are equal. From the definition it is clear that $G_1\sim G_2 \Rightarrow DG_1\sim DG_2$ for any partial derivation D.

Thus the difference with classical analysis is that the symbol $=$ of classical analysis splits into two different symbols: the equality in $\mathscr{G}(\Omega)$ [denoted by $=$; of course $G_1 = G_2 \Rightarrow GG_1 = GG_2 \forall G\in\mathscr{G}(\Omega)$] and the association (which is a much weaker form of equality, incoherent with the multiplication).

Let us illustrate this fact by two basic examples.

(a) Let $H\in\mathscr{G}(\mathbb{R})$ be such that the restrictions of H to $(-\infty,0[$ and to $]0,+\infty)$ are, respectively, 0 and 1, and that H is everywhere bounded in a natural sense; we say that H is a Heaviside generalized function. Let n be an integer.

From the definition it follows that $H^n\sim H$. But $H^n\neq H$. [Indeed let us assume by absurd $H^n = H$. This would imply $H^{n-1}H' = (1/n)H'$, thus $HH' = \frac{1}{2}H'$: multiplication by H gives $H^2H' = \frac{1}{2}HH'$ and thus $\frac{1}{3}H' = \frac{1}{2}\frac{1}{2}H'$, i.e., $\frac{1}{3} = \frac{1}{4}$!]

(b) Consider the two possible formulations of Burger's equation,

$$u_t + uu_x = 0, \tag{5}$$

$$u_t + uu_x \sim 0, \tag{6}$$

and seek for shock wave solutions of the form

$$u(x,t) = \Delta u H(x - ct) + u_l, \quad \Delta u = u_r - u_l.$$

Then (6) gives exactly the classical weak solutions. Now any solution $u\in\mathscr{G}(\mathbf{R})$ of (5) should also be a solution of the equation $uu_t + u^2u_x = 0$. Since the velocity of the shocks for this last equation are (obvious computation) different from the ones of the shock of (5) we get at once that (5) has no shock solutions.

In classical analysis microscopic phenomena such as the behavior of a Heaviside step function at its point of discontinuity are ignored. This is the origin of the ambiguity of the multiplication of distributions. In our theory this behavior is taken into account; as a consequence we have several different Heaviside generalized functions that all differ by their microscopic behavior at 0. They are all associated between themselves but not equal ($H^n\neq H$).

III. SHOCK WAVE SOLUTIONS OF SYSTEMS IN NONCONSERVATIVE FORM

We interpret naturally (2) and (3) in the form

$$\rho_t + (\rho u)_x \sim 0, \quad (\rho u)_t + (\rho u^2)_x \sim \sigma_x,$$
$$\sigma_t + u\sigma_x \sim k^2 u_x, \tag{2'}$$
$$\rho_0(u_t + uu_x) \sim \sigma_x, \quad \sigma_t + u\sigma_x \sim k^2 u_x. \tag{3'}$$

We look for solutions of the form (steady shocks)

$$\rho(x,t) = \Delta\rho H(x - ct) + \rho_l, \quad \Delta\rho = \rho_r - \rho_l \neq 0,$$
$$u(x,t) = \Delta u K(x - ct) + u_l, \quad \Delta u = u_r - u_l \neq 0, \tag{7}$$
$$\sigma(x,t) = \Delta\sigma L(x - ct) + \sigma_l, \quad \Delta\sigma = \sigma_r - \sigma_l \neq 0,$$

where $H,K,L\in\mathscr{G}(\mathbb{R})$ are three Heaviside generalized functions; $c\in\mathbb{R}$ is the velocity of the shock, which is assumed to be constant as well as $\rho_r, \rho_l, u_r, u_l, \sigma_r, \sigma_l$. We expose the detail of the computations for (3'); since these computations are repetitive we shall not give them in detail later. Setting (7) in (3') we get at once

$$\begin{aligned} &-c\Delta u H' + (\Delta u)^2 HH' + u_l \Delta u H' \sim (\Delta\sigma/\rho_0)K', \\ &-c\Delta\sigma K' + \Delta u\Delta\sigma HK' + u_l\Delta\sigma K' \sim k^2\Delta U H'. \end{aligned} \tag{8}$$

We know that $HH' \sim \frac{1}{2}H'$ and that $H' \sim K'$ (H' and K' are two Dirac generalized functions). Thus the first line of (8) implies

$$-c + \Delta u/2 + u_l = \Delta\sigma/\rho_0\,\Delta u, \tag{9}$$

which is the classical Rankine–Hugoniot condition. Now the second line of (8)—since it is in nonconservative form—gives something new. Denoting by $\delta\in\mathscr{G}(\mathbb{R})$ any Dirac generalized function ($H' \sim K' \sim \delta$) it gives

$$HK' \sim A\delta, \quad A\in\mathbb{R} \text{ depending on } c, u_l, k^2, \Delta u, \Delta\sigma, \tag{10}$$

and

$$-c + u_l = k^2\frac{\Delta u}{\Delta\sigma} - A\,\Delta u. \tag{11}$$

Equations (9)–(11) are the jump conditions for (3'). They are ambiguous since they depend on the value of A. This ambiguity comes from the product $u\sigma_x$ in (3'). This ambiguity is not acceptable from a physical viewpoint but it is inherent to (3'). Thus (3') does not reflect completely the physical phenomenon: some additional physical piece of information is needed to get rid of the ambiguity; this will be done in Sec. V. Substituting (7) into (2') we obtain (10), (11), and the two other equations

$$k^2\frac{\Delta u}{\Delta\sigma}-\frac{1}{2}\left(\frac{1}{\rho_l}+\frac{1}{\rho_r}\right)\frac{\Delta\sigma}{\Delta u}=\left(A-\frac{1}{2}\right)\Delta u,$$
$$\rho_r\rho_l(\Delta u)^2+\Delta\sigma\Delta\rho=0. \tag{12}$$

We assume the solutions σ, u of (2') and (3') to be the same. This implies that the number A occurring in (2') and (3') is the same and

$$\frac{2}{\rho_0}=\frac{1}{\rho_l}+\frac{1}{\rho_r}. \tag{13}$$

We have obtained the value of ρ_0, but we have not yet any piece of information on A. Note that (13) had already been used in elastoplastic codes, on the basis of numerical tests.

IV. A NONCONSERVATIVE FORM OF HYDRODYNAMICS

Recall that we have two concepts of equalities: the strong one denoted by $=$, and the weak one denoted by $\sim$ (called association). One ascertains by examples [for instance by studying various forms of the systems (2), (3)] that in a system, if we write some equations (in general not all of them so as to get existence of solution) with the (strong) equality and the other ones with the association, the set of solutions is considerably reduced (relatively to the system written only with association) and sometimes we may obtain uniqueness of solutions. Such a uniqueness would be welcome to get rid of the ambiguity in Sec. III due to the quantity A coming from the "product of distribution $Y\delta$." Here we exploit this remark in the case of system (4). In the second and third equation of (4) a viscosity is neglected; this viscosity is more important in the third equation (which further includes thermal effects) than in the second one. This gives us the idea to rewrite (4) in the more precise form

$$\rho_t+(\rho u)_x=0,\quad (\rho u)_t+(\rho u^2+p)_x=0,$$
$$\mathscr{E}_t+(\mathscr{E}u+pu)_x\sim 0,\quad p=\phi(\rho,I), \tag{4'}$$

in which we use the association symbol only once. Note that a formulation like (4') does not make sense within distribution theory, in which one has only at hand one kind of equality. A deep physical reason to use (4') is not clear; probably because it has only the value of an approximation. See the Note added in proof for a clearer formulation.

Setting $v=(1/\rho)$ (v is the specific volume) and writing the state law under the form $I=\varphi(v,p)$ one obtains the following theorem.

Theorem 1: The system (4') is equivalent to the system

$$v_t+uv_x-vu_x=0,\quad u_t+uu_x+vp_x=0,$$
$$\left(\frac{1}{v}\frac{\partial\varphi}{\partial p}(v,p)\right)\cdot(p_t+up_x)+\left(p+\frac{\partial\varphi}{\partial v}(v,p)\right)u_x\sim 0. \tag{14}$$

Proof: It is an easy computation following exactly computations of classical analysis except that one is not allowed to multiply the equations with associations. Elimination of the energy by means of the state law in (4') gives the system

$$\rho_t+(\rho u)_x=0,\quad u_t+p_x/\rho+uu_x=0,$$
$$(\rho D_2\Psi)\cdot(p_t+up_x)+pu_x-\rho^2u_xD_1\Psi\sim 0,$$

in which we set $I=\Psi(\rho,p)$. Using v one gets at once (14). □

In the case of aerodynamics, $I=p/(\gamma-1)\rho$, $\gamma\in\mathbb{R}$, and then the third equation in (14) is

$$p_t+up_x+\gamma pu_x\sim 0. \tag{15}$$

Most media may be modeled by a Mie–Gruneisen state law $p=\gamma\rho I+F(v)$, $\gamma>0$ and F a positive function of v. The third equation in (14) becomes

$$p_t+up_x+[(\gamma+1)p-F(v)-vF'(v)]u_x\sim 0. \tag{15'}$$

If (14) were written with three association symbols then it would have ambiguous jump conditions (it is in nonconservative form); since it is equivalent to the conservative system (4') then it is clear it has nonambiguous jump conditions. The solutions of (14) of the form (steady shocks)

$$v(x,t)=\Delta vH(x-ct)+v_l,\quad \Delta v=v_r-v_l,$$
$$u(x,t)=\Delta uK(x-ct)+u_l,\quad \Delta u=u_r-u_l, \tag{16}$$
$$p(x,t)=\Delta pL(x-ct)+p_l,\quad \Delta p=p_r-p_l,$$

with H, K, and L Heaviside generalized functions, are characterized by the following theorem.

Theorem 2: v,u,p in (16) are solutions of (14) if and only if $H=K=L$ [plus obviously the relations obtained by substituting (16) with $H=K=L$ in (14); these are the classical jump conditions of (4)].

Proof: The result follows from the first two equations in (14). Indeed the first one gives

$$(-c+u_l+\Delta uK)H'-(\Delta uK')H$$
$$-v_l(\Delta u/\Delta v)K'=0.$$

This equation gives H as a function of K since the equation $a(x)y'+b(x)y+c(x)=0$ is solved in $\mathscr{G}(\mathbb{R})$ as in the classical case. The second equation in (14) gives a similar result. Finally after some computations one obtains $H=K=L$. □

V. THE JUMP FORMULAS IN ELASTICITY

For the same reason as above we naturally state the system (1) with the equality in $\mathscr{G}(\mathbb{R}^2)$ for mass and momentum conservation and with the association for the other equations. One may reproduce the proofs of Theorem 1 and Theorem 2 in which p is replaced by $p-S=-\sigma$, since it is based only on the equalities [in $\mathscr{G}(\mathbb{R}^2)$] in the first two equations. One obtains that v, u, and σ are represented by exactly the same Heaviside generalized functions. As a consequence the following theorem is established.

Theorem 3: The jump conditions of the systems (2) and (3), when considered as models in elasticity, are given by the formulas in Sec. III in which $A=\frac{1}{2}$.

Another argument for this result is presented in Ref. 9.

Remark 1: In Appendix B we show that in other physical circumstances one has values of the quantities like A that may be completely different from $\frac{1}{2}$; there we develop an example in elastoplasticity. The same conclusion is obtained in hydrodynamics with viscosity.[10]

Remark 2: Equations (9) and (11) give at once

$$c-\frac{u_r+u_l}{2}=-\frac{1}{\rho_0}\frac{\Delta\sigma}{\Delta u},$$
$$c-\frac{u_r+u_l}{2}=-k^2\frac{\Delta u}{\Delta\sigma}+\left(A-\frac{1}{2}\right)\Delta u.$$

With the value $A = \frac{1}{2}$ one gets

$$(c - (u_r + u_l)^2/2) = k^2/\rho_0 ,$$

which is used by experimentalists to obtain the value of k^2 from measurements of c, u_r, u_l, and ρ_0.

VI. NUMERICAL SCHEMES, COMPARISON WITH EXPERIMENTS

The fact that the quantities v, u, p in hydrodynamics or v, u, σ in elasticity are represented by the same Heaviside generalized function (we say that these quantities "vary in phase" on a shock) is very interesting as it facilitates the construction of efficient numerical schemes. In this way we have obtained numerical schemes adapted to Hooke's law and more generally to the approximation of nonconservative terms appearing in hydrodynamics and elastodynamics. These results have been used in the generation of the LEA elastoplastic code and in modifications on the HULL code. In hydrodynamics the system (14) [with (15) or (15') as third equation] is very convenient from the numerical viewpoint. In particular, one obtains good pressure predictors that are essential to generate performing hydrodynamics codes. See Refs. 4, 7, and 8.

In the case of the system (3) one can even obtain mathematical proofs of the convergence of the numerical schemes to a generalized solution $u,\sigma \in \mathcal{G}(\mathbb{R}^2)$; see Refs. 4–6.

Comparison of the jump conditions obtained in Sec. V with the experiments has been (indirectly) done in Ref. 4 Chap. 5, through comparison with classical codes (which are known to agree with the experimental results). The conclusion is a very good agreement.

Note added in proof: A more natural method to resolve the ambiguity consists in stating the basic laws of physics with the equality in $\mathcal{G}$ (thus postulating their validity even in the small width of the shock) and in stating the constitutive equations with the association (thus accepting their possible nonvalidity inside the width of the shock, but also excluding too singular behaviors). Although different from (4'), this formulation leads easily to (14) when the unknowns are assumed to be piecewise continuous.

ACKNOWLEDGMENT

This work has been done under contracts with the Centre d'Etudes de Gramat (CEG), Délégation Générale pour l'Armement.

APPENDIX A: A MATHEMATICAL CONSTRUCTION

We give here a very simplified definition of the new generalized functions, in which the inclusion of $\mathcal{D}'(\Omega)$ into $\mathcal{G}(\Omega)$ is dropped but which is enough for the application in this paper. We define a set $\mathcal{E}_M[\Omega]$ (here the letters $\mathcal{E}$ and M form a unique symbol and are not to be dissociated) as the set of the maps

$$R:\ ]0,\ 1] \times \Omega \to \mathbb{C}$$
$$\epsilon,\ x \qquad R(\epsilon,x)$$

such that (i) for any $\epsilon > 0$, the map $x \to R(\epsilon,x)$ is a C^∞ function of the variable $x \in \Omega$, and (ii) if

$$D = \frac{\partial^{|k|}}{\partial x_1^{k_1} \cdots \partial x_n^{k_n}}$$

(D = identity if $|k| = 0$) and if K is any compact subset of Ω, then there exists an integer $N \in \mathbb{N}$, and two constants $c > 0$ and $\eta > 0$, $0 < \eta < 1$, such that

$$\sup_{x \in K} |DR(\epsilon,x)| \leqslant c/\epsilon^N ,$$

if $0 < \epsilon < \eta$.

If $R \in \mathcal{E}_M[\Omega]$ then it is obvious that $DR \in \mathcal{E}_M[\Omega]$. Note also that $\mathcal{E}_M[\Omega]$ is a vector space and that the pointwise product R_1R_2 is an element of $\mathcal{E}_M[\Omega]$ provided $R_1 \in \mathcal{E}_M[\Omega]$ and $R_2 \in \mathcal{E}_M[\Omega]$.

We denote by $\mathcal{N}[\Omega]$ the set of all the elements $R \in \mathcal{E}_M[\Omega]$ that have the property that for all D and K (as above) there exists an integer $N \in \mathbb{N}$ such that $\forall_q \geqslant N\ \exists c_q, \eta_q > 0$ such that

$$\sup_{x \in K} |DR(\epsilon,x)| \leqslant c_q \epsilon^q ,$$

if $0 < \epsilon < \eta_q$.

If $R \in \mathcal{N}[\Omega]$ then it is obvious that $DR \in \mathcal{N}[\Omega]$. $\mathcal{N}[\Omega]$ is a vector subspace of $\mathcal{E}_M[\Omega]$ but we have much more: if $R_1 \in \mathcal{E}_M[\Omega]$ and $R_2 \in \mathcal{N}[\Omega]$ their pointwise product R_1R_2 is in $\mathcal{N}[\Omega]$ (i.e., $\mathcal{N}[\Omega]$ is an ideal of the algebra $\mathcal{E}_M[\Omega]$).

We define the generalized functions on Ω as the elements of the quotient algebra

$$\mathcal{G}(\Omega) = \mathcal{E}_M[\Omega]/\mathcal{N}[\Omega] .$$

In the classical construction of $\mathbb{R}$ from $\mathbb{Q}$ by the method of Cauchy sequences the set $\mathcal{E}_M[\Omega]$ is replaced by the set of all Cauchy sequences of rational numbers and the set $\mathcal{N}[\Omega]$ is replaced by the set of all null sequences (i.e., the sequences that converge to 0) of rational numbers. If $f \in \mathcal{C}^\infty(\Omega)$ then we consider the map $R \in \mathcal{E}_M[\Omega]$ defined by $R(\epsilon,x) = f(x)$ and thus we obtain at once an inclusion of $\mathcal{C}^\infty(\Omega)$ into $\mathcal{G}(\Omega)$. Now let $\rho \in \mathcal{D}(\mathbb{R})$ be a $\mathcal{C}^\infty$ function such that $\int \rho(x)dx = 1$. Then the class in $\mathcal{G}(\mathbb{R})$ of the map $R(\epsilon,x) = (1/\epsilon)\rho(x/\epsilon)$ is an element of G of $\mathcal{G}(\mathbb{R})$ that is associated with the Dirac delta function [i.e., $\forall \Psi \in \mathcal{D}(\mathbb{R})$,

$$\int R(\epsilon,x)\Psi(x)dx \to \Psi(0)$$

when $\epsilon \to 0$]. Two generalized functions $G_1, G_2 \in \mathcal{G}(\Omega)$ are associated (i.e., $G_1 \sim G_2$) if there are representatives R_1, R_2 of G_1, G_2, respectively, such that $\forall \Psi \in \mathcal{D}(\Omega)$

$$\int_\Omega (R_1(\epsilon,x) - R_2(\epsilon,x))\Psi(x)dx \to 0$$

when $\epsilon \to 0$. (Then the same holds for any representatives of G_1 and G_2.)

APPENDIX B: ELASTOPLASTIC SHOCK WAVES

There are shock waves in which the medium passes from the elastic state into the plastic state; before the shock the medium is elastic and after the shock it is plastic. These shocks may be modeled by the system (1) in which μ depends on S: $\mu(S) = \mu_0 \in \mathbb{R}$ if $|S| \leqslant S_0$ and $\mu(S) = 0$ if $|S| = S_0$ (thus $|S|$ cannot reach values $> S_0$).

If, as done intuitively within our theory, we consider that the shocks have an infinitely small—but not exactly

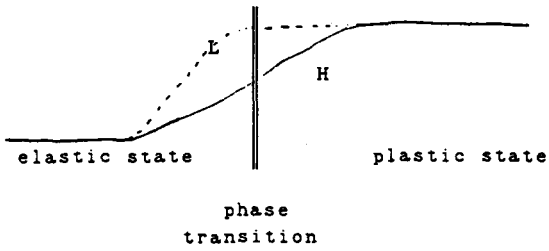


FIG. 1. "Microscopic aspect" of an elastoplastic shock wave.

null—width, then the Heaviside functions of the quantities like σ,u,ρ have the following aspect. (See Fig. 1.)

The continuous line corresponds to any Heaviside function H of $u,\rho,p,\mathscr{E}$: they vary throughout the shock. The dotted line corresponds to the Heaviside function L of S: it varies only in the elastic state. Thus one sees easily, in case the phase transition takes place rather early within the shock, that L and H are quite different: the Dirac function L' has its support located in a region in which H is nearly equal to 0. Thus we have $HL' \sim B\delta$ with $B > 0$ rather close to 0 (this affects the jump conditions, which are observable effects). This shows that L and H need absolutely to be considered as quite different mathematical objects, while there is only one Heaviside function in distribution theory.

[1]J. F. Colombeau, *New Generalized Functions and Multiplication of Distributions* (North-Holland, Amsterdam, 1984), p. 375.
[2]J. F. Colombeau, *Elementary Introduction to New Generalized Functions* (North-Holland, Amsterdam, 1985), p. 280.
[3]J. F. Colombeau and A. Y. Le Roux, "Generalized functions and products appearing in equations of physics," to be submitted.
[4]J. J. Cauret, Thèse, Bordeaux, 1986.
[5]J. J. Cauret, J. F. Colombeau, and A. Y. Le Roux, "Solutions généralisées discontinues de problèmes hyperboliques non conservatifs," C. R. Acad. Sci. (Paris) **302**, 435 (1986).
[6]J. J. Cauret, J. F. Colombeau, and A. Y. Le Roux, "Discontinuous generalized solutions of nonlinear nonconservative hyperbolic equations," in submission.
[7]J. F. Colombeau and A. Y. Le Roux, "Numerical techniques in elastodynamics," Proceedings of an advanced research workshop held in Saint Etienne, 1986, *Lecture Notes in Mathematics*, Vol. 1270 (Springer, Berlin, 1987), pp. 103–114.
[8]J. F. Colombeau and A. Y. Le Roux, "Numerical methods for hyperbolic systems in nonconservative form, using products of distributions," *Advances in Computer Methods for PDE's* VI (IMACS, 1987), pp. 28–37.
[9]J. F. Colombeau, A. Y. Le Roux, and B. Perrot, "Multiplications de distributions et ondes de choc elastiques ou hydrodynamiques en dimension un," C. R. Acad. Sci. (Paris) **305**, 453 (1987).
[10]J. Aragona, J. F. Colombeau, and C. K. Raju, "Products of distributions and shocks in viscous media," in submission.

Studies in space-times admitting two spacelike Killing vectors

M. Halilsoy
Nuclear Engineering Department, King Abdulaziz University, P. O. Box 9027, Jeddah-21413, Saudi Arabia

(Received 25 June 1986; accepted for publication 30 September 1987)

Some properties of the space-times admitting two spacelike Killing vectors are studied. In particular, using harmonic maps the degree of freedom on the M' manifold is exploited to add scalar and electromagnetic fields to Bonnor's nonsingular solution. It is also shown that for vacuum space-times the noncommutativity of two spacelike Killing vectors is incompatible with the self-similarity requirement and such a self-similar vacuum space-time has no Taub–NUT equivalent extension.

I. INTRODUCTION

Static axially symmetric fields with two commuting Killing vectors were considered first by Weyl[1] who, under the assumed symmetry, presented a general solution to the Einstein field equations. Einstein and Rosen[2] studied later the intrinsically similar fields with two commuting spacelike Killing vectors which amounted to the cylindrically symmetrical gravitational waves. The imploding–exploding waves interpretation, as an example of scattering of such cylindrical waves, was given by Marder.[3] The same metric was considered independently both by Weber and Wheeler[4] and by Bonnor.[5] In particular, Bonnor gave an example of a nonsingular solution to the field equations that might be interpreted as a cosmological model of interest. The term "nonsingular," however, must be taken cautiously since, as explained in detail by Bonnor and by Weber and Wheeler, the fact that the metric tensor should behave like $1/r$ and the Riemann tensor like $1/r^2$ (as the requirements of asymptotic flatness) is not satisfied in such a solution. The well-known result that no nonsingular colliding plane wave (CGW) space-time exists[6] could be anticipated from the above case due to the inherent similarily between planar and cylindrical geometries. In other words both of these geometries are represented by the same metrics in a particular choice of coordinates but the boundary conditions differ and therefore the difference is a global one.

In this paper we make use of the same cylindrically symmetric metric to generate radiation sources, such as electromagnetic and massless scalar fields. The method we adopt in the solution generating technique was presented briefly earlier[7] and for the sake of completeness we shall review it here. In this new approach of harmonic maps we reduce the general relativistic problem to the one of classical field theory and it is our belief that this method adds considerable elegance and simplicity when compared to the other existing methods. An effective Lagrangian is introduced via the harmonic maps between the suitably chosen Riemannian manifolds. For a general review of the physics of harmonic maps we refer to the paper of Misner,[8] whereas for the mathematical aspects the paper of Eells and Sampson[9] provides the proper references to be consulted. To a certain extent we shall make use of the cylindrical wave line element with single polarization due to Einstein and Rosen,

$$ds^2 = e^{2(\gamma-\Psi)}(dt^2 - d\rho^2) - \rho^2 e^{-2\Psi}\, d\phi^2 - e^{2\Psi}\, dz^2, \qquad (1)$$

where the metric functions Ψ and γ are only functions of t and ρ. The vacuum field equations, for later reference, are given by

$$\Psi_{\rho\rho} + (1/\rho)\Psi_\rho - \Psi_{tt} = 0, \qquad (2)$$

$$\gamma_{\rho\rho} - \gamma_{tt} = \Psi_t^2 - \Psi_\rho^2, \qquad (3)$$

whereas the integrability conditions are

$$\gamma_\rho = \rho(\Psi_\rho^2 + \Psi_t^2), \qquad (4)$$

$$\gamma_t = 2\rho\Psi_\rho\Psi_t. \qquad (5)$$

Solution of this set of equations is usually carried out by solving (2) first. This is the cylindrical wave equation that admits wave solutions. Following Bonnor, we solve (2) by the method of complex translation discovered first by Appell in 1887. Having known that $\Psi = (\rho^2 - t^2)^{-1/2}$ forms a solution, then Appell's theorem states that the real part of the complex function, $\Psi = [\rho^2 - (t - ic)^2]^{-1/2}$ ($c = \text{const}$), is also a solution to (2). Bonnor's final solution is expressed by

$$\Psi = [u + (u^2 + w^2)^{1/2}]^{1/2}[u^2 + w^2]^{-1/2},$$
$$\gamma = \frac{-\rho^2(u^2 - w^2)}{2(u^2 + w^2)^2} + \frac{1}{4c^2}\left[\frac{u}{(u^2 + w^2)^{1/2}} + 1\right], \qquad (6)$$

where

$$u = \rho^2 - t^2 + c^2, \quad w = 2ct.$$

The regularity of this solution should be understood in the sense that no metric function or scalars from the Riemann tensor diverge for $\rho \to \infty$ and $t \to \pm\infty$, where the ranges are $0 \leqslant \rho < \infty$, $-\infty < t < +\infty$.

II. METHOD FOR GENERATING NEW SOLUTIONS

It can easily be verified that Eqs. (2) and (3) follow from the variational principle of the action

$$I[\Psi,\gamma,\lambda] = \int [\gamma_\rho\lambda_\rho - \gamma_t\lambda_t - \lambda(\Psi_\rho^2 - \Psi_t^2)]\, d\rho\, dt, \qquad (7)$$

where $\lambda = \rho$ is to be imposed as a coordinate condition subsequent to the variation. We recall from the theory of harmonic mappings of Riemannian manifolds that this action is in the form of an energy functional

$$E[f] = \int g'_{AB}\frac{\partial f^A}{\partial x^b}\frac{\partial f^B}{\partial x^b} g^{ab}|g|^{1/2}\, d^2x, \qquad (8)$$

 0022-2488/88/020320-07$02.50

where the manifolds M and M' are chosen, respectively, by

$$M:\ ds^2 = d\rho^2 - dt^2 + \lambda^2\, d\phi^2$$
$$= g_{ab}\, dx^a\, dx^b \quad (a,b:\ 1,2,3), \tag{9}$$
$$M':\ ds'^2 = (d\lambda/\lambda)d\gamma - d\Psi^2$$
$$= g'_{AB}\, dx'^A\, dx'^B \quad (A,B:\ 1,2,3). \tag{10}$$

Note that since $\lambda = \rho$ is a coordinate condition and ϕ is a cyclic variable, the effective dimensions of both M and M' are two. The map is represented by $f^A = \{\Psi,\gamma\}: M \to M'$, and the stationary conditions

$$\delta E\,[f^A] = 0 \tag{11}$$

are the statements that the maps are harmonic. As a side remark we note that in contrast to various maps in mathematics, harmonic maps explicitly can not be known *a priori* until Eqs. (11) are solved explicitly. Naturally all the information expected from a standard variational principle can also be extracted from the harmonic map action as well. For instance, besides the stationary requirement one may check the stability of a general relativistic manifold by studying the second variation due to Jacobi. In this paper we shall restrict ourselves only to the first variations. The theory of harmonic maps in general relativity can be summarized in the following: choose two Riemannian manifolds (9) and (10) in such a way that when the energy functional (8) is constructed from them, its stationary requirements (11) coincide exactly with the vacuum Einstein equations under consideration.[10]

Now, if f^A is a solution of the field equations (11), then a new solution f^K is obtained as a function of f^A from the isometry (invariance) of the line element of the M' manifold. This amounts to

$$ds'^2 = g'_{AB}(f)\,df^A\,df^B = \tilde{g}'_{KL}(\tilde{f})\,d\tilde{f}^K\,d\tilde{f}^L, \tag{12}$$

which yields the implicit relations

$$g'_{AB}(f)\frac{\partial f^A}{\partial \tilde{f}^K}\frac{\partial f^B}{\partial \tilde{f}^L} = \tilde{g}'_{KL}(\tilde{f}). \tag{13}$$

We note that such an isometry does not necessarily imply that the metric tensor $\tilde{g}'_{KL}$ has the same dimensions as that of $\tilde{g}'_{AB}$. In particular, we shall consider the case where the ranges of the indices K,L are larger than A,B and we shall interpret this as a problem of embedding. The mathematical details of embeddings are not our purpose here. We would like rather to make use of these concepts in order to yield tangible results that may prove useful in physics, and particularly in general relativity. We face embeddings in particular when we want to generate solutions with radiation sources from known solutions of vacuum. The method of isometry applies best to the two-dimensional problems and the reason for this may be connected with the existence of conformal techniques and analyticity in this particular dimension.

In order to obtain a new vacuum solution from a known solution we have to find Killing vectors of the corresponding M' manifold at hand. However, not every Killing vector yields a significantly new solution other than the original one. It is instructive at this point to mention a particularly well-known example. Stationary symmetrical gravitational fields (SAS) can be handled as a reduced formalism due to Ernst[11] or equivalently by the method due to Matzner and Misner.[12] In the latter method reduced Einstein equations are obtained from the variational principle of the map, $f^A = \{X,Y\}: M \to M'$, where

$$M:\ ds^2 = d\rho^2 + dz^2 + \rho^2\, d\phi^2, \tag{14}$$
$$M':\ ds'^2 = (dX^2 + dY^2)/X^2. \tag{15}$$

The Lagrangian density and the corresponding three Killing vectors of this M' are given by

$$L = \rho[(\nabla X)^2 + (\nabla Y)^2]/X^2, \tag{16}$$

and

$$\xi_1 = \partial_Y,$$
$$\xi_2 = 2XY\,\partial_X + (Y^2 - X^2)\partial_Y, \tag{17}$$
$$\xi_3 = X\,\partial_X + Y\,\partial_Y,$$

respectively. The new solution that is generated from the isometry can be expressed by

$$\frac{df^A}{dt} = \alpha\xi_i f^A \quad (i = 1,2,3), \quad f^A = \{X,Y\} \quad (0 \leqslant t \leqslant 1), \tag{18}$$

where α is a constant, and t is a continuous parameter. For $t = 0$ we recover our old solution f^A, whereas for $t = 1$, the new solution $\tilde{f}^A$ is obtained. It turns out in this example that ξ_3 leads to a scale factor and therefore the isometry it generates results in the identity of two solutions (old and new). Similarly, ξ_1 also is not very interesting, but the vector ξ_2 results in a significant, new solution. Linear combinations of Killing vectors may lead to interesting results so that such cases should be investigated as well. The three Killing vectors in the example above in fact arise from the invariance under fractional linear transformations with three parameters when the harmonic map Lagrangian is expressed in Ernst's complex formulation. We have already stated that the two methods are equivalent.

The proposed method of generating new solutions can naturally be extended to cover the cases of different dimensional isometric transformations. We can imbed an M' manifold into a new manifold of higher dimensionality such that the new dimensions can be interpreted as the energy-momentum tensor due to some radiation sources. The idea of imbedding the configuration space into a larger dimension does not emerge here as a novel one, since the same procedure had been employed in particle physics long ago.[13] To conclude this section we would also like to add that the transitive property of isometric transformations provides us with additional means to find possible isometric solutions. Only when isometry is applied to a unique solution in a particular class of solutions (such as Schwarzschild, Kerr) does it fail to yield anything new.

III. RADIATION SOURCES WITHOUT SOURCES

In this section we shall exploit the degrees of freedom on the M' manifold to generate electromagnetic and massless scalar radiations as the source to a modified gravitational background. Let $\Gamma^{\mu}_{\nu\rho}$ be the Christoffel symbol of a pure gravitational space-time, so that the geodesic equation is giv-

en by

$$\frac{d^2x^\mu}{d\tau^2}+\Gamma^\mu_{\nu\rho}\frac{dx^\nu}{d\tau}\frac{dx^\rho}{d\tau}=0. \tag{19}$$

When an Einstein–Maxwell (EM) solution is generated from a vacuum solution the new geodesic equation should read

$$\frac{d^2x^\mu}{d\tau^2}+\Gamma'^\mu_{\nu\rho}\frac{dx^\nu}{d\tau}\frac{dx^\rho}{d\tau}=\frac{q}{m}F^\mu_\alpha u^\alpha, \tag{20}$$

where F^μ_α stands for the "induced" (as transmuted) electromagnetic (e.m.) field and Γ' is the new Christoffel symbol. The two geodesic equations should coincide, in reality, since the space-time has undergone a dual interpretation under which the actual physics of the overall process must remain invariant. Each of our results can be stated as a theorem.

Theorem 1: Given that the vacuum Einstein equations can be represented by the harmonic map between the metrics

$$ds^2=d\rho^2-dt^2+\lambda^2\,d\phi^2, \tag{21}$$

$$ds'^2=(1/\lambda)d\lambda\,d\gamma-d\Psi^2, \tag{22}$$

then a system of Einstein–Maxwell (EM) coupled equations can be represented by the harmonic map between

$$ds^2=d\rho^2-dt^2+\lambda^2\,d\phi^2, \tag{23}$$

and

$$ds'^2=(1/\lambda)d\lambda\,d\gamma-(d\mu^2+e^{-\mu}\,dA^2). \tag{24}$$

The space-time metric reads

$$ds^2=e^{2\gamma-\mu}(dt^2-d\rho^2)-\rho^2e^{-\mu}\,d\phi^2-e^\mu\,dz^2 \tag{25}$$

and the isometry of the line elements (22) and (24) yields the constraint condition

$$d\mu^2+e^{-\mu}\,dA^2=4\,d\Psi^2. \tag{26}$$

(Note that we have introduced the factor 4 by scaling the functions μ and A by by $\frac{1}{2}$.)

Proof: The Lagrangian density obtained from the map between the metrics in (21) and (22) is given by

$$L_1=(\lambda_\rho\gamma_\rho-\lambda_t\gamma_t)-\lambda(\Psi_\rho^2-\Psi_t^2) \tag{27}$$

which yields the vacuum equations (2) and (3). The Lagrangian density obtained from the map between (23) and (24) is given by

$$L_2=(\lambda_\rho\gamma_\rho-\lambda_t\gamma_t)-\lambda\,[\mu_\rho^2-\mu_t^2+e^{-\mu}(A_\rho^2-A_t^2)]. \tag{28}$$

The Euler–Lagrange (EL) equation, $\delta L_2/\delta\gamma=0$, holds true by virtue of the choice $\lambda=\rho$. Next is the equation $\delta L_2/\delta A=0$, being equivalent to

$$(\rho e^{-\mu}A_t)_t-(\rho e^{-\mu}A_\rho)_\rho=0, \tag{29}$$

which stands for the only nontrivial Maxwell equation. To verify this, define the e.m. four-potential

$$A_\mu=(0,0,0,A), \tag{30}$$

so that $F_{tz}=A_t$ and $F_{\rho z}=A_\rho$ are the nonvanishing components of the e.m. field tensor. It can be checked that the source-free Maxwell equation

$$\partial_\mu(\sqrt{-g}F^{\mu\nu})=0 \tag{31}$$

coincides with (29), where the metric is (25).

Finally, the remaining EL equation, $\delta L_2/\delta\mu=0$, turns out to be identical with the EM equation,

$$\mu_{\rho\rho}-\mu_{tt}+(1/\rho)=\tfrac{1}{2}(A_t^2-A_\rho^2)e^{-\mu}. \tag{32}$$

This completes the proof that if L_1 describes a system of vacuum equations, then L_2 describes an EM system.

In order to see the significance of the foregoing theorem we generate some new EM solutions from the vacuum ones. To this end we solve first the constraint condition (26). It turns out that this equation, similar to taking the roots of unity, possesses a large class of solutions where μ and A are expressed as functions of Ψ. A particular integral of the constraint equation,

$$\begin{aligned}e^\mu&=(\alpha+\tfrac{1}{2}\beta^2)\operatorname{sech}^2\Psi,\\ A&=2(\alpha+\tfrac{1}{2}\beta^2)^{1/2}\tanh\Psi,\end{aligned} \tag{33}$$

where α,β are nonzero constants, was reported a long time ago by Misra.[14] In addition to this solution we present two more classes of solutions as follows.

(i) Let $A=2b_0e^{\mu/2}$, where $b_0=$ const. The integration of the constraint equation yields

$$\mu=\pm 2(1+b_0^2)^{-)1/2}\Psi.$$

In contrast to Misra's solution (33), this new solution has the feature that it has vacuum Einstein as a limit.

(ii) Letting $A=k\mu$ ($k=$ const), we obtain from the constraint equation the transcendental relation

$$\frac{(1+k^2e^{-\mu})^{1/2}+1}{(1+k^2e^{-\mu})^{1/2}-1}=\exp[\pm 2\Psi+2(1+k^2e^{-\mu})^{1/2}]. \tag{34}$$

Being transcendental, this expression cannot be inverted for μ analytically in terms of Ψ. In fact, the constraint condition (26) possesses a large class of solutions sharing this transcendental nature.

Theorem 2: Given that the harmonic map between the manifolds (21) and (22) yields vacuum equations, then Einstein-massless scalar field equations can be generated from the map between the metrics

$$ds^2=d\rho^2-dt^2+\lambda^2\,d\phi^2, \tag{35}$$

$$ds'^2=(1/\lambda)d\lambda\,d\gamma-(d\mu^2+k\,d\phi^2), \tag{36}$$

where k is the coupling constant.

Proof: The Lagrangian density for this map is given by

$$L_3=(\lambda_\rho\gamma_\rho-\lambda_t\gamma_t)-\lambda\,[\mu_\rho^2-\mu_t^2+k(\phi_\rho^2-\phi_t^2)]. \tag{37}$$

We must show now that L_3 describes an Einstein-scalar system whereas the space-time metric is still (25). The constraint relation is expressed now by

$$d\mu^2+kd\phi^2=4\,d\Psi^2. \tag{38}$$

(Note here also that 4 is a scale factor.)

EL equations for the scalar field ϕ are given by

$$\partial_\mu(\sqrt{-g}g^{\mu\nu}\phi_\nu)=0 \tag{39}$$

or equivalently

$$(\rho\phi_t)_t-(\rho\phi_\rho)_\rho=0. \tag{40}$$

Einstein-scalar equations are obtained by the conditions $\delta L_3/\delta\mu=0$ and $\delta L_3/\delta\lambda=0$, and therefore L_3 forms a La-

grangian for an Einstein-scalar system.

By making use of the constraint condition it is not hard to obtain Einstein-scalar solutions. Choosing $\mu = 2\Psi \cos a_0$, $k^{1/2}\phi = 2\Psi \sin a_0$ ($a_0 = \text{const}$), by virtue of the vacuum Eqs. (2), (3), Einstein scalar equations are satisfied. The constant a_0 here plays the role of a phase constant which removes the scalar field for $a_0 = 0$. In the case of static spherically symmetric scalar fields, the corresponding solution obtained by similar means employed here is the Newman–Janis–Winicour (NJW) solution.[15] In a routine manner the uniqueness argument of NJW can be extended to the scalar solution obtained here in the cylindrically symmetric geometry.

Theorem 3: The two foregoing theorems (1) and (2) can be combined to yield a Lagrangian for the EM-scalar field system. The harmonic map will now be between the manifolds,

$$ds^2 = d\rho^2 - dt^2 + \lambda^2 d\phi^2, \tag{41}$$

$$ds'^2 = (1/\lambda)d\lambda\, d\gamma - [d\mu^2 + e^{-\mu} dA^2 + k\, d\phi^2]. \tag{42}$$

Proof: The effective Lagrangian density of the map between the given manifolds (41) and (42) will be

$$L_4 = \lambda_\rho \gamma_\rho - \lambda_t \gamma_t - \lambda\,[\mu_\rho^2 - \mu_t^2 + e^{-\mu}(A_\rho^2 - A_t^2) + k(\phi_\rho^2 - \phi_t^2)] \tag{43}$$

and the constraint condition will be given by

$$d\mu^2 + e^{-\mu} dA^2 + k\, d\phi^2 = 4\, d\Psi^2. \tag{44}$$

EL equations for L_4 with respect to each function will yield all EM-scalar field equations. The proof follows therefore from the foregoing theorems.

The following solution, for example, solves the constraint condition (44) and therefore constitutes also a solution for the EM-scalar system,

$$\begin{aligned} &\mu = 2\Psi \cos b_0, \\ &A = 4 \exp(\Psi \cos b_0)\cdot \cos c_0 \tan b_0, \\ &k^{1/2}\phi = 2\Psi \sin b_0 \sin c_0 \\ &\qquad (b_0, c_0\text{: constants}). \end{aligned} \tag{45}$$

One observes simply that $c_0 = 0$ implies that only the e.m. field exists and $b_0 = 0$ leaves only the scalar field. Vacuum is recovered for $b_0 = 0 = c_0$.

Finally we would like to note that the e.m. field adopted in the foregoing solutions was of the form $A_\mu = \delta_A^x A = (0,0,0,A)$. This may be extended to the case with two nonvanishing components, given as $A_\mu = \delta_\mu^x A + \delta_\mu^y B = (0,0,B,A)$. By this latter choice, however, the constraint condition to be solved becomes

$$d\mu^2 + e^{-\mu} dA^2 + (1/\rho)e^{\mu} dB^2 = 4\, d\Psi^2, \tag{46}$$

whose particular integrals are rather involved compared with the former case where $B = 0$.

IV. TWO REMARKS ON BONNOR'S SOLUTION

(1) In this section we derive an equation for the timelike geodesics where the space-time element is being projected onto the (ρ,t) plane. In other words we simplify the general geodesic equation

$$\frac{d^2x^\mu}{d\tau^2} + \Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{d\tau}\frac{dx^\beta}{d\tau} = 0 \tag{47}$$

for the particular case of $\phi = z = 0$, and where the cylindrical radius is to be parametrized by t. For this purpose we choose the following variational principle to yield directly the projected geodesic equation:

$$I = \int ds = \int e^Z (1 - \dot\rho^2)^{1/2}\, dt, \tag{48}$$

where $\dot\rho = d\rho/dt$, and $Z = \gamma - \Psi$. As it is already implied by this reduced action principle we can study the cases for $\dot\rho < 1$, i.e., the timelike geodesics. The resulting equation for geodesics is obtained as

$$\ddot\rho = (\dot\rho^2 - 1)\left(\dot\rho\frac{\partial Z}{\partial t} + \frac{\partial Z}{\partial \rho}\right). \tag{49}$$

Unfortunately, the relative simplicity of this equation does not help in the search for an analytic solution for ρ as a function of time. The difficulty originates from the rather complicated form of $Z \equiv \gamma - \Psi$, in Bonnor's solution. A numerical solution, however, can be achieved by assigning values for $\dot\rho$ in the interval $0 < \dot\rho < 1$ and plotting the resulting ρ for arbitrary values of the running time. In this way we find the trajectory of a particle in the nonsingular cosmological model given by Bonnor.

(2) Our second remark concerns the physical meaning of the nonzero constant c in Bonnor's solution (6). (Note that we have fixed the other constant b that appears in the original solution[5] by $b = 1$.) We want to explain that this constant c (and b) is not connected with the topology of the cosmological model. The degree of harmonic maps for the case of S^2 into S^2, as had been shown by Eells and Sampson, turns out to be finite and gives the number of windings that the base manifold is being wrapped. The energy of the map also emerges as proportional to the same topological integer. The integer property of the map arises from the uniqueness requirements of the rotational components of the map. All such nice topological features, however, can hardly find room in general relativity. The reason can be attributed to the noncompact, hyperbolic nature of Riemannian manifolds. To see the inherent difference between the compact and noncompact manifolds, from the physics point of interest, we refer to the analysis of Hirayama *et al.*[16] In this reference it is explained that for Heisenberg's ferromagnet the number of slips of the spin vector equals the degree of the harmonic map. The same analysis, on the other hand, when applied to the Weyl (or TS) class of gravitational fields, yields a divergent result. Having learned also from the two-dimensional field theories[17] that the topological class does not change in the course of time, we can handle Bonnor's cosmological model as a one-dimensional field theory on a flat background. An index can be defined for Bonnor's Ψ field by an expression proportional to $\int_0^\infty \rho \Psi_\rho\, d\rho = 1 \sinh^{-1}(\infty)$, which diverges unless an infinite factor is subtracted.

In the Weyl case, the scalar field propagating on flat space is given by $\Psi = \tanh^{-1}\xi = \delta \tanh^{-1} x$. Here, ξ is the real version of the Ernst potential, x is one of the prolate

spheroidal coordinates ($1<x<\infty$), and δ is the Weyl parameter. A topological index could be defined from Ψ, provided $\Psi(\infty)-\Psi(1)$ is a finite number. It turns out that before one accepts δ as the topological degree one has to divide (or subtract) by an infinite factor, since $\Psi(1)$ diverges.

Comparing the two cases it seems that Bonnor's solution is the first member of a larger family, yet to be discovered and the corresponding parameter of Weyl's δ will characterize the topological class, albeit in some ambiguous way.

V. THERE IS NO HYPERSURFACE NONORTHOGONAL SELF-SIMILAR COSMOLOGICAL VACUUM MODEL

The general space-time geometry that describes cylindrical gravitational waves with the cross polarization term is given by[18]

$$ds^2 = e^{2(\gamma-\Psi)}(dt^2-d\rho^2) - e^{2\Psi}(dz+w\,d\phi)^2-\rho^2e^{-2\Psi}\,d\phi^2, \tag{50}$$

which is considered as a generalization of the Einstein–Rosen metric. From the inherent identity between cylindrical and planar geometries this metric can be transformed into the metric that describes colliding plane gravitational waves. This latter metric due to Szekeres[19] is given by

$$ds^2 = 2e^{-M}\,du\,dv - e^{-U}[e^{V}\cosh W\,dx^2 + e^{-V}\cosh W\,dy^2 - 2\sinh W\,dx\,dy]. \tag{51}$$

(1) It is our purpose to show now that this metric admits no self-similar solutions, simply because whenever it does, it turns out to be diagonalized. By the self-similar solution, here we imply that all metric functions depend functionally on a single harmonic function $\sigma = e^{-U}$, where $\sigma_{uv}=0$, or in the case of the metric (50), σ satisfies $\sigma_{\rho\rho}+(1/\rho)\sigma_\rho-\sigma_{tt}=0$. Let us note that although the choice of harmonic variables is not an imperative one, the structure of Einstein equations suggests that such a choice facilitates the formalism to a great extent.[20]

The self-similar vacuum Einstein equations are obtained from the harmonic map between the manifolds

$$M:\ ds^2 = d\sigma^2, \tag{52}$$

$$M':\ ds'^2 = dW^2+\cosh^2 W\,dV^2. \tag{53}$$

The metric function M, which does not appear in the map, turns out to satisfy a quadrature equation that, as a requirement of complete integrability, must admit a solution. The self-similar Lagrangian and equations are given in the following:

$$L = W'^2+\cosh W\cdot V'^2, \tag{54}$$

$$V'\cosh^2 W = a_0 = \text{const}, \tag{55}$$

$$W'' = a_0^2(\sinh W)/(\cosh^3 W) \tag{56}$$

$$\left(' \equiv \frac{d}{d\sigma}\right).$$

Solutions for V and W take the form

$$e^{2V} = \frac{b_0+a_0\tanh b_0\sigma}{b_0-a_0\tanh b_0\sigma}, \tag{57}$$

$$\sinh W = \left[1-\left(\frac{a_0}{b_0}\right)^2\right]^{1/2}\sinh b_0\sigma \tag{58}$$

$$(b_0=\text{const}).$$

However, it can be observed by the coordinate transformation

$$x = \bar{x}\cos(\alpha/2)(+\bar{y}\sin(\alpha/2),\qquad y=\bar{x}\sin(\alpha/2)+\bar{y}\cos(\alpha/2) \tag{59}$$

that the metric function W can be set to zero. The choice of α that accomplishes this task is

$$\alpha = \tan^{-1}[(b_0/a_0)^2-1]. \tag{60}$$

(2) As the second point we would like to check whether the space-time metric with two spacelike Killing vectors admits a Taub–NUT-like solution. To this end we consider the Ernst equation in the coordinates[21,22]

$$\tau = u(1-v^2)^{1/2}+v(1-u^2)^{1/2},\qquad \sigma = u(1-v^2)^{1/2}-v(1-u^2)^{1/2}. \tag{61}$$

The simplest Ernst potential $\xi=\tau$ turns out to be the Khan–Penrose (KP)[23] solution for the CGW. From the experience of SAS space-times one obtains, by taking $\xi=e^{i\alpha}\tau$ ($\alpha=\text{const}$), the Taub–NUT solution. If the same replacement is carried out here, for the space-time with two spacelike Killing vectors, the resulting solution turns out to be diagonalizable. Thus the Taub–NUT type solution does not exist for the metric under investigation. For the cylindrically symmetrical line element the same proof can be done by employing the similar type of coordinates to (61),

$$2\tau = [(1+t)^2-\rho^2]^{1/2}+[(1-t)^2-\rho^2]^{1/2},\qquad 2\sigma = [(1+t)^2-\rho^2]^{1/2}-[(1-t)^2-\rho^2]^{1/2}. \tag{62}$$

As a matter of fact, a more general result can be proved in this line: whenever the real and the imaginary parts of the Ernst potential are functionally related (i.e., one can be expressed in terms of the other) then the metric reduces to a diagonal one.

Finally, we explore the possible self-similar cosmological vacuum model in the presence of two commuting Killing vectors. In the metric above we take $W=0$, and express the remaining metric functions as functions of a common harmonic function. Since the proof is rather simple, we shall just content ourselves by stating the result that such a self-similar cosmology happens to be the Kasner[24] cosmology. Any other form of solution must be transformable into Kasner solution by a coordinate transformation.

VI. KILLING VECTORS OF THE M MANIFOLD

Obviously the metric used by Bonnor [Eq. (6)] can be transformed into the plane wave space-time by the following identifications:

$$2^{1/2}u = t-\rho,\quad M=2(\Psi-\gamma),\quad V=2\Psi-\ln\rho,\qquad 2^{1/2}v = t+\rho,\quad e^{-U}=\rho,\quad z\to x,\quad \phi\to y, \tag{63}$$

so that the resulting space-time metric reads

$$ds^2 = 2e^{-M}\,du\,dv - e^{-U}(e^{V}dx^2 + e^{-V}dy^2). \tag{64}$$

This is the particular case of the Szekeres metric (51) when the cross polarization term is suppressed. The vacuum Einstein equations for this more general line element are obtained from the harmonic maps[21,22] between the manifolds

$$ds^2 = 2\,du\,dv, \tag{65}$$

$$ds' = e^{-U}[2\,dU\,dM + dU^2 - dW^2 - dV^2\cosh^2 W]. \tag{66}$$

From the metric functions, U is chosen as a coordinate condition and the determination of M is reduced to quadratures. By setting $W=0$, first, the metric of M' takes the form

$$ds'^2 = e^{-U}[2\,dU\,dM + dU^2 - dV^2]. \tag{67}$$

The problem now is to determine the nontrivial Killing vectors of this line element which will aid in generating a new solution from the old one. It can be verified that this geometry admits a nontrivial Killing vector

$$\xi = \xi^A\frac{\partial}{\partial X^A} = V\frac{\partial}{\partial M} + U\frac{\partial}{\partial V}. \tag{68}$$

We shall proceed now to obtain the new solution $(\tilde{U},\tilde{M},\tilde{V})$ generated from a known solution (U,M,V) by the isometry of this Killing vector. The isometry equation is given by

$$\dot{X}_A = \alpha\xi X_A, \tag{69}$$

where α is a new parameter. Upon substituting ξ one obtains

$$\dot{U}=0,\quad \dot{M}=\alpha V,\quad \dot{V}=\alpha U. \tag{70}$$

Imposing now the initial ($t=0$) and the image ($t=1$) conditions of the isometry, the new solution is expressed by

$$\begin{aligned}&\tilde{U}=U,\quad \tilde{V}=V+\alpha U,\\ &\tilde{M}=M+\alpha V+\tfrac{1}{2}\alpha^2 U.\end{aligned} \tag{71}$$

Choosing as (U,M,V) the nonsingular solution of Bonnor, by this isometry we obtain a new solution with an additional parameter. The same isometry has been employed elsewhere to generate new scalar plane waves.[25]

The foregoing method of isometries can equivalently be handled in the Ernst formalism. Defining the complex potential by

$$\eta = \frac{\sinh V\cosh W - i\sinh W}{\cosh V\cosh W + 1}, \tag{72}$$

the following equality holds true:

$$\frac{4\,d\eta\,d\bar{\eta}}{(1-\eta\bar{\eta})^2} = dW^2 + \cosh^2 W\cdot dV^2. \tag{73}$$

The left-hand side of this equality coincides exactly with the M' manifold of the Ernst Lagrangian. Thus any isometry of the rhs corresponds to an isometry of the lhs and vice versa. For instance, the isometry

$$\eta\to\eta' = [1+\eta(i\beta-1)]/[1+\eta(i\beta+1)] \tag{74}$$

with the real parameter β, which is known as the Ehlers[26] transformation, can directly be adopted in the generation of a new cosmological model. However, our line of search will follow an alternative route, rather than employing well-known results. Once a pair (V,W) of solutions is known we shall proceed to generate a new pair $(\tilde{V},\tilde{W})$ by employing the isometry

$$d\tilde{W}^2 + d\tilde{V}^2\cosh^2\tilde{W} = dW^2 + dV^2\cosh W. \tag{75}$$

This can be achieved by making use of the general Killing vector, namely,

$$\begin{aligned}\xi = (c_1e^{V} + c_2e^{-V})\frac{\partial}{\partial W}\\ + [c_3 - (c_1e^{V} - c_2e^{-V})\tanh W]\frac{\partial}{\partial V},\end{aligned} \tag{76}$$

where c_1, c_2, and c_3 are arbitrary constants. As an example, we shall obtain the new solution corresponding to the linear combination of the two Killing vectors

$$\xi_{(1)} = e^{-V}\left(\frac{\partial}{\partial W} + \tanh W\frac{\partial}{\partial V}\right), \tag{77}$$

$$\xi_{(2)} = e^{V}\left(\frac{\partial}{\partial W} - \tanh W\frac{\partial}{\partial V}\right), \tag{78}$$

in accordance with the relation

$$\dot{X}_A = (\alpha_0\xi_{(1)} + \beta_0\xi_{(2)})X_A, \tag{79}$$

where α_0 and β_0 are constants. We obtain equivalently the pair

$$\dot{V} = \tanh W(\alpha_0e^{-V} - \beta_0e^{V}), \tag{80}$$

$$\dot{W} = \alpha_0e^{-V} + \beta_0e^{V}. \tag{81}$$

After tedious calculations one obtains the new solution

$$\begin{aligned}\sinh\tilde{W} &= \cosh\alpha\sinh W\\ &\quad + \tfrac{1}{2}\sinh\alpha\cosh W(\beta e^{V} + \beta^{-1}e^{-V}),\end{aligned} \tag{82}$$

$$\begin{aligned}\beta\cosh\tilde{W}e^{\tilde{V}} &= \sinh\alpha\sinh\tilde{W}\\ &\quad + \tfrac{1}{2}\cosh\tilde{W}[\beta e^{V}(\cosh\alpha+1)\\ &\quad \times\beta^{-1}e^{-V}(\cosh\alpha-1)],\end{aligned} \tag{83}$$

where the new parameters α,β are defined by

$$\alpha = 2(\alpha_0\beta_0)^{1/2},\quad \beta = (\beta_0/\alpha_0)^{1/2}.$$

It is readily observed that in the limit $\alpha=0$ we recover the old solution (V,W), but otherwise we have a new solution $(\tilde{V},\tilde{W})$ generated from the isometry. The constant β emerges as a scale parameter for the functions e^{V} and $e^{\tilde{V}}$ and therefore it can be washed out from the solution.

At this stage we can also check whether a solution with $W\neq 0$ can be generated from a known solution with $W=0$.[27] For this purpose our isometry takes the form

$$\sinh\tilde{W} = \sinh\alpha\cosh V, \tag{84}$$

$$\cosh\tilde{W}e^{\tilde{V}} = \cosh\alpha\cosh V + \sinh V. \tag{85}$$

After some simple algebra it can be observed that the corresponding space-time metric diagonalizes under the hyperbolic rotation

$$\begin{aligned}x' &= x\cosh(\alpha/2) + y\sinh(\alpha/2),\\ y' &= x\sinh(\alpha/2) + y\cosh(\alpha/2),\end{aligned} \tag{86}$$

and as a result such a solution does not exist. The method of isometries fails to add a cross term to a diagonal metric but it maps a given solution into a new one.

VII. CONCLUSION

We have shown that space-times admitting two spacelike Killing vectors admit dual interpretations and to this end, the method of harmonic maps proves to be a useful technique. What seems a more important question however, is whether such dual properties of the vacuum fields have any physical significance beyond mathematics. Consider, for instance, a proton and a neutron in a given vacuum field that admits the e.m. field via dual interpretation. The apparent paradox between the geodesics equations of proton and neutron will be resolved provided their mass difference is attributed to an e.m. origin.

The method of isometries in the M' manifold provides a promising feature and a useful alternative to already existing methods in general relativity. As a matter of fact, the method of harmonic maps applies to any theory whose Lagrangian is expressed in pure kinetic form. Self-dual SU(N) field equations and instantons in classical field theory provide such examples, to mention a few. Further, in the instanton problem the base manifold is the four-dimensional Euclidean manifold with definite metric that can be mapped onto a sphere and the degree of harmonic maps results in a topologically significant number.

[1]H. Weyl, Ann. Phys. (Leipzig) **54**, 117 (1917).
[2]A. Einstein and N. Rosen, J. Franklin Inst. **223**, 43(1937).
[3]L. Marder, Nature **235**, 379 (1972).
[4]J. Weber and J. A. Wheeler, Rev. Mod. Phys. **29**, 509 (1957).
[5]W. B. Bonnor, J. Math. Mech. **6**, 203 (1957).
[6]F. J. Tipler, Phys. Rev. D **22**, 2929 (1980).
[7]M. Halilsoy, Phys. Lett. A **84**, 404 (1981).
[8]C. W. Misner, Phys. Rev. D **18**, 4510 (1978).
[9]J. Eells, Jr. and J. H. Sampson, Am. J. Math. **86**, 109 (1964).
[10]Y. Nutku, Ann. Inst. H. Poincaré A **21**, 175 (1974).
[11]F. J. Ernst, Phys. Rev. **167**, 1175 (1968).
[12]R. Matzner and C. W. Misner, Phys. Rev. **154**, 1229 (1967). See also, G. Neugebauer and D. Kramer, Ann. Phys. (Leipzig) **24**, 62 (1969).
[13]Y. Ne'eman, Rev. Mod. Phys. **37**, 227 (1965).
[14]M. Misra, Proc. Cambridge Philos. Soc. **58**, 711 (1962).
[15]E. Newman, A. Janis, and J. Winicour, Phys. Rev. Lett. **20**, 878 (1968).
[16]M. Hirayama, H. Chia Tze, J. Ishida, and T. Kawabe, Phys. Lett. A **66**, 352 (1978).
[17]J. Honerkamp, A. Patani, M. Schlindwein, and Q. Shafi, Lett. Nuovo Cimento **15**, 97 (1976).
[18]P. Jordan, J. Ehlers, and W. Kundt, Abh. Akad. Wiss. Mainz Math. Naturwiss. Kl. **1960**, 2.
[19]P. Szekeres, J. Math. Phys. **13**, 286 (1972).
[20]M. Halilsoy, Lett. Nuovo Cimento **37**, 231 (1983).
[21]Y. Nutku and M. Halil, Phys. Rev. Lett. **39**, 1379 (1977).
[22]M. Halil, J. Math. Phys. **20**, 120 (1979).
[23]K. A. Khan and R. Penrose, Nature **229**, 185 (1971).
[24]E. Kasner, Am. J. Math. **43**, 217 (1921).
[25]M. Halilsoy, Lett. Nuovo Cimento **44**, 544 (1985).
[26]J. Ehlers, in *Les theories relativistes de la gravitation* (CNRS, Paris, 1959).
[27]M. Halilsoy, Phys. Lett. A **84**, 359 (1981).

Classification of the space-times admitting a constant curvature synchronization

C. Bona
Departament de Fisica, Universitat Illes Balears, 07071 Palma de Mallorca, Spain

B. Coll
Chaire de Physique Mathématique, Collège de France, Paris, France

(Received 31 July 1986; accepted for publication 26 August 1987)

A constant curvature synchronization is a foliation of the space-time by spacelike hypersurfaces which are spaces of constant curvature for the induced metric. The classification of the space-times admitting such a synchronization is given according to the dimension and structure of all the isometry groups that can act with orbits tangent to it.

I. INTRODUCTION

Let F be a non-null foliation of the space-time. An F-Killing vector field is a (tangent to F) vector field of the space-time that is a Killing field for the metric induced on every leaf of F by the space-time metric. The F-Killing vector fields are not necessarily Killing fields for the space-time metric, but they form a finite Lie algebra over the ring Φ of the constant functions on F. They are the generators of the (connected) group of symmetries of the space-time called intrinsic symmetries by Collins,[1] who seems to have been the first to consider them. They have been studied with some detail in the works by Krasinski,[2] Collins and Szafron,[3] and the present authors.[4,5]

A foliation S whose leaves are spacelike hypersurfaces is called a synchronization. The dimension of the Lie algebra of S-Killing vector fields is $\leqslant 6$ and when it is 6, the leaves of S are Riemann spaces of constant curvature. The space-times admitting synchronizations with leaves of constant curvature are called, for short, CCS space-times. The CCS space-times may play an important role in cosmology, not only as weaker, intermediate versions of the cosmological principle,[5] but also as background spaces for anisotropic and/or inhomogeneous cosmologies[3]; they are also of interest in the study of compact matter distributions.[6]

The object of this paper is to classify the CCS space-times according to the (continuous connected) isometry groups attached to S, that is to say, according to the maximal subalgebras of the S-Killing fields which are also Killing fields of the space-time.

The classification is given in Sec. II in the form of three tables (corresponding to the sign of the scalar curvature of the leaves of S). They are the tables of the (conjugate classes of) subalgebras of the S-Killing algebra.

Some cases are common to the three tables: the existence of nontrivial isometry groups is compatible with the change of sign of the scalar curvature of the leaves of S. This fact, already pointed out by Krasinski[2] in a less general situation, may be of interest in cosmology.

Section III deals with the problem of constructing explicit models pertaining to a given class. One step in that direction is the obtaining of the scalar and vector invariants under every S-Killing subalgebra. Finally, it is noted that there are two particular cases in which the Killing algebra of a CCS space-time containing one given S-Killing subalgebra must contain the whole S-Killing algebra. This implies that these exceptional S-Killing subalgebras cannot be maximal.

II. CLASSIFICATION OF CCS SPACE-TIMES

Throughout this work, the tensor elements defined on the space-time V_4 are noted with a caret in order to distinguish them from those which will be defined on submanifolds. Also, we denote, in general, tensors and cotensors associated by the metric with the same letter and, whenever confusion is possible, tensors will be distinguished with an asterisk "*."

The space-time V_4 is endowed with a Lorentzian structure $(V_4, \hat{g})$, $\hat{g}$ being chosen with signature 2ϵ, $\epsilon = \pm 1$. In order to avoid difficulties related essentially with derived algebras, $(V_4, \hat{g})$ is supposed to be of differentiability class C^∞.

No hypothesis is made on the energy tensor T associated to $\hat{g}$, so that the Einstein equations do not carry restrictions on the generality of the metrics considered here; in fact, they are not used at all.

A. Metric form

In a domain Ω of the space-time $(V_4,\hat{g})$, *a synchronization* S is a foliation of Ω by spatial hypersurfaces called instants. Let ϵg denote the metric induced by $\hat{g}$ on every instant of S; a constant curvature synchronization is a synchronization S for which every instant Σ, considered as a three-dimensional Riemannian manifold (Σ,g), is of constant curvature:

$$\mathrm{Ric}(g) = 2\kappa g. \tag{1}$$

Space-times admitting a constant curvature synchronization, or CCS space-times, have been studied in Ref. 4.

A local chart in Ω is adapted to S if it is of the form $\{t,y^i\}$ $(i = 1,2,3)$, where $t = \text{const}$ is a local equation for S. It is clear that the scalar curvature κ in (1) is constant on every instant of S, so that we have $\kappa \equiv \kappa(t)$ in any adapted local chart. In such a chart, the metric $\hat{g}$ of a CCS space-time may be written[4]

$$\hat{g} = (\epsilon\alpha^2 + a\cdot a)dt \otimes dt + a \bar{\otimes} dt + \epsilon g, \tag{2}$$

the (positive definite) metric g obeying Eq. (1).[7]

 0022-2488/88/020327-05$02.50

According to the equivalence theorem for (simply connected) constant curvature space forms, there always exist in V_4 adapted local charts in which g in (2) takes the same (time-dependent) specified form. As far as explicit calculations are implied, we shall use in this paper the so-called almost-Cartesian local charts (t, x^i) or their related cylindrical and spherical ones, denoted, respectively, by (t,ρ,ϕ,z) and (t,r,θ,ϕ). They are defined in such a way that, in the spherical local charts, the metric g takes the specified form

$$g = R^2\{[1/(1-kr^2)]dr\otimes dr + r^2(d\theta\otimes d\theta + \sin^2\theta\, d\phi\otimes d\phi)\}, \tag{3}$$

where $k = 0, \pm 1$ and the function $R(t)$ is related to κ by $R^2\kappa = k$, R being arbitrary when $\kappa = 0$.

Let us remember that, under a local foliated isometry (an isometry preserving S), the metric form (2) changes to

$$\hat{g} = -[\epsilon\alpha^2 - (a+b)\cdot(a+b)]dt\otimes dt + (a+b)\tilde{\otimes}dt + \epsilon g, \tag{4}$$

where b is the one-form associated by g to the vector field b^* defining the spatial component of the derived diffeomorphism.[4]

B. S-Killing fields and isometry group

The Killing vector fields admitted by the three-dimensional metrics g on the instants of S are called S-Killing vector fields.[4] For CCS space-times, they form a six-dimensional Lie algebra, say A, over the ring Φ of the constant functions on the instants of S. In almost-Cartesian local charts, A is generated by the following set of S-Killing fields:

$$q_i \equiv (1-kr^2)^{1/2}\frac{\partial}{\partial x^i}, \quad j_i \equiv -\epsilon_{irs}x^r\frac{\partial}{\partial x^s}, \tag{5}$$

ϵ_{irs} being the Levi-Civita symbol and $i,r,s = 1,2,3$. The structure of A, which easily follows from (5), is given by

$$[j_i,j_j] = \epsilon_{ijr}j_r, \quad [j_i,q_j] = \epsilon_{ijr}q_r, \quad [q_i,q_j] = k\epsilon_{ijr}j_r, \tag{6}$$

so that it depends on the signature k of the curvature κ: A is isomorphic to SO(4), E(3), or SO(3,1), when k is equal to 1, 0, or -1, respectively.

The algebra A generates the identity component U of the corresponding isometry group G of g. The metric g being definite, G consists of two connected components, $G = C_2\cdot U$, where C_2 is the cyclic group generated by a parity transformation.

We are interested in the connected isometry groups of the CCS space-time metric $\hat{g}$ which leave invariant every instant of S. The corresponding Lie algebra of Killing fields is therefore tangent to S and then it is a subalgebra $\underline{A}$ of the S-Killing algebra A. The CCS space-times can then be classified according to the maximal subalgebra $\underline{A}$ of the S-Killing algebra A they admit as a (foliated) Killing algebra.

C. Subalgebras of the S-Killing algebra

Let us consider a subgroup $\underline{G}$ of G and denote by σ the realization of G defining its (effective) action on the instants of S. Let $\{t, x^i\}$ and $\{t,y^i\}$ be two local charts related by an element $\underline{a}$ of the subgroup $\underline{G}$, $y = \sigma(x,\underline{a}) \equiv \sigma_{\underline{a}}(x)$, and let $\{t,u^i\}$ and $\{t,v^i\}$ be, respectively, the transformed local charts by an arbitrary element p of the full group G, $u = \sigma_p(x)$, $v = \sigma_p(y)$; we have then $bp = pa$ for some b in such a way that when $a\in\underline{G}$ then $b\in\underline{G}'$ (the subgroup conjugate to $\underline{G}$ by $p\in G$). Considered as isometry transformation groups on S, two groups are equivalent iff they are conjugate under the full isometry group $G = C_2\cdot U$. Their Lie algebras, $\underline{A}$ and $\underline{A}'$, related by the adjoint action Ad of G on A, belong thus to the same conjugate class, say $(\underline{A})$.

As stated in the preceding subsection, the classification of the CCS space-times amounts to the classification of the subalgebras of A (see Ref. 8) and we have seen here that this classification must be performed according to the conjugate classes of A under G. We list now the results for the three algebras given by (6).

1. Lorentz algebra

The conjugate classes of the subalgebras of the Lorentz algebra $A\sim$SO(3,1) were obtained in a pioneering work by Finkelstein[9] and later by Bacry and Kihlberg.[10] Their results are given (with an adapted notation) in Table I. As usual in relativistic theories, the Bianchi notation is given for the three-dimensional classes.

2. Euclidean algebra

The conjugate classes of the subalgebras of the Euclidean algebra $A\sim$E(3) have been more recently studied by Beckers *et al.*[11] Their list is given in Table II.

3. Rotation algebra

In the case of the four-dimensional rotation algebra $A\sim$SO(4), the conjugate classes can be computed from its direct product structure, SO(4) $\sim$SO(3) $\otimes$ SO(3). Patera *et al.*[12] have recently given a standard method for computing the subalgebras of a direct product starting from the subalgebras of both factors, but in the present case a direct calculation is possible, owing to the simplicity of the subalgebras of SO(3). The result is given in Table III.

TABLE I. Conjugate classes of subalgebras of SO(3,1) under the isometry group $G\equiv$SO(3,1)$\circ C_2$. The number between brackets indicates the dimension.

Notation	Structure	Lie algebra generators
(6)	SO(3,1)	j_i, q_i $(i=1,2,3)$
(4)	SG(2)	j_1-q_2, j_2+q_1, j_3, q_3
$(3)_\infty$	*B* V	j_1-q_2, j_2+q_1, q_3
$(3)_\lambda$	*B* VII$_h$ $(h=\lambda^2)$	j_1-q_2, j_2+q_1, $j_3+\lambda q_3$ $(\lambda>0)$
$(3)_0$	*B* VII$_0$, E(2)	j_1-q_2, j_2+q_1, j_3
(3)	*B* IX, SO(3)	j_i $(i=1,2,3)$
(3)′	*B* VIII, SO(2,1)	q_1, q_2, j_3
(2)	Abelian	j_3, q_3
(2)′		j_1-q_2, j_2+q_1
(2)″	non-Abelian	j_1-q_2, q_3
$(1)_\infty$	trivial	q_3
$(1)_\lambda$		$j_3+\lambda q_3$ $(\lambda>0)$
$(1)_0$		j_3
(1)		j_1-q_2

TABLE II. Conjugate classes of subalgebras of E(3) under the isometry group $G \equiv E(3) \circ C_2$. The number between brackets indicates the dimension.

Notation	Structure	Lie algebra generators
(6)	E(3)	$j_i,\ q_i \quad (i=1,2,3)$
(4)	$E(2) \otimes T(1)$	$j_3,\ q_i \quad (i=1,2,3)$
$(3)_\infty$	B I, Abelian	$q_i \quad (i=1,2,3)$
$(3)_\lambda$	B VII_0, E(2)	$q_1,\ q_2,\ j_3+\lambda q_3 \quad (\lambda>0)$
$(3)_0$		$q_1,\ q_2,\ j_3$
(3)	B IX, SO(3)	$j_i \quad (i=1,2,3)$
(2)	Abelian	$j_3,\ q_3$
(2)′		$q_1,\ q_2$
$(1)_\infty$		q_3
$(1)_\lambda$	trivial	$j_3+\lambda q_3 \quad (\lambda>0)$
$(1)_0$		j_3

III. EXPLICIT CONSTRUCTION OF SPACE-TIME MODELS

A. Killing fields

In Ref. 4, we gave the necessary and sufficient conditions for an S-Killing field m^* to be a Killing field; in adapted local charts when the metric $\hat{g}$ admits the form (2), these conditions are

$$L(m^*)\alpha = 0, \tag{7}$$

$$\partial_t m^* = [a^*, m^*], \tag{8}$$

where a^* is the vector field associated by g to the one-form a in Eq. (2).

The scalar equation (7) reveals the fact that, if a subalgebra $\underline{A}$ is a Killing algebra, then α is a scalar invariant under $\underline{A}$. The vector equation (8) being linear in a^*, it follows that the vector field a^* is of the form

$$a^* = \underline{a}^* + p^*, \tag{9}$$

where $\underline{a}^*$ is an invariant vector field under $\underline{A}$:

$$[\underline{a}^*, m^*] = 0, \quad \forall m^* \in \underline{A}, \tag{10}$$

and p^* is any particular solution of the system

$$\partial_t m^* = [p^*, m^*], \quad \forall m^* \in \underline{A}. \tag{11}$$

TABLE III. Conjugate classes of subalgebras of SO(4) under the isometry group $G \equiv SO(4) \circ C_2$. The number between brackets indicates the dimension.

Notation	Structure	Lie algebra generators
(6)	SO(4)	$j_i,\ q_i \quad (i=1,2,3)$
(4)	$SO(3) \otimes SO(2)$	$j_1+q_1,\ j_2+q_2,\ j_3,\ q_3$
(3)′	B IX, SO(3)	$j_i+q_i \quad (i=1,2,3)$
(3)		$j_i \quad (i=1,2,3)$
(2)	Abelian	$j_3,\ q_3$
$(1)_\lambda$	trivial	$j_3+\lambda q_3 \quad (0<\lambda<1)$
$(1)_0$		j_3

TABLE IV. Scalar and vector invariants under the subalgebras of SO(3,1).

SO(3,1) subalgebra	Scalar invariants	Vector invariants
(6)	...	...
(4)	...	d^*l
$(3)_\infty$	...	$d^*l,\ e^{-l}(j_1-q_2),\ e^{-l}(j_2+q_1)$
$(3)_\lambda$	...	d^*l $e^{-l}(\sin(l/\lambda)(j_1-q_2)-\cos(l/\lambda)(j_2+q_1))$ $e^{-l}(\cos(l/\lambda)(j_1-q_2)+\sin(l/\lambda(j_2+q_1))$
$(3)_0$	l	d^*l
(3)	r	d^*r
(3)′	z	d^*z
(2)	ρ	$d^*\rho,\ j_3,\ q_3$
(2)′	l	$d^*l,\ j_1-q_2,\ j_2+q_1$
(2)″	x	$d^*x,\ d^*l,\ e^{-l}(j_1-q_2)$
$(1)_\infty$	$\rho,\ \phi$	
$(1)_\lambda$	$\rho,\ \lambda\phi+h$	$d^*\rho,\ j_3,\ q_3$
$(1)_0$	$\rho,\ z$	
(1)	$x,\ l$	$d^*x,\ d^*l,\ j_1-q_2$

In order to obtain explicit CCS space-time models pertaining to a given class ($\underline{A}$) in Tables I–III, one may use the (nontrivial) scalar and vector invariants $\{s_m, v_n\}$ under $\underline{A}$ to construct the metric $\hat{g}$ as given in (2) with

$$\alpha = \alpha(s_m, t), \quad a = p + \sum_n \gamma_n(s_m, t) v_n, \tag{12}$$

where α and γ are arbitrary functions of their arguments. The particular solution p^* of (11) can be obtained in terms of the time derivatives of the Killing fields m^* in every case.

B. Scalar and vector invariants

The evaluation of the scalar and vector invariants has been made alternatively in the almost-Cartesian, cylindrical, or spherical local charts defined in Sec. I, according to the nature of the subalgebra considered.

Tables IV, V, and VI give them for the subalgebras of SO(3,1), E(3), and SO(4), respectively. The dimension of the orbits is $d = 3 - n$, n being the number of independent (nontrivial) scalar invariants. The functions l, g, and h are given by

TABLE V. Scalar and vector invariants under the subalgebras of E(3).

E(3) subalgebra	Scalar invariants	Vector invariants
(6)	...	...
(4)	...	q_3
$(3)_\infty$	...	$q_i \quad (i=1,2,3)$
$(3)_\lambda$	...	$q_3,\ \sin(z/\lambda)q_1+\cos(z/\lambda)q_2,$ $\cos(z/\lambda)q_1-\sin(z/\lambda)q_2$
$(3)_0$	z	q_3
(3)	r	d^*r
(2)	ρ	$d^*\rho,\ j_3,\ q_3$
(2)′	z	$q_i \quad (i=1,2,3)$
$(1)_\infty$	$\rho,\ \phi$	
$(1)_\lambda$	$\rho,\ z+\lambda\phi$	$d^*\rho,\ j_3,\ q_3$
$(1)_0$	$\rho,\ z$	

TABLE VI. Scalar and vector invariants under the subalgebras of SO(4).

SO(4) subalgebra	Scalar invariants	Vector invariants
(6)	...	...
(4)	...	$j_3 - q_3$
(3)′	...	$j_i - q_i \quad (i = 1,2,3)$
(3)	r	d^*r
(2)	ρ	
$(1)_\lambda$	$\rho,\ \lambda\phi + g$	$d^*\rho,\ j_3,\ q_3$
$(1)_0$	$\rho,\ z$	

$$l \equiv \ln(\sqrt{1+r^2} - z), \quad g \equiv \arctan(z/\sqrt{1-r^2}), \quad h \equiv \operatorname{arctanh}(z/\sqrt{1+r^2}). \tag{13}$$

C. Exceptional cases

Let us give a closer look at the class (3)′ of conjugate subalgebras of SO(4) (Table III). The S-Killing generators are

$$j_i + q_i \quad (i = 1,2,3), \quad [j_i + q_i, j_j + q_j] = \epsilon_{ijk}(j_k + q_k). \tag{14}$$

Let us consider now a CCS space-time with positive curvature in a domain Ω and of class (3)′ everywhere in Ω. This implies that the algebra of Killing vectors tangent to S is generated by the following set of Killing vectors m:

$$m_r = C_{ri}(t)(j_i + q_i) \quad (r = 1,2,3), \quad [m_r, m_s] = \epsilon_{rst} m_t. \tag{15}$$

Comparing Eqs. (14) and (15), we get the following algebraic restrictions on the functions $C_{ri}(t)$:

$$\epsilon_{rsu} C_{uk} = \epsilon_{ijk} C_{ir} C_{js}. \tag{16}$$

It follows that the matrix of coefficients C_{ij} must be orthogonal, that is,

$$C \cdot C^T = C^T \cdot C = I, \tag{17}$$

where C^T stands for the transpose of the matrix C. Taking the time derivative of (17), one gets that the matrix $\omega(t)$, defined as

$$\omega = C^T \cdot (\partial_t C), \tag{18}$$

must be antisymmetric.

It is easy to check that the vector p, constructed as

$$p = \sum_i f_i(t)\,(j_i + q_i), \quad f_i(t) \equiv -\tfrac{1}{2}\epsilon_{irs}\omega_{rs}, \tag{19}$$

is a solution of the system (11) with m_r given by (15). The most general metric form for such a CCS space-time can be obtained from Eqs. (12) and (19) and the scalar and vector invariants for the case (3)′ in Table III:

$$\hat{g} = -[\epsilon\alpha^2 + a \cdot a]dt \otimes dt + a \tilde{\otimes} dt + \epsilon g, \quad \alpha = \alpha(t), \quad a = \sum_n [\gamma_n(t)(j_n - q_n) + f_n(t)(j_n + q_n)]. \tag{20}$$

Note that the vector a in (20) is an S-Killing vector field. One can then transform the metric (20) by a local foliated isometry with an associated vector field $b = -a$ so that, allowing for Eq. (4), the transformed metric is

$$\hat{g} = -\epsilon\alpha^2(t)dt \otimes dt + \epsilon g, \tag{21}$$

with g given by Eq. (3). The metric (21) is of class (6) (Friedmann–Robertson–Walker space-time).

It is then clear that any CCS space-time of positive curvature admitting a subalgebra $\underline{A}$ of class (3)′ as a Killing algebra must admit the whole S-Killing algebra A. It follows that the subalgebras contained in classes (3)′ or (4) in Table III cannot be maximal. These are two exceptional cases, as it is possible to construct CCS space-time models with a maximal S-Killing subalgebra pertaining to any of the remaining classes in Tables I–III.

D. Comments

As pointed out at the beginning of this section, the foliated isometry classes of CCS space-times that we have obtained in Sec. II correspond, in principle, to connected Lie subgroups of $G = C_2 \cdot U$, that is to say, to Lie subgroups of U. Nevertheless, some classes correspond, in fact, to total (disconnected) subgroups of G; this is the case, in particular, for one-dimensional subgroups when the one-form associated by $\hat{g}$ to the generating Killing field is integrable: if $v \equiv \exp\{tv^*\}$ acts isometrically and $v \wedge dv = 0$, then $C_2 \cdot \{v\}$ acts also isometrically.

A finer classification may be obtained by dividing every one of our foliated isometry classes according to the isometry groups acting transversally to S. One of the basic elements of this problem (the table of the subalgebras of the Poincaré group) is known[12]; nevertheless, such a classification depends essentially on the concrete time dependence of the quantities appearing in (12) and, for this reason, we have avoided it here. A complete isometry classification of this kind has been given in Ref. 5 for the particular case of CCS space-times for which the first and the second fundamental forms of the instants of S admit the same isometry group (Stephani universes).

Finally, let us remark that the presence of nontrivial isometries on a CCS space-time may be compatible with changes of sign of the curvature $\kappa(t)$. Despite the fact that the S-Killing algebra A changes, the three algebras SO(3,1), E(3), and SO(4) have common subalgebras (see Tables I–III) which may act isometrically on the space-time. This is clearly the case for the SO(3) subalgebra, as was already pointed out by Krasinski.[2]

This implies that our classification scheme is a local one. Its extension to a global classification may require some supplementary conditions arising from the Einstein field equations and the physical interpretation of the stress-energy tensor, as is the case with the global approach made in Ref. 2.

ACKNOWLEDGMENT

The authors acknowledge financial support by the Spanish Comision Asesora para la Investigacion Cientifica y Tecnica (CAICYT) under Project No. 1005/84.

[1] C. B. Collins, Gen Relativ. Gravit. **10**, 925 (1979).
[2] A. Krasinski, Gen Relativ. Gravit. **13**, 1021 (1981).
[3] C. B. Collins and D. A. Szafron, J. Math. Phys. **20**, 2347, 2354, 2362 (1979).

[4]C. Bona and B. Coll, J. Math. Phys. **26**, 1583 (1985).
[5]C. Bona and B. Coll, Gen. Relativ. Gravit. (to be published).
[6]C. Bona, J. Stela, and P. Palou, Gen Relativ. Gravit. **19**, 179 (1987).
[7] $\tilde{\otimes}$ denotes the symmetrization of the tensorial product and $a \cdot a$ stands for $g(a,a)$. The lapse α and the shift a are, respectively, scalar and vector fields on every instant of S.
[8]This is exact up to the exceptional cases in which a given subalgebra $\mathcal{A}$ cannot be maximal for CCS space-times. These cases are discussed at the end of the paper.
[9]D. Finkelstein, Phys. Rev. D **100**, 924 (1955).
[10]H. Bacry and A. Kihlberg, J. Math. Phys. **10**, 2132 (1969).
[11]J. Beckers, J. Patera, M. Perroud, and P. Winternitz, J. Math Phys. **18**, 72 (1977).
[12]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **16**, 1597, 1615 (1975).

Asymptotic decay of the free field covariance on a Riemannian manifold via heat kernel techniques

Gian Fabrizio De Angelis and Diego de Falco
Dipartimento di Fisica, Università di Salerno, I-84100 Salerno, Italy

Glauco Di Genova
I.S.A.S., International School for Advanced Studies, Strada Costiera, 11, I-34014 Trieste, Italy

(Received 3 February 1987; accepted for publication 16 September 1987)

It is demonstrated, via recent global estimates for the heat kernel found by Li and Yau [Acta Matematica **156**, 153 (1986)], that the covariance of the mass m free field decays exponentially, at a rate m, with the geodesic distance on any complete, noncompact Riemannian manifold with non-negative Ricci curvature. The rate of the exponential decay is larger than m on simply connected manifolds with negative sectional curvature.

I. INTRODUCTION

We are interested in generalizing to Riemannian manifolds the standard Euclidean field theory.[1] While some encouraging results have been obtained in this direction,[2] we believe in the necessity of deeper understanding of the Riemannian counterparts of well-established facts in the conventional (flat Euclidean) theory in order to provide a firm basis for a constructive approach to quantum field theory on a curved space-time.

Here we study the long distance behavior of the covariance

$$C = (-\Delta_M + m^2)^{-1} \tag{1}$$

corresponding to the mass m free field on a complete noncompact Riemannian manifold M, where Δ_M is the Laplace–Beltrami operator on M.

In the flat case ($M = \mathbb{R}^\nu$) the covariance (1) has an integral kernel $C(x,y)$ satisfying

$$\lim_{|x-y|\to\infty} -\frac{1}{|x-y|}\ln C(x,y) = m, \tag{2}$$

i.e., $C(x,y)$ decays exponentially at a rate m. Equation (2) identifies the inverse correlation length of the field with the mass parameter m appearing in (1).

On a curved Riemannian manifold the situation may be different. For instance, if M is the hyperbolic three-dimensional space H_3 with constant curvature $-1/R^2$, then

$$C(x,y) = \frac{\exp(-(m^2 + R^{-2})^{1/2} d(x,y))}{4\pi R \sinh(d(x,y)/R)}, \tag{3}$$

where $d(x,y)$ is the geodesic distance between x and y. This covariance kernel, when the analog of the limit (2) is performed, yields

$$\lim_{d(x,y)\to\infty} -\frac{1}{d(x,y)}\ln C(x,y) = (m^2 + R^{-2})^{1/2} + R^{-1} > m. \tag{4}$$

Equation (4) shows that in general there will be no good *a priori* reason for observing an inverse correlation length equal to the "mass parameter" m.

However, we demonstrate that on a complete, noncompact, smooth Riemannian manifold M with non-negative Ricci curvature, the covariance decays exponentially at a rate m, exactly as in the flat Euclidean case. For instance, such a result holds for the "imaginary time" counterpart of the Kruskal space-time.

On the other hand, when M is a simply connected manifold with strictly negative sectional curvature we demonstrate that the covariance kernel decays exponentially at a rate strictly larger than m.

Our strategy is based on the heat kernel representation for the covariance kernel[3]

$$C(x,y) = \int_0^\infty e^{-m^2 t} p(t,x,y)\,dt, \tag{5}$$

where, of course, $p(t,x,y)$ is the heat kernel on M. Thus our problem concerns decay estimates for the Laplace transform of the heat kernel. The long distance behavior of the covariance is then related by Theorem 1 to the asymptotic behavior of the heat kernel for large t and $d(x,y)$. We point out that this kind of asymptotic control on the heat kernel is not deducible from the usual Molchanov small time asymptotics.[4]

When the Ricci curvature is non-negative, the necessary information about the heat kernel can be obtained from recent global estimates found by Li and Yau,[5] which in this case acquire a particularly simple form. Instead, when the sectional curvature is strictly negative and the manifold is simply connected, we use a monotonicity property of the heat kernel with respect to the curvature and an exact representation, found by Elworthy and Truman,[6] to derive a simple estimate for the decay rate. In this way we control the asymptotic behavior of the heat kernel and then the long distance decay of the covariance.

II. THE ASYMPTOTIC BEHAVIOR OF THE HEAT KERNEL DETERMINES THE LONG DISTANCE DECAY OF THE COVARIANCE

Let M be a complete noncompact smooth Riemannian manifold, $d(x,y)$ the geodesic distance between x and y, $p(t,x,y)$ the heat kernel on M, and $s(\epsilon|x,y) = [1/d(x,y)] \times \ln p(\epsilon d(x,y),x,y)$ for $\epsilon > 0$.

Theorem 1: Suppose the following.

(i) For every $\epsilon > 0$ there exists

 0022-2488/88/020332-04$02.50

$$\lim_{d(x,y)\to\infty} s(\epsilon|x,y) = s(\epsilon),$$

uniformly on any compact interval $[a,b] \subset (0,\infty)$.

(ii) There exist $R^* > 0$, $K^* > 0$, $0 < a^* < b^*$, and $\delta^* > 0$ such that $d(x,y) > R^*$ implies both

$$e^{d(x,y)s(\epsilon|x,y)} \leqslant K^*$$

for any $\epsilon > b^*$ and $\theta\epsilon - s(\epsilon|x,y) \geqslant f(\theta) + \delta^*$ for any $0 < \epsilon < a^*$, $\theta > 0$, where

$$f(\theta) = \inf_{\epsilon>0} (\theta\epsilon - s(\epsilon)).$$

Then

$$\lim_{d(x,y)\to\infty} \frac{1}{d(x,y)} \ln \int_0^\infty e^{-\theta t} p(t,x,y)\,dt = -f(\theta). \quad (6)$$

Proof: First we note that, by changing variables,

$$\lim_{d(x,y)\to\infty} \frac{1}{d(x,y)} \ln \int_0^\infty e^{-\theta t} p(t,x,y)\,dt$$
$$= \lim_{d(x,y)\to\infty} \frac{1}{d(x,y)} \ln \int_0^\infty \exp(-d(x,y) \times (\theta\epsilon - s(\epsilon|x,y)))\,d\epsilon,$$

so that the proof rests on the fact that the integrand $\exp(-d(x,y)(\theta\epsilon - s(\epsilon|x,y)))$ develops a very high peak in the large $d(x,y)$ limit (Laplace asymptotic method).

For convenience, we divide the proof into two steps.

Step (i):

$$\liminf_{d(x,y)\to\infty} \frac{1}{d(x,y)} \ln \int_0^\infty \exp(-d(x,y) \times (\theta\epsilon - s(\epsilon|x,y)))\,d\epsilon \geqslant -f(\theta). \quad (7)$$

In fact, for every $\delta > 0$ there exist both $\bar{\epsilon} > 0$ such that $\theta\bar{\epsilon} - s(\bar{\epsilon}) < f(\theta) + \delta/3$ and a compact interval $\Delta_\delta \ni \bar{\epsilon}$ such that $\theta\epsilon - s(\epsilon) \leqslant \theta\bar{\epsilon} - s(\bar{\epsilon}) + \delta/3 < f(\theta) + 2\delta/3$ for any $\epsilon \in \Delta_\delta$.

By using the uniform convergence of $s(\epsilon|x,y)$ to $s(\epsilon)$ on Δ_δ, there exists $r(\delta) > 0$ such that if $d(x,y) > r(\delta)$, then

$$\int_0^\infty \exp(-d(x,y)(\theta\epsilon - s(\epsilon|x,y)))\,d\epsilon \geqslant |\Delta_\delta| \exp(-d(x,y)(f(\theta) + \delta));$$

this inequality leads to (7).

Step (ii):

$$\limsup_{d(x,y)\to\infty} \frac{1}{d(x,y)} \ln \int_0^\infty \exp(-d(x,y) \times (\theta\epsilon - s(\epsilon|x,y)))\,d\epsilon \leqslant -f(\theta). \quad (8)$$

Let us choose $\delta = \delta^*$ and let $[a,b]$ be a compact interval containing Δ_{δ^*} with $a < a^*$ and $b > b^*$; moreover, we suppose that $d(x,y) > \max(R^*, r(\delta^*))$. On one hand we have

$$\int_a^b \exp(-d(x,y)(\theta\epsilon - s(\epsilon|x,y)))\,d\epsilon \geqslant |\Delta_{\delta^*}| \exp(-d(x,y)(f(\theta) + \delta^*));$$

on the other hand,

$$\int_{(0,a)\cup(b,\infty)} \exp(-d(x,y)(\theta\epsilon - s(\epsilon|x,y)))\,d\epsilon \leqslant a \exp(-d(x,y)(f(\theta) + \delta^*)) + \frac{K^*}{\theta d(x,y)} \exp(-\theta b d(x,y)).$$

Then, for any $\eta \in (0,1)$ there exists $r > 0$ such that

$$\int_0^\infty \exp(-d(x,y)(\theta\epsilon - s(\epsilon|x,y)))\,d\epsilon \leqslant (1+\eta) \int_a^b \exp(-d(x,y)(\theta\epsilon - s(\epsilon|x,y)))\,d\epsilon, \quad (9)$$

provided that $0 < a < a^*$, $b > [f(\theta) + \delta^*]/\theta$, and $d(x,y) > r$. Furthermore, since $s(\epsilon|x,y)$ is uniformly convergent to $s(\epsilon)$ on $[a,b]$, then for each fixed $\delta > 0$ there exists a suitable $r > 0$ such that

$$\int_a^b \exp(-d(x,y)(\theta\epsilon - s(\epsilon|x,y)))\,d\epsilon \leqslant \int_a^b \exp(-d(x,y)(\theta\epsilon - s(\epsilon) - \delta))\,d\epsilon \leqslant \int_a^b \exp(-d(x,y)(f(\theta) - \delta))\,d\epsilon = (b-a)\exp(-d(x,y)(f(\theta) - \delta))$$

for $d(x,y) > r$. By exploiting the last inequality and Eq. (9), Eq. (8) follows immediately.

The reader acquainted with modern statistical mechanics will easily recognize that Theorem 1 is essentially the statement of the thermodynamical equivalence between the canonical ensemble and the microcanonical ensemble.[7]

In the flat case ($M = \mathbf{R}^\nu$), where

$$p(t,x,y) = (4\pi t)^{-\nu/2} \exp(-d(x,y)^2/4t), \quad (10)$$

all the hypotheses of Theorem 1 can be easily verified and result in

$$s(\epsilon) = -1/4\epsilon, \quad f(\theta) = \theta^{1/2}. \quad (11)$$

On a general nonflat Riemannian manifold the task is not as easy, except for the few exceptional cases where the heat kernel is explicitly known. Nevertheless, all that is really needed are some good upper and lower estimates for the heat kernel exhibiting the same structure as in the flat case.

III. LI AND YAU BOUNDS AND ASYMPTOTIC DECAY OF THE COVARIANCE

Throughout this section we restrict ourselves to a complete, noncompact, smooth Riemannian manifold M with non-negative Ricci curvature. We denote by ν the dimension of M. For such a manifold the Li and Yau bounds[5] acquire the following forms.

(i) Upper bound (Ref. 5, p. 175). For each $\eta > 0$ there exists $C(\eta) > 0$ such that

$$p(t,x,y) \leqslant C(\eta) V_x(t^{1/2})^{-1} \exp(-d(x,y)^2/(4+\eta)t), \quad (12)$$

for any t,x,y; here $V_x(r)$ is the Riemannian volume of the geodesic ball centered at $x \in M$ with radius r.

(ii) Lower bound (Ref. 5, p. 182). For each $\eta > 0$ there exists $C(\eta) > 0$ such that

$$p(t,x,y)\geqslant C(\eta)^{-1}V_x(t^{1/2})^{-1}\exp(-d(x,y)^2/(4-\eta)t), \tag{13}$$

for any t,x,y.

In Eqs. (12) and (13) $C(\eta)\to\infty$ as $\eta\to 0$. In the following we will denote by $V_*(r)$ the volume of the Euclidean ball of radius r in $\mathbb{R}^\nu$.

Lemma 1: If $0<R_1<R_2$ then

$$V_x(R_2)/V_x(R_1)\leqslant V_*(R_2)/V_*(R_1).$$

Proof: See Ref. 5, p. 177.

Lemma 2: For each $a>0$ and for each fixed $x\in M$ the result is

$$\lim_{d(x,y)\to\infty}\frac{1}{d(x,y)}\ln V_x(a^{1/2}d(x,y)^{1/2})=0.$$

Proof: Trivial by using Lemma 1.

Lemma 3: Hypothesis (i) of Theorem 1 is satisfied by the heat kernel on M; moreover, $s(\epsilon)=-1/4\epsilon$, as in the flat case $M=\mathbb{R}^\nu$.

Proof: By choosing a compact interval $[a,b]\subset(0,\infty)$ from the estimates (12) and (13) it follows that

$$\begin{aligned}-\frac{\ln C(\eta)}{d(x,y)}-\frac{\ln V_x(b^{1/2}d(x,y)^{1/2})}{d(x,y)}+\left(\frac{1}{4}-\frac{1}{4-\eta}\right)\frac{1}{a}\\ \leqslant s(\epsilon|x,y)+\frac{1}{4\epsilon}\leqslant\frac{\ln C(\eta)}{d(x,y)}-\frac{\ln V_x(a^{1/2}d(x,y)^{1/2})}{d(x,y)}\\ +\left(\frac{1}{4}-\frac{1}{4+\eta}\right)\frac{1}{a}\end{aligned}$$

for any $\epsilon\in[a,b]$, $\eta>0$, $x,y\in M$.

Now, for any fixed δ we can choose a suitable $\eta>0$ such that the inequalities

$$\left(\frac{1}{4}-\frac{1}{4+\eta}\right)\frac{1}{a}<\frac{\delta}{2},\quad -\left(\frac{1}{4}-\frac{1}{4-\eta}\right)\frac{1}{a}>-\frac{\delta}{2}$$

hold. Then we obtain, for $\epsilon\in[a,b]$,

$$\begin{aligned}-\frac{\ln C(\eta)}{d(x,y)}-\frac{\ln V_x(b^{1/2}d(x,y)^{1/2})}{d(x,y)}-\frac{\delta}{2}\\ <s(\epsilon|x,y)+\frac{1}{4\epsilon}\\ <\frac{\ln C(\eta)}{d(x,y)}-\frac{\ln V_x(a^{1/2}d(x,y)^{1/2})}{d(x,y)}+\frac{\delta}{2},\end{aligned}$$

and the uniform convergence of $s(\epsilon|x,y)$ to $1/4\epsilon$ on the compact interval $[a,b]$ follows by Lemma 2.

Lemma 4: Hypothesis (ii) of Theorem 1 is satisfied by the heat kernel on M.

Proof: Step (i): For any fixed $K^*>0$ there exists an $R^*>0$ such that $d(x,y)>R^*$ implies $p(t,x,y)\leqslant K^*$ for every $t>0$. In fact, for any fixed $\eta>0$ there exists $C(\eta)>0$ such that

$$p(t,x,y)\leqslant C(\eta)V_x(t^{1/2})^{-1}\exp(-d(x,y)^2/(4+\eta)t).$$

Then, for $t\geqslant R^2$ the result is

$$p(t,x,y)\leqslant C(\eta)V_x(R)^{-1}$$

and step (i) follows from the choice of a large enough R^2; on the other hand, when $0<t<R^2$, we can use Lemma 1 to obtain

$$V_x(t^{1/2})^{-1}\leqslant V_*(t^{1/2})^{-1}V_x(R)^{-1}V_*(R)\leqslant\text{const}\,t^{-\nu/2}.$$

In turn, this inequality and the relation

$$\sup_{t>0}t^{-\nu/2}\exp(-d(x,y)^2/(4+\eta)t)=\text{const}\,d(x,y)^{-\nu}$$

imply step (i) by the choice of $d(x,y)$ large enough.

Step (ii): For each $\delta^*>0$ there exist $a^*>0, R^*>0$ such that $d(x,y)>R^*$ implies $\theta\epsilon-s(\epsilon|x,y)>\theta^{1/2}+\delta^*$ for any $\epsilon\in(0,a^*)$. In fact, for any fixed $\eta>0$ there exists $C(\eta)>0$ such that

$$\begin{aligned}&\theta\epsilon-s(\epsilon|x,y)-\theta^{1/2}\\ &\geqslant\theta\epsilon+\frac{1}{(4+\eta)\epsilon}-\theta^{1/2}\\ &\quad-\frac{\ln V_x(\epsilon^{1/2}d(x,y)^{1/2})}{d(x,y)}-\frac{\ln C(\eta)}{d(x,y)}.\end{aligned}$$

On the other hand, for $\epsilon\in(0,a^*)$ Lemma 1 implies

$$V_x(\epsilon^{1/2}d(x,y)^{1/2})\geqslant(\epsilon/a^*)^{\nu/2}V_x(a^{*1/2}d(x,y)^{1/2}).$$

Thus, by using the obvious inequality $\ln x\geqslant 1-x^{-1}$ for $x\in(0,1)$, we obtain

$$\begin{aligned}&-s(\epsilon|x,y)-\theta^{1/2}\\ &\geqslant\theta\epsilon+\frac{1}{(4+\eta)\epsilon}-\theta^{1/2}+\frac{\nu}{2d(x,y)}\left(1-\frac{a^*}{\epsilon}\right)\\ &\quad+\frac{\ln V_x(a^{*1/2}d(x,y)^{1/2})}{d(x,y)}-\frac{\ln C(\eta)}{d(x,y)}\\ &=\theta\epsilon+\frac{1}{2(4+\eta)\epsilon}-\theta^{1/2}\\ &\quad+\left(\frac{1}{2(4+\eta)}-\frac{\nu a^*}{2d(x,y)}\right)\frac{1}{\epsilon}+\frac{\nu}{2d(x,y)}\\ &\quad+\frac{\ln V_x(a^{*1/2}d(x,y)^{1/2})}{d(x,y)}-\frac{\ln C(\eta)}{d(x,y)}.\end{aligned}$$

Now, by choosing a suitable a^* and R^* the result is

$$\theta\epsilon-1/2(4+\eta)\epsilon-\theta^{1/2}>2\delta^*,\quad\text{for any }\epsilon\in(0,a^*),$$

$$1/2(4+\eta)-a^*/2d(x,y)\geqslant 0,\quad\text{for any } d(x,y)>R^*.$$

Then, by using Lemma 1 we obtain

$$\frac{\nu}{2d(x,y)}+\frac{\ln V_x(a^{*1/2}d(x,y)^{1/2})}{d(x,y)}-\frac{\ln C(\eta)}{d(x,y)}>-\delta^*$$

for large enough $d(x,y)$ and step (ii) follows.

Since all the hypotheses of Theorem 1 are satisfied, we obtain the following.

Theorem 2: Let M be a complete, noncompact, smooth Riemannian manifold with non-negative Ricci curvature. Then, for each fixed $x\in M$,

$$\lim_{d(x,y)\to\infty}\frac{1}{d(x,y)}\ln\int_0^\infty e^{-m^2t}p(t,x,y)dt=-m. \tag{14}$$

IV. LOWER BOUND ON THE DECAY RATE OF THE COVARIANCE ON NEGATIVELY CURVED MANIFOLDS

For ν-dimensional Riemannian manifolds whose curvatures are everywhere negative, Theorem 2 fails in general. Indeed, let us consider the case when M is a ν-dimensional simply connected manifold whose sectional curvature is bounded from above by $-R^{-2}$. Then in this case the result is[8]

$$p(t,x,y) \leqslant p_R(t,x,y),$$

where $p_R(t,x,y)$ is the heat kernel on the constant curvature simply connected space form H_ν whose curvature is identically $-R^{-2}$.

When the curvature is everywhere negative, the heat kernel admits the following representation[6]:

$$p(t,x,y) = (4\pi t)^{-\nu/2}\Xi_y(x)^{-1/2}\exp(-d(x,y)^2/(4t)) \times \mathrm{E}\left(\exp\int_0^t W(y_s)ds\right),$$

where $\Xi_y(x)$ is the Ruse invariant, $W(x) = \Xi_y(x)^{1/2}\Delta_M\Xi_y(x)^{-1/2}$, and E stands for the expectation with respect to the Brownian bridge y_s between x and y $(0 \leqslant s \leqslant t)$.

When we consider the ν-dimensional space H_ν with constant curvature R^{-2} the result is

$$\Xi_y(x) = (R\,d(x,y)^{-1}\sinh(R^{-1}d(x,y)))^{\nu-1}$$

and

$$W(x) = -(\nu-1)^2/4R^2 + [(\nu-1)(\nu-3)/4] \times (d(x,y)^{-2} - R^{-2}\sinh(R^{-1}d(x,y))^{-2}) \geqslant -(\nu-1)^2/4R^2.$$

This explicit expression for $W(x)$ allows us to derive the following estimates:

$$W(x) \geqslant -(\nu-1)^2/(4R^2) \quad (\nu>2),$$
$$W(x) \geqslant -(1+N)/(4R^2) \quad (\nu=2),$$

where

$$N = \sup[x^{-2} - (\sinh x)^{-2}] > 0.$$

Then, by using the estimate for $\nu > 2$, we obtain

$$p(t,x,y) \leqslant p_R(t,x,y) \leqslant (4\pi t)^{-\nu/2}\exp(-d(x,y)^2/4t)(R\,d(x,y)^{-1} \times \sinh(R^{-1}d(x,y)))^{(1-\nu)/2} \times \exp(-t(\nu-1)^2/4R^2). \tag{15}$$

This implies

$$s(\epsilon) \leqslant -\frac{1}{4\epsilon} - \frac{(\nu-1)^2}{4R^2} + \frac{1-\nu}{2R};$$

consequently, in Eq. (6) we obtain the result

$$f(\theta) = \inf_{\epsilon>0}\{\theta\epsilon - s(\epsilon)\} \geqslant (\theta + (\nu-1)^2/4R^2)^{1/2} + (\nu-1)/2R > \theta^{1/2}.$$

By using the bound (15) it is possible to estimate the value of $s(\epsilon)$ appearing in Theorem 1. Thus we have proved the following statement concerning the *bound on the decay rate*: Let M be a ν-dimensional $(\nu>2)$ simply connected, noncompact, smooth Riemannian manifold with sectional curvature bounded from above by the negative constant $-R^2$. Then

$$\limsup_{d(x,y)\to\infty}\frac{1}{d(x,y)}\ln\int_0^\infty e^{-m^2t}p(t,x,y)dt \leqslant -\left(m^2 + \frac{(\nu-1)^2}{4R^2}\right)^{1/2} - \frac{\nu-1}{2R} < -m,$$

i.e., the covariance on M decays exponentially at a rate strictly larger than m.

A similar result holds for $\nu = 2$ through the use of the corresponding estimate for $W(x)$.

[1] J. Glimm and A. Jaffe, *Quantum Physics. A Functional Integral Point of View* (Springer, Berlin, 1981); F. Guerra, L. Rosen, and B. Simon, "The $P(\varphi)_2$ Euclidean quantum field theory as classical statistical mechanics," Ann. Math. **101**, 111 (1975).

[2] G. F. De Angelis, D. de Falco, and G. Di Genova, "Random fields on Riemannian manifolds: A constructive approach," Commun. Math. Phys. **103**, 297 (1986); G. F. De Angelis, D. de Falco, and G. Di Genova, "The 'Euclidean' approach to quantum fields in nonflat space-time," in *Quantum Field Theory*, edited by F. Mancini (North-Holland, Amsterdam, 1986).

[3] R. M. Wald, "On the Euclidean approach to quantum field theory in curved space-time," Commun. Math. Phys. **70**, 221 (1979).

[4] S. A. Molchanov, "Diffusion processes and Riemannian geometry," Russ. Math. Survey **30**, 1 (1975); C. Bellaiche, "Le comportement asymptotique de $p(t,x,y)$ quand $t \to 0$," Soc. Math. France Astérisque **84–85**, 151 (1981).

[5] P. Li and S. T. Yau, "On the parabolic kernel of the Schrödinger operator," Acta Matematica **156**, 153 (1986).

[6] K. D. Elworthy, *Stochastic Differential Equations on Manifold* (Cambridge U. P., Cambridge, 1982).

[7] D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969).

[8] A. Debiard, B. Gaveau, and E. Mazet, "Théorème de comparaison en géométrie," J. Res. Inst. Kyoto Univ. **12.2**, 391 (1976).

Infinite-dimensional homogeneous manifolds with translational symmetry and nonlinear partial differential equations

J. Szmigielski[a)]
Department of Physics and Astronomy, University of Georgia, Athens, Georgia 30602

(Received 7 November 1986; accepted for publication 7 October 1987)

A construction relating scale-invariant, nonlinear partial differential equations and the orbits of the group of translations on infinite-dimensional homogeneous manifolds is proposed. It presupposes that the homogeneous manifold M has translations and scale transformations among its automorphisms. This assumption is made to ensure that the set of orbits of the group of translations is invariant under scale transformations. What is important is that the proposed construction provides a way of inducing scale covariance of derived nonlinear equations for which the set of orbits can be identified with a set of solutions of these equations. Application to the derivation of the potential Korteweg–de Vries (KdV) and the sine–Gordon equation yields as an important intermediate result the construction of their respective Lax pairs out of the commuting vectors tangent to the orbits. In yet another application, an infinite-dimensional, scale-invariant Riccati equation is derived. The latter is known to be related to the potential Kadomtsev–Petviashvilli (KP) equation. The orbit leading to the Riccati equation is then computed in different charts. The transition functions between charts are shown to generate the nonlinear term in the potential KP equation. Also, the relation to the Zakharov–Shabat dressing method is briefly discussed.

I. INTRODUCTION

The inverse scattering method as applied to solving nonlinear differential equations can be included among the important techniques of mathematical physics. One of its essential features is that the equation under study is rewritten as the compatibility condition for a system of partial linear differential equations that usually depends upon an additional *spectral parameter*.[1] The role of this additional parameter is particularly clear in the Zakharov–Shabat *dressing* method,[2] a generalization of the original formulation of the inverse scattering method.[3] One starts from an overdetermined system of differential equations written in matrix form as

$$i\frac{\partial}{\partial x}\Psi = U\Psi, \tag{1.1a}$$

$$i\frac{\partial}{\partial y}\Psi = V\Psi, \tag{1.1b}$$

where Ψ is assumed to be an invertible $n\times n$ matrix. The compatibility condition for (1.1a) and (1.1b) takes the form of the zero-curvature condition

$$\frac{\partial}{\partial x}V - \frac{\partial}{\partial y}U + i[U,V] = 0. \tag{1.1c}$$

This equation is nonlinear, but as long as U and V are not subject to any additional conditions, (1.1c) has as an immediate solution,

$$U = i\left(\frac{\partial}{\partial x}\Psi\right)\Psi^{-1}, \quad V = i\left(\frac{\partial}{\partial y}\Psi\right)\Psi^{-1}, \tag{1.2}$$

where Ψ can be any invertible function of x and y. In fact, even the stronger results holds; by Frobenius' theorem (1.1a) and (1.1b) are equivalent to (1.1c). However, solving (1.1c) becomes considerably more difficult, once U and V satisfy some additional conditions. If, for example, U and V have to satisfy the following conservation law:

$$\frac{\partial}{\partial x}U + \frac{\partial}{\partial y}V = 0, \tag{1.3}$$

then finding an appropriate Ψ in (1.2) requires more effort than before. What makes the Zakharov–Shabat method so attractive is the way in which the constraints, like (1.3), are incorporated. This is done, in their method, by introducing an extra complex variable λ and making U and V depend upon λ in a special way. For instance, in order to incorporate (1.3), it suffices to seek a solution $U(\lambda)$, $V(\lambda)$ to (1.1c) in the form[4]

$$U(\lambda) = A/(\lambda+1), \quad V = B/(1-\lambda). \tag{1.4}$$

It then follows that A and B satisfy (1.2) as well as (1.3). Now, given a solution to (1.1a) and (1.1b) with U and V given by (1.4), it is known that new solutions can be produced by finding transformations of Ψ that preserve that λ dependence of U and V. For (1.4), these transformations are found by solving the Riemann–Hilbert problem.[4] This method is called *dressing* and is particularly useful in generating solitonlike solutions.[1]

An overwhelming majority of nonlinear differential equations to which the inverse scattering method has been successfully applied comes from low-dimensional models, usually involving one space and one time variable. An important question is as follows: To what extent can the inverse scattering method be extended to higher dimensions? One problem that automatically arises in this context is how to account for space-time symmetries of those higher-dimensional differential equations. For instance, if one has in mind relativistic equations, then the Poincaré group should leave the solution set of an appropriate generalization of (1.1a)–(1.1c) invariant. The only known nontrivial examples of rel-

a) Present address: Department of Mathematics, University of Virginia, Charlottesville, Virginia 22903.

 0022-2488/88/020336-11$02.50

ativistic equations that admit a Lax representation are the self-dual Yang–Mills equations[5,6] as well as certain supersymmetric Yang–Mills equations.[7–9]

The purpose of this paper is to develop a systematic treatment of Lax representations admitting certain covariance groups. In particular, the nonlinear partial differential equations presented below are invariant under a semidirect product of translations and scale transformations.

In Sec. II A, a method is proposed that, in principle, can be used to derive nonlinear differential equations, covariant under a semidirect product $Q = T \circledS W$ of the translation group T and a group W. In anticipation of Sec. III, this is followed by a discussion of the relevant properties of the loop algebra LSL(2,C).

In Sec. III, two examples are presented. One is the potential KdV equation that is derived from an action of an Abelian group T on $\mathrm{LSL}(2,\mathbb{C})/P$, P being a certain parabolic subgroup LSL(2,C). The second example is the sine–Gordon equation. Choosing a distribution (in the sense of Frobenius) on the tangent bundle of $(\mathrm{LSL}(2,\mathbb{C}) \times \mathrm{LSL}(2,\mathbb{C}))$, the sine–Gordon equation is derived from the action of the translation group, restricted to this distribution.

In Sec. IV, the potential KP equation is studied in relation to the action of the translation group on the infinite-dimensional Grassmannian manifold introduced by Sato[10] and studied by Segal and Wilson[11] and Dorfmeister *et al.*[12]

This work was motivated by a preliminary report by Wilson.[13]

II. INFINITE-DIMENSIONAL MANIFOLDS AND NONLINEAR EQUATIONS

A. The covariant orbits as a source of covariant equations

Let G be an infinite-dimensional Lie group modeled on a complex Banach space. As an example, one can take any loop group LG associated to a finite-dimensional, semisimple group G.[14] The choice of the modeling space will be important insofar as it controls the class of solutions one can generate through the method described in this paper. In this presentation, however, we will not address this problem.

The Lie algebra g of G is assumed to admit the following Banach space decomposition:

$$g = h \oplus v \quad \text{(direct sum of vector spaces)}, \tag{2.1}$$

where both h and v are closed, Banach–Lie subalgebras. Let V, H denote closed Banach–Lie subgroups of G with Lie algebras v and h, whose only intersection $V \cap H$ is the unit element I.

We will study actions of the additive group $\mathbb{C}^k$ (also called translations), where k is some integer on the homogeneous space of left cosets $M = G/V$. In fact, general Banach–Lie group theory[15] gives the global existence of the manifold M, as well as that of the vector fields arising from the left G action on M. Let $[g]$ denote the equivalence class containing g. Here G acts on M as follows:

$$x \cdot [g] = [xg], \quad \text{where } x,g \in G.$$

The following assumptions are made.

(i) The subalgebra v has an Abelian Banach–Lie subalgebra γ with $\dim \gamma \geqslant k$.

(ii) There is a group homomorphism D from $S^1 \to \mathrm{Aut}\, G$, i.e.,

$$D(\rho) \in \mathrm{Aut}\, G, \quad \text{for } \rho \in S^1$$

and

$$D(\rho_1)D(\rho_2) = D(\rho_1\rho_2); \tag{2.2}$$

moreover, both H and V are invariant under $D(\rho)$.

(iii) The map $D(\rho)$ induces a linear map $D(\rho)_*$ of the tangent space T_eG onto itself that is *diagonalizable* and the following holds:

$$g = \bigoplus_{j \in Z} g_j, \tag{2.3}$$

and

$$D(\rho)_* z_j = \rho^j z_j, \quad \text{for } z_j \in g_j. \tag{2.4}$$

(iv) The subalgebra γ is invariant under $D(\rho)$.

By virtue of (ii) we can define an action of $D(\rho)$ on M by setting

$$D(\rho) \cdot [g] \equiv [D(\rho)g] \equiv [g(\rho)]. \tag{2.5}$$

Next, we consider a nonlinear representation of the group of translations $\mathbb{C}^k$ by means of the following procedure.

By splitting (2.3), we write

$$\gamma = \bigoplus_{i \in Z} \gamma_i, \quad \gamma_i \subset g_i$$

and assume that $\dim \gamma_i \leqslant 1$.

Denote by $p_1, p_2, \ldots, p_k$ a basis of $\mathbb{C}^k$. Then we choose a sequence

$$i \to a(i), \quad \text{for } i = 1,\ldots,k,$$

such that

$$\dim \gamma_{a(i)} = 1.$$

Each choice of the sequence $(a(1),\ldots,a(k))$ defines an imbedding $\hat{a}$ of $\mathbb{C}^k$ into γ. Note that the choice of a sequence $i \to a(i)$ amounts to picking a k-dimensional subalgebra (= subspace, since γ is Abelian) of γ, invariant under $D(\rho)$. On the group level we obtain an imbedding of $\mathbb{C}^k$ into Γ, where Γ is connected Lie subgroup of G with Lie algebra γ,

$$\mathbb{C}^k \xrightarrow{\hat{a}} \Gamma.$$

Indeed, let $P_{a(1)},\ldots,P_{a(k)}$ be a basis for

$$\gamma_a = \bigoplus_{i=1}^{k} \gamma_{a(i)},$$

then $\mathbf{t} = t_1 p_1 + \cdots + t_k p_k$ is mapped by $\hat{a}$ into

$$\underline{t} = \exp \sum_{i=1}^{k} t_i P_{a(i)} \in \Gamma. \tag{2.6}$$

Since Γ acts on M, we obtain in this way an action of $\mathbb{C}^k$ on M, namely $\underline{t} \cdot [g] = [tg]$. Denote by $\mathbb{C}^k \cdot [g]$ the orbit of $\mathbb{C}^k$ through $[g]$ and by $O^k(M)$ the set of all orbits, i.e.,

$$O^k(M) = M/\mathbb{C}^k. \tag{2.7}$$

Then $O^k(M)$ is invariant under the scale transformations

$D(\rho)$ as it can easily be seen from the following computation:

$$\begin{aligned} D(\rho)\cdot \underline{t}\cdot[g] &= [D(\rho)\underline{t}D(\rho)g] \\ &= \left[\exp\left(\sum_{i=1}^{k} t_i D(\rho)_* P_{a(i)}\right) g(\rho)\right] \\ &= \left[\exp\left(\sum_{i=1}^{k} t_i P_{a(i)} \rho^{a(i)}\right) g(\rho)\right] \\ &= \underline{t}(\rho)\cdot[g(\rho)], \end{aligned}$$

where $\underline{t}(\rho) = t_1\rho^{j(1)}p_1 + \cdots + t_k\rho^{j(k)}p_k$.

From the above, it follows that $D(\rho)$ *maps the orbit going through the point* $[g]$ *into the orbit passing through the point* $[g(\rho)]$.

Now, assume that $g = g_H g_V$, $g_H \in H$, $g_V \in V$. Then for small t's,

$$\underline{t}g = g_H(\underline{t})g_V(\underline{t}), \tag{2.8}$$

where $g_H(\underline{t}) \in H$, $g_V(\underline{t}) \in V$. Hence we can set

$$\underline{t}\cdot[g] = g_H(\underline{t}) \tag{2.9}$$

with the understanding that (2.9) holds true for a properly chosen neighborhood of zero in $\mathbb{C}^k$ (this set depends on g_H).

Consider the set of commuting vector fields (infinitesimal transformations) on M defined by

$$X_i([g]) = \frac{d}{d\epsilon}\exp(\epsilon P_{a(i)})\cdot[g]|_{\epsilon=0}, \quad i = 1,\ldots,k. \tag{2.10}$$

This set defines an *integrable distribution*[16] on M. As long as Γ has trivial intersection with the stabilizer of $[g]$, locally the orbit $\mathbb{C}^k\cdot[g]$ furnishes an integral manifold for this distribution. Every vector tangent to the orbit $\mathbb{C}^k\cdot[g]$ at a point $\underline{t}\cdot[g]$ is a linear combination of the vectors

$$X_i(\underline{t}\cdot[g]) = \partial_i(\underline{t}\cdot[g]), \tag{2.11}$$

where $\partial_i = \partial/\partial t_i$.

With the help of (2.9) we can identify the commuting vector fields X_i, when restricted to the orbits $\mathbb{C}^k\cdot[g]$, with elements of the Lie algebra h by defining

$$(g_H(\underline{t}))^{-1}{}_* X_i(\underline{t}\cdot[g]) \equiv L_i(\underline{t}\cdot[g]). \tag{2.12}$$

At this point, however, we would like to emphasize a local (chart dependent) character of the above statement. In general, the coefficients of the L_i will have singularities reflecting the topology of M [see also (2.8)].

If G can be represented as a subgroup of the group of linear operators acting on some vector space, then

$$L_i(\underline{t}\cdot[g]) = g_H(\underline{t})^{-1}\,\partial_i(g_H(\underline{t})),$$

which is understood as the multiplication of linear operators.

The commutativity of the vector fields (2.10) translates now into the zero-curvature condition

$$\partial_i L_j(\underline{t}\cdot[g]) - \partial_j L_i(\underline{t}\cdot[g]) + [L_i(\underline{t}\cdot[g]), L_j(\underline{t}\cdot[g])] = 0. \tag{2.13}$$

The commutativity of the vector fields for $L_i(\underline{t}\cdot[g])$ generates, in general, nonlinear differential equations. The type of nonlinearity, however, depends on the particular chart used to evaluate (2.11), as it is illustrated in Sec. IV.

We now discuss properties of these nonlinear equations due to scale covariance.

Since h is invariant under $D(\rho)$, and it is also graded relative to $D(\rho)$, one can write

$$h = \bigoplus_{n\in W} h_n,$$

where $W \subset Z$. This implies that $L_i(\underline{t}\cdot[g])$ is graded, i.e.,

$$L_i(\underline{t}\cdot[g]) = \sum_{n\in W}\sum_{m=1}^{\dim h_n} \beta^n_{i,m}(\underline{t}\cdot[g])h^m_n, \tag{2.14}$$

where for fixed n, $\{h^m_n\}$ denotes a basis of h_n.

From (2.10), it follows that

$$\begin{aligned} D(\rho)_* X_i([g]) &= \frac{d}{d\epsilon} D(\rho)\exp(\epsilon P_{a(i)})\cdot[D(\rho)g]|_{\epsilon=0} \\ &= \frac{d}{d\epsilon}\exp(\epsilon\rho^{a(i)}P_{a(i)})\cdot[g(\rho)]|_{\epsilon=0} \\ &= \rho^{a(i)}X_i([g(\rho)]). \end{aligned} \tag{2.15a}$$

Hence

$$D(\rho)_* L_i(\underline{t}\cdot[g]) = L_i(\underline{t}(\rho)\cdot[g(\rho)])\rho^{a(i)}. \tag{2.15b}$$

Upon combining (2.14) and (2.15b), we arrive at the following transformation rule for the coefficients $\beta^n_{i,m}(\underline{t}\cdot[g])$:

$$\beta^n_{i,m}(\underline{t}[g]) = \rho^{a(i)}\rho^{-n}\beta^n_{i,m}(\underline{t}(\rho)\cdot[g(\rho)])$$

or, equivalently,

$$\beta^n_{i,m}(\underline{t}\cdot[g(\rho)]) = \rho^{n-a(i)}\beta^n_{i,m}(\underline{t}(\rho^{-1})\cdot[g]). \tag{2.16}$$

Since (2.13) is equivalent to an infinite system of nonlinear equations for the coefficients $\beta^n_{i,m}$, by writing (2.13) at the point $[g(\rho)]$, we see that (2.16) describes an action of the scale transformations on the space of solutions to (2.13). The above considerations are summed up in the following.

Proposition 1: The zero-curvature equation (2.13) is invariant under the group of scale transformation $D(\rho)$. The explicit action of $D(\rho)$ on the components of $L(\underline{t}\cdot[g])$ is given by (2.16).

In order to successfully implement the above method, one has to be able to actually compute $X_i(\underline{t}\cdot[g])$, or, at least, establish what form $X_i(\underline{t}\cdot[g])$ may take. The same, of course, may be said about $L_i(\underline{t}[g])$. Two concrete methods will be employed to derive the differential equations. In Sec. IV, the commuting vector fields will be computed explicitly. In Sec. III, however, another method will be used.

Rewrite (2.8) as

$$gg_V^{-1}(\underline{t}) = (\underline{t})^{-1}g_H(\underline{t}), \tag{2.17}$$

and notice that (2.17) can be considered as a map from the orbit $\mathbb{C}^k\cdot[g]$ into a submanifold of G, which is defined by the image of

$$\varphi\colon\ \varphi(\underline{t}\cdot[g]) = gg_V^{-1}(\underline{t}).$$

Indeed, (2.17) does not depend on g_V [see (2.8)], hence φ is well defined as a map from M into G. Next, by identifying the tangent bundle TG with $G\times g$ through the left translations on the group, we see that the vector fields

$$M_i(\underline{t}\cdot[g]) = (\varphi^{-1}(\underline{t}\cdot[g]))_*\,\partial_i\varphi(\underline{t}\cdot[g]), \tag{2.18}$$

defined for $i = 1,\ldots,k$, form the image under φ_* of the vector

fields $X_i(\underline{t}\cdot[g])$. It is important to notice that $M_i(\underline{t}\cdot[g])$ belong to v.

Hence if we use $L_i(\underline{t}\cdot[g])$ instead of $X_i(\underline{t}\cdot[g])$, then we can compare the vectors tangent to the orbit $\mathbb{C}^k\cdot[g]$ with those tangent to the image of φ. From (2.18), it follows that

$$M_i(\underline{t}\cdot[g]) = L_i(\underline{t}\cdot[g]) - g_H^{-1}(\underline{t})P_{a(i)}g_H(\underline{t}). \tag{2.19}$$

The latter can be used to prove that φ is locally an immersion. To this end, it suffices to show that $\varphi_*(\underline{t})$ is a one-to-one map. To prove it, let us consider

$$\sum_{i=1}^{i=k} c_i M_i(\underline{t}\cdot[g]) = 0,$$

where not all of the c_i's are zero. However, in this case, by (2.19) we obtain

$$\sum_{i=1}^{i=k} c_i L_i(\underline{t}\cdot[g]) = g_H(\underline{t})\left(\sum_{i=1}^{i=k} c_i P_i\right) g_H^{-1}(\underline{t})$$

or

$$g_H(\underline{t})\sum_{i=1}^{i=k} c_i L_i(\underline{t}\cdot[g]) g_H^{-1}(\underline{t}) = \left(\sum_{i=1}^{i=k} c_i P_i\right).$$

The left-hand side of the above being in h and the right-hand side being in v implies that both are actually zero. This contradicts the injectivity of $\hat{a}$.

As before, the $M_i(\underline{t}\cdot[g])$'s satisfy the zero-curvature condition

$$\partial_i M_j(\underline{t}\cdot[g]) - \partial_j M_i(\underline{t}\cdot[g]) + [M_i(\underline{t}\cdot[g])] = 0. \tag{2.20}$$

Recall that v is also invariant under $D(\rho)$, thereby it can be written as

$$v = \bigoplus_{l\in U} v_l,$$

where $U\in Z$. Therefore each $M_i(\underline{t}\cdot[g])$ can be decomposed as follows:

$$M_i(\underline{t}\cdot[g]) = \sum_{l\in U}\sum_{m=1}^{\dim v_l} a_{i,m}^l(\underline{t}\cdot[g])v_l^m, \tag{2.21}$$

where for fixed l, $\{v_l^m\}$ denotes a basis of v_l.

Applying $D(\rho)$ to $\varphi(\underline{t}\cdot[g])$, we see that

$$D(\rho)\varphi(\underline{t}\cdot[g]) = \varphi(\underline{t}(\rho)\cdot[g(\rho)]). \tag{2.22}$$

Hence

$$D(\rho)_* M_i(\underline{t}\cdot[g]) = \rho^{a(i)} M_i(\underline{t}(\rho)\cdot[g(\rho)]),$$

and the transformation rule for $\alpha_{i,m}^l$ reads

$$a_{i,m}^l(\underline{t}\cdot[g]) = \rho^{a(i)-l} a_{i,m}^l(\underline{t}(\rho)\cdot[g]),$$

which in turn leads to

$$a_{i,m}^l(\underline{t}\cdot[g(\rho)]) = \rho^{l-\alpha(i)}\alpha_{i,m}^l(\underline{t}(\rho^{-1})\cdot[g]), \tag{2.23}$$

where $l\in U$, $i = 1,\ldots,k$. The zero-curvature condition (2.20), when written in terms of $\alpha_{i,m}^l$, is equivalent to a system of nonlinear equations, which is invariant under the transformation (2.23). However, for the concrete cases presented in Sec. III, this system turns out to be *finite*. The appearance of only finitely many equations makes the use of (2.17) as a tool for generating equations more practical than that of (2.13).

We would like to close this section with several remarks about the implementation of the scaling invariance. In fact, we have used a rather restricted class of the scaling transformations, namely those with $|\rho| = 1$. If one wants to include ρ's of absolute values different from unity, one is likely to find that $D(\rho)_*$ cannot be defined on the whole tangent space $g = T_eG$ [see (iii)]. Indeed, consider an element $n\in g$ such that

$$n = \sum_{k=-\infty}^{\infty} n_k z_k, \quad \text{for } z_k\in g_k.$$

Then, formally at least,

$$D(\rho)_* n = \sum_{k=-\infty}^{\infty} n_k \rho^k z_k.$$

However, in general, in a given Banach norm on g, the series on the right will not converge. This difficulty is indicative of the problems one is likely to encounter when dealing with general scaling transformations. We will return to this problem at the end of Sec. III B.

B. Loop algebras and loop groups

This section contains a review of results concerning loop algebras and loop groups.[14,17] They will be instrumental in Sec. III. First, we give a very concrete example of a loop algebra. Consider the Lie algebra sl(2,C) whose entries are trigonometric polynomials, i.e.,

$$L(\mathrm{sl}(2,\mathbb{C})) = L\otimes \mathrm{sl}(2,\mathbb{C}),$$

where L stands for the algebra of trigonometric polynomials, that is, the finite Fourier series. It is convenient to introduce the standard basis of sl(2,C):

$$e = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad f = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad h = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Consequently, a basis for $L(\mathrm{sl}(2,\mathbb{C}))$ is defined by setting

$$e(k) = \begin{pmatrix} 0 & \lambda^k \\ 0 & 0 \end{pmatrix}, \quad f(k) = \begin{pmatrix} 0 & 0 \\ \lambda^k & 0 \end{pmatrix}, \quad h(k) = \begin{pmatrix} \lambda^k & 0 \\ 0 & -\lambda^k \end{pmatrix}, \tag{2.24}$$

where $\lambda\in S^1$.

In terms of the basis $\{e(k),f(k),h(k);\ k\in\mathbb{Z}\}$ the commutation relations for $L(\mathrm{sl}(2,\mathbb{C}))$ are as follows:

$$\begin{aligned} &[h(j),e(k)] = 2e(j+k), \\ &[h(j),f(k)] = -2f(j+k), \\ &[e(j),f(k)] = h(j+k), \quad [h(j),h(k)] = 0, \end{aligned} \tag{2.25}$$

where $j,k\in\mathbb{Z}$, and all remaining commutators vanish.

In Sec. III, we will use the following subalgebra[18] of $L(\mathrm{sl}(2,\mathbb{C}))$. Define

$$P_{2j+1} = e(2j+1) + f(2j+1) = \begin{pmatrix} 0 & \lambda^{2j+1} \\ \lambda^{2j+1} & 0 \end{pmatrix}, \tag{2.26a}$$

$$X_{2j+1} = -e(2j+1) + f(2j+1) = \begin{pmatrix} 0 & \lambda^{2j+1} \\ -\lambda^{2j+1} & 0 \end{pmatrix}, \tag{2.26b}$$

$$X_{2j} = h(2j) = \begin{pmatrix} \lambda^{2j} & 0 \\ 0 & -\lambda^{2j} \end{pmatrix}, \tag{2.26c}$$

for all $j\in\mathbb{Z}$.

Denote by g the vector space spanned by the elements of the set $\{P_{2j+1}, X_j; j\in\mathbb{Z}\}$. Then g is also a Lie algebra with the following commutation relations:

$$[P_{2j+1}, P_{2k+1}] = 0, \quad [P_{2j+1}, X_k] = 2X_{2j+1+k}, \tag{2.27a}$$

$$[X_j, X_k] = \begin{cases} 2(-1)^{j+k}P_{j+k}, & \text{if } j+k\in 2\mathbb{Z}+1, \\ 0, & \text{otherwise.} \end{cases} \tag{2.27b}$$

From the above commutation relations, it follows that g is a $\mathbb{Z}$-graded Lie algebra, i.e.,

$$g = \bigoplus_{j\in\mathbb{Z}} g_j, \quad \text{where } [g_j, g_k] \subset g_{j+k}.$$

Also (2.27a) and (2.27b) imply that g has three Abelian subalgebras

$$\gamma^{(0)} = \{P_{2j+1},\ j\in\mathbb{Z}\},$$
$$\gamma^{(1)} = \{X_{2j+1},\ j\in\mathbb{Z}\},$$
$$\gamma^{(2)} = \{X_{2j},\ j\in\mathbb{Z}\}.$$

Denote $\gamma^{(0)}\equiv\gamma$. This algebra will be employed in Sec. III.

So far, everything has been purely algebraic, i.e., whenever the summation appeared, it was understood to mean a finite sum. One can complete the loop algebra $L(\mathrm{sl}(2,\mathbb{C}))$, relative to some convenient topology. Since we want to remain within the Banach setting, we can start from a Banach algebra of functions on S^1 and then consider all matrices from $\mathrm{sl}(2,\mathbb{C})$ whose entries belong to this algebra. However, if we want to make use of the method presented in Sec. II A, we have to make sure that all the assumptions of this method are satisfied. In particular, the existence of the Banach space decomposition (2.1) has to be checked once h and v have been chosen. Below we present a general setting which, for all the examples we know of, ensures that (2.1)–(2.5) hold true.

We shall follow Appendix A of Goodman and Wallach.[19]

Let A_w be the space of functions

$$f(\lambda) = \sum_{n\in\mathbb{Z}} a(n)\lambda^n,$$

$\lambda\in S^1$, such that

$$\|f\|_w = \sum_{n\in\mathbb{Z}} |a(n)|w(n) < \infty, \tag{2.28}$$

where w is a weight, that is, w: $\mathbb{Z}\to(0,\infty)$ and $w(k+m)\leqslant w(k)w(m)$.

With the norm $\|\cdot\|$, pointwise multiplication of functions and *-operations given by complex conjugation, A_w is a commutative Banach *-algebra. Here, A_w is *decomposing*, i.e., for f as above,

$$f_+(\lambda) = \sum_{n>0} a(n)\lambda^n, \quad f_-(\lambda) = \sum_{n<0} a(n)\lambda^n$$

belong to A_w and $f = f_+ + f_-$.

Later on we will use the following projection operators:

$$(Pf_+)(\lambda) = \sum_{n>0} a(n)\lambda^n, \quad (Pf_-)(\lambda) = \sum_{n<0} a(n)\lambda^n$$

and

$$(P_0 f)(\lambda) = a(0).$$

Now, let $M_2(A_w)$ denote the algebra of 2×2 matrices with entries in A_w. By defining a norm

$$\|T\| = \left\{\sum_{i,j=1}^{2} \|T_{ij}\|^2\right\}^{1/2}, \tag{2.29}$$

$M_2(A_w)$ is made into a Banach algebra. Next, we define $\mathrm{SL}_2(A_w) = \{g\in M_2(A_w)\colon \det g = 1\}$. This is a closed subgroup of the group of invertible elements of the Banach algebra $M_2(A_w)$. It can be shown (see Lemma A.2 in Goodman and Wallach), that $\mathrm{SL}_2(A_w)$ is a complex lie subgroup of the group invertible elements of $M_2(A_w)$ with Lie algebra $\mathrm{sl}_2(A_w) = \{x\in M_2(A_w)\colon \mathrm{tr}(x) = 0\}$. The Fourier-series analog of the unipotent group of lower-triangular matrices with diagonal entries 1 and the group of upper-triangular matrices are defined by

$$H(A_w) = \{g\in M_2(A_w)\colon P_+ g = 0,\ P_0 g = I, \text{ and } \det g = 1\}, \tag{2.30}$$

$$V(A_w) = \{g\in M_2(A_w)\colon P_- g = 0,\ \det g = 1\}. \tag{2.31}$$

Set $Q_- = P_0 + P_-$, $Q_+ = P_0 + P_-$. Then, by definition,

$$H(A_w)\subset Q_- M_2(A_w), \quad V(A_w)\subset Q_+ M_2(A_w).$$

Both $Q_+M_2(A_w)$ and $Q_-M_2(A_w)$ are Banach algebras, and by repeating the previous procedure, as applied to $\mathrm{SL}_2(A_w)$, one can show that $H(A_w)$ and $V(A_2)$ are Banach Lie groups with the Banach Lie algebras,

$$h(A_w) = \{x\in P_- M_2(A_w)\colon \mathrm{tr}\, x = 0\},$$
$$v(A_w) = \{x\in Q_+ M_2(A_w)\colon \mathrm{tr}\, x = 0\}.$$

To prove that they are both Lie subgroups of $\mathrm{SL}_2(A_w)$, it suffices to show that both are immersed as manifolds in $\mathrm{SL}_2(A_w)$. This can be easily checked at the origin. Indeed, at the origin, both $h(A_w)$ and $v(A_w)$ have topological complements (in fact, they are complements of each other). This guarantees that $H(A_w)$ and $V(A_w)$ are *locally immersed* at the origin in $\mathrm{SL}_2(A_w)$.[20] Then, by translating the origin to any other point of either $H(A_w)$ or $V(A_w)$, this holds true at every point. Hence $H(A_w)$ and $V(A_w)$ are (injectively) immersed as Banach manifolds in $\mathrm{SL}_2(A_w)$. This proves that they are Lie subgroups of $\mathrm{SL}_2(A_w)$.

III. EXPLICIT EXAMPLES

In this section, two concrete applications of the scheme described in the first part of the previous section are presented.

A. Potential KdV equation

Let $\tilde{g}$ denote the completion (in the sense of Sec. II B) of g, g being defined by (2.27a)–(2.27c).

In accordance with Sec. II A, we decompose $\tilde{g}$ into the subalgebras h and v:

$$\tilde{g} = h\oplus v, \tag{3.1}$$

where

$a \in v$

iff

$$a(\lambda) = \alpha e(-1) + \sum_{k=0}^{\infty} (a_k^{(1)} h(2k) + a_k^{(2)} e(2k+1) + a_k^{(3)} f(2k+1)),$$

and

$b \in h$

iff

$$b(\lambda) = \beta f(-1) + \sum_{k=1}^{\infty} (b_k^{(1)} h(-2k) + b_k^{(2)} e(-(2k+1)) + b_k^{(3)} f(-(2k+1))).$$

The subalgebras h and v differ but slightly from $h(A_w)$ and $v(A_w)$. Indeed, v has only one more element $e(-1)$, and this makes v a maximal parabolic subalgebra of $\tilde{g}$. Similarly, h differs by one element from $h(A_w)$. Following the discussion in Sec. II B [see e.g., (2.30) and (2.31)], we can introduce three Banach–Lie groups $\tilde{G}$, H, and V, the Lie algebras of which are, respectively, $\tilde{g}$, h, and v. Now consider $\gamma = \gamma^0$ and an action of $\mathbb{C}^2$ on $\tilde{G}/V$. The imbedding $\hat{a}$ of $\mathbb{C}^2$ into γ is defined by choosing a sequence $1 \to a(1) = 1$, $2 \to a(2) = 3$ (see Sec. II A). This means that $\mathbf{t} = t_1 p_1 + t_3 p_3$ is mapped by $\hat{a}$ [see (2.6)] into

$$\underline{t} = \exp(t_1 P_1 + t_3 P_3) \in \Gamma. \tag{3.2}$$

Then, let $[g]$ be an element in M and suppose $[g]$ be acted upon by $\underline{t}$ to the left. Assume, in addition, that $g = g_H g_V$. Then (at least for small t's), from (2.9) we obtain

$$\underline{t} \cdot [g] = g_H(\underline{t}) \tag{3.3a}$$

where

$$\underline{t}\, g_H g_V = g_H(\underline{t}) g_V(\underline{t}). \tag{3.3b}$$

The last equation describes the flow $\underline{t}g$ in terms of the motion $g_V(\underline{t})$ along the fiber and the motion $g_H(\underline{t})$ on the manifold M.

From (2.26a), we have

$$P_1 = e(1) + f(1), \quad P_3 = e(3) + f(3). \tag{3.4}$$

We can parametrize the first few terms of $g_H(\underline{t})$ as follows [see (2.30)]:

$$g_H(\underline{t}) = I + \beta f(-1) + b(-2) + o(-3), \tag{3.5}$$

where $b(-2) \equiv \lambda^{-2}(aI - \alpha/2\sigma_3)$, σ_3 is the third Pauli matrix, and $o(n)$ contains only powers of λ^k, $k \leqslant n$. Consequently,

$$g_H^{-1}(\underline{t}) = I - \beta f(-1) - b(-2) + o(-3). \tag{3.6}$$

With the help of (3.5) and (3.4), we can write

$$L_1(\underline{t} \cdot [g]) = \partial_1 \beta f(-1) + o(-2), \tag{3.7}$$

and also

$$g_H^{-1}(\underline{t}) P_1 g_H(\underline{t}) = (I - \beta f(-1) - b(-2) + o(-3)) \times \sigma_1(1)(I + \beta f(-1) + b(-2) + o(-3)), \tag{3.8}$$

where $\sigma_1(1) = \lambda \sigma_1$.

By straightforward algebra, (3.8) reduces to

$$g_H^{-1}(\underline{t}) P_1 g_H(\underline{t}) = \sigma_1(1) + \beta\sigma_3 + \alpha e(-1) - \beta^2 f(-1) - \alpha f(-1) + o(-2). \tag{3.9}$$

The last equation, along with (3.7), allows one to compute $M_1(\underline{t} \cdot [g])$, defined earlier by (2.19), yielding

$$M_1(\underline{t} \cdot [g]) = -(\sigma_1(1) + \beta\sigma_3 + \alpha e(-1)) + (\beta^2 + \alpha + \partial_1 \beta) f(-1) + o(-2).$$

However, $M_1(\underline{t} \cdot [g])$ belongs to v, hence, by virtue of (3.1), the coefficient at $f(-1)$, as well as all of the coefficients in $o(-2)$, have to vanish. Thus $M_1(\underline{t} \cdot [g])$ contains only *finitely many* terms,

$$M_1(\underline{t} \cdot [g]) = -(\sigma_1(1) + \beta\sigma_3 + \alpha e(-1)), \tag{3.10}$$

where

$$\beta^2 + \alpha + \partial_1 \beta = 0. \tag{3.11}$$

Similarly, to compute $M_3(\underline{t} \cdot [g])$, we write

$$L_3(\underline{t} \cdot [g]) = \partial_3 \beta f(-1) + o(-2), \tag{3.12}$$

and with the help of (3.9) we obtain

$$\begin{aligned} g_H^{-1}(\underline{t}) P_3 g_H(\underline{t}) &= \lambda^2 g_H^{-1}(\underline{t}) P_1 g_H(\underline{t}) \\ &= \sigma_1(3) + \beta\sigma_3(2) - (\alpha + \beta^2) f(1) + \alpha e(1) + \gamma\sigma_3 \\ &\quad - \delta e(-1) - \sigma f(-1) + o(-2), \end{aligned} \tag{3.13}$$

where some yet to be specified parameters γ, δ, σ have been introduced. Since $M_3(\underline{t} \cdot [g])$ belongs to v, we can right away eliminate the coefficient at $f(-1)$, thereby obtaining

$$\sigma + \partial_3 \beta = 0. \tag{3.14}$$

Furthermore, a relation between γ and δ can be established by the use of the following method. Compute the determinant of $g_H^{-1}(\underline{t}) P_3 g_H(\underline{t})$. It reads $\det \sigma_1(3) = -\lambda^6$, if one evaluates it directly. However, by making use of (3.13) and (3.14), we arrive at $-\lambda^6 + \lambda^2(2\beta\gamma - \delta - \alpha(\alpha + \beta^2) + \partial_3 \beta) + o(0)$. Hence

$$2\beta\gamma - \delta - \alpha(\alpha + \beta^2) + \partial_3 \beta = 0. \tag{3.15}$$

Using (2.19), with the help of (3.12) and (3.13), we can express $M_3(\underline{t} \cdot [g])$ as follows:

$$M_3(\underline{t} \cdot [g]) = \lambda^2 M_1(\underline{t} \cdot [g]) + (\alpha + \beta^2) f(1) - \gamma\sigma_3 + \delta e(-1). \tag{3.16}$$

The remaining part of the derivation consist of (i) the computation of the zero-curvature condition (2.20) for $M_1(\underline{t} \cdot [g])$ and $M_3(\underline{t} \cdot [g])$, and (ii) the elimination of α and δ by the use of the constraints (3.11) and (3.15).

Upon substitution of (3.10) and (3.16) into (2.20) and performing a standard computation, we arrive at a Laurent polynomial in λ; by virtue of (2.20), the coefficients of this polynomial have to vanish. All of these coefficients, along with the corresponding elements of the basis of v, are listed below:

$$(\partial_1 \beta - \partial_1 \beta)\sigma_3(2) = 0;$$

$$(\partial_1 \alpha + 2\gamma) e(1) = 0;$$

$$(\partial_1^2 + 2\beta\, \partial_1 \beta - 2\gamma) f(1) = 0;$$

$$(\partial_3 \beta - \partial_1 \gamma + \partial_1 \beta \alpha + \delta)\sigma_3 = 0;$$

$$(-\partial_3 \alpha - \partial_1 \delta + 2\gamma\alpha + 2\beta\delta) e(-1) = 0.$$

Hence, in particular, we obtain

$$\partial_1\alpha = -2\gamma, \tag{3.17}$$

$$\partial_3\beta = \partial_1\gamma - \alpha\,\partial_1\beta - \delta. \tag{3.18}$$

After substituting (3.17) into (3.18), it follows from (3.11) and (3.15) that

$$\partial_3\beta = \tfrac{1}{2}\partial_1^2(\beta^2 + \partial_1\beta) + 2(\beta^2 + \partial_1\beta) \times\partial_1\beta - \partial_3\beta - \beta\,\partial_1(\partial_1\beta + \beta^2),$$

which in turn is equivalent to the *potential* KdV equation

$$\partial_3\beta = \tfrac{1}{4}\partial_1^3\beta + \tfrac{3}{2}(\partial_1\beta)^2. \tag{3.19}$$

B. The sine–Gordon equation

Consider the subgroup G' of $\tilde{G}$ defined as follows:

$$G' = \left\{ g\in\tilde{G},\ g(\lambda) = \begin{pmatrix} a(\lambda) & b(\lambda) \\ \bar{b}(\lambda) & \bar{a}(\lambda) \end{pmatrix} \right\}, \tag{3.20}$$

where the Fourier coefficients of $a(\lambda)$ and $\bar{a}(\lambda)$ are complex conjugates of each other. The Lie algebra g' of G' consists of elements of the form

$$\begin{pmatrix} i\alpha(\lambda) & \beta(\lambda) \\ \bar{\beta}(\lambda) & -i\alpha(\lambda) \end{pmatrix},$$

where $\alpha(\lambda)$ has real Fourier coefficients. This algebra has a natural decomposition

$$g' = g'_- \oplus g'_+ = g'\cap h(A) \oplus g'\cap v(A). \tag{3.21}$$

Now consider a direct product

$$G'\times G' \equiv G = \{(g_1,g_2);\ g_1,g_2\in G'\}. \tag{3.22}$$

The Lie algebra of a direct product is the direct sum of the corresponding Lie algebras, i.e.,

$$g = (g',g'). \tag{3.23}$$

Define

$$h = (g'_-, g'_+) \tag{3.24}$$

and, respectively,

$$v = \{(x,x);\ x\in g'\}. \tag{3.25}$$

Lemma 1: Let h,v be given by (3.24) and (3.25). Then $g = h\oplus v$.

Proof: One must show that any $(x_1,x_2)\in g$ can be uniquely decomposed in terms of elements from h and v. Let $(x_1,x_2)\in g$ and choose

$$x_-\in g'_- \quad\text{and}\quad x_+\in g'_+,$$

such that $x_1 - x_2 = x_- + x_+$. Moreover, set $x = x_1 - x_-$. Then

$$x_+ + x = x_+ + x_1 - x_- = x_1 - (x_1 - x_2) = x_2$$

and

$$(x_1,x_2) = (x_-,x_+) + (x,x)$$

follows. This sum is direct, since

$$(x,x)\in(g'_-, g'_+)$$

implies $x = 0$.

Denote by V,H the connected subgroups of G whose Lie algebras are v and h and consider the homogeneous manifold $M = G/V$. Note that now the whole Abelian subalgebra γ (Ref. 21) belongs to v. In addition, this time γ has coefficients in R. The imbedding $\hat{a}$ of $\mathbb{R}^2$ into γ is defined by choosing $a(1) = -1$, $a(2) = 1$, which means that $t\in\mathbb{R}^2$ is mapped into

$$\underline{t} = \exp(t_-P_{-1} + t_+P_1)\in\Gamma. \tag{3.26}$$

Now $\underline{t}$ acts on G diagonally, i.e.,

$$\underline{t}(g_1,g_2) \equiv (\underline{t}g_1, \underline{t}g_2). \tag{3.27}$$

Assume that $(g_1,g_2)\in G$ and

$$g_1 = g_{1-}g_V, \quad g_2 = g_{2+}g_V, \tag{3.28}$$

where $(g_{1-},g_{2+})\in H$ and $(g_V,g_V)\in V$. Since for small t's, we have $\underline{t}(g_1,g_2)\in HV$, there exists $g_V(\underline{t})\in V$ such that

$$\underline{t}g_1 = (\underline{t}g_1)_-g_V(\underline{t}), \tag{3.29a}$$

$$\underline{t}g_2 = (\underline{t}g_2)_+g_V(\underline{t}). \tag{3.29b}$$

By eliminating $g_V(\underline{t})$, we can rewrite (3.29a) and (3.29b) as follows:

$$(\underline{t}g_1)(\underline{t}g_2)^{-1} = (\underline{t}g_1)_-(\underline{t}g_2)_+^{-1}. \tag{3.30}$$

It is clear that (3.30) is equivalent to (3.29). We would like to point out that (3.30) is a special case of the Riemann–Hilbert problem.[2] The Riemann–Hilbert problem appears within the context of the Zakharov–Shabat method as a means of proliferating solutions to nonlinear differential equations (see the Introduction). In that setting, our $\underline{t}$ would represent a vacuum solution that becomes dressed through (3.30).

To make contact with the preceding example of the KdV equation, we can spell out the local description of the orbit $\mathbb{R}^2\cdot[g]$:

$$\underline{t}\cdot[g] = ((\underline{t}g_1)_-,(\underline{t}g_2)_+), \tag{3.31}$$

where $g = (g_1,g_2)$. Correspondingly, the vector fields $L_i(\underline{t}\cdot[g])$ take the form

$$L_-(\underline{t}\cdot[g]) = ((\underline{t}g_1)_-^{-1}\partial_{-1}(\underline{t}g_1)_-, (\underline{t}g_2)_+^{-1}\partial_{-1}(\underline{t}g_2)_+), \tag{3.32a}$$

$$L_+(\underline{t}\cdot[g]) = ((\underline{t}g_1)_-^{-1}\partial_{+1}(\underline{t}g_1)_-, (\underline{t}g_2)_+^{-1}\partial_{+1}(\underline{t}g_2)_+), \tag{3.32b}$$

where $\partial_{-1} = \partial/\partial t_-$, $\partial_{+1} = \partial/\partial t_+$.

As discussed in Sec. II A, one can map the orbit $\mathbb{R}^k\cdot[g]$ into the submanifold of G defined by (2.17). The counterpart of (2.17) is given by

$$\psi(\underline{t}\cdot[g]) = (g_1g_V^{-1}(\underline{t}), g_2g_V^{-1}(\underline{t})).$$

The vector fields tangent to this submanifold, when pulled back to the origin of the group G, read

$$\begin{aligned} M_-(\underline{t}[g]) &= (\psi^{-1}(\underline{t}\cdot[g])\partial_{-1}\psi(\underline{t}\cdot[g])) \\ &= (g_V(\underline{t})\partial_{-1}g_V^{-1}(\underline{t}), g_V(\underline{t})\partial_{-1}g_V^{-1}(\underline{t})) \\ &= ((\underline{t}g_1)_-^{-1}\partial_{-1}(\underline{t}g_1)_- - (\underline{t}g_1)_-^{-1}P_{-1}(\underline{t}g_1)_-, \\ &\quad (\underline{t}g_2)_+^{-1}\partial_{-1}(\underline{t}g_2)_+ - (\underline{t}g_2)_+^{-1}P_{-1}(\underline{t}g_2)_+), \end{aligned}$$

where use has been made of (3.29a) and (3.29b). The condition that $M_-(\underline{t}\cdot[g])$ belongs to v means that

$$(\underline{t}g_1)_-^{-1}\partial_{-1}(\underline{t}g_1)_- - (\underline{t}g_1)_-^{-1}P_{-1}(\underline{t}g_1)_-$$

$$= (\underline{t}g_2)_+^{-1}\partial_{-1}(\underline{t}g_2)_+ - (\underline{t}g_2)_+^{-1}P_{-1}(\underline{t}g_2)_+. \tag{3.33}$$

From (2.27), we obtain

$$P_{-1} = e(-1) + f(-1).$$

Set

$$(\underline{t}g_2)_+^{-1} = \begin{pmatrix} e^{i\alpha} & 0 \\ 0 & e^{-i\alpha} \end{pmatrix}(I + O(\lambda)), \tag{3.34}$$

where $O(\lambda)$ contains only positive powers of λ. Upon substituting (3.34) into (3.33) and comparing powers of λ, we find that

$$M_-(\underline{t}\cdot[g]) = -(\cos 2\alpha\,\sigma_1(-1) - \sin 2\alpha\,\sigma_1(-1), \cos 2\alpha\,\sigma_1(-1) - \sin 2\alpha\,\sigma_2(-1)). \tag{3.35}$$

A similar computation performed for

$$M_+(\underline{t}\cdot[g]) = \psi(\underline{t}\cdot[g])^{-1}\partial_1\psi(\underline{t}\cdot[g])$$

gives

$$M_+(\underline{t}\cdot[g]) = -(\sigma_1(1) + i\partial_1\alpha\sigma_3, \sigma_1(1) + i\partial_1\alpha\sigma_3). \tag{3.36}$$

The final step is to evaluate the zero-curvature condition (2.20). A straightforward computation shows that the above equation is equivalent to

$$(\partial_{-1}\partial_{+1}\alpha - 2\sin 2\alpha)\sigma_3 = 0, \tag{3.37}$$

which in turn produces the sine–Gordon equation

$$\partial_{-1}\partial_{+1}\alpha = 2\sin 2\alpha. \tag{3.38}$$

This equation is invariant under the following transformation:

$$\alpha^{\text{new}}(t_-, t_+) = \alpha(\rho t_-, \rho^{-1}t_+),$$

which agrees with (2.23), if one sets $l = -1$, $a(1) = -1$ [see also (3.35)].

We would like to remark that the above formula holds only for real ρ. In our setting, this implies that ρ can take on two discrete values $\{-1, 1\}$. However, it is possible to extend our discussion to other real ρ's by considering those elements $g \in G$ for which $g(\rho) \in G$, where $g(\rho)(\lambda) = g(\rho\lambda)$. For instance, let us choose the weight function w [see (2.28)] to be equal to one. With this choice of w, A_w becomes a Banach space of maps, $S^1 \to \mathbb{C}$, with absolutely convergent Fourier series. Then we can consider a subset of A_w that consists of real analytic functions, that is, functions analytic in some annulus containing S^1. Take g to be one of such functions, g is therefore analytic in the annulus $1/r < |\lambda| < r$ for some $r > 1$. Thus $g(\rho) \in A_w$ if $\rho < r$.

Finally, (3.10) along with (3.16), (3.35), and (3.36) are Lax pairs for the KdV and the sine–Gordon equation. They have been derived from commuting vector fields of infinitesimal translations on infinite-dimensional manifolds.

IV. SCALE INVARIANT RICCATI EQUATION ON AN INFINITE-DIMENSIONAL GRASSMANIAN MANIFOLD

In this section, an account is given of another example illustrating how one can relate nonlinear differential equations to an action of the translation group on an infinite-dimensional manifold.

Consider the space $H = L^2(S^1)$ of all square-integrable complex-valued functions on the unit circle $S^1 = \{\lambda \in \mathbb{C};\ |\lambda| = 1\}$. This is a separable Hilbert space with the scalar product $(\cdot|\cdot)$ given by

$$(f,g) = \frac{1}{2\pi}\int_0^{2\pi} f\,\overline{(e^{i\varphi})}\,g(e^{i\varphi})\,d\varphi. \tag{4.1}$$

One defines two closed subspaces H_+ and H_-, the former being spanned by the basis elements $\{\lambda^k, k \geq 0\}$ and the latter by $\{\lambda^k, k < 0\}$. Furthermore, let V^+ denote the set of Hilbert–Schmidt operators from H_+ to H_-, and V^- the set of all bounded operators from H_- to H_+. With any pair (x,y) such that $x \in V^+$, $y \in V^-$, one can associate a closed subspace $(x{:}y)$ of H by setting,[12]

$$(x{:}y) = \text{closed span of the columns of the matrix } \begin{pmatrix} I - yx \\ -x \end{pmatrix}, \tag{4.2}$$

where the matrix is written relative to the splitting $H = H_+ \oplus H_-$. Another pair $(x',y') \in V^+ \times V^-$ gives the same closed subspace if and only if there is an invertible operator $g \in \mathrm{GL}(H_+)$ such that

$$\begin{pmatrix} I - yx \\ -x \end{pmatrix} = \begin{pmatrix} I - y'x' \\ -x' \end{pmatrix} g. \tag{4.3}$$

From (4.3), it follows that $x = x'g$, and, upon substituting x for $x'g$ in the upper block of (4.2), one obtains

$$g = I - (y - y')x. \tag{4.4}$$

Since g is invertible, $I - (y - y')$ has to be invertible for the pair (x',y') to represent the same subspace as (x,y). If $I - (y - y')x$ is invertible, x' can be found from the relation

$$x' = x[I - (y - y')x]^{-1}. \tag{4.5}$$

The set X of all closed subspaces of H constructed via (4.2) can be shown to be a Hilbert manifold. Here X is an infinite-dimensional analog of the Grassman manifold. A more complete statement reads[12] as follows.

Theorem 1: (i) The sets $U_a = \{(x{:}a);\ x \in V^+\}$, $a \in V^-$, together with the maps

$$\Psi_a:\ V^+ \to U_a,\quad x \to (x{:}a),$$

are charts of X.

(ii) The "change of coordinates"

$$\varphi_{ba} = \Psi_b^{-1} \circ \Psi_a^{-1}(U_a \cap U_b) \to \Psi_b^{-1}(U_a \cap U_b),$$

is given by

$$\varphi_{ba}(x) = x[I - (a - b)x]^{-1}.$$

(iii) For any $a \in V^-$, the chart U_a is open and dense in X.

From (i), it is clear that X is a Hilbert manifold: X is modeled on the Hilbert space of Hilbert–Schmidt operators.

A. C^∞ action on X

Let $\underline{t}$ be a finite sequence of numbers from $\mathbb{C}$, i.e., $\underline{t} = (t_1, t_2, \ldots)$. The set of all such $\underline{t}$'s forms a vector space. We denote this vector space by $\mathbb{C}^\infty$. The action of $\underline{t}$ on X is given by

$$\underline{t} = \exp \sum_{k=1} t_k P^k, \tag{4.6}$$

where P_k is the kth power of the shift operator P ($P\lambda^m = \lambda^{m+1}$ for each $m \in \mathbb{Z}$). Hence an imbedding of $\mathbb{C}^\infty$ into the group of automorphisms of X is accomplished by the choice of the Abelian algebra (see Sec. II A)

$$\gamma = \bigoplus_{i \in \mathbb{Z}} \mathbb{C} P^i \tag{4.7}$$

and the map

$$i \to a(i) = i, \quad \text{for } i \geq 1. \tag{4.8}$$

To describe explicitly the action of $\underline{t}$ on X, we write the matrix of $\underline{t}$ as follows:

$$\underline{t} = \begin{pmatrix} a(\underline{t}) & b(\underline{t}) \\ 0 & d(\underline{t}) \end{pmatrix}, \tag{4.9}$$

where the block entries are written relative to the splitting $H = H_+ \oplus H_-$. Now let W be a point in X. Hence $W = (x:y)$ for some $(x,y) \in V^+ \times V^-$. Then, by using (4.2), we define a new point $\underline{t} \cdot W$ by setting $\underline{t} \cdot W =$ closed span of the columns of the matrix

$$\underline{t} \begin{pmatrix} I - yx \\ -x \end{pmatrix} a(\underline{t})^{-1}. \tag{4.10}$$

Carrying out the multiplication of matrices in (4.10), we obtain

$$\underline{t} \cdot W = (x_L(\underline{t}) : y(\underline{t})), \tag{4.11}$$

where

$$x_L(\underline{t}) = d(\underline{t}) x a(\underline{t})^{-1}, \tag{4.12a}$$

$$y(\underline{t}) = a(\underline{t}) y d(\underline{t})^{-1} + b(\underline{t}) d(\underline{t})^{-1}. \tag{4.12b}$$

Formulas (4.12a) and (4.12b) describe an action of the group $\mathbb{C}^\infty$ on $V^+ \times V^-$. It is important to notice that (4.12a) itself describes a *linear* action of $\mathbb{C}^\infty$ on V^+. However, from the point of view of the manifold X, this linear action on V^+ *has to be accompanied by a compatible change of charts as defined by* $y(\underline{t})$. This point of view will be explored more thoroughly in Sec. IV C.

In the remaining part of this section we study the action of $\mathbb{C}^\infty$ in the particular chart U_0.

Recall the definitions

$$U_0 = \{(x:0);\ x \in V^+\}$$

and

$$\Psi_0 \colon x \to (x:0).$$

Furthermore, set $y = 0$ in (4.12b), then, from (4.12a) and (4.12b),

$$\underline{t} \cdot W = \underline{t} \cdot (x:0) = (x_L(\underline{t}), b(\underline{t}) d(\underline{t})^{-1}). \tag{4.13}$$

The last equation means that $d(\underline{t}) x a^{-1}(\underline{t})$ is the coordinate of the point $\underline{t} \cdot W$ in the chart $U_{q(t)}$, where $q(t) = b(\underline{t}) d(\underline{t})^{-1}$. Now assuming that $\underline{t} \cdot W$ is in the domain of the chart U_0, we can write (4.13) in terms of the U_0 chart. Indeed, from Theorem 1 (ii), we obtain

$$\begin{aligned} \underline{t} \cdot (x:0) &= (x_L(\underline{t}) [I - b(\underline{t}) d(\underline{t})^{-1} x_L(\underline{t})]^{-1} : 0) \\ &= (x_L(\underline{t}) [I - b(\underline{t}) x a(\underline{t})^{-1}]^{-1} : 0). \end{aligned}$$

Hence

$$\underline{t} \cdot x \equiv x(\underline{t}) = x_L(\underline{t}) [I - b(\underline{t}) d(\underline{t})^{-1} x_L(\underline{t})]^{-1} \tag{4.14}$$

describes for small $\underline{t}$'s a *nonlinear representation of the group of translations* on V^+.

On the other hand, $\underline{t} \cdot x$ is a point of the orbit $\mathbb{C}^\infty \cdot W$, if written in the U_0 chart. Hence from (4.14) we can compute the vector fields $X_i(\underline{t} \cdot W)$ defined by (2.10) and (2.11), which are tangent to the orbit $\mathbb{C}^\infty \cdot W$. A simple computation shows that in the chart U_0,

$$\begin{aligned} X_i(\underline{t} \cdot W) &= \partial_i x(\underline{t}) \\ &= P^i{}_- x(\underline{t}) - x(\underline{t}) P^i{}_+ + x(\underline{t}) s(i) x(\underline{t}), \\ & \quad i \geq 1, \end{aligned} \tag{4.15}$$

where

$$P^i = \begin{pmatrix} P^i{}_+ & s(i) \\ 0 & P^i{}_- \end{pmatrix}.$$

Therefore the Riccati equation (4.15) describes the vector fields tangent to the orbit $\mathbb{C}^\infty \cdot W$. A generic role this equation plays in relating the KP equation to the action of $\mathbb{C}^\infty$ on X is discussed in Ref. 12. Moreover, in the same paper, it has been shown that one of the matrix entries of $x(\underline{t})$, namely $x(\underline{t})_{-1,0} = (\lambda^{-1} | x(\underline{t}) | \lambda^0)$, satisfies the potential KP equation

$$\begin{aligned} \tfrac{3}{4} \partial_2^2 x(\underline{t})_{-1,0} = \partial_1 [& \partial_3 x(\underline{t})_{-1,0} - \tfrac{1}{4} \partial_1^3 x(\underline{t})_{-1,0} \\ & + \tfrac{3}{2} (\partial_1 x(\underline{t})_{-1,0})^2]. \end{aligned} \tag{4.16}$$

B. Scale invariance of the Riccati equation

The remainder of this section is devoted to the study of the scaling properties of the Riccati equation (4.15), as well as the differential equation related to the linear representation (4.12a).

First, consider the operator $D(\rho) \in \mathrm{GL}(H)$, $\rho \in \mathbb{C}$, $|\rho| = 1$, defined by the equation

$$(D(\rho) f)(\lambda) = f(\rho \lambda), \quad \text{for } f \in H = L^2(S^1). \tag{4.17}$$

The matrix representation for this operator reads

$$D(\rho) = \begin{pmatrix} \cdot & & & & 0 \\ & \rho^1 & & & \\ & & \rho^0 & & \\ & & & \rho^{-1} & \\ 0 & & & & \cdot \end{pmatrix}. \tag{4.18}$$

The corresponding diagonal blocks (relative to H_+ and H_-) are denoted as $D_+(\rho)$ and $D_-(\rho)$.

Next we consider the adjoint action of $D(\rho)$ on γ,

$$\begin{aligned} (D(\rho) P^k D(\rho)^{-1}) f(\lambda) &= (D(\rho) P^k) f(\rho^{-1} \lambda) \\ &= D(\rho) \lambda^k f(\rho^{-1} \lambda) = \rho^k \lambda^k f(\lambda) \\ &= \rho^k (P^k f)(\lambda). \end{aligned}$$

That is,

$$D(\rho) P^k D(\rho^{-1}) = \rho^k P^k, \tag{4.19}$$

which agrees with (2.4).

Finally, define an action of $D(\rho)$ on X by setting

$$D(\rho)\cdot(x{:}y) = (D_-(\rho)xD_+^{-1}(\rho){:}D_+(\rho)yD_-^{-1}(\rho))$$
$$\equiv(x(\rho){:}y(\rho)). \tag{4.20}$$

From (4.11), we obtain that

$$D(\rho)\cdot\underline{t}\cdot(x{:}y) = D(\rho)\cdot(x_L(\underline{t}){:}y(\underline{t}))$$
$$= (D_-(\rho)x_L(\underline{t})D_+^{-1}(\rho){:}$$
$$D_+(\rho)y(\underline{t})D_-^{-1}(\rho)). \tag{4.21}$$

However, with the help of (4.12a) and (4.12b), we obtain

$$D_-(\rho)x_L(\underline{t})D_+^{-1}(\rho)$$
$$= D_-(\rho)d(\underline{t})D_-^{-1}(\rho)x(\rho)D_+(\rho)a(\underline{t})^{-1}D_+^{-1}(\rho).$$

To interpret the latter formula, with the help of (4.9), we compute

$$D(\rho)\underline{t}D(\rho)^{-1}$$
$$= \begin{pmatrix} D_+(\rho)a(\underline{t})D_+^{-1}(\rho) & D_+(\rho)b(\underline{t})D_-^{-1}(\rho) \\ 0 & D_-(\rho)d(\underline{t})D_-^{-1}(\rho) \end{pmatrix}. \tag{4.22}$$

By virtue of (4.6) and (4.19), (4.22) can be written as

$$D(\rho)\underline{t}D^{-1}(\rho) = \underline{t}(\rho). \tag{4.23}$$

From (4.23), it follows that

$$D_-(\rho)x_L(\underline{t})D_+^{-1}(\rho) = d(\underline{t}(\rho))x(\rho)a^{-1}(\underline{t}(\rho))$$

holds. Similarly, by (4.12b), (4.22), and (4.23),

$$D_+(\rho)y(\underline{t})D_-^{-1}(\rho)$$
$$= D_+(\rho)a(\underline{t})yd^{-1}(\underline{t})D_-^{-1}(\rho)$$
$$+ D_+(\rho)b(\underline{t})d(\underline{t})D_-^{-1}(\rho)$$
$$= a(\underline{t}(\rho))y(\rho)d(\underline{t}(\rho))^{-1} + b(\underline{t}(\rho))d(\underline{t}(\rho))^{-1}.$$

Hence

$$D(\rho)\cdot\underline{t}\cdot(x{:}y) = \underline{t}(\rho)\cdot(x(\rho){:}y(\rho)). \tag{4.24}$$

The upshot is clearly the same as in Sec. II.

Corollary 1: The set of all orbits

$$O^\infty(X) = X/\mathbb{C}^\infty$$

is invariant under the scale transformation $D(\rho)$.

Let $D(\rho)_*$ denote the induced action of $D(\rho)$ on vectors tangent to X.

Corollary 2:

$$D(\rho)_* X_i(\underline{t}\cdot W) = \rho^i X_i(\underline{t}(\rho)\cdot W(\rho)),$$

for $i \geqslant 1$.

Writing the above formula at the point $\underline{t}(\rho^{-1})\cdot W$ and using (4.21) and (4.15), we arrive at

$$D(\rho)_* X_i(\underline{t}(\rho^{-1})\cdot W) = \rho^i X_i(\underline{t}\cdot W(\rho))$$

or

$$X_i(\underline{t}\cdot W(\rho)) = \rho^{-i}D_-(\rho)X_i(\underline{t}(\rho^{-1})\cdot W)D_+^{-1}(\rho),$$

which is the counterpart of (2.16).

C. Group-theoretical linearization of the Riccati equation

This part contains a detailed study of the linear representation of the group $\mathbb{C}^\infty$ on V^+, given by (4.12a). The whole formalism that has been developed so far is equally well applicable to this case. Here V^+ can be thought of as a manifold M (or X), with the global canonical chart V^+. Then the orbit

$$\underline{t}_L\cdot x \equiv x_L(\underline{t}) = d(\underline{t})xa(\underline{t})^{-1} \tag{4.25}$$

gives rise to the vector fields tangent to it,

$$X_i(\underline{t}_L\cdot x) = \partial_i x_L(\underline{t}) = P^i_- x_L(\underline{t}) - x_L(\underline{t})P^i_+ . \tag{4.26}$$

Clearly, (4.26) corresponds to the linear part of (4.15). Moreover, the following holds.

Theorem 2: The orbit $x_L(\underline{t})$ satisfies the linearized potential KP equation [see (4.16)],

$$\tfrac{3}{4}\partial_2^2 x_L(\underline{t}) = \partial_1[\partial_3 x_L(\underline{t}) - \tfrac{1}{4}\partial_1^3 x_L(\underline{t})].$$

Proof: From (4.26), by a straightforward computation, one obtains

(i) $3\partial_2^2 x_L(\underline{t}) = 3P^4_- x_L(\underline{t}) - 6P^2_- x_L(\underline{t})P^2_+ + 3x_L(\underline{t})P^4_+ ;$

(ii) $4\partial_1\partial_3 x_L(\underline{t}) = 4P^4_- x_L(\underline{t}) - 4P^3_- x_L(\underline{t})P_+ - 4P_- x_L(\underline{t})P^3_+ + 4x_L(\underline{t})P^4_+ ;$

(iii) $-\partial_1^4 x_L(\underline{t}) = -P^4_- x_L(\underline{t}) + 4P^3_- x_L(\underline{t})P_+ - x_L(\underline{t})P^4_+ - 6P^2_- x_L(\underline{t})P^2_+ + 4P_- x_L(\underline{t})P^3_+ .$

Adding (ii) and (iii), we obtain (i).

Corollary 3: Each matrix entry of $x_L(\underline{t})$ satisfies the linearized potential KP equation.

Finally, recall that $x(\underline{t})_{-1,0}$ satisfies the potential KP equation. From (4.14) and Theorem 2, it follows that

$$x(\underline{t})_{-1,0} = \sum_{k=0}^{\infty} x_L(\underline{t})_{-1,k}([I - b(\underline{t})xa(\underline{t})^{-1}]^{-1})_{k,0}. \tag{4.27}$$

Hence, all the solutions to the potential KP equation derived from the proposed scheme are linear combinations of the elements $x_L(\underline{t})_{-1,k}$ which satisfy the linearized KP equation, with coefficients given by the transition matrix $\varphi_{Oq(\underline{t})}$ [Theorem 1 (ii) and (4.14)], for $q(\underline{t}) = b(\underline{t})d(\underline{t})^{-1}$.

V. SUMMARY AND CONCLUSIONS

This paper is to be viewed as an attempt to describe nonlinear partial differential equations admitting certain covariance groups from the perspective of infinite-dimensional geometry. The study is restricted here to those equations that are invariant under a semidirect product of translations and scale transformations.

The method proposed requires that an infinite-dimensional manifold M have among its automorphisms both translations and scale transformations. One studies a set of submanifolds of M that is invariant under both translations and scale transformations. Such a set is provided by the union of orbits of the group of translations. By restricting the corresponding vector fields (infinitesimal translations) to these orbits and exploring their commutativity, one obtains nonlinear equations, invariant under translations and scale transformations. The potential KdV equation as well as the sine–Gordon equation are derived as an illustration. These

equations are well known; their derivations, however, are different from the standard ones involving the matrix Lax formulation. In particular, the method presented here starts from the action of an Abelian group and Lax pairs are constructed from the commuting vector fields generated by this action.

Moreover, as has been alluded to in Sec. III, our construction exhibits some commonalities with the Zakharov–Shabat dressing method, more precisely, with the so-called vacuum dressing[4] and the Riemann–Hilbert problem associated to it. However, as the example of the potential KdV equation suggests, more general factorization problems than those appearing in connection to the sine–Gordon equation or principal chiral models are needed. In those cases, one considers the factorization of a matrix defined on the unit circle into the matrix analytic inside the unit circle and another matrix analytic outside. This situation should be contrasted with (3.3b) and (2.8).

Finally, Sec. IV contains a discussion of the Riccati equation (4.15) obtained via the same method using the infinite-dimensional Grassmannian manifold (4.2). The quadratic nonlinearity of the latter equation is discussed from the point of view of the geometry of this manifold. Explicit formulas relating the quadratic nonlinearity to the transition functions between different charts are derived and presented in the form of (4.27). It would be very interesting to compare the results of Sec. IV with the recent studies of systems of nonlinear *ordinary* differential equations related to finite-dimensional homogeneous spaces.[22–25] Our method has been successfully applied to several other examples (see Ref. 26 for more). What remains unclear to us is the exact range of applicability of our method. In particular, we consider it to be a very interesting question as to which extent one can apply the results of the paper to nonlinear equations invariant under the Poincaré group.

ACKNOWLEDGMENTS

The author is grateful to Dr. J. Harnad for bringing Ref. 13 to his attention. He would like to express his gratitude to Professor R. L. Anderson and Professor J. Dorfmeister for creating a stimulating atmosphere in which this work was performed, and it was Professor J. Dorfmeister's help and persistence which saw this work to its end. Also, a great deal of help in preparing this manuscript, extended to the author by Dr. M. Adams and Mr. D. Boye, is gratefully appreciated.

[1]See, e.g., M. J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform* (SIAM, Philadelphia, PA, 1981); S. Novikov, S. V. Manakov, L. P. Pitaevskii, and V. E. Zakharov, *Theory of Solitons* (Consultants Bureau, New York, 1984).
[2]V. E. Zakharov and A. B. Shabat, Zh. Eksp. Teor. Fiz. **61**, 118 (1971).
[3]C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, Phys. Rev. Lett. **19**, 1095 (1967).
[4]V. E. Zakharov and A. V. Mikhailov, Sov. Phys. JETP **47**, 1017 (1978).
[5]A. A. Belavin and V. E. Zakharov, Phys. Lett. B **73**, 53 (1978).
[6]K. Pohlmeyer, Commun. Math. Phys. **72**, 34 (1980).
[7]E. Witten, Phys. Lett. B **77**, 394 (1978).
[8]J. Harnad and S. Shnider, "Constraints and field equations for ten-dimensional super Yang–Mills theory," Stevens Institute of Technology, preprint, 1985.
[9]E. Abdalla, M. Forger, and M. Jacque, "Higher conservation laws for ten-dimensional supersymmetric Yang–Mills theories," CERN-TH.4431/86.
[10]M. Sato, RIMS Kokyuroku **439**, 30 (1980).
[11]G. Segal and G. Wilson, "Loop groups and equations of KdV type," Publ. IHES **61**, 1985.
[12]J. Dorfmeister, E. Neher, and J. Szmigielski, "Automorphisms of Banach manifolds associated with the KP equation" (to appear).
[13]G. Wilson, C. R. Acad. Sci. (Paris) **299**, 587 (1984).
[14]G. Segal and A. Prestley, *Loop Groups and their Representations* (Oxford U. P., New York, 1986).
[15]N. Bourbaki, *Groupes et Algèbras de Lie* (Hermann, Paris, 1972), Chaps. 2 and 3.
[16]C. Chevalley, *Theory of Lie Groups* (Princeton U. P., Princeton, NJ, 1954).
[17]V. G. Kac, *Infinite Dimensional Lie Algebras* (Cambridge U. P., Cambridge, 1985).
[18]This subalgebra is in fact isomorphic to $L(\mathrm{sl}(2,\mathbb{C}))$.
[19]R. Goodman and N. R. Wallach, J. Math. **347**, 69 (1984).
[20]S. Lang, *Differential Manifolds* (Springer, Berlin, 1985), Proposition 2, p. 27.
[21]The algebra γ is identified with $\mathrm{diag}(\gamma,\gamma)$.
[22]R. L. Anderson, Lett. Math. Phys. **4**, 1 (1980).
[23]R. L. Anderson, J. Harnad, and P. Winternitz, Lett. Math. Phys. **5**, 143 (1981).
[24]J. Harnad, P. Winternitz, and R. L. Anderson, J. Math. Phys. **24**, 1062 (1983).
[25]M. A. De Olmo, M. A. Rodriguez, and P. Winternitz, J. Math. Phys. **27**, 14 (1985).
[26]J. Szmigielski, Ph.D. thesis, University of Georgia, 1986.

Lax conjecture about the eigenspeed of the KdV equation

Xu Bang Qing
Institute of Atomic Energy, Academia Sinica, Beijing, People's Republic of China

Wang Yong Jiu
Changsha Railway Institute, Changsha, People's Republic of China

Shao Chang Gui
Wuhang Engineering College, Wuhang, People's Republic of China

(Received 13 March 1986; accepted for publication 7 October 1987)

The Lax conjecture for the KdV equation $u_t + 6uu_x + u_{xxx} = 0$ is proved. Let u be the solution of the KdV equation, which is defined for all x and t and vanishes at $x = \pm\infty$. Then there exists a discrete set of positive numbers $c_1,...,c_N$—called the eigenspeeds of u—and sets of phase shifts $\theta_j^{\pm}$ such that

$$\lim_{t\to\pm\infty} u(x+ct,t) = \begin{cases} S(x-\theta_j^{\pm},c_j), & \text{if } c = c_j, \\ 0, & \text{if } c \neq c_j, \end{cases}$$

where S is a solitary wave [P. D. Lax, Commun. Pure Appl. Math. **21**, 467 (1968)].

I. INTRODUCTION AND RESULTS

Kruskal and Zabusky,[1-3] by analyzing numerically computed solutions of the KdV equation, observed two important phenomena: (i) for very large time t, the solution forming many solitons travels toward the right side and, yielding a "oscillatory tail," travels toward the left side; and (ii) the solution is quite stable such that any soliton preserves its shape and velocity upon collision with any other soliton. Gardner *et al.*[4] developed (ii) in detail, in which they established the remarkable inverse method. This method is well known and has had great success. About (i), Lax gives a precise formulation.

Let $u(x,t)$ be the solution of the Cauchy problem of the KdV equation:

$$\begin{aligned} &u_t(x,t) + 6u(x,t)u_x(x,t) + u_{xxx}(x,t) = 0, \\ &u(x,0) = u_0(x), \quad -\infty < x < \infty. \end{aligned} \tag{1}$$

Then there exist constants c_j, $j = 1,2,...,N$, called eigenspeeds of $u(x,t)$, and phase shifts $\theta_j^{\pm}$, $j = 1,2,...,N$, such that

$$\lim_{t\to\pm\infty} u(x+ct) = \begin{cases} S(x-\theta_j^{\pm},c_j), & \text{if } c = c_j, \\ 0, & \text{if } c \neq c_j, \end{cases}$$

where S is a solitary wave. Lax[5] asserted the following.

Equation (1) is called the evolutional equation corresponding to the problem of the eigenvalue of

$$\psi_{xx} + [k^2 + u(x,t)]\psi = 0, \tag{2}$$

where $u(x,t)$ is the solution of Eq. (1). If the c_j are the eigenspeeds of the solution of Eq. (1), $-c_j/4$ will be the eigenvalue of Eq. (2), and vice versa, i.e., if $-k_j^2$ is the eigenvalue of Eq. (2), $4k_j^2$ will be the eigenspeed of the solution of the equation.

The first half of this assertion is proved by Lax.[5] Tanaka[6,7] gave a demonstration about this conjecture. This paper gives a new demonstration improving on Tanaka's result both by allowing more general data and by getting more complete results. Tu[8] also hopes that this problem can be fully resolved.

Theorem 1: Let the initial data $u_0(x)$ of the Cauchy problem [Eq. (1)] satisfy the following conditions:

$$\begin{aligned} &u_0(x) \in D^4(x), \\ &u_0^j(x) = o(|x|^{-10}), \quad j \leq 4, \quad \text{as } x \to \pm\infty, \end{aligned} \tag{3}$$

$$\int_{-\infty}^{\infty} u_0(x)dx > 0. \tag{4}$$

Then, corresponding to the eigenvalue $-k_j^2$ of (2), there must be N eigenspeeds c_j of (1) with

$$c_j = 4k_j^2, \quad j = 1,2,...,N.$$

This theorem will be proved using the inverse scattering method.

II. INVERSE SCATTERING METHOD

The inverse scattering method has been stated in many papers,[9,10,11] and here we only describe it in outline.

In the pair of problems (1) and (2), the solution of (1) is called the potential of (2). When $t = 0$, the potential $u(x,0) = u_0(x)$ is known, by which we may obtain the following scattering data:

(a) $-k_1^2, -k_2^2, ..., -k_N^2$,

$k_1 < k_2 < \cdots < k_N$ (eigenvalues),

(b) c_j^2, $j = 1,2,...,N$ (normalizing constants corresponding to eigenvalues $-k_j^2$),

(c) $b(k)$ (the reflection coefficient).

The method for obtaining scattering data from the potential is called the direct scattering method. If the scattering coefficients evolve in time according to the rules

$$\begin{aligned} &k_j(t) = k_j, \quad c_j^2(t) = c_j^2 e^{8k_j^3 t}, \quad j = 1,2,...,N, \\ &b(k,t) = b(k)e^{8ik^3 t}, \end{aligned}$$

then the potential of Eq. (2) corresponding to these scattering coefficients will be the solution of the Cauchy problem of Eq. (1) under the initial data $u(x,0) = u_0(x)$. The method for obtaining the potential in terms of the known scattering

 0022-2488/88/020347-06$02.50

coefficients is called the inverse scattering method. The topic of this method is to establish the Gel'fand–Levitan–Marchenko (GLM) equation using the scattering coefficients:

$$-\infty<x<\infty,\quad y\geqslant 0,\quad t\geqslant 0,$$

$$B_d(x+y,t)=\sum_{j=1}^{N} c_j^2 e^{8k_j^3 t}e^{-2k_j(x+y)},$$

$$B_c(x+y,t)=\frac{1}{\pi}\int_{-\infty}^{\infty} b(k)e^{8ik^3t+2ik(x+y)}\,dk,$$

$$B(x+y,t)=2B_d(x+y,t)+B_c(x+y,t),$$

$$K(x,y,t)+\int_0^{\infty} B(x+y+z,t)K(x,z,t)dz+B(x+y,t)=0 \quad \text{(GLM equation)}, \tag{5}$$

$$u(x,t)=-\frac{\partial}{\partial x}K(x,0,t). \tag{6}$$

Cohen[11] has demonstrated that, under initial data satisfying (3), we have

$$b(k)\in C^7(k),\quad |b^{(m)}(k)|=o(|k|^{-4}),$$
$$\text{as } |k|\to\infty,\quad m<7, \tag{7}$$

and that the solution of Eq. (1) can really be resolved by (5) and (6) with

$$K(x,y,t)\in L^1[0,\infty)\cap L^{\infty}[0,\infty)\in L^2[0,\infty),$$
$$-\infty<x<\infty,\quad t\geqslant 0.$$

Zabusky[12] has proved that, under initial data satisfying (4), at least one of the eigenvalues of Eq. (2) exists: Thus the first conclusion of Theorem 1 holds and the remainder requires only that we prove the second conclusion. We shall prove it in two steps. First, let $b(k)\equiv 0$, i.e., let $u(x,t)$ be an N-soliton solution, and, next, let $b(k)\not\equiv 0$.

Theorem 2: Suppose that $u_d(x,t)$ is the N-soliton solution of Eq. (1), i.e., where $B_c(x,t)\equiv 0$, under initial data satisfying (3) and (4). Then there must exist N eigenspeeds c_j, $j=1,2,\ldots,N$, with

$$c_j=4k_j^2.$$

Proof: See the Appendix.

III. SOME LEMMAS

Before discussing the case $b(k)\not\equiv 0$, we state some lemmas.

Lemma 1: Set

$$\Omega(r,t)=\frac{1}{2\pi}\int_{-\infty}^{\infty} b(k)e^{8ik^3t+ikr}\,dk,$$

where $(r,t)\in(-\infty,\infty)\times[0,\infty)$, $R(k)\in C^1(-\infty,\infty)$, $|R(k)|\leqslant M_1$, $|R'(k)|\leqslant M_2$, $-\infty<k<\infty$, and $R(k)=O(|k|^{-2})$, as $k\to\infty$. Then we have, for any $r\geqslant r_0$ uniformly,

$$\Omega(r,t)\to 0,\ \text{as } t\to\infty. \tag{8}$$

Proof: Setting $k'=kt^{1/3}$ (for brevity we still write k' as k), it follows that

$$\Omega(r,t)=\frac{1}{2\pi t^{1/3}}\int_{-\infty}^{\infty} R\left(\frac{k}{t^{1/3}}\right)e^{i(8k^3+kr/t^{1/3})}\,dk. \tag{9}$$

Taking an apropose large positive constant M, we divide the integral into three parts:

$$\Omega(r,t)=\frac{1}{2\pi t^{1/3}}\left\{\int_{-M}^{M}+\int_{M}^{\infty}+\int_{-\infty}^{-M}\right\}R\left(\frac{k}{t^{1/3}}\right)\times e^{i(8k^3+kr/t^{1/3})}\,dk$$
$$\equiv A+B+C. \tag{10}$$

We estimate these integrals respectively. First,

$$|A|\leqslant\frac{1}{2\pi t^{1/3}}\left|\int_{-M}^{M}R\left(\frac{k}{t^{1/3}}\right)e^{i(8k^3+kr/t^{1/3})}\,dk\right|$$
$$\leqslant\frac{1}{2\pi t^{1/3}}\int_{-M}^{M}\left|R\left(\frac{k}{t^{1/3}}\right)\right|dk\leqslant\frac{\bar{M}}{2\pi t^{1/3}}, \tag{11}$$

where $\bar{M}=MM_1$. Next,

$$B=\frac{1}{2\pi t^{1/3}}\int_{M}^{\infty}R\left(\frac{k}{t^{1/3}}\right)e^{i(8k^3+kr/t^{1/3})}\,dk$$
$$=\frac{1}{2\pi t^{1/3}}\left\{-R\left(\frac{M}{t^{1/3}}\right)e^{i(8M^3+Mr/t^{1/3})/i(24M^2+r/t^{1/3})}\right.$$
$$+\int_{M}^{\infty}48kR\left(\frac{k}{t^{1/3}}\right)e^{i(8k^3+kr/t^{1/3})/i(24k^2+r/t^{1/3})^2}\,dk$$
$$\left.-\int_{M}^{\infty}R'\left(\frac{k}{t^{1/3}}\right)e^{i(8k^3+kr/t^{1/3})/i(24k^2r/t^{1/3})}\,dk\right\}.$$

Thus it follows that

$$|B|\leqslant\frac{1}{2\pi t^{1/3}}\left\{\left|\frac{R(M/t^{1/3})}{24M^2+r/t^{1/3}}\right|\right.$$
$$+\int_{M}^{\infty}\frac{48k\,|R(k/t^{1/3})|}{(24k^2+r/t^{1/3})^2\,dk}$$
$$\left.+\int_{M}^{\infty}\frac{|R'(k/t^{1/3})|}{(24k^2+r/t^{1/3})dk}\right\}.$$

When $r\geqslant r_0$, for an apropose large t we have $24k^2+r/t^{1/3}\geqslant 24k^2-1$. Hence it follows that

$$|B|\leqslant\frac{1}{2\pi t^{1/3}}\left\{\frac{M_1}{24M^2-1}+\int_{M}^{\infty}\frac{48kM_1}{(24k^2-1)^2\,dk}\right.$$
$$\left.+\int_{M}^{\infty}\frac{M_2t^{-1/3}}{(24k^2-1)dk}\right\}$$
$$\leqslant\frac{1}{2\pi t^{1/3}}\left\{\frac{M_1}{24M^2-1}\right.$$
$$\left.+\int_{M}^{\infty}\frac{48kM_1}{(24k^2-1)^2\,dk}+\int_{M}^{\infty}\frac{M_2t^{-1/3}}{(24k^2-1)dk}\right\}$$
$$\leqslant\frac{1}{2\pi t^{1/3}}\left\{\frac{2M_1}{24M^2-1}+\ln\left(\frac{\sqrt{24}M+1}{24M-1}\right)\frac{M_2}{2\sqrt{24}}\right\}. \tag{12}$$

Similarly, we get

$$|C|\leqslant\frac{1}{2\pi t^{1/3}}\left\{\frac{2M_1}{24M^2-1}+\ln\left(\frac{24M+1}{24M-1}\right)\frac{M_2}{2\sqrt{24}}\right\}. \tag{13}$$

Combining (11)–(13), eventually we have

$$|\Omega_{(r,t)}|\leqslant c_0/t^{1/3},\quad c_0=\text{const.}$$

Therefore, for $r\geqslant r_0$ uniformly,

$$|\Omega(r,t)|\to 0,\ \text{as } t\to\infty.$$

The lemma is proved.

Corollary: Suppose that $k_0 \neq 0$ is real and that n is an arbitrary positive integer; then we have, for $r \geqslant r_0$ uniformly,

$$\Omega^{(n)}(r,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} R(k)\frac{e^{8ik^3t+ikr}}{(k_0-ik)^n}dk \to 0, \quad \text{as } t\to\infty. \quad (14)$$

Lemma 2: If the function $\Omega(r,t)$ is defined as Lemma 1, where $R(k)\in C^2(-\infty,\infty)$ and $R'(k)$ and $R''(k)$ are summable, then

$$\int_0^{\infty}\Omega(r,t)dr \to 0, \quad \text{as } t\to\infty.$$

Proof: Since $|R'(k)|\leqslant C_1$, $|R''(k)|\leqslant C_2$, $-\infty < k < \infty$. Picking an appropriately large positive number M, we divide the integral into two parts:

$$\int_0^{\infty}\Omega(r,t)dr = \left(\int_0^M + \int_M^{\infty}\right)dr\,\Omega(r,t) \equiv A + B.$$

From Lemma 1 it is seen that for an appropriately large t there is a constant C such that

$$|\Omega(r,t)| \leqslant C/t^{1/3}.$$

Thus

$$|A| \leqslant \int_0^{\infty}|\Omega(r,t)|dr \leqslant \frac{CM}{t^{1/3}}. \quad (15)$$

After twice integrating by parts for B, we obtain

$$\begin{aligned} B = &\int_M^{\infty}dr\frac{1}{2\pi}\int_{-\infty}^{\infty}dk(-1)R''(k)\frac{e^{i(8k^3t+kr)}}{(24k^2t+r)^2} \\ &+\int_M^{\infty}dr\frac{1}{2\pi}\int_{-\infty}^{\infty}dk\,3R'(k)48kt\frac{e^{i(8k^3t+kr)}}{(24k^2t+r)^3} \\ &+\int_M^{\infty}dr\frac{1}{2\pi}\int_{-\infty}^{\infty}dk\,R(k)48t\frac{e^{i(8k^3t+kr)}}{(24k^2t+r)^3} \\ &+\int_M^{\infty}dr\frac{1}{2\pi}\int_{-\infty}^{\infty}dk(-3)R(k)(48kt)^2 \\ &\times\frac{e^{i(8k^3t+kr)}}{(24k^2t+r)^4} \\ \equiv &B_1+B_2+B_3+B_4. \end{aligned}$$

Then these integrals are estimated, respectively, as follows:

$$\begin{aligned} |B_1| &\leqslant \frac{1}{2\pi}\int_{-\infty}^{\infty}dk\,|R''(k)|\int_M^{\infty}dr\frac{1}{(24k^2t+r)^2} \\ &\leqslant \frac{C_2}{2\pi}\int_{-\infty}^{\infty}dk\frac{1}{(24k^2t+M)} \leqslant \frac{C_2}{2(24t)^{1/2}}. \end{aligned} \quad (16)$$

Because the estimates of B_2, B_3, and B_4 are similar, we estimate only B_2. As in Lemma 1, taking constants $M_1<0$, $M_2>0$, we divide the integral into three sections

$$\begin{aligned} B_2 = &\int_M^{\infty}dr\frac{1}{2\pi}\left(\int_{M_1}^{M_2}+\int_{M_2}^{\infty}+\int_{-\infty}^{M_1}\right)dk\,3R'(k)48kt \\ &\times\frac{e^{i(8k^3t+kr)}}{(24k^2t+r)^3} \equiv B_{2,1}+B_{2,2}+B_{2,3}. \end{aligned}$$

Since the estimation of $B_{2,1}$ is easy, we consider only $B_{2,2}$ and $B_{2,3}$. It follows immediately that

$$\begin{aligned} |B_{2,2}| &\leqslant \int_M^{\infty}dr\frac{1}{2\pi}\int_{M_2}^{\infty}dk\frac{3|R(k)|48t}{(24k^2t+r)^3} \\ &\leqslant \frac{1}{2\pi}\int_{M_2}^{\infty}dk\,3|R'(k)|48t\int_{M_2}^{\infty}dr\frac{1}{(24k^2t+r)^3} \\ &\leqslant \frac{36C_1}{\pi}\int_{M_2}^{\infty}dk\frac{48kt}{(24k^2t+M)^2} \leqslant \frac{36C_1}{24M_2^2t+M}. \end{aligned} \quad (17)$$

Similarly,

$$\begin{aligned} |B_{2,3}| &\leqslant \frac{1}{2\pi}\int_{-\infty}^{M_1}dk\,3|R'(k)|48kt\int_M^{\infty}dr\frac{1}{(24k^2t+r)^3} \\ &\leqslant \frac{36C_1}{\pi}\int_{-\infty}^{M_1}dk\frac{48kt}{(24k^2t+M)^2} \leqslant \frac{36C_1}{24M_1^2t+M}. \end{aligned} \quad (18)$$

Combining (15)–(18), we obtain

$$\int_0^{\infty}\Omega(r,t)dr \to 0, \text{ as } t\to\infty.$$

The lemma is proved.

It is not hard to prove the following corollaries.

Corollary 1: Suppose $\Omega^{(n)}(r,t)$ is defined as (14); then, at $t\to\infty$,

$$\int_0^{\infty}\Omega^{(n)}(r,t)dr \to 0.$$

Corollary 2: Suppose $\Omega^{(n)}(r,t)$ is defined as (14). Then, for $r\geqslant r_0$ uniformly,

$$\begin{aligned} &\int_0^{\infty}\cdots\int_0^{\infty}\Omega^{(j_1)}(r+r_1,t)\Omega^{(j_2)}(r_1+r_2,t)\cdots\Omega^{(j_n)} \\ &\quad\times(r_n+r,t)\,dr_1dr_2\cdots dr_n \to 0, \quad \text{as } t\to\infty. \end{aligned}$$

where $j_1, j_2, \ldots, j_n$ are any positive integers.

Corollary 3: Let $f(r)$ be a bounded function in $[0,\infty)$; then

$$\int_0^{\infty}drf(r)\Omega(r,t) \to 0, \quad \text{as } t\to\infty.$$

IV. EIGENSPEED OF THE GENERAL SOLUTION

We turn now to the proof of Theorem 1. Let $C[0,\infty)$ be a normed space consisting of the bounded continous function with the norm

$$\|f\| = \sup_{y>0}|f(y)|.$$

It is not hard to verify that this space is a B space. We defined the linear parameter operators $\mathbb{C}_{x,t}, \mathbb{D}_{x,t}$ in $C[0,\infty)$:

$$(\mathbb{C}_{x,t}f)(y) = \int_0^{\infty}B_c(x+y+z,t)f(z)dz,$$

$$(\mathbb{D}_{x,t}f)(y) = \int_0^{\infty}2B_d(x+y+z,t)f(z)dz,$$

where $b(k)$ satisfies condition (7). In view of the corollary of Lemma 2, it is readily known that $\mathbb{C}_{x,t}$ and $\mathbb{D}_{x,t}$ are really linear operators.

For these operators, we establish the theorem of Neumann expansion. For this purpose, first of all, we describe some lemmas.

Lemma 3:

$\|\mathbb{C}_{x,t}\| \to 0, \quad -\infty < x < \infty, \quad \text{as } t \to \infty.$

Proof: By definition, we have

$$\|\mathbb{C}_{x,t}\| = \sup_{\|f\|=1} \sup_{y>0} \left| \int_0^\infty \frac{1}{\pi} \int_{-\infty}^\infty R(k) e^{8ik^3 t + ik(x+y+z)} \times dk f(z) dz \right|.$$

According to Corollary 3 of lemma 2, it follows that

$$\sup|\cdot| \to 0, \quad \text{as } t \to \infty.$$

Thus

$$\|\mathbb{C}_{x,t}\| \to 0, \quad \text{as } t \to \infty.$$

The lemma is proved.

Lemma 4: $(I + \mathbb{D}_{x,t})^{-1}$ exists for $-\infty < x < \infty$, $t \geqslant 0$, and fixed x_0:

$$\|(I + \mathbb{D}_{x,t}) - 1\| \leqslant M, \quad \text{if } x \geqslant x_0,$$

where M is constant.

Proof: By definition, we have

$$\|(I + \mathbb{D}_{x,t})^{-1}\| = \sup_{\|f\|=1} \|(I + \mathbb{D}_{x,t})^{-1} f\|. \tag{19}$$

Setting

$$(I + \mathbb{D}_{x,t})^{-1} f = K(x,y,t) \equiv \overline{K}_{x,t}(y), \tag{20}$$

$$K(x,y,t) = \sum_{j=1}^N e^{-2k_j y} C^j(x,t) + f(y), \tag{21}$$

it follows from (20) and (21) that

$$C^j(x,t) + 2c_j^2 e^{8k_j^3 t} e^{-2k_j x} \sum_{l=1}^N \frac{C^l(x,t)}{2(k_j + k_l)} = -\int_0^\infty 2c_j^2 e^{8k_j^3 t} e^{-2k_j(x+z)} f(z) dz \equiv R_j(x,t), \quad j = 1,2,\ldots,N.$$

Thus

$$C^l(x,t) = (\det A_l)/(\det A), \tag{22}$$

where A is the coefficient matrix of equation, and A_l is a matrix formed by A substituted in the right-hand side for the lth column. Substituting (22) into (21), it easily follows from (18) that

$$\|(I + \mathbb{D}_{x,t})^{-1}\| = \sup_{\|f\|=1} \sup_{y>0} \left| \sum_{i=1}^N \frac{e^{-2k_i y} \det A_i}{\det A} + f(y) \right| \leqslant M, \quad \text{if } x \geqslant x_0,$$

where M is constant. The lemma is proved.

From Lemmas 3 and 4, we may conclude the following lemma.

Lemma 5: For $x \geqslant x_0$, there is a t_0, and when $t \geqslant t_0$, we have

$$\|(I + \mathbb{D}_{x,t})^{-1} \mathbb{C}_{x,t}\| < 1.$$

Theorem 3: In $C[0,\infty)$, the inverse of the operator $(I + \mathbb{D}_{x,t})^{-1}\mathbb{C}_{x,t})$ exists and

$$[I + (I + \mathbb{D}_{x,t})^{-1}\mathbb{C}_{x,t}]^{-1} = \sum_{n=0}^\infty (-1)^n [(I + \mathbb{D}_{x,t})^{-1}\mathbb{C}_{x,t}]^n. \tag{23}$$

Proof: From Lemma 4, it is known that $(I + \mathbb{D}_{x,t})^{-1}$ is a bounded linear operator in $C[0,\infty)$ and $(I + \mathbb{D}_{x,t})^{-1}\mathbb{C}_{x,t}$ also is true. Hence it follows at once from Lemma 5 that (22) holds under the convergence of operatoral norm.[13] The theorem is proved.

Putting

$$K(x,y,t) = K_{x,t}(y),$$

$$-B_c(x+y,t) = Q_{x,t}(y), \quad -B_d(x+y,t) = W_{x,t}(y)/2,$$

it is known from Lemma 1 that $W_{x,t}(y)$, $Q_{x,t}(y)$ belong to $C[0,\infty)$. Thus the GLM equation [Eq. (5)] is the operational equation in $C[0,\infty)$, which may be written as

$$(I + \mathbb{D}_{x,t} + C_{x,t}) K_{x,t} = W_{x,t} + Q_{x,t}. \tag{24}$$

Theorem 4: Let the scattering coefficient $b(k)$ satisfy (7); then there exists a unique solution in $C[0,\infty)$ satisfying the GLM equation, and

$$K_{x,t}(y) = \sum_{n=0}^\infty (-1)^n [(I + \mathbb{D}_{x,t})^{-1}\mathbb{C}_{x,t}]^n \times [K_d(x,t) + (I + \mathbb{D}_{x,t})^{-1} Q_{x,t}]. \tag{25}$$

Proof: Acting the operator $(I + \mathbb{D}_{x,t})^{-1}$ on Eq. (24), from Theorem 2 it follows that

$$(I + (I + \mathbb{D}_{x,t})^{-1}\mathbb{C}_{x,t}) K_{x,t} = (I + \mathbb{D}_{x,t})^{-1}(W_{x,t} + Q_{x,t}),$$

$$K_{x,t}(y) = (I + (I + \mathbb{D}_{x,t})^{-1}\mathbb{C}_{x,t})^{-1}(I + \mathbb{D}_{x,t})^{-1}(W_{x,t} + Q_{x,t}).$$

Furthermore, considering Theorem 3, it is known at once that $K(x,y,t) = K_{x,t}(y) \in C[0,\infty)$ exists and (25) holds. The theorem is proved.

We may obtain the explicit expression of $K_{x,t}(y)$ using the method of Theorem 2 and Lemma 4. To do this, we define

$$\Omega_n^{(j)}(x+y,t) = \frac{1}{2\pi} \int_{-\infty}^\infty b(k) \frac{e^{8ik^3 t + ik(x+y)}}{(k_j - ik)^n\, dk},$$

$$\Omega_{n,m}^{j,l}(x+y,t) = \frac{1}{2\pi} \int_{-\infty}^\infty b(k) \frac{e^{8ik^3 t + ik(x+y)}}{(k_j - ik)^n (k_l - ik)^m\, dk}.$$

Likewise, A expresses the coefficient matrix as in Lemma 2. Then

$$K(x,y,t) = \sum_{j=1}^N e^{-2k_j y}\{a_1^{(j)}(x,t) - a_2^{(j)}(x,t) + a_3^{(j)}(x,t) + \cdots\} + \sum_{j=1}^N \Omega_1^{(j)}(x+y,t) e^{-2k_j x}\{a_1^{(j)}(x,t) - a_2^{(j)}(x,t) + \cdots\} + \sum_{j=1}^N \int_0^\infty \Omega_0^{(j)}(y+z,t)\Omega_1^{(j)}(z+x,t) dz\, e^{-2k_j x}\{a_1^{(j)}(x,t) - a_2^{(j)}(x,t) + \cdots\}, \tag{26}$$

where

$$a_n^{(j)}(x,t) = \det A_{i,n}/\det A.$$

Here $A_{i,n}$ is the matrix formed by matrix A substituting the column

$$n=1, \quad -c_j^2 e^{8k_j^3 t} e^{-2k_j x},$$

$$n=2, \quad -c_j^2 e^{8k_j^3 t} e^{-2k_j x} \Omega_1^{(j)}(2x,t),$$

$$n=3, \quad -c_j^2 e^{8k_j^3 t} e^{-2k_j x} \sum_{l=1}^{N} \Omega_{1,1}^{(l,j)}(2x,t) a_1^{(l)}(x,t),$$

$$n=4, \quad -c_j^2 e^{8k_j^3 t} e^{-2k_j x}\left[\sum_{l=1}^{N} a_2^{(l)}(x,t) e^{-2k_l x} \Omega_{1,1}^{(l,j)}(2x,t) + \int_0^{\infty} \Omega_1^{(j)}(z+x,t)\Omega_0^{(j)}(z+x,t)dz\right],$$

$$j=1,2,\ldots,N,$$

into the l th column. Differentiating the GLM equation with respect to x and then using an abstract version,

$$(I+\mathbb{D}_{x,t}+\mathbb{C}_{x,t})R_{x,t} = S_{x,t}+T_{x,t}, \tag{27}$$

where

$$R_{x,t}(y) = \frac{\partial}{\partial x}K(x,y,t),$$

$$S_{x,t}(y) = -\frac{\partial}{\partial x}B(x+y,t),$$

$$T_{x,t}(y) = -\int_0^{\infty}\left(\frac{\partial}{\partial x}B(x+y+z,t)\right)K(x,z,t)dz.$$

We obtain from Lemma 2 and Theorem 4 that $S_{x,t}(y)$ and $T_{x,t}(y)$ belong to $C[0.\infty)$.

By analogy with Theorem 4, the following theorem holds.

Theorem 5: If scattering coefficients satisfy (7), then there exists a unique solution in $C[0,\infty)$ satisfying Eq. (27), and

$$\frac{\partial}{\partial x}K(x,y,t) = \sum_{n=0}^{\infty}(-1)^n[(I+\mathbb{D}_{x,t})^{-1}\mathbb{C}_{x,t}]^n \frac{\partial}{\partial x}K(x,\ldots,t) + (I+\mathbb{D}_{x,t})^{-1}(S_{x,t}+T_{x,t}). \tag{28}$$

Theorem 6: Suppose that $K(x,y,t)$ is the solution of the GLM equation under scattering coefficients satisfying (7) and that $k_d(x,y,t)$ is the solution where $b(k)\equiv 0$. Then, for $x\geqslant x_0$, $y\geqslant 0$ uniformly, as $t\to\infty$,

$$K(x,y,t)\to K_d(x,y,t), \quad \frac{\partial}{\partial x}K(x,y,t)\to\frac{\partial}{\partial x}K_d(x,y,t).$$

Proof: Because $b(k)$ satisfies (7), it follows from Theorem 4 that (26) holds. Moreover, one obtains, by rearranging these terms,

$$K^{(n)}(x,y,t) = K_d(x,y,t) + \{\cdots\} + K_c^{(n)}(x,y,t),$$

where $K^{(n)}(x,y,t)$ is the sum of the first n terms of a series solution which converges uniformly to $K(x,y,t)$, as $n\to\infty$, and $K_c^{(n)}(x,y,t)$ is the same when $B_d(x+y,t)\equiv 0$. From Lemmas 1 and 2 and their corollaries, it is not hard to verify that $K_c^{(n)}(x,y,t)$ and $\{\cdots\}$ converge uniformly to zero, as $t\to\infty$, for $x\geqslant x_0$, $y\geqslant 0$. Hence, for $x\geqslant x_0$, $y\geqslant 0$ uniformly,

$$\lim_{t\to\infty}[K(x,y,t)-K_d(x,y,t)] = \lim_{t\to\infty}\lim_{n\to\infty}[K^{(n)}(x,y,t)-K_d(x,y,t)] = \lim_{n\to\infty}\lim_{t\to\infty}[K^{(n)}(x,y,t)-K_d(x,y,t)] = 0.$$

Similarly we have, for $x\geqslant x_0$, $y\geqslant 0$ uniformly,

$$\frac{\partial}{\partial x}K(x,y,t)\to\frac{\partial}{\partial x}K_d(x,y,t), \quad \text{as } t\to\infty.$$

The theorem is proved.

Theorem 7: Suppose that $u(x,t)$ and $u_d(x,t)$ are the solution and N-soliton solution of Eq. (1) under the initial data $u_0(x)$ satisfying (3) and (4), respectively. Then we have, for $x\geqslant x_0$ uniformly,

$$u(x,t)\to u_d(x,t), \quad \text{as } t\to\infty.$$

Proof: Owing to the fact that when the initial data satisfy (3) and (4), scattering coefficients will satisfy (7), we obtain

$$u(x,t) = -\frac{\partial}{\partial x}K(x,0,t),$$

$$u_d(x,t) = -\frac{\partial}{\partial x}K_d(x,0,t).$$

When $t\to-\infty$, we may establish Theorem 7 similarly to Theorem 6. The theorem is proved.

Combining Theorem 2 and Theorem 7, we may directly obtain the proof of Theorem 1. In fact, the conclusion of Theorem 1 holds for $u_d(x,t)$, and, on the other hand, since $u(x,t)\to u_d(x,t)$, as $t\to\pm\infty$, for $x\geqslant x_0$ uniformly, therefore

$$\lim_{t\to\pm\infty}u(x+ct,t) = \lim_{t\to\pm\infty}u_d(x+ct,t) = \begin{cases} S(x-\theta_j^{\pm},c_j), & \text{if } c=c_j, \\ 0, & \text{if } c\neq c_j. \end{cases}$$

Thus Theorem 1 is proved.

APPENDIX: THE PROOF OF THEOREM 2

Proof: First, we give the expression of the N-soliton solution. The GLM equation is as follows:

$$K_d(x,y,t) + \int_0^{\infty} 2B_d(x+y+z,t)K_d(x,z,t)dz + 2B_d(x+y,t) = 0. \tag{A1}$$

Set

$$K_d(x,y,t) = \sum_{i=1}^{N} 2h_i(x,t)e^{-k_i(x+y)}. \tag{A2}$$

Substituting (A2) into (A1) we get

$$h_i(x,t) + \sum_{n=1}^{N} h_n(x,t)\frac{c_n^2}{k_n+k_i}e^{8k_i^3 t}e^{-(k_n+k_i)x} = -c_i^2 e^{8k_i^3 t - k_i x}, \quad i=1,2,\ldots,N.$$

Let A be the coefficient matrix of the above equation,

$$A_{ij} = \delta_{ij} + [c_i^2/(k_i + k_j)]e^{8k_j^3 t} e^{-(k_i + k_j)x},$$

and let A_k be the matrix formed by A, the k th column of which is the derivative of the k th column of matrix A with respect to x. Then

$$h_i(x) = (\det A_i)e^{2k_i x}/\det A,$$

so that

$$K_d(x,y,t) = \sum_{i=1}^{N} 2\frac{\det A_i}{\det A} e^{k_i(x-y)}.$$

According to the method of Hirota,[14] $u_d(x,t)$ may be represented as

$$u_d(x,t) = -2\ln(f)_{xx} = -2[f_{xx}/f - (f_x/f)^2],$$

where

$$f = \sum_{\mu=0,1} \exp\left(\sum_{1\le i<j\le N} B_{ij}\mu_i\mu_j + \sum_{i=1}^{N}\mu_i X_i\right),$$

here $\Sigma_{\mu=0,1}$ implies the summation over all possible combinations of $\mu_1 = 0,1,\ldots,\mu_N = 0.1$,

$$X_i = 2k_i x - \beta_i k_i t + r_i, \quad \beta_i = 8k_i^2,$$

$$r_i = \ln(c_i^2/2k_i), \quad e^{B_{ij}} = ((k_i - k_j)/(k_i + k_j))^2.$$

To obtain the exponding form of f, we adopt the symbol $C_N^{(pj)}(X,B)$, which expresses, taking any p numbers $X_{i_1}, X_{i_2},\ldots,X_{i_p}$ of $X_1, X_2,\ldots,X_N$, the summation of all terms:

$$[2(k_{i_1} + k_{i_2} + \cdots + k_{i_p})]^j \exp\left\{\sum_{1\le m<l\le p} B_{i_m i_l} + \sum_{m=1}^{P} X_m\right\}.$$

Then

$$f(X,t) = 1 + \sum_{i=1}^{N}\exp(X_i) + C_N^{(2,0)}(X,B) + \cdots + C_N^{(N,O)}(X,B),$$

$$f_x(X,t) = \sum_{i=1}^{N} 2k_i\exp(X_i) + C_N^{(2,1)}(X,B) + \cdots + C_N^{(N,1)}(X,B),$$

$$f_{xx}(X,t) = \sum_{i=1}^{N}(2k_i)^2\exp(X_i) + C_N^{(2,2)}(X,B) + \cdots + C_N^{(N,2)}(X,B).$$

We consider now the limitation of $u_d(x + ct,t)$, when $t\to\infty$. It is not hard to verify that, as $t\to\infty$,

$$\exp(2k_j(x + ct) - \beta_j k_j + r_j) \to \begin{cases} 0, & c < 4k_j^2, \\ \exp(2k_j x + r_j), & c = 4k_j^2, \\ \infty, & d > 4k_j^2. \end{cases}$$

We use the sign $\overline{X}_j$ to express X_j in which x is replaced by $x + ct$.

(1) When $4k_{J-1}^2 < c < 4k_J^2$, $J = 2,3,\ldots,N$,

$$\exp(\overline{X}_j) \to \begin{cases} \infty, & j < J-1, \\ 0, & j \ge J, \end{cases} \quad \text{as } t\to\infty.$$

In this case, as $t\to\infty$, the terms containing the factors $\exp(\overline{X}_1 + \overline{X}_2 + \cdots + \overline{X}_{J-1})$ dominate within all terms. Hence, as $t\to\infty$,

$$u_d(x + ct,t)\to 0.$$

(2) When $c = 4k_J^2$, $J = 2,3,\ldots,N$, as $t\to\infty$,

$$\exp(\overline{X}_j) \to \begin{cases} \infty, & j \ll J-1, \\ \exp(2k_j x + r_j), & j = J, \\ 0, & j \gg J+1. \end{cases}$$

In this case, the terms containing the factors $\exp(\Sigma_{j=1}^{J-1}\overline{X}_j)$ or $\exp(\Sigma_{j=1}^{J}\overline{X}_j)$ dominate within all terms. Hence, as $t\to\infty$,

$$u_d(x + ct,t) \to -8k_J^2\exp\left(\sum_{1\le i<j\le J-1} B_{ij}\right)\exp\left(\sum_{1\le i<j\le J} B_{ij} + 2k_J x + r_J\right)\left[\exp\left(\sum_{1\le i<j\le J-1} B_{ij}\right) + \exp\left(\sum_{1\le i<j\le J} B_{ij} + 2k_J x + r_J\right)\right]^{-2} \to \text{soliton}.$$

(3) When $4k_J^2 < c$, $u_d(x + ct,t)\to 0$. as $t\to\infty$. Thus the proof is the same as in (1).

(4) When $c < 4k_1^2$, we have, as $t\to\infty$,

$$\exp(\overline{X}_j)\to 0, \quad j = 1,2,\ldots,N.$$

Therefore, as $t\to\infty$,

$$u_d(x + ct,t)\to 0.$$

(5) When $c = 4k_1^2$, we have, as $t\to\infty$,

$$u_d(x + ct) \to -2(2k_1)^2\exp(2k_1 x + r_1) \times [1 + \exp(2k_1 x + r_1)]^{-2} \to \text{soliton}.$$

When $t\to -\infty$, the proof is similar. Thus Theorem 2 is proved.

[1] N. J. Zabusky, in *Mathematical Models in Physical Science*, edited by S. Drobot (Prentice–Hall, Englewood Cliffs, NJ, 1963).
[2] N. J. Zabusky and M. D. Kruskal, Phys. Rev. Lett. **15**, 240 (1965).
[3] N. J. Zabusky, *A Synergetic Approach to Problems of Dispersive Propagation and the Interaction Nonlinear Partial Differential Equation*, (Academic, New York, 1965).
[4] C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, Phys. Rev. Lett. **19**, 1095 (1967).
[5] P. D. Lax, Comm. Pure Appl. Math. **21**, 467 (1968).
[6] S. Tanaka, Publ. Res. Inst. Math. Sci. Kyoto Univ. **8**, 419 (1973).
[7] S. Tanaka, Publ. Res. Inst. Math. Sci. Kyoto Univ. **10**, 367 (1975).
[8] K. Z. Tu, Ying Yong Shuxue Yu Jisuan Shuxue **21** (1979).
[9] A. C. Scott and D. W. McLaughlin, Proc. IEEE **61** (1973).
[10] V. I. Kapman, Phys. Scr. **20**, 462 (1979).
[11] A. Cohen, Arch. Rat. Mech. Anal. **71**, 143 (1979).
[12] N. J. Zabusky, Phys. Rev. **168**, 124 (1968).
[13] A. E. Taylor, *Introduction to Functional Analysis* (Wiley, New York, 1958).
[14] H. Hirota, J. Math. Phys. **14**, 810 (1973).

Duals and propagators: A canonical formalism for nonlinear equations

D. G. Cacuci
Oak Ridge National Laboratory, Engineering Physics and Mathematics Division, Oak Ridge, Tennessee 37831

R. B. Perez
University of Tennessee, Department of Nuclear Engineering, Knoxville, Tennessee 37996

V. Protopopescu
Oak Ridge National Laboratory, Engineering Physics and Mathematics Division, Oak Ridge, Tennessee 37831

(Received 20 March 1987; accepted for publication 31 August 1987)

A new formalism for solving general nonlinear equations is proposed. Given a nonlinear operator equation, a *dual* operator is constructed, canonically, similar to the adjoint operator in linear theory. By using this dual operator, advanced and retarded propagators (analogs of the Green's functions in linear theory) are generated. These propagators satisfy the customary reciprocity and semigroup properties and yield the formal solution of the original nonlinear problem. It is further shown that these propagators can be obtained either by solving a linear equation that still contains an implicit dependence on the problem's solution or by solving a closed-form nonlinear integral equation from which the problem's solution itself is completely eliminated. The formalism is canonical in the sense that its applicability is not affected by the particularities of the nonlinear operator, boundary conditions, and underlying phase space. All aspects of applying this formalism to nonlinear problems are illustrated analytically on the Riccati equation.

I. INTRODUCTION

Triggered by the pioneering work of Zabusky and Kruskal,[1] and Gardner, Greene, Kruskal, and Miura[2] on the Korteweg–de Vries equation, the study of integrable nonlinear partial differential equations has since been established as a new branch of mathematical physics. So far, the general effort in this branch has been focused on attempting to cast the nonlinear (evolution) equation under consideration as a nonlinear Hamiltonian system and, if successful, to make further progress by investigating special properties such as (complete) integrability, applicability of inverse scattering transforms, and Lax-pair representations. As amply documented in many articles and several treatises (see, e.g., Ref. 3 and references therein), the fundamental motivation underlying this effort has been the attempt to solve nonlinear equations by using the powerful and refined methods of the *theory for linear operators.*

Although the inverse scattering transform method has been successfully used to solve (or, at least, to test the integrability of) particular types of nonlinear evolution equations, as yet there is no canonical procedure to construct the inverse scattering problem (or, alternatively, a Lax pair[4]) for a general nonlinear system. A useful, but still not general, approach towards establishing such a procedure has been proposed by Chen, Lee, and Liu.[5] The connection between nonlinear partial differential equations solvable by inverse scattering transform techniques and nonlinear ordinary differential equations of Painlevé type (see, e.g., Ref. 3) has been recently extended to include a certain class of nonlocal integrodifferential equations.[6] Concepts of nonlocality have been also used, in other contexts, for transforming a local nonlinear problem into a nonlocal linear one.[7] Recently, a new general class of integrable nonlinear partial differential equations on an infinite three-dimensional (3-D) lattice has been found from the compatibility condition of a linear nonhomogeneous system of equations that is invariant under a certain nonlocal integral transformation.[8] Further information about a given nonlinear equation can be obtained by exploiting its "algebraic properties," i.e., the Lie group that leaves it invariant.[9,10]

It is important to note that several of the aforementioned methods use, rather extensively, various linearized (variational, tangential, or gradient) equations[4,5,10,11] associated with the nonlinear problem under consideration. Most of these investigations were aimed at discovering whether these nonlinear (evolution) equations possessed a Hamiltonian structure or not; occasionally, these investigations used variational equations for various conserved functionals of the field variables rather than variational equations obtained directly from the nonlinear equations underlying the problem. However, to our knowledge, none of these investigations used the integrated variational equations that will prove essential to the new formalism we develop in this work.

This new formalism (see also Ref. 12) is for solving physical problems represented by general nonlinear operator equations of any dimensionality, on a finite or infinite domain, and involving nonlinear boundary and/or initial conditions. The detailed mathematical description of this general nonlinear problem is presented in Sec. II. In Sec. III, we construct the *dual* operator, which is shown to play for the nonlinear problem described in Sec. II a role that is simi-

 0022-2488/88/020353-09$02.50

lar to the role the customary adjoint operator plays for linear problems. This dual operator is fundamental to introducing and constructing the advanced and retarded propagators for the nonlinear problem. We construct these propagators in Sec. IV, show that they generalize the Green's functions used for solving linear problems, and use them to derive the closed-form integral representation of the solution to the nonlinear problem described in Sec. II. In Sec. V, we introduce the closed-form integral equations satisfied by these propagators. Section VI presents functional-series expansions for both the propagators and the solution to the nonlinear problem; these series expansions appear as generalizations of the customary perturbation theory expansions. The usefulness of the formalism presented in Secs. II–VI is illustrated in Sec. VII by applying it to the Riccati equation—a nonlinear equation with a known analytical solution. Finally, the conclusions are summarized in Sec. VIII.

II. STATEMENT OF THE PROBLEM

Consider that the physical problem is represented by a general nonlinear equation of the form

$$N(u(x),\alpha(x),x)=0, \quad \text{in } \Omega, \tag{1}$$

where the following conditions apply.

(1) $x=(x_1,\ldots,x_n)\in\Omega\subset R^n$ is the phase-space position vector; Ω, henceforth referred to as the phase space of the problem, is an open domain with piecewise smooth boundary $\partial\Omega$. If the problem is time dependent, the set $(x_1,\ldots,x_n)$ is understood to include the time variable t among the independent variables.

(2) $u\in H_1$ is the state vector; for notational simplicity, u is considered in this work to be a scalar-valued function, although the derivations to follow remain in principle unchanged even if u is a vector-valued function; H_1 is a Hilbert space endowed with an inner product denoted by $\langle\ ,\ \rangle$.

(3) $\alpha(x)=(\alpha_1,\ldots,\alpha_I)$ is the vector of parameters appearing in the description of the physical problem.

(4) N: $D(N)\subset H_1\to H_2$ is a nonlinear operator acting on u, with domain $D(N)$ dense in H_1; H_2 is a Hilbert space endowed with an inner product denoted by $[\ ,\ ]$. The nonhomogeneous (source) terms in Eq. (1) are represented by $N(0,\alpha(x),x)\in H_2$; of course, if Eq. (1) is homogeneous, then $N(0,\alpha,x)=0$. In this work, N is considered to contain (combinations of) integral, multiplicative, and differential operators. Therefore a set of boundary/initial conditions is needed to specify $D(N)$. These conditions are denoted by

$$B(u,\alpha)=0, \quad \text{on } \partial\Omega, \tag{2}$$

where the components of the vector-valued operator B are partial differential operators with coefficients defined in the whole of Ω; nonhomogeneous boundary terms are represented by the nonzero components of $B(0,\alpha)$.

III. CONSTRUCTION OF THE DUAL OPERATOR $N^+(u)$

The crux of representing the solution of Eqs. (1) and (2) in an integral form is the construction of an operator $N^+(u)$ analogous to the adjoint operator in linear theory, and henceforth referred to as the operator *dual* to $N(u)$. It is expected that, unlike the adjoint in linear theory, this dual cannot have a "global" definition but must depend parametrically on u. Since the main purpose of this work is to develop the algorithmic structure of the formalism, we explicitly assume that the formal operations to follow make sense.

The operator $N^+(u)$ is required to satisfy the relationship

$$[N(u),v]=\langle u,N^+(u)v\rangle, \tag{3}$$

and to act linearly on any v in the domain

$$D(N^+(u))=\{v\in H_2 \mid \exists h\in H_1 \text{ with } \langle u,h\rangle=[N(u),v], \text{ for any } u\in D(N)\}. \tag{4}$$

As will be seen in the following, the operator $N^+(u)$ can be uniquely defined to satisfy Eqs. (3) and (4) for a large class of nonlinear operators $N(u)$. The construction of $N^+(u)$ starts from the fundamental theorem of calculus in nonlinear functional analysis; this theorem states[13] that

$$N(u_0+h)-N(u_0)=\int_0^1 \delta N(u_0+\epsilon h;h)\,d\epsilon, \tag{5}$$

where

$$\delta N(u_0;h)\equiv \left.\frac{dN(u_0+\epsilon h)}{d\epsilon}\right|_{\epsilon=0} \tag{6}$$

is the first Gâteaux variation of $N(u)$ at u_0 with increment h, for $u_0,h\in D(N)$.

Setting $u_0=0$ and $h=u$ in Eq. (5) and replacing the resulting equation for $N(u)$ in the left side of Eq. (3) gives

$$[N(u),v]=\left[N(0)+\int_0^1 \delta N(\epsilon u;u)\,d\epsilon,v\right]. \tag{7}$$

Comparing the right sides of Eqs. (7) and (3) shows that $N^+(u)$ can be constructed if and only if it is possible to extract a linear u dependence from the first term of the inner product appearing on the right side of Eq. (7). This is possible if and only if $N(u)$ admits a first Gâteaux derivative $N'(u_0)$ at $u_0\in D(N)$, in which case

$$\delta N(u_0;h)=N'(u_0)h, \tag{8}$$

where $N'(u_0)$ operates *linearly* on (the second argument) $h\in D(N)$. The *parametric* dependence of N' on u_0 is, in general, nonlinear.

In view of the linearity property expressed by Eq. (8), the right side of Eq. (7) can be written as an inner product of the form $\langle u,\cdot\rangle$ in H_1 by writing a Green's formula for the inner product $[N'(u)h,v]$, i.e.,

$$[N'(u)h,v]=\langle h,[N'(u)]^*v\rangle+\{\sigma(u)(h,v)\}, \tag{9}$$

where

$$v\in\{v\in H_2 \mid \exists w\in H_1 \text{ with } \langle h,w\rangle=[N'(u)h,v], \text{ for any } h\in D(N)\},$$

and $\{\sigma(u)(h,v)\}$ represents the bilinear form on the boundary spaces that arise from the customary integrations by

parts.[14] Note that $\sigma(u)$ depends, in general, nonlinearly on u.

The equations $N'(u)h=0$ and $[N'(u)]^*v=0$ are usually referred to as the direct and adjoint *variational* (gradient, tangential) equations associated with Eq. (1). Such variational equations have been used to some extent in previous works[3,4,10,11] to investigate the (possibly Hamiltonian) structure of the original nonlinear equation. Variational methods have also been used to generate closed equations for the one-particle Green's functions to obtain functionals of the solution in field and many-body[15] theories, and for sensitivity analysis[16] of general nonlinear systems. However, the variational equations themselves cannot be used to define a propagator similar to the propagators (Green's functions) for linear systems.

Replacing u by ϵu in Eq. (9) and integrating over ϵ from 0 to 1 gives

$$[L(u)h,v]=\langle h,L^*(u)v\rangle+\{\Sigma(u)(h,v)\}, \tag{10}$$

where the operators $L(u)$, $L^*(u)$, and $\Sigma(u)$ are defined as

$$L(u)\equiv\int_0^1 N'(\epsilon u)d\epsilon, \tag{11}$$

$$L^*(u)\equiv\int_0^1 [N'(\epsilon u)]^*d\epsilon, \tag{12}$$

and

$$\Sigma(u)\equiv\int_0^1 \sigma(\epsilon u)d\epsilon. \tag{13}$$

Just like the operators $N'(u)$, $[N'(u)]^*$, and $\sigma(u)$, the integrated operators $L(u)$, $L^*(u)$, and $\Sigma(u)$ still act linearly on the vector h although, in general, they depend nonlinearly on u. Note the important property

$$L(u)u=\int_0^1 N'(\epsilon u)u\,d\epsilon=N(u)-N(0), \tag{14}$$

which highlights the fact that, unlike the variational equations, the integrated operator $L(u)$ is related *directly* to the original nonlinear operator $N(u)$; this property is essential for the development of propagators for nonlinear systems, as will be shown in the following.

Equation (10) is valid for any $h\in H_1$ and $v\in H_2$; in particular, it is valid for $h\equiv u$. Replacing h in Eq. (10) by u, using Eq. (14), and comparing the resulting equation with Eq. (3) shows that the dual operator $N^+(u)$ can be uniquely represented as

$$N^+(u)=(u/\|u\|^2)\,\overline{N(0)}+L^*(u)+S(u). \tag{15}$$

Note that the right-hand side of (15) is singular for $u=0$, thus properly accounting for the nonhomogeneous (source) term $N(0)$. In Eq. (15), $S(u)$ is a symbolic notation for the operator that generates the boundary form $\{\Sigma(u)(u,v)\}$, and the overbar denotes complex conjugation. Of course, if u is a vector-valued function, the first term on the right side of Eq. (15) needs to be rewritten in terms of the transposed product of the column vector $N(0)$ with the row vector $u/\|u\|^2$. However, the formalism (although more complicated) remains canonical and leads to a unique expression for the dual $N^+(u)$. The uniqueness of the representation (15) is ensured by the requirement that Eq. (3) be satisfied in a functional sense, i.e., together with the functional derivative (9) considered in the limit $h\to u$.

In practice, the explicit expressions of $L(u)$ and $L^*(u)$ are always needed, but the explicit expression of the dual $N^+(u)$, i.e., Eq. (15), is not; the main purpose of writing Eq. (15) was to highlight the formal correspondence between the dual $N^+(u)$ and the customary adjoint operator encountered in linear functional analysis.

IV. PROPAGATORS FOR NONLINEAR OPERATOR EQUATIONS

A. The advanced propagator $G^*(u(x);x,x')$

Taking $L^*(u)$ as defined by Eq. (12), consider $G^*(u(x);x,x')$ to be the (unique) solution of the linear system

$$L^*(u(x))G^*(u(x);x,x')=\delta(x-x'),\quad \text{in } \Omega, \tag{16}$$

$$\gamma^*(G^*(u(x);x,x'))=0,\quad \text{on } \partial\Omega. \tag{17}$$

where the operator γ^*, acting linearly on G^*, represents boundary and/or initial conditions for G^* on $\partial\Omega$, as specified below.

Forming the inner products of Eqs. (1) and (16) with G^* and u, respectively, subtracting one from the other, and using the definition of the dual $N^+(u)v$ given in Eq. (15) shows that

$$\begin{aligned}&[N(u(x)),G^*(u(x);x,x')]\\&\quad-\langle u(x),L^*(u(x))G^*(u(x);x,x')\rangle\\&\quad=0-u(x')=\langle u(x),(u/\|u\|^2)\,\overline{N(0)}G^*(u(x);x,x')\rangle\\&\qquad+\{\Sigma(u(x))(u(x),G^*(u(x);x,x'))\}.\end{aligned}$$

The initial/boundary conditions represented by $\gamma^*(G^*)=0$ in Eq. (17) are now chosen to attain, optimally, the following objectives: (a) to ensure that Eqs. (16) and (17) constitute a well-posed problem for G^*; (b) to simplify the procedure for solving Eqs. (16) and (17) to obtain $G^*(u(x);x,x')$; and (c) to simplify the expression of the residual terms on $\partial\Omega$ after using (i) the (possibly nonlinear) initial/boundary conditions $B(u)=0$ for $u(x)$ given in Eq. (2), and (ii) the boundary/initial conditions for $G^*(u(x);x,x')$ given by Eq. (17). Note that these residual terms, henceforth denoted by $\{\Sigma(u)(u,G^*)\}_{B(u)=0,\ \gamma^*(G^*)=0}$, remain linear in G^* after using Eqs. (2) and (17) in Eq. (18). Although the selection of the initial/boundary conditions $\gamma^*(G^*)=0$ is not unique in general, the physical problem modeled by the original nonlinear system [i.e., Eqs. (1) and (2)] often provides clear guidelines for selecting $\gamma^*(G^*)=0$.

Following the above procedure to select the boundary/initial conditions $\gamma^*(G^*)=0$, and using these conditions together with Eq. (2) in the previous equation yields the solution $u(x)$ for the original nonlinear system as follows:

$$\begin{aligned}u(x')=&-\langle u(x),(u/\|u\|^2)\,\overline{N(0)}G^*(u(x);x,x')\rangle\\&-\{\Sigma(u(x))(u(x),\\&G^*(u(x),x,x'))\}_{B(u)=0,\ \gamma^*(G^*)=0}.\end{aligned} \tag{18}$$

Equation (18) highlights the role of $G^*(u(x);x,x')$ as an "advanced propagator" for the solution $u(x)$, since G^* acts

to propagate the source term $N(0)$ and the initial/boundary conditions for u, from $u(x)$ to $u(x')$, just as the customary propagator acts for the linear problems in field theory. Note, though, that—in contradistinction to the case of these linear problems—the advanced propagator $G^*(u(x);x,x')$ for the nonlinear system represented by Eqs. (1) and (2) depends implicitly and, in general, nonlinearly on the solution $u(x)$. We notice that when Eq. (1) is a true initial value problem (i.e., the volume and boundary sources reduce to initial conditions only), Eq. (18) reduces to

$$u(x',t') = \int u(x,0)G^*(u(x,0);x,0,x',t')dx.$$

On this form, the semigroup property of G^* is readily available, namely,

$$G^*(u(x,0);x,0,x'',t'') = \int dx'\, G^*(u(x,0);x,0,x',t')G^*(u(x',t');x',t',x'',t''). \tag{19}$$

B. The retarded propagator $G(u(x);x,x')$

Consider next the solution $G(u(x);x,x')$ of the linear system

$$L(u(x))G(u(x);x,x') = \delta(x-x'), \quad \text{in } \Omega, \tag{20}$$

$$\gamma(G(u(x);x,x')) = 0, \quad \text{on } \partial\Omega, \tag{21}$$

where the operator $\gamma(G)$, linear in G, represents initial and/or boundary conditions for G on $\partial\Omega$ chosen such that the replacement of $\gamma(G)=0$ and $\gamma^*(G^*)=0$ in the bilinear form $\{\Sigma(u)(G,G^*)\}_{\gamma(G)=0,\ \gamma^*(G^*)=0}$ causes the resulting (surface) integral over $\partial\Omega$ to vanish.

Forming the inner products of Eq. (16) with $G(u(x),x,x'')$ [or, alternatively, of Eq. (20) with $G^*(u(x),x,x'')$] and using Eq. (10) together with Eqs. (17) and (21) leads to

$$G^*(u(x);x,x') = G(u(x');x',x). \tag{22}$$

Equation (22) is the reciprocity relationship between the advanced and the retarded propagators G^* and G, respectively. Although G and G^* depend on u, this reciprocity relationship is identical to the reciprocity relationship satisfied by the advanced and retarded propagators encountered in linear problems. This is not surprising since by construction G and G^* satisfy *linear* equations. Furthermore, it follows from Eqs. (22) and (18) that the solution $u(x)$ of the nonlinear problem represented by Eqs. (1) and (2) can also be represented in terms of the retarded propagator G by an expression similar to Eq. (18). These expressions give the solution $u(x)$ of the nonlinear system represented by Eqs. (1) and (2) in an integral form, in terms of the advanced or retarded propagators, respectively. Both of these integral forms are formally similar to those encountered in linear problems as can be found in field theory.[15] But, in contradistinction to those linear problems, Eq. (18) and its retarded propagator counterpart are *nonlinear integral equations* since both the retarded and the advanced propagators depend parametrically on the solution u.

V. INTEGRAL EQUATIONS FOR PROPAGATORS

A. Integral equation satisfied by the forward propagator

Consider that $u_0(x)$ is known, and consider the two systems satisfied, respectively, by the propagators

$$G_0(x'',x) \equiv G(u_0(x'');x'',x)$$

and

$$G_u^*(x'',x') \equiv G^*(u(x'');x'',x'),$$

i.e.,

$$L(u_0(x''))G_0(x'',x) = \delta(x''-x), \quad \text{in } \Omega, \tag{23}$$

$$\gamma(G_0(x'',x)) = 0, \quad \text{on } \partial\Omega, \tag{24}$$

and

$$L^*(u(x''))G_u^*(x'',x') = \delta(x''-x'), \quad \text{in } \Omega, \tag{25}$$

$$\gamma^*(G_u^*(x'',x')) = 0, \quad \text{on } \partial\Omega. \tag{26}$$

Forming the inner products of Eqs. (25) and (23) with $G_0(x'',x)$ and $G_u^*(x'',x')$, respectively, yields

$$[L(u_0(x''))G_0(x'',x), G_u^*(x'',x')] = G_u^*(x,x'), \tag{27}$$

and

$$\langle G_0(x'',x), L^*(u(x''))G_u^*(x'',x')\rangle = G_0(x',x) = G_0^*(x,x'), \quad \text{by reciprocity.} \tag{28}$$

Subtracting Eq. (28) from Eq. (27) gives

$$G_u^*(x,x') - G_0^*(x,x') = [L(u_0(x''))G_0(x'',x), G_u^*(x'',x')] - \langle G_0(x'',x), L^*(u(x''))G_u^*(x'',x')\rangle. \tag{29}$$

Writing Eq. (10) for the particular case when $u = u_0(x'')$, $h = G_0(x'',x)$, and $v = G_u^*(x'',x')$ shows that the first term on the right side of Eq. (29) can be recast as

$$[L(u_0(x''))G_0(x'',x), G_u^*(x'',x')] = \langle G_0(x'',x), L^*(u_0(x''))G_u^*(x'',x')\rangle + \{\Sigma(u_0(x''))(G_0(x'',x), G_u^*(x'',x'))\}_{\gamma(G_0)=0,\ \gamma^*(G_u^*)=0}. \tag{30}$$

Therefore replacing (30) in (29) and using the reciprocity relationship to replace $G_0(x'',x)$ by $G_0^*(x,x'')$ in the resulting expression leads to

$$G_u^*(x,x') = G_0^*(x,x') + \langle G_0^*(x,x''), [L^*(u_0(x'')) - L^*(u(x''))]G_u^*(x'',x')\rangle + \{\Sigma(u_0(x''))(G_0(x'',x), G_u^*(x'',x'))\}_{\gamma(G_0)=0,\ \gamma^*(G_u^*)=0}. \tag{31}$$

Equation (31) is the closed-form nonlinear integral equation satisfied by the forward propagator.

B. Integral equation satisfied by the backward propagator

As before, consider that $u_0(x,t)$ is known, and consider the systems satisfied, respectively, by the propagators

$$G_u(x'',x) \equiv G(u(x'');x'',x)$$

and

$$G_0^*(x'',x) \equiv G^*(u_0(x''),x'',x'),$$

i.e.,

$$L(u(x''))G_u(x'',x) = \delta(x''-x), \quad \text{in } \Omega, \tag{32}$$

$$\gamma(G_u(x'',x)) = 0, \quad \text{on } \partial\Omega, \tag{33}$$

and

$$L^*(u_0(x''))G_0^*(x'',x') = \delta(x''-x'), \quad \text{in } \Omega, \tag{34}$$

$$\gamma^*(G_0^*(x'',x')) = 0, \quad \text{on } \partial\Omega. \tag{35}$$

Taking the inner products of (32) and (34) with $G_0^*(x'',x')$ and $G_u(x'',x)$, respectively, gives

$$[L(u(x''))G_u(x'',x), G_0^*(x'',x')] = G_0^*(x,x') = G_0(x',x), \tag{36}$$

by reciprocity, and

$$\langle G_u(x'',x), L^*(u_0(x''))G_0^*(x'',x')\rangle = G_u(x',x). \tag{37}$$

Proceeding as before, i.e., using Eq. (10) to recast the left side of Eq. (37) in terms of L and subtracting Eq. (36) from Eq. (37), leads to the following nonlinear integral equation for the backward propagator $G_u(x',x)$:

$$G_u(x',x) = G_0(x',x) + [(L(u_0(x'')) - L(u(x')))G_u(x'',x), G_0(x',x'')] - \{\Sigma(u_0(x''))(G_u(x'',x), G_0^*(x'',x'))\}_{\gamma(G_u)=0,\ \gamma^*(G_0^*)=0}. \tag{38}$$

It is essential to note that the integral equations (31) and (38) for the advanced and the retarded propagator, respectively, are *exact*, and their nonlinear character reflects *exactly* the nonlinearities in the original system represented by Eq. (1). Thus Eqs. (31) and (38) are fundamentally different—both as to the way they were derived and as to their final forms—from the nonlinear equations expressing the Green's functions for one, two, or many particles in many-body and field theories[15]; in these theories, the nonlinearities appearing in the respective Green's functions are not intrinsic, but are introduced as a result of the *approximations* needed to close the respective equations.

In general, Eqs. (31) and (38) for the advanced and retarded propagators G_u^* and G_u, respectively, are nonlinear integrodifferential equations. However, the order of the highest derivative appearing in Eqs. (31) and (38) will be lower than the order of the highest derivative appearing in the original nonlinear system represented by Eq. (1); in many cases Eqs. (31) and/or (38) actually reduce to purely integral equations.

Although exact and retaining the full nonlinear character of the original problem [i.e., Eqs. (1) and (2)], Eqs. (31) and (38) may still be difficult to solve in practice. Nevertheless, Eqs. (31) and/or (38) together with Eq. (18) appear to open possibilities of applying well-established functional-analytic tools (such as fixed-point arguments and, as we shall present in the next section, perturbative series expansions) in ways not feasible for the original nonlinear problem.

VI. SERIES EXPANSIONS

Consider that $u_0(x)$ represents an arbitrary, but known, point in phase space, and that $h(x) \equiv u(x) - u_0(x)$ represents the difference between $u_0(x)$ and the solution $u(x)$ of Eq. (1). The equation for $h(x)$ is obtained by applying Eq. (5) to Eq. (18). This gives

$$h(x') = -u_0(x') - \langle u_0 + h, [(u_0+h)/\|(u_0+h)\|^2]\,\overline{N(0)}G_u^*(x,x')\rangle - \{\Sigma(u_0+h)(u_0+h, G_u^*(x,x'))\}_{B(u)=0,\ \gamma^*(G_u^*)=0}. \tag{39}$$

The relationship between the advanced propagator $G_u^*(x,x')$ and $h(x)$ is obtained by applying Eq. (5) to Eq. (31). This yields

$$G_u^*(x,x') = G_0^*(x,x') + \left\langle G_0^*(x,x''), \left(\int_0^1 L^{*\prime}(u_0+\epsilon h)h\,d\epsilon\right) G_u^*(x'',x')\right\rangle + \{\Sigma(u_0(x''))(G_0(x'',x), G_u^*(x'',x'))\}_{\gamma(G_0)=0,\ \gamma^*(G_u^*)=0}. \tag{40}$$

The counterparts of Eqs. (39) and (40), i.e., the equations that relate $h(x)$ and the retarded propagator $G_u(x,x')$, are similarly obtained.

Note that Eq. (40) is linear in G_u^* and can, in principle, be solved by standard linear methods, e.g., by iteration, in terms of the known quantities u_0 and G_0^*. The kernel of this integral equation, though, depends parametrically on h. In practice, $L^{*\prime}(u_0+\epsilon h)$ may be further expanded in powers of h (using Taylor series for operators[13]); such an expansion is useful when $L^*(u)$ has a polynomial-type dependence on u since then only a finite number of operators multilinear in h appear in the expansion. Replacing the resulting expression for G_u^* in Eq. (39) yields a closed-form nonlinear integral equation for h involving only the known quantities u_0 and G_0^*. Note that $\Sigma(u_0+h)$ in Eq. (39) may also need to be expanded in powers of h, i.e.,

$$\Sigma(u_0+h)=\Sigma(u_0)+\int_0^1 \Sigma'(u_0+\epsilon h;h)d\epsilon$$

$$=\Sigma(u_0)+\sum_{n=1}^{\infty}\Sigma^{(n)}(u_0)h^n.$$

The expansions represented by Eqs. (39) and (40) represent generalizations of the Lippman–Schwinger-type expansions, on the one hand, and invariant embedding, on the other hand. In particular, these equations can be used to obtain closed-form expressions for perturbation diagrams. Also, when u_0 is a nominal solution of the system represented by Eqs. (1) and (2), Eq. (39) can be used to perform an efficient, comprehensive, and global sensitivity analysis to assess the variations in u_0 (or any function thereof) induced by any variations in the system's parameters.

VII. ILLUSTRATIVE EXAMPLE: THE RICCATI EQUATION

It is instructive to illustrate the application of the formalism developed in Secs. II–VI to a simple example whose solution is available analytically, in closed form, namely the Riccati equation. In addition to being tractable analytically, the Riccati equation occurs naturally in many physical applications such as control theory,[17] the Toda lattice,[18] or the theory of reflection on random media.[19]

Thus we consider the Riccati equation

$$N(u)\equiv\frac{du}{dt}+bu^2-c=0,\quad t\in[0,t_f],\quad b>0,\quad c>0, \tag{41}$$

$$u(0)=u_i>0, \tag{42}$$

and we choose the spaces H_1, H_2 as $H_1=H_2=L_2([0,t_f];dt)$. The unique solution of Eqs. (41) and (42) is

$$u(t)=\frac{u_i+(c/b)^{1/2}\tanh[t(bc)^{1/2}]}{1+u_i(b/c)^{1/2}\tanh[t(bc)^{1/2}]}. \tag{43}$$

The Gâteaux differential of $N(u)$ in Eq. (41) is the operator $N'(u)$ acting on h as

$$N'(u)h=\frac{dh}{dt}+2buh.$$

Therefore the predual and antidual of $N(u)$ are given by

$$L(u)h=\frac{dh}{dt}+buh, \tag{44}$$

and

$$L^*(u)v=-\frac{dv}{dt}+buv, \tag{45}$$

respectively. Thus the advanced propagator $G_u^*(t,t')$ is the solution of

$$L^*(u)G_u^*(t,t')=-\frac{dG_u^*(t,t')}{dt}+bu(t)G_u^*(t,t')=\delta(t-t'), \tag{46}$$

$$G_u^*(t,t')=0,\quad \text{at } t=t_f \quad (\text{i.e., for } t>t'). \tag{47}$$

As expected, Eq. (46) is a (first-order, nonhomogeneous) linear equation for $G_u^*(t,t')$; its solution is

$$G_u^*(t,t')=H_+(t'-t)\exp\left[\int_{t'}^{t}bu(\tau)d\tau\right], \tag{48}$$

where $H_+(t'-t)$ is the unit step (Heaviside) function defined as

$$H_+(t'-t)=\begin{cases}0, & \text{for } t'<t,\\ 1, & \text{for } t'\geqslant t.\end{cases} \tag{49}$$

Performing the calculations indicated prior to Eq. (18) gives

$$0-u(t')=\int_0^{t_f}\left(\frac{du}{dt}G_u^*(t,t')+bu^2G_u^*(t,t')-cG_u^*(t,t')\right)dt$$

$$-\int_0^{t_f}\left(-u\frac{dG_u^*(t,t')}{dt}+u^2bG_u^*(t,t')\right)dt$$

$$=[u(t)G_u^*(t,t')]_0^{t_f}-\int_0^{t_f}cG_u^*(t,t')dt.$$

This expression can be further reduced by using Eq. (47) and interchanging t and t' to obtain

$$u(t)=\int_0^{t_f}cG_u^*(t',t)dt'+u_iG_u^*(0,t). \tag{50}$$

Replacing $G_u^*(t',t)$ by its expression given in Eq. (48) and taking into account that $H_+(t-t')$ is zero for $t'>t$ leads to the following implicit closed form for $u(t)$:

$$u(t)=\exp\left(-\int_0^t bu(\tau)d\tau\right)\times\left[u_i+c\int_0^t dt'\exp\left(\int_0^{t'}bu(\tau)d\tau\right)\right]. \tag{51}$$

Equation (51) represents a fixed-point form of the original Riccati equation. It is easy to verify that the solution $u(t)$ given by Eq. (51) indeed satisfies the original Riccati equation together with the respective initial condition.

The retarded propagator $G_u(t,t'')$ satisfies the linear system

$$L(u)G_u(t,t'')=\frac{dG_u(t,t'')}{dt}+bu(t)G_u(t,t'')=\delta(t-t''), \tag{52}$$

$$G_u(0,t'')=0,\quad \text{at } t=0 \quad (\text{i.e., for } t<t''). \tag{53}$$

Solving Eqs. (52) and (53) yields

$$G_u(t,t'') = H_+(t-t'')\exp\left[\int_t^{t''} bu(\tau)d\tau\right]. \quad (54)$$

It is apparent from Eqs. (54) and (48) that the advanced and retarded propagators satisfy the reciprocity relationship given in Eq. (22), i.e., $G_u^*(t'',t') = G_u(t',t'')$. Of course, the closed-form solutions for the propagators are not needed to obtain this reciprocity relationship; it can be obtained directly by taking the inner products of Eqs. (46) and (52) with $G_u(t,t'')$ and $G_u^*(t,t')$, respectively, subtracting one inner product from the other, and taking into account the initial and final time conditions given in Eqs. (47) and (53). Using this reciprocity relationship in Eq. (50) gives the alternative form of the solution $u(t)$ in terms of the retarded propagator as

$$u(t) = \int_0^{t_f} cG_u(t,t')dt' + u_i G_u(t,0). \quad (55)$$

The usefulness of the series expansions presented in Sec. VI can be highlighted by comparison to the customary perturbation/iteration methods used for nonlinear equations. Recall that, in their most powerful setting, the perturbation/iteration methods start by considering the nonlinear system to be a perturbation of the linear system obtained by neglecting the nonlinear terms. Thus, for the Riccati equation given by Eqs. (41) and (42), the corresponding linear system is obtained by neglecting the nonlinear term bu^2, so the starting point for the perturbation/iteration procedure is the linear equation

$$N_0(u) \equiv \frac{du_0}{dt} - c = 0, \quad u_0(0) = u_i. \quad (56)$$

The solution $u_0(t) = u_i + ct$ of Eq. (56) is then used as the starting point in the perturbation/iteration procedure

$$N_n(u) \equiv \frac{du_n}{dt} - c = -bu_{n-1}^2, \quad u_n(0) = u_i. \quad (57)$$

Solving Eq. (57) successively for $n = 1,2,...$ yields the successive approximations $u_1(t),u_2(t),...$, for the solution $u(t)$ of the Riccati equation. The explicit expressions of the first two iterates of $u(t)$ are

$$u_1 = (t) = u_i + ct - b(u_i^2 t + cu_i t^2 + c^2 t^3/3), \quad (58)$$

and

$$u_2(t) = u_i + ct - b(u_i^2 t + cu_i t^2 + c^2 t^3/3) + b^2(u_i^3 t^2 + \tfrac{4}{3}u_i^2 ct^3 + \tfrac{2}{3}c^2 u_i t^4 + \tfrac{2}{15}c^3 t^5) - b^3(\tfrac{1}{3}u_i^4 t^3 + \tfrac{1}{2}cu_i^3 t^4 + \tfrac{1}{3}c^2 u_i^2 t^5 + \tfrac{1}{3}c^3 u_i t^6 + \tfrac{1}{63}c^4 t^7). \quad (59)$$

Thus the perturbation/iteration procedure yields a power series expansion, in powers of b, of the exact solution $u(t)$ of the Riccati equation. Specifically, the nth approximation $u_n(t)$ of the exact solution $u(t)$ represents the power series expansion of $u(t)$ up to and including the terms in b^{2^n-1}. This fact can be readily verified by comparing the above expressions of u_0, u_1, and u_2 with the direct expansion in powers of b of the exact solution given by Eq. (43), i.e.,

$$u(t) = u_i + ct - b(u_i^2 t + u_i t^2 c + t^3 c^2/3) + b^2(u_i^3 t^2 + \tfrac{4}{3}u_i^2 ct^3 + \tfrac{2}{3}c^2 u_i t^4 + \tfrac{2}{15}c^3 t^5) - b^3(\tfrac{1}{3}u_i^4 t^3 + \tfrac{1}{2}cu_i^3 t^4 + \tfrac{1}{3}c^2 u_i^2 t^5 + \tfrac{1}{3}c^3 u_i t^6 + \tfrac{1}{63}c^4 t^7) + O(b^4) \equiv u_0 + \sum_{i=1}^{\infty} b^i u_i(t). \quad (60)$$

The functions G_0 and G_0^* are obtained using $u_0 = u_i + ct$ in Eqs. (23) through (26); this gives

$$G_0(t,t') = H_+(t-t') \times \exp[bu_i(t'-t) + bc[(t')^2 - t^2]/2], \quad (61)$$

and

$$G_0^*(t,t') = H_+(t'-t) \times \exp[bu_i(t-t') + bc[t^2 - (t')^2]/2], \quad (62)$$

respectively. Of course, Eqs. (61) and (62) could also have been obtained by substituting $u_0 = u_i + ct$ in Eqs. (54) and (48), respectively, and performing the integrations over τ. As expected, G_0 and G_0^* satisfy the general reciprocity relation between propagators as given by Eq. (22).

In view of the expressions for $L^*(u)$, $L(u)$, and noting that $\{\Sigma(u_0)(G_0,G_u^*)\}$ vanishes because of the respective initial and final-time conditions for this example [cf., Eqs. (46), (47), (52), and (53)], the integral equation satisfied by the forward propagator, i.e., Eq. (31), reduces to

$$G_u^*(t,t') = G_0^*(t,t') + \int_0^{t_f} G_0^*(t,t'') \times [-bh(t'')]G_u^*(t'',t')dt'', \quad (63)$$

where $h(t) \equiv u(t) - u_0(t)$. Solving Eq. (63) by iteration gives

$$G_u^*(t,t') = G_0^*(t,t') + \int_0^{t_f} G_0^*(t,\tau_1)[-bh(\tau_1)]G_0^*(\tau_1,t')d\tau_1 + \int_0^{t_f}\int_0^{t_f} G_0^*(t,\tau_1)[-bh(\tau_1)]G_0^*(\tau_1,\tau_2) \times [-bh(\tau_2)]G_0^*(\tau_2,t')d\tau_2\,d\tau_1 + \int_0^{t_f}\int_0^{t_f}\int_0^{t_f} G_0^*(t,\tau_1)[-bh(\tau_1)]G_0^*(\tau_1,\tau_2) \times [-bh(\tau_2)]G_0^*(\tau_2,\tau_3)[-bh(\tau_3)]G_0^*(\tau_3,t')d\tau_3\,d\tau_2\,d\tau_1 + O(h^4). \quad (64)$$

For this example, the expression for $h(t)$, given in general by Eq. (39), becomes

$$h(t) = -u_0(t) + \int_0^{t_f} cG_u^*(t',t)dt' + u_i G_u^*(0,t). \quad (65)$$

Replacing now Eq. (64) in Eq. (65) gives

$$h(t) = -u_0(t) + \int_0^{t_f} c\Big\{G_0^*(t',t) + \int_0^{t_f} G_0^*(t',\tau_1)[-bh(\tau_1)]G_0^*(\tau_1,t)d\tau_1 + \int_0^{t_f}\int_0^{t_f} G_0^*(t',\tau_1)[-bh(\tau_1)]G_0^*(\tau_1,\tau_2)$$
$$\times[-bh(\tau_2)]G_0^*(\tau_2,t)d\tau_2\,d\tau_1 + \int_0^{t_f}\int_0^{t_f}\int_0^{t_f} G_0^*(t',\tau_1)[-bh(\tau_1)]G_0^*(\tau_1,\tau_2)[-bh(\tau_2)]G_0^*(\tau_2,\tau_3)$$
$$\times[-bh(\tau_3)]G_0^*(\tau_3,t)d\tau_3\,d\tau_2\,d\tau_1\Big\}dt' + u_i\Big\{G_0^*(0,t) + \int_0^{t_f} G_0^*(0,\tau_1)[-bh(\tau_1)]G_0^*(\tau_1,t)d\tau_1$$
$$+ \int_0^{t_f}\int_0^{t_f} G_0^*(0,\tau_1)[-bh(\tau_1)]G_0^*(\tau_1,\tau_2)[-bh(\tau_2)]G_0^*(\tau_2,t)d\tau_2\,d\tau_1 + \int_0^{t_f}\int_0^{t_f}\int_0^{t_f} G_0^*(0,\tau_1)[-bh(\tau_1)]$$
$$\times G_0^*(\tau_1,\tau_2)[-bh(\tau_2)]G_0^*(\tau_2,\tau_3)[-bh(\tau_3)]G_0^*(\tau_3,t)d\tau_3\,d\tau_2\,d\tau_1\Big\}dt' + O(h^4). \tag{66}$$

Note that the expansions given in Eqs. (64) and (66) are valid for any b; each iteration brings in the product (bhG_0^*), so that equal-order truncations on h, b, or G_0^* are equivalent, but only as long as Eqs. (39) and (40) are the basis for the iterative solution. Otherwise, if G_0^* is also expanded in powers of b, the resulting expansion is severely restricted to small time values by the appearance of secular terms. This fact becomes apparent by noting that the expressions of G_0^* and G_0 [cf. Eqs. (61) and (62)] already contain summed up powers of b in the respective exponential terms; in other words, the secular terms are already summed up, into a convergent expression for large t, in the forms of G_0^* and G_0. Thus the expansions shown in Eqs. (64) and (66) and, by implication, those given by Eqs. (39) and (40) have a considerably larger range of validity than the customary perturbation/iteration procedures for nonlinear equations.

It is also important to note that the (approximation to the) solution $u(t)$ obtained by either adding $u_0(t)$ to $h(t)$ given by Eq. (66), or by replacing Eq. (64) in Eq. (50) and performing the respective integration, is completely equivalent to the corresponding iterate of the fixed-point (implicit) form solution represented by Eq. (51). For example, the first iterate $u_1(t)$ is obtained by using Eqs. (64) or (66) truncated at $O(h^2)$; this gives

$$u_1(t) = \exp\Big(-\int_0^t bu_0(\tau)d\tau\Big)\times\Big[u_i + c\int_0^t dt' \exp\int_0^{t'} bu_0(\tau)d\tau\Big]$$
$$= \exp\Big[-b\Big(u_i t + c\frac{t^2}{2}\Big)\Big]\times\Big[u_i + c\int_0^t dt' \exp b\Big(u_i t' + c\frac{(t')^2}{2}\Big)\Big], \tag{67}$$

which is identical to the expression of the first iterate of the fixed-point solution represented by Eq. (51). Comparing Eqs. (67) and (58) reveals clearly that Eq. (67), obtained from the generalized series expansions presented in Sec. VI, represents a vast improvement over the customary perturbation theory result, since all of the secular terms are now summed up into a bounded expression involving decaying exponentials. Furthermore, the $(n+1)$th iterate of the integral, fixed-point solution given by our new formalism is

$$u_{n+1}(t) = \exp\Big[-\int_0^t bu_n(\tau)\Big]\times\Big\{u_i + c\int_0^t d\tau \exp\int_0^\tau bu_n(y)dy\Big\}. \tag{68}$$

Equation (68) clearly shows that asymptotically, as $t\to\infty$, these iterates give

$$\lim_{t\to\infty} u_{n+1}(t)u_n(t) = c/b, \quad \text{for any } n,$$

which is the *exact* asymptotic behavior of the *exact* solution $u(t)$ of the Riccati equation; this fact can be readily seen from the analytical solution given in Eq. (43).

To test our formalism, we have performed a numerical comparison between Eq. (68) and the form

$$u_{n+1}(t) = \int_0^t [c - bu_n^2(s)]ds + u_i, \quad u_0(t) = u_i, \tag{69}$$

which is obtained by applying the customary Picard iteration method to the differential (Riccati) equation (41). Our preliminary results indicate that Eq. (68) is superior to Eq. (69) regarding convergence (for the same accuracy), stability, and CPU computer time. The detailed results of this comparison, and comparisons to other numerical methods, are expected to be published elsewhere.[20]

VIII. CONCLUSIONS

We have presented in this paper a new and canonical formalism for finding solutions to equations involving nonlinear (differential, multiplicative, integral) operators. This formalism is not limited, *per se*, to special boundary and/or initial conditions, or to a special structure (e.g., evolution) of the underlying nonlinear operators. Our method relies on the construction and exploitation of the *dual* operator, which has been shown to be the analog of the adjoint operator in linear theory. This dual operator enables us to construct and use the advanced and retarded nonlinear propagators; these propagators generalize the customary Green's functions in linear theory. Note that the construction of the dual operator requires the existence of the Gâteaux derivatives of the nonlinear operators underlying the original problem; thus the applicability of our formalism is limited by this regularity requirement.

Although obtained from the variational operators

$[N'(u)]^*$ and $N'(u)$ by a simple functional integration, the operators $L^*(u)$ and $L(u)$ are the fundamental and natural operators for developing our formalism as the nonlinear analog to the Green's function method in linear theory. Also essential to developing our formalism is the observation that the relationship $L(u)u = N(u) - N(0)$ is satisfied by $L(u)$, but not by $N'(u)$. Thus, as shown by Eq. (18), the forward and backward propagators G_u^* and G_u, which are solutions of equations involving the operators $L^*(u)$ and $L(u)$, carry all the information needed to solve the original nonlinear equation. In this way, the propagators G_u^* and G_u generalize the customary Green's functions from linear theory; in particular, when the original problem is linear, G_u^* and G_u reduce to the usual Green's functions. In this case, the dual operator $N^+(u)$ becomes the actual adjoint of $N(u)$, $L^*(u)$ becomes the formal adjoint of $N(u)$, and so forth. Furthermore, L, L^*, N^+ become independent of u and, consequently, so do G^* and G.

We have shown that the advanced and retarded propagators G_u^* and G_u satisfy a reciprocity relationship analogous to that satisfied by the customary Green's functions. We have further shown that these propagators can be obtained as solutions of equations that are both nonlinear and integrodifferential—in general—but where the order of the highest derivative will always be lower than the order of the highest derivative appearing in the original nonlinear system. Note, though, that in many cases, including the Korteweg–de Vries equation, the integrodifferential equation for one or both of the propagators reduces to a purely integral equation.

We further noted that these integrodifferential (or integral) equations for the propagators are *exact*, and their nonlinear character reflects *exactly* the nonlinearities present in the original system. This is in contradistinction to nonlinearities that may appear in the expressions of the Green's functions in many-body and field theories, where such nonlinearities are not intrinsic to these theories but are introduced as a result of approximations needed to close the respective equations.

Using the advanced or retarded propagators, we have converted the solution of the original boundary/initial value problem into an integral form. In principle, such a conversion is always advantageous, even if the resulting integral form is nonlinear. This is both because (a) the contraction principle and/or fixed-point theorems could be applied to this integral form (but not to the original nonlinear boundary/initial value problem) to prove existence and uniqueness conditions, and (b) most numerical analysis and computational methods are comparatively more mature and less difficult to implement for integral equations than for differential ones. Note, though, that because the nonlinearities of the original problem are inherently and exactly incorporated in the integral equations produced by our formalism, these integral equations may still be difficult to solve in practice.

The formal character of the derivations underlying our formalism is underscored by the fact that we have not addressed rigorously the issues of existence, well posedness, and uniqueness; these issues will be addressed in detail elsewhere. Furthermore, a direct comparison between our formalism and the few available methods for solving nonlinear equations (e.g., inverse scattering transform, similarity) is rendered difficult at this stage. This is because, in general, these methods directly yield, when applicable, the solution of the original problem, while our formalism—although clearly more general—yields the solution in terms of the advanced or retarded propagator which, in turn, satisfies a nonlinear equation. Our current research is aimed both at performing such comparisons[21] for the Korteweg–de Vries equation and at investigating several promising practical applications of this formalism to solve nonlinear boundary/initial value problems.

ACKNOWLEDGMENTS

We are particularly indebted to Dr. A. Zucker for continuous and understanding support.

We gratefully acknowledge partial financial support from the U.S. Department of Energy under Contract No. DE-AC05-840R21400 with Martin Marietta Energy Systems, Inc.

[1]N. J. Zabusky and M. D. Kruskal, Phys. Rev. Lett. **15**, 240 (1965).
[2]C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, Phys. Rev. Lett. **19**, 1095 (1967).
[3]F. Calogero and A. Degasperis, *Spectral Transforms and Solitons, I*, (North-Holland, Amsterdam, 1982).
[4]P. D. Lax, Commun. Pure Appl. Math. **28**, 141 (1975).
[5]H. H. Chen, Y. C. Lee, and C. J. Liu, Phys. Scr. **20**, 490 (1979).
[6]B. Grammaticos, B.Dorizzi, and A. Ramani, Phys. Rev. Lett. **53**, 1 (1984).
[7]T. Nishigori, Physica A **111**, 323 (1982); A Roy Chowdhury, A. K. Dey, and S. Roy, Phys. Scr. **32**, 161 (1985).
[8]F. W. Nijhoff, Lett. Math. Phys. **39**, 235 (1985).
[9]W. Oevel and A. S. Fokas, J. Math. Phys. **25**, 918 (1984).
[10]B. Fuchssteiner, Prog. Theor. Phys. **65**, 861 (1981).
[11]K. M. Case, Phys. Rev. Lett. **55**, 445 (1985).
[12]D. G. Cacuci, R. B. Perez, and V. Protopopescu, *Propagators for General Nonlinear Equations*, ORNL/TM-10016 (Oak Ridge National Laboratory, Oak Ridge, TN, 1986).
[13]M. Z. Nashed, "Differentiability and related properties of nonlinear operators: Some aspects of the role of differentials in nonlinear functional analysis," in *Nonlinear Functional Analysis and Applications*, edited by L. B. Rall (Academic, New York, 1971).
[14]E. A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations* (McGraw-Hill, New York, 1955).
[15]J. Schwinger, Proc. Natl. Acad. Sci. **37**, 452, 455 (1951); G. Baym and L. P. Kadanoff, Phys. Rev. **124**, 287 (1961); P.C. Martin, E. D. Siggia, and H. A. Rose, Phys. Rev. A **8**, 423 (1973).
[16]D. G. Cacuci, J. Math. Phys. **22**, 2794, 2803 (1981).
[17]J. L. Lions, *Optimal Control of Systems Governed by Partial Differential Equations* (Springer, New York, 1971).
[18]A. K. Common and D. E. Roberts, J. Phys. A **19**, 1889 (1986).
[19]W. E. Kohler, SIAM J. Appl. Math. **46**, 464 (1986).
[20]Y. Y. Azmy, D. G. Cacuci, and V. Protopopescu, in preparation.
[21]D. G. Cacuci and O. Karakashian, in preparation.

Dirac bracket revisited

G. J. Ruggeri
Departamento de Física, Facultad de Ciencias, Universidad Central de Venezuela, Caracas, Venezuela

(Received 1 October 1986; accepted for publication 13 August 1987)

Among all possible singular (Lie) brackets in classical dynamics of discrete systems, the matrix of Dirac brackets between phase-space coordinates is uniquely characterized by having the symplectic (Lagrange) matrix as a generalized inverse. This result is used to prove an explicit representation of the Dirac bracket in terms of its singular functions.

I. INTRODUCTION

Since the pioneering work of Dirac, Bergmann, Anderson, and others on constrained Hamiltonian systems in classical mechanics and field theory, there has been continuous interest in the structure of the ensuing dynamical scheme. In particular, a number of papers[1] have been devoted to elucidating the structural features of the Dirac bracket which, as we know, emerges as the natural classical (Lie) bracket for dynamical systems with second class constraints.

The Dirac bracket, however, seems to be only a particular instance of a wide set of (singular) classical brackets one can think of.[2] So the question naturally arises of how to distinguish (uniquely, if possible) the Dirac bracket among the class of structures which share the following properties.[3]

(a) For any two dynamical variables A, B (functions of the phase-space coordinates ξ^μ, $\mu = 1,...,R$) the bracket is defined by

$$\{A,B\} \equiv f^{\mu\nu}\,\partial_\mu A\,\partial_\nu B. \tag{1}$$

Here and in what follows $\partial_\mu \equiv \partial/\partial\xi^\mu$ and the sum convention is implied over repeated indices (Greek or otherwise).

The (non-necessarily constant) matrix $f \equiv (f^{\mu\nu})$ is (b) skew symmetric, (c) satisfies the Jacobi identity, and (d) is singular. Its rank, $R - C$ say, is necessarily an even number.

These properties confer to the space of dynamical variables the structure of a function group in the sense of Lie. It is known that the singularity of f is completely accounted for by the existence of C independent functions ϕ^i, $i = 1,...,C$, which are in involution with any dynamical variable; that is,

$$\{\phi^i,\xi^\mu\} = 0, \tag{2}$$

for all values of i and μ. These functions are called singular.[4] For the Dirac bracket the singular functions are those which express the second class constraints.

Equation (2) is equivalent to

$$f^{\mu\nu}\,\partial_\nu\phi^i = 0, \tag{3}$$

which shows the set $(\partial_\nu\phi^i)$ as a null vector of f labeled by the index i. The most general null vector is a combination of the form $a_i\,\partial_\mu\phi^i$ for some set of coefficients a_i. Any of these combinations can only be 0 for all μ if $a_i = 0$ for all i.

II. DIRAC BRACKET

We shall show below that the matrix of Dirac brackets can be uniquely characterized as that matrix (of order $R = 2N$) which satisfies, besides properties (a)–(d) above, the equation

$$f\epsilon f = f. \tag{4}$$

Here $\epsilon = (\epsilon_{\alpha\beta})$ is the symplectic (or Lagrange) matrix of order $2N$. It is defined by

$$\epsilon_{\alpha\beta} \equiv \sum_{m=1}^{N} (\delta_{\alpha,N+m}\delta_{\beta,m} - \delta_{\alpha,m}\delta_{\beta,N+m}). \tag{5}$$

Equation (4) can be described as saying that the Lagrange matrix is a *regular* (skew symmetric) generalized inverse of the matrix of Dirac brackets. For a singular bracket that is the nearest we can get to the well-known fact that the Lagrange matrix is the inverse of the matrix of Poisson brackets. In other words, both the matrices of Poisson and Dirac brackets share the matrix ϵ as a common generalized inverse. This fact is a simple and compact characterization of Dirac brackets which shows neatly its conspicuous relation to Poisson brackets.

Equation (4) can be freed from any specific system of coordinates in phase space. In fact, that equation remains invariant under regular transformations of coordinates if f and ϵ are, respectively, transformed as contravariant and covariant tensors. Under these transformations ϵ loses its canonical form (5) and becomes, in general, a nonconstant matrix. It continues to be, however, a regular matrix that satisfies the Lagrange identity. These two features are all that we really need in Eq. (4). Lagrange and Jacobi identities are, of course, closely connected. In fact, if relation (4) is granted the first identity implies the second. (See the Appendix.[5])

Define now the skew symmetric matrix $\ell = (\ell_{\mu\nu})$ through the identification

$$\ell \equiv \epsilon - \epsilon f \epsilon \tag{6}$$

This matrix, which cannot vanish identically, is annihilated by f; that is

$$f\ell = \ell f = 0. \tag{7}$$

But then $\ell_{\mu\nu}$ must be a bilinear combination of the null vectors of f, i.e.,

$$\ell_{\mu\nu} = \partial_\mu\phi^i\,C_{ij}\partial_\nu\phi^j. \tag{8}$$

Here (C_{ij}) is also a skew symmetric matrix. Substituting Eq. (8) in Eq. (6), f can be inverted as

$$f^{\mu\nu} = \epsilon^{\mu\nu} - \epsilon^{\mu\alpha}\,\partial_\alpha\phi^i\,C_{ij}\,\partial_\beta\phi^j\,\epsilon^{\beta\nu}, \tag{9}$$

where $\epsilon^{\alpha\beta}$ is a general entry of the matrix of Poisson brackets. Equation (9) is already the standard form for the Dirac bracket. It remains only to show that the matrix (C_{ij}) inverts

 0022-2488/88/020362-03$02.50

the matrix of the Poisson brackets between the singular functions, that is,

$$\{\phi^i,\phi^j\}_P \equiv \epsilon^{\alpha\beta}\partial_\alpha\phi^i\,\partial_\beta\phi^j. \tag{10}$$

This, in fact, is a consequence of the independence of the null vectors of f, because the relations

$$0 = f^{\mu\nu}\,\partial_\nu\phi^k = \epsilon^{\mu\alpha}\,\partial_\alpha\phi^i(\delta_i^k - C_{ij}\{\phi^j,\phi^k\}_P),$$

valid for any μ and k, imply the vanishing of the quantity inside round brackets.

III. AN EXPLICIT REPRESENTATION

The Dirac bracket in its standard form, Eq. (9), depends rather implicitly both on the structure of the Poisson bracket and on the functional form of the singular functions. We show in this section how these two features can be separated, giving to the Dirac bracket a completely explicit expression.

Consider the equality

$$f^{\alpha\beta} = (1/a)f^{\alpha\beta\sigma_1\sigma_2\cdots\sigma_c}\,\partial_{\sigma_1}\phi^1\,\partial_{\sigma_2}\phi^2\cdots\partial_{\sigma_c}\phi^c. \tag{11}$$

Here

$$f^{\alpha\beta\sigma_1\sigma_2\cdots\sigma_c} \equiv \eta_c\,\delta^{\alpha\beta\sigma_1\cdots\sigma_c}_{\mu\nu\lambda_1\cdots\lambda_c}\epsilon^{\mu\nu}\epsilon^{\lambda_1\lambda_2}\cdots\epsilon^{\lambda_{c-1}\lambda_c} \tag{12a}$$

with

$$\eta_c = [2^{(C+2)/2}((C+2)/2)!]^{-1}, \tag{12b}$$

and a is a function to be determined. It will be shown below that Eq. (11) represents, in fact, the Dirac bracket. Equation (12a) contains the so-called generalized delta symbol defined as

$$\delta^{\alpha_1\alpha_2\cdots\alpha_n}_{\beta_1\beta_2\cdots\beta_n} = \sum_P (-1)^p\delta^{\alpha_1}_{P\beta_1}\delta^{\alpha_2}_{P\beta_2}\cdots\delta^{\alpha_n}_{P\beta_n}. \tag{13}$$

The sum on the right-hand side of Eq. (13) is over all permutations P of the lower indices (β), where p is the parity of P with respect to the "natural" order of these indices. This is the one they have at the left-hand side. In fact, due to the obvious equality $\delta^{\alpha_1\cdots\alpha_n}_{\beta_1\cdots\beta_n} = \delta^{\beta_1\cdots\beta_n}_{\alpha_1\cdots\alpha_n}$, there is complete symmetry between upper and lower indices. Then the symbol vanishes unless both sets of indices coincide modulo a permutation. It is also skew symmetric under the interchange of any two upper (lower) indices. Thus the numerical factor η_c in the right-hand side of Eq. (12a) reduces all equal terms that come from the interchange of indices within each factor ϵ and from permutations between these factors.

The symbol $f^{\alpha\beta\sigma_1\cdots\sigma_c}$ is then completely skew symmetric with respect to the interchange of any two of its indices. This already gives $f^{\alpha\beta}\,\partial_\beta\phi^i = 0, i = 1,2,...,C$. [That the most general null vector of $f^{\alpha\beta}$ in Eq. (11) is of the form $a_i\,\partial_\beta\phi^i$ can be easily proved from Eq. (15) below.]

We now show that there exists a unique function a that makes $f^{\alpha\beta}$, as given in Eq. (11), fulfill Eq. (4). In fact, the contracted expression $\epsilon_{\sigma\alpha}f^{\alpha\beta\sigma_1\cdots\sigma_c}$ contains two types of terms. The first type occurs when $\epsilon_{\sigma\alpha}$ encounters a factor $\epsilon^{\alpha\beta}$ (or $\epsilon^{(\beta\alpha)}$). This gives rise to δ^β_σ. A moment of reflection shows that these terms add up to

$$2((C+2)/2)\eta_c\left(\delta^{\sigma_1\cdots\sigma_c}_{\mu_1\cdots\mu_c}\epsilon^{\mu_1\mu_2}\cdots\epsilon^{\mu_{c-1}\mu_c}\right)\delta^\beta_\sigma.$$

The second type of term corresponds to contractions of the form $\epsilon_{\sigma\alpha}\epsilon^{\alpha\sigma_i} = \delta^{\sigma_i}_\sigma$, $i = 1,...,C$. It is seen then that choosing

$$a = \eta_{c-2}\left(\delta^{\sigma_1\cdots\sigma_c}_{\mu_1\cdots\mu_c}\epsilon^{\mu_1\mu_2}\cdots\epsilon^{\mu_{c-1}\mu_c}\right)\partial_{\sigma_1}\phi^1\cdots\partial_{\sigma_c}\phi^c, \tag{14}$$

the f in Eq. (11) satisfies

$$\epsilon_{\sigma\alpha}f^{\alpha\beta} = \delta^\beta_\sigma - h^\beta_i\,\partial_\sigma\phi^i, \tag{15}$$

for some set of coefficients h^β_i. This completes the proof because Eqs. (4) and (15) are equivalent.

An alternative expression for a can be obtained as follows. Contracting Eq. (15) with $\partial_\beta\phi^j$ gives $(h^\beta_i\,\partial_\beta\phi^j - \delta^j_i)\partial_\sigma\phi^i = 0$ for all values of σ. Then

$$h^\beta_i\,\partial_\beta\phi^j = \delta^j_i. \tag{16}$$

Equations (15) and (16) lead now to

$$\mathrm{Tr}(\epsilon f) = \delta^\beta_\beta - \delta^i_i = R - C. \tag{17}$$

Combining Eqs. (11) and (17) we finally get

$$a = (R-C)^{-1}\epsilon_{\beta\alpha}f^{\alpha\beta\sigma_1\cdots\sigma_c}\,\partial_{\sigma_1}\phi^1\cdots\partial_{\sigma_c}\phi^c. \tag{18}$$

An interesting feature of the representation (11), as regards the dynamical role of the Dirac bracket, is that the singular functions and the Hamiltonian will enter in an almost identical way. The difference comes from the function a and disappears when it happens to be (or can be made) a constant.[6] This suggests, as an alternative point of view of the Hamilton–Dirac equations, to treat both the singular functions and the Hamiltonian as the same type of mathematical objects. The resulting dynamical scheme is formally similar to the one advocated by Nambu a number of years ago.[7]

We point out finally that an explicit representation in terms of its singular functions (which do not have to be always constraints in the sense of Dirac) can be found for any singular classical bracket as defined in the Introduction.

APPENDIX: LAGRANGE AND JACOBI IDENTITIES

For skew symmetric matrices ϵ and f, Lagrange and Jacobi identities are, respectively, equivalent to the vanishing of the differential expressions

$$L_{\alpha\beta\gamma} \equiv \partial_\alpha\epsilon_{\beta\gamma} + \partial_\gamma\epsilon_{\alpha\beta} + \partial_\beta\epsilon_{\gamma\alpha},$$
$$J^{\alpha\beta\gamma} \equiv f^{\alpha\mu}\,\partial_\mu f^{\beta\gamma} + f^{\gamma\mu}\,\partial_\mu f^{\alpha\beta} + f^{\beta\mu}\,\partial_\mu f^{\gamma\alpha},$$

for any set of three distinct indices α, β, γ. These identities are dual structural features for matrices ϵ and f related by Eq. (4), exactly as they are in the case of the Poisson bracket. In fact, one can prove directly the equality

$$J^{\alpha\beta\gamma} = f^{\alpha\mu}f^{\beta\nu}f^{\gamma\lambda}L_{\mu\nu\lambda}, \tag{A1}$$

so that the Lagrange identity for ϵ implies the Jacobi identity for f. As here f is singular the converse is not true in general. A proof of Eq. (A1) proceeds very much as in the Poisson bracket case; for brevity we omit the details.

[1] The literature is abundant. The following is just a sample of it. P. A. M. Dirac, Can. J. Math. **2**, 129 (1950); P. G. Bergmann and I. Goldberg, Phys. Rev. **98**, 531 (1955); N. Mukunda and E. C. G. Sudarshan, J. Math. Phys. **9**, 411 (1968); S. Shanmugadhasan, *ibid.* **14**, 677 (1973).

[2] Other instances are some of the structures introduced by Y. Nambu, Phys. Rev. D **7**, 2405 (1973). In this respect see G. J. Ruggeri, Int. J. Theor. Phys. **12**, 287 (1975).

[3]Answers to this problem, distinct from ours, have been given (implicitly or explicitly) in the literature. See, for example, the papers of P. G. Bergmann and I. Goldberg and N. Mukunda and E. C. G. Sudarshan in Ref. 1.

[4]Also called, sometimes, Casimir invariant functions.

[5]Equation (4) admits also the transformations $\epsilon_{\alpha\beta} \to \tilde{\epsilon}_{\alpha\beta} = \epsilon_{\alpha\beta} + \partial_\alpha g_i\, \partial_\beta \phi^i - \partial_\beta g_i\, \partial_\alpha \phi^i$, for arbitrary g_i. This preserves the Lagrange identity. It can be seen that this type of regauging is needed to make room for (Dirac's) canonical transformations. See also, G. J. Ruggeri, "A variational principle for Hamiltonian triplets, Sect. II," submitted for publication.

[6]This occurs, for instance, when there are only two (second class) constraints. In this case $a = \{\phi^1, \phi^2\}_P$, which can always be made equal to 1 by choosing appropriately the form of the constraints.

[7]Y. Nambu in Ref. 2.

[8]G. J. Ruggeri, "On general Hamiltonian dynamics for finite systems," submitted for publication.

Behavior of dynamic soliton solutions in nonintegrable extended Klein–Gordon systems by means of a state plane technique

H. Tateno
Electrotechnical Laboratory, 1-1-4 Umezono, Sakuramura, Niiharigun, Ibaraki, Japan

(Received 16 July 1987; accepted for publication 23 September 1987)

The dynamic behavior of soliton solutions in nonintegrable extended Klein–Gordon systems as a continuum containing a dissipation term and an external force term, expressed by $\phi_{xx} - \phi_{tt} - F(\phi) = G\phi_t - J_B$, is investigated geometrically in a state plane by transforming the equation into three basic equations, each of which is associated with the derivative with respect to x, t, and ϕ, respectively. The initial and the boundary conditions are imposed so that the solution approaches asymptotically the stationary solitary-wave solution as $|x|$ and $|t|$ approach infinity. In the above treatment, the waves consisting of the pair of ϕ_x and ϕ_t are divided into two components, the traveling wave component, $V(x,t)$, and the others. The state plane is then constructed by the coordinates consisting of $V(x,t)$ and ϕ. The traveling wave component is defined with ϕ_T by introducing nonlinear coordinates $\Xi(x,t) = X(x,t) - uT(x,t)$ instead of the linear coordinates $\xi = x - ut$, and travels with a constant velocity of u in the coordinates. Their properties are investigated for various cases of wave interactions. The method for analyzing soliton interactions is described in detail. It is first shown that the solutions to the basic equations presented are in agreement with the well-known solutions for two soliton interactions in the pure sine–Gordon system. Next, as an example, the analytical method is applied to the extended sine–Gordon system, the behavior of the local distortions produced during the soliton–antisoliton interaction because of existence of the moving singularities is described, and details of the mechanism and properties of them are clarified. It is finally shown that the additional dissipation term proportional to ϕ_{xxt} acts to smooth the distortions resulting from the disappearance of the moving singularities.

I. INTRODUCTION

The study of the dynamic behavior of solutions in nonlinear Klein–Gordon systems has proved stimulating and useful in many area of physics.[1,2] The more general systems are the ones in which an external force term and a dissipation term are taken into consideration in the above systems, and the equation is then expressed for the continuum by

$$\phi_{xx} - \phi_{tt} - F(\phi) = G\phi_t - J_B, \tag{1.1}$$

where G is the dissipation coefficient, and J_B is a uniformly applied external force. The equation for the discrete systems is expressed by the finite difference equation corresponding to Eq. (1.1). Then, an expression such as Eq. (1.1) has usually been applied at the continuum limit of the discrete systems. However, the difference equation has also been used for the analysis of continuum systems, especially when the equation is not integrable.[3–11] Such an application of the difference equation for nonintegrable continuum systems seems to be due to the lack of an appropriate theoretical way to treat them exactly at present.

The integrability condition for Eq. (1.1) is given by

$$\phi_{xt} = \phi_{tx}. \tag{1.2}$$

On the contrary, if the equation is not integrable, there may exist at least one position $(x,t) = (x_c,t_c)$ not satisfying Eq. (1.2), i.e.,

$$\phi_{xt} \neq \phi_{tx}, \tag{1.3}$$

which suggests that there should exist a singularity there.

It is shown that the computer simulation for such nonintegrable systems using the corresponding finite difference equation usually brings some emission of radiation in dynamical variations irrespective of whether we take account of G and J_B or not.[3,5–10] Such a simulation is useful for discrete systems, from which we may obtain the exact solution, although time discreteness is still included. However, strictly speaking, we cannot use this method as a means of obtaining exact solutions in the continuum system because there must exist at least at one position satisfying condition Eq. (1.3) for the nonintegrable system, however small the mesh. A perturbation method[1,2,12–14] is normally used for the continuum system if the perturbation is small, but it is powerless for a large perturbation. Moreover, since the method is usually developed from a solution obtained under an integrable condition, and since there is no room to include the effect of the nonintegrability during the development of the perturbation process, the results naturally have no connection with the nonintegrability condition. It is also apparent that it would be useless to make an effort to transform Eq. (1.1) so as to be associated with a linear equation. Thus we understand that we cannot use any of above methods to see the exact behavior of solutions. It is also expected that there are new phenomena based upon nonintegrability that cannot be derived by the above methods if we take account of the nonintegrability condition. Here, we intend to investigate the exact behavior of solutions for the systems as a continuum described by Eq. (1.1) under the condition that the solution approaches asymptotically stationary solitary-wave solutions as both $|x|$ and $|t|$ approach infinity by using a method in which the nonintegrability condition is included. The geo-

 0022-2488/88/020365-21$02.50

metrical method in a state plane (phase plane), which consists of the relation between ϕ and ϕ_t, is a useful means for the investigation of exact solutions on the stationary state to the extended sine–Gordon system.[4,15,16] It is shown that such a state plane technique can be extended to develop to the nonstationary state from the stationary state, and as a result it has been shown that the local distortion based upon the nonintegrability can exist in the extended sine–Gordon system.[17] The most typical feature of this treatment is that we can derive three basic equations, each of which is associated with the derivative with respect to x, t, and ϕ, respectively, and is equivalent to Eq. (1.1). We are then effectively treating the ordinary differential equations instead of the partial differential equation. Thus we can construct the solution starting at the stationary state, and continue to the nonstationary state with conventional analytical techniques for ordinary differential equations.

In this paper we present a detailed version of the state plane technique for solving the equation on the extended Klein–Gordon systems as a continuum expressed by Eq. (1.1) under the condition that there exists a stationary solitary-wave solution as x and t approach infinity, and we also present some results for the application to the extended sine–Gordon system. The plan of this paper is as follows: In Sec. II, we derive three basic equations equivalent to Eq. (1.1), each of which is associated with the derivative with respect to x, t, and ϕ, respectively. In Sec. III, the expression for ϕ_t and ϕ_x is divided into the traveling wave component and other wave components with introduction of nonlinear coordinates. In Sec. IV, the equation describing the traveling wave component is derived in association with the state plane, and the properties of the solution in the stationary state are clarified. In Sec. V, the functions for describing the nonlinear coordinates are introduced and their properties are discussed, and in Sec. VI, the nature of the nonlinear coordinates is clarified. In Sec. VII, the soliton–antisoliton and the soliton–soliton interaction solutions are derived from our basic equations for the pure sine–Gordon system to confirm that our theory is appropriate. In Sec. VIII, the soliton–antisoliton and the soliton–soliton interaction solutions are constructed qualitatively on the state plane based upon the traveling wave component. In Sec. IX, an analytical method for the soliton–antisoliton interaction is developed. In Sec. X, the analytical method is applied to the extended sine–Gordon system, and the behavior of the solution is discussed in detail, where we investigate the local distortions of waves produced by moving singularities. In Sec. XI, the mechanism of the local distortions is clarified, and the solution is compared with the solution for the system having an additional dissipation term proportional to ϕ_{xxt}. In Sec. XII, we summarize our results.

II. EQUATION ON STATE PLANE

We can consider two kinds of phase planes for Eq. (1.1) consisting of (ϕ, ϕ_t) and (ϕ, ϕ_x), respectively, since we are treating two-dimensional problems consisting of x and t. However, since ϕ_t and ϕ_x are not independent of each other, it is convenient to define a state plane common to these planes. For this purpose we express the solution in terms of ϕ_t and ϕ_x instead of ϕ as

$$\phi_t = V(x,t)g(x,t), \tag{2.1}$$

$$\phi_x = -V(x,t)h(x,t)/u, \tag{2.2}$$

where $V(x,t)$ is a state variable, and $g(x,t)$ and $h(x,t)$ are other variables. From Eqs. (2.1) and (2.2), we can construct a state plane consisting of the relation between $V(x,t)$ and ϕ.

In Eqs. (2.1) and (2.2), if $g(x,t)$ and $h(x,t)$ are unity and not dependent upon the values of x and t, ϕ_t and ϕ_x construct a stationary solution in Eq. (1.1) that describes the waves traveling with a constant velocity of u, since they satisfy

$$\phi_t + u\phi_x = 0. \tag{2.3}$$

In this special case, the state plane is identical to the phase plane consisting of (ϕ, ϕ_t).

The solution ϕ to Eq. (1.1) is determined by designating x and t. We can then consider that ϕ_x and ϕ_t are determined through ϕ by designating x and t, if we keep phase planes of (ϕ, ϕ_x) and (ϕ, ϕ_t) in mind. Then ϕ_{xx} and ϕ_{tt} are expressed by

$$\phi_{xx} = \phi_x \frac{\partial \phi_x}{\partial \phi}, \quad \phi_{tt} = \phi_t \frac{\partial \phi_t}{\partial \phi}.$$

Integrating Eq. (1.1) with respect to ϕ and using the above relations, we obtain

$$\int \phi_x \, d(\phi_x) - \int \phi_t \, d(\phi_t) = \int [F(\phi) + G\phi_t - J_B] d\phi. \tag{2.4}$$

For convenience, we assume throughout this paper that $J_B \geqslant 0$, and that $F(\phi)$ repeats exactly same shape with a certain period of ϕ. The extended sine–Gordon system, the extended multiple sine–Gordon system, and so on satisfy the latter condition.

We may regard Eq. (1.1) as an equation on the transmission line for electromagnetic waves.[2,4] Then, the first term and the second term on the left in Eq. (2.4) denote the magnetic energy per unit length supplied to the transmission line by waves and the electric energy per unit length stored in it, respectively. The first term on the right denotes the energy stored per unit length in the distributed nonlinear nondissipative element $F(\phi)$, the second term energy dissipation per unit length due to G, and the last term the energy per unit length supplied to the waves by the distributed direct current J_B.

Differentiating Eq. (2.1) with respect to x and Eq. (2.2) with respect to t, and equating them, i.e., using the integrability condition Eq. (1.2), we obtain

$$\left[\frac{\partial V}{\partial \phi}\right]_t = \left[\frac{\partial V}{\partial \phi}\right]_x + \frac{u}{h(x,t)g(x,t)}\left(g_x + \frac{h_t}{u}\right), \tag{2.5}$$

where

$$\left[\frac{\partial V}{\partial \phi}\right]_t = \frac{V_x}{\phi_x}, \tag{2.6}$$

$$\left[\frac{\partial V}{\partial \phi}\right]_x = \frac{V_t}{\phi_t}. \tag{2.7}$$

If we integrate Eq. (2.6), keeping t constant, and integrate Eq. (2.7), keeping x constant with respect to ϕ, respectively, we obtain $V(x,t)$ for either case. Thus we can set

$$\left[\frac{\partial V}{\partial \phi}\right]_t = \left[\frac{\partial V}{\partial \phi}\right]_x = \frac{\partial V}{\partial \phi}. \tag{2.8}$$

Inserting Eq. (2.8) into (2.5), we obtain the conservative relation

$$g_x + h_t/u = 0. \tag{2.9}$$

In Eq. (2.9), the nonintegrability condition is not included in $g(x,t)$ and $h(x,t)$ in Eq. (2.9). This results from use of Eq. (1.2).

We insert in part Eqs. (2.1) and (2.2) into Eq. (2.4), taking account of Eqs. (2.6)–(2.8). As a result, we obtain the following equation on the state plane:

$$\beta(x,t)V(x,t)\frac{\partial V}{\partial \phi} - \zeta(x,t)V(x,t) = \eta(\phi), \tag{2.10}$$

where

$$\beta(x,t) = h^2(x,t)/u^2 - g^2(x,t),$$
$$\zeta(x,t) = (1/u)h_x + g_t + Gg(x,t),$$
$$\eta(\phi) = F(\phi) - J_B.$$

Since the relation in Eq. (2.8) is derived irrespective of the integrability condition (1.2), Eq. (2.10) may satisfy condition (2.3). Actually, if $\beta(x,t)$ becomes zero at certain ranges of x and t in Eq. (2.10), it is possible for singularity to exist there. Thus it is understood that the nonintegrability condition is normally included in Eq. (2.10).

From Eqs. (2.1), (2.2), and (2.6)–(2.8), we obtain the following two relations necessary to transform Eq. (2.10) into the expressions for real space:

$$V(x,t)\frac{\partial V}{\partial \phi} = \frac{u^2}{h^2(x,t)}\left[\phi_{xx} - \frac{1}{h(x,t)}h_x\phi_x\right] \tag{2.11}$$

$$= \frac{1}{g^2(x,t)}\left[\phi_{tt} - \frac{1}{g(x,t)}g_t\phi_t\right]. \tag{2.12}$$

By using Eqs. (2.1) and (2.2) for $V(x,t)$ and by using Eqs. (2.11) and (2.12) for $V(x,t)\partial V/\partial\phi$, respectively, Eq. (2.10) is transformed into the following expressions on real space:

$$\phi_{xx} + \chi_1(x,t)\phi_x = \frac{h^2(x,t)}{u^2}\frac{\eta(\phi)}{(x,t)}, \tag{2.13}$$

$$\phi_{tt} - \chi_2(x,t)\phi_t = g^2(x,t)\frac{\eta(\phi)}{\beta(x,t)}, \tag{2.14}$$

where

$$\chi_1(x,t) = \frac{h(x,t)}{u}\frac{\zeta(x,t)}{\beta(x,t)} - \frac{h_x}{h(x,t)},$$

$$\chi_2(x,t) = g(x,t)\frac{\zeta(x,t)}{\beta(x,t)} + \frac{g_t}{g(x,t)}.$$

We can regard Eqs. (2.13) and (2.14) as ordinary differential equations consisting of derivatives with respect to x and t, respectively, each of which is exactly equivalent to Eq. (1.1), if the functional forms of $g(x,t)$ and $h(x,t)$ are known. Then, we can first calculate, for instance, the solution for ϕ from either Eq. (2.13) or Eq. (2.14) using any of the conventional analytical techniques for an ordinary differential equation, and then we can calculate $V(x,t)$ from either Eq. (2.1) or Eq. (2.2). Equation (2.13) can be used to obtain the solution with changing x, where t is fixed, and Eq. (2.14) can be used to obtain the solution with changing t, where x is fixed.

III. WAVE COMPONENT

In order to see the meaning of $V(x,t)$, we divide ϕ_t and ϕ_x, not ϕ, into two components as follows:

$$\phi_t = (\phi_t)^{(t)} + (\phi_t)^{(r)}, \tag{3.1}$$

$$\phi_x = (\phi_x)^{(t)} + (\phi_x)^{(r)}, \tag{3.2}$$

where it is noted that each component on the right-hand side of the above expressions does not mean the derivative with respect to the subscript, t or x, but expresses the quantity as defined by

$$(\phi_t)^{(t)} = V(x,t),$$

$$(\phi_t)^{(r)} = [g(x,t) - 1](\phi_t)^{(t)},$$

$$(\phi_x)^{(t)} = -V(x,t)/u,$$

$$(\phi_x)^{(r)} = [h(x,t) - 1](\phi_x)^{(t)},$$

each of which belongs to the same state of ϕ. It is also noted that $(\phi_x)^{(t)}$ and $(\phi_t)^{(t)}$ satisfy

$$(\phi_t)^{(t)} + u(\phi_x)^{(t)} = 0. \tag{3.3}$$

From the analogy between Eqs. (3.3) and (2.3), we can expect $(\phi_x)^{(t)}$ and $(\phi_t)^{(t)}$ to construct a traveling wave component. However, since they interact with $(\phi_x)^{(r)}$ and $(\phi_t)^{(r)}$, they change their shape as either x or t is changed, insofar as we treat the nonstationary state. That is, the conservation relation on the (x,t) coordinates such as Eq. (2.3) is not satisfied in Eq. (3.3).

We introduce nonlinear coordinates $X(x,t)$ and $T(x,t)$ so as to satisfy a conservative relation between $(\phi_x)^{(t)}$ and $(\phi_t)^{(t)}$ in that coordinate system as follows:

$$(\phi_t)^{(t)} = \phi_T, \quad (\phi_x)^{(t)} = \phi_X. \tag{3.4}$$

Such a state as Eq. (3.4) in Eq. (3.3) is realized if there is a way to change $T(x,t)$ without changing $X(x,t)$ and also if there is a way to change $X(x,t)$ without changing $T(x,t)$ by changing appropriately the values of x and t. Since the functional form of $T(x,t)$ is apparently different from $X(x,t)$ throughout the whole range of x and t, the state as in Eq. (3.4) naturally exists. From Eqs. (2.1), (2.2), and (3.4), the relationships among $T(x,t)$, $g(x,t)$, and ϕ_t, and among $X(x,t)$, $h(x,t)$, and ϕ_x are then given by

$$\frac{\partial T}{\partial t} = \frac{1}{V(x,t)}\phi_t = g(x,t), \tag{3.5}$$

$$\frac{\partial X}{\partial x} = \frac{u}{V(x,t)}\phi_x = h(x,t). \tag{3.6}$$

Integrating Eq. (3.5) with respect to t and Eq. (3.6) with

respect to x, we obtain

$$T(x,t) = \int^t \frac{1}{V(x,t')} \frac{\partial \phi}{\partial t'} dt'$$
$$= \int^t g(x,t')dt' + T_0(x), \tag{3.7}$$
$$X(x,t) = -\int^x \frac{u}{V(x',t)} \frac{\partial \phi}{\partial x'} dx'$$
$$= \int^x h(x',t)dx' + X_0(t), \tag{3.8}$$

where $T_0(x)$ and $X_0(t)$ are arbitrary functions of x and t, respectively. We shall see in Sec. V that $X_0(t)$ and $T_0(t)$ are associated with singularities in the state plane. We also introduce a $\Xi(x,t)$ coordinate having a framework moving with a constant velocity u in $X(x,t)$ space, which is defined by

$$\Xi(x,t) = -u \int^{\phi} \frac{d\phi'}{V(x,t)} = X(x,t) - uT(x,t). \tag{3.9}$$

By using Eq. (3.9), Eq. (3.4) is also rewritten as

$$(\phi_t)^{(t)} = -u \frac{\partial \phi}{\partial \Xi}, \quad (\phi_x)^{(t)} = \frac{\partial \phi}{\partial \Xi}, \tag{3.10}$$

where we have used the relation

$$\left[\frac{\partial \phi}{\partial \Xi}\right]_X = \left[\frac{\partial \phi}{\partial \Xi}\right]_T = \frac{\partial \phi}{\partial \Xi},$$

which is derived in the same way as the relation in Eq. (2.8). Equation (3.10) indicates that $\phi(x,t)$ and $V(x,t)$ are a function of $\Xi(x,t)$. Thus we can replace them by $\phi(\Xi)$ and $V(\Xi)$, respectively. If the solution approaches the stationary state, $T(x,t)$ will approach t and $X(x,t)$ will approach x, because both $g(x,t)$ and $h(x,t)$ should approach unity. As a result, $\Xi(x,t)$ will approach $\xi = x - ut$. Thus we anticipate that $T(x,t)$, $X(x,t)$, and $\Xi(x,t)$ will be the extensions of t, x, and ξ, respectively. It will be shown in Sec. VI that these anticipations come true.

Throughout this paper, we assume that the solution approaches an asymptotically stationary solitary-wave solution as both $|x|$ and $|t|$ approach infinity. We refer to the components having (t) and (r) in the superscript as the (t) component and the (r) component of the waves, respectively. Then, the (r) component must disappear as both $|x|$ and $|t|$ approach infinity from the expression of its definition so that ϕ_t and ϕ_x may satisfy Eq. (2.3). This indicates that the (r) component is substantially confined in a restricted region in real space. Thus we expect that the (r) component will construct a standing wave and/or a decaying radiation.

IV. TRAVELING WAVE COMPONENT

We can rewrite Eq. (1.1) using the $\Xi(x,t)$ coordinates as

$$V_{\Xi} = -u[F'(\Xi) + GV(\Xi) - J_B]/(1-u^2), \tag{4.1}$$

where

$$F'(\Xi) = F(\phi(\Xi)) + F^{(r)}(\Xi),$$
$$F^{(r)}(\Xi) = -(\phi_x)_x^{(r)} + (\phi_t)_t^{(r)} + (\Xi_x - 1)V_{\Xi}/u$$
$$+ (\Xi_t + u)V_{\Xi} + G(\phi_t)^{(r)},$$

and $F^{(r)}(\Xi)$ is the contribution from the (r) component. Since there is an infinite number of combinations of x and t for a given value of either $X(x,t)$ or $T(x,t)$ in $\Xi(x,t)$ coordinates, which are, from their definition, different than ξ coordinates, the state of the system is not determined by designating only the value of $\Xi(x,t)$ that is determined by a pair of certain values of $X(x,t)$ and $T(x,t)$, and it is also needed to designate both the values of x and t. This means that $V(\Xi)$ and $\phi(\Xi)$ are in general not single-valued functions of the value of $\Xi(x,t)$.

The equation on the state plane can be written, from Eqs. (3.10) and (4.1), as

$$\frac{\partial V}{\partial \phi} = \frac{u^2}{1-u^2} \frac{F'(\Xi) + GV(\Xi) - J_B}{V(\Xi)}. \tag{4.2}$$

Equation (4.2) is another expression of Eq. (2.10) using the coordinate system of $\Xi(x,t)$, and is compared with the expression for stationary solitary waves in the extended sine–Gordon system.[15,16] We understand that the solution to Eq. (4.2) has singular points if the denominator and the numerator vanish simultaneously. Then, the singular points correspond to $|\Xi(x,t)| \to \infty$, where $V(\Xi) \to 0$.

In Eq. (4.2), we treat the first-order differential equation instead of the original second-order partial differential equation on real plane. As a result, the stationary solution is limited to a certain area connecting two singular points in the state plane,[2,15,16] although the solution expands infinitely in the conventional treatment on real space. From the above facts, we can investigate geometrically the exact solution in the state plane easier than doing this in real space. That is, once we know the field of directions in the state plane on Eq. (4.2), which is the direction solutions move with increasing (x,t), we can visually clarify the properties of solutions in the plane qualitatively almost with a freehand drawing, and can easily calculate the necessary solution quantitatively.[15,16] The other feature is that the initial and boundary conditions are imposed at the slopes $\partial V/\partial \phi$ of the singular points. Thus it enables us to set initially the well-defined exact stationary solitary wave solution in a finite area on the state plane, developing to the nonstationary state at the finite values of x and t, where the entire solutions still remain in a certain area in the state plane not so different from the solution for the original stationary state.

In Eq. (4.1) if both $|x|$ and $|ut|$ are much larger than unity, the (r) component may be disregarded, since the situation is regarded as the stationary state. It is confirmed in Sec. VI that $X_0(t)$ and $T_0(x)$ become constants under these circumstances. Thus Eq. (4.1) is rewritten as

$$(1-u^2)\phi_{\xi\xi} + uG\phi_{\xi} = -u[F(\phi) - J_B]. \tag{4.3}$$

Equation (4.3) is in agreement with the result derived directly by using ξ in Eq. (1.1). Then, Eq. (4.2) is rewritten as follows:

$$\frac{\partial V^{(s)}}{\partial \phi} = \frac{u^2}{1-u^2} \frac{F(\phi) + GV^{(s)}(\xi) - J_B}{V^{(s)}(\xi)}, \tag{4.4}$$

where (s) means the stationary state. The solution to Eq. (4.4) has a singular point as well as Eq. (4.2), if the denominator and the numerator vanish simultaneously. The singular point corresponds to the state at $|\xi| = \infty$ in Eq. (4.2), where $V^{(s)}(\xi) = 0$.

The singularity is determined from the linearized equation of Eq. (4.2), that is,

$$(1-u^2)\psi_{\xi\xi} + uG\psi_\xi = \left.\frac{dF}{d\phi}\right|_{\phi_0}\psi, \tag{4.5}$$

where $\psi = \phi - \phi_0$, and ϕ_0 is ϕ at the singular point. The solution to Eq. (4.5) can be written as

$$\psi = \psi_{0,\pm}\ \exp(k_{0,\pm}\xi), \tag{4.6}$$

where $\psi_{0,\pm}$ are constants, and

$$k_{0,\pm} = \pm a_{0,\pm}/u,$$

$$a_{0,\pm} = d_1 \mp d_2,$$

$$d_1 = \frac{u}{2}\left[\left(\frac{uG}{1-u^2}\right)^2 + \frac{4}{1-u^2}\left.\frac{dF}{d\phi}\right|_{\phi_0}\right], \tag{4.7}$$

$$d_2 = (u^2/2)[G/(1-u^2)]. \tag{4.8}$$

If $u<0$ we replace u by $-u$ throughout this paper. As a result, $a_{0,\pm}$ is replaced by $-a_{0,\mp}$, and then $k_{0,\pm}$ by $-k_{\mp}$. Thus $a_{0,\pm}$ is chosen so as to be always positive. The d_1^2 corresponds to the discriminant of Eq. (4.5). First, we assume $u^2<1$. If $(dF/d\phi)|_{\phi 0}>0$ in Eq. (4.7), the singularity is a saddle point, and if $(dF/d\phi)|_{\phi 0}<0$, the singularity is a node or a spiral point depending upon whether the sum of two terms in the square root in Eq. (4.7) is positive or negative. Next, we assume $u^2>1$. If $(dF/d\phi)|_{\phi 0}>0$ in Eq. (4.7), the singularity is a node or a spiral point depending upon whether the sum of two terms in the square root in Eq. (4.7) is positive or negative. The typical patterns of solution curves around the singular points are qualitatively depicted in Fig. 1 for $(dF/d\phi)|_{\phi 0}>0$ and Fig. 2 for $(dF/d\phi)|_{\phi 0}<0$. The patterns of solution curves around the singular points are characterized by the properties of the field of directions as follows.

(i) The solution curves are perpendicular to the ϕ axis, that is, the line for

$$V^{(s)}(\xi) = 0 \tag{4.9}$$

when the solution curves cross the line, because $\partial V^{(s)}/\partial\phi$ is then infinity there from Eq. (4.4).

(ii) The solution curves are parallel to the ϕ axis when they cross the curve for

$$V^{(s)}(\xi) = [J_B - F(\phi)]/G, \tag{4.10}$$

because $\partial V^{(s)}/\partial\phi$ is then zero from Eq. (4.4).

(iii) The curves for Eqs. (4.9) and (4.10) divide the state plane into four areas with different sign combinations for $V^{(s)}(\xi)$ and ϕ as depicted in Figs. 1 and 2.

(iv) The ϕ component of the field of directions increases monotonically as the solution curves leave the ϕ axis as seen from the first expression in Eq. (3.10) for the stationary state, in which $(\phi_t)^{(t)}$ and $\Xi(x,t)$ are replaced by $V^{(s)}(\xi)$ and ξ, respectively.

(v) The $V^{(s)}(\xi)$ component of the field of directions increases monotonically as the solution curves leave the curves described by Eq. (4.10) as seen from the following expression

$$V_\xi^{(s)} = -u[F(\phi) + GV^{(s)}(\xi) - J_B]/(1-u^2), \tag{4.3'}$$

which is obtained from Eq. (4.3).

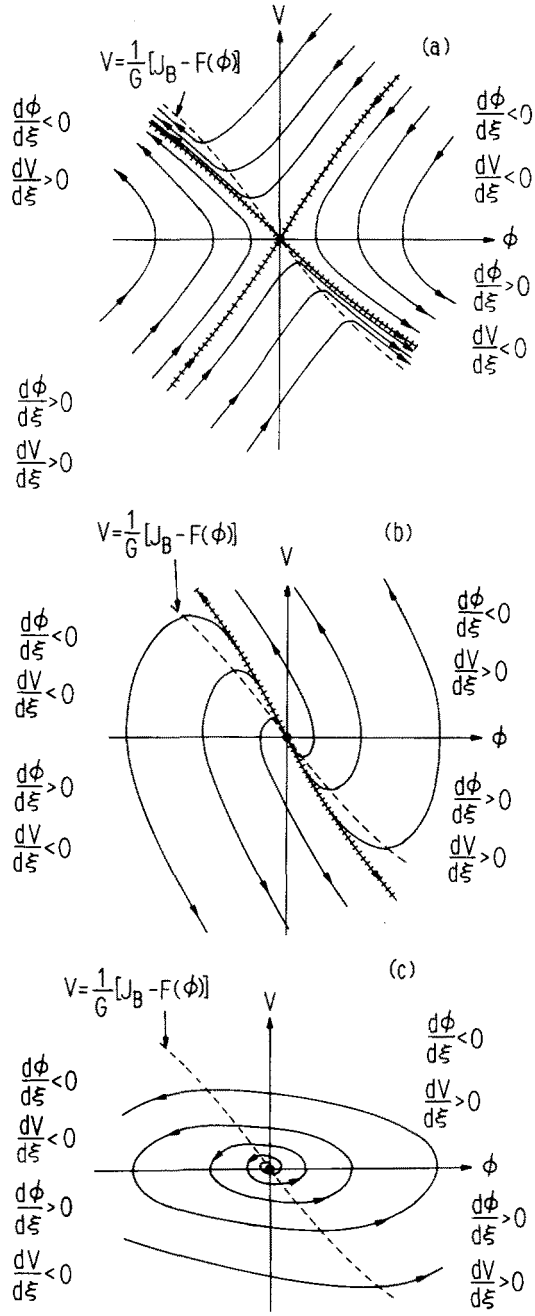


FIG. 1. The patterns of solution curves around singular points on the stationary solitary wave when $dF/d\phi>0$. The patterns of solution curves are classified by hatched curves ⟼⟼. The dot denotes the singular point and the arrowhead on the curve denotes the field of directions: (a) a saddle point, (b) a node, (c) a spiral point.

(vi) The solution curves have a certain slope θ expressed by

$$\theta = \tan^{-1}[u^2/(1-u^2)]G,$$

at infinitely large absolute values of V.

(vii) Since $F(\phi)$ is a periodic function, $(dF/d\phi)|_{\phi 0}$ is changed alternatively from plus to minus with increasing ϕ_0. Accordingly, the saddle point and the node or the spiral point appear alternatively in a certain range of J_B.

(viii) If J_B is larger than the maximum value of $|F(\phi)|$,

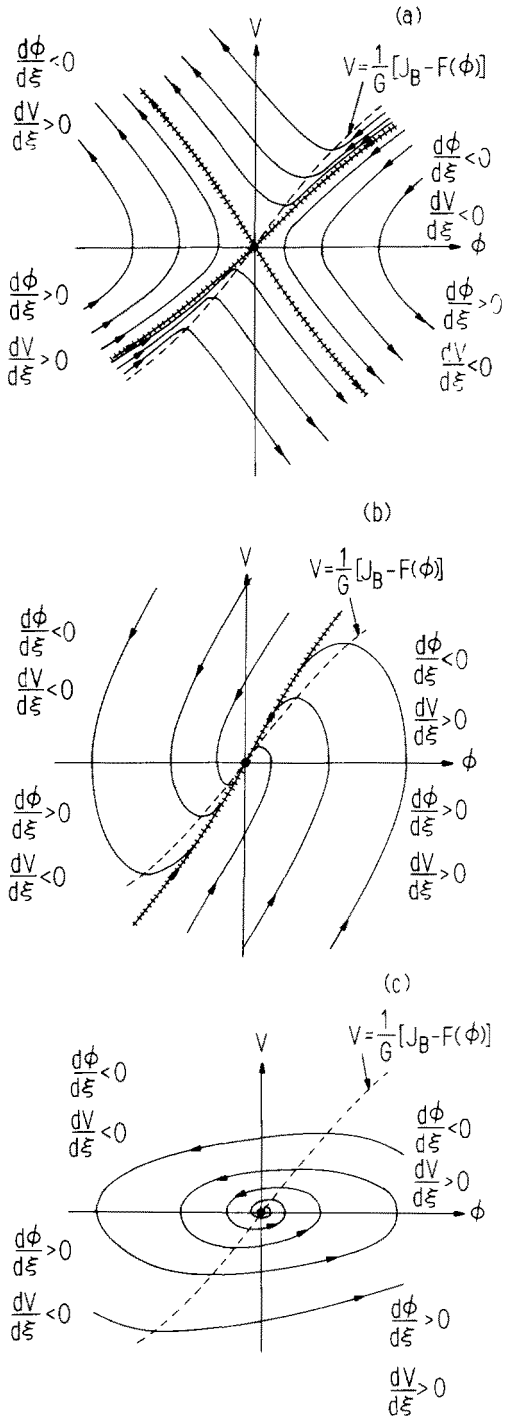


FIG. 2. The patterns of solution curves around singular points on the stationary solitary wave when $dF/d\phi<0$. The patterns of solution curves are classified by hatched curves +++++ . The dot denotes the singular point and the arrowhead on the curve denotes the field of directions: (a) a saddle point, (b) a node, (c) a spiral point.

there exists no stationary solitary-wave solution, since every singular point disappears.

Throughout this paper, we also assume that $(dF/d\phi)_{\phi 0}$ at $|\xi|=\infty$ is positive, and that $u^2<1$ since the solution for $u^2>1$ does not construct the stable stationary wave solution as far as $J_B\neq 0$ and $G\neq 0$. Even if $J_B=G=0$, the solution for $u^2>1$ is outside our interest on account of periodic radi-

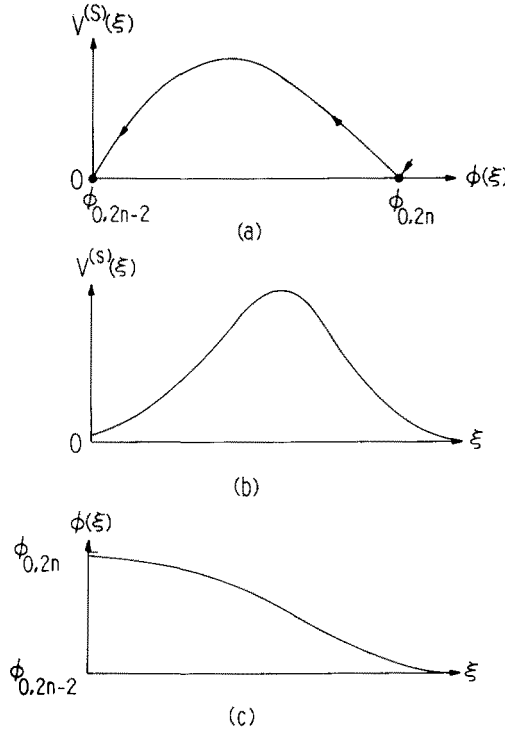


FIG. 3. Stationary solitary-wave solution: (a) solution in state plane, (b) $V^{(s)}(\xi)$ vs ϕ, (c) $\phi(\xi)$ vs ξ.

ation solutions. Here ϕ_0 is distinguished by whether it is concerned with an even or odd singular point counted from the origin, i.e., $\phi_0=\phi_{0,2n}$ or $\phi_{0,2n-1}$, where n is an integer denoting the position of the singular point. Here, we can choose the stable singular point $(\phi,V^{(s)}(\xi))=(\phi_{0,2n},0)$ to be a saddle point that corresponds to the vacuum state. The point $(\phi_{0,2n-1},0)$ is naturally a spiral point or a node (if $G=J_B=0$, the singularity is a vortex). Thus the stationary solitary wave solution can be constructed by finding a solution that starts at a saddle point, ending at another saddle point by making reference to Fig. 1(a), for instance, as depicted in Fig. 3(a). The slopes at the saddle point, $(\partial V^{(s)}/\partial\phi)|_{\phi 0}$, are significant since they are concerned with initial and boundary conditions, and are written as

$$\left.\frac{\partial V^{(s)}}{\partial\phi}\right|_{\phi_{0,2n-2}}=a_{0,-},\quad \left.\frac{\partial V^{(s)}}{\partial\phi}\right|_{\phi_{0,2n}}=-a_{0,+}, \tag{4.11}$$

if the solution is in the region between $(\phi_{0,2n-2},0)$ and $(\phi_{0,2n},0)$. The arrowhead on the curve in Fig. 3(a) denotes the field of directions, where we assume $u>0$, which means the forward wave. If $u<0$, which means the backward wave, it is a straightforward matter to show from Eqs. (3.10) and (4.3′) that the arrowhead points in the opposite direction to Fig. 3(a). The effects of x and t on the field of directions are understood through $\xi=x-ut$. Accordingly, if we observe the forward wave, increasing x has the same effect as decreasing t, and if we observe the backward wave, increasing x has the same effect as increasing t. From Fig. 3(a), we understand that if we observe the forward wave solution in the state plane, the solution leaves the saddle point $(\phi,V^{(s)}(\xi))=(\phi_{0,2n},0)$ toward the upper left with increasing ξ from $-\infty$, and finally reaches the saddle point

$(\phi_{0,2n-2},0)$ from the upper right at $\xi \to +\infty$. The relation between $V^{(s)}(\xi)$ and ξ is depicted in Fig. 3(b), and the relation between $\phi(\xi)$ and ξ is depicted in Fig. 3(c), respectively.

Consider the more generalized extended Klein–Gordon system in the $(X(x,t), T(x,t))$ coordinates described by

$$\phi_{XX} - \phi_{TT} - F'(X,T) = G\phi_T - J_B. \tag{4.12}$$

It is understood that Eq. (4.1) is in agreement with the expression for traveling waves in Eq. (4.12). Accordingly, it may be considered that $(\phi_t)^{(t)}$ constructs the traveling wave having a constant velocity u in an extended Klein–Gordon system with $F'(\Xi)$ in the $\Xi(x,t)$ coordinates instead of $F(\phi)$ in the ξ coordinates.

V. PROPERTIES OF $g(x,t)$ AND $h(x,t)$

If there is no cause to disturb a solitary wave anywhere, the wave keeps the stationary state, where $g(x,t) = h(x,t) = 1$. Given the proper causal conditions, the wave is changed from the stationary state when approaching its source, and there appears an (r) component, so that both $g(x,t)$ and $h(x,t)$ deviate from unity. Accordingly, we may regard their deviation as a measure of the disturbance of the wave. The above wave constructs an incoming wave toward the disturbance source. On the other hand, an outgoing wave from the source may exist, approaching asymptotically the stationary state while losing the (r) component.

We expand $g(x,t)$ using arbitrary functions $g_1(x)$ and $g_2(t)$, and expand $h(x,t)$ using arbitrary functions $h_1(x)$ and $h_2(t)$ as

$$g(x,t) = \sum_i b_i^{(1)}[g_1(x)]^i + \sum_j b_j^{(2)}[g_2(t)]^j + \sum_i \sum_j b_{ij}[g_1(x)]^i[g_2(t)]^j, \tag{5.1}$$

$$h(x,t) = \sum_i d_i^{(1)}[h_1(x)]^i + \sum_j d_j^{(2)}[h_2(t)]^j + \sum_i \sum_j d_{ij}[h_1(x)]^i[h_2(t)]^j, \tag{5.2}$$

where i and j are integers, and $b_i^{(1)}, b_j^{(2)}, b_{ij}, d_i^{(1)}, d_j^{(2)}$, and d_{ij} are constants. The derivatives of Eq. (5.1) with respect to x and of Eq. (5.2) with respect to t are expressed by

$$g_x = \sum_i \left\{ i \left[b_i^{(1)} + \sum_j b_{ij}[g_2(t)]^j \right] [g_1(x)]^{i-1} \right\} \frac{dg_1}{dx}, \tag{5.3}$$

$$h_t = \sum_j \left\{ j \left[d_j^{(2)} + \sum_i d_{ij}[h_1(x)]^i \right] [h_2(t)]^{j-1} \right\} \frac{dh_2}{dt}, \tag{5.4}$$

respectively. It is seen from Eqs. (5.3) and (5.4) that g_x and h_t are proportional to dg_1/dx and dh_2/dt, respectively. Since $|x|$ and $|t|$ approach infinity at the same time, $g(x,t)$ and $h(x,t)$ also approach unity, respectively. If only x approaches infinity, both $g(x,t)$ and $h(x,t)$ should become at most only a function of t, and if only t approaches infinity, they should become at most only a function of x. That is, $dg_1/dx \to 0$ and $dh_1/dx \to 0$ as $|x| \to \infty$, and $dg_2/dt \to 0$ and $dh_2/dt \to 0$ as $|t| \to \infty$. We set then for convenience

$$g_1(|x|) = h_1(|x|) = 1 \quad (|x| \to \infty),$$
$$g_2(|t|) = h_2(|t|) = 1 \quad (|t| \to \infty).$$

Thus if $|x|$ approaches infinity, while t is kept to a certain value, g_x approaches zero from Eq. (5.3). We see then that h_t should also approach zero from Eq. (2.9). This means from Eq. (5.4) that at an arbitrary value of t, we obtain the condition

$$\frac{dh_2}{dt} = 0, \tag{5.5}$$

or

$$\sum_j j \left[d_j^{(2)} + \sum_i d_{ij} \right] [h_2(t)]^{j-1} = 0. \tag{5.6}$$

However, we cannot adopt Eq. (5.6) because t must be fixed at a particular value then. Since Eq. (5.5) is adopted for any value of x, $h(x,t)$ must be only a function of x, i.e.,

$$h(x,t) = h(x). \tag{5.7}$$

If $|t|$ approaches infinity, h_t approaches zero from Eq. (5.4). From Eq. (2.9), g_x should also approach zero at an arbitrary value of x. From Eq. (5.3), we obtain

$$\frac{dg_1}{dx} = 0. \tag{5.8}$$

Since Eq. (5.8) is adopted for any value of t in Eq. (5.1), $g(x,t)$ must be only a function of t, i.e.,

$$g(x,t) = g(t). \tag{5.9}$$

We assume that the effect of the disturbance on the wave is most dominant at the origin, i.e., $(x,t) = (0,0)$. Then, the disturbance keeps weakening with increasing $|x|$ and $|t|$ at the same time so that the wave tends toward the pure traveling wave. This indicates that the (r) component should be confined effectively in a restricted region around the origin. Next, with increasing only $|x|$ (or $|t|$) more than a certain value, the (r) component keeps decreasing toward zero whatever the value of $|t|$ (or $|x|$). Thus at the infinitely large value of $|x|$ (or $|t|$) the (r) component may be disregarded compared with the traveling wave component in Eq. (4.1). In this situation, $F'(\Xi)$ is replaced by $F(\phi)$ in Eq. (4.2), which is rewritten as

$$V_X = -u[F(\phi) + GV(\Xi) - J_B]/(1-u^2) \quad (|x| \gg 1), \tag{5.10}$$

$$V_T = u^2[F(\phi) + GV(\Xi) - J_B]/(1-u^2) \quad (|ut| \gg 1), \tag{5.11}$$

where we have used the relation between Eqs. (3.4) and (3.10). However, it is noted that the state at $|x| \gg 1$ or at $|ut| \gg 1$ is generally different from the stationary state in Eq. (4.1), because even if $|x|$ approaches infinity, the state should be influenced by $g(t)$ and $X_0(t)$, and even if $|ut|$ approaches infinity, the state should be influenced by $h(x)$ and $T_0(x)$. That is, we cannot simply replace $\Xi(x,t)$ by ξ in this situation. Thus we expect that our approach will present us with more extensive information than the conventional theory even in the limit of small amplitude of the wave. Under such a condition, we can eliminate $F'(\Xi) = F(\phi)$ from Eqs. (2.10) and (4.2). As a result, we obtain the following equa-

tions for determining the functional forms of $g(t)$ and $h(x)$:

$$[g^2(t)-1]\left.\frac{\partial V}{\partial \phi}\right|_{\phi_0} - G[g(t)-1] + \frac{dg}{dt} = 0 \quad (|x| \to \infty), \tag{5.12}$$

$$\frac{1}{u}[h^2(x)-1]\left.\frac{\partial V}{\partial \phi}\right|_{\phi_0} - \frac{dh}{dx} = 0 \quad (|ut| \to \infty). \tag{5.13}$$

Here we note that if a sign of u is selected in Eq. (4.2) it is impossible to choose the region in which the forward wave and the backward wave component can coexist, because $V(\Xi)$ then represents only a kind of traveling wave component determined by the sign of u there. Accordingly, if we wish to treat two components with different signs, we must set them in separate regions respective to the signs. We first direct our attention only to the forward wave, i.e., $u>0$. We can then set such a region either for $t<0$ and $x<0$ or for $t>0$ and $x>0$ in physical reality. Since $(\partial V^{(s)}/\partial\phi)|_{\phi 0}$ is independent of both x and t from Eq. (4.11), $(\partial V/\partial\phi)|_{\phi 0}$ should only approach a function of t as $|x| \to \infty$. Thus we can set $(\partial V/\partial\phi)|_{\phi 0}$ in Eq. (5.12) by introducing $\omega_{\pm}(t)$ to (4.11) as

$$\left.\frac{\partial V}{\partial \phi}\right|_{\phi_0} = \pm \omega_{\pm}(t) a_{0,\mp} \quad (|x| \to \infty), \tag{5.14}$$

where $\omega_{\pm}(t)$ are positive, and the upper and the lower signs mean that the values of t are positive and negative, respectively. On the other hand, since $(\partial V/\partial\phi)|_{\phi 0}$ should be at most only a function of x since $|t| \to \infty$, we can also set $(\delta V/\partial\phi)|_{\phi 0}$ in Eq. (5.13) by introducing $\gamma_{\pm}(x)$ to Eq. (4.11) as

$$\left.\frac{\partial V}{\partial \phi}\right|_{\phi_0} = \mp \gamma_{\pm}(x) a_{0,\pm} \quad (|ut| \to \infty), \tag{5.15}$$

where $\gamma_{\pm}(x)$ are positive, and the upper and the lower signs mean that t is positive and negative, respectively.

As $|t|$ approaches infinity in Eq. (5.14), $\omega_{\pm}(t)$ approaches unity, since we always observe the far side of the wave measured from the origin in this process since the condition $|x| \gg |ut|$ is always preserved, that is, the situation should be in a stationary state. For convenience, we assume that $\omega_{\pm}(t)$ are equal to unity irrespective of the value of t as an ideal case, though we will see later that such an assumption does not satisfy the condition at $t=0$ in the strict sense of the word. The $g(t)$ for $t>0$ and $t<0$, $g_{\pm}(t)$, can then be solved from Eq. (5.10) as

$$g_{\pm}(t) = \begin{cases} \pm [B_{\pm} \tanh(B_{\pm} a_{0,\mp} t) - A_{\pm}], & (5.16) \\ \pm [B_{\pm} \coth(B_{\pm} a_{0,\mp} t) - A_{\pm}], & (5.17) \end{cases}$$

respectively, where $A_{\pm} = G/2a_{0,\mp}$ and $B_{\pm} = 1 \pm A_{\pm}$. We will see in Sec. IX that Eq. (5.17) is significant since it can be used as a means to obtain the exact solution for the soliton–antisoliton interaction by means of a suitable coordinate transformation.

Notice that if $B_- \leqslant 0$ in Eq. (5.16) or Eq. (5.17), $g_-(t)$ does not satisfy the initial condition so that $g_-(t)$ may not converge to unity with decreasing t to minus infinity. Thus we may consider that the coexistence region of solutions is limited so as to satisfy the condition where $0 \leqslant A_- < 1$.

On the other hand, even if $|x|$ approaches infinity in Eq. (5.15), we cannot say that $\gamma_{\pm}(x)$ approaches unity, since we always observe the origin side of the wave in this process because of $|ut| \gg |x|$ in Eq. (5.13), and the wave continues to be influenced by disturbances around $x=0$. The $h(x)$ for $t>0$ and $t<0$, $h_{\pm}(x)$, can be solved without any assumption in Eq. (5.13), and are written as

$$h_{\pm}(x) = \begin{cases} (\pm)\tanh\left[\dfrac{a_{0,\pm}}{u}\displaystyle\int_0^x \gamma_{\pm}(x')dx'\right], & (5.18) \\ (\pm)\coth\left[\dfrac{a_{0,\pm}}{u}\displaystyle\int_0^x \gamma_{\pm}(x')dx'\right], & (5.19) \end{cases}$$

where the $(\pm)$ signs are applied to $x>0$ and $x<0$, respectively, and the other $\pm$ signs are applied to $t>0$ and $t<0$, respectively.

Finally, if we only observe the backward wave, we can set $u<0$ for the region $t<0$ and $x>0$ and for the region $t>0$ and $x<0$. Then, we still obtain the same expressions as Eqs. (5.16)–(5.19) by replacing u by $-u$ in Eqs. (5.12) and (5.13). Thus the phenomenon on the backward wave becomes identical to the one on the forward wave.

VI. NONLINEAR COORDINATES ON SINGULAR POINTS

At this stage, Eqs. (3.7) and (3.8) can be rewritten as

$$T(x,t) = \int^t g(t')dt' + T_0(x), \tag{6.1}$$

$$X(x,t) = \int^x h(x')dx' + X_0(t). \tag{6.2}$$

However, the functional forms of $T_0(x)$ and $X_0(t)$ are not yet known. The solution to Eq. (4.2) has singular points at $|x| = \infty$ and $|t| = \infty$. From Eq. (3.7), we can define the following quantity at $|t| \to \infty$:

$$T_{0,\gamma}(x) = \mp \lim_{t \to \pm\infty} \int_{\pm\infty}^t \frac{1}{V(x,t')}\frac{\partial\phi}{\partial t'}dt'. \tag{6.3}$$

Since $V(x,t)$ converges to zero as $t \to \pm\infty$, Eq. (6.3) can be developed as

$$T_{0,\gamma}(x) = \lim_{t \to \pm\infty} \frac{\partial\phi}{\partial V} = \mp \frac{1}{\gamma_{\pm}(x) a_{0,\pm}}, \tag{6.4}$$

which is the reciprocal of the slope at the singular point at $t \to \pm\infty$. Next, from Eq. (3.8), we can define the following quantity at $|x| \to \infty$:

$$X_{0,\omega}(t) = \pm u \lim_{x \to \pm\infty} \int_{\pm}^x \frac{1}{V(x',t)}\frac{\partial\phi}{\partial x'}dx'. \tag{6.5}$$

Equation (6.5) can be developed as

$$X_{0,\omega}(t) = -\lim_{x \to \pm\infty} u\frac{\partial\phi}{\partial V} = \mp \frac{u}{\omega_{\pm}(t) a_{0,\mp}}, \tag{6.6}$$

which is proportional to the reciprocal of the slope at the singular point at $x \to \pm\infty$. We finally define $\Xi_{0,\infty}(x,t)$ from Eqs. (3.9), (6.4), and (6.6) as

$$\Xi_{0,\infty}(x,t) = \pm u\left[\frac{1}{\gamma_{\pm}(x) a_{0,\pm}} - \frac{1}{\omega_{\pm}(t) a_{0,\mp}}\right]. \tag{6.7}$$

Since $T_{0,\gamma}(x)$ and $X_{0,\omega}(t)$ are derived from the definition of $T(x,t)$ and $X(x,t)$, respectively, $T_{0,\gamma}(x)$, $X_{0,\omega}(t)$, and

$\Xi_{0,\infty}(x,t)$ can be included in $T_0(x)$, $X_0(t)$, and $\Xi(x,t)$, respectively.

If we need to choose the function form of $g_{\pm}(t)$ as the same kind of expression as Eq. (5.17), we see that $g_{\pm}(t) \to \infty$ as $t \to \pm\infty$, and then we obtain $T(x,t) \to \mp\infty$, because of

$$\lim_{t\to\pm 0} T(x,t) = \pm \lim_{t\to\pm 0} (\ln|t|)/a_{0,\mp} \to \mp\infty, \tag{6.8}$$

that is, the state at $t=0$ describes a singular point in Eq. (4.2). On the other hand, if we need to choose Eq. (5.19) for $h_{\pm}(x)$, we see that $h_{\pm}(x) \to \infty$ as $x \to \pm 0$, and then we obtain $X(x,t) \to \mp\infty$. The state at $x=0$ describes a singular point. Thus we can also define expressions such as Eqs. (6.4), (6.6), and (6.7) for the state at $t = \pm 0$ and $x = \pm 0$ as follows:

$$T_{0,0}(x) = \pm \lim_{t\to\pm 0} \int_{\pm 0}^{t} \frac{1}{V(x,t')} \frac{\partial\phi}{\partial t'} dt'$$

$$= \lim_{t\to\pm 0} \frac{\partial\phi}{\partial V}, \tag{6.9}$$

which is the reciprocal of the slope at the singular point at $t = \pm 0$. At $x=0$,

$$X_{0,0}(t) = \mp u \lim_{x\to\pm 0} \int_{\pm 0}^{x} \frac{1}{V(x',t)} \frac{\partial\phi}{\partial x'} dx'$$

$$= -\lim_{x\to\pm 0} u \frac{\partial\phi}{\partial V}, \tag{6.10}$$

which is proportional to the reciprocal of the slope at the singular point at $x = \pm 0$. We can also define $\Xi_{0,0}(x,t)$ from Eqs. (3.9), (6.9), and (6.10) as follows:

$$\Xi_{0,0}(x,t) = X_{0,0}(t) - uT_{0,0}(x). \tag{6.11}$$

Equations (6.9)–(6.11) can be also included in $T_0(x)$, $X_0(t)$, and $\Xi(x,t)$, respectively. There is another kind of singular point at the position satisfying $\beta(x,t)=0$ as described in detail in Sec. XI, the effect of which can be included in $T_0(x)$ and $X_0(t)$ as well.

If we take such effects as $T_0(x)$ and $X_0(t)$ into consideration in Eqs. (6.1) and (6.2), respectively, we see that $V(\Xi)$ is in general not a single-valued function of ϕ on account of their variation. However, in the stationary state, $T_0(x)$ and $X_0(t)$ become constants since $\gamma_{\pm}(x)$ and $\omega_{\pm}(t)$ should approach constants as $|x| \to \infty$ and $|t| \to \infty$, respectively. Moreover, $\beta(x,t)$ also approaches $1/u^2 - 1$, being a constant. Thus $\Xi(x,t)$ is replaced by ξ. Then $V(\xi)$ becomes the single-valued function of ξ, i.e., of ϕ. As a result, $V(\Xi)$ is written as $V(\Xi) = V^{(s)}(\xi)$.

VII. CONSTRUCTION OF $V(\Xi)$

A. Soliton–antisoliton interaction

If we refer to the kink solution of ϕ decreasing with increasing x as the soliton as described, for instance, in Fig. 3(c), the antisoliton naturally becomes the kink solution of ϕ increasing with increasing x.

Consider the properties of $V(\Xi)$ around a singular point in adopting Eq. (5.18) for $X(x,t)$ in Eq. (6.2) and Eq. (5.17) for $T(x,t)$ in Eq. (6.1). As t approaches plus (or minus) infinity, $T(x,t)$ approaches plus (or minus) infinity

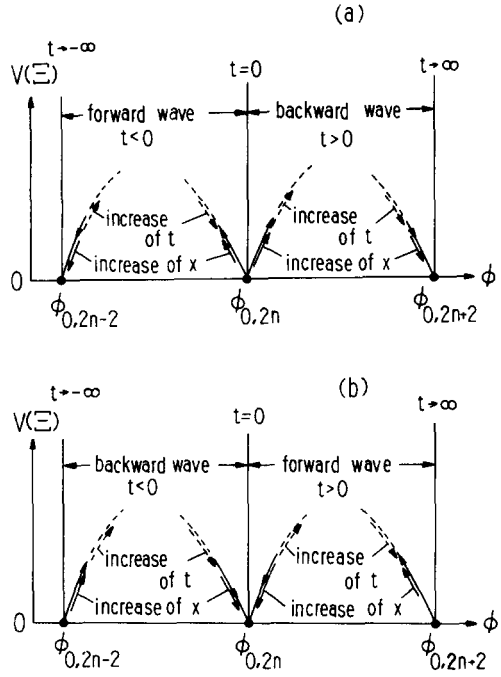


FIG. 4. Relationship between $V(x,t)$ and ϕ in the two soliton interaction, where the field of directions are denoted by the arrowhead on the curve: (a) combination of the forward wave for $t<0$ and the backward wave for $t>0$, (b) combination of the backward wave for $t<0$ and forward wave for $t>0$.

and then the solution to Eq. (4.2) approaches a singular point, that is, zero. On the other hand, it is seen from Eq. (6.8) that as t approaches zero from the side of $t<0$, $T(x,t)$ approaches $+\infty$, and as t approaches zero from the side of $t>0$, $T(x,t)$ approaches $-\infty$. Then, the solution to Eq. (4.2) again approaches a singular point. Moreover, since $g_{\pm}(t) \to \infty$ and $V(\Xi) \to 0$ as $t \to \pm 0$, ϕ_t in Eq. (2.1) can keep a finite value at $t=0$, where we choose the value of ϕ as $\phi_{0,2n}$. From the above facts and from analogy with the stationary solution in Fig. 3(a), the relations between $V(\Xi)$ and ϕ can be depicted qualitatively in Fig. 4, if $V(\Xi)$ is chosen so as to connect the singular points, $(\phi_{0,2n-2},0)$ and $(\phi_{0,2n+2},0)$, in the state plane. Then, we have to treat two soliton solutions: the one is constructed in the region between $\phi_{0,2n-2}$ and $\phi_{0,2n}$, and the other in the region between $\phi_{0,2n}$ and $\phi_{0,2n+2}$. For convenience, insofar as we do not give a special notice, we assume that the singular points at $|T(x,t)| = \infty$ are saddle points. From the analogy of the stationary solution $V^{(s)}(\xi)$, we understand that $V(\Xi)$ is positive everywhere, as shown in Fig. 4. If there exist a forward wave for $t<0$ and a backward wave for $t>0$, as depicted in Fig. 4(a), then $\phi_x < 0$ for $t<0$, $\phi_x = 0$ at $t=0$, and $\phi_x > 0$ for $t>0$ from Eq. (2.2). On the other hand, if there exist a backward wave for $t<0$ and a forward wave for $t>0$, as depicted in Fig. 4(b), then $\phi_x > 0$ for $t<0$, $\phi_x = 0$ for $t=0$, and $\phi_x < 0$ for $t>0$ from Eq. (2.2). Note that ϕ_x at $x=0$ is always zero since $h_{\pm}(0) = 0$ there. Existence of the respective regions of $\phi_x < 0$ and $\phi_x > 0$ means that we treat both the soliton and the antisoliton. In this case, the singular point at $|T_{\pm}(x,t)| = \infty$ which is the node and spiral point can also be considered [see Eq. (11.2)]. If the singular point is a spiral point, the above explanation for ϕ_x is amended to include the radiation component around the singular points

as suspected from the characteristics in Figs. 1(c) and 2(c). If the singular points are nodes, then there exist two cases: one is where no radiation component exists as in the case for the saddle points and the other is where extremely attenuated radiation associated with a region of the opposite sign of ϕ_t close to the singular points exists. Thus it is understood that the forward wave and the backward wave are associated with the soliton and the antisoliton, respectively, irrespective of whether the radiation component is included or not.

In Fig. 4, the direction the solution moves with increasing x or t is determined based upon the field of directions denoted by the arrowhead on the solution curves associated with the increasing direction of $\Xi(x,t)$. Then, the direction the solution moves with increasing $X(x,t)$ or $T(x,t)$ is determined based upon the following expression:

$$\Xi(x,t) = X(x,t) \mp |u| T(x,t), \tag{7.1}$$

where the upper sign is used for the forward wave and the lower sign for the backward wave. The field of directions may be expressed by the slope in the relation between $V(\Xi)$ and ϕ. The following expressions exist according to whether we use the $(X(x,t),T(x,t))$ coordinates or the (x,t) coordinates:

$$\frac{\partial V}{\partial T}\left(\frac{\partial \phi}{\partial T}\right)^{-1} \text{ or } \frac{\partial V}{\partial t}\left(\frac{\partial \phi}{\partial t}\right)^{-1}$$

and

$$\frac{\partial V}{\partial X}\left(\frac{\partial \phi}{\partial X}\right)^{-1} \text{ or } \frac{\partial V}{\partial x}\left(\frac{\partial \phi}{\partial x}\right)^{-1}.$$

Using the relations between $T(x,t)$ and t in Eq. (3.5) and between $X(x,t)$ and x in Eq. (3.6), we obtain the following relations:

$$\frac{\partial V}{\partial T}\left(\frac{\partial \phi}{\partial T}\right)^{-1} = \frac{\partial V}{\partial t}\left(\frac{\partial \phi}{\partial t}\right)^{-1} = \frac{\partial V}{\partial \phi},$$

$$\frac{\partial V}{\partial X}\left(\frac{\partial \phi}{\partial X}\right)^{-1} = \frac{\partial V}{\partial x}\left(\frac{\partial \phi}{\partial x}\right)^{-1} = \frac{\partial V}{\partial \phi}.$$

Thus we understand that the field of directions associated with the $\Xi(x,t)$ coordinate is in agreement with the one associated with the ξ coordinate. Therefore, we can use the sign combinations used for the stationary solution as depicted in Fig. 1(a) in determining the field of directions even for such a nonstationary state. Accordingly, it can be said in Fig. 4 that, for instance, increasing t from $-\infty$ $V(\Xi)$ leaves the singular point $(\phi_{0,2n-2},0)$ toward the upper right, approaching the singular point $(\phi_{0,2n},0)$ from the upper left as t approaches -0, and that $V(\Xi)$ leaves the singular point $(\phi_{0,2n},0)$ toward the upper right by increasing t from $+0$, and finally reaching the singular point $(\phi_{0,2n+2},0)$ from the upper left at $t = +\infty$. From the above process, it can be seen that the direction the solution moves with increasing t does not depend upon whether we are observing the forward wave or the backward wave.

Consider a situation where t is negative and is fixed at a value. If x is negative, $V(\Xi)$ constructs the forward wave in Fig. 4(a). That is, it leaves the singular point $(\phi_{0,2n},0)$ toward the upper left with increasing x from $-\infty$, and reaching the position corresponding to $x = 0$, where the value of ϕ shows a minimum before reaching $\phi_{0,2n-2}$. If x is positive, then we treat the backward wave in Fig. 4(b). Increasing x from zero, the solution keeps moving toward the right, finally reaching the point $(\phi_{0,2n},0)$ at $x = +\infty$ from the upper left. Thus the state at $x = +\infty$ becomes the same as the state at $x = -\infty$. Next, consider the situation where t is positive and fixed at a value. If x is negative, $V(\Xi)$ constructs a backward wave in Fig. 4(a). That is, it leaves the singular point $(\phi_{0,2n},0)$ at $x = -\infty$ toward the upper right with increasing x from $-\infty$, finally reaching the position corresponding to $x = 0$, where the value of ϕ exhibits a maximum before reaching the value being $\phi_{0,2n+2}$. If x is positive, we are to treat the forward wave in Fig. 4(b). With increasing x from zero, $V(\Xi)$ moves toward the left, approaching the singular point $(\phi_{0,2n},0)$ from the upper right, and finally reaching it at $x = +\infty$. In the above processes, the maximum or the minimum value of at $x = 0$ depends upon the value of t, that is, when $ut \ll -1$, the minimum value of ϕ is very close to $\phi_{0,2n-2}$, moving toward the right with increasing t. When $t \gg 1$, the maximum value of ϕ is very close to $\phi_{0,2n+2}$, moving toward the left with decreasing t. When $t = 0$, the value of ϕ becomes $\phi_{0,2n}$ irrespective of the value of x.

In Fig. 4(a), combining the forward wave that evolves from $\phi_{0,2n-2}$ to $\phi_{0,2n}$ for $t \leqslant 0$ with the backward wave at $t = 0$ that evolves from $\phi_{0,2n}$ to $\phi_{0,2n+2}$ for $t \geqslant 0$ at a certain value of x, we observe the situation at the position that the solitary forward wave coming from $x \ll -1$ is reflected at the open end of the semi-infinitely long system, where $x = 0$, coming back to $x \ll -1$. This is because the phase velocity $w(x,t)$, which is defined by

$$w(x,t) = \frac{dx}{dt} = -\frac{\phi_t}{\phi_x}, \tag{7.2}$$

and which also corresponds to the line impedance in the equivalent transmission line of the system, is infinity at $x = 0$ irrespective of the value of t, where it is noted in the equivalent transmission line that ϕ_t and $-\phi_x$ denote the voltage between the line and the current along the line, respectively. In Fig. 4(b), combining the backward wave solution evolving from $\phi_{0,2n-2}$ to $\phi_{0,2n}$ for $t \leqslant 0$ with the forward wave one developed from $\phi_{0,2n}$ to $\phi_{0,2n+2}$ for $t \geqslant 0$ at a position, we observe the wave at a position in the semi-infinitely long system where the solitary backward wave coming from $x \gg 1$ is reflected at $x = 0$, coming back to $x \gg 1$. Next, our attention is directed to the combination of Figs. 4(a) and 4(b). Combining the state between $\phi_{0,2n-2}$ and $\phi_{0,2n}$ for $t \leqslant 0$ in Fig. 4(a) with the one between $\phi_{0,2n}$ and $\phi_{0,2n+2}$ for $t \geqslant 0$ in Fig. 4(b), we construct the forward wave component in a soliton–antisoliton interaction. In this case, at $t < 0$ we observe the forward wave only in the region of $x < 0$, and at $t > 0$ we observe the forward wave only in the region of $x > 0$. We can also construct the backward wave component in the soliton–antisoliton interaction by combining the solution between $\phi_{0,2n-2}$ and $\phi_{0,2n}$ in Fig. 4(b) for $t \leqslant 0$ with the one between $\phi_{0,2n}$ and $\phi_{0,2n+2}$ in Fig. 4(a) for $t \geqslant 0$. In this case, at $t < 0$ we observe the backward wave only in the region of $x > 0$, and at $t > 0$ we observe the backward wave only in the region of $x < 0$. From the above explanations, we understand that the soliton–antisoliton interaction is equivalent to the state of the reflection at the open end of the semi-infinitely long system. Combining the state between $\phi_{0,2n-2}$ and $\phi_{0,2n}$

in Fig. 4(a) with the state between $\phi_{0,2n-2}$ and $\phi_{0,2n}$ in Fig. 4(b), the situation exhibits a spatial distribution in the soliton–antisoliton interaction at a value of t being negative, where the soliton and the antisoliton are coming near to each other. Combining the state between $\phi_{0,2n}$ and $\phi_{0,2n+2}$ in Fig. 4(a) with the state between $\phi_{0,2n}$ and $\phi_{0,2n+2}$ in Fig. 4(b), the situation exhibits a spatial distribution in the soliton–antisoliton interaction at a value of t being positive, where the soliton and the antisoliton are going away to each other.

If we direct our attention to ϕ_x, its polarity in the region between $\phi_{0,2n-2}$ and $\phi_{0,2n}$ is opposite to the polarity in the region between $\phi_{0,2n}$ and $\phi_{0,2n+2}$ for a given value of x, because the sign of u in the former region is opposite to the sign in the latter one in Eq. (2.2).

B. Soliton–soliton interaction

Here, we take account of Eq. (5.19) for $X(x,t)$ and of Eq. (5.16) for $T(x,t)$. We define $I(\Xi)$ as

$$I(\Xi) = -V(\Xi)/u$$

in Eq. (2.1), and then $-I(\Xi)$ represents the traveling wave component of the current along the equivalent transmission line. Equations (2.2) and (2.1) are rewritten as

$$\phi_x = I(\Xi)h(x), \tag{7.3}$$

$$\phi_t = -uI(\Xi)g(t). \tag{7.4}$$

Since Eqs. (5.19) for $h_\pm(x)$ and (5.16) for $g_\pm(t)$ have functionally the same properties as Eqs. (5.17) for $g_\pm(t)$ and (5.18) for $h_\pm(x)$, respectively, that is, $I(\Xi)$ in Eqs. (7.3) and (7.4) plays the same role as $V(\Xi)$ for the soliton–antisoliton interaction. Accordingly, the roles of x and t in Fig. 4 are also exchanged. Thus we understand that the polarity of $V(\Xi)$ between $\phi_{0,2n-2}$ and $\phi_{0,2n}$ becomes opposite to the one between $\phi_{0,2n}$ and $\phi_{0,2n+2}$, although the polarity of $I(\Xi)$ between $\phi_{0,2n-2}$ and $\phi_{0,2n}$ is same as the one between $\phi_{0,2n}$ and $\phi_{0,2n+2}$, and moreover that the state of $\phi = \phi_{0,2n}$ is always realized at $x = 0$ this time. This indicates that the kink solution of ϕ for these waves is either an increasing or a decreasing function of x from $\phi_{0,2n-2}$ to $\phi_{0,2n+2}$. Thus we understand that such a soliton represents the soliton–soliton interaction. In this case, we may regard the position $x = 0$ as the short end in the semi-infinitely long system, because of $w(0,t) = 0$ in Eq. (7.2). We also understand that the soliton–soliton interaction is equivalent to the state of the reflection at the short end in the semi-infinite system. It is noted, however, as far as dissipation and external forces are taken into consideration, the soliton–soliton interaction as described above does not take place under the initial conditions and the boundary condition imposed here insofar as we take account of G and J_B. This is because in the infinitely long system the soliton at $x \gg 1$ must travel in the same direction with the soliton at $x \ll -1$ with the same speed because of the existence of the uniformly applied external force and the dissipation.

In the above, we only discuss the traveling wave component of the solution. The (r) component is obtained, for instance, by subtracting the traveling wave component from the net solution of ϕ_t obtained from Eq. (2.13) or (2.14). The analytical method for the net solution will be developed in Sec. IX for the soliton–antisoliton interaction.

VIII. PURE SINE–GORDON SYSTEM

The properties of the pure sine–Gordon system are well known. Therefore we first apply our basic equations to the pure sine–Gordon system to verify the certainty of our theory. In this case, we can set $G = J_B = 0$ and $F(\phi) = \sin\phi$. Then, $a_{0,\pm}$ are rewritten as

$$a_{0,\pm} = a_0 = u/(1-u^2)^{1/2}.$$

We assume that

$$F'(\Xi) = \sin\phi \tag{8.1}$$

in Eq. (4.2). This assumption means that Eq. (4.2) is in agreement with Eq. (4.4), that is, the expression of the traveling wave component in the state plane is not influenced by the existence of the (r) component. However, it is again noted that the form of $V(\Xi)$ is quite different from that of $V^{(s)}(\xi)$ in real space. Under such an assumption, Eq. (4.2) reduces to

$$\frac{\partial V}{\partial \phi} = \frac{a_0^2 \sin\phi}{V(\Xi)}. \tag{8.2}$$

The solution to Eq. (8.2) is given by

$$V(\Xi) = \pm 2a_0 \sin(\phi/2), \tag{8.3}$$

which means that $V(\Xi)$ is the single-valued function of ϕ. Thus we can rewrite $V(\Xi)$ as $V(\phi)$.

On the other hand, Eq. (2.10) reduces to

$$\frac{\partial V}{\partial \phi} = \frac{1}{\beta(x,t)}\left[\frac{\sin\phi}{V(\phi)} + \frac{1}{u}\frac{dh}{dx} + \frac{dg}{dt}\right]. \tag{8.4}$$

We see from Eqs. (8.2) and (8.3) that $(\partial V/\partial\phi)|_{\phi 0}$ becomes independent of both x and t as either $|T|$ or $|X|$ approaches infinity. This means that we do not need to take account of the effect of $\gamma_\pm(x)$ nor that of $\omega_\pm(t)$, that is, $\gamma_\pm(x) = \omega_\pm(t) = 1$ in Eqs. (5.14) and (5.15). As a result, $(\partial V/\partial\phi)|_{\phi 0}$ is equal to $\pm a_0$, which is its value in the stationary state. Accordingly, we can use Eqs. (5.16) or (5.17) for $g_\pm(t)$ as an exact solution to Eq. (5.12).

Equations (5.16)–(5.19) are then rewritten as

$$g(t) = \begin{cases} \pm \tanh(a_0 t), & (8.5) \\ \pm \coth(a_0 t), & (8.6) \end{cases}$$

$$h(x) = \begin{cases} \pm \tanh(a_0 x/u), & (8.7) \\ \pm \coth(a_0 x/u). & (8.8) \end{cases}$$

We eliminate $\sin\phi$ from Eqs. (8.2) and (8.4). Then, we obtain

$$\frac{\partial V}{\partial \phi} = a_0 \frac{\Lambda^2(x,t) - 1}{\Lambda^2(x,t) + 1}, \tag{8.9}$$

where, if we adopt Eqs. (8.6) and (8.7),

$$\Lambda(x,t) = [\sinh(a_0 t)]/[u\cosh(a_0 x/u)], \tag{8.10}$$

and if we adopt Eqs. (8.5) and (8.8),

$$\Lambda(x,t) = [\cosh(a_0 t)]/[u\sinh(a_0 x/u)]. \tag{8.11}$$

We eliminate $V(\phi)$ from Eqs. (8.3) and (8.9). Then, we obtain the well-known expressions

$$\phi = \pm 4\tanh^{-1}[\Lambda(x,t)] \tag{8.12}$$

for two soliton interactions.[18] Thus we understand that the combination of Eqs. (8.10) and (8.12) is in agreement with the soliton–antisoliton interaction, and that of Eqs. (8.11) and (8.12) is in agreement with the soliton–soliton interactions. The assumption in Eq. (8.1) is correct for the pure sine–Gordon system. The breather solution is also obtained if u is a pure imaginary number in Eq. (8.10).

However, for other well-known nonlinear Klein–Gordon systems such as the ϕ^4 equation, the multiple sine–Gordon equations, and so on, we cannot obtain the correct solution using an assumption, such as $F'(\Xi)=F(\phi)$, that is different from the pure sine–Gordon system. If $V(\Xi)$ is weakened or strengthened in the state plane during the interaction, the equation is deformed from

$$\frac{\partial V}{\partial \phi}=\frac{a_0^2 F(\phi)}{V(\Xi)}, \tag{8.13}$$

which is obtained by replacing $\sin\phi$ by $F(\phi)$ in Eq. (8.2). This is the situation for the above nonlinear Klein–Gordon systems that are not integrable. Accordingly, we may regard Eq. (8.13) as a condition of the firm solidity of the wave in a pure nonlinear Klein–Gordon system.

IX. ANALYTICAL METHOD FOR SOLITON–ANTISOLITON INTERACTION

A. Energy flow and its related quantities

Before entering into the main issue, we need to describe the energy flow during the soliton–antisoliton interaction and quantities concerned with it. Consider the equivalent transmission line of an extended Klein–Gordon systems.[2,4] The sum of the energy flow of the soliton and the antisoliton $\epsilon^{(t)}$ is obtained by integrating the instantaneous power flow $p^{(t)}(\Xi)$ passing through a position x_0 along the equivalent transmission line defined by[19]

$$\begin{aligned} p^{(t)}(\Xi) &= -(\phi_x)^{(t)}(\phi_t)^{(t)}, \\ &= \frac{1}{u}V(\Xi)\frac{\partial\phi}{\partial T}, \end{aligned} \tag{9.1}$$

with respect to t from $-\infty$ to $+\infty$ through $T(x_0,t)$, i.e.,

$$\begin{aligned} \epsilon^{(t)} &= \int_{-\infty}^{+\infty} P^{(t)}(\Xi)\frac{\partial T}{\partial t}dt \\ &= \frac{1}{u}\int_{\phi_{0,2n-2}}^{\phi_{0,2n+2}} V(\Xi)\,d\phi . \end{aligned} \tag{9.2}$$

We integrate Eq. (4.2) with respect to t from $-\infty$ to $+\infty$ through $T(x_0,t)$ by noticing that $V(\Xi)=0$ there, where we do not take account of the directionality. Then, we obtain

$$\int_{\phi_{0,2n-2}}^{\phi_{0,2n+2}} V(\Xi)\,d\phi=\frac{2J_B(\phi_{0,2n}-\phi_{0,2n-2})}{G}, \tag{9.3}$$

where we have used the following relations:

$$\begin{aligned} &\int_{\phi_{0,2n-2}}^{\phi_{0,2n+2}} F'(\Xi)\,d\phi=\int_{\phi_{0,2n-2}}^{\phi_{0,2n+2}} F(\phi)\,d\phi=0, \\ &\phi_{0,2n}-\phi_{0,2n-2}=\phi_{0,2n+2}-\phi_{0,2n}. \end{aligned} \tag{9.4}$$

Equation (9.4) means that there is no energy stored in the nonlinear element [either $F'(\Xi)$ or $F(\phi)$] after the whole process is complete. It is straightforward to see that the physically significant systems such as the sine–Gordon, the multiple sine–Gordon, and so on, irrespective of whether they have the dissipation term and the external force term, satisfy the condition for the middle expression in Eq. (9.4). The condition on the left-hand side in Eq. (9.4) is also simply derived from the fact that $F'(\Xi)$ is regarded as $F(\phi)$ just around the singular point. Inserting Eq. (9.3) into Eq. (9.2), we obtain the following expression:

$$\epsilon^{(t)}=2J_B(\phi_{0,2n}-\phi_{0,2n-2})/(uG). \tag{9.5}$$

In the limit to the stationary state, $(\phi_t)^{(t)}$ is replaced by $V^{(s)}(\xi)$, $(\phi_x)^{(t)}$ by $-V^{(s)}(\xi)/u$, $X(x,t)$ by x, and $T(x,t)$ by t. Accordingly, we again obtain Eq. (9.3) for the stationary state if we put two stationary waves between $\phi_{0,2n-2}$ and $\phi_{0,2n+2}$. Thus we understand that $\epsilon^{(t)}$ is twice the energy of the stationary solitary wave. In the case of the extended sine–Gordon system, $\epsilon^{(t)}$ is

$$\epsilon^{(t)}=4\pi J_B/(uG).$$

We differentiate Eq. (9.1) with respect to ϕ. Then, we obtain

$$\frac{\partial p^{(t)}}{\partial\phi}=\frac{2}{u}\frac{\partial V}{\partial T}. \tag{9.6}$$

We integrate Eq. (9.6) with respect to t from $-\infty$ to $+\infty$ through $T(x,t)$ like Eq. (9.2). The result shows zero because of $V(\Xi)=0$ at $t=\pm\infty$, and is also in agreement with the one for the limit at the stationary state. From the above two facts, we may add the condition that the integration of $\partial^2 p(t)/\partial\phi^2$ with respect to t from $-\infty$ to $+\infty$ through $T(x,t)$ should also be in agreement with the one for the stationary state. Thus we obtain the following relation between the slopes at $t=\pm\infty$:

$$\gamma_+(x)a_{0,+}+\gamma_-(x)a_{0,-}=a_{0,+}+a_{0,-}=2d_1. \tag{9.7}$$

B. Coordinate transformation

Here, we choose the combination of Eq. (5.17) for $g(t)$ and Eq. (5.18) for $h(x)$. However, on account of the assumption $\omega_\pm(t)=1$ we will see that Eq. (5.17) does not express the correct value of $g(t)$ as clarified later insofar as we take account of the dissipation and the external force. However, we can derive the exact solution by taking advantage of Eq. (5.17) as described below. We replace $g_\pm(t)$ in Eq. (5.17) with $g'_\pm(t)$ to distinguish it from the exact value. We transform the t coordinate into newly introduced $T_\pm$ coordinates so as to satisfy $g'_\pm(t)=g_\pm(T_\pm)$. In this case, we also need to transform the x coordinate into $X_\pm$ coordinates so that we make the states in the $(X_\pm,T_\pm)$ coordinates identical to the state of the exact solution in the (x,t) coordinates. Then, we can also impose the following identity condition for the phase velocity $w(x,t)$:

$$\frac{dx}{dt}=\frac{dX_\pm}{dT_\pm}. \tag{9.8}$$

Then, the following relation should hold at $t=T_\pm=0$:

$$\frac{dx}{dt}=\frac{dX_+}{dT_+}=\frac{dX_-}{dT_-}. \tag{9.8'}$$

The transformation from the x coordinate into the $X_\pm$ coordinates is made as follows:

$$a_{0,\pm}\gamma_\pm(x) = d_1 \mp d_2\alpha_\pm(X_\pm)\,, \tag{9.9}$$

by introducing $\alpha_\pm(X_\pm)$. Since Eq. (9.9) should satisfy Eq. (9.7), we can set

$$\alpha_+(X_+) = \alpha_-(X_-) = \alpha(x)\,. \tag{9.10}$$

If we integrate Eq. (9.9), we obtain

$$\int_0^x a_{0,\pm}\gamma_\pm(x')\,dx' = \int_0^{X_\pm} d_1\left[1 \mp \frac{d_2}{d_1}\alpha_\pm(X'_\pm)\right]dX'_\pm\,. \tag{9.11}$$

By using Eq. (9.11), $h_\pm(x)$ is transformed into the following value in $X_\pm$ coordinates:

$$h_\pm(x) = \tanh\left(\int_0^{X_\pm} d_1\left[1 \mp \frac{d_2}{d_1}\alpha_\pm(X'_\pm)\right]dX'_\pm\right). \tag{9.12}$$

C. Solution at $t=0$

We can rewrite Eq. (2.13) or Eq. (2.14) as

$$\frac{\partial\phi_t}{\partial x} + \frac{h_\pm(x)}{u\beta_\pm(x,t)}\left(\frac{1}{u}\frac{dh_\pm}{dx} + \frac{dg_\pm}{dt} + Gg_\pm(t)\right)\phi_t = -\frac{g_\pm(t)h_\pm(x)}{u\beta_\pm(x,t)}\eta(\phi)\,, \tag{9.13}$$

where

$$\beta_\pm(x,t) = h^2_\pm(x)/u^2 - g^2_\pm(t)\,.$$

From the analogy of the exact value $g_\pm(t)$ with $g'_\pm(t)$, we can impose the conditions just around $t=0$ that $g^2_\pm(t) \gg h^2_\pm(x)$ and

$$\left|\frac{dg_\pm}{dt}\right| \gg G|g_\pm(t)| \gg \left|\frac{1}{u}\frac{dh_\pm}{dx}\right|\,.$$

Then, Eq. (9.13) reduces to

$$\frac{\partial\phi_t}{\partial x} - \frac{h_\pm(x)}{u}\frac{1}{g^2_\pm(t)}\frac{dg_\pm}{dt}\phi_t = 0\,. \tag{9.14}$$

Since it is seen from Eq. (5.12) that the form of $g_\pm(t)$ must be the same as that of $g'_\pm(t)$ around $t=0$ except that the effects of $\omega_\pm(t)$ are included in $g_\pm(t)$. Thus $g_\pm(t)$ and $[1/g^2_\pm]dg_\pm/dt$ around $t=0$ in Eq. (9.14) can be expressed from the analogy with Eq. (5.17) by

$$g_\pm(t) \simeq \pm B'_\pm \coth[B'_\pm\omega_\pm(0)a_{0,\mp}t] \simeq \pm 1/[\omega_\pm(0)a_{0,\mp}t]\,, \tag{9.15}$$

$$\frac{1}{g^2_\pm(t)}\frac{dg_\pm}{dt}\bigg|_{t\to\pm 0} = \mp\omega_\pm(0)a_{0,\pm}\,, \tag{9.15'}$$

where

$$B'_\pm = 1 \pm A_\pm/\omega_\pm(0)\,.$$

We insert Eqs. (9.15) and (9.15') into (9.14), and take account of the following continuity condition at $t = \pm 0$:

$$\phi_t|_{t\to+0} = \phi_t|_{t\to-0} = \phi_t|_{t=0}\,. \tag{9.16}$$

As a result, we obtain

$$\omega_+(0)a_{0,-} = \omega_-(0)a_{0,+}\,.$$

Moreover, $\omega_\pm(0)a_{0,\mp}$ can be chosen to be equal to the arithmetic mean of $a_{0,+}$ and $a_{0,-}$ as clarified later, i.e.,

$$\omega_\pm(0)a_{0,\mp} = d_1\,. \tag{9.17}$$

Taking Eqs. (9.16) and (9.17) into consideration in Eq. (9.14), we obtain

$$h_+(x) = h_-(x) = h(x)\,. \tag{9.18}$$

Equation (9.18) indicates that the form of $h_\pm(x)$ is irrespective of whether t is positive or not. Taking Eqs. (5.18) and (9.7) into consideration with Eq. (9.18), we obtain

$$\gamma_\pm(x)a_{0,\pm} = d_1\,. \tag{9.19}$$

Thus we see that $\gamma_\pm(x)$ are constants. Since $u>0$ in either state for $t>0$ and $x>0$ or for $t<0$ and $x<0$, and since $u<0$ in either state for $t>0$ and $x<0$ or for $t<0$ and $x>0$, Eq. (9.14) is finally expressed by

$$\frac{\partial\phi_t}{\partial x}\bigg|_{t=0} (\pm)\frac{d_1}{u}h(x)\phi_t\bigg|_{t=0} = 0\,, \tag{9.20}$$

where $(\pm)$ is used for the regions in $x>0$ and $x<0$, respectively, and u is always positive as a result of replacement of u by $-u$ if $u<0$. The solution to Eq. (9.20) is given by

$$\phi_t|_{t=0} = \phi_{t,0}\exp\left((\mp)\int_0^x \frac{d_1}{u}h(x')\,dx'\right), \tag{9.21}$$

where $\phi_{t,0}$ is a constant.

The relation between x and $X_\pm$ is then expressed from Eqs. (9.11) and (9.19) by

$$x = X_\pm \mp \frac{d_2}{d_1}\int_0^{X_\pm}\alpha_\pm(X'_\pm)\,dX'_\pm\,. \tag{9.22}$$

Next, we use $g'_\pm(t)$ instead of $g_\pm(t)$ in Eq. (9.14) and try to derive an exact solution by transforming the (x,t) coordinate into the $(X_\pm, T_\pm)$ coordinates. At $t=0$, Eq. (9.14) is rewritten as

$$\frac{d\phi_t^{(\pm)}}{\partial x}\bigg|_{t\to\pm 0} (\pm)\frac{h(x)}{u}a_{0,\mp}\,\phi_t^{(\pm)}\bigg|_{t\to\pm 0} = 0\,, \tag{9.23}$$

where the superscripts $(\pm)$ on ϕ_t denote the states being $t>0$ and $t<0$, respectively. The solutions to Eq. (9.23) are expressed by

$$\phi_t^{(\pm)}|_{t\to\pm 0} = \phi_{t,0}\exp\left((\mp)\int_0^x \frac{a_{0,\mp}}{u}h(x')\,dx'\right). \tag{9.24}$$

It is apparent that Eq. (9.24) does not satisfy the continuity condition, i.e.,

$$\phi_t^{(+)}|_{t\to+0} \neq \phi_t^{(-)}|_{t\to-0} \tag{9.25}$$

since $a_{0,-} \neq a_{0,+}$, except for the case $G = J_B = 0$. This indicates that Eq. (9.24) is changed with a step at $t=0$ without satisfying Eq. (9.16). So it is natural to consider that the functional form of $\omega_\pm(t)$ in the exact solution should be adjusted so as to remove the step at $t=0$ expressed by Eq. (9.25), that is, so as to satisfy the continuity condition [Eq. (9.16)]. Accordingly, it is also natural to consider that the

values of $\omega_{\pm}(t)a_{0,\mp}$ included in Eq. (9.14) are adjusted at $t=0$ so as to become the arithmetic mean of $a_{0,+}$ and $a_{0,-}$, i.e., satisfying Eq. (9.17), so that the value of $\phi_t|_{t=0}$ is adjusted to be the geometric mean of $\phi_t^{(+)}|_{t\to+0}$ and $\phi_t^{(-)}|_{t\to-0}$.

Consider the above adjustment in more detail. From Eq. (9.24) $\phi_t^{(+)}|_{t\to+0}$ is equal to $\phi_t^{(-)}|_{t\to-0}$ at $x=0$, and $\phi_t^{(+)}|_{t\to+0}$ becomes smaller than $\phi_t^{(-)}|_{t\to-0}$ at any value of $|x|\neq 0$ since $a_{0,-}>a_{0,+}$. Therefore the relation between $\phi_t^{(+)}|_{t\to+0}$ and $\phi_t^{(-)}|_{t\to-0}$ at an arbitrary value of $|x|$, except for $x=0$, can be shown by states 1 and 1′ in Fig. 5, respectively. Thus we see that none of these states can be identical to each other because of G and J_B. As a first step, we transform the x coordinates into $X_{\pm}$ coordinates using Eq. (9.22). Equation (9.14), in which $g_{\pm}(t)$ is replaced by $g'_{\pm}(t)$, that is, another expression of Eq. (9.23), is rewritten as

$$\left.\frac{\partial\phi_t^{(\pm)}}{\partial X_{\pm}}\right|_{t\to\pm0} - \left.\frac{h(x)}{u}\frac{1}{g'^2_{\pm}(t)}\frac{dg'_{\pm}}{dt}\frac{dx}{dX_{\pm}}\phi_t^{(\pm)}\right|_{t\to\pm0} = 0, \tag{9.26}$$

and then the positions of $\phi_t^{(\pm)}|_{t\to\pm0}$ are moved from state 1 to 2 and from state 1′ to 2′ as shown in Fig. 5, respectively. In this process the x coordinates are moved up to $X_{\pm}$ coordinates, while t coordinate is kept unchanged. We next transform the t coordinates into the $T_{\pm}$ coordinates using the relation $g'_{\pm}(t)=g_{\pm}(T_{\pm})$ so as to satisfy Eq. (9.8) by keeping $X_{\pm}$ coordinates unchanged, so that the state at the point $(x,t)=(x,\pm0)$ may be identical to the state at the point $(X_{\pm},T_{\pm})=(X_{\pm},0)$. However, we cannot say that $(dX_+/dT_+)|_{T_+\to0}$ is equal to $(dX_-/dT_-)|_{T_-\to0}$ yet. Taking account of Eq. (9.8), we can rewrite Eq. (9.26) as

$$\left.\frac{\partial\phi_t^{(\pm)}}{\partial X_{\pm}}\right|_{T_{\pm}\to0} - \left.\frac{h(X_{\pm})}{u}\frac{1}{g^2_{\pm}(T_{\pm})}\frac{dg_{\pm}}{dT_{\pm}}\phi_t^{(\pm)}\right|_{T_{\pm}\to0} = 0. \tag{9.27}$$

In this stage, the positions representing $\phi_t^{(\pm)}$ are moved from state 2 to 3 and from state 2′ to 3′, respectively. We consider the process by which the $T_{\pm}$ coordinates approach the t coordinates and the $X_{\pm}$ coordinates approach the x coordinates in Eq. (9.27). In this process the solution moves along the dashed curve, reaching state 4. As a result, since $X_{\pm}$ and $T_{\pm}$ in Eq. (9.27) are replaced by x and t, respectively, state 4 describes the exact solution in the (x,t) coordinates, i.e., Eq. (9.20). However, since we are looking for the states for the $X_{\pm}$ coordinates, the states have to move from state 3 to 5 and from state 3′ to 5′, respectively. Consider the relationship between $\phi_t^{(\pm)}$ and $\phi_{T_{\pm}}$, expressed by

$$\phi_t^{(\pm)} = \frac{dT_{\pm}}{dt}\phi_{T_{\pm}}. \tag{9.28}$$

In order that the states at $T_{\pm}$ may be identical to the state at t, it is necessary that $\phi_t^{(\pm)}=\phi_{T_{\pm}}=\phi_t$, i.e., $t\to T_{\pm}$. It is also necessary to move the x coordinates to the $X_{\pm}$ coordinates from Eq. (9.8′). Such a process has already been taken in the movement from state 1 (1′) to 2 (2′). At the limit $t\to T_{\pm}$, the results are in agreement with the results ob-

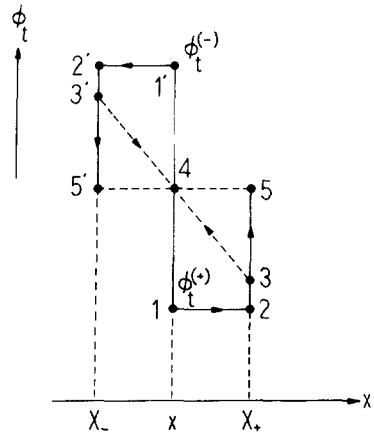


FIG. 5. Relationship between $\phi_t^{(+)}|_{t\to+0}$ and $\phi_t^{(-)}|_{t\to-0}$.

tained by replacing t by $T_{\pm}$ and x by $X_{\pm}$ in Eq. (9.20). Thus the states at 5 and 5′ can be written as

$$\left.\frac{\partial\phi_t}{\partial X_{\pm}}\right|_{T_{\pm}\to0} - \left.\frac{h(X_{\pm})}{u}\frac{1}{g^2_{\pm}(T_{\pm})}\frac{dg_{\pm}}{dT_{\pm}}\phi_t\right|_{T_{\pm}\to0} = 0. \tag{9.27′}$$

It is also possible to rewrite the state in a form exhibiting the relation between x and $X_{\pm}$ by inserting Eq. (9.8′) into Eq. (9.20):

$$\left.\frac{\partial\phi_t}{\partial X_{\pm}}\right|_{T_{\pm}\to0} - \left.\frac{h(x)}{u}\frac{1}{g^2_{\pm}(t)}\frac{dg_{\pm}}{dt}\frac{dx}{dX_{\pm}}\phi_t\right|_{T_{\pm}\to0} = 0. \tag{9.29}$$

The solution to Eq. (9.29) is expressed by

$$\phi_t|_{T_{\pm}\to0} = \phi_{t,0}\exp\left((\mp)\int_0^{X_{\pm}}\frac{d_1}{u}h(x)\times\left(1\mp\frac{d_2}{d_1}\alpha_{\pm}(X'_{\pm})\right)dX'_{\pm}\right), \tag{9.30}$$

where

$$\int_0^{X_+}\alpha_+(X'_+)\,dX'_+ = \int_0^{X_-}\alpha_-(X'_-)\,dX'_- = \int_0^{X}\alpha(x')\,dx', \tag{9.31}$$

because of the identity of each state.

D. Net solution

If we insert Eq. (9.31) in Eq. (9.22), we obtain

$$x=(X_++X_-)/2. \tag{9.32}$$

Since Eq. (9.22) is independent of t, we can apply Eq. (9.32) at any value of t. The time difference for traveling between x and $X_{\pm}$, $\tau(x,t)$, is expressed by

$$\tau(x,t)=l(x)/w(x,t),$$

where $l(x)$ is the distance between x and $X_{\pm}$ and is given by

$$l(x)=\frac{d_2}{d_1}\int_0^x\alpha(x')\,dx'.$$

By dividing Eq. (9.22) by $w(x,t)$, the relation between t and $T_\pm$ can be written as

$$t = T_\pm \mp \tau(x,t) . \tag{9.33}$$

It is noted that $\tau(x,0) = \tau(0,t) = 0$ since $|w(x,0)| = |w(0,t)| = \infty$.

In numerical calculations, it is more convenient to use Eq. (2.14) than Eq. (2.13), since ϕ is the increasing function of t between $\phi_{0,2n-2}$ and $\phi_{0,2n+2}$ in the soliton–antisoliton interaction, except for the region around the singular point at $|X_\pm|$ or $|T_\pm| \to \infty$ if it is a node or a spiral point. Consider the following expression, in which $g_\pm(t)$ is replaced by $g'_\pm(t)$ in Eq. (2.14):

$$\phi_{tt}^{(\pm)} - \chi'_2(x,t)\phi_t^{(\pm)} = g'^2_\pm(t)\eta(\phi)/\beta'_\pm(x,t) , \tag{9.34}$$

$$\chi'_2(x,t) = \frac{g'_\pm(t)}{\beta'_\pm(x,t)}\left(\frac{1}{u}\frac{dh_\pm}{dx} + \frac{dg'_\pm}{dt} + Gg'_\pm(t)\right) + \frac{1}{g'_\pm(t)}\frac{dg'_\pm}{dt} , \tag{9.35}$$

$$\beta'_\pm(x,t) = h^2_\pm(x)/u^2 - g'^2_\pm(t) . \tag{9.36}$$

We transform the x coordinates into the $X_\pm$ coordinates using Eq. (9.22), and then the state is moved into 1 (1′) into 2 (2′) in Fig. 5. Next, we transform the t coordinates into $T_\pm$ coordinates using the relation $g'_\pm(t) = g_\pm(T_\pm)$, and then the state is moved from 2 (2′) to 3 (3′) in Fig. 5 so as to satisfy the identity condition [Eq. (9.8)]. The result shows

$$\frac{\partial\phi_t^{(\pm)}}{\partial T_\pm} - \left(\frac{g_\pm(T_\pm)}{\beta'_\pm(X_\pm,t)} \times\left(\frac{1}{u}\frac{dh_\pm}{dX_\pm} + \frac{dg_\pm}{dT_\pm} + Gg(T_\pm)\frac{dt}{dT_\pm}\right) + \frac{1}{g^2_\pm(T_\pm)}\frac{dg_\pm}{dT_\pm}\right)\phi_t^{(\pm)} = \frac{g^2_\pm(T_\pm)\eta(\phi)}{\beta'_\pm(X_\pm,t)}\frac{dt}{dT_\pm} . \tag{9.37}$$

At the limit $t \to T_\pm$ [from 3 (3′) into 5 (5′) in Fig. 5], we can rewrite Eq. (9.37) as

$$\frac{\partial\phi_t}{\partial T_\pm} - \chi_2(X_\pm,T_\pm)\phi_t = g^2_\pm(T_\pm)\frac{\eta(\phi)}{\beta_\pm(X_\pm,T_\pm)} . \tag{9.38}$$

If we replace $T_\pm$ by t and $X_\pm$ by x in Eq. (9.38), we obtain Eq. (2.14). Thus we understand that the solution to Eq. (9.38) is in agreement with the solution in Eq. (2.14). It is also understood from the above process that finding the solution satisfying the continuity condition at $t = T_\pm = 0$ in Eq. (9.34) by regarding the value of x as X_- and the value of t as T_-, if $t < 0$, and the value of x as X_+ and the value of t as T_+, if $t > 0$, is equivalent to solving Eq. (9.38), thus satisfying the same continuity condition as the above at $T_\pm = t = 0$. Thus the waveforms of ϕ_t in the $(X_\pm, T_\pm)$ coordinates are calculated by starting first at the point in the state plane $(\phi,\phi_t) = (\phi_{0,2n-2},0)$ with increasing T_-, ending at the point $(\phi_{0,2n},\phi_t|_{T-\to 0})$, and, next, starting the calculation at the point $(\phi_{0,2n+2},0)$ with decreasing T_+, ending at the point $(\phi_{0,2n},\phi_t|_{T+\to 0})$ so as to be $\phi_t|_{T+\to 0} = \phi_t|_{T-\to 0}$. The waveforms in the (x,t) coordinates are obtained by using Eqs. (9.32) and (9.33) from the result in the $(X_\pm, T_\pm)$ coordinates.

X. APPLICATION TO EXTENDED SINE–GORDON SYSTEM

The coexistence region of a soliton and an antisoliton for the extended sine–Gordon system is determined to satisfy $0 \leqslant A_- < 1$ in Eq. (5.17), and is depicted in Fig. 6 with the unshaded region in the relation between G and J_B. The dashed curve shows the boundary between the destructive and the nondestructive collision calculated by Mclaughin *et al.*[13] using a perturbation technique. The circle and the triangle also show the boundary in a long cavity having open ends obtained using a computer simulation technique by Erné *et al.*[9] The circle shows the boundary for J_B below which the solution decays into either a radiation or a static soliton after the soliton collides with the open end, which is equivalent to the soliton–antisoliton interaction, and the triangle shows the boundary for J_B above which the solution switches to a radiation. Since our region represents coexistence of the soliton and the antisoliton at the limits where $|x| \to \infty$ and $t \to -\infty$, it includes the above boundary. It should be noted that the coexistence region disappears for G larger than 0.85.

The numerical integration on Eq. (9.38) is made by using the Runge–Kutta method, except for a narrow region around the position satisfying the condition $\beta(x,t) = 0$, because the result due to the Runge–Kutta method sometimes diverges there. Accordingly, we use Euler's method in the narrow region. As a result, we always obtain satisfactory results with it.

The one example of solution curves for $G = J_B = 0$, i.e., for the pure sine–Gordon system calculated by the method

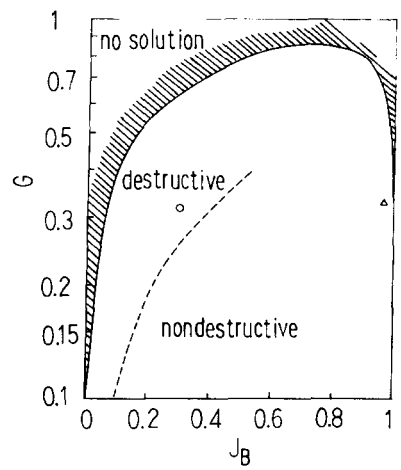


FIG. 6. Coexistence region of the soliton and the antisoliton in the extended sine–Gordon system. The dashed curve shows the boundary between the destructive and nondestructive collision and the circle and the triangle show the boundary between the destructive collision by McLaughlin *et al.* using a perturbation technique,[13] and nondestructive collision of the soliton with the open ends of a long cavity obtained by Erné *et al.*[9] with computer simulation. The coexistence region includes the destructive and nondestructive collision boundary.

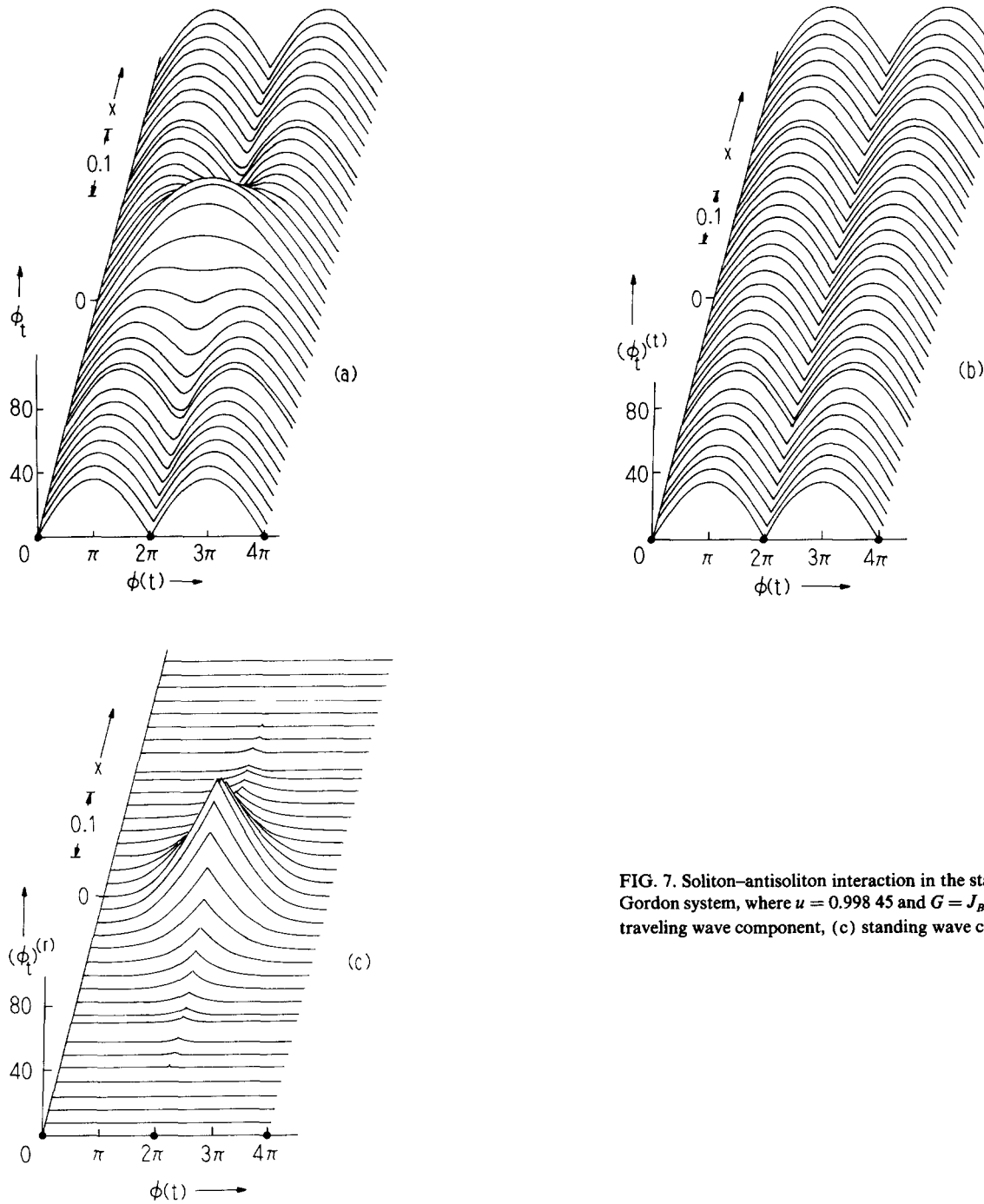


FIG. 7. Soliton–antisoliton interaction in the state plane to the pure sine–Gordon system, where $u = 0.998\,45$ and $G = J_B = 0$: (a) net solution, (b) traveling wave component, (c) standing wave component.

described in D in Sec. IV, is depicted in Fig. 7(a) in the relation between ϕ_t and $\phi(t)$ as changing parameters x, where $u = 0.998\,45$ for later reference. For $|x| \gg 1$, the solutions are completely separated into two isolated states expressing existence of the stationary soliton and the stationary antisoliton. With decreasing $|x|$, the coupling of the soliton and the antisoliton increases and brings completely the one state at $t = 0$. The solution curves in Fig. 7(a) are divided into the traveling wave component $(\phi_t)^{(t)}$ and the (r) component $(\phi_t)^{(r)}$ as depicted in Figs. 7(b) and 7(c), respectively. It is seen that $(\phi_t)^{(t)}$ is independent of the value of x in the (ϕ,ϕ_t) plane, that is, the single-valued function of ϕ as expected. The net waveforms in real space are depicted in Fig. 8(a), and are in agreement with those calculated from Eqs. (8.10) and (8.12). The traveling wave component and the (r) component are depicted in Figs. 8(b) and 8(c), respectively. The (r) component constructs a standing wave without radiation. As the soliton and the antisoliton come near each other, the traveling wave component decays to be transformed into the standing wave component. At $x = 0$ and $t = 0$, it is completely replaced by $(\phi_t)^{(r)}$.

Solutions for the extended sine–Gordon system are numerically analyzed for $G = 0.018$ throughout this paper. The relations between $\phi_t|_{T_\pm \to 0}$ and $|X_\pm|$ are depicted with solid curves in Fig. 9 for $J_B = 0.4$. The dashed curve shows the relation between $\phi_t|_{t=0}$ and $|x|$ calculated from Eq.

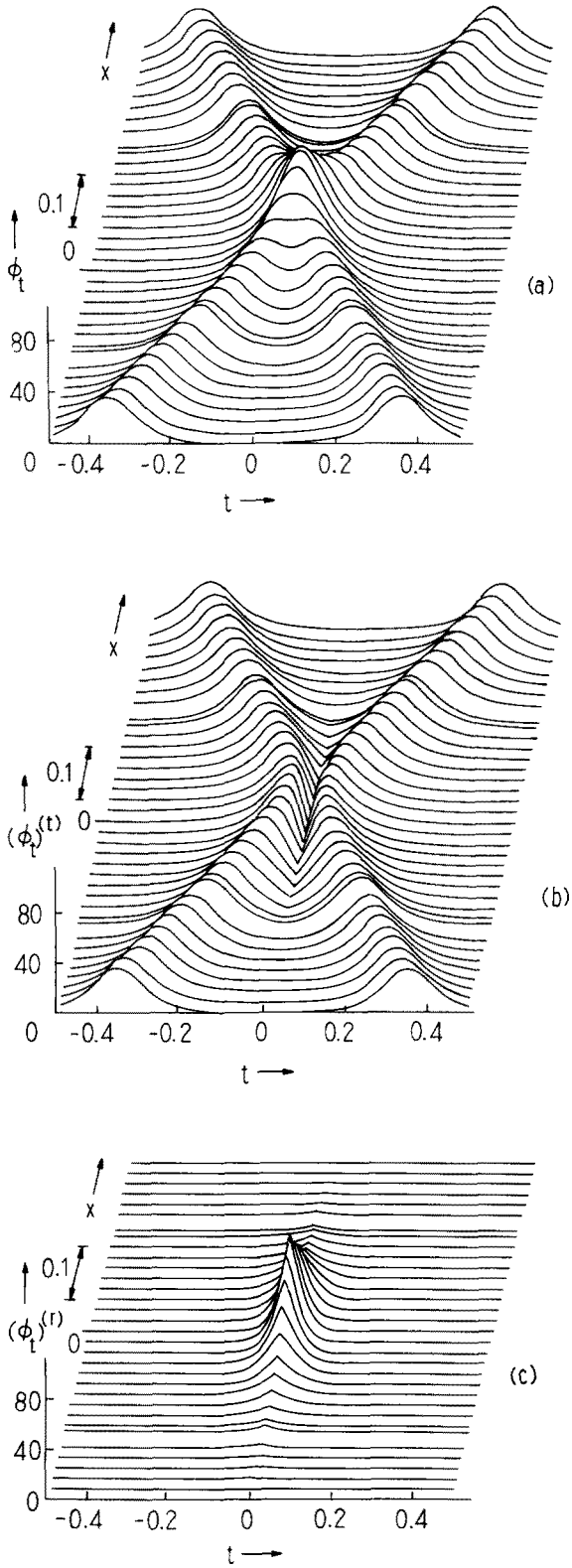


FIG. 8. Waveforms of the solution in Fig. 7; (a) net waveforms, (b) traveling wave component, (c) standing wave component.

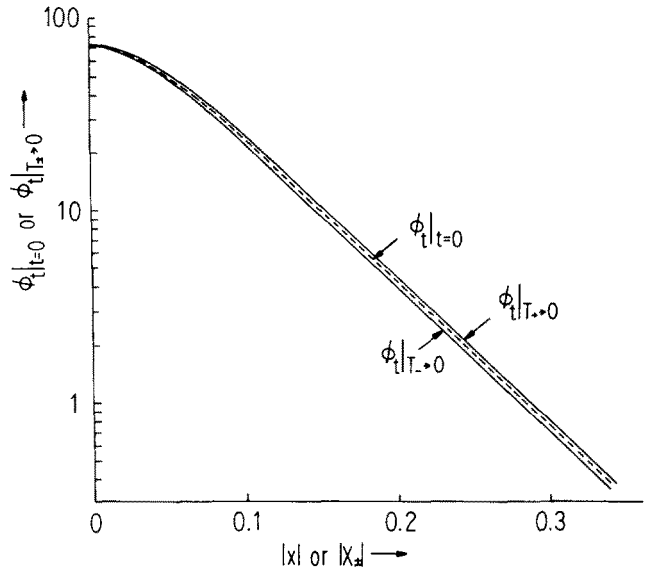


FIG. 9. Relationship between $\phi_t|_{t=0}$ and $|X_\pm|$, where $G=0.018$ and $J_B=0.4$. The dashed curve is the exact solution obtained by $x=(X_+ + X_-)/2$, and is in agreement with the analytical solution expressed by Eq. (9.21).

(9.21), and is in agreement with the one calculated from Eq. (9.32) by using the solid curves. The solution curves in the relation between ϕ_t and ϕ and the waveforms for $J_B=0.4$ are depicted in Figs. 10 and 11, respectively. While the soliton and the antisoliton are far from the origin of each other, they travel almost with a constant velocity. Once they come near to a certain extent, a wedge-shaped distortion is generated in each wave at the same time at the position satisfying $\beta(x,t)=0$. The distortion moves to the direction opposing the corresponding soliton movement. As they come closer to each other, the distortion disappears, so that there can be no position satisfying the condition $\beta(x,t)=0$. After the centers of the waves pass through each other, a thorn-shaped distortion is generated at the position satisfying $\beta(x,t)=0$. The distortion moves in the direction opposing the corresponding soliton movement, and disappears when it arrives around the center of the wave. As the waves go to each other, they approach the stationary state. Since $u=0.998\,45$ in this case, the waveform can be directly compared with the one for $G=J_B=0$ in Fig. 8. It is noted that $(\phi_t)^{(t)}$ no longer becomes a single-valued function of ϕ because of the effects of $\omega_\pm(t)$ and of the local distortions around $\beta(x,t)=0$. We also notice from Figs. 7(c) and 10(c) that $\partial(\phi_t)^{(r)}/\partial\phi$ approaches zero as $|t|$ approaches infinity whatever the value of x is, which means the disappearance of the (r) component.

In the above calculation, the time step δt is chosen to be 5×10^{-3}. To examine the influence of the time step, it is also changed with several steps from 5×10^{-3} to 5×10^{-4}. The results are depicted in Fig. 12, and do not depend upon the values of them.

The numerical analyses are also performed for $J_B=0.1$ and 0.6, respectively. For $J_B=0.1$, the solution is almost in agreement with the solution for the pure sine–Gordon system with same value of u except that the weaker distortions appear at the position satisfying the condition $\beta(x,t)=0$. For $J_B=0.6$, the solution is more strongly distorted than for $J_B=0.4$ around the position satisfying $\beta(x,t)=0$. As a result, the traveling wave component is appreciably weakened in a region where $t<0$ and is appreciably strengthened in a

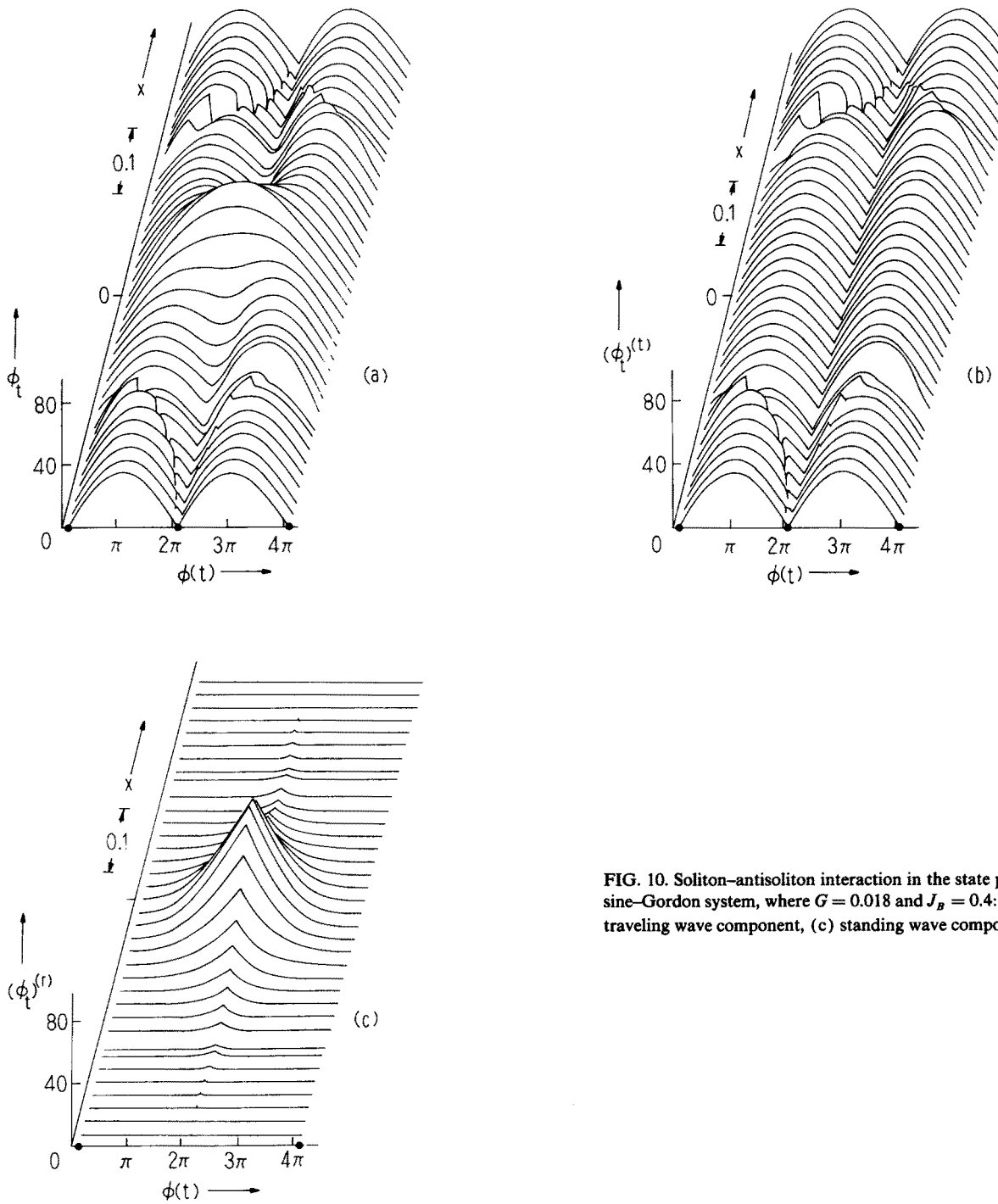


FIG. 10. Soliton–antisoliton interaction in the state plane to the extended sine–Gordon system, where $G = 0.018$ and $J_B = 0.4$: (a) net solution, (b) traveling wave component, (c) standing wave component.

region where $t > 0$, both of which are mainly in a region of x around $x = 0$ not satisfying the condition $\beta(x,t) = 0$.

The conventional computer simulation, by taking advantage of the corresponding finite difference equation, usually brings some emission of radiation when the system is disturbed as in the cases of a soliton–soliton interaction in the extended sine–Gordon system,[20] and of the soliton–antisoliton interaction in the ϕ^4 system,[5,6] in the double sine–Gordon system,[3,5] and even in the pure sine–Gordon system.[3] Moreover, some analyses have been made of the transmission of a soliton at a microshort put at a position in the extended sine–Gordon system by using a perturbation technique[12,13,21] and computer simulation.[21] Although in the perturbation technique no radiation is observed, in the computer simulation some radiation is observed. In our case, for the extended sine–Gordon system, no radiation takes place but local distortions.

XI. LOCAL DISTORTION IN EXTENDED SINE–GORDON SYSTEM

We linearize Eq. (2.14) around a singular point at $|t| = \infty$ for the extended sine–Gordon system. The result

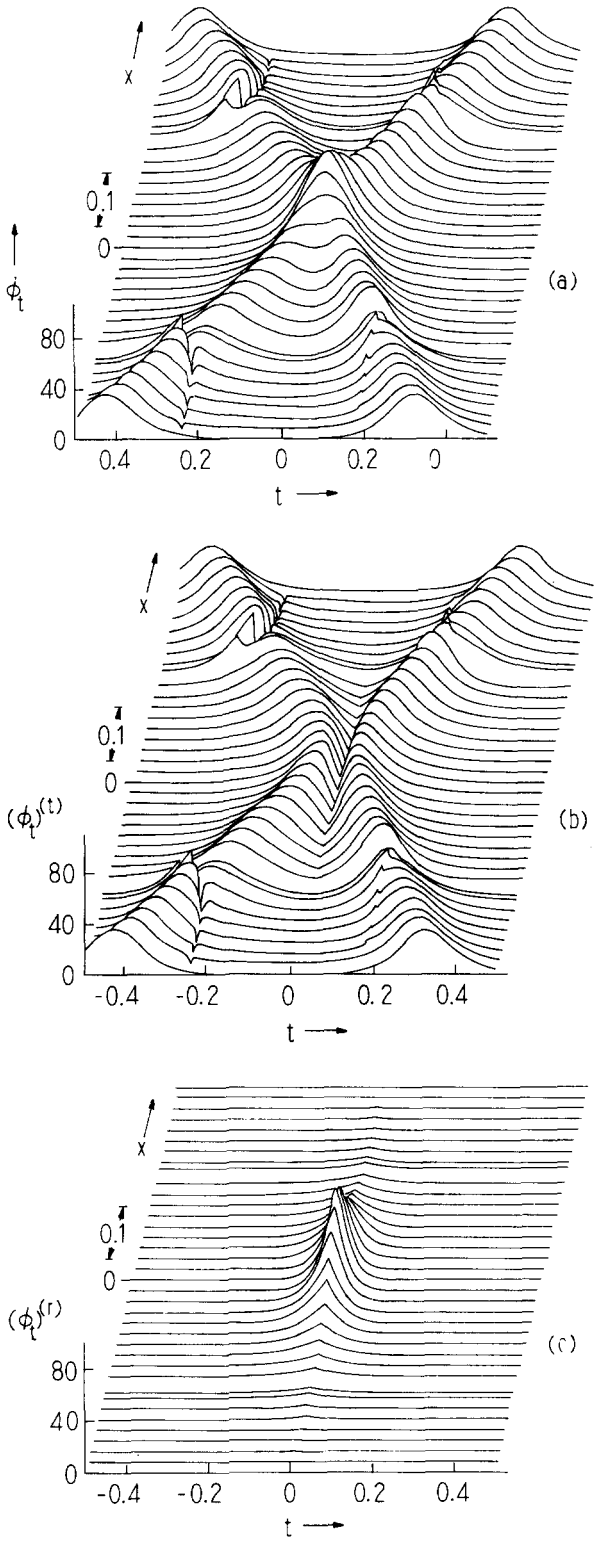


FIG. 11. Waveforms of the solution in Fig. 10: (a) net waveforms, (b) traveling wave component, (c) standing wave component.

shows

$$\psi_{tt} - \frac{u}{h^2(x) - u^2}\left(\frac{dh}{dx} + uG\right)\psi_t - \frac{u^2}{h^2(x) - u^2}(\cos\phi_0)\psi = 0. \tag{11.1}$$

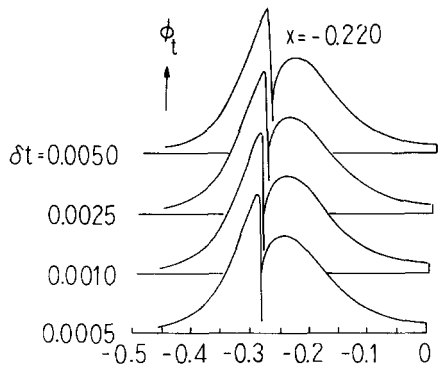


FIG. 12. Effect of the time step δt on the soliton–antisoliton interaction in the extended sine–Gordon system, where $G = 0.018$ and $J_B = 0.4$.

The discriminant $D(x)$ of Eq. (11.1) becomes

$$D(x) = \left(\frac{u}{h^2(x) - u^2}\right)^2\left(\frac{dh}{dx} + uG\right)^2 + \frac{4u^2}{h^2(x) - u^2}(\cos\phi_0). \tag{11.2}$$

If $|x|$ is smaller than a certain value, say $|x_{c0}|$, where x_{c0} satisfies $h^2(x_{c0}) = u^2$, the singular point at $|t| \to \infty$ is a node or a spiral point from Eq. (11.2), and the relation $h^2(x) < u^2 g^2(t)$ is always preserved, since $h^2(x)$ is an increasing function of $|x|$ toward unity and since $g^2(t)$ is a decreasing function of $|t|$ toward unity. The former means that an arbitrary solution around the singular point can always reach the singular point by increasing $|t|$ to infinity from the property of solution curves for the node or the spiral point, for instance, as seen from Fig. 1(b) or Fig. 1(c), or Fig. 2(b) or Fig. 2(c). The latter means a state satisfying $h^2(x) = u^2 g^2(t)$, i.e., $\beta(x,t) = 0$, can never be realized with a change of t. Under this condition of $|x|$, we obtain a smoother solution curve starting at the point $(\phi,\phi_t) = (\phi_{0,2n-2},0)$ at $t = -\infty$, passing $\phi_{0,2n}$ at $t = 0$ with a certain value of ϕ_t, and ending at the point $(\phi_{0,2n+2},0)$ at $t = +\infty$. The solution curves around $x = 0$ in Figs. 8 and 11 correspond to this kind of solution.

If $|x|$ is larger than $|x_{c0}|$, $h(x)$ becomes larger than u. In

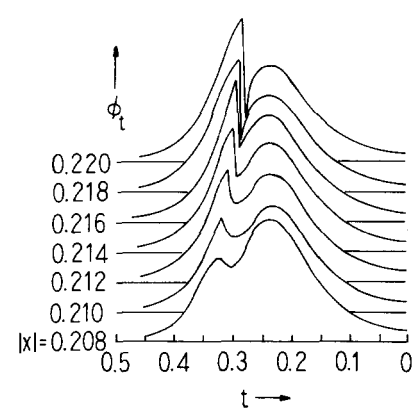


FIG. 13. Detailed growth process of the wedge-shaped distortion in the extended sine–Gordon system.

this case, the singularity at $|t| \to \infty$ is a saddle point from Eq. (11.2). Then, we understand that almost every solution curve is repelled before reaching it, except four routes from the property of solution curves for the saddle point, for instance, as seen from Fig. 1(a) or Fig. 2(a). This suggests that the route connecting $\phi_{0,2n-2}$ and $\phi_{0,2n+2}$ should be very limited. Moreover, with decreasing $|t|$, the situation can always be changed from $h^2(x) > u^2 g^2(t)$ to $h^2(x) < u^2 g^2(t)$ via the state

$$h^2(x_c) - u^2 g^2(t_c) = 0, \tag{11.3}$$

at $t = t_c(x)$, where $x = x_c(t)$. Equation (11.3) means that the position (x_c, t_c) is moving. Thus the moving singularity is established there. Around $t_c(x)$, the solution in the state plane is characterized from Eq. (2.10) by

$$[\eta(\phi) + \zeta(x,t)V(x,t)]/\beta(x,t). \tag{11.4}$$

When t approaches $t_c(x)$, $\beta(x,t)$ approaches zero. Then, the numerator in Eq. (11.4) should also approach zero in order to exhibit a definite value of the solution at $t_c(x)$. The method of bringing the denominator to zero is generally different from the method of bringing the numerator to zero, and they depend upon the values of x, J_B, and G. Thus the distortion is forced to be concentrated around $t_c(x)$. It is rather exceptional that no such distortions appear for the pure sine–Gordon system. This results from the fact that $V(\Xi)$ happens to become a single-valued function of ϕ. Thus in Eq. (11.4) for the pure sine–Gordon system the method of bringing the denominator to zero is always the same as the method of bringing the numerator to zero irrespective of the value of x, which, as well as Eq. (8.2), brings no existence of the moving singularity, and enables us to integrate Eq. (1.1).

The detailed growth process of the wedge-shaped distortion for $G = 0.018$ and $J_B = 0.4$ is depicted in Fig. 13. Once $|x|$ exceeds $|x_{c0}|$, a new peak appears on the left-hand side of the main peaks, and then the new peak and the valley between these peaks grow quickly as moving backward. Thus the discontinuity to produce the wedge-shaped distortion is established in a short time. We may also have the effect of this kind of singularities, that is, the moving singularities at $x_c(t)$ and $t_c(x)$ included in $X_0(t)$ and $T_0(x)$.

Finally, we add on the left-hand side of Eq. (1.1) the extra dissipation term expressed by $(1/R)\phi_{xxt}$, where R is a positive constant.[4] That is,

$$\phi_{xx} - \phi_{tt} - \sin\phi = -(1/R)\phi_{xxt} + G\phi_t - J_B. \tag{11.5}$$

The equation on the state plane corresponding to Eq. (11.5) is written as

$$\beta(x,t)\frac{\partial V}{\partial \phi} - \left(\frac{\eta(\phi)}{V(x,t)} + \zeta(x,t)\right) = -\frac{g(t)h^2(x)}{Ru^2}\left[V(x,t)\frac{\partial^2 V}{\partial \phi^2} + \frac{u}{h^2(x)} \times \left(\frac{1}{u}h^2(x)\frac{\partial V}{\partial \phi} - \frac{dh}{dx}\right)\frac{\partial V}{\partial \phi}\right]. \tag{11.6}$$

Equation (11.6) can be used for the analysis of the fluxon–antifluxon collision in the uniformly biased Josephson transmission line, along which the quasiparticle flow is taken into consideration. We assume that R is larger than a value, so that the effect of the right-hand side of Eq. (11.6) may be neglected as the first-order approximation, if the value of t is much different from t_c. However, as t approaches $t_c(x)$, the value of $\partial V/\partial \phi$ may be changed appreciably. This brings the increase of the absolute value of $\partial^2 V/\partial \phi^2$. Thus the effect of the right-hand side in Eq. (11.6) becomes most remarkable just around $t_c(x)$. In this case, the quantity concerned with the moving singularity is the coefficient on $\partial^2 V/\partial \phi^2$. However, the coefficient never vanishes around $t_c(x)$, that is, there appears no singularity around there. This indicates that the existence of R acts to smooth the local distortion.

XII. SUMMARY

Dynamic behavior of solutions in nonlinear Klein–Gordon systems with a dissipation and an external force, referred to as extended Klein–Gordon systems, is treated geometrically in a state plane by transforming the equation into three basic equations, each of which is associated with the derivative with respect to x, t, and ϕ, respectively.

Instead of ϕ, the solution is treated with the forms of ϕ_t and ϕ_x. They are divided into the sum of the traveling wave component expressed by $V(x,t)$ and the other wave component, respectively. Thus $V(x,t)$ is expressed by the form representing the stationary waves of a more generalized Klein–Gordon system consisting of nonlinear coordinates.

The boundary and initial conditions are imposed on the slopes at singular points at infinity on the traveling wave component in the state plane under the conditions that there exists initially either a stationary soliton or an antisoliton at $|x| = \infty$, and that finally either the stationary soliton or the antisoliton is again established at $|x| = \infty$. From the initial condition in $t \to -\infty$, the coexistence condition of solutions is determined. The solution is treated by dividing the region of t (or x) into two, t (or x) $\geqslant 0$ and t (or x) $\leqslant 0$, because of the existence of the singularity for $V(x,t)$ at t (or x) $= 0$.

It is shown from the properties of the field of directions that the soliton–antisoliton interaction is equivalent to the state of the reflection at the open end of the semi-infinite system, and the soliton–soliton interaction is equivalent to the state of the reflection at the short end of the equivalent semi-infinite transmission line to the system. However, as far as the external force and the dissipation are taken into consideration, the soliton–soliton interaction cannot be realized under the initial and the boundary conditions imposed here.

To verify the appropriateness of our model, it is first applied to the soliton–antisoliton interaction and the soliton–soliton interaction in the pure sine–Gordon system, and the well-known expressions are derived for them. The analytical method is described in detail for the soliton–antisoliton interaction in extended Klein–Gordon systems. As an example for the application of the analytical method, the extended sine–Gordon system is treated. From the boundary conditions, a definite region is clarified that would make possible coexistence of the soliton and the antisoliton in the relation between the dissipation coefficient and the external force. It is also found that when the soliton approaches an antisoliton up to a certain distance a local wedge-shaped distortion is generated in the soliton and in the antisoliton at the same time. The distortion moves to the direction opposing the soli-

ton movement, and disappears before the collision at the center takes place. When the soliton and the antisoliton recede from each other up to a certain distance, a thorn-shaped distortion is generated in front of each wave. When the distortion reaches the vicinity of the center of the wave, it disappears. Such distortions appear around the position satisfying the condition $\beta(x,t) = 0$, where the moving singularities appear, and the properties and the mechanisms of them are clarified in detail. It is shown that no such distortions appear in the pure sine–Gordon system on account of disappearance of the moving singularity, because $V(x,t)$ becomes a single-valued function of ϕ. The detailed mechanism of the distortion is clarified from the properties of the moving singularity in our derived equations.

It is finally proved that if the dissipation term proportional to ϕ_{xxt} is added to Eq. (1.1), the effect acts to smooth the distortions.

ACKNOWLEDGMENTS

The author is grateful to S. Sakai, K. Tomizawa, N. Hashizume, and H. Takayasu for helpful discussions.

[1]For instance, *Solitons and Condensed Matter Physics, Springer Series in Solid-State Sciences*, Vol. 8, edited by A. R. Bishop and T. Schneider (Springer, Berlin, 1978); *Solitons, Topics in Current Physics*, Vol. 17, edited by R. K. Bullough and P. J. Caudrey (Springer, Berlin, 1980).

[2]*Solitons in Action*, edited by K. Longren and A. C. Scott (Academic, New York, 1978).

[3]R. K. Bullough, P. J. Caudrey, and H. M. Gibbs, in *Solitons, Topics in Current Physics*, Vol. 17, edited by R. K. Bullough and P. J. Caudrey (Springer, Berlin, 1980).

[4]A. C. Scott, *Active and Nonlinear Wave Propagation in Electronics* (Wiley–Interscience, New York, 1970).

[5]M. J. Ablowitz, M. D. Kruskal, and J. F. Ladik, SIAM J. Appl. Math. **35**, 428 (1979).

[6]D. K. Cambell, J. F. Schonfeld, and C. A. Wingate, Physica D **9**, 1 (1983).

[7]M. Peyrard and D. K. Campbell, Physica D **9**, 33 (1983).

[8]K. Nakajima, Y. Onodera, T. Nakamura, and R. Sato, J. Appl. Phys. **45**, 4095 (1974).

[9]S. N. Erné and R. D. Parmentier, J. Appl. Phys. **51**, 5025 (1980).

[10]J. F. Curie, S. E. Trullinger, A. R. Bishop, and J. A. Krumhansl, Phys. Rev. B. **15**, 5567 (1977).

[11]P. S. Lomdahl, O. H. Soerensen, and P. L. Christiansen, Phys. Rev. B. **25**, 5737 (1982).

[12]M. B. Fogel, S. E. Trullinger, A. R. Bishop, and J. A. Krumhansl, Phys. Rev. B. **15**, 1978 (1977).

[13]D. Mclaughlin and A. C. Scott, Phys. Rev. A **18**, 1652 (1978).

[14]P. W. Kitchenside, A. L. Mason, R. K. Bullough, and P. J. Candrey, in *Solitons and Condensed Matter Physics, Springer Series in Solid-State Sciences*, Vol. 8, edited by A. R. Bishop and T. Schneider (Springer, Berlin, to be published).

[15]H. Tateno and S. Sakai, Jpn. J. Appl. Phys. **21**, L7 (1982).

[16]H. Tateno and S. Sakai, Jpn. J. Appl. Phys. **22**, 161 (1983).

[17]H. Tateno, Phys. Rev. B **36**, 2170 (1987).

[18]J. Rubinstein, J. Math. Phys. **11**, 258 (1970).

[19]S. Sakai and H. Tateno, Jpn. J. Appl. Phys. **22**, 1374 (1983).

[20]K. Nakajima, Y. Sawada, and Y. Onodera, J. Appl. Phys. **46**, 5272 (1975).

[21]M. Salerno, M. P. Soerensen, O. Kovgaard, and P. L. Christiansen, Wave Motion **5**, 49 (1983).

Soliton solutions in an integrable chiral model in 2+1 dimensions

R. S. Ward
Department of Mathematical Sciences, University of Durham, Science Laboratories, South Road, Durham DH1 3LE, United Kingdom

(Received 25 August 1987; accepted for publication 14 October 1987)

There is a modified SU(2) chiral model in 2 + 1 dimensions which is integrable. It admits multisoliton solutions, in which the solitons move at constant velocity, and pass through one another without scattering or changing shape.

I. INTRODUCTION

Most of the well-known examples of integrable systems admitting soliton solutions live in 1 + 1 dimensions. Some of these, such as the sine–Gordon equation and the chiral field equation,[1] are Lorentz invariant. That is to say, they are invariant under the action of SO(1,1) acting on space-time. In 2 + 1 dimensions, many examples of integrable systems are known (for example, the Kadomtsev–Petviashvili and Davey–Stewartson equations), but they are very far from being SO(2,1) invariant. There do exist integrable rotationally invariant systems in more than one space dimension [for example, Bogomolny–Prasad–Sommerfield (BPS) monopoles in three-space], but the "solitons" in these examples do not move (or to put it another way, the introduction of time dependence seems to destroy the integrability). There is no known example of a Lorentz-invariant system admitting genuine soliton solutions in more than one space dimension.

This paper does not remedy that situation, but describes a system which, in a sense, comes close to doing so. It is a modification of the SU(2) chiral field equation

$$\eta^{\mu\nu}\partial_\mu(J^{-1}\partial_\nu J)=0. \tag{1}$$

Here J is a map from $\mathbb{R}^3$ to SU(2), and should be thought of as a two-by-two unitary matrix of functions of the coordinates x^μ on $\mathbb{R}^3$. Greek indices range over the values 0, 1, 2, and ∂_μ denotes the partial derivative with respect to x^μ. The Einstein summation convention applies throughout, to Greek indices only. The tensor $\eta^{\mu\nu}$ is the (inverse) Minkowski metric, given by $\eta^{\mu\nu}=\mathrm{diag}(-1,1,1)$.

So Eq. (1) is Lorentz invariant in (2 + 1)-dimensional space-time. It does not seem to be integrable. But a modification is suggested by the observation that the self-dual Yang–Mills (sdYM) equations in four dimensions can be written in a form[2] which resembles (1), and the sdYM equations *are* integrable. So one can take the SU(2) sdYM equations in 2 + 2 dimensions (in 3 + 1 dimensions they are inconsistent), and reduce to 2 + 1 dimensions by requiring that the fields be independent of one of the coordinates. This yields the equation

$$\eta^{\mu\nu}\partial_\mu(J^{-1}\partial_\nu J)+V_\alpha\varepsilon^{\alpha\mu\nu}\partial_\mu(J^{-1}\partial_\nu J)=0, \tag{2}$$

where $\varepsilon^{\alpha\mu\nu}$ is the alternating tensor (with $\varepsilon^{012}=1$), and V_α is a constant unit vector. This equation is the subject of the paper. It is integrable, and as we shall see, it admits soliton solutions.

II. CONSERVATION OF ENERGY

In this section, it is shown that the system (2) admits an energy functional that is positive definite and conserved. But first, a couple of other remarks about (2) are in order. Equations (2) are not Lorentz invariant, because of the vector V_α [there is a residual symmetry group which is SO(2) or SO(1,1), depending on whether V_α is timelike or spacelike]. The solutions of (2) behave relativistically in some ways, and nonrelativistically in others: this will be illustrated later. The equation has the form

$$\Box J=\text{poly in } J,\ J^{-1},\text{ and } \partial_\mu J,$$

where $\Box=\eta^{\mu\nu}\partial_\mu\partial_\nu$ is the wave operator; so one sees that it is hyperbolic and quasilinear, and its characteristics are the light cones. The leading (second-order) term in the equation, namely $\Box J$, is Lorentz invariant; the symmetry-breaking term involving V_α contains only first derivatives of J.

The usual energy-momentum tensor relevant to the original equation (1) is

$$T_{\mu\nu}=(-\delta^\alpha_\mu\delta^\beta_\nu+\tfrac{1}{2}\eta_{\mu\nu}\eta^{\alpha\beta})\mathrm{tr}(J^{-1}J_\alpha J^{-1}J_\beta), \tag{3}$$

where δ^α_μ denotes the Kronecker delta, tr denotes trace, and J_α denotes $\partial_\alpha J$. Equation (1) implies that $T_{\mu\nu}$ is divergence-free. However, if we impose (2), then the divergence of $T_{\mu\nu}$ is

$$\partial^\mu T_{\mu\nu}=-\tfrac{1}{3}V_\nu\varepsilon^{\alpha\beta\gamma}\,\mathrm{tr}(J^{-1}J_\alpha J^{-1}J_\beta J^{-1}J_\gamma). \tag{4}$$

So $T_{\mu\nu}$ is not conserved, and neither, in general, is the energy-momentum vector $P_\mu=T_{\mu 0}$. Clearly the divergence of P_μ vanishes if and only if $V_0=0$.

Now it was mentioned above that V_μ can either be timelike (i.e., $\eta^{\mu\nu}V_\mu V_\nu=-1$) or spacelike (i.e., $\eta^{\mu\nu}V_\mu V_\nu=1$). When Eq. (2) was discussed by Manakov and Zakharov,[3] they (in effect) chose V_μ to be timelike. But this choice is incompatible with $V_0=0$. It seems hard to find an energy functional in this case, and it may be that one does not exist. From now on, therefore, let us take V_μ to be spacelike, with $V_0=0$; to be specific, let V_μ have components (0,1,0).

Consequently, we have $\partial^\mu P_\mu=0$, and so the energy

$$E=\int_{x^0}P_0\,dx^1\,dx^2,$$

which is the integral of the energy density P_0 over the spacelike plane $x^0=\text{const}$, is independent of x^0: energy is conserved. [Something has to be said about boundary condi-

 0022-2488/88/020386-04$02.50

tions: we require that J be everywhere smooth, and that $J = K + O(r^{-1})$ as $r \to \infty$, where K denotes a constant SU(2) matrix, and $r = \{(x^1)^2 + (x^2)^2\}^{1/2}$ measures distance in two-space.] It is worth emphasizing that E is a positive-definite functional of the field J.

The fact that the energy for solutions of (2) is the same as that for solutions of (1) is perhaps not surprising, since the additional term in (2) is analogous to a background magnetic field in classical mechanics,[4] and so does not affect the energy. A magnetic field influences the motion of a charged particle, but does no work on it.

III. THE ONE-SOLITON SOLUTION

In this section we examine a one-soliton solution of (2). Its shape is constant in time; all it does is move at constant velocity. It is given in terms of a complex parameter μ (which determines the velocity) and an arbitrary rational function f of one complex variable (which determines the shape). The expression for the matrix J is fairly simple, and is given in Appendix A [Eq. (A8)]. It suffices here to exhibit the energy density P_0,

$$P_0 = \Lambda |f'|^2 (1 + |f|^2)^{-2}, \tag{5}$$

where Λ is the constant

$$\Lambda = \tfrac{1}{2}|\mu|^{-4}|\mu - \bar{\mu}|^2 (1 + |\mu|^2)^2,$$

and where f is a rational meromorphic function of

$$\omega = x + \tfrac{1}{2}\mu(t + y) + \tfrac{1}{2}\mu^{-1}(t - y). \tag{6}$$

[The space-time coordinates x^μ are written as (t,x,y); and f' denotes the derivative of f.]

If $\mu = \pm i$, the $\omega = x \pm iy$, and the configuration is static. The energy, obtained by integrating (5), is $E = 8\pi N$, where N is the degree of the rational function f. These solutions are, of course, well-known: they correspond to the "instantons" of the two-dimensional O(3) σ model (or CP^1 model); or, in alternative language, they are the harmonic maps from S^2 (compactified $\mathbb{R}^2$) into SU(2). Roughly speaking (and generically), the solution looks like N static lumps, at arbitrary positions in the xy plane. For values of μ other than $\pm i$, there is time dependence, and the lumps move. But they all move with the same velocity, and so it makes sense, in this case, to regard the whole multilump configuration as a single soliton. To keep things simple, let us take $f(\omega) = \omega$. Thus $N = 1$, i.e., there is just one peak in the energy density. From (5), we see that this lump is located at $\omega = 0$, and from (6) we can compute its velocity. What one obtains is the following.

Write $\mu = me^{i\theta}$, and restrict $\sin\theta$ to be nonzero (i.e., μ to be nonreal). The velocity of the soliton in the xy plane is

$$(v_x, v_y) = \left(-\frac{2m\cos\theta}{1+m^2}, \frac{1-m^2}{1+m^2} \right) \tag{7}$$

and its speed v is given by

$$v^2 = 1 - (4m^2 \sin^2\theta)/(1+m^2)^2.$$

So $v < 1$: the lump can travel at any speed less than that of light. Let γ denote the usual relativistic factor $\gamma = (1 - v^2)^{-1/2}$. From (5) one deduces that the lump has total energy $E = 8\pi\gamma \sin^2\theta$, that its height is $8\gamma^2 \sin^4\theta$, that its width in the x direction is 1, and that its width in the y direction is $(\gamma \sin^2\theta)^{-1}$.

Let us examine two special cases, namely motion in the y direction and in the x direction. If $\theta = \pm\frac{1}{2}\pi$ (i.e., μ is pure imaginary), then the lump moves in the y direction, and its behavior is relativistic: the total energy is $8\pi\gamma$ (the rest-energy 8π multiplied by $\gamma > 1$), the x width is 1 (the same as at rest), and the y width is γ^{-1} (Lorentz–Fitzgerald contraction in the direction of motion). If, on the other hand, $m = 1$ (i.e., $\mu = e^{i\theta}$), then the lump moves in the x direction with speed $\cos\theta$, and its behavior is not relativistic. Its energy is $8\pi|\sin\theta|$, which is *less* than the rest energy; its height $8\sin^2\theta$ decreases if its speed is increased. For motion in other directions, the behavior is intermediate between these two cases.

To sum up, in the one-soliton sector, one has soliton solutions whose shape depends on $4N + 4$ real parameters (rational functions of degree N), and which move with any constant velocity (as long as the speed is less than that of light). The anisotropy of Eq. (2) shows up, in particular, in the way that the energy of the soliton depends on the direction of its motion.

IV. A TWO-SOLITON SOLUTION

There is a good candidate for an n-soliton, namely the solution (A7) derived in Appendix A. (It has not been checked in general that this has all the desired properties, such as smoothness.) Let us examine the two-soliton solution in more detail.

The $n = 2$ solution depends on two complex constants μ_1 and μ_2, and two holomorphic functions f_1 and f_2. Take these functions to be

$$f_k(\omega_k) = \omega_k + c_k,$$

where c_1 and c_2 are complex constants [cf. $f(\omega) = \omega$ in the previous section]. One's first guess, from the analysis of the previous section, might be that the corresponding solution represents two lumps L_1 and L_2, and that L_k travels at a velocity determined by μ_k according to Eq. (7). This is indeed what happens. Moreover, L_k travels along a straight line determined by μ_k and c_k: the two lumps do not scatter off each other. There is no change of direction or "phase shift" when they pass each other. One can see this as follows.

Fix a constant value of the time t, with $|t|$ very large. Then $|f_k|$ will be very large, except near the point ξ_k in the xy plane at which $\omega_k + c_k$ vanishes. It follows that the energy density P_0 is small except near these two points ξ_1 and ξ_2. This can be made more precise by setting $m_a^1 = (0,1)$, in the notation of Appendix A. (In effect, this amounts to taking the asymptotic limit $|t| \to \infty$ while keeping our eyes on the lump L_2, so that $|f_1| \to \infty$ but f_2 remains finite.) In this limit, P_0 is given by the simple expression

$$P_0 = B(A + |f_2|^2)^{-2}, \tag{8}$$

where A and B are the positive constants

$$A = |(\mu_1 - \bar{\mu}_2)/(\mu_1 - \mu_2)|^2,$$

$$B = \tfrac{1}{2}A\,|\mu_2|^{-4}(1 + |\mu_2|^2)^2|\mu_2 - \bar{\mu}_2|^2.$$

One sees from (8) that P_0 has a local maximum at the point

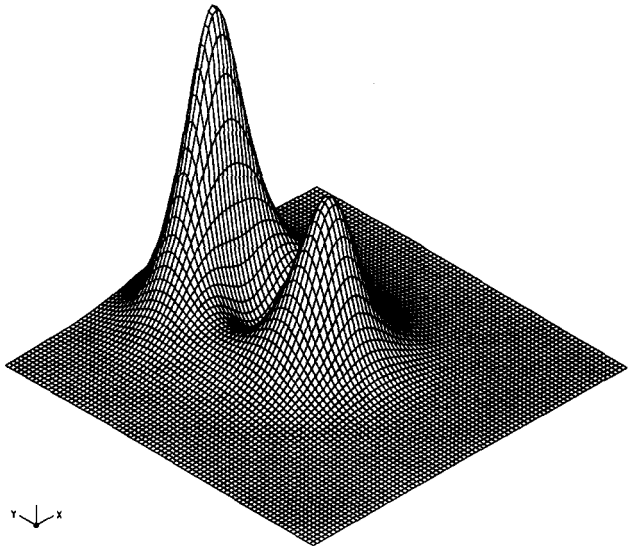


FIG. 1. A two-soliton configuration at time $t = -10$. The lump at the origin (in the middle of the square) is stationary, and the other (taller) one is moving towards it along the y axis.

ξ_2 where f_2 vanishes. By definition, the lump L_2 is located at this point.

Now as t varies, the point ξ_2 moves along the straight line (referred to above) given by

$$x + \tfrac{1}{2}\mu_2(t+y) + \tfrac{1}{2}\mu_2^{-1}(t-y) + c_2 = 0.$$

So the lump L_2 moves along this line at constant velocity, and does not scatter. Neither, of course, does L_1.

By way of example, let us look at the special case $c_1 = c_2 = 0$, $\mu_1 = i$, $\mu_2 = 2i$. The resulting expressions for J and for the energy density are given in Appendix B, in order to illustrate the sort of expressions that occur. The first lump L_1 remains stationary at the origin, while L_2 moves along the y axis with speed $\frac{3}{5}$. At time $t = 0$ the two lumps coincide, and form a single peak. Figure 1 represents a snapshot of the solution at time $t = -10$.

V. CONCLUDING REMARKS

It would be nice to have a Lorentz-invariant soliton system in $2+1$ (or, even better, $3+1$) dimensions, but no such system is known. The equation discussed in this paper is a modification of a Lorentz-invariant one. Its solitons maintain their shape and do not scatter off one another; this point was also emphasized by Manakov and Zakharov[3] in their study of the related equation with V_α timelike. Note that the localized solitons of the Kadomtsev–Petviashvili (KP) equation also do not scatter.

It is worth pointing out that in the "unmodified" chiral equation (1), scattering *does* occur: this follows, for example, from a study of the closely related O(3) σ model (or CP^1 model) in $2+1$ dimensions.[5]

APPENDIX A: GENERAL SOLITON SOLUTIONS

The method for generating soliton solutions is that of the "Riemann problem with zeros," developed by Zakharov and his co-workers. What follows is an adaptation of the procedure as described by Forgács *et al.*[6]

Let ζ be a complex parameter, and (u,v,x) real variables (coordinates on $\mathbb{R}^3$). Let A and B be two-by-two anti-Hermitian trace-free matrices, depending on u, v, x but not on ζ. Consider the set of linear equations

$$\begin{aligned}(\zeta\partial_x - \partial_u)\psi &= A\psi,\\ (\zeta\partial_v - \partial_x)\psi &= B\psi,\end{aligned} \tag{A1}$$

where $\psi = \psi(u,v,x)$ is a two-by-two matrix satisfying $\det\psi = 1$, and the reality condition

$$\psi(u,v,x,\bar{\zeta})^* = \psi(u,v,x,\zeta)^{-1}. \tag{A2}$$

Here * denotes complex conjugate transpose. It is easily checked that (A2) is consistent with (A1): each equation in (A1) is in effect "real." But of course the system (A1) is overdetermined, and in order for a solution ψ to exist, A and B have to satisfy integrability conditions, which are

$$B_x = A_v, \quad A_x - B_u - [A,B] = 0. \tag{A3}$$

If we put $J(u,v,x) = \psi(u,v,x,0)^{-1}$, then Eq. (A3), and (A1) at $\zeta = 0$, give

$$\partial_x(J^{-1}\partial_x J) - \partial_v(J^{-1}\partial_u J) = 0. \tag{A4}$$

Now set $u = \frac{1}{2}(t+y)$, $v = \frac{1}{2}(t-y)$: then (A4) becomes Eq. (2), with $x^\mu = (t,x,y)$, and with $V_\mu = (0,1,0)$. Note also that J takes values in SU(2), by virtue of (A2).

This means that we can generate solutions of (2) with the aid of the overdetermined linear system (A1). To obtain the solutions that we want, we assume that $\psi(\zeta)$ (the dependence on u, v, x is to be understood) has the form

$$\psi_{ab}(\zeta) = \delta_{ab} + \sum_{k=1}^{n} (\zeta - \mu_k)^{-1} n_a^k m_b^k. \tag{A5}$$

Here a, b range over 1, 2 and label the rows and columns of ψ; $\mu_1,\ldots,\mu_n$ are complex constants; and $n_1^1,\ldots,m_1^1,\ldots$ are complex-valued functions of u, v, x (and not of ζ).

First, let us impose the unitarity condition (A2). The product

$$\sum_{b=1}^{2} \psi_{ab}(\zeta)\psi_{bc}(\bar{\zeta})^*$$

should equal δ_{ac}, so its apparent poles at each of the $\bar{\mu}_l$ should be removable. This is the case if and only if

$$n_a^k = -\sum_{l=1}^{n} (\Gamma^{-1})^{kl}\bar{m}_a^l,$$

where Γ^{-1} is the inverse of

$$\Gamma^{kl} = \sum_{a=1}^{2} (\bar{\mu}_k - \mu_l)^{-1}\bar{m}_a^k m_a^l. \tag{A6}$$

Then it is easily checked that this necessary condition for unitarity (A2) is also sufficient.

Next, we impose (A1), the crucial point being that the matrices A and B defined by (A1) should be independent of ζ. Again, requiring that apparent poles be removable gives equations (differential equations for the m_a^k), and these are satisfied if $m_a^k = m_a^k(\omega_k)$ depends on x, u, v only via the combination

$$\omega_k = x + \mu_k u + \mu_k^{-1} v$$

(i.e., m_a^k is a holomorphic function of ω_k). There is a homo-

geneity property, namely that if we multiply $m_a^k = (m_1^k, m_2^k)$ by a function λ_k, then ψ does not change. So we may set $m_1^k = 1$, $m_2^k = f_k$.

The solution ψ therefore depends on n constants μ_k and n holomorphic functions $f_k = f_k(\omega_k)$. The latter are only "locally" holomorphic at the moment, i.e., not entire: we have not yet imposed the requirement of smoothness and the boundary condition.

The above is not quite in its final form, since, as it stands,

$$\alpha := \det \psi = \prod_{k=1}^{n} \frac{\bar{\mu}_k}{\mu_k};$$

but dividing by the square root of this constant α achieves $\det \psi = 1$ without disturbing (A1) or (A2). So finally we get an expression for (the inverse of) J by evaluating $\psi(\zeta)$ at $\zeta = 0$:

$$(J^{-1})_{ab} = \alpha^{-1/2}\left(\delta_{ab} + \sum_{k,l} \mu_k^{-1} (\Gamma^{-1})^{kl} \bar{m}_a^l m_b^k\right). \tag{A7}$$

By way of example, take the case $n = 1$, with $m_a = (1, f)$. Then Γ (a one-by-one matrix) is

$$\Gamma = (1 + |f|^2)/(\bar{\mu} - \mu)$$

(μ must be nonreal), and (A7) gives

$$J = \frac{1}{|\mu|(1+|f|^2)} \begin{bmatrix} \mu + \bar{\mu}|f|^2 & (\mu - \bar{\mu})f \\ (\mu - \bar{\mu})\bar{f} & \bar{\mu} + \mu|f|^2 \end{bmatrix}. \tag{A8}$$

This expression is smooth everywhere, and satisfies the boundary condition at infinity, provided f is a rational function of

$$\omega = x + \tfrac{1}{2}\mu(t+y) + \tfrac{1}{2}\mu^{-1}(t-y).$$

To compute the energy density P_0 is just a matter of algebra: one gets

$$P_0 = \tfrac{1}{2}|\mu - \bar{\mu}|^2 |\mu|^{-4} (1 + |\mu|^2)^2 |f'|^2 (1 + |f|^2)^{-2}, \tag{A9}$$

where f' is the derivative of f.

APPENDIX B: A TWO-SOLITON SOLUTION

In the formulas of Appendix A, put $n = 2$, $\mu_1 = i$, $\mu_2 = 2i$. Then the unitary matrix J is given, in terms of the two holomorphic functions f_1 and f_2, by

$$\begin{aligned} J_{11} &= \Delta^{-1}\{|f_1 f_2|^2 + 16 f_1 \bar{f}_2 + 4\bar{f}_1 f_2 \\ &\qquad - 9|f_1|^2 - 9|f_2|^2 + 1\}, \\ J_{21} &= 6\Delta^{-1}\{|f_1|^2 \bar{f}_2 - \bar{f}_1 |f_2|^2 + \bar{f}_1 - \bar{f}_2\}, \\ J_{22} &= \bar{J}_{11}, \quad J_{12} = -\bar{J}_{21}, \end{aligned}$$

where

$$\Delta = |f_1 f_2|^2 + 9|f_1|^2 + 9|f_2|^2 - 8 f_1 \bar{f}_2 - 8\bar{f}_1 f_2 + 1.$$

Take $f_1(\omega_1) = \omega_1$ and $f_2(\omega_2) = \omega_2$. Since Δ is nowhere zero, J is smooth everywhere; and it satisfies the required boundary condition at spatial infinity. The energy density is

$$\begin{aligned} P_0 = 9\Delta^{-2} \operatorname{Re}\{&25|\omega_1|^4/2 + 8|\omega_2|^4 - 18\omega_1^2 \bar{\omega}_2^2 \\ &+ 2\omega_1^2 + 2\omega_2^2 - 36\omega_1 \bar{\omega}_2 - 4\omega_1 \omega_2 + 25|\omega_1|^2 \\ &+ 16|\omega_2|^2 + \tfrac{25}{2}\}. \end{aligned} \tag{B1}$$

It is this function which is plotted in Fig. 1, at time $t = -10$. It is easy to see from (B1) that for $|t|$ large, the energy density is peaked at the two points $(x,y) = (0,0)$ and $(x,y) = (0, -3t/5)$: one lump is stationary at the origin, and the other moves along the y axis at constant speed $\frac{3}{5}$.

[1] V. E. Zakharov and A. V. Mikaĭlov, Sov. Phys. JETP **47**, 1017 (1978).
[2] K. Pohlmeyer, Commun. Math. Phys. **72**, 37 (1980).
[3] S. V. Manakov and V. E. Zakharov, Lett. Math. Phys. **5**, 247 (1981).
[4] Y. S. Wu and A. Zee, Nucl. Phys. B **272**, 322 (1986).
[5] R. S. Ward, Phys. Lett. B **158**, 424 (1985).
[6] P. Forgács, Z. Horvath, and L. Palla, Nucl. Phys. B **229**, 77 (1983).

The evolution of perturbations of the renormalized long wave equation[a)]

C. V. Easwaran and S. R. Majumbar
Department of Mathematics and Statistics, University of Calgary, Calgary, Alberta, Canada

(Received 18 November 1986; accepted for publication 17 June 1987)

The evolution of the solitary wave solution of the perturbed renormalized long wave equation $u_t + 6uu_x - u_{xxt} = \epsilon u$ is considered using two timing and matched asymptotic expansions. As in the case of the perturbed KdV equation, it is found that behind the slowly varying solitary wave there are two distinct regions, a near tail and a far tail. The far tail is given by an exponentially decaying expression in contrast to the KdV far tail which is oscillatory.

I. INTRODUCTION

In recent years there has been considerable interest in the study of perturbed evolution equations. These equations govern physical phenomena such as wave propagation in a slowly varying medium, waves in a channel of varying cross section, and solitary waves moving along a sloping beach. Karpman[1] studied the perturbed KdV equation by using perturbations on the inverse scattering method, while Grimshaw[2] and Johnson[3] used a series expansion to determine the evolution of the perturbed KdV equation. In the most recent work Smyth[4] used two timing and matched asymptotic expansions to determine the evolution of the perturbed KdV equation

$$u_t + 6uu_x + u_{xxx} = \epsilon u, \tag{1}$$

where ϵ is a small parameter. He finds two distinct regions behind the slowly varying solitary wave: (i) a near tail that eventually breaks up into new solitary waves and that together with the soliton conserves the KdV mass, and (ii) a far tail that makes no contribution, to $O(\epsilon)$, to the mass conservation.

Our aim in this paper is to examine the evolution of solitary waves of the perturbed renormalized long wave (RLW) equation (or the BBM equation after Benjamin *et al.*[5])

$$u_t + 6uu_x - u_{xxt} = \epsilon u, \tag{2}$$

using the asymptotic expansion techniques introduced by Smyth.[4] The RLW equation, with $\epsilon = 0$ in (2), is an alternative model equation for long waves and derives from the Boussinesq equation following the same assumptions used to derive the KdV equation (in fact, one can use $u_t \simeq -cu_x$ in the dispersive correction term, remaining within the approximation for long waves, to obtain the RLW equation; here c is the wave velocity). Recent interest in the RLW equation comes from the numerical experiments showing the inelastic scattering properties of its solitary waves. In fact, the RLW equation has only three nontrivial conservation laws depending smoothly on u and its derivatives whereas the KdV equation has an infinite number of conservation laws. However, in certain theoretical investigations, the RLW equation is superior as a model for long waves (see Benjamin *et al.*[5] for a discussion of regularity properties of the RLW equation compared to the KdV equation for the same initial data).

Since the perturbed evolution equations are crucial in many physical phenomena, it is worth investigating how a different model equation (in this case the RLW instead of KdV) affects the system properties. In this paper we have chosen the form ϵu on the right of (2) simply for the sake of brevity. One could, for example, consider instead $-\epsilon u$ if a small damping is present or $-\epsilon u_{xx}$ in the presence of heat conduction. For all these latter cases, however, the analysis is essentially the same.

As in the case of the perturbed KdV equation it will be found that the slowly varying solitary wave solution of (2) does not conserve mass. It is then assumed that there is a "near tail" region just behind the solitary wave caused by a mass flux from it. Behind the "near tail" there will be another "far tail" region governed by the linearized form of (2). The essential difference between the present analysis of the perturbed RLW equation and the perturbed KdV equation as given in Smyth[4] will be in the far tail region. The far tail will be found to be exponentially decaying as $x \to \infty$ while for the perturbed KdV it is oscillatory, given by an Airy function.

Both Grimshaw and Johnson also note that the slowly varying soliton expansion for the perturbed KdV equation is nonuniform as $x \to -\infty$. They then introduce a new expansion assumed to be valid in the entire region behind the soliton. Physically, two different regions behind the soliton are considered more favorable, since the near tail is due to mass flux from the soliton while the far tail results from initial conditions and does not contribute to mass conservation.

II. THE SOLITARY WAVE

We will assume that the solution of (2) consists of a main solitary wave with slowly varying parameters given by the expansion

$$u = u^0(\theta,T) + \epsilon u^1(\theta,T) + \cdots, \tag{3}$$

where

$$T = \epsilon t, \quad \theta = x - C(T)/\epsilon, \qquad C_T = \omega_0(T) + \epsilon^2\omega_2(T) + \cdots. \tag{4}$$

Substituting (3) in (2) the zeroth-order equation is

$$-\omega_0 u^0_\theta + 6u^0u^0_\theta - \omega_0 u^0_{\theta\theta\theta} = 0. \tag{5}$$

This has a solitary wave solution

$$u^0 = \eta(T)\operatorname{sech}^2[\tfrac{1}{2}\theta], \tag{6}$$

where

[a)] This paper was presented at the Seventh Canadian Symposium on Fluid Dynamics held in Sackville, New Brunswick, 2–4 June 1986.

 0022-2488/88/020390-03$02.50

$$\eta(T) = \tfrac{1}{2}\omega_0(T). \tag{7}$$

It is worth observing that the solitary wave speed depends linearly on the amplitude unlike the KdV soliton for which $\omega_0(T) = 4(\eta_T^2)$.

One can show by simple conservation arguments that the solitary wave (6) does not conserve mass. This is because the perturbed RLW equation (2) has the energy conservation law

$$\frac{1}{2}\frac{d}{dt}\left[\int_{-\infty}^{\infty}(u^2 + u_x^2)dx\right] = \epsilon\int_{-\infty}^{\infty} u^2\,dx, \tag{8}$$

if $u \to 0$ as $x \to \pm\infty$.

Using (6) we have

$$\eta_T = \tfrac{5}{6}\eta$$

or

$$\eta = \eta_0 e^{(5/6)T}, \quad \text{where } \eta_0 \text{ is a constant.} \tag{9}$$

Equation (2) also has a mass conservation equation

$$\frac{d}{dt}\int_{-\infty}^{\infty} u\,dx = \epsilon\int_{-\infty}^{\infty} u\,dx. \tag{10}$$

But using (9) we have

$$\frac{d}{dt}\int_{-\infty}^{\infty} u\,dx = \frac{10}{3}\epsilon\eta, \tag{11}$$

$$\epsilon\int_{-\infty}^{\infty} u\,dx = 4\epsilon\eta. \tag{12}$$

Thus we see that the slowly varying solitary wave (6) conserves energy, but not mass. To make up for this it will be assumed that there is a tail region behind the solitary wave.

We now take up a detailed formal asymptotic analysis, following Smyth.[4] First it will be shown that the expansion (3) is not uniformly valid as $x \to -\infty$. The $O(\epsilon)$ equation resulting from substituting (3) in (2) is

$$\omega_0 u^1_{\theta\theta\theta} + 6u^0u^1_\theta - \omega_0 u^1_\theta + 6u^0_\theta u^1 - u^0_{T\theta\theta} + u^0_T = u^0. \tag{13}$$

This has adjoint

$$\omega_0 v_{\theta\theta\theta} + 6u^0 v_\theta - \omega_0 v_\theta = 0. \tag{14}$$

Multiplying (13) by v, (14) by u^0, and adding, we get

$$\omega_0\left[(v_{\theta\theta}u^1)_\theta + (u^1_{\theta\theta}v)_\theta - (v_\theta u^1_\theta)_\theta\right] + 6(u^0u^1v)_\theta - \omega_0(u^1v)_\theta = (u^0 - u^0_T + u^0_{T\theta\theta})v \tag{15}$$

and, integrating from $-\infty$ to ∞ w.r.t. θ,

$$-\omega_0\left[u^1v - u^1v_{\theta\theta} - u^1_{\theta\theta}v - u^1_\theta v_\theta\right]_{-\infty}^{\infty} = \int_{-\infty}^{\infty}(u^0 - u^0_T + u^0_{T\theta\theta})v\,d\theta. \tag{16}$$

We require that $u^1 \to 0$ as $\theta \to \infty$ and that u^1 is bounded as $\theta \to -\infty$. The bounded solutions of (14) are $v = u^0$ and $v = 1$. When $v = u^0$, from (16) we have

$$0 = \int_{-\infty}^{\infty}(u^0 - u^0_T + u^0_{T\theta\theta})u^0\,d\theta. \tag{17}$$

Using (6) this gives

$$\eta_T = \tfrac{5}{6}\eta \tag{18}$$

as previously found.

Now consider the solution $v = 1$. If we assume that $u^1 \to 0$ as $\theta \to -\infty$ we would have, from (16),

$$0 = \int_{-\infty}^{\infty}(u^0 - u^0_T + u^0_{T\theta\theta})d\theta, \tag{19}$$

which would give an expression for η different from (18). We thus see that as $\theta \to -\infty$, $u^{(1)}$ does not tend to zero but tends to a constant value given by

$$2\eta u^1 = \int_{-\infty}^{\infty}(u^0 - u^0_T + u^0_{T\theta\theta})d\theta = \tfrac{2}{3}\eta. \tag{20}$$

Thus as $\theta \to -\infty$, $u^{(1)}$ tends to a constant $\frac{1}{3}$, although $u^0 \to 0$, and the expansion (3) is not uniformly valid. This will be rectified by matching the expansion (3) with an outer expansion.

III. THE NEAR TAIL

The outer expansion for the near tail region just behind the soliton is assumed to depend on the slow scales $X = \epsilon x$, $T = \epsilon t$. Thus an expansion of the form

$$u = \epsilon V_1(X,T) + \epsilon^2 V_2(X,T) + \cdots \tag{21}$$

is considered.

From the perturbed RLW equation (2) we have

$$(V_1)_T = V_1, \tag{22}$$

hence

$$V_1(X,T) = A(X)e^T, \tag{23}$$

where $A(X)$ is to be determined by matching with the inner solution. The matching has to be done with a moving solitary wave which has a speed

$$\omega_0 = 2\eta(T). \tag{24}$$

Hence we have, using (10),

$$\frac{dX}{dT} = 2\eta(T) = 2\eta_0 e^{(5/6)T}. \tag{25}$$

Integrating, we get

$$\epsilon x = \tfrac{12}{5}\eta_0(e^{(5/6)T} - 1). \tag{26}$$

Thus the solitary wave is at position X when

$$T = \tfrac{6}{5}\ln\left[1 + \tfrac{5}{12}x/\eta_0\right]. \tag{27}$$

Since $u_1 \to \frac{1}{3}$ as $\theta \to -\infty$, the matching requires that

$$\tfrac{1}{3} = A(X)e^T, \tag{28}$$

when T is given by (27).

Thus we see that

$$A(X) = 1/3\left[1 + \tfrac{5}{12}x/\eta_0\right]^{6/5}. \tag{29}$$

We notice that the near tail expansion cannot be valid for all times because of the exponential growth in time in (23). In the case of perturbed KdV equation the near tail breaks up eventually into new solitons and we expect the same behavior to take place here, too. We do not go into the details of the breakup since it has to be determined numerically. It can be checked at this stage that the near tail together with the solitary wave conserves the RLW mass to $O(\epsilon)$.

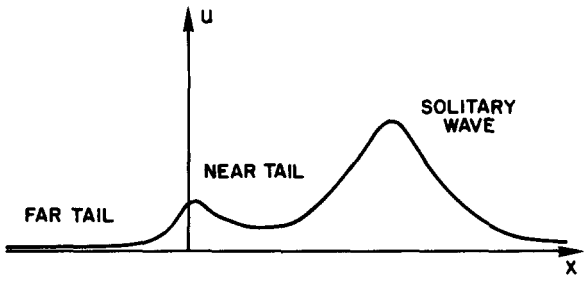


FIG. 1. Perturbed RLW solitary wave.

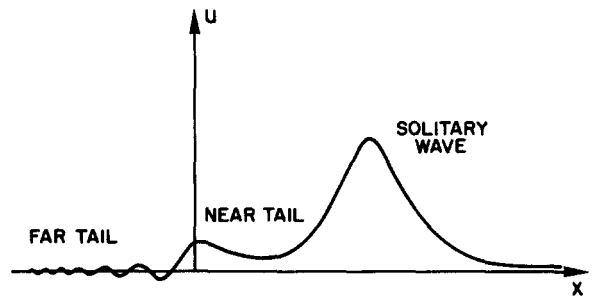


FIG. 2. Perturbed KdV solitary wave.

IV. THE FAR TAIL

Since the solitary wave started at $x = 0$ we expect that the near tail will extend from $x = 0$ till $x = x_s$, the position of the solitary wave. The region $x < 0$ will be called a far tail and it will be assumed to have an expansion of the form

$$u = \epsilon U_1(x,t,T) + \epsilon^2 U_2(x,t,T) + \cdots . \tag{30}$$

Using this in (2), we get the first-order equation

$$(U_1)_t - (U_1)_{xxt} = 0. \tag{31}$$

This has a similarity solution of the form

$$U_1 = B(t,T)e^x . \tag{32}$$

The function $B(t,T)$ may be determined by matching with the near tail solution at $x = 0$. Using (23) and (29) we find

$$\tfrac{1}{3}e^T = B(t,T),$$

hence

$$U_1 = \tfrac{1}{3}e^{(T+x)}. \tag{33}$$

This far tail expression differs significantly from the KdV far tail found in Smyth.[4] There it is seen that

$$U_1 = \frac{e^T}{3\eta_0}\int_{-\infty}^{x/(3t)^{1/2}} \mathrm{Ai}(s)ds, \tag{34}$$

where Ai is the Airy function.

Figure 1 is a schematic diagram of the perturbed RLW solitary wave. Figure 2 shows the perturbed KdV solitary wave (see Smyth[4]). As we see, the significant difference occurs in the far tail—whereas the KdV far tail is oscillatory, the RLW far tail decays exponentially as $x \to -\infty$.

So far we have only examined $O(\epsilon)$ equations for the far tail. The long time evolution of the far tail is also undetermined at this stage. Nevertheless one can make certain observations without going into greater detail. Using the expansion (30) at the second order we get, for the perturbed KdV equation,

$$(U_2)_{xx} + (U_2)_t + 6U_1(U_1)_x + (U_1)_T = U_1 \tag{35}$$

and, for the RLW,

$$-(U_2)_{txx} + (U_2)_t + 6U_1(U_1)_x - (U_1)_{Txx} + (U_1)_T = U_1. \tag{36}$$

Thus we see that the derivatives of U_1 are involved at the second order. Because of the nature of the Airy function the expansion (30) leads to secularities at this order for the KdV equation. However, no such secularities arise for the RLW equation because of the exponential decay of the derivatives of (34) as $x \to -\infty$.

ACKNOWLEDGMENT

This research was supported by Grant A5294 from the Natural Sciences and Engineering Research Council of Canada.

[1] V. I. Karpman, "Soliton evolution in the presence of perturbations," Phys. Scr. **20**, 462 (1979).
[2] R. Grimshaw, "Slowly varying solitary waves, I-Kortweg–de Vries equation," Proc. R. Soc. London Ser. A **368**, 359 (1979); in *Nonlinear Waves*, edited by L. Debnath (Cambridge U. P., Cambridge, 1983), p. 44.
[3] R. Johnson, "On the asymptotic solution of the K–dV equation with slowly varying coefficients," J. Fluid Mech. **60**, 813 (1971).
[4] N. F. Smyth, Ph.D. thesis, California Institute of Technology, 1984.
[5] T. B. Benjamin, J. L. Bona, and J. T. Mahony, "Model equations for long waves in nonlinear dispersive systems," Phil. Trans. R. Soc. London Ser. A **272**, 47 (1972).

Galilean quantum kinematics

Jorge Krause
Facultad de Física, Pontificia Universidad Católica de Chile, Casilla 6177, Santiago 22, Chile

(Received 1 October 1986; accepted for publication 7 October 1987)

The Galilean symmetry of a free particle in one-dimensional space is examined under the scope of non-Abelian quantum kinematics. Within the Hilbert space that carries the regular ray representation of the Galilei group the Schrödinger operator appears as one of the three fundamental invariants of the extended kinematic algebra. By means of a superselection rule the physical Hilbert subspaces of the system are identified, in which a complementary ray representation of the Galilean transformation produces the time-dependent Schrödinger equation, and the Feynman space-time propagator. The quantization approach used in this paper is purely group theoretic and relativistic.

I. INTRODUCTION

This paper is a brief report of work in progress concerning the physical assessment of non-Abelian quantum kinematics.[1] Continuing previous work, we devote this paper to the quantum kinematic theory of the Galilei group. In another paper[2] belonging to this line of research the quantum kinematic formalism of non-Abelian Lie groups (i.e., Ref. 1) was applied successfully to the one-dimensional harmonic oscillator. In fact, it was shown in Ref. 2 that the usual quantum model of the harmonic oscillator can be obtained directly from the exclusive kinematic consideration of its Newtonian symmetries.

It is the aim of this paper to repeat, step by step, the same kind of group-quantization process in the case of a scalar Newtonian free particle moving in one-dimensional space. Undoubtedly, the Galilei group (even if one keeps aside rotations[3] and improper transformations) plays such an outstanding role in "nonrelativistic" quantum mechanics[4] that the *direct quantization* of this group will help us to understand the physical possibilities of quantum kinematics, in general. (For the underlying physical motivation of the present paper, we refer the reader to our previous work.[1,2])

The organization of this paper follows that of Ref. 2. In Sec. II we present a rather brief account of the (well-known) regular ray representations of the Galilei group. Next, the extended Lie algebra and the non-Abelian commutators (namely, the extended kinematic algebra) are also discussed in Sec. II. Section III introduces the superselection rules that permit the identification of the Hilbert subspace of physical significance. In Sec. IV we discuss the complementary space-time representation of the Galilei group, and we finish our group-theoretic analysis obtaining the time-dependent Schrödinger equation and the corresponding space-time propagator of a free particle in one-dimensional space. Section V contains some concluding remarks.

II. REGULAR RAY REPRESENTATIONS AND KINEMATIC ALGEBRA OF THE GALILEI GROUP IN ONE-DIMENSIONAL SPACE

Let us introduce a two-dimensional space-time $\{(t,x)\}$, in which the "Galilei" group G_1 becomes realized as the following transformation of variables:

$$t' = t + q^1, \quad x' = x + q^2 - q^3 t. \tag{2.1}$$

Clearly, G_1 is the Newtonian relativity group[5] of a free particle (in one space dimension) and Eq. (2.1) entails the special Newtonian relativity theory of the system.

Next, we shall quantize this system directly through its Galilean symmetries. Our first task is to settle the *regular ray representations* of G_1. Since these are well known, we present this issue in a very sketchy fashion, at least in order to introduce our notation. The group manifold $M(G_1)$ is defined by $-\infty < q^a < +\infty$, $a = 1,2,3$, and the group law[6] is as follows:

$$\begin{aligned} q''^1 &= g^1(q';q) = q'^1 + q^1, \\ q''^2 &= g^2(q';q) = q'^2 + q^2 - q'^3 q^1, \\ q''^3 &= g^3(q';q) = q'^3 + q^3. \end{aligned} \tag{2.2}$$

Hence Lie's infinitesimal right operators in $M(G_1)$ are given by

$$X_1 = \partial_1, \quad X_2 = \partial_2, \quad X_3 = \partial_3 - q^1 \partial_2; \tag{2.3}$$

wherefrom one obtains the familiar Lie algebra[7]

$$[X_1,X_2] = 0, \quad [X_2,X_3] = 0, \quad [X_3,X_1] = X_2. \tag{2.4}$$

Since G_1 is unimodular one defines the Hilbert space $\mathscr{H}(G_1)$, which carries the ("true") regular representation of G_1, as the set of all complex functions $\psi(q^1,q^2,q^3)$ defined on $M(G_1)$, and such that[8]

$$\langle\psi|\psi\rangle = \mu_0 \int\int\int dq^1\, dq^2\, dq^3 |\psi(q^1,q^2,q^3)|^2 < \infty, \tag{2.5}$$

where μ_0 is an arbitrary normalization real constant. Within the *rigged* Hilbert space structure associated with $\mathscr{H}(G_1)$ let us introduce a complete orthogonal continuous basis $\{|q^2,q^2,q^3\rangle\}$ (cf. Ref. 1):

$$\langle q'^1,q'^2,q'^3|q^1,q^2,q^3\rangle = \mu_0^{-1}\delta(q'^1 - q^1)\delta(q'^2 - q^2)\delta(q'^3 - q^3), \tag{2.6}$$

$$\mu_0 \int\int\int dq^1\, dq^2\, dq^3 |q^1,q^2,q^3\rangle\langle q^1,q^2,q^3| = I, \tag{2.7}$$

where I stands for the identity operator in $\mathscr{H}(G_1)$. Plainly, one has $\psi(q^1,q^2,q^3) = \langle q^1,q^2,q^3|\psi\rangle$, for every vector $|\psi\rangle \in \mathscr{H}(G_1)$.

An admissible two-cocycle of G_1 is given by[9,10]

$$\phi_k(q';q) = \tfrac{1}{2}k\,[q'^2 q^3 - q'^3(q^2 + q^1 q^3)], \tag{2.8}$$

which is well known indeed. Hence the right-exponent gen-

 0022-2488/88/020393-05$02.50

erators[9] associated with $\phi_k(q';q)$ [namely,

$$r_a^{(k)}(q) = \lim_{q'\to 0} \partial'_a \phi_k(q';q), \quad a = 1,2,3]$$

are given by

$$r_1^{(k)}(q) = 0, \quad r_2^{(k)}(q) = \tfrac{1}{2}kq^3, \quad r_3^{(k)}(q) = -\tfrac{1}{2}k(q^2 + q^1q^3). \tag{2.9}$$

In this manner, one is ready to introduce a set of unitary operators $U_k(q)$, corresponding to the *ray extensions* of the regular representation carried by $\mathscr{H}(G_1)$. We define these operators as follows (cf. also Ref. 2):

$$U_k(q'^1,q'^2,q'^3)|q^1,q^2,q^3\rangle = e^{ik/2[q'^2q^3 - q'^3(q^2 + q^1q^3)]} \times |q'^1 + q^1, q'^2 + q^2 - q'^3q^1, q'^3 + q^3\rangle, \tag{2.10}$$

$$U_k(q^1,q^2,q^3)|0,0,0\rangle = |q^1,q^2,q^2\rangle, \tag{2.11}$$

wherefrom one gets a regular ray representation of G_1. These operators are consistent with the representation of the identity operator I stated in Eq. (2.11). Let us remark that the same Hilbert space $\mathscr{H}(G_1)$, which carries the regular vector representation of G_1, carries all the ray extensions of this representation. (As a matter of fact, this theorem corresponds to a general property of all Lie groups.) We have now finished with the regular ray representations of G_1.

Now we come to one of the main points of non-Abelian quantum kinematics, since we are ready to consider the extended Lie algebra and the generalized canonical commutators of G_1. First, let us consider the generators $P_a^{(k)}$, $a = 1,2,3$, of the unitary ray operators $U_k(q)$. These are defined in the usual manner, i.e.,

$$U_k(0 + \delta q) = I - (i/\hbar)\delta q^a P_a^{(k)}. \tag{2.12}$$

Then, by means of the general formula (cf. Ref. 2)

$$P_a^{(k)}|q\rangle = i\hbar[X_a(q) + ir_a^{(k)}(q)]|q\rangle, \tag{2.13}$$

one gets the *extended Lie algebra* of G_1, as follows[11]:

$$[P_1^{(k)},P_2^{(k)}] = 0, \quad [P_1^{(k)},P_3^{(k)}] = i\hbar P_2^{(k)}, \quad [P_3^{(k)},P_2^{(k)}] = i\hbar^2 k. \tag{2.14}$$

We next perform the *complete quantization* of the Galilei group G_1 defining the following spectral integrals (cf. Ref. 1):

$$Q^a = \mu_0 \int\int\int dq^1\, dq^2\, dq^3 |q^1,q^2,q^3\rangle q^a \langle q^1,q^2,q^3|, \quad a = 1,2,3. \tag{2.15}$$

Obviously, one has $[Q^a,Q^b] = 0$, $Q^{a+} = Q^a$, and $Q^a|q^1,q^2,q^3\rangle = q^a|q^1,q^2,q^3\rangle$. Furthermore, it can be shown that these operators transform "covariantly" under the representative ray operators of G_1; i.e., one has

$$U_k^+(q)Q^1U_k(q) = q^1 + Q^1, \quad U_k^+(q)Q^2U_k(q) = q^2 + Q^2 - q^3Q^1, \quad U_k^+(q)Q^3U_k(q) = q^3 + Q^3, \tag{2.16}$$

[cf. Eqs. (2.2)] even if one considers a ray extension of the regular representation. (This fact is a consequence of the general functional properties of two-cocycles and is valid, in general, for all Lie groups.) Hence it follows that the *non-Abelian canonical commutators* are given by the general formula (Ref. 1)

$$[Q^a,P_b^{(k)}] = i\hbar R^a{}_b(Q) = i\hbar \lim_{q\to 0} \partial_b g^a(q;Q), \tag{2.17}$$

which, in the present instance, yields

$$\begin{aligned} &[Q^1,P_1^{(k)}] = i\hbar, \quad [Q^2,P_1^{(k)}] = 0, \quad [Q^3,P_1^{(k)}] = 0,\\ &[Q^1,P_2^{(k)}] = 0, \quad [Q^2,P_2^{(k)}] = i\hbar, \quad [Q^3,P_2^{(k)}] = 0,\\ &[Q^1,P_3^{(k)}] = 0, \quad [Q^2,P_3^{(k)}] = -i\hbar Q^1, \quad [Q^3,P_3^{(k)}] = i\hbar. \end{aligned} \tag{2.18}$$

The set of commutators presented in Eqs. (2.14) and (2.18) define the *extended kinematic algebra* of G_1.

Let us derive some interesting results from this kinematic algebra. First, it follows that $\mathscr{H}(G_1)$ contains *three linearly independent operators that are invariant operators* of G_1 (we omit the details of this calculation),

$$S_k = P_1^{(k)} + (1/2\hbar k)[P_2^{(k)}]^2, \tag{2.19}$$

$$A_k = P_2^{(k)} + \hbar k Q^3, \tag{2.20}$$

$$B_k = Q^1P_2^{(k)} + P_3^{(k)} - \hbar k Q^2. \tag{2.21}$$

The reader can easily check these operators against their fundamental properties: $[S_k,P_a^{(k)}] = 0$, $[A_k,P_a^{(k)}] = 0$, $[B_k,P_a^{(k)}] = 0$ $(a = 1,2,3)$. Furthermore, every other invariant operator of G_1 in $\mathscr{H}(G_1)$ must be a function of S_k, A_k, and B_k, exclusively.

Of course, the most interesting invariant operator is S_k, because it contains the term $P_1^{(k)}$, which is the generator of time translation symmetry. Moreover, it must be borne in mind that S_k is the only invariant operator pertaining to the extended Lie algebra (2.14); namely, S_k is the only Casimir operator of G_1, in the usual sense of this notion. The invariant operators A_k and B_k are *new*, since they stem from the *quantization* of G_1 through Eq. (2.15). We refer to S_k as the *Schrödinger operator* of the system (Ref. 2), and we write

$$H_k = (2\hbar k)^{-1}(P_2^{(k)})^2, \tag{2.22}$$

to denote the familiar Hamiltonian. The Schrödinger operator is endowed with the fundamental property

$$U_k^+(q)S_kU_k(q) = S_k, \tag{2.23}$$

which entails the Galilean invariance of the underlying quantum model. (Let us observe that neither the Hamiltonian H_k nor $P_1^{(k)}$ are Galilean invariants.) It is clear that, since $P_2^{(k)}$ is the generator of space translation symmetry, one interprets k as $\hbar k = m$, where m is the *mass* of the particle.[11]

Finally, let us remark that for the other two invariant operators one has

$$[A_k,B_k] = i\hbar m. \tag{2.24}$$

III. SUPERSELECTION RULES

In order to arrive at the physical interpretation of the quantum kinematic model of G_1, we shall adopt the same postulate already used in Ref. 2; namely we consider $\mathscr{H}(G_1)$ as an incoherent Hilbert space, and introduce the following superselection rule: *The allowable physical states of the sys-*

tem correspond to simultaneous eigenkets of the Schrödinger operator S_k and one of the other invariant operators (A_k or B_k) previously found. The motivation for these superselection rules is the same as in our previous work (Ref. 2). In this paper we shall not indulge on the physical meaning of A_k and B_k. (We defer this interesting matter to another opportunity.) Hence, for the sake of concreteness, in the sequel we will concentrate our attention on the compatible set $\{S_k, A_k\}$ exclusively. In other words, in the present paper, the boost operator Q^3 will be used only in *reducing* the regular ray representation of G_1 via the postulated superselection rule connected with A_k [cf. Eq. (2.20)] and S_k.

One solves rather easily the eigenvalue problems of S_k and A_k within the Hilbert space $\mathscr{H}(G_1)$. Since $\{P_1^{(k)}, P_2^{(k)}, Q^3\}$ is a complete set of commuting self-adjoint operators, after some simple manipulations, one calculates the general form of the simultaneous eigenkets of S_k and A_k. Thus one gets

$$|E,p;\psi\rangle = \int_{-\infty}^{\infty} dp_2(p_2) \left| p_1 = E - \frac{p_2^2}{2m}, p_2, q^3 = \frac{p-p_2}{m} \right\rangle$$
$$= \int\int \frac{dq^1\, dq^2}{(2\pi\hbar)^{3/2}} \exp\left[\frac{i}{\hbar}\left(Eq^1 + \frac{1}{2}pq^2\right)\right] \int_{-\infty}^{\infty} dp_2\, \psi(p_2) \exp\left[-\frac{i}{\hbar}\left(\frac{p_2^2 q^1}{2m} - \frac{1}{2}p_2 q^2\right)\right] \left| q^1, q^2, q^3 = \frac{p-p_2}{m} \right\rangle, \tag{3.1}$$

where we have adjusted $\mu_0 = (2\pi h)^{-1}$, and where the kets $|p_1,p_2,q^3\rangle$ define a new basis in $\mathscr{H}(G_1)$. These are plane-wave packets that belong in the Hilbert subspace $\mathscr{H}_{Ep} \subset \mathscr{H}(G_1)$, and are such that

$$S_k |E,p;\psi\rangle = E\,|E,p;\psi\rangle, \tag{3.2}$$
$$A_k |E,p;\psi\rangle = p|E,p;\psi\rangle. \tag{3.3}$$

Each physical Hilbert subspace $\mathscr{H}_{Ep}$ has just one degree of freedom; i.e., the $\psi(p_2)$'s are arbitrary complex functions, provided they are consistent with the following scalar product:

$$\langle E',p';\psi|E,p;\phi\rangle = \delta(E'-E)\delta(p'-p)(\psi|\phi), \tag{3.4}$$

where, clearly,

$$(\psi|\phi) = \int_{-\infty}^{\infty} dp_2\, \psi^*(p_2)\phi(p_2). \tag{3.5}$$

IV. COMPLEMENTARY RAY REPRESENTATIONS OF G_1 IN NEWTONIAN SPACE-TIME

It is our aim to obtain the ordinary wave mechanical description of the system. Therefore, we next introduce a space-time group-theoretic formalism that is well adapted to the Galilean relativity theory stated in Eqs. (2.1). Following the same idea used in Ref. 2, we introduce new kets $|t,x\rangle \in \mathscr{H}(G_1)$ (rigged), so that, by construction, one has

$$U_k(q^1,q^2,q^3)|t,x\rangle = e^{i\rho_k(q^1,q^2,q^3;t,x)}|t+q^1, x+q^2-q^3t\rangle, \tag{4.1}$$

where ρ_k is a real-valued *phase function*. If $|\psi\rangle \in \mathscr{H}(G_1)$, one sets $\psi(t,x) = \langle t,x|\psi\rangle$, and thus one gets

$$\langle t,x|U_k^+(q^1,q^2,q^3)|\psi\rangle = e^{-i\rho_k(q^1,q^2,q^3;t,x)}\psi(t+q^1, x+q^1-q^3t), \tag{4.2}$$

which entails a *complementary ray representation* of the group G_1 of space-time symmetry transformations stated in Eqs. (2.1). (A general method for the construction of an admissible phase function ρ_k will be discussed in a forthcoming paper.[12]) It is immediate that ρ_k must be endowed with the following fundamental properties:

$$\rho_k(0,0,0;t,x) = 0, \tag{4.3}$$
$$\rho_k(q^1,q^2,q^3;t,x) + \rho_k(q'^1,q'^2,q'^3;t+q^1,x+q^2-q^3t) - \rho_k(q'^1+q^1,q'^2+q^2-q'^3q^1,q'^3+q^3;t,x) = (m/2\hbar)[q'^2q^3 - q'^3(q^2+q^1q^3)] \tag{4.4}$$

in order to be consistent with the group property of the $U_k(q)$'s. (These two properties are enough for calculating an allowable phase function ρ_k associated with the two-cocycle ϕ_k.) We will use the following phase function for G_1:

$$\rho_k(q^1,q^2,q^3;t;x) = -(m/2\hbar)(2x+q^2-q^3t)q^3, \tag{4.5}$$

which the reader can check against Eqs. (4.3) and (4.4). A set of associated *phase generators* of the complementary ray representation[12] is defined as follows (cf. also Ref. 2):

$$\rho_a^{(k)}(t,x) = \lim_{q\to 0} \partial_a \rho_k(q^1,q^2,q^3;t,x), \tag{4.6}$$

with $a = 1,2,3$. So, in the present case, we get

$$\rho_1^{(k)}(t,x) = 0,\quad \rho_2^{(k)}(t,x) = 0,\quad \rho_3^{(k)}(t,x) = -(m/\hbar)x. \tag{4.7}$$

Now, it can be shown that a necessary and sufficient condition for the kets $|t,x\rangle$ to be endowed with the required property (4.1) is that they have the following general form[12] (cf. Ref. 2):

$$|t,x\rangle = \int\int\int \frac{dq^1\, dq^2\, dq^3}{2\pi\hbar}\, \xi_k^*(t+\bar{q}^1, x+\bar{q}^2-\bar{q}^3t) \times e^{i\rho_k(\bar{q}^1,\bar{q}^2,\bar{q}^3;t,x)}|q^1,q^2,q^3\rangle, \tag{4.8}$$

where the $\bar{q}$'s denote the group-inversion formulas for the parameters; namely,

$$\bar{q}^1 = -q^1,\quad \bar{q}^2 = -q^2 - q^1q^3,\quad \bar{q}^3 = -q^3. \tag{4.9}$$

(The reader can prove the sufficiency of condition (4.8) quite directly.) Thus far, the *generating wave function* $\xi_k(t,x)$ is completely arbitrary and remains at our disposal.

We observe that an infinitesimal transformation of the $|t,x\rangle$'s yields immediately the following space-time realiza-

tions of the generators:

$$P_1^{(k)}|t,x\rangle = i\hbar\,\partial_t|t,x\rangle, \quad P_2^{(k)}|t,x\rangle = i\hbar\,\partial_x|t,x\rangle,$$
$$P_3^{(k)}|t,x\rangle = (-i\hbar t\,\partial_x + mx)|t,x\rangle, \tag{4.10}$$

for whatever function $\xi_k(t,x)$ one may use in (4.8). Therefore the Schrödinger operator acting on these kets becomes the familiar operator (cf. also Ref. 2)

$$S_k|t,x\rangle = (i\hbar\,\partial_t - (\hbar^2/2m)\partial_x^2)|t,x\rangle. \tag{4.11}$$

Of course, the generating wave function ξ_k (t,x) must be identified on physical ground. In this sense, it seems rather natural to demand that the space-time kets $|t,x\rangle$ themselves correspond to physically realizable states (as we have done in Ref. 2). Thus, according to Eqs. (3.1) an (4.8), we look for kets that have simultaneously the following two general forms:

$$|t,x;E,p\rangle = \int\int\int \frac{dq^1\,dq^2\,dq^3}{2\pi\hbar} \times \xi^*_{Ep}[t-q^1,(x-q^2)+q^3(t-q^1)] \times \exp\left[\frac{im}{2\hbar}[2x-q^2+q^3(t-q^1)]q^3\right]|q^1,q^2,q^3\rangle$$

$$= \int\int \frac{dq^1\,dq^2}{(2\pi\hbar)^{3/2}} \exp\left[\frac{i}{\hbar}\left(Eq^1+\frac{pq^2}{2}\right)\right] \times \int_{-\infty}^{+\infty} dp_2\,\psi_{Ep_2}(t,x) \times \exp\left[-\frac{i}{\hbar}\left(\frac{p_2^2q^1}{2m}-\frac{p_2q^2}{2}\right)\right] \times \left|q^1,q^2,q^3=\frac{p-p_2}{m}\right\rangle; \tag{4.12}$$

i.e., we demand the fulfillment of Eq. (4.1), and, also,

$$S_k|t,x;E,p\rangle = E|t,x;E,p\rangle, \tag{4.13}$$
$$A_k|t,x;E,p\rangle = p|t,x;E,p\rangle. \tag{4.14}$$

The solution of this problem follows.

It is clear that ψ_{Ep_2} has to satisfy

$$(i\hbar\,\partial_t - (\hbar^2/2m)\partial_x^2)\psi_{Ep_2}(t,x) = E\psi_{Ep_2}(t,x), \tag{4.15}$$

whose elementary solutions are

$$\psi_{Ep_2}(t,x) = \exp\left[\frac{i}{\hbar}\left(E-\frac{p_2^2}{2m}\right)t\right]\left[\alpha(p_2)\exp\left[-\frac{i}{\hbar}p_2x\right] + \beta(p_2)\exp\left[\frac{i}{\hbar}p_2x\right]\right]. \tag{4.16}$$

Substituting these plane waves into Eq. (4.12), and considering the projection on the q basis, i.e., $\langle q^1,q^2,q^3|t,x;E,p\rangle|$, after some manipulations, one concludes that α must be a constant and β must be zero. Thus one gets the answer

$$\xi^*_{Ep}(t,x) = \frac{m\alpha}{\sqrt{2\pi\hbar}}\exp\left[\frac{i}{\hbar}\left[\left(E-\frac{p^2}{2m}\right)t+px\right]\right], \tag{4.17}$$

where α is a normalization constant.

So we obtain the physical space-time kets $|t,x;E,p\rangle$, which carry the complementary ray representations of the Galilei group G_1, with $\mathscr{H}_{Ep}\subset\mathscr{H}(G_1)$, in the following two equivalent forms:

$$|t,x;E,p\rangle = \frac{m\alpha}{2\pi\hbar}\int\int\int dq^1\,dq^2\,dq^3 \times\exp\left[\frac{-i}{\hbar}\left(E-\frac{p^2}{2m}\right)(t-q^1)\right] \times\exp\left[\frac{-i}{\hbar}p[(x-q^2)+q^3(t-q^1)]\right] \times\exp\left[\frac{im}{2\hbar}[2x-q^2+q^3(t-q^1)]q^3\right]|q^1,q^2,q^3\rangle$$
$$= \alpha\int_{-\infty}^{\infty} dp_2\exp\left[\frac{i}{\hbar}\left[\left(E-\frac{p_2^2}{2m}\right)t+p_2x\right]\right] \times\left|p_1=E-\frac{p_2^2}{2m},p_2,q^3=\frac{p-p_2}{m}\right\rangle. \tag{4.18}$$

In this fashion, for any given physical state $|E,p;\psi\rangle\in\mathscr{H}_{Ep}$ [cf. Eq. (3.1)], one has

$$\langle t,x;E',p'|E,p;\psi\rangle = \delta(E'-E)\delta(p'-p)\psi_E(t,x), \tag{4.19}$$

where the wave function $\psi_E(t,x)$ is given by

$$\psi_E(t,x) = \alpha m\exp\left(\frac{i}{\hbar}Et\right)\int_{-\infty}^{+\infty} dp_2\psi(p_2) \times\exp\left[\frac{i}{\hbar}\left(\frac{p_2^2}{2m}t-p_2x\right)\right], \tag{4.20}$$

as it should be. Therefore the time-dependent Schrödinger equation reads

$$\langle t,x;E',p'|S_k|E,p;\psi\rangle = \delta(E'-E)\delta(p'-p)\,(-\hbar\,\partial_t - (\hbar^2/2m)\partial_x^2)\psi_E(t,x)$$
$$= \delta(E'-E)\delta(p'-p)E\psi_E(t,x). \tag{4.21}$$

The "regularization" of the theory is automatic because of the delta functions, which come from the superselection rules.

Finally, let us discuss briefly the space-time kernel of a Newtonian free particle in the context of quantum kinematics. Setting $t'=t$, from Eq. (4.18) one gets

$$\langle t,x';E',p'|t,x;E,p\rangle = (2\pi\hbar m\alpha^2)\delta(E'-E)\delta(p'-p)\delta(x'-x). \tag{4.22}$$

Hence we normalize $\alpha=(2\pi\hbar m)^{-1/2}$. Obviously, since

$$\langle t',x';E',p'|S_k|t,x;E,p\rangle = (i\hbar\,\partial_t - (\hbar^2/2m)\partial_x^2)\langle t',x';E',p'|t,x;E,p\rangle$$
$$= E\langle t',x';E',p'|t,x;E,p\rangle, \tag{4.23}$$

the space-time Feynman propagator of the system is given simply by

$$\langle t',x';E',p'|t,x;E,p\rangle = \delta(E'-E)\delta(p'-p)\exp\left[\frac{i}{\hbar}E(t'-t)\right]\int_{-\infty}^{\infty}\frac{dp_2}{2\pi\hbar} \times\exp\left[-\frac{i}{\hbar}\left[\frac{p_2^2}{2m}(t'-t)-p_2(x'-x)\right]\right], \tag{4.24}$$

which yields immediately the familiar Green's function of the time-dependent Schrödinger equation of a free particle in one-dimensional space[13] (suitably modulated by the delta functions, as a consequence of the superselection rules).

As a matter of principle, it is interesting to observe that, according to Eq. (4.18), the kernel (4.24) can be calculated also by means of a triple integral that is manifestly a *Hurwitz invariant integral over the group manifold.* (In Ref. 2 we have discussed a tentative interpretation of this fundamental group-theoretic result.)

V. CONCLUDING REMARKS

In this paper we have discussed the main features of the group-quantization process of a scalar Newtonian free particle through its Galilean symmetries. To be sure, the fact that one is able to *deduce* the correct form of the Schrödinger equation from the assumed Newtonian symmetries of the system is not as striking in the case of a free particle as it was in the case of the simple harmonic oscillator. Nevertheless, it should be understood that the emphasis of this work has been on the kinematic *method of group-theoretic quantization.* Indeed, we would like to stress some facts concerning this issue.

(1) The Newtonian symmetry group of the harmonic oscillator and that of a free particle (i.e., E_2 and G_1, respectively) are completely different Lie groups.

(2) In both instances, one obtains the Schrödinger operator, the time-dependent Schrödinger equation, and the space-time kernel of the system, following exactly the same group-theoretic steps.

(3) The time-dependent Schrödinger equation corresponds precisely to the space-time realization of one of the fundamental invariants of the Newtonian point symmetry group of the system.

(4) The Feynman propagator is given by the space-time transition amplitudes $\langle t',x'|t,x\rangle$, where the kets $|t,x\rangle$ satisfy the superselection rule and carry an irreducible complementary space-time representation of the relativity group of the system. (The quantum kinematic formalism is intrinsically relativistic indeed.)

(5) The treatment of the problem of "quantization" in these examples is purely group theoretic, since only the assumed symmetries of the system participate in the construction of the desired quantum model.

(6) The "principle of correspondence" is not an indispensable conceptual device for the correct quantization (neither of a free particle nor of the harmonic oscillator). Instead, one uses a "symmetry principle," and *calculates* the correct form of the Hamiltonian.

How general is the meaning of these results, and how far can one push into physics the idea that quantum theory is essentially a *mechanics of symmetries*[14] remains to be seen. Needless to say, such an achievement would be of great interest in the realm of high energy physics.

ACKNOWLEDGMENTS

This paper was supported in part by DIUC Grant No. 50/87 and by FONDECYT Grant No. 670/87.

[1] J. Krause, J. Phys. A **18**, 1309 (1985).

[2] J. Krause, J. Math. Phys. **27**, 2922 (1986).

[3] Work is in progress concerning SU(2) quantum kinematics; for a previous report, cf. J. Krause, "SO(3) commutators for angular momentum and rotation observables," in *Group Theoretical Methods in Physics*, edited by G. Denardo *et al.* (Springer, Berlin, 1984).

[4] The literature on unitary ray representations of the Galilei group is rather extensive; cf., for instance, V. Bargmann, Ann. Math. **59**, 1 (1954) (which is the "classic" on the issue); J. Voisin, J. Math. Phys. **6**, 1519 (1965); H. J. Bernstein, J. Math. Phys. **8**, 406 (1967). The subject has been studied mainly in connection with the Galilean invariance of the Schrödinger equation of a free particle; see M. Hammermesh, Ann. Phys. (NY) **9**, 518 (1960); also A. G. Nikitin and V. I. Fushchich, Teor. Mat. Fiz. **44**, 34 (1980), and references therein. An intersting survey of this subject can be found in J. M. Lévy-Leblond, "Galilei group and Galilean invariance," in *Group Theory and Its Applications*, edited by E. M. Loebl (Academic, New York, 1971), Vol. II.

[5] J. M. Lévy-Leblond, Riv. Nuovo Cimento **4**, 99 (1974) presents a nice discussion of Newtonian relativity.

[6] G. Racah, Ergeb. Exakten Naturwiss. **37**, 28 (1965).

[7] By the way, let us observe that Eq. (2.4) is also the Lie algebra of the familiar Heisenberg–Weyl group, whose regular and ray representation theory have been worked out long ago. (The author acknowledges the referee for this remark.) Compare M. Stone, Proc. Natl. Acad. Sci. USA **16**, 172 (1930); and J. von Newumann, Math. Ann. **104**, 570 (1931). A number of subtle mathematical aspects of the Heisenberg–Weyl group is considered by P. Cartier, *Algebraic Groups and Discontinuous Subgroups*, Proceedings of Symposia in Pure Mathematics, Vol. 9 (Am. Math. Soc., Providence, RI, 1966) pp. 361–383. We mention this coincidence here, although in the present paper we do not elaborate on the relation between G_1 and the Heisenberg–Weyl group.

[8] Compare, for instance, M. A. Naimark and A. I. Stern, *Theory of Group Representations* (Springer, New York, 1982).

[9] For a simple description of the techniques of two-cocycle calculus, cf. J. Krause, J. Math. Phys. **28**, 2268 (1987).

[10] J. C. Houard, J. Math. Phys. **18**, 502 (1977); also, J. M. Lévy-Leblond, in Refs. 4 and 5, and H. J. Bernstein, in Ref. 4.

[11] Compare, V. Bargmann, in Ref. 4.

[12] Compare, J. Krause, *Complementary Spacetime Representations for Non-Abelian Quantum Kinematics*, Proceedings of the 5th Chilean Symposium on Physics, Santiago, 1986 (SOCHIFI, Santiago, 1986), pp. 671–680.

[13] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).

[14] There are several contributions following this idea in the recent literature, although from quite different perspectives. Compare R. T. Prosser, J. Math. Phys. **24**, 548 (1983), and references therein. For a very enlightening discussion of the epistemological significance of group theory in quantum mechanics, see J. M. Lévy-Leblond, in Ref. 5.

On the generalized multiplication and addition of complex numbers

Yoel Tikochinsky[a)]
Mullard Radio Astronomy Observatory, Cavendish Laboratory, Madingley Road, Cambridge CB3 OHE, United Kingdom

(Received 12 April 1987; accepted for publication 5 August 1987)

Let $x \circ y$ and $x \oplus y$ denote complex valued generalized multiplication and addition of complex numbers. If (a) multiplication is associative, (b) addition is associative and commutative, (c) there exists a complex number e satisfying $x \oplus e = x$ for all x, (d) multiplication is distributive with respect to addition, and (e) both operations are analytic functions in the two variables, then there exists a mapping $x' \to x$ such that $(x \circ y)' = x' \cdot y'$ and $(x \oplus y)' = x' + y'$. This result is used in the derivation of the multiplication and addition laws for quantum amplitudes.

I. INTRODUCTION

The purpose of this paper is to give a physicist's approach to the solution of the following problem. We seek two complex valued functions $f(x,y)$ and $g(x,y)$ of two complex variables x and y such that (a) $f(x,y)$ is associative, (b) $g(x,y)$ is associative and commutative, (c) there exists a complex number e satisfying $g(x,e) = x$ for all x, and (d) $f(x,y)$ is distributive with respect to $g(x,y)$. The two functions may be thought of as two operations, namely, $f(x,y) = x \circ y$ (generalized multiplication) and $g(x,y) = x \oplus y$ (generalized addition). It is evident that regular multiplication and addition of complex numbers, with $e = 0$, satisfy all the above. What is, perhaps, less evident is that conditions (a)–(d) together with the assumption of analyticity in both variables x and y, (essentially) determine the functions $f(x,y)$ and $g(x,y)$. We shall show (at least locally) that these functions are equivalent (isomorphic) to regular multiplication and addition. That is, there exists a one-to-one transformation $x' = H(x)$ such that both equations

$$H[f(x,y)] = H(x) \cdot H(y) \quad \text{or} \quad (x \circ y)' = x' \cdot y' \tag{1a}$$

and

$$H[g(x,y)] = H(x) + H(y) \quad \text{or} \quad (x \oplus y)' = x' + y' \tag{1b}$$

are satisfied. Although analyticity is sufficient to prove the result (1), it is not clear that it is necessary. That a strong condition on the functions f and g is required in order to fulfill (1) is demonstrated by the following example. Let $x = x_R + ix_I$. Then $g(x,y) = x + y$, $e = 0$, and $f(x,y) = x_R y_R + ix_I y_I$ satisfy (a)–(d) but not (1). In the course of proving Eq. (1), we shall assume without further qualifications that all the operations involved are valid, at least locally. Thus, for example, we shall assume that an equation of the form $f(x) = y$ can always be (locally) inverted to find $x = x(y)$.

The motivation for the above theorem arose in the process of deriving[1] Feynman's laws for combining quantum mechanical amplitudes.[2] Indeed, let $f(x,y)$ denote the amplitude for the process $C \to B \to A$ with partial amplitudes $\langle B|C\rangle = y$ and $\langle A|B\rangle = x$ [Fig. 1(a)]. Then, by the associative law for combining processes in series [Fig. 1(b)],

$$f[x, f(y,z)] = f[f(x,y), z]. \tag{2}$$

Similarly, if $g(x,y)$ denotes the amplitude for the process displayed in Fig. 2(a) then, clearly,

$$g(x,y) = g(y,x), \tag{3}$$

and by the associative law for combining processes in parallel [Fig. 2(b)],

$$g[x, g(y,z)] = g[g(x,y), z]. \tag{4}$$

Let e denote the amplitude for a blocked process. Then the amplitude for the process depicted in Fig. 2(a), with the lower branch blocked, must satisfy

$$g(x,e) = x. \tag{5}$$

Consider now the process displayed in Fig. 3. Viewed as two processes in series [Fig. 3(a)], the amplitude is $f[g(x,y), \lambda]$. On the other hand, viewed as a process in parallel [Fig. 3(b)], the amplitude is $g[f(x,\lambda), f(y,\lambda)]$. Hence the distributive law is

$$f[g(x,y), \lambda] = g[f(x,\lambda), f(y,\lambda)]. \tag{6}$$

In terms of the quantum mechanical amplitudes, Eq. (1) now reads as follows: (a) the amplitude for processes occurring in series is the product of the amplitudes for each part; (b) the amplitude for processes occurring in parallel is the sum of the amplitudes for the different alternatives. Of course, by choosing $H(x) \neq x$ one can change the *form* of the laws but their content remains invariant. The situation is analogous to that prevailing in probability theory.[3] One can choose (if one wishes) to work with $\log p_i$ instead of p_i thereby changing the form of the laws for combining probabilities without changing their content.

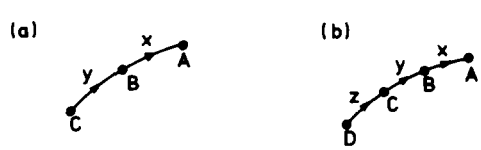


FIG. 1. Processses in series.

[a)] Permanent address: Racah Institute of Physics, The Hebrew University of Jerusalem, Jerusalem 91904, Israel.

 0022-2488/88/020398-02$02.50

FIG. 2. Processes in parallel.

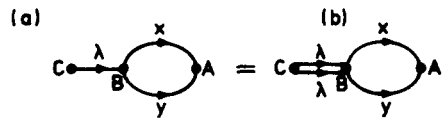


FIG. 3. Combined processes.

II. DERIVATION OF EQ. (1)

Differentiating Eq. (4) with respect to x and y in turn and taking the ratio of the resulting equations, we have

$$\frac{g_1[x,g(y,z)]}{g_2[x,g(y,z)]}=\frac{g_1(x,y)}{g_2(x,y)}g_1(y,z)\,. \tag{7}$$

Now set $y=c=$ const in the equation above, replace $g(c,z)$ in the resulting equation by y and solve $g(c,z)=y$ for z to obtain

$$\frac{g_1(x,y)}{g_2(x,y)}=\frac{g_1(x,c)}{g_2(x,c)}g_1(c,z(y))\,. \tag{8}$$

Let

$$k(x)=\int_e^x\frac{g_1(x,c)}{g_2(x,c)}\,dx \quad\text{and}\quad l(y)=\int^y\frac{dy}{g_1(c,z(y))}\,. \tag{9}$$

Then Eq. (8) reads

$$\frac{(\partial g/\partial x)(x,y)}{(\partial g/\partial y)(x,y)}=\frac{k'(x)}{l'(y)}\,. \tag{10}$$

The vanishing of the Jacobian $\partial[g(x,y),k(x)+l(y)]/\partial(x,y)$ means that $g(x,y)$ and $[k(x)+l(y)]$ are functionally dependent. Thus[4]

$$g(x,y)=h[k(x)+l(y)]\,. \tag{11}$$

We now apply commutativity [Eq. (3)] to Eq. (11) to obtain $k(x)+l(y)=k(y)+l(x)$. Hence

$$k(x)-l(x)=k(y)-l(y)=\text{const}, \tag{12}$$

and since $l(y)$ is defined up to a constant [see Eq. (9)], we can choose the constant in Eq. (12) to be 0. Thus $k(x)=l(x)$ and

$$g(x,y)=h[k(x)+k(y)]\,. \tag{13}$$

Inserting the result (13) into the associative law (4), we have

$$h[k(x)+k(g(y,z))]=h[k(g(x,y))+k(z)]\,. \tag{14}$$

The substitution $x=e$ now leads, by use of Eqs. (9) and (5), to the result

$$k[g(y,z)]=k(y)+k(z)\,, \tag{15}$$

which is of the form (1b) (see the Note added in proof).

Let $\tilde{x}=k(x)$, $\tilde{g}(x,y)=k[g(x,y)]=G(\tilde{x},\tilde{y})$, and $\tilde{f}(x,y)=k[f(x,y)]=F(\tilde{x},\tilde{y})$. In terms of the new notation, Eqs. (15), (2), and (6) read, respectively,

$$G(\tilde{y},\tilde{z})=\tilde{y}+\tilde{z}\,, \tag{16}$$

$$F[\tilde{x},F(\tilde{y},\tilde{z})]=F[F(\tilde{x},\tilde{y}),\tilde{z}]\,, \tag{17}$$

and

$$F(\tilde{x}+\tilde{y},\tilde{\lambda})=F(\tilde{x},\tilde{\lambda})+F(\tilde{y},\tilde{\lambda})\,. \tag{18}$$

We now turn to determine the function $F(\tilde{x},\tilde{y})$. Differentiating Eq. (18) with respect to $\tilde{x}$, we have

$$F_1(\tilde{x}+\tilde{y},\tilde{\lambda})=F_1(\tilde{x},\tilde{\lambda})=m(\tilde{\lambda})\,. \tag{19}$$

Hence, with the aid of the relation $F(0,\tilde{\lambda})=0$, which follows from (18), we secure

$$F(\tilde{x},\tilde{\lambda})=\tilde{x}\cdot m(\tilde{\lambda})\,. \tag{20}$$

Inserting this result into the associative law (17), we find

$$\tilde{x}\cdot m[F(\tilde{y},\tilde{z})]=F(\tilde{x},\tilde{y})\cdot m(\tilde{z})=\tilde{x}\cdot m(\tilde{y})\cdot m(\tilde{z})\,, \tag{21}$$

and hence

$$m[F(\tilde{y},\tilde{z})]=m(\tilde{y})\cdot m(\tilde{z})\,. \tag{22}$$

Equation (22) is precisely Eq. (1a) with $H(x)=m[k(x)]$.

In order to show that Eq. (1b) is also fulfilled, we have to demonstrate that

$$m[G(\tilde{x},\tilde{y})]=m(\tilde{x}+\tilde{y})=m(\tilde{x})+m(\tilde{y})\,. \tag{23}$$

Indeed, by Eq. (22), $F(\tilde{x},\tilde{y})=F(\tilde{y},\tilde{x})$. Hence employing Eqs. (18) and (20),

$$\begin{aligned}F(\tilde{x}+\tilde{y},\tilde{\lambda})&=F(\tilde{\lambda},\tilde{x}+\tilde{y})=\tilde{\lambda}\cdot m(\tilde{x}+\tilde{y})\\&=F(\tilde{\lambda},\tilde{x})+F(\tilde{\lambda},\tilde{y})=\tilde{\lambda}\cdot[m(\tilde{x})+m(\tilde{y})]\,,\end{aligned}$$

which proves (23).

Note added in proof: Condition (c), namely, the requirement that a unit element e with respect to generalized addition exists, is in fact superfluous. Indeed, from Eq. (14) we have $k(g(y,z))-k(z)=k(g(x,y))-k(x)\equiv c(y)$. Hence, by commutativity, $k(g(x,y))=k(x)+c(y)=k(y)+c(x)$. Thus $k(x)-c(x)=k(y)-c(y)=$ const, and by adjusting the lower limit e in Eq. (9) the constant can be made to vanish. We therefore have $c(x)=k(x)$ and Eq. (15) follows. Substituting $z=e$ in Eq. (15) and using Eq. (9) we get $k(g(y,e))=k(y)+k(e)=k(y)$. Thus $g(y,e)=y$ for any y [Eq. (5)], and the existence and uniqueness of the unit element e is established.

ACKNOWLEDGMENT

I am indebted to Dan Shalitin for very helpful discussions.

[1] Y. Tikochinsky (unpublished, 1987).
[2] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics* (Addison–Wesley, Reading, MA, 1965), Vol. III.
[3] R. T. Cox, Am. J. Phys. **14**, 1 (1946).
[4] J. Aczel, *Lectures on Functional Equations and their Applications* (Academic, New York, 1966).

One-dimensional quantum three-wave interaction model with two fermionic fields and one bosonic field

Yu-Peng Wang
Institute of Physics, Chinese Academy of Sciences, P.O. Box 603, Beijing, People's Republic of China

Fu-Cho Pu
Center for Theoretical Physics, CCAST (World Laboratory) and Institute of Physics, Chinese Academy of Sciences, P.O. Box 603, Beijing, People's Republic of China

(Received 23 July 1987; accepted for publication 23 September 1987)

The quantum three-wave interaction (3WI) model with two fermionic fields and one bosonic field is studied on the basis of quantum inverse scattering method (QISM). The Hamiltonian for this model is diagonalized and an infinite number of conserved quantities is obtained.

I. INTRODUCTION

Three-wave interaction (3WI) models play an important role in various physical applications. The classical 3WI models were extensively studied by many authors.[1,2] The quantum cases were considered with the quantum inverse scattering method (QISM) in Ref. 3 and the Bethe ansatz in Ref. 4, respectively. In this paper, we quantize the components of the 3WI model as two fermionic fields and one bosonic field which is relevant to the Lee model in quantum field theory. This problem has been raised in Ref. 4. In the present paper, we show that this model is also completely integrable and the infinite number of conserved quantities can be obtained. Further, we diagonalize the Hamiltonian by finding the common eigenstates of all conserved quantities.

Equations of motion for the 3WI model are expressed as follows:

$$\begin{aligned} &Q_{1t} + C_1 Q_{1x} = -igQ_3^* Q_2, \\ &Q_{2t} + C_2 Q_{2x} = -igQ_3 Q_1, \\ &Q_{3t} + C_3 Q_{3x} = -igQ_1^* Q_2, \end{aligned} \tag{1.1}$$

where the Q_α's are the wave fields, C_α is group velocity of field Q_α ($\alpha = 1,2,3$), and g is a real constant.

The Hamiltonian for this is

$$H = \int \left\{ \sum_{\alpha=1}^{3} C_\alpha Q_\alpha^* \frac{1}{i} \frac{\partial}{\partial x} Q_\alpha + g(Q_1^* Q_3^* Q_2 + Q_2^* Q_3 Q_1) \right\} dx. \tag{1.2}$$

The fields Q_α's satisfy the following commutation and anticommutation relations:

$$\begin{aligned} &[Q_3(x), Q_\alpha(y)] = 0, \quad [Q_3(x), Q_\alpha^*(y)] = \delta(x-y)\delta_{3\alpha}, \\ &\{Q_j(x), Q_i(y)\} = 0, \quad \{Q_j(x), Q_i^*(y)\} = \delta(x-y)\delta_{ij}, \end{aligned} \tag{1.3}$$

where $i,j = 1,2$, Q_1 and Q_2 are the fermionic fields, and Q_3 is the bosonic one.

II. *R* MATRIX AND COMMUTATION RELATIONS

The associated auxiliary linear equation for the 3WI model in the QISM takes the form

$$\frac{\partial}{\partial x}\phi(x,\lambda) = :\mathscr{L}(x,\lambda)\phi(x,\lambda):, \tag{2.1}$$

where

$$\mathscr{L}(x,\lambda) = \begin{bmatrix} i d_1 \lambda & l_3 Q_3 & l_2 Q_2 \\ l_3 Q_3^* & i d_2 \lambda & l_1 Q_1 \\ l_2 Q_2^* & l_1 Q_1^* & i d_3 \lambda \end{bmatrix}$$

is a super operator matrix,[5] in which (ij), $(3\,3)$ are even elements and $(3j)$, $(j\,3)$ are odd elements, λ is, as usual, denoting the spectral parameter, and

$$l_1 = \frac{g}{\sqrt{\beta_{12}\beta_{13}}}, \quad l_2 = \frac{g}{\sqrt{\beta_{12}\beta_{23}}}, \quad l_3 = \frac{ig}{\sqrt{\beta_{13}\beta_{23}}},$$
$$C_3 > C_2 > C_1, \quad \beta_{\alpha\alpha'} = C_{\alpha'} - C_\alpha = d_\alpha - d_{\alpha'}.$$

The lattice form of Eq. (2.1) is

$$\phi_{n+1}(\lambda) = :L_n(\lambda)\phi_n(\lambda):,$$

$$L_n(\lambda) = \begin{bmatrix} 1 + i d_1 \lambda \Delta & l_3 Q_{3n} & l_2 Q_{2n} \\ l_3 Q_{3n}^* & 1 + i d_2 \lambda \Delta & l_1 Q_{1n} \\ l_2 Q_{2n}^* & l_1 Q_{1n}^* & 1 + i d_3 \lambda \Delta \end{bmatrix}, \tag{2.2}$$

where Δ is the lattice spacing, and $Q_{\alpha n} = Q_\alpha(x_n)\Delta$. We impose the boundary conditions for the wave fields,

$$\lim_{|x|\to\infty} Q_\alpha(x) = 0. \tag{2.3}$$

The monodromy matrix is

$$T_N(\lambda) = L_{N-1}(\lambda)L_{N-2}(\lambda)\cdots L_{-N}(\lambda). \tag{2.4}$$

As $N\Delta \to \infty$, $\Delta \to 0$, we define

$$T(\lambda) = \lim V^{-N}(\lambda) T_N(\lambda) V^{-N}(\lambda), \tag{2.5}$$

where

$$V(\lambda) = \begin{bmatrix} 1 + i d_1 \lambda \Delta & 0 & 0 \\ 0 & 1 + i d_2 \lambda \Delta & 0 \\ 0 & 0 & 1 + i d_3 \lambda \Delta \end{bmatrix}.$$

The solution for the Yang–Baxter equation

$$R(\lambda,\mu) L_n(\lambda) \underset{s}{\otimes} L_n(\mu) = L_n(\mu) \underset{s}{\otimes} L_n(\lambda) R(\lambda,\mu) \tag{2.6}$$

is

$$R(\lambda,\mu) = AI + BP_s, \tag{2.7}$$

 0022-2488/88/020400-03$02.50

where $\underset{s}{\otimes}$ denotes the super direct product defined in Ref. 5,

$$A=\frac{ic}{\lambda-\mu+ic},\quad B=\frac{\lambda-\mu}{\lambda-\mu+ic},\quad c=\frac{g^2}{\beta_{12}\beta_{13}\beta_{23}},$$

$$P_s=\begin{bmatrix}1&0&0&0&0&0&0&0&0\\0&0&0&1&0&0&0&0&0\\0&0&0&0&0&0&1&0&0\\0&1&0&0&0&0&0&0&0\\0&0&0&0&1&0&0&0&0\\0&0&0&0&0&0&0&1&0\\0&0&1&0&0&0&0&0&0\\0&0&0&0&0&1&0&0&0\\0&0&0&0&0&0&0&0&-1\end{bmatrix},$$

and I is the 9×9 unit matrix. From Eq. (2.6), it follows that

$$R(\lambda,\mu)T_N(\lambda)\underset{s}{\otimes}T_N(\mu)=T_N(\mu)\underset{s}{\otimes}T_N(\lambda)R(\lambda,\mu). \tag{2.8}$$

By taking the limit $N\to\infty$, $\Delta\to 0$ carefully, we obtain the following equation:

$$R_+(\lambda,\mu)T(\lambda)\underset{s}{\otimes}T(\mu)=T(\mu)\underset{s}{\otimes}T(\lambda)R_-(\lambda,\mu), \tag{2.9}$$

where

$$R_\pm=\begin{bmatrix}\alpha&0&0&0&0&0&0&0&0\\0&\pm\eta&0&\beta&0&0&0&0&0\\0&0&\pm\eta&0&0&0&\beta&0&0\\0&\gamma&0&\mp\eta&0&0&0&0&0\\0&0&0&0&\alpha&0&0&0&0\\0&0&0&0&0&\pm\eta&0&\beta&0\\0&0&\gamma&0&0&0&\mp\eta&0&0\\0&0&0&0&0&\gamma&0&\mp\eta&0\\0&0&0&0&0&0&0&0&\xi\end{bmatrix},$$

$$\alpha=\frac{1}{\lambda-\mu},\quad \beta=\frac{\lambda-\mu+ic}{(\lambda-\mu+i\epsilon)^2},\quad \gamma=\frac{1}{\lambda-\mu+ic},\quad \eta=i\pi\delta(\lambda-\mu),\quad \xi=-(\lambda-\mu-ic)/(\lambda-\mu)(\lambda-\mu+ic).$$

If we write $T(\lambda)$ in the form

$$T(\lambda)=\begin{bmatrix}a_1(\lambda)&b_3(\lambda)&b_2(\lambda)\\c_3(\lambda)&a_2(\lambda)&b_1(\lambda)\\c_2(\lambda)&c_1(\lambda)&a_3(\lambda)\end{bmatrix}, \tag{2.10}$$

we can write some of the commutation relations explicitly from Eq. (2.9),

$$\begin{aligned}&[a_\alpha(\lambda),a_{\alpha'}(\lambda)]=0,\quad [a_1(\lambda),c_1(\lambda)]=0,\quad a_1(\lambda)c_3(\mu)=[(\lambda-\mu-ic)/(\lambda-\mu-i\epsilon)]c_3(\mu)a_1(\lambda),\\&a_1(\lambda)c_2(\mu)=[(\lambda-\mu-ic)/(\lambda-\mu-i\epsilon)]c_2(\mu)a_1(\lambda),\quad [a_3(\lambda),c_3(\mu)]=0,\\&a_3(\lambda)c_2(\mu)=[(\lambda-\mu-ic)/(\lambda-\mu-i\epsilon)]c_2(\mu)a_1(\lambda),\quad a_3(\lambda)c_1(\mu)=[(\lambda-\mu-ic)/(\lambda-\mu+i\epsilon)]c_1(\mu)a_3(\lambda).\end{aligned} \tag{2.11}$$

III. INFINITE NUMBER OF CONSERVED QUANTITIES, EIGENSTATES

From the Neumann expansion for $T(\lambda)$,

$$T(\lambda)=1+\sum_{n=1}^{\infty}\int :V(z_1)V(z_2)\cdots V(z_n):\theta(z_1>\cdots>z_n)dz_1\cdots dz_n,$$

$$V(x)=\begin{bmatrix}0&e^{-i\beta_{12}\lambda x}l_3Q_3(x)&e^{-i\beta_{13}\lambda x}l_2Q_2(x)\\e^{i\beta_{12}\lambda x}l_3Q_3^*(x)&0&e^{-i\beta_{23}\lambda x}l_1Q_1(x)\\e^{i\beta_{13}\lambda x}l_2Q_2^*(x)&e^{i\beta_{23}\lambda x}l_1Q_1^*(x)&0\end{bmatrix}, \tag{3.1}$$

we obtain the asymptotic expansions for $a_1(\lambda), a_3(\lambda)$ when $\lambda \to \infty$,

$$a_1(\lambda) = 1 - \frac{ic}{\lambda} N - \frac{ic}{\lambda^2 \beta_{12}} P_3 - \frac{ic}{\lambda^2 \beta_{13}} P_2 - \frac{ic}{\beta_{12}\beta_{23}\lambda^2} V - \frac{1}{2}\frac{c^2}{\lambda^2} N(N-1) + O\left(\frac{1}{\lambda^3}\right), \tag{3.2}$$

$$a_3(\lambda) = 1 - \frac{ic}{\lambda} M - \frac{ic}{\lambda^2 \beta_{13}} P_2 - \frac{ic}{\lambda^2 \beta_{23}} P_1 - \frac{ic}{\beta_{13}\beta_{23}\lambda^2} V - \frac{1}{2}\frac{c^2}{\lambda^2} M(M-1) + O\left(\frac{1}{\lambda^3}\right),$$

where

$$M = \int (Q_2^* Q_2 + Q_1^* Q_1) dx,$$

$$N = \int (Q_2^* Q_2 + Q_3^* Q_3) dx,$$

$$P_\alpha = -i \int Q_\alpha^* \frac{\partial}{\partial x} Q_\alpha \, dx,$$

$$V = g \int (Q_1^* Q_3^* Q_2 + Q_2^* Q_3 Q_1) dx.$$

We define the vacuum by $Q_\alpha |0\rangle = 0$. We have

$$b_\alpha |0\rangle = 0, \quad a_\alpha(\lambda)|0\rangle = |0\rangle. \tag{3.3}$$

In order to show that the a_α's are generating functionals of infinite conserved quantities, we define

$$\begin{aligned} K_1(\lambda) &= C_1 \beta_{23} a_3(\lambda) + C_3 \beta_{12} a_1(\lambda), \\ K_2(\lambda) &= \beta_{23} a_3(\lambda) + \beta_{12} a_1(\lambda), \end{aligned} \tag{3.4}$$

for which the coefficients of λ^{-2} give the Hamiltonian (1.2) and momentum, respectively. It is easy to show that from Eq. (2.9)

$$[K_1(\lambda), a_\alpha(\mu)] = 0, \quad [K_1(\lambda), K_2(\mu)] = 0. \tag{3.5}$$

Let M, N, H, and $P = P_1 + P_2 + P_3$ be four of the infinite conserved quantities. The state defined by

$$\begin{aligned} |n_1,n_2,n_3,0\rangle &= C_1(\lambda_1) C_1(\lambda_2) \cdots C_1(\lambda_{n_1}) \\ &\quad \times C_2(\mu_1) C_2(\mu_2) \cdots C_2(\mu_{n_2}) \\ &\quad \times C_3(k_1) C_3(k_2) \cdots C_3(k_{n_3}) |0\rangle \\ &= \prod_{i=1}^{n_1} C_1(\lambda_i) \prod_{j=1}^{n_2} C_2(\mu_j) \prod_{l=1}^{n_3} C_3(k_l) |0\rangle, \end{aligned} \tag{3.6}$$

is an eigenstate of the infinite conserved quantities, where the orders of noncommutative operators in products are explicitly indicated. In fact, from (2.10), (2.11), and (3.3) we obtain

$$\begin{aligned} a_1(\lambda)|n_1,n_2,n_3,0\rangle &= \prod_{j=1}^{n_2} \frac{\lambda - \mu_j - ic}{\lambda - \mu_j} \prod_{l=1}^{n_3} \frac{\lambda - k_l - ic}{\lambda - k_l} |n_1,n_2,n_3,0\rangle, \\ a_3(\lambda)|n_1,n_2,n_3,0\rangle &= \prod_{i=1}^{n_1} \frac{\lambda - \lambda_i - ic}{\lambda - \lambda_i} \prod_{j=1}^{n_2} \frac{\lambda - \mu_j - ic}{\lambda - \mu_j} |n_1,n_2,n_3,0\rangle, \\ &\quad \lambda \neq \lambda_i, \mu_j, k_l. \end{aligned} \tag{3.7}$$

We can show that

$$\begin{aligned} N|n_1,n_2,n_3,0\rangle &= (n_2 + n_3)|n_1,n_2,n_3,0\rangle, \\ M|n_1,n_2,n_3,0\rangle &= (n_1 + n_2)|n_1,n_2,n_3,0\rangle, \\ H|n_1,n_2,n_3,0\rangle &= \left(C_1\beta_{23}\sum_{i=1}^{n_1}\lambda_i + C_2\beta_{13}\sum_{j=1}^{n_2}\mu_j + C_3\beta_{12}\sum_{l=1}^{n_3}k_l\right) \\ &\quad \times |n_1,n_2,n_3,0\rangle, \\ P|n_1,n_2,n_3,0\rangle &= \left(\beta_{23}\sum_{i=1}^{n_1}\lambda_i + \beta_{13}\sum_{j=1}^{n_2}\mu_j + \beta_{12}\sum_{l=1}^{n_3}k_l\right)|n_1,n_2,n_3,0\rangle. \end{aligned} \tag{3.8}$$

IV. CONCLUSION

In conclusion, we will discuss degeneracy of the eigenstates and the problem of existence of boundstates for our model. From Eq. (2.11), we can see that the states $C_1(\lambda)C_3(\lambda)|0\rangle$ and $C_2(\lambda)|0\rangle$ have the same eigenvalue for all the conserved quantities. In fact, we can define a ket state

$$\begin{aligned} |n_1,n_2,n_3,n\rangle &= \prod_{i=1}^{n_1} C_1(\lambda_i) \prod_{j=1}^{n_2-n} C_2(\mu_j) \\ &\quad \times \prod_{m=n_2-n+1}^{n_2} C_1(\mu_m) C_3(\mu_m) \\ &\quad \times \prod_{l=1}^{n_3} C_3(k_l)|0\rangle. \end{aligned} \tag{4.1}$$

It is an eigenstate of $a_1(\lambda), a_3(\lambda)$ with eigenvalue independent of n.

From (2.9), we see that

$$C_j(\lambda) C_j(\lambda + ic) = 0, \quad j = 1,2. \tag{4.2}$$

It leads to the conclusion of nonexistence of bound states for Q_1 and Q_2 particles that is in agreement with the results of Ref. 4. The bound state of $n - Q_3$ particles is given by

$$\begin{aligned} &\prod_{l=1}^{n} C_3(k_l)|0\rangle, \\ &k_l = \mu - i(2l - n - 1)c, \quad l = 1,2,\ldots,n. \end{aligned} \tag{4.3}$$

But this bound state has no extra bound energy for Hamiltonian (1.2).

[1] D. J. Kaup, A. Reiman, and A. Bers, Rev. Mod. Phys. **51**, 275 (1979).
[2] V. E. Zakharov and S. V. Manakov, Sov. Phys. JETP **42**, 842 (1976).
[3] P. P. Kulish, Physica D **18**, 360 (1986).
[4] M. Wadati and K. Ohkuma, J. Phys. Soc. Jpn. **53**, 1229, 2889 (1984).
[5] F-C. Pu and B-H. Zhao, Phys. Rev. D **10**, 2253 (1984).

The classification of complete sets of operators commuting with the Dirac operator in Minkowski space-time

N. Kamran
Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, N2L 3G1 Canada

M. Légaré
Centre de Recherches Mathématiques, Université de Montréal, Case Postale 6128, Succ. "A," Montréal, Québec, H3C 3J7 Canada

R. G. McLenaghan
School of Mathematics and Computing, Western Australian Institute of Technology, Perth, Western Australia and Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, N2L 3G1 Canada

P. Winternitz
Centre de Recherches Mathématiques, Université de Montréal, Case Postale 6128, Succ. "A," Montréal, Québec, H3C 3J7 Canada

(Received 22 July 1987; accepted for publication 9 September 1987)

Under the action of the Poincaré group P(1,3) the three-, four-, and five-dimensional vector spaces of formally self-adjoint first-order matrix differential operators commuting among themselves and with the Dirac operator are classified. This gives a complete classification of the maximal subspaces of the vector space of first-order formally self-adjoint matrix differential operators commuting with the Dirac operator, which form an Abelian Lie algebra under the commutator.

I. INTRODUCTION

While a considerable amount of work has gone into the formulation of a complete theory of separation of variables for *scalar* differential operators such as the Hamilton–Jacobi and Laplace–Beltrami operators on Riemannian manifolds,[1–3] the case of *matrix* differential operators has received much less attention.

Of all the matrix differential operators of mathematical physics, it is the Dirac operator that has been studied most.

Indeed, central to the analytic solution of the relativistic hydrogen atom problem stands the fact that in Minkowski space-time the Dirac equation is solvable by separation of variables when expressed in spherical coordinates and a fixed frame of reference (as opposed to a moving one). In this context, the separability property is underlied by the existence of a triple $\mathbf{J} = (J_x, J_y, J_z)$ of first-order matrix differential operators *commuting* with the Dirac operator, namely, the total angular momentum operator well known to Dirac himself.[4]

The operators J_x, J_y, and J_z give the infinitesimal action of the subgroup SO(3,R) of the group P(1,3) of geometrical symmetries of the Dirac equation on four-component spinors. [The operators J_x, J_y, and J_z are the Lie derivatives with respect to the generators L_x, L_y, and L_z of so(3,R) in its standard vector field representation on $\mathbb{R}^4$.] The separable solutions are eigenspinors of J with the separation constants as eigenvalues.

In their analysis of Chandrasekhar's[5] separation of variables procedure for the Dirac equation in the Kerr solution of the Einstein field equation,[1] Carter and McLenaghan[6] recognized that the commuting operators associated with the separability of the Dirac operator need not be Lie derivatives with respect to the generators of the underlying two-dimensional Abelian isometry group.

Carter and McLenaghan[6] indeed obtained an operator which is constructed from the irreducible valence-2 Killing–Yano tensor existing in the Kerr solution and which admits the separable solutions as eigenspinors; their construction has been extended to the case of the neutrino operator in a more general context.[7] (The existence of such an operator was subsequently proved to be a sufficient condition for Dirac separability.[8])

These results enabled Kalnins, Miller, and Williams[9] to analyze the separability in oblate spheroidal coordinates of the Dirac equation in Minkowski space-time by taking the flat limit of the Kerr solution. In contrast with the case of spherical coordinates, separability here requires a moving frame, which turns out to be adapted to the eigenvectors of the Killing–Yano tensor.

We should also mention the work of Cook,[10] who gave a list of the known separable coordinate systems and frames and attempted to formulate a Stäckel theory for the Dirac operator in Minkowski space-time in analogy with the case of the Hamilton–Jacobi and Laplace–Beltrami operators.

Our purpose in this paper is to classify under the action of the Poincaré group P(1,3) the maximal three- and four-dimensional vector spaces of first-order matrix differential operators commuting among themselves and with the Dirac operator. Here, "maximal" means a vector space that is not contained in a higher-dimensional vector space of first-order matrix differential operators commuting among themselves

 0022-2488/88/020403-09$02.50

and with the Dirac operator. This will lay the ground work for the classification of the separable coordinates and frames for the Dirac operator in Minkowski space-time, which will be given in a forthcoming paper.

In Sec. II we recall some known results on first-order matrix differential operators commuting with the Dirac operator in Minkowski space-time and derive necessary and sufficient conditions for a set of such operators to span a three- or four-dimensional Abelian Lie algebra. In Sec. III we present our method for classifying these Lie algebras of matrix differential operators under the action of the Poincaré group P(1,3) and give an exhaustive list of the equivalence classes with a representative for each class. In the latter step, we use the results of Ref. 11, where the closed continuous subgroups of the Poincaré group P(1,3) are classified in conjugacy classes.

II. FIRST-ORDER OPERATORS COMMUTING WITH THE DIRAC OPERATOR

On four-dimensional Minkowski space $[M_4, g := \mathrm{diag}(1,-1,-1,-1)]$ we consider the Dirac operator

$$\mathbf{H}_D := i\gamma^\alpha \nabla_\alpha - mI_4, \tag{2.1}$$

where ∇_α denotes the covariant differentiation operator on four-spinors and $\{\gamma^\alpha, \alpha = 0,1,2,3\}$ is a set of Dirac matrices associated to the Minkowski metric g by the anticommutation relations $\gamma^\alpha\gamma^\beta + \gamma^\beta\gamma^\alpha = 2g^{\alpha\beta}I_4$.

Let $\mathscr{D}$ be the vector space of formally self-adjoint first-order differential operators on the $C^\infty(M_4)$-module of C^∞ sections of the bundle associated via the $D(\frac{1}{2},0)\oplus D(0,\frac{1}{2})$ representation of SL(2,C) to the spin bundle $\mathscr{S}$ of (M_4,g) and let $\mathscr{A}$ be the subspace of $\mathscr{D}$ consisting of first-order operators commuting with the Dirac operator.

The following result of McLenaghan and Spindel[12] describes the elements of $\mathscr{A}$.

Proposition 1: The most general formally self-adjoint first-order operator $\mathbf{K}$ commuting with the Dirac operator $\mathbf{H}_D$ is given by

$$\mathbf{K} = [i(k^\alpha I_4 + c\gamma^\alpha) + k^{\alpha\beta}\gamma_5\gamma_\beta + k^{\alpha\beta\gamma}\gamma_{[\beta}\gamma_{\gamma]}]\nabla_\alpha + kI_4 - \tfrac{3}{4}{}^*k^\alpha_{;\alpha}\gamma_5 + \tfrac{1}{3}{}^*k^{\alpha\beta}{}_{;\alpha}\gamma_\beta - \tfrac{i}{4}k_{\beta;\gamma}\gamma^{[\beta}\gamma^{\gamma]}, \tag{2.2}$$

where

$$k_{\alpha\beta} = k_{[\alpha\beta]}, \quad k_{\alpha\beta\gamma} = k_{[\alpha\beta\gamma]}, \tag{2.3}$$

$$k_{(\alpha;\beta)} = k_{\alpha(\beta;\gamma)} = k_{\alpha\beta(\gamma;\delta)} = c_{,\alpha} = k_{,\alpha} = 0, \tag{2.4}$$

$${}^*k^{\alpha\beta} = \tfrac{1}{2}\epsilon^{\alpha\beta\gamma\delta}k_{\gamma\delta}, \quad {}^*k^\alpha = \tfrac{1}{6}\epsilon^{\alpha\beta\gamma\delta}k_{\beta\gamma\delta}, \tag{2.5}$$

where $\epsilon^{\alpha\beta\gamma\delta}$ denotes the volume element on M_4 and $\gamma_5 := \frac{1}{24}\epsilon_{\alpha\beta\gamma\delta}\gamma^\alpha\gamma^\beta\gamma^\gamma\gamma^\delta$.

We remark that (k^α) is a Killing vector, $(k_{\alpha\beta})$ is a valence-2 Killing–Yano tensor (or a Penrose–Floyd tensor), and $(k_{\alpha\beta\gamma})$ is a valence-3 Killing–Yano tensor.[13] We thus have the following expressions in Cartesian coordinates:

$$k_\alpha = T_\alpha + R_{\alpha\beta}x^\beta, \quad R_{\alpha\beta} = R_{[\alpha\beta]}, \tag{2.6}$$

$$k_{\alpha\beta} = C_{\alpha\beta} + \epsilon_{\alpha\beta\gamma\delta}D^\gamma x^\delta, \quad C_{\alpha\beta} = C_{[\alpha\beta]}, \tag{2.7}$$

$$k_{\alpha\beta\gamma} = \epsilon_{\alpha\beta\gamma\delta}(A^\delta + Bx^\delta). \tag{2.8}$$

We observe that $\mathscr{A}$ has the following direct sum decomposition:

$$\mathscr{A} = \mathscr{P}\oplus\mathscr{Q}\oplus\mathscr{J}, \tag{2.9}$$

where we have introduced the one-dimensional subspaces $\mathscr{Q} := \mathrm{span}\{\mathbf{H}_D\}$ and $\mathscr{J} := \mathrm{span}\{I_4\}$ of $\mathscr{A}$. The elements of $\mathscr{P}$ will be denoted by $\mathbf{K}(k,\Pi,\Lambda)$, where we let

$$k := k^\alpha\frac{\partial}{\partial x^\alpha}, \quad \Pi := \frac{1}{2}k_{\alpha\beta}\,dx^\alpha\wedge dx^\beta, \quad \Lambda = \tfrac{1}{6}k_{\alpha\beta\gamma}\,dx^\alpha\wedge dx^\beta\wedge dx^\gamma. \tag{2.10}$$

Our purpose is to classify under the action of the Poincaré group P(1,3) the three- and four-dimensional subspaces of $\mathscr{P}$ that form an Abelian Lie algebra under the commutator and are maximal, i.e., not contained in a higher-dimensional subspace of $\mathscr{P}$ that forms an Abelian Lie algebra under the commutator.

The condition for two operators $\mathbf{K}^{(1)}$ and $\mathbf{K}^{(2)}$ in $\mathscr{P}$ to commute is given by the following result, which immediately follows from the discussion of Sec. 5 in Ref. 12.

Proposition 2: The commutator of two first-order operators $\mathbf{K}^{(1)}$ and $\mathbf{K}^{(2)}$ in $\mathscr{P}$ vanishes if and only if

$$\Pi^{(1)} = \lambda\Pi^{(2)}, \quad \Lambda^{(1)} = \lambda\Lambda^{(2)}, \tag{2.11}$$

$$[k^{(1)},k^{(2)}] = 0, \tag{2.12}$$

$$\mathscr{L}_{k^{(1)}}\Pi^{(2)} - \mathscr{L}_{k^{(2)}}\Pi^{(1)} = 0, \quad \mathscr{L}_{k^{(1)}}\Lambda^{(2)} - \mathscr{L}_{k^{(2)}}\Lambda^{(1)} = 0, \tag{2.13}$$

where λ is a constant.

From Proposition 2 and the expressions (2.6)–(2.8), we immediately obtain the following corollaries.

Corollary 1: Every three-dimensional (respectively, four-dimensional) subspace of $\mathscr{P}$ that forms an Abelian Lie algebra under the commutator admits a basis of the form $(\mathbf{K}^{(a)}(k^{(a)},0,0),\mathbf{K}^{(3)}(k^{(3)},\Pi,\Lambda))$, where $a = 1,2$ [respectively, $(\mathbf{K}^{(a)}(k^{(a)},0,0),\mathbf{K}^{(4)}(k^{(4)},\Pi,\Lambda))$, where $a = 1,2,3,$].

Corollary 2: Necessary and sufficient conditions for the basis elements $\mathbf{K}^{(a)}$, where $a = 1,2,3$ (respectively, $a = 1,2,3,4$), to commute are

$$T^{(a)\alpha}R^{(b)\gamma}{}_\alpha - T^{(b)\alpha}R^{(a)\gamma}{}_\alpha = 0, \tag{2.14}$$

$$R^{(a)\alpha}{}_\beta R^{(b)\gamma}{}_\alpha - R^{(b)\alpha}{}_\beta R^{(a)\gamma}{}_\alpha = 0, \tag{2.15}$$

where $a,b = 1,2,3$ (respectively, $a,b = 1,2,3,4$), and

$$\epsilon_{\alpha\beta\gamma\rho}D^\gamma T^{(a)\rho} + C_{\rho\beta}R^{(a)\rho}{}_\alpha + C_{\alpha\rho}R^{(a)\rho}{}_\beta = 0, \tag{2.16}$$

$$D^\gamma(\epsilon_{\alpha\beta\gamma\rho}R^{(a)\rho}{}_\delta + \epsilon_{\rho\beta\gamma\delta}R^{(a)\rho}{}_\alpha + \epsilon_{\alpha\rho\gamma\delta}R^{(a)\rho}{}_\beta) = 0, \tag{2.17}$$

$$B\epsilon_{\alpha\beta\gamma\rho}T^{(a)\rho}+A^{\delta}(\epsilon_{\rho\beta\gamma\delta}R^{(a)\rho}{}_{\alpha}-\epsilon_{\rho\alpha\gamma\delta}R^{(a)\rho}{}_{\beta}+\epsilon_{\rho\alpha\beta\delta}R^{(a)\rho}{}_{\gamma})=0, \tag{2.18}$$

where $a,b=1,2$ (respectively, $a,b=1,2,3$).

III. THE CLASSIFICATION

We first study the case of maximal three-dimensional subspaces $\mathscr{P}_3$ of $\mathscr{P}$ which form an Abelian Lie algebra under the commutator.

Let $\mathscr{L}_2:=\text{span}\{k^{(1)},k^{(2)}\}$ and $\mathscr{L}_3:=\text{span}\{k^{(1)},k^{(2)},k^{(3)}\}$. We have $\dim\mathscr{L}_2=2$ and $\dim\mathscr{L}_3\geqslant 2$ in order for $\mathscr{P}_3$ to be three-dimensional and $k^{(3)}\in C(\mathscr{L}_2):=\text{centralizer}(\mathscr{L}_2)$ in the Lie algebra p(1,3) of the Poincaré group P(1,3).

Using the adjoint action of P(1,3) on p(1,3), we partition the set of two-dimensional Abelian subalgebras $\mathscr{L}_2$ of p(1,3) into orbits, as was done in Ref. 11. For each orbit representative $\mathscr{L}_2=\text{span}\{k^{(1)},k^{(2)}\}$, we apply Corollary 2 to obtain the valence-2 and valence-3 Killing–Yano tensors $(k_{\alpha\beta})$ and $(k_{\alpha\beta\gamma})$ that ensure the commutation property.

If $\dim\mathscr{L}_3=2$, there exists a basis for $\mathscr{P}_3$ of the form $(\mathbb{K}^{(1)}(k^{(1)},0,0),\mathbb{K}^{(2)}(k^{(2)},0,0),\mathbb{K}^{(3)}(0,\Pi,\Lambda))$. We use the second and third exterior powers of $\exp(N(\mathscr{L}_2))$, where $N(\mathscr{L}_2):=\text{normalizer}(\mathscr{L}_2)$ in p(1,3) to cast $(k_{\alpha\beta})$ and $(k_{\alpha\beta\gamma})$ in normal form.

If $\dim\mathscr{L}_3=3$, we first use the adjoint action of $\exp(N(\mathscr{L}_2))$ to partition the set of one-dimensional subalgebras of the (additive) quotient $C(\mathscr{L}_2)/\mathscr{L}_2$ into orbits. [Note that the action of $\exp(N(\mathscr{L}_2))$ is well defined in the quotient $C(\mathscr{L}_2)/\mathscr{L}_2$ since $\text{Ad}(\exp N(\mathscr{L}_2))$ maps $C(\mathscr{L}_2)$ onto $C(\mathscr{L}_2)$ and $\mathscr{L}_2$ onto $\mathscr{L}_2$.] For each orbit representative span $\{[k^{(3)}]\}$, we check that the valence-2 and valence-3 Killing–Yano tensors $(k_{\alpha\beta})$ and $(k_{\alpha\beta\gamma})$ obtained above do not satisfy the commutation conditions (2.16)–(2.18) of Corollary 2, with $a=3$. If these conditions were satisfied, the triple $\{\mathbb{K}^{(1)},\mathbb{K}^{(2)},\mathbb{K}^{(3)}\}$ would not span a maximal subspace forming an Abelian Lie algebra under the commutator and thus would have to be rejected. If the commutation conditions are not satisfied, it remains to be verified that there is no element $[k^{(4)}]$ of $C(\mathscr{L}_2)/\mathscr{L}_2$ such that the quadruple $\{\mathbb{K}^{(1)},\mathbb{K}^{(2)},\mathbb{K}^{(3)},\mathbb{K}^{(4)}(k^{(4)},0,0)\}$ spans a four-dimensional subspace that forms an Abelian Lie algebra under the commutator. Indeed, if such a $[k^{(4)}]$ existed, the triple $\{\mathbb{K}^{(1)},\mathbb{K}^{(2)},\mathbb{K}^{(3)}\}$ would again not span a maximal three-dimensional subspace forming an Abelian Lie algebra under the commutator and thus would have to be rejected. If no such $[k^{(4)}]$ exists, the span is maximal and one uses the second and third exterior powers of $\exp(N(\mathscr{L}_2)\cap N(\text{span}\{[k^{(3)}]\}))$ to cast $(k_{\alpha\beta})$ and $(k_{\alpha\beta\gamma})$ in normal form. The result of this procedure is presented in Table I, where we have used the action of the discrete subgroup $\{\text{Id}_{M_4},\exp(\pi L_3)\}$ of P(1,3) to put sharper restrictions on the range of the parameters labeling the equivalence classes.

The case of four-dimensional subspaces $\mathscr{P}_4$ of $\mathscr{P}$ is treated following a procedure similar to the one just described for three-dimensional subspaces. The results are recorded in Table II.

In Tables I and II, we have used the following notation for the generators of p(1,3) in its standard vector field representation on M_4:

$$L_i:=-\frac{1}{2}\epsilon_{ijk}M_{jk},\quad K_i:=M_{0i},\quad P_\alpha:=-\frac{\partial}{\partial x^\alpha}, \tag{3.1}$$

where

$$M_{\mu\nu}:=x_\nu\frac{\partial}{\partial x^\mu}-x_\mu\frac{\partial}{\partial x^\nu}, \tag{3.2}$$

where the indices i and α take the values 1,2,3 and 0,1,2,3, respectively.

Finally, it should be mentioned that one can also classify under P(1,3) the maximal subspaces of dimension 5 of $\mathscr{P}$ which form an Abelian Lie algebra under the commutator, using the above method. One obtains subspaces of the form span $\{\mathbb{K}^{(1)}(P_0,0,0),\mathbb{K}^{(2)}(P_1,0,0),\mathbb{K}^{(3)}(P_2,0,0),\mathbb{K}^{(4)}(P_3,0,0),\mathbb{K}^{(5)}(0,\Pi,\Lambda)\}$, where

$$\Pi=\tfrac{1}{2}C_{\alpha\beta}\,dx^\alpha\wedge dx^\beta,\quad \Lambda=\tfrac{1}{8}\epsilon_{\alpha\beta\gamma\delta}A^{\delta}\,dx^\alpha\wedge dx^\beta\wedge dx^\gamma,$$

and the parameters $C_{\alpha\beta}$ and A^{δ} take the values listed in Appendix F.

This exhausts the possible dimensions for maximal subspaces of $\mathscr{P}$ which form an Abelian Lie algebra under the commutator.

IV. CONCLUSIONS

We have classified under the action of the Poincaré group P(1,3) the maximal vector spaces of formally self-adjoint first-order matrix differential operators commuting among themselves and with the Dirac operator. These vector spaces will be crucial in the classification of the separable coordinate systems for the Dirac equation, as illustrated in the following example.[9]

Consider four-dimensional Minkowski space M_4 in oblate spheroidal coordinates (t,r,α,ϕ) defined as

$$\begin{aligned}&x^0=t,\quad x^1=(r^2+\hat{a}^2)^{1/2}\sin\alpha\cos\phi,\\&x^2=(r^2+\hat{a}^2)^{1/2}\sin\alpha\sin\phi,\quad x^3=r\cos\alpha.\end{aligned} \tag{4.1}$$

When expressed in the above coordinates and in an appropriately chosen moving frame, the Dirac equation is solvable by separation of variables.[5] The separable solutions are eigenspinors of the commuting operators $\mathbb{K}^{(1)}(\partial/\partial t,0,0)$, $\mathbb{K}^{(2)}(\partial/\partial\phi,0,0)$, and $\mathbb{K}^{(3)}(0,\frac{1}{2}D_{ij}\,dx^i\wedge dx^j,0)$, where

$$(D^i_j):=\begin{pmatrix}0&0&0&-\hat{a}\\0&0&-x^3&x^2\\0&x^3&0&-x^1\\\hat{a}&-x^2&x^1&0\end{pmatrix}, \tag{4.2}$$

with the separation constants as eigenvalues. The frame required for separation of variables is precisely (up to normalizations) the frame of eigenvectors of (D^i_j) corresponding to the eigenvalues $\pm ir$ and $\pm i\hat{a}\cos\theta$. The complete set $\{\mathbb{K}^{(1)},\mathbb{K}^{(2)},\mathbb{K}^{(3)}\}$ corresponds to case 7 of Table I, with $a=0$ and $C_{03}=\hat{a}$.

TABLE I. Three-dimensional subspaces span $\{\mathbf{K}^{(1)}(k^{(1)},0,0),\mathbf{K}^{(2)}(k^{(2)},0,0),\mathbf{K}^{(3)}(k^{(3)},\Pi,\Lambda)\}$.

Case	$k^{(1)}$	$k^{(2)}$	$k^{(3)}$	Π	Λ	Parameters	Values
1	L_3	K_3	0	$\cos\phi\, dx^0\wedge dx^3+\sin\phi\, dx^1\wedge dx^2$	$\frac{1}{6}B\epsilon_{\alpha\beta\gamma\delta}x^\delta dx^\alpha\wedge dx^\beta\wedge dx^\gamma$	B,ϕ	$B\in\mathbf{R},0\leqslant\phi<\pi$
2	L_3	K_3	0	0	$\frac{1}{6}\epsilon_{\alpha\beta\gamma\delta}x^\delta dx^\alpha\wedge dx^\beta\wedge dx^\gamma$	...	...
3	L_2+K_1	L_1-K_2	$\cos\theta(P_0-P_3)$	$\sin\theta[\sin\phi(dx^0+dx^3)\wedge dx^1+\cos\phi\{(dx^0+dx^3)\wedge(x^2dx^1-x^1dx^2)+(x^0+x^3)dx^1\wedge dx^2\}]$	$\frac{1}{6}\epsilon_{\alpha\beta\gamma\delta}x^\delta dx^\alpha\wedge dx^\beta\wedge dx^\gamma$	θ,ϕ	$0\leqslant\theta,\phi\leqslant\pi$
4	L_2+K_1	L_1-K_2	0	0	$\frac{1}{6}\epsilon_{\alpha\beta\gamma\delta}x^\delta dx^\alpha\wedge dx^\beta\wedge dx^\gamma$	...	...
5	L_3	P_0-P_3	$b(P_0+P_3)$	$C_{03}\,dx^0\wedge dx^3+(x^0+x^3)dx^1\wedge dx^2+(dx^0+dx^3)\wedge(x^2dx^1-x^1dx^2)$	$(A^3dx^0-A^0dx^3)\wedge dx^1\wedge dx^2$	(C_{03},A^0,A^3,b)	See Appendix A
6	L_3	P_3	aP_0	$C_{03}\,dx^0\wedge dx^3-x^0dx^1\wedge dx^2+dx^0\wedge(x^1dx^2-x^2dx^1)$	$(A^3dx^0-A^0dx^3)\wedge dx^1\wedge dx^2$	C_{03},A^0,A^3,a	$C_{03},A^0,A^3,a\in\mathbf{R}$
7	L_3	P_0	aP_3	$C_{03}\,dx^0\wedge dx^3+x^3dx^1\wedge dx^2+(x^1dx^2-x^2dx^1)\wedge dx^3$	$(A^3dx^0-A^0dx^3)\wedge dx^1\wedge dx^2$	C_{03},A^0,A^3,a	$C_{03},A^0,A^3,a\in\mathbf{R}$
8	K_3	P_1	aP_2	$C_{12}\,dx^1\wedge dx^2+x^2dx^0\wedge dx^3+(x^0dx^3-x^3dx^0)\wedge dx^2$	$(A^2dx^1-A^1dx^2)\wedge dx^0\wedge dx^3$	C_{12},A^1,A^2,a	$A^1,A^2,a\in\mathbf{R}$ $C_{12}\geqslant0$
9	L_2+K_1	P_0-P_3	$a(L_1-K_2)$	$(x^0+x^3)dx^1\wedge dx^2+(x^1dx^2-x^2dx^1)\wedge(dx^0+dx^3)$	$-dx^0\wedge dx^1\wedge dx^3$	a	$a\in\mathbf{R}$
10	L_2+K_1	P_0-P_3	$a(L_1-K_2)$	$\cos\phi(dx^0+dx^3)\wedge dx^1+\sin\phi(dx^0+dx^3)\wedge dx^2$	$-dx^0\wedge dx^1\wedge dx^3$	a,ϕ	$a\in\mathbf{R}^*$ $0\leqslant\phi<2\pi$
11	L_2+K_1	P_0-P_3	$\epsilon(L_1-K_2)$	0	$-dx^0\wedge dx^1\wedge dx^3$	ϵ	$\epsilon^2=1$
12	L_2+K_1	P_0-P_3	bP_2	$(x^0+x^3)dx^1\wedge dx^2+(x^1dx^2-x^2dx^1)\wedge(dx^0+dx^3)$	$-dx^0\wedge dx^1\wedge dx^3$	b	$b\in\mathbf{R}^*$
13	L_2+K_1	P_0-P_3	P_2	$(x^0+x^3)dx^1\wedge dx^2+(x^1dx^2-x^2dx^1)\wedge(dx^0+dx^3)$	$-A^0(dx^0+dx^3)\wedge dx^1\wedge dx^2$	A^0	$A^0\in\mathbf{R}$
14	L_2+K_1	P_0-P_3	$c(L_1-K_2+\epsilon P_2)$	$(x^0+x^3)dx^1\wedge dx^2+(x^1dx^2-x^2dx^1)\wedge(dx^0+dx^3)$	$-A^2dx^0\wedge dx^1\wedge dx^3$	c,A^2,ϵ	$c>0$ or $A^2>0$ $\epsilon^2=1$
15	L_2+K_1	P_0-P_3	$c(L_1-K_2+\epsilon P_2)$	$(x^0+x^3)dx^1\wedge dx^2+(x^1dx^2-x^2dx^1)\wedge(dx^0+dx^3)$	$-A^0(dx^0+dx^3)\wedge dx^1\wedge dx^2$	c,A^0,ϵ	$c>0,A^0\in\mathbf{R}$ $\epsilon^2=1$
16	L_2+K_1	P_0-P_3	$c(L_1-K_2+\epsilon P_2)$	$C_{01}(dx^0+dx^3)\wedge dx^1+C_{02}(dx^0+dx^3)\wedge dx^2$	$-dx^0\wedge dx^1\wedge dx^3$	c,C_{01},C_{02},ϵ	$C_{01},C_{02}\in\mathbf{R}$ $c\in\mathbf{R}^*,\epsilon^2=1$
17	L_2+K_1	P_2	$\cos\theta(P_0-P_3)$	$\sin\theta\cos\phi(dx^0+dx^3)\wedge dx^2-x^1dx^0\wedge dx^3-(x^0dx^3-x^3dx^0)\wedge dx^1$	$-\sin\theta\sin\phi(dx^0+dx^3)\wedge dx^1\wedge dx^2-A^2dx^0\wedge dx^1\wedge dx^3$	θ,ϕ A^2	$0\leqslant\theta,\phi\leqslant\pi$ $A^2\in\mathbf{R}$
18	L_2+K_1	P_2	0	$-x^1dx^0\wedge dx^3-(x^0dx^3-x^3dx^0)\wedge dx^1$	$-A^2dx^0\wedge dx^1\wedge dx^3$	A^2	$A^2\in\mathbf{R}$
19	P_0-P_3	P_1	$c(L_1-K_2)$	$\frac{1}{2}C_{\alpha\beta}\,dx^\alpha\wedge dx^\beta$	$\frac{1}{6}\epsilon_{\alpha\beta\gamma\delta}A^\delta dx^\alpha\wedge dx^\beta\wedge dx^\gamma$	$C_{\alpha\beta},A^\delta,c$	See Appendix B
20	P_0-P_3	P_1	$d(L_1-K_2+P_0+P_3)$	$\frac{1}{2}C_{\alpha\beta}\,dx^\alpha\wedge dx^\beta$	$\frac{1}{6}\epsilon_{\alpha\beta\gamma\delta}A^\delta dx^\alpha\wedge dx^\beta\wedge dx^\gamma$	$C_{\alpha\beta},A^\delta,d$	See Appendix C
21	P_0	P_3	aL_3	$\frac{1}{2}C_{\alpha\beta}\,dx^\alpha\wedge dx^\beta$	$\frac{1}{6}\epsilon_{\alpha\beta\gamma\delta}A^\delta dx^\alpha\wedge dx^\beta\wedge dx^\gamma$	$C_{\alpha\beta},A^\delta,a$	See Appendix D

TABLE I. (*Continued.*)

Case	$k^{(1)}$	$k^{(2)}$	$k^{(3)}$	Π	Λ	Parameters	Values
22	P_1	P_2	bK_3	$\frac{1}{2}C_{\alpha\beta}\,dx^\alpha\wedge dx^\beta$	$\frac{1}{6}\epsilon_{\alpha\beta\gamma\delta}A^\delta\,dx^\alpha\wedge dx^\beta\wedge dx^\gamma$	$C_{\alpha\beta},A^\delta,b$	See Appendix E
23	$L_2+K_1+P_0+P_3$	P_0-P_3	aP_2	$(x^0+x^3)dx^1\wedge dx^2-2\,dx^2\wedge dx^3$ $+(x^1\,dx^2-x^2\,dx^1)\wedge(dx^0+dx^3)$	$-A^0(dx^0+dx^3)\wedge dx^1\wedge dx^2$ $-A^2\,dx^0\wedge dx^1\wedge dx^3$	a,A^0,A^2	$a,A^0,A^2\in\mathbf{R}$
24	$L_2+K_1+\epsilon P_2$	P_0-P_3	dP_2	$(x^0+x^3)dx^1\wedge dx^2+\epsilon\,dx^0\wedge dx^3$ $+(x^1\,dx^2-x^2\,dx^1)\wedge(dx^0+dx^3)$	$-A^2\,dx^0\wedge dx^1\wedge dx^3$	b,A^2 ϵ	$A^2>0$ or $b>0$ $\epsilon^2=1$
25	$L_2+K_1+\epsilon P_2$	P_0-P_3	bP_2	$(x^0+x^3)dx^1\wedge dx^2+\epsilon\,dx^0\wedge dx^3$ $+(x^1\,dx^2-x^2\,dx^1)\wedge(dx^0+dx^3)$	$-A^0(dx^0+dx^3)\wedge dx^1\wedge dx^2$	b,A^0 ϵ	$A^0\in\mathbf{R},b>0$ $\epsilon^2=1$
26	$L_2+K_1+\epsilon P_2$	P_0-P_3	$c(L_1-K_2-\epsilon P_1+\alpha P_2)$	$(x^0+x^3)dx^1\wedge dx^2+\epsilon\,dx^0\wedge dx^3$ $+(x^1\,dx^2-x^2\,dx^1)\wedge(dx^0+dx^3)$	$-A^2dx^0\wedge dx^1\wedge dx^3$	α,c,A^2 ϵ	$\alpha\in\mathbf{R},A^2\in\mathbf{R}^*$ $c>0,\epsilon^2=1$
27	$L_2+K_1+\epsilon P_2$	P_0-P_3	$c(L_1-K_2-\epsilon P_1+\alpha P_2)$	$(x^0+x^3)dx^1\wedge dx^2+\epsilon\,dx^0\wedge dx^3$ $+(x^1\,dx^2-x^2\,dx^1)\wedge(dx^0+dx^3)$	$-A^0(dx^0+dx^3)\wedge dx^1\wedge dx^2$	α,c,A^0 ϵ	$\alpha,A^0\in\mathbf{R}$ $c>0,\epsilon^2=1$
28	$L_2+K_1+\epsilon P_2$	P_0-P_3	$c(L_1-K_2-\epsilon P_1+\alpha P_2$	$C_{01}(dx^0+dx^3)\wedge dx^1$ $+C_{02}(dx^0+dx^3)\wedge dx^2$	$-dx^0\wedge dx^1\wedge dx^3$	α,c,C_{01} C_{02},ϵ	$C_{01},C_{02},\alpha\in\mathbf{R}$ $C\in\mathbf{R}^*,\epsilon^2=1$

TABLE II. Four-dimensional subspaces span $\{\mathbf{K}^{(1)}(k^{(1)},0,0), \mathbf{K}^{2}(k^{(2)},0,0), \mathbf{K}^{(3)}(k^{(3)},0,0), \mathbf{K}^{(4)}(0,\Pi,\Lambda)\}$.

Case	$k^{(1)}$	$k^{(2)}$	$k^{(3)}$	Π	Λ	Parameters	Values
1	L_2+K_1	L_1-K_2	P_0-P_3	$(x^0+x^3)dx^1\wedge dx^2$ $+(x^1dx^2-x^2dx^1)\wedge(dx^0+dx^3)$	$-A^0(dx^0+dx^3)\wedge dx^1\wedge dx^2$	A^0	$A^0\in\mathbf{R}$
2	L_2+K_1	L_1-K_2	P_0-P_3	$(dx^0+dx^3)\wedge dx^2$	$-(dx^0+dx^3)\wedge dx^1\wedge dx^2$	...	...
3	L_2+K_1	L_1-K_2	P_0-P_3	$(dx^0+dx^3)\wedge dx^2$	0	...	...
4	L_2+K_1	L_1-K_2	P_0-P_3	0	$(dx^0+dx^3)\wedge dx^1\wedge dx^2$	...	...
5	L_3	P_0	P_3	$C_{03}\,dx^0\wedge dx^3+C_{12}\,dx^1\wedge dx^2$	$dx^0\wedge dx^1\wedge dx^2$	C_{03},C_{12}	$C_{03},C_{12}\in\mathbf{R}$
6	L_3	P_0	P_3	$C_{03}\,dx^0\wedge dx^3+C_{12}\,dx^1\wedge dx^2$	$dx^3\wedge dx^1\wedge dx^2$	C_{03},C_{12}	$C_{03},C_{12}\in\mathbf{R}$
7	L_3	P_0	P_3	$\sin\phi\,dx^0\wedge dx^3+\cos\phi\,dx^1\wedge dx^2$	$(\epsilon\,dx^0-dx^3)\wedge dx^1\wedge dx^2$	ϵ,ϕ	$0\leqslant\phi<2\pi,\epsilon^2=1$
8	L_3	P_0	P_3	0	$(\epsilon\,dx^0-dx^3)\wedge dx^1\wedge dx^2$	ϵ	$\epsilon^2=1$
9	L_3	P_0	P_3	$\sin\phi\,dx^0\wedge dx^3+\cos\phi\,dx^1\wedge dx^2$	0	ϕ	$0\leqslant\phi<\pi$
10	K_3	P_1	P_2	$C_{03}\,dx^0\wedge dx^3+C_{12}\,dx^1\wedge dx^2$	$dx^1\wedge dx^0\wedge dx^3$	C_{03},C_{12}	$C_{03},C_{12}\in\mathbf{R}$
11	K_3	P_1	P_2	$\sin\phi\,dx^0\wedge dx^3+\cos\phi\,dx^1\wedge dx^2$	0	ϕ	$0\leqslant\phi<2\pi$
12	L_2+K_1	P_0-P_3	P_2	$\sin\phi(dx^0+dx^3)\wedge dx^1$ $+\cos\phi(dx^0+dx^3)\wedge dx^2$	$-dx^0\wedge dx^1\wedge dx^3$	ϕ	$0\leqslant\phi<2\pi$
13	L_2+K_1	P_0-P_3	P_2	0	$-dx^0\wedge dx^1\wedge dx^3$	...	...
14	L_2+K_1	P_0-P_3	P_2	$\sin\theta\sin\phi(dx^0+dx^3)\wedge dx^1$ $+\sin\theta\cos\phi(dx^0+dx^3)\wedge dx^2$	$-\cos\theta(dx^0+dx^3)\wedge dx^1\wedge dx^2$	θ,ϕ	$0\leqslant\phi<2\pi,0\leqslant\theta<\pi/2$
15	L_2+K_1	P_0-P_3	$L_1-K_2+P_2$	$C_{01}(dx^0+dx^3)\wedge dx^1$ $+C_{02}(dx^0+dx^3)\wedge dx^2$	$(dx^0+dx^3)\wedge dx^1\wedge dx^2$	C_{01},C_{02}	$C_{01},C_{02}\in\mathbf{R}$
16	P_0-P_3	P_1	$L_1-K_2+P_0+P_3$	$C_{01}(dx^0+dx^3)\wedge dx^1$ $+C_{02}(dx^0+dx^3)\wedge dx^2$	$(dx^0+dx^3)\wedge dx^1\wedge dx^2$	C_{01},C_{02}	$C_{01},C_{02}\in\mathbf{R}$
17	P_0-P_3	P_1	$L_1-K_2+P_0+P_3$	$\sin\phi(dx^0+dx^3)\wedge dx^1$ $+\cos\phi(dx^0+dx^3)\wedge dx^2$	0	ϕ	$0\leqslant\phi<2\pi$
18	P_0-P_3	P_1	$L_1-K_2+P_0+P_3$	$C_{01}(dx^0+dx^3)\wedge dx^1$ $+C_{02}(dx^0+dx^3)\wedge dx^2$	$dx^0\wedge dx^2\wedge dx^3$	C_{01},C_{02}	$C_{01},C_{02}\in\mathbf{R}$

ACKNOWLEDGMENTS

It is a pleasure to thank Professor D. Z. Djokovic for a useful discussion.

One of us (ML) thanks the Centre de Recherches Mathématiques for its kind hospitality. This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada, the fonds FCAR du Gouvernement du Québec pour les chercheurs sans affiliation institutionnelle, and the Ontario–Québec Exchange Program.

APPENDIX A: VALUES OF (C_{03},A^0,A^3,b) FOR CASE 5 OF TABLE I

The values taken by the parameters (C_{03},A^0,A^3,b) for Case 5 of Table I are

5.1. $(C_{03},0,A^3,b)$

5.2. $(C_{03},A^0,0,b)$

5.3. $(C_{03},\epsilon,\epsilon,b)$

5.4. $(C_{03},A^0,-A,\epsilon)$

5.5. $(\epsilon,A^0,-A^0,0)$

5.6. $(0,A^0,-A^0,0)$

5.7. $(C_{03},0,0,\epsilon)$

5.8. $(\epsilon,0,0,0)$

5.9. $(0,0,0,0)$;

$$C_{03},b\in\mathbf{R},\quad A^0,A^3\in\mathbf{R}^*,\quad \epsilon^2=1.$$

APPENDIX B: VALUES OF $(C_{\alpha\beta})$, (A^5), AND c FOR CASE 19 OF TABLE I

The values taken by the parameters $(C_{01},C_{02},C_{13},C_{23},C_{03},C_{12},A^0,A^1,A^2,A^3,c)$ for case 19 of Table I are

19.1. $(0,0,1,C_{23},0,0,A^0,A^1,A^2,A^3,c)$; $c,C_{23}\in\mathbf{R}^*$, $A^0,A^1,A^2\in\mathbf{R}$, $A^3>0$.

19.2. $(1,C_{02},0,0,0,0,A^0,A^1,A^2,A^3,c)$; $c,C_{02}\in\mathbf{R}^*$, $A^0,A^1,A^2\in\mathbf{R}$, $A^3>0$.

19.3 $(1,C_{02},1,C_{02},0,0,0,A^1,A^2,A^3,c)$; $c,C_{02}\in\mathbf{R}^*$, $A^1,A^2\in\mathbf{R}$, $A^3>0$.

19.4. $(1,C_{02},1,C_{02},0,0,A^0,A^1,A^2,0,c)$; $c,C_{02}\in\mathbf{R}^*$, $A^1,A^2\in\mathbf{R}$, $A^0>0$.

19.5. $(1,C_{02},1,C_{02},0,0,A^0,\cos\phi,\sin\phi,A^0,c)$; $c,C_{02}\in\mathbf{R}^*$, $A^0>0$, $0<\phi<2\pi$.

19.6. $(1,C_{02},1,C_{02},0,0,A^0,0,0,A^0,\epsilon)$; $C_{02}\in\mathbf{R}^*$, $A^0>0$, $\epsilon^2=1$.

19.7. $(1,C_{02},1,C_{02},0,0,1,A^1,A^2,-1,c)$; $c\in\mathbf{R}^*$, $A^1,A^2\in\mathbf{R}$.

19.8. $(0,1,C_{13},C_{23},0,0,A^0,A^1,A^2,A^3,c)$; $c,C_{13}\in\mathbf{R}^*$, $C_{23},A^1,A^2,A^3\in\mathbf{R}$, $A^0>0$.

19.9. $(1,C_{02},0,C_{23},0,0,A^0,A^1,A^2,A^3,c)$; $c,C_{23}\in\mathbf{R}^*$, $C_{02},A^1,A^2,A^3\in\mathbf{R}$, $A^0>0$.

19.10. $(1,0,1,C_{23},0,0,A^0,A^1,A^2,A^3,c)$; $c,C_{23}\in\mathbf{R}^*$, $A^1,A^2,A^3\in\mathbf{R}$, $A^0>0$.

19.11. $(1,C_{02},1,0,0,0,A^0,A^1,A^2,A^3,c)$; $c,C_{02}\in\mathbf{R}^*$, $A^1,A^2,A^3\in\mathbf{R}$, $A^0>0$.

19.12. $(0,0,0,1,0,0,A^0,A^1,A^2,A^3,c)$; $c\in\mathbf{R}^*$, $A^1,A^2,A^3\in\mathbf{R}$, $A^0>0$.

19.13. $(0,1,0,0,0,0,A^0,A^1,A^2,A^3,c)$; $c\in\mathbf{R}^*$, $A^1,A^2,A^3\in\mathbf{R}$, $A^0>0$.

19.14. $(0,1,0,1,0,0,0,A^1,A^2,A^3,c)$; $c\in\mathbf{R}^*$, $A^1,A^2\in\mathbf{R}$, $A^3>0$.

19.15. $(0,1,0,1,0,0,A^0,A^1,A^2,0,c)$; $c\in\mathbf{R}^*$, $A^1,A^2\in\mathbf{R}$, $A^0>0$.

19.16. $(0,1,0,1,0,0,A^0,\cos\phi,\sin\phi,A^0,c)$; $c\in\mathbf{R}^*$, $A^0>0$, $0<\phi<2\pi$.

19.17. $(0,1,0,1,0,0,A^0,0,0,A^0,\epsilon)$; $A^0>0$, $\epsilon^2=1$.

19.18. $(0,1,0,1,0,0,1,A^1,A^2,-1,c)$; $c\in\mathbf{R}^*$, $A^1,A^2\in\mathbf{R}$.

19.19. $(1,0,-1,C_{23},0,0,A^0,A^1,A^2,A^3,c)$; $c,C_{23}\in\mathbf{R}^*$, $A^1,A^2,A^3\in\mathbf{R}$, $A^0>0$.

19.20. $(1,C_{02},-1,0,0,0,A^0,A^1,A^2,A^3,c)$; $c,C_{02}\in\mathbf{R}^*$, $A^1,A^2,A^3\in\mathbf{R}$, $A^0>0$.

19.21. $(1,1,-1,1,0,0,A^0,A^1,A^2,A^3,c)$; $c\in\mathbf{R}^*$, $A^1,A^2,A^3\in\mathbf{R}$, $A^0>0$.

19.22. $(1,-1,-1,-1,0,0,A^0,A^1,A^2,A^3,c)$; $c\in\mathbf{R}^*$, $A^1,A^2,A^3\in\mathbf{R}$, $A^0>0$.

19.23. $(0,0,1,0,0,0,A^0,A^1,A^2,A^3,c)$; $c\in\mathbf{R}^*$, $A^1,A^2,A^3\in\mathbf{R}$, $A^0>0$.

19.24. $(1,0,0,0,0,0,A^0,A^1,A^2,A^3,c)$; $c\in\mathbf{R}^*$, $A^1,A^2,A^3\in\mathbf{R}$, $A^0>0$.

19.25. $(1,0,1,0,0,0,0,A^1,A^2,A^3,c)$; $c\in\mathbf{R}^*$, $A^1,A^2\in\mathbf{R}$, $A^3>0$.

19.26. $(1,0,1,0,0,0,A^0,A^1,A^2,0,c)$; $c\in\mathbf{R}^*$, $A^1,A^2\in\mathbf{R}$, $A^0>0$.

19.27. $(1,0,1,0,0,0,A^0,\cos\phi,\sin\phi,A^0,c)$; $c\in\mathbf{R}^*$, $0<\phi<2\pi$, $A^0>0$.

19.28. $(1,0,1,0,0,0,A^0,0,0,A^0,\epsilon)$; $A^0>0$, $\epsilon^2=1$.

19.29. $(1,0,1,0,0,0,1,A^1,A^2,-1,c)$; $c\in\mathbf{R}^*$, $A^1,A^2\in\mathbf{R}$.

19.30. $(0,1,C_{13},-1,0,0,A^0,A^1,A^2,A^3,c)$; $c,C_{13}\in\mathbf{R}^*$, $A^1,A^2,A^3\in\mathbf{R}$, $A^0>0$.

19.31. $(C_{01},1,0,-1,0,0,A^0,A^1,A^2,A^3,c)$; $c,C_{01}\in\mathbf{R}^*$, $A^1,A^2,A^3\in\mathbf{R}$, $A^0>0$.

19.32. $(1,1,1,-1,0,0,A^0,A^1,A^2,A^3,c)$; $c\in\mathbb{R}^*$, $A^1,A^2,A^3\in\mathbb{R}$, $A^0\geqslant 0$.

19.33. $(1,-1,1,1,0,0,A^0,A^1,A^2,A^3,c)$; $c\in\mathbb{R}^*$, $A^1,A^2,A^3\in\mathbb{R}$, $A^0\geqslant 0$.

19.34. $(0,0,0,0,\cos\theta,\sin\theta,0,A^1,A^2,A^3,c)$; $A^1,A^2\in\mathbb{R}$, $A^3\in\mathbb{R}^*$, $0\leqslant\theta<\pi$, $c>0$.

19.35. $(0,0,0,0,\cos\theta,\sin\theta,A^0,A^1,A^2,0,c)$; $A^1,A^2\in\mathbb{R}$, $A^0\in\mathbb{R}^*$, $0\leqslant\theta<\pi$, $c>0$.

19.36. $(0,0,0,0,\cos\theta,\sin\theta,\epsilon,A^1,A^2,\epsilon',c)$; $A^1,A^2\in\mathbb{R}$, $0\leqslant\theta<\pi$, $\epsilon^2=1=\epsilon'^2$.

19.37. $(0,0,0,0,\cos\theta,\sin\theta,0,A^1,A^2,0,1)$; $A^1,A^2\in\mathbb{R}$, $0\leqslant\theta<\pi$.

19.38. $(\cos\theta,\sin\theta,-\cos\theta,-\sin\theta,0,0,0,0,0,A^3,c)$; $c\in\mathbb{R}^*$, $0\leqslant\theta<\pi$, $A^3>0$.

19.39. $(\cos\theta,\sin\theta,-\cos\theta,-\sin\theta,0,0,A^0,0,0,0,c)$; $c\in\mathbb{R}^*$, $0\leqslant\theta<\pi$, $A^0>0$.

19.40. $(\cos\theta,\sin\theta,-\cos\theta,-\sin\theta,0,0,0,1,0,0,1,c)$; $c\in\mathbb{R}^*$, $0\leqslant\theta<\pi$.

19.41. $(\cos\theta,\sin\theta,-\cos\theta,-\sin\theta,0,0,0,1,\cos\phi,\sin\phi,0,c)$; $c\in\mathbb{R}^*$, $0\leqslant\theta<\pi$, $0<\phi<2\pi$, $\phi\neq\pi$.

19.42. $(0,0,0,0,0,0,0,0,0,0,1,c)$; $c>0$.

19.43. $(0,0,0,0,0,0,1,0,0,0,c)$; $c>0$.

19.44. $(0,0,0,0,0,0,1,0,0,1,1)$.

19.45. $(0,0,0,0,0,0,0,0,\cos\phi,\sin\phi,0,1)$; $0<\phi<2\pi$, $\phi\neq\pi$.

APPENDIX C: VALUES OF $(C_{\alpha\beta})$, (A^δ), AND d FOR CASE 20 OF TABLE I

The values taken by the parameters $(C_{01},C_{02},C_{13},C_{23},C_{03},C_{12},A^0,A^1,A^2,A^3,d)$ for case 20 of Table I are

20.1. $(\cos\phi,C_{02},\sin\phi,C_{23},0,C_{12},A^0,A^1,A^2,A^3,d)$; $d\in\mathbb{R}^*$, $C_{02},C_{23},C_{12},A^\alpha\in\mathbb{R}$, $0\leqslant\phi<2\pi$.

20.2. $(C_{01},\cos\phi,-C_{01},\sin\phi,C_{03},0,A^0,A^1,A^2,A^3,d)$; $d\in\mathbb{R}^*$, $C_{01},C_{03},A^\alpha\in\mathbb{R}$, $0\leqslant\phi<2\pi$.

20.3. $(0,C_{02},0,-C_{02},1,C_{12},A^0,A^1,A^2,A^3,d)$; $d\in\mathbb{R}^*$, $C_{02},C_{12},A^\alpha\in\mathbb{R}$.

20.4. $(C_{01},0,-C_{01},0,0,1,A^0,A^1,A^2,A^3,d)$; $d\in\mathbb{R}^*$, $C_{01},A^\alpha\in\mathbb{R}$.

20.5. $(C_{01},C_{02},-C_{01},-C_{02},0,0,0,1,A^2,0,d)$; $C_{01},C_{02}\in\mathbb{R}$, A^2, $d\in\mathbb{R}^*$.

20.6. $(C_{01},C_{02},-C_{01},-C_{02},0,0,\cos\phi,0,A^2,\sin\phi,d)$; $d\in\mathbb{R}^*$, $C_{01},C_{02},A^2\in\mathbb{R}$, $0\leqslant\phi<2\pi$.

20.7. $(C_{01},C_{02},-C_{01},-C_{02},0,0,A^0,0,A^2,-A^0,1)$; $A^2\in\mathbb{R}^*$, $C_{01},C_{02},A^0\in\mathbb{R}$.

APPENDIX D: VALUES OF $(C_{\alpha\beta})$, (A^δ), AND a FOR CASE 21 OF TABLE I

The values taken by the parameters $(C_{01},C_{02},C_{13},C_{23},C_{03},C_{12},A^0,A^1,A^2,A^3,a)$ for case 21 of Table I are

21.1. $(0,0,0,1,C_{03},C_{12},A^0,A^1,A^2,A^3,a)$; $C_{03},C_{12},A^\alpha\in\mathbb{R}$, $a>0$.

21.2. $(0,1,0,0,C_{03},C_{12},A^0,A^1,A^2,A^3,a)$; $C_{03},C_{12},A^\alpha\in\mathbb{R}$, $a>0$.

21.3. $(0,1,C_{13},0,C_{03},C_{12},A^0,A^1,A^2,A^3,a)$; $C_{13},C_{03},C_{12},A^\alpha\in\mathbb{R}$, $a>0$.

21.4. $(0,1,0,1,C_{03},C_{12},0,A^1,A^2,A^3,a)$; $C_{03},C_{12},A^1,A^2\in\mathbb{R}$, $A^3\in\mathbb{R}^*$, $a>0$.

21.5. $(0,1,0,1,C_{03},C_{12},A^0,A^1,A^2,0,a)$; $C_{03},C_{12},A^1,A^2\in\mathbb{R}$, $A^0\in\mathbb{R}^*$, $a>0$.

21.6. $(0,1,0,1,\sin\chi\sin\theta\cos\phi,\sin\chi\sin\theta\sin\phi,A^0,\sin\chi\cos\theta,\cos\chi,A^0,a)$;

$A^0\in\mathbb{R}$, $0\leqslant\phi<2\pi$, $0\leqslant\theta\leqslant\pi$, $0\leqslant\chi\leqslant\pi$, $a>0$.

21.7. $(0,1,0,1,0,0,A^0,0,0,A^0,1)$; $A^0\in\mathbb{R}$.

21.8. $(0,1,0,1,C_{03},C_{12},\epsilon,A^1,A^2,-\epsilon,a)$; $C_{03},C_{12},A^1,A^2\in\mathbb{R}$, $\epsilon^2=1$, $a>0$.

21.9. $(0,0,0,0,C_{03},C_{12},0,0,A^2,1,a)$; $C_{03},C_{12}\in\mathbb{R}$, $a\in\mathbb{R}^*$, $A^2>0$.

21.10. $(0,0,0,0,C_{03},C_{12},1,0,A^2,0,a)$; $C_{03},C_{12}\in\mathbb{R}$, $a\in\mathbb{R}^*$, $A^2>0$.

21.11. $(0,0,0,0,\sin\theta\cos\phi,\sin\theta\sin\phi,1,0,\cos\theta,\epsilon,a)$; $a\in\mathbb{R}^*$, $0\leqslant\phi<2\pi$, $0\leqslant\theta\leqslant\pi/2$.

21.12. $(0,0,0,0,C_{03},C_{12},0,0,A^2,0,1)$; $C_{03},C_{12}\in\mathbb{R}$, $A^2>0$.

APPENDIX E: VALUES OF $(C_{\alpha\beta})$, (A^δ), AND b FOR CASE 22 OF TABLE I

The values taken by the parameters $(C_{01},C_{02},C_{13},C_{23},C_{03},C_{12},A^0,A^1,A^2,A^3,b)$ for case 22 of Table II are

22.1. $(0,0,0,1,C_{03},C_{12},A^0,A^1,A^2,A^3,b)$; $C_{03},C_{12},A^\delta\in\mathbb{R}$, $b>0$.

22.2. $(0,1,0,0,C_{03},C_{12},A^0,A^1,A^2,A^3,b)$; $C_{03},C_{12},A^\delta\in\mathbb{R}$, $b>0$.

22.3. $(0,1,C_{13},0,C_{03},C_{12},A^0,A^1,A^2,A^3,b)$; $C_{13},C_{03},C_{12},A^\delta\in\mathbb{R}$, $b>0$.

22.4. $(0,1,0,1,C_{03},C_{12},0,A^1,A^2,A^3,b)$; $C_{03},C_{12},A^1,A^2\in\mathbf{R}$, $A^3\in\mathbf{R}^*$, $b>0$.

22.5. $(0,1,0,1,C_{03},C_{12}A^0,A^1,A^2,0,b)$; $C_{03},C_{12},A^1,A^2\in\mathbf{R}$, $A^0\in\mathbf{R}^*$, $b>0$.

22.6. $(0,1,0,1,\sin\chi\sin\theta\cos\phi,\sin\chi\sin\theta\sin\phi,A^0,\sin\chi\cos\theta,\cos\chi,A^0,b)$;
$0\leqslant\phi<2\pi$, $0\leqslant\theta\leqslant\pi$, $0\leqslant\chi\leqslant\pi$, $A^0\in\mathbf{R}$, $b>0$.

22.7. $(0,1,0,1,0,0,A^0,0,0,A^0,1)$; $A^0\in\mathbf{R}$.

22.8. $(0,1,0,1,C_{03},C_{12},1,A^1,A^2,-1,b)$; $b\in\mathbf{R}^*$, $C_{03},C_{12},A^1,A^2\in\mathbf{R}$.

22.9. $(0,0,0,0,C_{03},C_{12},0,0,A^2,1,b)$; $b\in\mathbf{R}^*$, $C_{03},C_{12}\in\mathbf{R}$, $A^2\geqslant 0$.

22.10. $(0,0,0,0,C_{03},C_{12},1,0,A^2,0,b)$; $b\in\mathbf{R}^*$, $C_{03},C_{12}\in\mathbf{R}$, $A^2\geqslant 0$.

22.11. $(0,0,0,0,\sin\theta\cos\phi,\sin\theta\sin\phi,1,0,\cos\theta,\epsilon,b)$; $b\in\mathbf{R}^*$, $0\leqslant\phi<2\pi$, $0\leqslant\theta<\pi/2$, $\epsilon^2=1$.

22.12. $(0,0,0,0,0,0,1,0,0,1,\epsilon)$; $\epsilon^2=1$.

APPENDIX F: VALUES OF $(C_{\alpha\beta})$ AND (A^b) FOR THE FIVE-DIMENSIONAL SUBSPACES OF $\mathscr{P}$

The values taken by the parameters $(C_{01},C_{02},C_{13},C_{23},C_{03},C_{12},A^0,A^1,A^2,A^3)$ for the equivalence classes of five-dimensional (d) subspaces of $\mathscr{P}$ are

5d.1. $(0,0,0,0,\sin\phi,\cos\phi,0,0,A^2,A^3)$; $A^3\in\mathbf{R}^*$, $A^2\geqslant 0$.

5d.2. $(0,0,0,0,\sin\phi,\cos\phi,A^0,0,A^2,0)$; $A^0\in\mathbf{R}^*$, $A^2\geqslant 0$.

5d.3. $(0,0,0,0,\sin\phi,\cos\phi,\epsilon,0,A^2,\epsilon')$; $A^2\geqslant 0$, $\epsilon^2=1=\epsilon'^2$.

5d.4. $(0,0,0,0,\sin\phi,\cos\phi,0,0,A^2,0)$; $A^2\geqslant 0$.

5d.5. $(1,0,-1,0,0,0,0,0,0,A^3)$; $A^3>0$.

5d.6. $(1,0,-1,0,0,0,1,0,0,1)$.

5d.7. $(1,0,-1,0,0,0,0,\sin\phi,\cos\phi,0)$; $0\leqslant\phi<2\pi$.

5d.8. $(1,0,-1,0,0,A^0,0,0,-A^0)$; $A^0\geqslant 0$.

5d.9. $(0,0,0,0,0,0,0,0,0,0,1)$.

5d.10. $(0,0,0,0,0,0,0,1,0,0,0)$.

5d.11. $(0,0,0,0,0,0,0,1,0,0,1)$.

[1]L. P. Eisenhart, Ann. Math. **35**, 284 (1934).
[2]N. M. J. Woodhouse, Commun. Math. Phys. **44**, 9 (1975).
[3]E. G. Kalnins and W. Miller, Jr., SIAM J. Math. Anal. **14**, 126 (1983).
[4]P. A. M. Dirac, *The Principles of Quantum Mechanics* (Oxford U.P., London, 1958).
[5]S. Chandrasekhar, Proc. R. Soc. London Ser. A **349**, 571 (1976).
[6]B. Carter and R. G. McLenaghan, Phys. Rev. D **19**, 1093 (1979).
[7]N. Kamran and R. G. McLenaghan, J. Math. Phys. **25**, 1019 (1984).
[8]N. Kamran and R. G. McLenaghan, Bull. Cl. Sci. Acad. Belg. **LXX**, 596 (1984).
[9]E. G. Kalnins, W. Miller, Jr., and G. C. Williams, J. Math. Phys. **27**, 1893 (1986).
[10]A. H. Cook, Proc. R. Soc. London Ser. A **383**, 247 (1982).
[11]J. Patera, R.T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **17**, 977 (1976).
[12]R. G. McLenaghan and P. Spindel, Bull. Soc. Math. Belg. **XXXI**, 65 (1979).
[13]K. Yano and S. Bochner, *Curvature and Betti Numbers, Annals of Mathematics Studies*, No. 32 (Princeton U.P., Princeton, 1953).

The semiclassical limit of quantum dynamics. I. Time evolution

Sam L. Robinson[a)]
Department of Mathematics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

(Received 3 March 1987; accepted for publication 23 September 1987)

The $\hbar \to 0$ limit of the quantum dynamics determined by the Hamiltonian $H(\hbar) = -(\hbar^2/2m)\Delta + V$ on $L^2(\mathbb{R}^n)$ is studied for a large class of potentials. By convolving with certain Gaussian states, classically determined asymptotic behavior of the quantum evolution of states of compact support is obtained. For initial states of class C_0^1 the error terms are shown to have L^2 norms of order $\hbar^{1/2-\epsilon}$ for arbitrarily small positive ϵ.

I. INTRODUCTION

Since the origin of quantum mechanics more than 60 years ago there has been much effort applied to the understanding of the relationship between the theory and its classical counterpart. Not only is such an understanding important from a purely theoretical viewpoint, but the mathematical techniques developed in order to study this question have provided physicists and chemists with useful and powerful computational tools. In this paper we study the relation between the classical and quantal descriptions of the dynamical evolution in the so-called semiclassical limit. We obtain results for the semiclassical time evolution, which, although basically being known from the work of Maslov[1] and Maslov–Fedoriuk,[2] are proved by new methods that we extend in a forthcoming paper[3] to the scattering theory. Our proofs rely heavily on the results of Hagedorn[4–6] concerning the semiclassical behavior of certain Gaussian initial states.

We now introduce the assumptions on the potential V and state our main result. Our main theorem is concerned with the quantum evolution of certain initial states of compact support.

Assumption 1.1: We assume that $V \in C^{l+2}(\mathbb{R}^n,\mathbb{R})$ for some integer $l \geqslant 1 + n/2$ and that there exist positive constants m_1 and m_2 such that $|V(x)| \leqslant e^{m_1|x|^2}$ and $V(x) \geqslant -m_2$ for all $x \in \mathbb{R}^n$.

Under this assumption the quantum Hamiltonian $H(\hbar) = -(\hbar^2/2m)\Delta + V$ is essentially self-adjoint on the infinitely differentiable functions of compact support in $L^2(\mathbb{R}^n)$. The corresponding classical Hamiltonian is given by the function $H_{\rm cl}(a,\eta) = (1/2m)|\eta|^2 + V(a)$ on $\mathbb{R}^{2n}$. For any initial condition (a_0,η_0) in $\mathbb{R}^{2n}$ the system

$$\frac{\partial}{\partial t} a(t) = \frac{1}{m}\eta(t), \tag{1.1a}$$

$$\frac{\partial}{\partial t} \eta(t) = -\nabla V(a(t)) \tag{1.1b}$$

has a unique solution $(a(a_0,\eta_0,t), \eta(a_0,\eta_0,t))$ such that

$$(a(0),\eta(0)) = (a(a_0,\eta_0,0),\eta(a_0,\eta_0,0)) = (a_0,\eta_0).$$

The solution $(a(a_0,\eta_0,t),\eta(a_0,\eta_0,t))$ is bounded for all $t \in [0,T]$ for any finite $T \geqslant 0$ and is of class C^{l+1} in the initial condition (a_0,η_0) (see, e.g., Chaps. I and II of Lefschetz[7]).

Theorem 1.2: Suppose V satisfies Assumption 1.1. Let $S_0 \in C^3(\mathbb{R}^n,\mathbb{R})$, $f \in C_0^1(\mathbb{R}^n,\mathbb{C})$, and $\lambda \in (0,\frac{1}{2})$. Define $Q(\cdot,t)$: $\mathbb{R}^n \mapsto \mathbb{R}^n$ by

$$Q(q_0,t) = a(q_0,\nabla S_0(q_0),t)$$

and assume that $t \geqslant 0$ is fixed and such that $\det[(\partial Q/\partial q_0)(x,t)] \neq 0$ for all $x \in \mathrm{supp}(f)$. Then there exists $\delta > 0$ and a constant C independent of $\hbar$ such that

$$\left\| e^{-itH(\hbar)/\hbar}(e^{iS_0/\hbar} f) - \sum_j e^{i(\mu_j + S_j/\hbar)} \left| \det\left[\frac{\partial Q}{\partial q_0}(x_j,t)\right]\right|^{-1/2} f(x_j) \right\| \leqslant C\hbar^{\lambda},$$

for all $\hbar \leqslant \delta$. For fixed x, the summation is over all j such that $Q(x_j,t) = x$ and hence is necessarily finite. Here $S_j(x)$ denotes the action

$$S(x_j) = S_0(x_j) + \int_0^t \left[\frac{1}{2m}|\eta(x_j,\nabla S_0(x_j),\tau)|^2 - V(a(x_j,\nabla S_0(x_j),\tau))\right] d\tau,$$

and μ_j is an integer multiple of $\pi/2$ defined explicitly in the proof.

Remarks: (1) The constants C and δ are, in general, time dependent. The dependence of the rightmost function inside the norm [which is, of course, the norm of $L^2(\mathbb{R}^n, d^n x)$] on x is given implicitly by $x = Q(x_j,t)$.

(2) The number μ_j is related to the Keller–Maslov index[1,8] of the path

$$\{(a(x_j,\nabla S_0(x_j),\tau),\eta(x_j,\nabla S_0(x_j),\tau)): 0 \leqslant \tau \leqslant t\}$$

and arises naturally in our proof in connection with the branch of the square root of $\det[(\partial Q/\partial q_0)(x,t)]$.

(3) The theorem is not a new result. Indeed, one can view our theorem as a slight generalization of the leading-order term in Theorem 12.3 of Maslov–Fedoriuk.[2] However, we prove our theorem by a different method and our proof extends to the scattering theory.

II. NOTATION AND DEFINITIONS

Throughout this paper n denotes the space dimension, and $L^2(\mathbb{R}^n)$ is the Hilbert space of square integrable com-

[a)] Present address: Department of Mathematical Sciences, University of North Carolina, Wilmington, North Carolina 28403.

 0022-2488/88/020412-08$02.50

plex valued functions on $\mathbb{R}^n$ with inner product $\langle \cdot,\cdot \rangle$ and norm $\|\cdot\|$ is given by

$$\langle f,g\rangle = \int_{\mathbb{R}^n} \bar{f}(x)g(x)d^n x, \quad \|f\| = \langle f,f\rangle^{1/2}.$$

For Ω an open subset of $\mathbb{R}^n$, $\mathbb{E} = \mathbb{R}$ or $\mathbb{C}$, and K a non-negative integer, $C^k(\Omega,\mathbb{E})$ denotes the linear space of k-times continuously differentiable functions mapping Ω into $\mathbb{E}$. Here $C_0^k(\Omega,\mathbb{E})$ is the subspace of functions in $C^k(\Omega,\mathbb{E})$ with support compact and contained in Ω. We will denote the support of a function f by $\text{supp}(f)$. The quantum mechanical Hamiltonian $H(\hbar)$ is the operator $-(\hbar^2/2m)\Delta + V$ on $L^2(\mathbb{R}^n)$. Here, Δ is the n-dimensional Laplacian operator,

$$\Delta = \left(\frac{\partial}{\partial x_1}\right)^2 + \cdots + \left(\frac{\partial}{\partial x_n}\right)^2,$$

$\hbar$ is a small positive parameter (a dimensionless multiple of Planck's constant), m is a positive constant, and V is a real valued function on $\mathbb{R}^n$ viewed here as a multiplication operator on $L^2(\mathbb{R}^n)$. For potentials V satisfying Assumption 1.1 the operator $H(\hbar)$ is essentially self-adjoint on the domain $C_0^\infty(\mathbb{R}^n,\mathbb{C})$.

A multi-index α is an ordered n-tuple $(\alpha_1,\alpha_2,...,\alpha_n)$ of non-negative integers. The order of a multi-index α is given by $|\alpha| = \Sigma_{i=1}^n \alpha_i$ and the factorial of α by

$$\alpha! = \prod_{i=1}^n (\alpha_i!).$$

For $x = (x_1,x_2,...,x_n) \in \mathbb{R}^n$ the symbol x^α is defined by

$$x^\alpha = \prod_{i=1}^n x_i^{\alpha_i}.$$

Here D^α stands for the partial differential operator

$$D^\alpha = \left(\frac{\partial}{\partial x}\right)^\alpha = \frac{\partial^{|\alpha|}}{(\partial x_1)^{\alpha_1}(\partial x_2)^{\alpha_2}\cdots(\partial x_n)^{\alpha_n}}.$$

For $x \in \mathbb{R}^n$ or $\mathbb{C}^n$, $|x|$ denotes the Euclidean norm of x. We denote the usual inner product on $\mathbb{R}^n$ or $\mathbb{C}^n$ by

$$\langle u,v\rangle = \sum_{i=1}^n \bar{u}_i v_i$$

and let $\{e_i\}_{i=1}^n$ be the standard basis for $\mathbb{R}^n$ or $\mathbb{C}^n$. If $f \in C^1(\mathbb{R}^n,\mathbb{E})$, where $\mathbb{E}$ is $\mathbb{R}^n$ or $\mathbb{C}^n$, $\partial f/\partial x$ is the matrix $(\partial f_i/\partial x_j)$. For $f \in C^1(\mathbb{R}^n,\mathbb{R})$ we will usually write $f^{(1)}$ instead of ∇f to denote the gradient of f and if $f \in C^2(\mathbb{R}^n,\mathbb{R})$ we will write $f^{(2)}$ to denote the Hessian matrix $(\partial^2 f/\partial x_i\,\partial x_j)$. We will not distinguish row and column vectors in $\mathbb{R}^n$ or $\mathbb{C}^n$ in our formulas and hence matrix products must be interpreted in context. The symbol $\mathbb{1}$ will stand for the $n \times n$ identity matrix. For an $n \times n$ complex matrix A we will use the symbol $|A|$ to denote the matrix $(AA^*)^{1/2}$, where A^* is the adjoint (complex conjugate transpose) of A. The symbol $\mathcal{U}_A$ will denote the unique unitary matrix guaranteed by the polar decomposition theorem such that $A = |A|\mathcal{U}_A$, $|A| = (AA^*)^{1/2}$.

Following Hagedorn[6] we define generalized Hermite polynomials on $\mathbb{R}^n$ recursively as follows: We set $\tilde{\mathcal{H}}_0(x) = 1$ and $\tilde{\mathcal{H}}_1(v;x) = 2\langle v,x\rangle$, where v is an arbitrary nonzero vector in $\mathbb{C}^n$. For $v_1,...,v_m$ arbitrary nonzero vectors in $\mathbb{C}^n$ we set

$$\begin{aligned}\tilde{\mathcal{H}}_m(v_1,...,v_m;x) &= 2\langle v_m,x\rangle \tilde{\mathcal{H}}_{m-1}(v_1,...,v_{m-1};x) \\ &\quad - 2\sum_{i=1}^{m-1} \langle v_m,\bar{v}_i\rangle \\ &\quad \times \tilde{\mathcal{H}}_{m-2}(v_1,...,v_{i-1},v_{i-2},...,v_{m-1};x).\end{aligned}$$

The polynomials $\tilde{\mathcal{H}}_m$ are independent of the ordering of the vectors $v_1,...,v_m$. Given a complex invertible $n \times n$ matrix A and a multi-index α we define the polynomial

$$\begin{aligned}\mathcal{H}_\alpha(A;x) = \tilde{\mathcal{H}}_{|\alpha|}(&\mathcal{U}_A e_1,...,\mathcal{U}_A e_1, \\ &\mathcal{U}_A e_2,...,\mathcal{U}_A e_n;x),\end{aligned}$$

where the vector $\mathcal{U}_A e_i$ appears α_i times in the list of variables of $\tilde{\mathcal{H}}_{|\alpha|}$.

We will find it useful to consider complex $n \times n$ matrices A and B satisfying the following conditions:

$$A \text{ and } B \text{ are invertible,} \tag{2.1a}$$

$$BA^{-1} \text{ is symmetric,} \tag{2.1b}$$

$$\text{Re}(BA^{-1}) \text{ is strictly positive definite,} \tag{2.1c}$$

$$(\text{Re}(BA^{-1}))^{-1} = AA^*. \tag{2.1d}$$

[Here, symmetric means (real symmetric) + i(real symmetric).]

For complex $n \times n$ matrices A and B satisfying conditions (2.1), vectors a and $\eta \in \mathbb{R}^n$, multi-indices α, and positive $\hbar$ we define

$$\begin{aligned}\phi_\alpha(A,B,\hbar,a,\eta,x) &= (\pi\hbar)^{-n/4}(2^{|\alpha|}\alpha!)^{-1/2}[\det(A)]^{-1/2} \\ &\quad \times \mathcal{H}_\alpha(A;\hbar^{-1/2}|A|^{-1}(x-a)) \\ &\quad \times \exp\{-(1/2\hbar)\langle (x-a),BA^{-1}(x-a)\rangle \\ &\quad + (i/\hbar)\langle \eta,(x-a)\rangle\}.\end{aligned}$$

Here $|A|$ is the matrix $(AA^*)^{1/2}$ and the branch of the square root of $\det(A)$ will be specified in the context in which the functions ϕ_α are used. Whenever we write $\phi_\alpha(A,B,\hbar,a,\eta,x)$ we are assuming that the matrices A and B satisfy conditions (2.1). For fixed A, B, $\hbar$, a, and η the functions $\phi_\alpha(A,B,\hbar,a,\eta,x)$ form an orthonormal basis of $L^2(\mathbb{R}^n)$.

III. SOME PRELIMINARY LEMMAS

In this section we prove two rather technical lemmas on the small $\hbar$ asymptotics of certain integrals of a type we encounter frequently in Sec. IV. The reader may skip the roofs of these lemmas as the details are not needed in the sequel.

Lemma 3.1: Let Ω be an open subset of $\mathbb{R}^n$. Let $S \in C^3(\Omega,\mathbb{R})$ and let $g \in C_0^1(\mathbb{R}^n,\mathbb{C})$ be such that $\text{supp}(g) \subset \Omega$. Suppose T is a complex $n \times n$ matrix valued class C^1 function on Ω satisfying

(1) $T(x)$ is symmetric [(real symmetric) + i(real symmetric)],

(2) $\text{Re}(T(x))$ is strictly positive definite,

for all $x \in \Omega$. Define the square root of $\det[T(x) + iS^{(2)}(x)]$ for $x \in \Omega$ by analytic continuation along $\xi \in [0,1]$ of

$$(\det[\text{Re}(T(x)) + \xi i(\text{Im}(T(x)) + S^{(2)}(x))])^{1/2}$$

starting with a positive value for $\xi = 0$. For $\hbar > 0$ and $x \in \mathbb{R}^n$, let

$$F(\hbar,x) = \hbar^{-n/2} \int_\Omega g(q) \times \exp\left(-\frac{1}{2\hbar} \langle (x-q), T(q)(x-q) \rangle + \frac{i}{\hbar} [S(q) + \langle S^{(1)}(q), (x-q) \rangle] \right) d^n q$$

and

$$G(\hbar,x) = (2\pi)^{n/2} \times (\det[T(x) + iS^{(2)}(x)])^{-1/2} g(x) e^{iS(x)/\hbar}.$$

Then, given $\lambda \in (0, \frac{1}{2})$ there exists $\delta > 0$ and a constant C independent of $\hbar < \delta$ such that

$$\|F(\hbar,\cdot) - G(\hbar,\cdot)\| < C\hbar^\lambda,$$

for all $\hbar \in (0,\delta)$.

Proof: We first note that the determinant of $\mathrm{Re}(T(x)) + \xi i(\mathrm{Im}(T(x)) + S^{(2)}(x))$ is nonzero for all $\xi \in [0,1]$ by virtue of the fact that $\mathrm{Re}(T(x))$ is strictly positive definite and $\mathrm{Im}(T(x)) + S^{(2)}(x)$ is real symmetric. We next reduce to a special case by noting that there exist finitely many open balls Ω_k such that

$$\mathrm{supp}(g) \subset \bigcup_{k=1}^K \Omega_k \subset \Omega.$$

By introducing a C^1 partition of unity $\{h_k\}_{k=1}^K$ satisfying $\mathrm{supp}(h_k) \subset \Omega_k$ and $\Sigma_{k=1}^K h_k = 1$ on $\mathrm{supp}(g)$ we see by the triangle inequality that it suffices to prove the lemma only for the special case in which Ω is an open ball. Assuming this, one can easily show the existence of a constant $\epsilon > 0$ and a closed ball $\mathscr{K} \subset \mathbb{R}^n$ such that

$$\mathrm{supp}(g) \subset \{q \in \mathbb{R}^n : \mathrm{dist}(q, \mathrm{supp}(g)) < \epsilon\} \subset \mathscr{K} \subset \Omega.$$

We set $\gamma = \frac{1}{3}(1+\lambda)$, $\delta = \epsilon^{1/\gamma}$, and choose T_0 such that

$$0 < T_0 < \inf_{q \in \mathscr{K}} \inf_{\|x\|=1} \langle x, T(q)x \rangle.$$

By the choice of T_0, $\mathrm{Re}(T(q)) > T_0 \mathbf{1}$, for all $q \in \mathscr{K}$.

We note that there exists a constant C_γ independent of $\hbar \in (0,\delta)$ such that

$$\int_{|x| > \hbar^\gamma} \exp\left(-\frac{1}{2\hbar} T_0 |x|^2 \right) d^n x < C_\gamma \hbar^{(n+2)/2}.$$

The hypotheses on S, T, and g along with Taylor's theorem imply the existence of constants C_1, C_2, and C_3 independent of $\hbar \in (0,\delta)$ such that

$$|S(q) + \langle S^{(1)}(q), (x-q) \rangle - S(x) + \tfrac{1}{2} \langle (x-q), S^{(2)}(x)(x-q) \rangle| \leqslant C_1 |x-q|^3,$$

for all $x,q \in \mathscr{K}$;

$$|\exp\{-(1/2\hbar)\langle (x-q), T(q)(x-q)\rangle\} - \exp\{-(1/2\hbar)\langle (x-q), T(x)(x-q)\rangle\}| \leqslant C_2 \hbar^{3\gamma - 1} \exp\{-(1/2\hbar) T_0 |x-q|^2\},$$

for all $x,q \in \mathscr{K}$ with $|x-q| \leqslant \hbar^2$; and

$$\|g(x-y) - g(x)\|_{L^2(\mathbb{R}^n, d^n x)} \leqslant C_3 |y|,$$

for all $y \in \mathbb{R}^n$.

For $x \in \mathbb{R}^n$ and $\hbar > 0$, define

$$F_1(\hbar,x) = \hbar^{-n/2} g(x) \times \int_{\mathscr{K}} \exp\left(-\frac{1}{2\hbar} \langle (x-q), T(q)(x-q) \rangle + \frac{i}{\hbar} [S(q) + \langle S^{(1)}(q), (x-q) \rangle] \right) d^n q,$$

$$F_2(\hbar,x) = \hbar^{-n/2} g(x) \times \int_{\mathscr{K}} \exp\left(-\frac{1}{2\hbar} \langle (x-q), T(q)(x-q) \rangle + \frac{i}{\hbar} \left[S(x) - \frac{1}{2} \langle (x-q), S^{(2)}(x)(x-q) \rangle \right] \right) \times d^n q,$$

and

$$F_3(\hbar,x) = \hbar^{-n/2} g(x) \times \int_{\mathscr{K}} \exp\left(-\frac{1}{2\hbar} \langle (x-q), T(x)(x-q) \rangle + \frac{i}{\hbar} \left[S(x) - \frac{1}{2} \langle (x-q), S^{(2)}(x)(x-q) \rangle \right] \right) \times d^n q.$$

Then,

$$|F(\hbar,x) - F_1(\hbar,x)| \leqslant \hbar^{-n/2} \int_{\mathscr{K}} |g(q) - g(x)| \times \exp\left\{ -\frac{1}{2\hbar} T_0 |x-q|^2 \right\} d^n q \leqslant \hbar^{-n/2} \int_{\mathbb{R}^n} |g(x-y) - g(x)| \times \exp\left\{ -\frac{1}{2\hbar} T_0 |y|^2 \right\} d^n y,$$

thus

$$\|F(\hbar,\cdot) - F_1(\hbar,\cdot)\| \leqslant \hbar^{-n/2} \int_{\mathbb{R}^n} \|g(x-y) - g(x)\|_{L^2(\mathbb{R}^n, d^n x)} \times \exp\left\{ -\frac{1}{2\hbar} T_0 |y|^2 \right\} d^n y \leqslant \hbar^{-n/2} C_3 \int_{\mathbb{R}^n} |y| \exp\left\{ -\frac{1}{2\hbar} T_0 |y|^2 \right\} d^n y = C_3' \hbar^{1/2},$$

where C_3' is independent of $\hbar$. Similarly,

$$|F_1(\hbar,x) - F_2(\hbar,x)| \leqslant \hbar^{-n/2} |g(x)| \int_{\mathscr{K}} \exp\left\{ -\frac{1}{2\hbar} T_0 |x-q|^2 \right\}$$

$$\times\left|\exp\left\{\frac{i}{\hbar}\left[S(q)+\langle S^{(1)}(q),(x-q)\rangle\right]\right\}-\exp\left\{\frac{i}{\hbar}\left[S(x)-\frac{1}{2}\langle(x-q),S^{(2)}(x)(x-q)\rangle\right]\right\}\right| d^nq,$$

hence

$$\begin{aligned}
&\|F_1(\hbar,\cdot)-F_2(\hbar,\cdot)\| \\
&\quad\leqslant\hbar^{-n/2}\|g\|\sup_{x\in\mathrm{supp}(g)}\left[\int_{\mathscr{K}}\exp\left\{-\frac{1}{2\hbar}T_0|x-q|^2\right\}\left|\exp\left\{\frac{i}{\hbar}\left[S(q)+\langle S^{(1)}(q),(x-q)\rangle\right]\right\}\right.\right. \\
&\qquad\left.\left.-\exp\left\{\frac{i}{\hbar}\left[S(x)-\frac{1}{2}\langle(x-q),S^{(2)}(x)(x-q)\rangle\right]\right\}\right| d^nq\right] \\
&\quad\leqslant\hbar^{-n/2}\|g\|\sup_{x\in\mathrm{supp}(g)}\left[\hbar^{-1}\int_{\mathscr{K}}\exp\left\{-\frac{1}{2\hbar}T_0|x-q|^2\right\}\left(\left|S(q)+\langle S^{(1)}(q),(x-q)\rangle\right.\right.\right. \\
&\qquad\left.\left.\left.-S(x)+\frac{1}{2}\langle(x-q),S^{(2)}(x)(x-q)\rangle\right|\right) d^nq\right] \\
&\quad\leqslant\hbar^{-(n+2)/2}\|g\|C_1\int_{\mathbf{R}^n}|y|^3\exp\left\{-\frac{1}{2\hbar}T_0|y|^2\right\} d^ny=C_1'\hbar^{1/2},
\end{aligned}$$

where C_1' is independent of $\hbar$. Moreover,

$$\begin{aligned}
&\|F_2(\hbar,\cdot)-F_3(\hbar,\cdot)\| \\
&\quad\leqslant\hbar^{-n/2}\|g\|\sup_{x\in\mathrm{supp}(g)}\left[\int_{\mathscr{K}}\left|\exp\left\{-\frac{1}{2\hbar}\langle(x-q),T(q)(x-q)\rangle\right\}-\exp\left\{-\frac{1}{2\hbar}\langle(x-q),T(x)(x-q)\rangle\right\}\right| d^nq\right] \\
&\quad\leqslant\hbar^{-n/2}\|g\|\sup_{x\in\mathrm{supp}(g)}\left[C_2\hbar^{3\gamma-1}\int_{|x-q|<\hbar^{\gamma}}\exp\left\{-\frac{1}{2\hbar}T_0|x-q|^2\right\} d^nq+2\int_{|x-q|\geqslant\hbar^{\gamma}}\exp\left\{-\frac{1}{2\hbar}T_0|x-q|^2\right\} d^nq\right] \\
&\quad\leqslant\|g\|(C_2\hbar^{\lambda}+2C_{\gamma}\hbar).
\end{aligned}$$

Now we observe that

$$G(\hbar,x)=\hbar^{-n/2}g(x)\int_{\mathbf{R}^n}\exp\left(-\frac{1}{2\hbar}\langle(x-q),(T(x)+iS^{(2)}(x))(x-q)\rangle+\frac{i}{\hbar}S(x)\right) d^nq$$

(see, e.g., Theorem B, Sec. 1a of Bargmann[9]). Hence

$$\begin{aligned}
|G(\hbar,x)-F_3(\hbar,x)|&\leqslant\hbar^{-n/2}|g(x)|\int_{\mathbf{R}^n/\mathscr{K}}\exp\left(-\frac{1}{2\hbar}\langle(x-q),\mathrm{Re}(T(x))(x-q)\rangle\right) d^nq \\
&\leqslant\hbar^{-n/2}|g(x)|\int_{|x-q|\geqslant\hbar^{\gamma}}\exp\left(-\frac{1}{2\hbar}T_0|x-q|^2\right) d^nq\leqslant C_{\gamma}\hbar|g(x)|
\end{aligned}$$

and therefore

$$\|G(\hbar,\cdot)-F_3(\hbar,\cdot)\|\leqslant C_{\gamma}\hbar\|g\|.$$

The triangle inequality completes the proof of the lemma. ■

Lemma 3.2: Let Ω, S, g, and T be as in Lemma 3.1. Suppose $c(\hbar,\cdot)\in C(\Omega,\mathbb{C})$ and $P\in C(\Omega\times\mathbf{R}^n,\mathbb{C})$ are such that there exist $r,\beta\geqslant 0$ and constants C_r and C_β such that

$$|c(\hbar,q)|\leqslant C_r\hbar^{r/2},\quad\text{for all } q\in\Omega$$

and

$$|P(q,x)|\leqslant C_\beta|x-q|^{\beta},\quad\text{for all } q\in\Omega,\quad x\in\mathbf{R}^n.$$

Let

$$\begin{aligned}
F(\hbar,x)=\hbar^{-n/2}&\int_{\Omega}g(q)c(\hbar,q)P(q,x) \\
&\times\exp\left(-\frac{1}{2\hbar}\langle(x-q),T(q)(x-q)\rangle\right. \\
&\left.+\frac{i}{\hbar}\left[S(q)+\langle S^{(1)}(q),(x-q)\rangle\right]\right) d^nq.
\end{aligned}$$

Then, there exists $\delta>0$ and a constant C independent of $\hbar$ such that

$$\|F(\hbar,\cdot)\|\leqslant C\hbar^{(r+\beta)/2},$$

for all $\hbar\leqslant\delta$.

Proof: Choose T_0 as in the proof of Lemma 3.1 and let C_1 be such that

$$\|g(x-y)-g(x)\|_{L^2(\mathbf{R}^n,d^nx)}\leqslant C_1|y|,$$

for all $y\in\mathbf{R}^n$. Define

$$\begin{aligned}
F_1(\hbar,x)=\hbar^{-n/2}&g(x) \\
&\times\int_{\Omega}c(\hbar,q)P(q,x) \\
&\times\exp\left(-\frac{1}{2\hbar}\langle(x-q),T(q)(x-q)\rangle\right.
\end{aligned}$$

$$+\frac{i}{\hbar}\left[S(q)+\langle S^{(1)}(q),(x-q)\rangle\right]\Big)d^nq$$

and

$$F_2(\hbar,x)=\hbar^{-n/2}g(x)c(\hbar,x)P(x,x)\times\int_\Omega\exp\Big(-\frac{1}{2\hbar}\langle(x-q),T(q)(x-q)\rangle+\frac{i}{\hbar}\left[S(q)+\langle S^{(1)}(q),(x-q)\rangle\right]\Big)d^nq.$$

Note that $P(x,x)$ [and hence $F_2(\hbar,x)$] is zero if $\beta\neq 0$. Then,

$$|F(\hbar,x)-F_1(\hbar,x)|\leq C_rC_\beta\hbar^{(r-n)/2}\times\int_{\mathbf{R}^n}|g(x-y)-g(x)|\,|y|^\beta\times\exp\Big(-\frac{1}{2\hbar}T_0|y|^2\Big)d^ny.$$

Thus

$$\|F(\hbar,\cdot)-F_1(\hbar,\cdot)\|\leq C_rC_\beta C_1\hbar^{(r-n)/2}\int_{\mathbf{R}^n}|y|^{\beta+1}\exp\Big(-\frac{1}{2\hbar}T_0|y|^2\Big)d^ny=C_1'\hbar^{(r+\beta+1)/2},$$

where C_1' is independent of $\hbar$. Since

$$|c(\hbar,q)P(q,x)-c(\hbar,x)P(x,x)|\leq 2C_rC_\beta\hbar^{r/2}|x-q|^\beta,$$

for all $q\in\Omega$ we obtain

$$|F_1(\hbar,x)-F_2(\hbar,x)|\leq C_2\hbar^{(r+\beta)/2}|g(x)|,$$

for some C_2 independent of $\hbar$, and hence

$$\|F_1(\hbar,\cdot)-F_2(\hbar,\cdot)\|\leq C_2\hbar^{(r+\beta)/2}\|g\|.$$

The proof is completed by noticing that there is a constant C_3 independent of $\hbar$ such that

$$\|F_2(\hbar,\cdot)\|\leq C_3\hbar^{r/2},$$

if $\beta=0$, and

$$\|F_2(\hbar,\cdot)\|=0,$$

if $\beta\neq 0$.

IV. PROOF OF THE THEOREM

In this section we prove Theorem 1.2. Suppose V satisfies Assumption 1.1 and let $S_0\in C^3(\mathbf{R}^n,\mathbf{R})$. Given $T>0$ and $a_0,\eta_0\in\mathbf{R}^n$, the system of ordinary differential equations,

$$\frac{\partial}{\partial t}a(t)=\frac{1}{m}\eta(t),\tag{4.1a}$$

$$\frac{\partial}{\partial t}\eta(t)=-V^{(1)}(a(t)),\tag{4.1b}$$

$$\frac{\partial}{\partial t}A(t)=\frac{i}{m}B(t),\tag{4.1c}$$

$$\frac{\partial}{\partial t}B(t)=iV^{(2)}(a(t))A(t),\tag{4.1d}$$

$$\frac{\partial}{\partial t}S(t)=\frac{1}{2m}|\eta(t)|^2-V(a(t)),\tag{4.1e}$$

subject to the initial conditions $a(0)=a_0$, $\eta(0)=\eta_0$, $A(0)=\mathbf{1}$, $B(0)=\mathbf{1}$, and $S(0)=S_0(a_0)$, has a unique bounded solution for $t\in[0,T]$. We denote this solution by $[a(a_0,\eta_0,t),\eta(a_0,\eta_0,t),A(a_0,\eta_0,t),B(a_0,\eta_0,t),S(a_0,\eta_0,t)]$. By considering the system (4.1a) and (4.1b) we find that the functions a and η are of class C^{l+1} in a_0 and η_0. These facts are standard results from the theory of ordinary differential equations.[7] By Theorem 1.1 of Hagedorn[4] the matrices $A(a_0,\eta_0,t)$ and $B(a_0,\eta_0,t)$ satisfy conditions (2.1) for all $t\in[0,T]$ and are given by

$$A(a_0,\eta_0,t)=\frac{\partial}{\partial a_0}a(a_0,\eta_0,t)+i\frac{\partial}{\partial\eta_0}a(a_0,\eta_0,t),\tag{4.2a}$$

$$B(a_0,\eta_0,t)=\frac{\partial}{\partial\eta_0}\eta(a_0,\eta_0,t)-i\frac{\partial}{\partial a_0}\eta(a_0,\eta_0,t).\tag{4.2b}$$

From (4.2) and the remarks above it follows that A and B are of class C^l in a_0 and η_0.

Let Q be the mapping from $\mathbf{R}^n\times(0,T)$ into $\mathbf{R}^n$ defined by

$$Q(q_0,t)=a(q_0,S_0^{(1)}(q_0),t).$$

By the hypothesis on S_0 and the remarks above, Q is of class C^2 in the variable $q_0\in\mathbf{R}^n$. Let $f\in C^1(\mathbf{R}^n,\mathbf{C})$ have compact support. We now fix $t\in(0,T)$ and assume

$$\det\Big[\frac{\partial Q}{\partial q_0}(x,t)\Big]\neq 0,\tag{4.3}$$

for all $x\in\mathrm{supp}(f)$. Since t is fixed, we will omit reference to t where possible but it should be remembered that all estimates obtained in this section are t dependent.

Under the assumption (4.3), for each $x\in\mathrm{supp}(f)$ there exists an open ball $\mathcal{N}_x\ni x$ such that the mapping $Q\restriction\mathcal{N}_x$ is a class C^2 diffeomorphism of $\mathcal{N}_x$ onto $Q[\mathcal{N}_x]$. Since $\mathrm{supp}(f)$ is compact, some finite subcollection $\{\mathcal{N}_k: k=1,\ldots,K\}$ of the family $\{\mathcal{N}_x: x\in\mathrm{supp}(f)\}$ covers $\mathrm{supp}(f)$. We denote by Q_k the diffeomorphism $Q\restriction\mathcal{N}_x$ and define, for $q\in Q_k[\mathcal{N}_k]$,

$$p_k(q)=\eta(Q_k^{-1}(q),S_0^{(1)}(Q_k^{-1}(q)),t),$$
$$A_k(q)=A(Q_k^{-1}(q),S_0^{(1)}(Q_k^{-1}(q)),t),$$
$$B_k(q)=B(Q_k^{-1}(q),S_0^{(1)}(Q_k^{-1}(q)),t),$$
$$S_k(q)=S(Q_k^{-1}(q),S_0^{(1)}(Q_k^{-1}(q)),t).$$

It is well known and not difficult to show that

$$p_k(q)=\frac{\partial}{\partial q}S_k(q).\tag{4.4}$$

Note that the functions p_k, A_k, and B_k are of class C^2 in $q\in Q_k[\mathcal{N}_k]$ while S_k is of class C^3. Moreover, since $A_k(q)A_k(q)^*$ is strictly positive definite, the operational calculus shows that the matrix $|A_k(q)|$ is continuously differentiable with respect to $q\in Q_k[\mathcal{N}_k]$.

Proposition 4.1: Let $q\in Q_k[\mathcal{N}_k]$. Then

$$\det[B_k(q)+iS_k^{(2)}(q)A_k(q)]=\det\Big[\frac{\partial}{\partial q}Q_k^{-1}(q)\Big]\cdot\det[\mathbf{1}+iS_0^{(2)}(Q_k^{-1}(q))].$$

Proof: First note that, by differentiating the expression defining Q,

$$\frac{\partial}{\partial q} Q_k^{-1}(q) = \left[\frac{\partial a}{\partial a_0} + \frac{\partial a}{\partial \eta_0} S_0^{(2)}\right]^{-1},$$

and, differentiating (4.4),

$$S_k^{(2)}(q) = \left[\frac{\partial \eta}{\partial a_0} + \frac{\partial \eta}{\partial \eta_0} S_0^{(2)}\right] \frac{\partial}{\partial q} Q_k^{-1}(q),$$

where functions of a_0 and η_0 are assumed evaluated at $a_0 = Q_k^{-1}(q)$ and $\eta_0 = S_0^{(1)}(Q_k^{-1}(q))$. By (4.2) and partitioning of determinants,[10]

$$\begin{aligned}
\det[B_k(q) + iS_k^{(2)}(q)A_k(q)] &= \det\left[\left(\frac{\partial \eta}{\partial \eta_0} - i\frac{\partial \eta}{\partial a_0}\right) + i\left(\frac{\partial \eta}{\partial a_0} + \frac{\partial \eta}{\partial \eta_0} S_0^{(2)}\right)\left(\frac{\partial}{\partial q} Q_k^{-1}(q)\right)\left(\frac{\partial a}{\partial a_0} + i\frac{\partial a}{\partial \eta_0}\right)\right] \\
&= \det\left[\frac{\partial}{\partial q} Q_k^{-1}(q)\right] \cdot \det\begin{bmatrix} \frac{\partial \eta}{\partial \eta_0} - i\frac{\partial \eta}{\partial a_0} & \frac{\partial \eta}{\partial a_0} + \frac{\partial \eta}{\partial \eta_0} S_0^{(2)} \\ \frac{\partial a}{\partial \eta_0} - i\frac{\partial a}{\partial a_0} & \frac{\partial a}{\partial a_0} + \frac{\partial a}{\partial \eta_0} S_0^{(2)} \end{bmatrix} \\
&= \det\left[\frac{\partial}{\partial q} Q_k^{-1}(q)\right] \cdot \det\begin{bmatrix} \frac{\partial \eta}{\partial \eta_0} & \frac{\partial \eta}{\partial a_0} \\ \frac{\partial a}{\partial \eta_0} & \frac{\partial a}{\partial a_0} \end{bmatrix} \cdot \det\begin{bmatrix} \mathbf{1} & S_0^{(2)} \\ -i\mathbf{1} & \mathbf{1} \end{bmatrix} \\
&= \det\left[\frac{\partial}{\partial q} Q_k^{-1}(q)\right] \cdot \det[\mathbf{1} + iS_0^{(2)}(Q_k^{-1}(q))].
\end{aligned}$$

In the last step we have used the fact that the mapping

$$(a_0, \eta_0) \mapsto (a(a_0, \eta_0, t), \eta(a_0, \eta_0, t))$$

is a canonical transformation.[11] ■

Let $q \in Q_k[\mathcal{N}_k]$. Define the branch of the square root of $\det[A_k(q)]$ by analytic continuation along $\tau \in [0,t]$ of $(\det[A(Q_k^{-1}(q), S_0^{(1)}(Q_k^{-1}(q)), \tau)])^{1/2}$ starting with a value of 1 for $\tau = 0$. We determine the branches of

$$(\det[\mathbf{1} + iS_0^{(2)}(Q_k^{-1}(q))])^{1/2}$$

and

$$(\det[B_k(q)A_k(q)^{-1} + iS_k^{(2)}(q)])^{1/2}$$

by analytic continuation of

$$(\det[\mathbf{1} + \xi iS_0^{(2)}(Q_k^{-1}(q))])^{1/2}$$

and

$$(\det[\mathrm{Re}(B_k(q)A_k(q)^{-1}) + \xi i(\mathrm{Im}(B_k(q)A_k(q)^{-1}) + S_k^{(2)}(q))])^{1/2},$$

respectively, along $\xi \in [0,1]$ starting with positive values for $\xi = 0$. This determines the branch of

$$(\det[B_k(q) + iS_k^{(2)}(q)A_k(q)])^{1/2}$$

and the branch of $(\det[(\partial/\partial q)Q_k^{-1}(q)])^{1/2}$ is then determined by Proposition 4.1. Let $\mu_k(q)$ be the positive integer (mod 4) such that

$$\left(\det\left[\frac{\partial Q}{\partial q_0}(Q_k^{-1}(q))\right]\right)^{1/2} = e^{i(\pi/2)\mu_k(q)} \left|\det\left[\frac{\partial Q}{\partial q_0}(Q_k^{-1}(q))\right]\right|^{1/2}.$$

By continuity and the fact that $Q_k[\mathcal{N}_k]$ is pathwise connected the index $\mu_k(q)$ is independent of the choice of $q \in Q_k[\mathcal{N}_k]$. Hence to each $\mathcal{N}_k$ we can assign a unique (mod 4) index μ_k. Moreover, in light of the facts above we see that if $\mathcal{N}_k \cap \mathcal{N}_j$ is nonempty then $\mu_k = \mu_j$. We now introduce a partition of unity $\{h_k\}_{k=1}^K$ on $\cup_{k=1}^K \mathcal{N}_k$ satisfying $h_k \in C_0^1(\mathbb{R}^n, \mathbb{C})$, $\mathrm{supp}(h_k) \subset \mathcal{N}_k$, $\Sigma_{k=1}^K h_k = 1$ on $\mathrm{supp}(f)$ and set $f_k = fh_k$. The following proposition allows us to restrict our attention to a single $\mathcal{N}_k$ and f_k.

Proposition 4.2: Let $x \in \mathbb{R}^n$ be fixed. Let $\{x_j\}_{j=1}^J$ be the set of points in $\cup_{k=1}^K \mathcal{N}_k$ such that $Q(x_j) = x$ for $j = 1,2,\ldots,J$ and let μ_j' be defined to be equal to μ_k if $x_j \in \mathcal{N}_k$. Define

$$\begin{aligned}
S_j'(x) &= S_0(x_j) \\
&\quad + \int_0^t \left[\frac{1}{2m} |\eta(x_j, S_0^{(1)}(x_j), \tau)|^2 \right. \\
&\quad \left. - V(a(x_j, S_0^{(1)}(x_j), \tau))\right] d\tau
\end{aligned}$$

and let χ_k denote the characteristic function of $Q_k[\mathcal{N}_k]$. Then

$$\begin{aligned}
&\sum_{k=1}^K e^{iS_k(x)/\hbar + \mu_k \pi/2} \left|\det\left[\frac{\partial Q}{\partial q_0}(Q_k^{-1}(x))\right]\right|^{-1/2} \\
&\quad \times f_k(Q_k^{-1}(x))\chi_k(x) \\
&= \sum_{j=1}^J e^{iS_j'(x)/\hbar + i\mu_j' \pi/2} \left|\det\left[\frac{\partial Q}{\partial q_0}(x_j)\right]\right|^{-1/2} f(x_j).
\end{aligned}$$

Proof: We note that μ_j' is well-defined by the remarks preceding the proposition. Moreover, if $x_j \in \mathcal{N}_{k_1} \cap \mathcal{N}_{k_2}$, then $S_{k_1}(x) = S_{k_2}(x) = S_j'(x)$. The proposition follows by changing variables from $Q_k^{-1}(x)$ to x_j in the first summation. ■

The next lemma is a version of Theorem 1.2 for the func-

tions f_k. The proof requires the functions $\phi_\alpha(A,B,\hbar,a,\eta,x)$ defined in Sec. II. This lemma, Proposition 4.2, and the triangle inequality complete the proof of Theorem 1.2.

Lemma 4.3: Given $\lambda\in(0,\frac{1}{2})$, there exist a number $\delta>0$ and a constant C independent of $\hbar<\delta$ such that

$$\left|\left| e^{-iH(\hbar)/\hbar}(e^{iS_0/\hbar}f_k) - e^{iS_k/\hbar + i\mu_k\pi/2}\left|\det\left[\frac{\partial Q}{\partial q_0}(Q_k^{-1}(\cdot))\right]\right|^{-1/2} \times f_k(Q_k^{-1}(\cdot))\chi_k \right|\right| \leqslant C\hbar^\lambda,$$

for all $\hbar\in(0,\delta)$.

Proof: We will omit the subscript k at the risk of confusion with previously defined quantities. The branches of square roots appearing in the proof are determined according to the discussion following Proposition 4.2. For $x\in\mathbb{R}^n$ let

$$\mathscr{F}_\hbar(x) = (4\pi\hbar)^{-n/4} \times \int_{\mathscr{N}} (\det[\mathbf{1}+S_0^{(2)}(q_0)])^{1/2} f(q_0) e^{iS_0(q_0)/\hbar} \times \phi_0(\mathbf{1},\mathbf{1},\hbar,q_0,S_0^{(1)}(q_0),x)d^Nq_0.$$

By definition of ϕ_0 and Lemma 3.1, there exist δ_1 and C_1 independent of $\hbar$ such that

$$\|e^{-iH(\hbar)/\hbar}(e^{iS_0/\hbar}f) - e^{-iH(\hbar)/\hbar}\mathscr{F}_\hbar\| \leqslant C_1\hbar^\lambda, \tag{4.5}$$

for all $\hbar\in(0,\delta_1)$ and $t\geqslant 0$. By Theorem 1.1 of Hagedorn,[6] for any $q_0\in\mathbb{R}^n$ there exists a constant C_2 such that

$$\left|\left| e^{iH(\hbar)/\hbar}(e^{iS_0/\hbar}\phi_0(\mathbf{1},\mathbf{1},\hbar,q_0,S_0^{(1)}(q_0),\cdot)) - e^{iS(q_0,S_0^{(1)}(q_0),t)/\hbar}\sum_{|\alpha|=0}^{3(l-1)} c_\alpha(\hbar,q_0,t) \times \phi_\alpha(A(q_0,S_0^{(1)}(q_0),t),B(q_0,S_0^{(1)}(q_0),t), \hbar,a(q_0,S_0^{(1)}(q_0),t),\eta(q_0,S_0^{(1)}(q_0),t),\cdot)\right|\right| \leqslant C_2\hbar^{l/2}, \tag{4.6}$$

where $c_\alpha(\hbar,q_0,\tau)$ is the unique solution of the system of ordinary differential equations

$$\frac{\partial}{\partial\tau}c_\alpha(\hbar,q_0,\tau) = \sum_{|\beta|=0}^{3(l-1)}\sum_{|\mu|=3}^{l+1} -i\hbar^{(|\mu|-2)/2}(\mu!)^{-1}[D^\mu V] \times (a(q_0,S_0^{(1)}(q_0),\tau))b_{\alpha\beta\mu}(q_0,\tau)c_\beta(\hbar,q_0,\tau), \tag{4.7}$$

subject to the initial conditions $c_0(\hbar,q_0,0)=1$ and $c_\alpha(\hbar,q_0,0)=0$ for $|\alpha|=0$. The quantities $b_{\alpha\beta\mu}$ are defined by

$$b_{\alpha\beta\mu}(q_0,\tau) = \langle \phi_\alpha(A(q_0,S_0^{(1)}(q_0),\tau),B(q_0,S_0^{(1)}(q_0),\tau),1,0,0,\cdot), x^\mu\phi_\beta(A(q_0,S_0^{(1)}(q_0),\tau),B(q_0,S_0^{(1)}(q_0),\tau),1,0,0,\cdot)\rangle. \tag{4.8}$$

By the argument of Lemma 2.5 of Hagedorn[5] and the remark following that lemma there exists a constant C' such that the functions c_α satisfy

$$|c_0(\hbar,q_0,\tau)-1| \leqslant C'\hbar^{1/2}, \tag{4.9a}$$

$$|c_\alpha(\hbar,q_0,\tau)| \leqslant C'\hbar^{r/2}, \tag{4.9b}$$

for $3(r-1)<|\alpha|\leqslant 3r$ and for all $\tau\in[0,t]$.

From the proofs[4–6] of these facts we conclude that (4.6) and (4.9) hold uniformly for q_0 in a compact subset of $\mathbb{R}^n$. From (4.7), (4.8), and the differentiability properties discussed above we conclude that c_α is of class C^1 in the variable $q_0\in\mathbb{R}^n$. Define

$$c_\alpha(\hbar,q) = c_\alpha(\hbar,Q^{-1}(q),t)$$

and

$$\Phi_\alpha(\hbar,x,t) = (4\pi\hbar)^{-n/4}\int_{\mathscr{N}} (\det[\mathbf{1}+S_0^{(2)}(q_0)])^{1/2} f(q_0) \times e^{iS(q_0,S_0^{(1)}(q_0),t)/\hbar} c_\alpha(\hbar,q_0,t) \times \phi_\alpha(A(q_0,S_0^{(1)}(q_0),t),B(q_0,S_0^{(1)}(q_0),t),\hbar, a(q_0,S_0^{(1)}(q_0),t),\eta(q_0,S_0^{(1)}(q_0),t),x)d^nq_0.$$

By (4.6),

$$\left|\left| e^{-iH(\hbar)/\hbar}\mathscr{F}_\hbar - \sum_{|\alpha|=0}^{3(l-1)}\Phi_\alpha(\hbar,\cdot,t)\right|\right| \leqslant C_2\hbar^{1/2}. \tag{4.10}$$

Changing the variable of integration in the definition of Φ_α,

$$\Phi_\alpha(\hbar,x,t) = (4\pi\hbar)^{-n/4}\int_{Q[\mathscr{N}]} (\det[\mathbf{1}+S_0^{(2)}(Q^{-1}(q))])^{1/2} \times f(Q^{-1}(q))e^{iS(q)/\hbar} \times c_\alpha(\hbar,q)\phi_\alpha(A(q),B(q),\hbar,q,p(q),x) \times \left|\det\left(\frac{\partial Q^{-1}}{\partial q}(q)\right)\right| d^nq. \tag{4.11}$$

We now note that the functions $\mathscr{H}_\alpha$ appearing in the definition of $\phi_\alpha(A(q),B(q),\hbar,q,p(q),x)$ are of class C^1 in $q\in Q[\mathscr{N}]$ by virtue of the fact that the matrices $\mathscr{U}_{A(q)}$ and $|A(q)|$ are continuously differentiable with respect to $q\in Q[\mathscr{N}]$. Hence we can apply Lemmas 3.1 and 3.2 to each term in each of the Φ_α's [the function $(f\circ Q^{-1})_{\chi_{Q[\mathscr{N}]}}$ is of class C_0^1 on $\mathbb{R}^n$] and conclude, by (4.4), (4.9), (4.11), and Proposition 4.1 that there exist δ_3 and C_3 independent of $\hbar\in(0,\delta_3)$ such that

$$\left|\left| \sum_{|\alpha|=0}^{3(l-1)}\Phi_\alpha(\hbar,\cdot,t) - e^{iS(\cdot)/\hbar + i\mu\pi/2} \times \left|\det\left(\frac{\partial Q}{\partial q_0}(Q^{-1}(\cdot))\right)\right|^{-1/2} f(Q^{-1}(\cdot))\chi\right|\right| \leqslant C_3\hbar^\lambda, \tag{4.12}$$

for all $\hbar\in(0,\delta_3)$. Equations (4.5), (4.10), (4.12), and the triangle inequality complete the proof of the lemma. ■

ACKNOWLEDGMENTS

This paper is based on the author's Ph.D. dissertation "Semiclassical quantum dynamics" at Virginia Polytechnic Institute and State University under the direction of George A. Hagedorn. The author wishes to thank Professor Hagedorn for his guidance and assistance in the preparation of this paper.

[1]V. P. Maslov, *Théorie des Perturbations et Méthodes Asymptotiques* (Dunod, Paris, 1972).
[2]V. P. Maslov and M. V. Fedoriuk, *Semi-Classical Approximation in Quantum Mechanics* (Reidel, Boston, 1981).
[3]S. Robinson, "The semiclassical limit of quantum dynamics. II. Scattering theory," preprint, Virginia Polytechnic Institute, 1986.
[4]G. A. Hagedorn, "Semiclassical quantum mechanics I: the $\hbar \to 0$ limit for coherent states," Commun. Math. Phys. **71**, 77 (1980).
[5]G. A. Hagedorn, "Semiclassical quantum mechanics III: the large order asymptotics and more general states," Ann. Phys. (NY) **135**, 58 (1981).
[6]G. A. Hagedorn, "Semiclassical quantum mechanics IV: large order asymptotics and more general states in more than one dimension," Ann. Inst. H. Poincaré **42**, 363 (1985).
[7]S. Lefschetz, *Differential Equations: Geometric Theory* (Interscience, New York, 1957).
[8]J. B. Keller, "Corrected Bohr–Sommerfeld quantum conditions for nonseparable systems." Ann. Phys. (NY) **4**, 180 (1958).
[9]V. Bargmann, "On a Hilbert space of analytic functions and an associated integral transform," Commun. Pure Appl. Math. **XIV**, 187 (1961).
[10]A. S. Householder, *The Theory of Matrices in Numerical Analysis* (Dover, New York, 1974).
[11]V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, Berlin, 1978).

Curvature structure in general relativity

G. S. Hall and W. Kay
Department of Mathematics, University of Aberdeen, Edward Wright Building, Dunbar Street, Aberdeen AB9 2TY, Scotland

(Received 24 April 1987; accepted for publication 30 September 1987)

This paper investigates the extent to which the curvature structure of space-time determines the metric stucture. It continues the work of earlier papers by prescribing the curvature structure and the curvature covariant derivatives up to certain orders. It is shown that, with the exception of the so-called generalized *pp* waves, the curvature and its first and second covariant derivatives essentially determine the metric up to coordinate transformations.

I. INTRODUCTION

This paper continues the work of some earlier papers on the extent to which the curvature tensor components $R^{a}{}_{bcd}$ in some coordinate domain of a space-time (M,g) determine the metric. Under reasonable conditions on the curvature components, it has been shown that the metric is as tightly determined as it can be, the only ambiguity being a constant conformal factor.[1] This paper explores those space-times where these conditions do not hold (and where a greater ambiguity in the metric results) by investigating the extra restrictions forced on the metric by prescribing not only the curvature components but also the first covariant derivatives of the curvature. In some very special cases, the second covariant derivatives of the curvature must also be assumed given. This problem is dealt with in Sec. III–VI, Sec. II being given over to some preliminary notions.

The notation throughout will be conventional. It is convenient on many occasions to use a real null tetrad of vector fields, (l,n,x,y), where the only nonvanishing inner products are $l^{a}n_{a} = x^{a}x_{a} = y^{a}y_{a} = 1$. The Lorentz metric g will have signature $(-,+,+,+)$, Latin indices will take the values 0,1,2,3, whilst Greek indices will range over 1,2,3. Second-order symmetric tensors on space-time will often be described in terms of their Segre type and the notation and results here will be taken from Ref. 2. An asterisk will be used for the usual duality operation on a bivector and the *blade* of a simple bivector will have its usual meaning. The *canonical pair of blades* of a bivector will, for a nonsimple bivector, refer to the orthogonal timelike–spacelike pair of two-spaces it determines at each point, and for a simple bivector, refer to the blade of the bivector and that of its orthogonal complement. Throughout the paper, the curvature tensor will always be assumed nonzero at each point in the region under discussion. All structures on M will be assumed smooth (although this restriction is easily weakened).

II. RECURRENCE AND DECOMPOSABILITY

In this section, the concept of recurrence and the related concept of reducibility will be reviewed. Most of the results mentioned here are contained explicitly or implicitly in the literature[3–9] and so the discussion will be brief.

A nonzero tensor W on M will be called *recurrent* if on each chart, $W^{a\cdots b}_{c\cdots d;e} = W^{a\cdots b}_{c\cdots d}p_{e}$ (where a semicolon denotes a covariant derivative with respect to g) for some (co)vector field p called the *recurrence vector*. If W is recurrent then so is ψW for any real valued differentiable function ψ on M. An equivalent statement is that if $p,q \in M$ and c is any differentiable curve from p to q, the tensor obtained at q by parallelly propagating W_{p} from p to q is proportional to W_{q}. A recurrent tensor W may be locally scaled so as to be covariantly constant if and only if the recurrence vector is locally a gradient. It follows that a nonzero recurrent vector field k on M is everywhere either spacelike, timelike, or null and, being necessarily hypersurface orthogonal, may be locally scaled so that $k_{a;b} \propto k_{a}k_{b}$. If a recurrent vector field k is everywhere non-null, then it can always be locally scaled so that $k_{a;b} = 0$. A null, recurrent vector field k on U may be locally scaled to be covariantly constant if and only if $R^{a}{}_{bcd}k^{d} = 0$ on M.

If a recurrent bivector F is admitted then $\overset{*}{F}$ is also recurrent and if F is non-null the recurrence vector is again locally a gradient and F (and $\overset{*}{F}$) can be locally scaled so as to be covariantly constant. In the non-null case, the two independent null directions in the timelike member of its canonical blade pair are recurrent. Conversely, if two recurrent vector fields l and n are admitted, the timelike bivector $2l_{[a}n_{b]}$ is recurrent (and is covariantly constant if l and n are scaled so that $l^{a}n_{a} = 1$). A recurrent, null bivector is not necessarily locally scalable to being covariantly constant but its principal null direction is a recurrent null vector field. The type of a recurrent bivector (timelike, spacelike, null, nonsimple) is the same throughout M.

If M admits a recurrent non-null bivector with principal null directions l and n then the Ricci identity can be used to show that the Petrov type is either O or D with l and n as repeated principal null directions of the Weyl tensor and that l and n are Ricci eigendirections, the Segre type of the Ricci tensor being $\{(1,1)(11)\}$ or its degeneracy. If M admits a recurrent null vector field l then the Petrov type is either O or one of the algebraically special types with l as a repeated principal null direction and l is a Ricci eigenvector. In this case, the Petrov type is III, N, or O if and only if the Ricci scalar $R = 0$. In vacuum, a recurrent vector or bivector must be null and the Petrov type is N or III [and is necessarily N (*pp* waves) if a constant vector or bivector is admitted].

A space-time M will be called *locally decomposable* if each $p \in M$ is contained in a chart domain U in which the metric takes either of the forms (α) or (β),

$$
\begin{aligned}
&(\alpha)\quad ds^{2} = \epsilon\, dx^{0^{2}} + g_{\alpha\beta}\, dx^{\alpha}\, dx^{\beta}, \quad g_{\alpha\beta} = g_{\alpha\beta}(x^{\gamma}),\\
&(\beta)\quad ds^{2} = g_{AB}\, dx^{A}\, dx^{B} + g_{A'B'}\, dx^{A'}\, dx^{B'}, \qquad (1)\\
&\qquad g_{AB} = g_{AB}(x^{c}), \quad g_{A'B'} = g_{A'B'}(x^{c'}),
\end{aligned}
$$

 0022-2488/88/020420-08$02.50

where $A,B,C = 1,2$, $A', B', C' = 3,4$, and $\epsilon = \pm 1$. The $1+3$ decomposition in (α) and $2+2$ decomposition in (β) are necessarily either timelike/spacelike or vice versa. Equivalent conditions are that each $p \in M$ lies in a neighborhood V such that any of the following hold.

(i) Here V admits a non-null, recurrent vector field for (α) and a non-null, recurrent bivector field for (β). Both the vector and bivector can then be locally scaled so as to be covariantly constant.

(ii) The neighborhood V admits a second-order, symmetric, recurrent tensor T_{ab} with general Segre type $\{1,111\}$ but with at least one degeneracy and not proportional to g_{ab}. The tensor T_{ab} can be locally scaled so as to be covariantly constant because for such tensors $T_{ab}T^{ab} \neq 0$ and so the recurrence vector is locally a gradient. (This condition is equivalent to the existence of a second-order, symmetric, recurrent tensor P_{ab} not proportional to the metric and satisfying $P^a{}_b P^b{}_c = P^a{}_c$.)

(iii) Here V admits a non-null vector field X for case (α) [a pair of vector fields Y and Z spanning a non-null two-space at each $p \in V$ for case (β)] such that the one-dimensional (two-dimensional) distribution so defined on V is invariant under parallel transport along any curve in V. [This provides the geometrical interpretation for local reducibility. The distributions above are necessarily integrable and their complementary distributions enjoy the same properties as the original ones. The two families of submanifolds that result in each case are spanned by the recurrent vector field and its orthogonal complement in (α) and by the canonical blades of the recurrent bivector in (β) in condition (i) above and determine the Segre decomposition in (ii). They also provide the coordinates for Eq. (1).]

III. CURVATURE STRUCTURE

The curvature structure of space-time arises naturally in the principal (frame) bundle $L(M)$ from the connection on $L(M)$ that, in turn, is uniquely determined as the torsion-free connection compatible with the space-time metric g and there has some recent interest in the extent to which the curvature determines the metric that gave rise to it (for a bibliography see Ref. 10). Clearly, if σ is a positive constant, then g and σg give rise to the same connection and curvature structure on M and a sufficient condition on the curvature for this to be the only ambiguity in g, together with a statement of the possible extra ambiguities when this condition fails to be satisfied, was given in Ref. 1. It is, of course, assumed that the curvature components $R^a{}_{bcd}$ are given in some connected coordinate domain U and arise from some Lorentz metric g. Any other Lorentz metric g' compatible with this curvature must satisfy the algebraic conditions $g'_{e(a}R^e{}_{b)cd} = 0$ at each $p \in U$. The solution of this equation depends on the rank of $R_{abcd} \equiv g_{ae}R^e{}_{bcd}$ regarded as a 6×6 symmetric matrix and on the algebraic nature of the range space of the linear map from the vector space of bivectors at p to itself given in the obvious way by $F_{ab} \to R_{abcd}F^{cd}$. (Here and throughout indices will always be raised and lowered with the original, given metric g.) It turns out[1,10,11] that the blade of any simple bivector in the range of this map consists entirely of eigenvectors of g' with respect to g at p (and hence all have the same eigenvalue) whilst the same is true for each of the canonical blades of any nonsimple bivector in this range (but the two eigenvalues arising may be distinct). When the curvature is of sufficiently high rank one obtains the situation where all members of the tangent space to M at p are eigenvectors of g' with respect to g with the same eigenvalue and so g' is a multiple of g. The following cases arise at each $p \in M$.

(a) Rank = 1. Here the range space is spanned by a single bivector that is necessarily simple (since $R_{a[bcd]} = 0$).

(b) Rank = 2 or 3 but with the members of the range simple bivectors with a unique, common eigenvector k with zero eigenvalue.

(c) Rank = 2 but with the range spanned by an orthogonal timelike–spacelike pair of simple bivectors (with respect to g) whose blades are spanned, respectively, by the pairs (l,n) and (x,y) where l,n,x,y constitute a null tetrad with respect to g.

(d) All other cases.

The general solutions for g' at p in these cases are, respectively,

$$\begin{aligned}
&\text{(a)}\ g'_{ab} = \phi g_{ab} + \alpha u_a u_b + 2\beta u_{(a}v_{b)} + \gamma v_a v_b,\\
&\text{(b)}\ g'_{ab} = \phi g_{ab} + \theta k_a k_b,\\
&\text{(c)}\ g'_{ab} = \phi g_{ab} + 2\lambda l_{(a}n_{b)}\\
&\qquad\quad = 2(\phi+\lambda)l_{(a}n_{b)} + \phi(x_a x_b + y_a y_b),\\
&\text{(d)}\ g'_{ab} = \phi g_{ab},
\end{aligned} \tag{2}$$

where $\phi,\alpha,\beta,\gamma,\theta,\lambda \in \mathbb{R}$ (restricted to preseve Lorentz signature) and u,v span the orthogonal complement of the blade of the spanning simple bivector in (a). The rank may, of course, change from point to point, but only in a manner that is consistent with the lower semicontinuity of the rank function so that if the rank at p is n then the rank is not less than n in some neighborhood of p. Here, however, it will be assumed that either case (a), (b), (c), or (d) holds throughout the coordinate domain U concerned. In this case, it can be shown that the assumed smoothness of the original metric (and hence the curvature) means that one can choose U so that the vectors u and v in case (a), k in case (b), and l,n,x,y in case (c) can (and will) be chosen smoothly over U. The solutions g' in each case are then smooth tensors on U if and only if the associated functions $\phi,\alpha,\beta,\gamma,\theta$, and λ are smooth and this will be assumed to be the case.

It is important to note that the equation

$$R_{abcd}\omega^d = 0 \tag{3}$$

has exactly two independent solutions (u and v) in case (a), exactly one independent solution (k) in case (b), and no nonzero solutions in cases (c) and (d).

Further information about cases (c) and (d) is immediately available. If case (d) holds at each $p \in U$ then it has been shown that ϕ is necessarily constant on U (Ref. 1) and so in this case the "best possible" result is obtained and the (symmetric, metric) connection compatible with the curvature is uniquely determined.

If case (c) holds at each $p \in U$ the curvature tensor takes the form

$$R_{abcd} = a F_{ab} F_{cd} + b \overset{*}{F}_{ab} \overset{*}{F}_{cd}$$

$$(F_{ab} = 2l_{[a}n_{b]},\ \overset{*}{F}_{ab} = 2x_{[a}y_{b]}), \qquad (4)$$

with a,b nowhere zero smooth functions on U. Considering now the original metric g and denoting the corresponding covariant derivative by a semicolon a contraction of the Bianchi identity $R_{ab[cd;e]} = 0$ first with $l^a x^b l^c n^d x^e$ and then with $l^a x^b l^c n^d y^e$, and use of (4) yields $l_{a;b}x^a x^b = l_{a;b}x^a y^b = 0$. Similarly, contractions with $l^a x^b x^c y^d l^e$ and $l^a x^b x^c y^d n^e$ give $l_{a;b}y^a l^b = l_{a;b}y^a n^b = 0$. The symmetry of the curvature tensor under the exchanges $l \leftrightarrow n$ and $x \leftrightarrow y$ reveals a similar set of relations and one deduces that $l_{a;b}x^a = l_{a;b}y^a = 0$. Since $l_{a;b}l^a = 0$, one sees that l is recurrent and, similarly, n is recurrent. Neither l nor n is scalable to a constant because of the absence of solutions to (3) in this case. It follows that U is $2+2$ locally decomposable with the surface forming blades of F and $\overset{*}{F}$ spanning the submanifolds of decomposition (described in Sec. II) at each $p \in U$. Again referring to Sec. II, F and $\overset{*}{F}$ are covariantly constant non-null bivectors and $2l_{(a}n_{b)}$ and $x_a x_b + y_a y_b$ are covariantly constant tensors with Segré type $\{(1,1)(11)\}$. The local decomposability also implies that, in (4), the functions a and b depend, respectively, only on the coordinates in the submanifolds spanned by F and $\overset{*}{F}$. The Petrov type is D or O and the Ricci tensor is of Segré type $\{(1,1)(11)\}$ or its degeneracy.

Now suppose that another metric g' of the form (2c) is also compatible with the curvature obtained from g. Then l_a (or $l'^a = g'^{ab}l_b$) and n_a are recurrent with respect to g' also. Denoting covariant derivatives with respect to g' by a stroke and, by P^a_{bc}, the difference between the Christoffel symbols Γ^a_{bc} and $\{^a_{bc}\}$ formed from g and g', respectively, one has, using a convenient result in Ref. 12 (with g'^{ab} denoting the inverse matrix of g'_{ab}),

$$P^a_{bc} = \{^a_{bc}\} - \Gamma^a_{bc} = \tfrac{1}{2}g'^{ad}(g'_{db;c} + g'_{dc;b} - g'_{bc;d}). \qquad (5)$$

The recurrence conditions on l_a show that $l_a P^a_{bc} \propto l_b l_c$. This condition, together with (5), the covariant constancy of the tensors $2l_{(a}n_{b)}$ and $x_a x_b + y_a y_b$ and some appropriate contractions show that $(\phi + \lambda)_{,a}$ lies everywhere in the blade of F whilst $\phi_{,a}$ lies everywhere in the blade of $\overset{*}{F}$. Since these blades are the submanifolds of decomposition, the ambiguity in the metric represented by (2c) can be interpreted as a conformal scaling of the metrics $2l_{(a}n_{b)}$ *and* $x_a x_b + y_a y_b$ induced in the submanifolds of decomposition by the original metric g with the functions $\phi + \lambda$ and ϕ, respectively. If one chooses conformally null coordinates u and v in the blades of F and conformally Euclidean coordinates x^2 and x^3 in the blades of $\overset{*}{F}$ one can compute the curvatures of g and g' and find that they are equal if and only if

$$\eta\eta_{,uv} = \eta_{,u}\eta_{,v} \quad (\eta = \phi + \lambda),$$
$$\phi\phi_{,\xi\bar{\xi}} = \phi_{,\xi}\phi_{,\bar{\xi}} \quad (\zeta = x^2 + ix^3),$$

for which a solution is $\eta = \alpha(u)\beta(v)$, $\phi = \gamma(\xi)\delta(\bar{\xi})$ with α,β,γ, and δ arbitrary differentiable functions (but with ϕ and η positive functions to preserve Lorentz signature).

IV. PRESCRIBING THE CURVATURE AND ITS FIRST COVARIANT DERIVATIVES

Suppose now that the components $R^a{}_{bcd;e}$ as well as the components $R^a{}_{bcd}$ are given in U (and that, as before, these components are consistent with some Lorentz metric g). Then the algebraic relations $g'_{f(a}R^f{}_{b)cd;e} = 0$ for any alternative Lorentz metric g' will supply, in general, further eigenvectors for g' with respect to g and hence further information on g'. In this context, a useful result was given in Ref. 13 that says that if g and $g' = \psi g$ are conformally related metrics with the same curvature and curvature derivative (in the above sense) on a connected manifold N of dimension $\geqslant 3$ and where the original metric giving rise to these given components is of arbitrary signature, then ψ is constant on N.

In case (d), the above extra conditions give no further information since the prescription of $R^a{}_{bcd}$ determines g up to a constant conformal factor and so the first- (and all higher-) curvature derivatives are necessarily determined. In case (c), however, further information is obtained. In fact, one can write the equality of the curvature derivatives in the form

$$R^a{}_{bcd;e} = R^a{}_{bcd/e} \Leftrightarrow R^a{}_{fcd}P^f_{be} + R^a{}_{bfd}P^f_{ce} + R^a{}_{bcf}P^f_{de} - R^f{}_{bcd}P^a_{fe} = 0. \qquad (6)$$

If one substitutes into (6) the expression (5) for P^a_{bc} and the ambiguity relation (2c) for g' [and recalls the covariant constancy of the tensors $2l_{(a}n_{b)}$ and $x_a x_b + y_a y_b$ with respect to g] a contraction with $l_a n^b l^c n^d l^e$ shows that $(\phi + \lambda)_{,a}$ is orthogonal to l. A reversal of the roles of l and n then shows that $(\phi + \lambda)_{,a}$ is orthogonal to n. Since it has already been shown that $(\phi + \lambda)_{,a}$ is orthogonal to x and y, it follows that $\phi + \lambda$ is constant on U. A similar argument involving a contraction with $x_a y^b x^c y^d x^e$ then shows that ϕ is constant on U. It follows in this case that the prescription of the curvature and its first derivatives determines the metric up to constant conformal scalings in each of the submanifolds of decomposition. Also, with this ambiguity, $P^a_{bc} = 0$ and so the connection is uniquely determined. Hence the curvatures and curvature derivatives of all orders for g and g' are equal.

Now consider case (b). Here the curvature is of rank 2 or 3 and has a uniquely determined direction k satisfying $k_a R^a{}_{bcd} = 0$. If $k_a R^a{}_{bcd;e} \neq 0$ at each point of U then sufficient extra eigenvectors of any alternative metric g' are generated to ensure that g' is conformaly related to g on U. It then follows from the result mentioned at the beginning of this section in Ref. 13 that the conformal factor is necessarily constant and so the metric is determined up to a constant conformal factor. So suppose that $k_a R^a{}_{bcd;e} = 0$ on U with k chosen smoothly on U. It then follows that $R^a{}_{bcd}k_{a;e} = 0$ on U and so, by the uniqueness of the independent solutions to (3) in this case k is recurrent on U. From the results of Sec. II it may then be assumed that k is chosen to be covariantly constant on U with respect to g if U is contractible. Thus the prescription of $R^a{}_{bcd}$ and $R^a{}_{bcd;e}$ on U uniquely determines g up to a constant conformal factor in case (b) unless k is (proportional to) a covariantly constant vector field as above. However, k_a must also be recurrent with respect to g' (but the scaling required to make it covariantly constant

with respect to g' will, in general, be different from that required with respect to g). So suppose k is scaled so that $k_{a;b}=0$. Then the signs (or zeros) of $k_a k_b g^{ab}$ and $k_a k_b g'^{ab}$ are constant throughout U and the same for both metrics. Thus there are only three cases to be considered, namely when k is timelike, spacelike, or null. Suppose first that k is timelike (and scaled so that $k_a k_b g^{ab}=-1$ together with $k_{a;b}=0$). The fact that k_a is recurrent with respect to g' yields

$$k_c P^c_{ab} \propto k_a k_b \quad [\Rightarrow \phi_{,a} k^a = 0 \text{ and } (\phi-\theta)_{,a} \propto k_a], \tag{7}$$

where the implications follow after a contraction with a vector orthogonal to k with respect to g and hence g'. Now for each $p \in U$ there is a coordinate domain in U (still labeled U) containing p such that in this coordinate domain the original metric takes the form

$$ds^2 = -du^2 + h_{\alpha\beta}\, dx^\alpha\, dx^\beta, \tag{8}$$

where $k_a = u_{,a} = (1,0,0,0)$, $k^a = g^{ab}k_b = (-1,0,0,0)$, and $h_{\alpha\beta}$ is independent of u. Then (7) shows that ϕ is independent of u and $\phi-\theta$ is independent of the x^α. The alternative metric g' takes the form

$$ds'^2 = -(\phi-\theta)du^2 + \phi h_{\alpha\beta}\, dx^\alpha\, dx^\beta. \tag{9}$$

Now consider the positive definite metrics h and ϕh induced in any of the hypersurfaces $u=$ constant by the metrics g and g'. It is easily shown from (8) and (9), together with the fact that ϕ is independent of u, that the curvature tensors of type (1,3) derived from the metrics h and ϕh, in the coordinate system x^α, have the same components [necessarily equal to the components $R^{\alpha}{}_{\beta\gamma\delta}$ of the type (1,3) curvature tensor of g and g']. Now the theorem proved in Ref. 1 and used to deal with the case (d) in Sec. III, when applied in the case of a three-dimensional positive definite space shows that when the curvature is of rank 2 or 3 the metric is determined up to a constant conformal factor. The assumptions of the present case show that the rank of the induced curvature in the hypersurfaces $u=$ constant is either 2 or 3 and is compatible with both h and ϕh. Hence ϕ is independent of x^α and is thus constant on U. Equation (7) then shows that θ depends only on u. These conditions determine the ambiguity completely because with ϕ constant and θ an arbitrary smooth function of u in the above coordinates, g and g' necessarily have equal curvature components $R^{a}{}_{bcd}$ as well as equal first- (and all higher-) curvature derivatives. This is a consequence of the following results, the first of which was given in Ref. 12:

$$\begin{aligned} 0 &= R'^{a}{}_{bcd} - R^{a}{}_{bcd} \\ &= P^a_{bd;c} - P^a_{bc;d} + P^a_{fc}P^f_{db} - P^a_{fd}P^f_{cb}, \end{aligned} \tag{10}$$

$$P^a_{bc} = \frac{1}{2}(\phi-\theta)^{-1}\frac{d\theta}{du}\, k^a k_b k_c, \tag{11}$$

together with (6) and similar equations for the higher derivatives. Here, the $R'^{a}{}_{bcd}$ denote the curvature components derived from g'. It follows that the corresponding connections are equal if and only if θ is constant on U.

The case when k is spacelike is essentially the same and so the case when k is null on U will be considered now. Suppose k is scaled and coordinates x^a are chosen in a subset of U (still labeled U) so that[8]

$$ds^2 = g_{\alpha\beta}\, dx^\alpha\, dx^\beta - 2\, dx^3\, dx^0, \tag{12}$$

where the $g_{\alpha\beta}$ depend only on the x^α, $k_a = (0,0,0,1) = u_{,a}$ $(u=x^3)$, $k^a = (-1,0,0,0)$, and $k_{a;b}=0$. A similar argument to that given at the beginning of the timelike case shows that ϕ depends only on u. Next, one notes that the relation (3) with k null implies that the Weyl tensor corresponding to g (and g') is algebraically special (including type O) in the Petrov classification with k as a repeated principal null direction.[14,15] The same relation also shows that if x^a and y^a are spacelike vector fields orthogonal to each other and to k with respect to g (and hence g') then R_{abcd} may be decomposed into the bivectors $F_{ab} = 2k_{[a}x_{b]}$, $G_{ab} = 2k_{[a}y_{b]}$, and $H_{ab} = 2x_{[a}y_{b]}$ as

$$\begin{aligned} R_{abcd} &= R_1 F_{ab}F_{cd} + R_2 G_{ab}G_{cd} + R_3 H_{ab}H_{cd} \\ &\quad + R_4(F_{ab}G_{cd} + G_{ab}F_{cd}) \\ &\quad + R_5(F_{ab}H_{cd} + H_{ab}F_{cd}) \\ &\quad + R_6(G_{ab}H_{cd} + H_{ab}G_{cd}), \end{aligned} \tag{13}$$

where $R_1 \cdots R_6$ are smooth functions on U. It will also be assumed that x^a and y^a are normalized with respect to g and then (13) gives for the Ricci tensor (for either g or g'),

$$\begin{aligned} R_{ab} &= (R_1+R_2)k_a k_b + 2R_6 k_{(a}x_{b)} \\ &\quad - 2R_5 k_{(a}y_{b)} + R_3(x_a x_b + y_a y_b). \end{aligned} \tag{14}$$

The remainder of the argument then proceeds by considering Petrov types (and it is known that g and g' have the same Petrov type[15]). Also the Weyl tensor rank (as a 6×6 matrix) is 0 for type O, 2 for type N, 4 for type III, and 6 for types II or D. Using the lower semicontinuity of rank it follows that there exist open subsets A and B and U (one of which could be empty) such that the Weyl tensor is of Petrov type III, II, or D on A and N or O on B and that $A \cup B$ is dense in U. The subsets A and B will be considered separately.

Since k is a recurrent null vector field on U, it follows from Sec. II that the Petrov type at $p \in A$ is II or D if and only if $R \neq 0$ at p. If this is the case then $R \neq 0$ in some neighborhood of p and a contraction of (6) with $x^a y^b x^c y^d$ and use of (13) gives $R\phi_{,a}=0$. Hence ϕ is constant in some neighborhood of p. Substituting this back into (6) and performing some further algebraic manipulations gives $\theta_{,a} \propto k_a$ and so θ depends only on u in this neighborhood. If, on the other hand, the Petrov type at p is III, the assumption that $\phi_{,a} \neq 0$ at p and a contraction of (6) with $x_a n^b n^c y^d y^e$ gives $R_6 = 0$ in (13) at p. A similar contraction gives $R_5 = 0$ at p and the resulting Ricci tensor at p given by (14) is now *not* of Segre type {(31)} with zero eigenvalue, contradicting.[15] Hence $\phi_{,a}=0$ at all $p \in A$ and so ϕ is a constant on each (necessarily open) component of A. Further, since the Ricci tensor has Segre type {(31)} with zero eigenvalue and null eigendirection spanned by k one may choose x and y so that $R_1 + R_2 = R_3 = R_5 = 0$ in (14) at p. On substituting this into (13) and the resulting expression into (6) one finds that either $\theta_{,a}=0$ or $R_{ab}\theta^{,b}=0$ at p. In the latter case $\theta_{,a}x^a = \theta_{,a}k^a = 0$ at p and a further contraction of (6) with

$x_a n^b n^c y^d$ shows that $\theta_{,a} \propto k_a$ at p. Hence ϕ is constant on each component of A and θ depends on u only throughout A, giving the same type of metric ambiguity as was obtained in the case when k was non-null. Once again, the connections associated with g and g' are equal on A if and only if $\theta_{,a} = 0$ on A (that is, θ is constant on each component of A) and the curvatures and their corresponding covariant derivatives of all orders are equal, independently of $\theta(u)$. Hence no further information is available from higher-order curvature derivatives than the first.

In B the Ricci tensor, if nonzero, has Segre type $\{(211)\}$ with zero eigenvalue and null eigendirection spanned by k (Ref. 15) and the curvature has rank 2 as follows directly from algebraic considerations. Since the Petrov type in B is N or O it follows from a result in Ref. 7 (see Eq. (2.3) of this reference) that if $\overset{+}{F}$ is any null, complex, self-dual bivector with principal null direction k (which is necessarily recurrent since k is) then $\overset{+}{F}$ has vanishing skew derivative $\overset{+}{F}_{ab;[cd]} = \overset{+}{F}_{ab/[cd]} = 0$. This implies that the recurrence vector corresponding to $\overset{+}{F}$ is a gradient and hence that $\overset{+}{F}$ can be (complex) scaled so as to be covariantly constant with respect to g in a neighborhood of any $p \in B$. In such a neighborhood the line element for g may be written as the special case of (12) (Refs. 4 and 8),

$$ds^2 = -2\,du\,dv + dx^{1^2} + dx^{2^2} + H(x^1,x^2,u)du^2, \qquad (15)$$

where x^0 has been replaced by v. In this case, the condition that the curvatures of g and g' are equal does not restrict the ambiguities so tightly. In fact the curvatures are equal if and only if $\phi = \phi(u)$ and[16]

$$\theta(u,v,x^1,x^2) = (\tfrac{3}{4}\phi^{-1}\phi')^2 - \tfrac{1}{2}\phi'' - \tfrac{1}{4}\phi^{-1}\phi' F(u)(x^{1^2} + x^{2^2}) + F(u)v + C(u)\,y + D(u)z + E(u),$$

where a prime denotes differentiation with respect to u and C, D, E, and F are functions of u. The first covariant derivatives of the curvature are then equal if and only if $F = 2\phi'$. It is not clear, in this case, what further information is obtainable by supposing the equality of higher-order curvature derivatives.

In all cases considered in this section the Petrov type and the Segre type of the Ricci tensor are easily found (and the table in Ref. 15 may be helpful but contains some omissions that are corrected in Ref. 17).

V. CASE (a) AND THE PRESCRIPTION OF SECOND-ORDER CURVATURE DERIVATIVES

This section will be concerned with case (a) and it will be supposed that the curvature (now of rank 1) and its first covariant derivatives are given. If the first curvature derivatives introduce enough extra bivectors for any alternative metric to be determined up to a conformal factor then the result in Ref. 13 used before can be used again to show that the conformal factor is constant and so the metric is determined up to a constant conformal factor.

Suppose now that the first curvature derivatives introduce no bivectors other than that bivector F introduced by the curvature. Then F is recurrent with respect to any metric compatible with the curvature and so F is either spacelike, timelike, or null (with respect to any compatible metric) everywhere in U.

If F is timelike U may be chosen so that there exist null vector fields l and n such that $l^a n_a = 1$ and $F_{ab} = 2l_{[a}n_{b]}$ that may be extended to a null tetrad on U, (l, n, x, y). The recurrence (and hence covariant constancy) of F implies the same for $\overset{*}{F}_{ab} = 2x_{[a}y_{b]}$ and so $x_a + iy_a$ is recurrent with a recurrence vector which is a gradient on account of the Ricci identity and the relations $R_{abcd}x^d = R_{abcd}y^d = 0$. Hence x and y may be chosen as covariantly constant with respect to g in the above null tetrad, $x_{a;b} = y_{a;b} = 0$. (The null vector fields l and n are, of course, recurrent.) The vector fields x and y are then recurrent with respect to any alternative compatible metric g' which for this case may always be written in the form[15]

$$g'_{ab} = \phi g_{ab} + \alpha x_a x_b + \beta y_a y_b. \qquad (16)$$

One then finds using the usual notation and appropriate expression for P^a_{bc},

$$P^a_{bc}\,l_a x^c = P^a_{bc} l_a y^c = P^a_{bc} n_a x^c = P^a_{bc} n_a y^c = 0$$
$$[\Rightarrow \phi_{,a}x^a = \phi_{,a}y^a = (\phi+\alpha)_{,a}l^a = (\phi+\alpha)_{,a}n^a \quad (17)$$
$$= (\phi+\beta)_{,a}l^a = (\phi+\beta)_{,a}n^a = 0].$$

Since the curvature components and their first covariant derivatives are specified in this case, the appropriate expressions for P^a_{bc} and the results of the previous equation when substituted into (6) and contracted with $l_a n^b l^c n^d$ reveal that ϕ is constant on U (the specific form of the curvature tensor $R_{abcd} \propto F_{ab}F_{cd}$ is used here). It then follows from (17) that $\alpha_{,a}l^a = \alpha_{,a}n^a = \beta_{,a}l^a = \beta_{,a}n^a = 0$. Now, as mentioned in Sec. II, such space-times are decomposable with the vector field pairs (l,n) and (x,y) spanning the submanifolds of decomposition. Using the covariant constancy of x and y with respect to g one can choose coordinates adapted to the decomposition such that

$$g_{ab} = \begin{pmatrix} -\psi & 0 & 0 & 0 \\ 0 & \psi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$
$$x_a x_b + y_a y_b = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \qquad (18)$$

It then follows that

$$g'_{ab} = \phi \begin{pmatrix} -\psi & 0 & 0 & 0 \\ 0 & \psi & 0 & 0 \\ 0 & 0 & \alpha' & 0 \\ 0 & 0 & 0 & \beta' \end{pmatrix}, \qquad (19)$$

where $\psi = \psi(x^0,x^1)$, $\alpha' = \alpha'(x^2,x^3)$, $\beta' = \beta'(x^2,x^3)$ with x^0 and x^1 coordinates in the submanifolds spanned by l and n and x^2 and x^3 coordinates in the submanifolds spanned by x and y. Now the curvature components $R^a{}_{bcd}$ from the original metric g are zero unless a, b, c, d are each either 0 or 1. On computing the curvature components $R'^a{}_{bcd}$ from g' one

finds that they are zero unless either a, b, c, d are each equal to 0 or 1 [in which case they are the same as the curvature components of the two-dimensional Lorentz metric diag$(-\psi, \psi)$] or a, b, c, d are each equal to 2 or 3 [in which case they are the same as the curvature components of the two-dimensional positive-definite metric diag(α', β')]. One also finds that $R^{a}{}_{bcd} \equiv R'^{a}{}_{bcd}$ if a, b, c, d are each equal to 0 or 1. Hence $R^{a}{}_{bcd} = R'^{a}{}_{bcd}$ if and only if diag(α', B') is a flat metric. The final result here is that g is determined (apart from the unavoidable constant conformal factor) up to a two-dimensional flat positive-definite metric in the submanifold spanned by the dual $\overset{*}{F}$ of the bivector F appearing in the curvature tensor (the restrictions on the connection then being clear). The final form for the tensor P^{a}_{bc} can then be used to show in a straightforward way that the curvatures and curvature derivatives of all orders for g and g' are now equal. The case when F is spacelike is essentially the same and is omitted. In both these cases (where F is non-null) the Petrov type is D and the Ricci tensor has Segre type $\{(1,1)(11)\}$.

Now consider the case when F is null. Here, the Petrov type is N and the Ricci tensor has Segre type $\{(211)\}$ with zero eigenvalue. In this case the situation is more complicated. However, choosing a null tetrad of vector fields (l, n, x, y) such that $F_{ab} = 2l_{[a}y_{b]}$ the metric ambiguity is[15]

$$g'_{ab} = \phi g_{ab} + \alpha l_a l_b + 2\beta l_{(a}x_{b)} + \gamma x_a x_b, \tag{20}$$

where α, β, and γ are smooth functions on U. One can arrange that l and x are covariantly constant with respect to g and hence recurrent with respect to g'. Similar calculations to the above then show that if l_a is written as $u_{,a}$, ϕ depends only on u. More information can be obtained here by dividing up the ambiguities represented by this equation into three cases characterized by the Segre type of g'_{ab} with respect to g_{ab} [necessarily either $\{(1,11)1\}$, $\{(31)\}$ or $\{(21)1\}$ or their degeneracies[15]]. Using the equality of the curvature and its first covariant derivatives for g and g' it can be shown[16] that in the first two of these cases ϕ must be constant and that in all cases further information is available on the functions α, β, and γ. In fact one can choose coordinates u, v, x, y such that the original metric g and the alternative metrics corresponding to the above Segre types are, respectively,

$$ds^2 = -2\,du\,dv + dx^2 + dy^2 + H(u,y)du^2$$
(original metric g—see Ref. 4),

$$ds'^2 = -2\phi\,du\,dv + \phi(dx^2 + dy^2) + \phi H(u,y)du^2 + \gamma(u,x)dx^2$$
[Segre type $\{(1,11)1\}$ or its degeneracy; $\alpha = \beta = 0$ in (20), $\phi =$ const and $(\gamma + \phi)\gamma_{,uu} = \frac{1}{2}(\gamma_{,u})^2$],

$$ds'^2 = -2\phi\,du\,dv = \phi(dx^2 + dy^2) + \phi H(u,y)du^2 + (B(u) + C(x))du\,dx$$
[Segre type $\{(31)\}$; $\alpha = \gamma = 0$ in (20), $\phi =$ const and $2\beta = B + C$],

$$ds'^2 = -2\phi(u)du\,dv + \gamma(u,x)dx^2 + \phi(u)(dx^2 + dy^2) + (\phi(u)H(u,y) + \alpha(u,v,x,y))du^2$$
[Segre type $\{(21)1\}$ or its degeneracy; $\beta = 0$ in (20) and

$$\alpha(u,v,x,y) = D(u,x) - 2v\phi_{,u} + yE(u) + \tfrac{1}{4}\phi^{-1}y^2((\phi_{,u})^2 - \phi\phi_{,uu})$$

and

$$2F\alpha_{,xx} = 2FD_{,xx} = F_{,x}D_{,x} + (F_{,u})^2 - 2FF_{,uu}$$
$(F = \gamma + \phi)$].

A lengthy calculation using the specific forms of the curvature in terms of the covariantly constant bivector F, the covariant constancy of the appropriate vector fields and the identity $R^{a}{}_{[bcd]} = 0$ shows that the assumption of the equality of the curvature and its first derivatives for g and g', in the first two cases above, implies the equality of the second and higher derivatives. Hence no further information is available in this way. However, for the case of Segre type $\{(21)1\}$ (or its degeneracy), the same assumption does not imply the equality of the second-curvature derivatives and if the latter condition is imposed further restrictions are obtained, one of which is that ϕ is constant on U. The third- and higher-order curvature derivatives for g and g' are then identically equal.

The only other possibility in case (a) is that the first-curvature derivatives introduce bivectors other than that introduced by the curvature itself but not sufficient to determine g up to a conformal factor. In this case it is not difficult to show that it is impossible for the curvature to introduce a non-null simple bivector whilst its derivative augments this by adding only the dual of this bivector. As a consequence, the case in question here is such that the simultaneous equations

$$R_{abcd}k^{d} = 0, \quad R_{abcd;e}k^{d} = 0, \tag{21}$$

uniquely determine the vector field k up to a scaling. Now one assumes that the second-curvature derivatives $R^{a}{}_{bcd;ef}$ are specified in U. If $R^{a}{}_{bcd;ef}k_{d} \neq 0$ on U then the extra bivectors introduced by these second derivatives together with the algebraic restrictions $g'_{k(a}R^{k}{}_{b)cd;ef} = 0$ on any alternative metric determine the metric up to a conformal factor and hence by the result mentioned earlier[13] the metric is determined up to a constant conformal factor. If, on the other hand, $R^{a}{}_{bcd;ef}k^{d} = 0$ on U then it easily follows from (21) that $k^{d}{}_{;e}P^{e}$ satisfies (21) for any choice of the vector field P and so k is recurrent and hence [by the first equation of (21)] can be assumed scaled so that it is constant on U. If k is non-null on U then the arguments given in case (b) of Sec. IV go through as before with a metric ambiguity of the form (2b) with ϕ constant and $\theta = \theta(u)$. The only difference in the argument is in the proof of the constancy of ϕ. That ϕ is independent of u goes through as before and one again projects to the hypersurfaces $u =$ const and notes that the three-

dimensional metrics h and ϕh in these hypersurfaces have the same intrinsic curvature components (necessarily equal to $R^{\alpha}{}_{\beta\gamma\delta}$) and the same intrinsic first covariant derivatives of the curvature (necessarily equal to $R^{\alpha}{}_{\beta\gamma\delta;\epsilon}$). One now applies the result of Ref. 13 to conclude that ϕ is constant. (The previous proof relied on the three-dimensional form of the theorem in Ref. 1 but this cannot be used here because the intrinsic curvature in the hypersurfaces only has rank 1.) If k is null on U, the bivector F introduced by the curvature is either spacelike or null. In the former case, the Petrov type is D, the Ricci tensor has Segre type $\{(1,1)(11)\}$ and $R \neq 0$ and so the ambiguity is as in the corresponding type for k null in case (b) of Sec. IV. This follows from the prescription of the curvature and its first and second derivatives and no further information can be obtained from prescribing higher-order derivatives. In the latter case the Petrov type is N, the Ricci tensor has Segré type $\{(211)\}$ with zero eigenvalue, and the ambiguity is again as in the appropriate type of case (b) in the final part of Sec. IV.

VI. SUMMARY AND DISCUSSION

The work in the previous four sections can be summarized in the following way. Using the notation of Sec. III, if the curvature components satisfy the conditions of case (d) then the metric is determined up to a constant conformal factor and the connection is determined uniquely. This result is "best possible" in the sense that prescribing higher-order curvature derivatives gives no further information. If the curvature components satisfy case (c) the curvature components and their first-covariant derivatives fix the metric up to two constant (two-space conformal) factors and the connection is uniquely determined. Again the result is best possible in the above sense. If the curvature components satisfy case (b) then with exception of the case when k is null and the Petrov type N or O, the curvature components and their first covariant derivatives determine the metric to within either a constant conformal factor (and so the connection is uniquely determined) or to within a constant conformal factor together with an arbitrary (up to signature requirements) function of a single variable [the determination of the connection is then given by (11)]. These results are all best possible. In the special case, the curvature components and their first covariant derivatives give certain information about the nature of the ambiguity but it is not clear what further information is obtainable from the equality of the second- or higher-covariant derivatives so it cannot be claimed that this result is best possible. If the curvature components satisfy the conditions of case (a) the curvature components and their first- and second-order covariant derivatives determine the metric and connection up to ambiguities described in Sec. V, and the results are best possible except in one special case that is very similar to the special case obtained in case (b).

These special cases can now be discussed. They constitute a subclass of that class of metrics characterized by the existence of a covariantly constant, nowhere-zero, null bivector or, equivalently, by the existence of a covariantly constant, nowhere-zero, null vector field l and by being of Petrov type N or O (and so, necessarily, l is the repeated principal null direction of the Weyl tensor, if it is nonzero, and of the Ricci tensor, if it is nonzero, that necessarily has the form $R_{ab} \propto l_a l_b$). Referring to this class as the class of "generalized *pp* waves" the special cases are a proper subclass since they have a two-parameter holonomy group (with the curvature of rank 1 or 2) whereas the generalized *pp* waves could also have a one-parameter holonomy group. These comments should be compared with the (vacuum) result given in Ref. 10 that a nonflat vacuum metric is determined up to a constant conformal factor by its curvature components and their first covariant derivatives unless it is a (vacuum) *pp* wave. This result is easily recovered from the present scheme by noting that a nonzero vacuum Riemann tensor has, on account of its self-duality, even rank and so case (a) is impossible. Case (c) is also impossible *in vacuo* and case (b) can only arise if k is null, a *pp* wave resulting if k is scalable to a covariantly constant vector field. Returning to the general case, it might be remarked, somewhat informally, that the generalized *pp* wave space-times were the only space-times where the best possible ambiguity did not assume a particularly simple form.

The results mentioned so far in this summary are clearly related to the holonomy group structure of space-time and this is discussed in detail in Refs. 17 and 18.

Finally, it was pointed out in Sec. III that if σ was a positive constant then the metrics g and σg give rise to the same curvature structure. They also give rise to the same sequence of curvature derivatives. Thus with regard to the type of problem considered here, such an ambiguity is unavoidable. Suppose now that the generalized *pp* waves are excluded from the discussion and that the curvature and its derivatives are prescribed up to that point where the ambiguity in the metric is best possible in the sense described above (so that, at most, second-order derivatives are required). Then there is a sense in which the above unavoidable ambiguity is the only one up to "coordinate transformations." This is, of course, clear for cases (c) and (d) [two constant conformal factors being required in case (c), one for each of the two families of submanifolds of decomposition]. The required coordinate transformation is clear in case (a) where F is non-null and where no other bivectors are introduced by the first curvature derivatives. In the other cases, the ambiguity is either a constant conformal factor only or takes the general form 2(b). If k is non-null in 2 (b) then the ambiguity can be simulated, apart from the constant factor ϕ, by a coordinate transformation $x'^{\alpha} = x^{\alpha}$, $u' = f(u)$ with f a suitable function [see Eqs. (8) and (9)]. If k is null in 2(b) similar comments hold, this time using the coordinate transformation $x'^{\alpha} = x^{\alpha}$, $x'^{0} = x^{0} + g(x^{3})$ for some suitable function g [see Eqs. (12)]. Such coordinate transformations merely reflect the existence of curvature collineations in M which, whilst preserving the curvature structure (and its derivatives), generate the metric ambiguities (cf. Ref. 1).

ACKNOWLEDGMENTS

Both authors wish to thank Dr. J. R. Pulham and Dr. I. G. Craw in Aberdeen for many useful discussions.

One of the authors (W. K.) gratefully acknowledges the award of a Science and Engineering Research Council studentship.

[1]G. S. Hall, Gen. Relativ. Gravit. **15**, 581 (1983).
[2]G. S. Hall, J. Phys. A **9**, 541 (1976).
[3]A. G. Walker, Q. J. Math. (Oxford) **20**, 135 (1949).
[4]J. Ehlers and W. Kundt in *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962).
[5]R. Debever and M. Cahen, Bull. Acad. Belg. Cl. Sci. **47**, 491 (1962).
[6]J. Leroy and R. G. McLenaghan, Proc. R. Soc. London Ser. A **327**, 229 (1972).
[7]G. S. Hall, J. Phys. A. **10**, 29 (1977).
[8]D. Kramer, H. Stephani, M. A. H. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (V.E. B. Deutscher Verlag der Wissenschaften, Berlin, 1980).
[9]E. M. Patterson, Q. J. Math. (Oxford) **2**, 151 (1951).
[10]G. S. Hall, in *Classical General Relativity*, edited by W. Bonnor *et al.* (Cambridge U.P., Cambridge, 1984).
[11]G. S. Hall and C. B. G. McIntosh, Int. J. Theor. Phys. **22**, 469 (1983).
[12]N. Rosen, Ann. Phys. (NY) **22**, 1 (1963).
[13]S. Tanno, Tohuku Math. J. **19**, 245 (1967).
[14]C. B. G. McIntosh and E. H. van Leeuwen, J. Math. Phys. **23**, 1149 (1982).
[15]G. S. Hall, Class. Quant. Gravit. **1**, 545 (1984).
[16]W. Kay, Ph.D. thesis, University of Aberdeen, 1987.
[17]G. S. Hall and W. Kay, J. Math. Phys. **29**, 428 (1988).
[18]G. S. Hall, Gen. Relativ. Gravit., to appear.

Holonomy groups in general relativity

G. S. Hall and W. Kay
Department of Mathematics, University of Aberdeen, Edward Wright Building, Dunbar Street, Aberdeen, AB9 2TY, Scotland

(Received 24 April 1987; accepted for publication 30 September 1987)

The infinitesimal holonomy group structure of space-time is discussed and related to the Petrov type of the Weyl tensor and the algebraic (Segre) type of the energy-momentum tensor. The number of covariant derivatives of the curvature tensor required to determine the infinitesimal holonomy group is determined in each case and the complete classification scheme is tabulated. Some special cases of physical interest are investigated in more detail. A geometrical approach is followed throughout.

I. INTRODUCTION

The purpose of this paper is to give a straightforward and complete solution to the problem of classifying gravitational fields in general relativity with respect to their holonomy group. A full list of holonomy groups is given together with the algebraic details of the associated Weyl tensor (Petrov type) and the energy-momentum tensor (Segre type) and they are displayed in Table I. Previous papers have dealt with holonomy group classifications[1–7] but none are complete in the above sense and some contain cases that cannot occur. None of them gives a satisfactory account of the full implications for the energy-momentum tensor.

Throughout, (M,g) will denote a space-time with g a smooth Lorentz metric on the smooth space-time manifold M. The tangent space to M at $p\in M$ is denoted by T_pM and the curvature tensor on M is represented by its coordinate components $R^{a}{}_{bcd}$. At all points considered here, the curvature will be supposed nonzero. The symbol $\mathscr{L}$ will denote the proper (component of the identity of the) Lorentz group. Segre notation for the Ricci tensor types is taken from Refs. 8 and 9 whilst the results in Ref. 10 will also be drawn upon. Einstein's equations with zero cosmological constant are assumed.

II. HOLONOMY GROUPS

The concept of holonomy is, perhaps, best described in the frame bundle of M.[11] However, it will suffice for the present purposes to describe it within M, following Ref. 11. The Lorentz group $\mathscr{L}$ acts on T_pM as a Lie transformation group in an obvious way as does any Lie subgroup of $\mathscr{L}$. Consider the vector space of linear transformations spanned by the matrices $R^{a}{}_{bcd}X^cY^d$, $R^{a}{}_{bcd;e}X^cY^dZ^e$,..., all evaluated at $p\in M$, where $X,Y,Z\in T_pM$ and where a semicolon denotes a covariant derivative. This set has a natural structure of a Lie algebra that makes it a subalgebra of the Lie algebra of the Lorentz group. The unique connected Lie subgroup of $\mathscr{L}$ that arises naturally from it is called the *infinitesimal holonomy group of M at p* and denoted by $\Phi'(p)$. The dimension of this subalgebra is the dimension of $\Phi'(p)$ and if it can be spanned by matrices of the form $R^{a}{}_{bcd}X^cY^d$ only, the associated infinitesimal holonomy group is called *perfect*. The dimension of $\Phi'(p)$ is, of course, restricted by lower semicontinuity in the sense that for each integer m, the subset $\{p\in M: \dim \Phi'(p)\geqslant m\}$ is open in M. However, in this paper it will be assumed that one works in an open, connected, simply connected (and, where required, contractible) chart domain U of M in which $\dim \Phi'(p)$ is constant. In this case the group $\Phi'(p)$ can be realized in another way. For each $p\in U$ let $C(p)$ denote the set of all closed C^k $(1\leqslant k\leqslant\infty)$ curves in U with initial and final end point p. The set of all linear transformations $T_pM\to T_pM$ obtained by the parallel transport of each member of T_pM around $c\in C(p)\,\forall c\in C(p)$ constitutes a subgroup $\Phi(p)$ of the Lorentz group $\mathscr{L}$ that is independent of the differentiability index k and is called the *holonomy group of U* at p. The constancy of $\dim \Phi'(p)$ over U implies that the groups $\Phi(p)$ and $\Phi'(p)$ are equal at each $p\in U$. [It is, in fact, a consequence of the assumed constancy of $\dim \Phi'(p)$ over U and the simply connected and connected (and hence path-connected) properties of U that the holonomy, restricted holonomy, local holonomy, and infinitesimal holonomy groups of U at each $p\in U$ are equal and at distinct points of U are isomorphic.]

The Lie algebra of $\mathscr{L}$ can and will be identified with the six-dimensional vector space of bivectors in Minkowski space, where the Lie bracket operation is the usual commutation. Starting with a real null tetrad (l,n,x,y) in Minkowski space a basis for this Lie algebra is, in an obvious notation, $l\wedge x$, $l\wedge y$, $l\wedge n$, $x\wedge y$, $n\wedge x$, and $n\wedge y$. The connected subgroups of $\mathscr{L}$ can then be specified by listing the subalgebras of this Lie algebra (see, for example, Ref. 12) and there are fifteen convenient types for these subalgebras. The corresponding subgroups of $\mathscr{L}$ will, following Ref. 1, be labeled R_1–R_{15} and are defined in Table I.

Now let $p\in M$ and let A be the Lie algebra of bivectors associated with $\Phi'(p)$ $[=\Phi(p)]$. If $k\in T_pM$ is an eigenvector (with respect to the metric g at p) of each member of A then the one-dimensional subspace of T_pM spanned by k is invariant under each transformation of $\Phi'(p)$ and if, also, the eigenvalues to k are all zero, then k is a fixed point of each member of $\Phi'(p)$. Of course, if k possesses any nonzero eigenvalue, it is necessarily a null vector. The subgroups of $\mathscr{L}$ can then be described in terms of their invariant subspaces in T_pM.[12] In the present case the equality of $\Phi'(p)$ and $\Phi(p)$ at all $p\in U$ enables these results to be seen geometrically by means of parallel translation. In particular, if V is an invariant subspace of $\Phi'(p)$ then the parallel transport of V at any other point $q\in U$ is independent of the path used and produces an integrable distribution on U. So, for example, if

 0022-2488/88/020428-05$02.50

the holonomy group is R_{14} (see Table I) then one has a recurrent null vector field l on U ($l_{a;b} = l_a p_b$ for some vector field p^a on U) whilst if the holonomy group is R_3, R_4, R_8, or R_{11} a covariantly constant, null vector field l is admitted.

Regarding the holonomy group in the form $\Phi'(p)$, one constructs the Lie algebra from the curvature by accumulating "generating bivectors" for this algebra from the curvature tensor and its successive covariant derivatives. It is useful here to note that if at any stage the kth covariant derivative of the curvature produces no bivectors other than those already accumulated in the previous stages up to the $(k-1)$th covariant derivative, then the procedure is complete in the sense that all higher-covariant derivatives of the curvature will produce only bivectors linearly dependent on those already obtained. If such is the case then the bivectors produced must span a subalgebra of the Lorentz algebra that then gives the associated holonomy group. To see this note that for any $X,Y,Z,T \in T_pM$,

$$R^{a}{}_{bcd;e\cdots fg}X^cY^d\cdots Z^fT^g = (R^{a}{}_{bcd;e\cdots f}X^cY^d\cdots Z^fT^g)_{;g} - R^{a}{}_{bcd;e\cdots f}(X^cY^d\cdots Z^fT^g)_{;g}. \quad (1)$$

If one regards the left-hand side of (1) as representing a bivector obtained from the $(k+1)$th curvature derivative, then the first term on the right-hand side of (1) may be written as $(P^a{}_bT^g)_{;g}$, where the bivector $P^a{}_b$ arises from the kth curvature derivative. By the assumption $P^a{}_b$ may then be written as a linear combination of bivectors arising from curvature derivatives of order $\leqslant k-1$ and as a consequence all the terms on the right-hand side of (1) are linear combinations of bivectors that have been produced up to the kth and hence the $(k-1)$th curvature derivatives. Repeating this argument then easily shows that all bivectors produced at any higher-order curvature derivative than the kth is a linear combination of those already produced up to the $(k-1)$th derivative. This, together with the remark that the above result in no way depended on the dimension of the manifold in question or on the signature of the metric (nor even on the fact that a metric compatible with the curvature existed), completes the proof. This result is useful in calculations because it shows that if, at a certain stage in the above procedure the bivectors obtained do not form a subalgebra, then the next curvature derivative must contribute bivectors that are independent of those already generated.

III. CURVATURE STRUCTURE AND INFINITESIMAL HOLONOMY GROUPS

This is the main section of the paper and its aim is to give a complete description of infinitesimal holonomy groups in general relativity and to show how they are related to the Weyl and energy-momentum tensors of space-time. This description also enables one to see that one need never go beyond the second curvature derivatives in order to determine this group and, in fact, apart from some very special cases, the first derivatives suffice. [There is a precise sense in which the curvature components alone determine the group in the "general case" and this is discussed more fully in Sec. IV (vii).] The work will proceed in the chart domain U described in the previous section that is restricted by the assumption of the constancy of dim $\Phi'(p)$ and another technical restriction that will be discussed later.

Suppose then that the curvature components $R^{a}{}_{bcd}$ are given in the coordinate domain U and are supposed to arise from some Lorentz metric g on U. The metric g is not supposed given and it is known that the curvature components do not necessarily determine the metric g or the associated (symmetric) connection uniquely. However the curvature components impose severe restrictions on the metric and connection.[13] The *rank* of the curvature at $p \in U$ is the number of independent bivectors contributed to the Lie algebra of $\Phi'(p)$ from the curvature components alone, that is, bivectors of the form $R^{a}{}_{bcd}X^cY^d$, $X, Y \in T_pM$. The following cases can be distinguished.[13–15]

(a) Rank 1: Here the contributions to the holonomy from the curvature at p are multiples of a (necessarily simple) bivector F.

(b) Rank 2 or 3; but where all the bivectors contributed to the holonomy from the curvature at p have a unique common eigendirection $k \in T_pM$ with eigenvalue zero (and are hence all simple).

(c) Rank 2; but where the bivectors contributed by the curvature can be spanned by two simple bivectors that form an orthogonal timelike–spacelike pair with respect to any metric compatible with the curvature.

(d) All other cases: A study of the metric ambiguities in each case then shows[15] that the following are independent of the metric that gave rise to the curvature structure: (i) the nature (spacelike, timelike, or null) of the bivector F in case (a), (ii) the nature (spacelike, timelike, or null) of k in case (b) and the nature of the bivectors arising from the curvature in this case, and (iii) the fact that the (unique up to scaling) pair of simple bivectors that arise from the curvature in case (c) form an orthogonal, timelike–spacelike pair of bivectors. These results arise from the algebraic restrictions $g'_{e(a}R^{e}{}_{b)cd} = 0$ that any alternative metric g' must satisfy and the general solution of this equation for each of the above cases is given in Refs. 13–15. Similarly, if the curvature derivatives are specified one has the analogous conditions $g'_{e(a}R^{e}{}_{b)cd;f} = 0$. The extra specification of curvature derivatives leads, in general, to more bivector contributions to the holonomy and, by the above relation, restricts the alternative metrics further[10] and in such a way that the nature (in the above sense) of any additional bivectors so generated is independent of the alternative metric. Similar comments apply to higher-order curvature derivatives.

The technical restrictions mentioned earlier can now be discussed. It will be assumed that exactly one of the cases (a)–(d) holds throughout U. This is not so severe a restriction as it may seem and will be discussed in more detail elsewhere. It will also be clear that certain conditions of a similar nature are demanded of the curvature derivatives.

Suppose now that condition (d) holds in U. Then it follows[13] that the metric corresponding to this curvature is determined up to a constant conformal factor and hence that the connection is uniquely determined. Thus the infinitesimal holonomy group is uniquely determined at each $p \in U$ since the holonomy group is (see Sec. II). Since condition (d) excludes the possibility of a nontrivial solution $r \in T_pM$

of the equation $R^{a}{}_{bcd}r^{d}=0$ at any $p\in U$ (Refs. 13–15) it follows from the second column of Table I that the group in this case is either R_9, R_{12}, R_{14}, or R_{15}. The case R_9 is necessarily perfect. If condition (c) holds in U the curvature components determine the group uniquely and it is R_7 (perfect). If condition (b) holds throughout U then either the curvature first derivatives introduce only bivectors for which k is an eigenvector with zero eigenvalue, or other bivectors are introduced. In the former case a covariantly constant vector field proportional to k is admitted and the group is uniquely determined, being R_6 (perfect) or R_{10} [$\approx$SO(1,2)] if k is spacelike in U or R_{13} [$\approx$SO(3)] if k is timelike in U. If k is null in U the group is either R_8 (perfect) or R_{11} ($\approx \mathbf{E}_2$, the Euclidean group of the plane) according as the bivectors contributed by the curvature are all null or not all null. In the latter case the group is either R_9 or R_{15} and, again, the curvature and curvature first derivatives distinguish them. If condition (a) holds throughout U then either no new bivectors are introduced by the curvature first derivatives, in which case the group is either R_4 (for F spacelike), R_3 (for F null), or R_2 (for F timelike), or else other bivectors are introduced. In the latter case one specifies the first and second curvature derivatives and either the accumulated bivectors admit a unique common eigendirection with zero eigenvalue or they do not. In the first of these, the situation is as in case (b) with a covariantly constant vector field admitted and the group is then uniquely determined as either R_{10}, R_{11}, or R_{13} (F spacelike), R_6 or R_{10} (F timelike), or R_8 or R_{10} (F null). In the second of these, the first and second curvature derivatives uniquely determine the group, which is either R_9 or R_{15}.

The proof can be gathered together from the results of Ref. 10 and the one at the end of the last section using Table I as a guide. The following comments may be helpful. For each of the nontrivial infinitesimal holonomy groups R_2–R_{11} and R_{13} it can be shown, by performing null rotations or similar transformations, that in each case the canonical set of bivectors for the group given in the second column of Table I can be chosen in such a way that a bivector basis for the span of the curvature is as in column 3. Similar entries for column 3 for the R_{12} and R_{14} cases can be given but they are more complicated and are omitted (the algebraic structure in the R_{12} case is, in any case, independent of the curvature rank). The second column enables the Ricci tensor and its Segre type to be calculated. Also, if U admits a recurrent, null vector field l that is not scalable to a constant ($l_{a;b}=l_aP_b$ with P_a not a gradient) then $R^{a}{}_{bcd}l^{d}\neq 0$ and so this restricts the set of bivectors that are contributed by the curvature alone. Also, in this case, any bivector of the form $l\wedge q$ will, upon covariant differentiation, introduce only bivectors of the form $l\wedge q'$ for covectors q and q'. The group in this case is either R_2, R_6, R_7, R_9, R_{12}, or R_{14}. If only $l\wedge n$ and $x\wedge y$ occur in the curvature tensor the Bianchi identities show that one necessarily has the group R_7.[10] If a covariantly constant null vector field l is admitted then $R^{a}{}_{bcd}l^{d}=0$, $R^{a}{}_{bcd;e}l^{d}=0$, etc., and the group is R_3, R_4, R_8, or R_{11}. In the R_{11} case the bivector $x\wedge y$ must occur in the curvature tensor. Another useful comment is that if the curvature tensor and its first covariant derivatives contain only bivectors for which a particular null vector field l is an eigenvector then l is necessarily recurrent. This follows by covariantly differentiating an appropriate expression for the curvature that contains a bivector F for which l is an eigenvector and then showing that, for any vector P, $F_{ab;c}P^{c}$ is a linear combination of bivectors each of which admits l as an eigenvector. Finally if one has case (a) with F represented by $x\wedge y$ and if all the bivectors that arise from the first curvature derivatives have l as a null eigenvector with zero eigenvalue, then l may be scaled to a covariantly constant null vector field. This follows by covariantly differentiating the relation $R^{a}{}_{bcd}l^{d}=0$ and obtaining the equation $l_{a;b}=l_ar_b+n_as_b$ for covector fields r and s. A contraction with l^{a} then shows that $s=0$ and so l is recurrent and hence scalable to a constant since $R^{a}{}_{bcd}l^{d}=0$.

This information and more is collected together in Table I, where the infinitesimal holonomy groups are listed together with their Lie algebras and the corresponding Petrov type of the Weyl tensor and Segre type of the Ricci tensor. For these calculations, the table in Ref. 15 is relevant (but note the remark in the caption to Table I) as is the comment that if a recurrent null vector field is admitted then the Weyl tensor is algebraically special and, in particular is of type III, N, or O if and only if the Ricci scalar $R=0$. For the computation of the algebraic structure in Table I it is useful to note that, if $R^{a}{}_{bcd}k^{d}=0$ at p for some non-null $k\in T_pM$, then by working in the three-space of T_pM orthogonal to k one can exhibit a simple relationship between the Segre type of the Riemann tensor (regarded in the usual way as a linear transformation on the six-dimensional vector space of bivectors), the Segre type of the Ricci tensor (either through its eigenvector structure or the eigenbivector structure of the anti-self-dual part of the Riemann tensor[9]) and the Petrov type (through the eigenbivector structure of the Weyl tensor).[15,16] The proof can be accomplished either by direct calculation or by considering the three-dimensional duality relation in the three-space of T_pM orthogonal to k.

IV. FURTHER COMMENTS

(i) For a null Maxwell field, the Segre type of the Ricci tensor is $\{(211)\}$ with zero eigenvalue and so the only possibilities for the infinitesimal holonomy group in this case (with the corresponding Petrov types in brackets) are R_3 (N), R_8 (N or O), R_{10} (N), R_{14} (III), and R_{15} (algebraically special and not conformally flat). This mostly follows from Table I, it being easily shown that the Ricci eigenvalue is zero in these cases. In the R_{14} case, the Petrov type I is ruled out (since all null Maxwell fields are algebraically special from the Goldberg–Sachs theorem, see, for example, Ref. 17) the Petrov types II and D are impossible because of the existence of a recurrent null vector field, together with the equation $R=0$, and the Petrov types N and O are impossible because in these cases one necessarily has $R_{abcd}l^{d}=0$ for the recurrent Maxwell principal null direction l. (The curvature rank is, in fact, necessarily 4 as is then easily seen on algebraic grounds from the canonical form of its (necessarily type III) Weyl tensor). Since the principal null direction of a conformally flat null Maxwell field can be scaled to a covariantly constant vector field (see, for example, Ref. 17) this case is incompatible with R_{15}. These results agree with those in Refs. 4 and 7.

TABLE I. The first column gives the group label, following the convention in Ref. 1, but with the trivial subgroup R_1 of $\mathscr{L}$ and the full group $R_{15} \equiv \mathscr{L}$ omitted. The second column gives a basis of bivectors for the corresponding Lie algebra. The third column gives the rank of the curvature tensor and the bivectors it contributes to the corresponding Lie algebra. The fourth and fifth columns are the associated Segre type of the Ricci (or energy-momentum) tensor and the Petrov type. The set $\{l,n,x,y\}$ always denotes a null tetrad whilst, in R_{13}, $\{x,y,z\}$ is an orthonormal triad of spacelike vectors and in R_{10} u is a timelike and z a spacelike vector orthogonal to each other and to x. No attempt is made to list the possibilities for R_{14} since many pairings for the Segre and Petrov types seem to be consistent. Suffice it to say, in this case, that a recurrent null vector field l is admitted, that the Petrov type is algebraically special with l as a principal null direction and that l is a null Ricci eigendirection. This last fact means that the corresponding Segre type is $\{(1,1)11\}$, $\{2,11\}$, $\{31\}$ or a degeneracy thereof. The comments in Sec. III show that for R_{14}, the curvature rank is not 1. The entries in the R_{10} (rank 2 and 3 cases) may be taken as a correction to the entries in the "(ii) spacelike" case of Table 2 in Ref. 15. Finally, it is not claimed that all the possibilities listed here can actually exist.

Holonomy group				
Label	Bivectors	Curvature rank/ bivectors	Segre type	Petrov type
R_2	$l\wedge n$	1	$\{(1,1)(11)\}$	D
R_3	$l\wedge x$	1	$\{(211)\}$	N
R_4	$x\wedge y$	1	$\{(1,1)(11)\}$	D
R_5	$l\wedge n+\rho(x\wedge y)$	Impossible		
R_6	$l\wedge n, l\wedge x$	$1(l\wedge n)$	$\{(1,1)(11)\}$	D
R_6	$l\wedge n, l\wedge x$	2	$\{(31)\}$	III
R_6	$l\wedge n, l\wedge x$	2	$\{2(11)\}$	II
R_7	$l\wedge n, x\wedge y$	2	$\{(1,1)(11)\}$ or $\{(1111)\}$	D or O
R_8	$l\wedge x, l\wedge y$	$1(l\wedge x)$	$\{(211)\}$	N
R_8	$l\wedge x, l\wedge y$	2	$\{(211)\}$	N or O
R_9	$l\wedge n, l\wedge x, l\wedge y$	$1(l\wedge n)$	$\{(1,1)(11)\}$	D
R_9	$l\wedge n, l\wedge x, l\wedge y$	$2(l\wedge n, l\wedge x)$	$\{(31)\}$	III
R_9	$l\wedge n, l\wedge x, l\wedge y$	$2(l\wedge n, l\wedge x)$	$\{2(11)\}$	II
R_9	$l\wedge n, l\wedge x, l\wedge y$	3	$\{(31)\}$	III
R_9	$l\wedge n, l\wedge x, l\wedge y$	3	$\{2(11)\}$	II or D
R_9	$l\wedge n, l\wedge x, l\wedge y$	3	$\{(1,1)(11)\}$	II or D
R_{10}	$l\wedge n, l\wedge x, n\wedge x$	$1(l\wedge n)$	$\{(1,1)(11)\}$	D
R_{10}	$l\wedge n, l\wedge x, n\wedge x$	$1(l\wedge x)$	$\{(211)\}$	N
R_{10}	$l\wedge n, l\wedge x, n\wedge x$	$1(x\wedge z)$	$\{(1,1)(11)\}$	D
R_{10}	$l\wedge n, l\wedge x, n\wedge x$	$2(l\wedge n, l\wedge x)$	$\{(31)\}$	III
R_{10}	$l\wedge n, l\wedge x, n\wedge x$	$2(l\wedge n, l\wedge x)$	$\{2(11)\}$	II
R_{10}	$l\wedge n, l\wedge x, n\wedge x$	$2(l\wedge x, n\wedge x)$	$\{1,111\}$ or $\{z\bar{z}11\}$ or $\{z\bar{z}(11)\}$	I
R_{10}	$l\wedge n, l\wedge x, n\wedge x$	$2(l\wedge x, n\wedge x)$	$\{(1,1)11\}$ or $\{1,1(11)\}$	I or D
R_{10}	$l\wedge n, l\wedge x, n\wedge x$	$2(l\wedge x, n\wedge x)$	$\{211\}$	II
R_{10}	$l\wedge n, l\wedge x, n\wedge x$	$2(u\wedge x, u\wedge z)$	$\{1,111\}$	I
R_{10}	$l\wedge n, l\wedge x, n\wedge x$	$2(u\wedge x, u\wedge z)$	$\{(1,1)(11)\}$	D
R_{10}	$l\wedge n, l\wedge x, n\wedge x$	$2(u\wedge x, u\wedge z)$	$\{(1,1)11\}$ or $\{1,1(11)\}$	I or D
R_{10}	$l\wedge n, l\wedge x, n\wedge x$	3	$\{(1,11)1\}$	O
R_{10}	$l\wedge n, l\wedge x, n\wedge x$	3	$\{1,111\}$ or $\{(1,1)11\}$ or $\{1,1(11)\}$ or $\{z\bar{z}11\}$ or $\{z\bar{z}(11)\}$	I
R_{10}	$l\wedge n, l\wedge x, n\wedge x$	3	$\{1,1(11)\}$ or $\{(1,1)11\}$ or $\{(1,1)(11)\}$ or $\{1,(111)\}$ or $\{(1,11)1\}$	D
R_{10}	$l\wedge n, l\wedge x, n\wedge x$	3	$\{211\}$ or $\{(21)1\}$ or $\{2(11)\}$	II
R_{10}	$l\wedge n, l\wedge x, n\wedge x$	3	$\{31\}$ or $\{(31)\}$	III
R_{10}	$l\wedge n, l\wedge x, n\wedge x$	3	$\{(21)1\}$	N
R_{11}	$l\wedge x, l\wedge y, x\wedge y$	$1(x\wedge y)$	$\{(1,1)(11)\}$	D
R_{11}	$l\wedge x, l\wedge y, x\wedge y$	$2(l\wedge x, x\wedge y)$	$\{(31)\}$	III
R_{11}	$l\wedge x, l\wedge y, x\wedge y$	$2(l\wedge x, x\wedge y)$	$\{2(11)\}$	II or D
R_{11}	$l\wedge x, l\wedge y, x\wedge y$	3	$\{(31)\}$	III
R_{11}	$l\wedge x, l\wedge y, x\wedge y$	3	$\{2(11)\}$	II or D
R_{11}	$l\wedge x, l\wedge y, x\wedge y$	3	$\{(1,1)(11)\}$	II or D
R_{12}	$\{l\wedge x, l\wedge y, l\wedge n+\rho(x\wedge y)\}$	>2	$\{(31)\}$	III
R_{13}	$x\wedge y, y\wedge z, x\wedge z$	$1(x\wedge y)$	$\{(1,1)(11)\}$	D
R_{13}	$x\wedge y, y\wedge z, x\wedge z$	$2(x\wedge y, y\wedge z)$	$\{1,111\}$	I
R_{13}	$x\wedge y, y\wedge z, x\wedge z$	$2(x\wedge y, y\wedge z)$	$\{1,1(11)\}$	D
R_{13}	$x\wedge y, y\wedge z, x\wedge z$	3	$\{1,111\}$ or $\{(1,1)11\}$	I
R_{13}	$x\wedge y, y\wedge z, x\wedge z$	3	$\{1,1(11)\}$ or $\{(1,1)(11)\}$ or $\{(1,11)1\}$	D
R_{13}	$x\wedge y, y\wedge z, x\wedge z$	3	$\{1,(111)\}$	O
R_{14}	$\{l\wedge n, x\wedge y, l\wedge x, l\wedge y\}$	>2	$\{(1,1)11\}$ or $\{211\}$ or $\{31\}$ or a degeneracy of one of these types	algebraically special

(ii) For a non-null Maxwell field, the Segre type of the Ricci tensor is $\{(1,1)(11)\}$ where the nonzero eigenvalues differ only in sign. The infinitesimal holonomy groups allowed in this case, together with their Petrov types, are R_7 (O), R_{14} (III or N), and R_{15}. (The R_9 possibility is ruled out since it has $R \neq 0$ and the others cannot occur since they have a zero Ricci eigenvalue.) In the R_7 case the curvature rank is 2 whereas it is 4 in the R_{14} (N) case and 3 or 4 in the R_{14} (III) case, as follows from elementary algebraic considerations. These results agree with those in Refs. 4 and 7.

(iii) For perfect fluids with $p \geqslant 0$, $\rho \geqslant 0$, where p is the isotropic pressure and ρ the proper density, the Segre type of the Ricci tensor is $\{1,(111)\}$ with the timelike eigenvalue nonzero and distinct from the spacelike one. Hence null Ricci eigenvectors are forbidden and the infinitesimal holonomy group and Petrov type are either R_{10} (D) or R_{15} (all Petrov types), the curvature rank being 3 in the R_{10} case. If it is required that $p \neq \rho$, then the spacelike eigenvalue is nonzero and only R_{15} survives.

(iv) The vacuum cases are not included in Table I. They are however well known.[1-4] The self-duality of the vacuum Riemann tensor forces the rank of the curvature tensor and the dimension of the infinitesimal holonomy group to be even and the only possibilities, and their Petrov types, are R_8 (N), R_{14} (III), and R_{15} (all Petrov types) with curvature ranks 2 (for R_8), 4 (for R_{14}), and 2, 4, or 6 (for R_{15}), respectively. The curvature components and their first covariant derivatives suffice to determine the holonomy group.

(v) For a (nonvacuum) Einstein space the self-duality of the curvature tensor again forces the rank of the curvature tensor and the dimension of the infinitesimal holonomy group to be even. This, together with the absence of any vector field k satisfying $R_{abcd}k^d = 0$ (since this would imply $R = 0$) means that the only possibilities are R_7, R_{14}, and R_{15}. For the group R_7, the Petrov type must be D because if it were O, one would have a space of constant curvature and hence a contradiction since the group R_7 has dimension 2. The group R_{14} cannot occur as can be seen from the following argument. Let l be a recurrent null vector field (which is necessarily admitted) and note that, since $R \neq 0$, the Petrov type is II or D. The Ricci identity on l and the Einstein space condition imply that

$$R_{abcd}l^d = G_{ab}l_c \quad (R_{abcd} = C_{abcd} + \tfrac{1}{6}Rg_{a[c}g_{d]b}), \quad (2)$$

where G is some bivector. Upon substituting into (2) the canonical forms for the Weyl tensor C_{abcd} for Petrov types II and D (see, for example, Ref. 18) and performing an obvious contraction, one obtains the contradiction $R = 0$. Hence a nonvacuum Einstein space can either have the infinitesimal holonomy group R_7 (curvature rank 2 and Petrov type D) or R_{15} (all Petrov types).

(vi) The group R_5 cannot occur as an infinitesimal holonomy group since it is generated by a nonsimple bivector, in contradiction to the relation $R^a{}_{[bcd]} = 0$. (Similar comments, together with some remarks in Sec. III, show that R_{12} cannot have curvature rank 1.) If one imposes the dominant energy conditions (see for example, Ref. 19) then R_{12} is ruled out because in this case the Segre type of the Ricci (and hence energy-momentum) tensor is necessarily $\{(31)\}$ and this type can never satisfy the dominant energy conditions.[8,9] Hence if these conditions are imposed, the Lie algebra of the holonomy group can always be spanned by simple bivectors.

(vii) Returning to the remark made at the beginning of Sec. III, consider the topological space $\Gamma^r(L)$ consisting of the set L of C^r Lorentz metrics on M ($r \geqslant 3$) together with the Whitney C^r topology (further details on jet bundles and Whitney topologies can be found in Ref. 20). Then there exists an open dense subset W of $\Gamma^r(L)$ such that the curvature associated with each $g \in W$ belongs to the case (d) defined earlier at each $p \in M$.[21] In this sense one can say that in the general case the curvature components determine the metric (up to a constant conformal factor), the connection and hence the infinitesimal holonomy group at each $p \in M$. In fact W can be chosen so that the infinitesimal holonomy group is necessarily equal to $\mathscr{L}$ ($\equiv R_{15}$) at each $p \in M$ as follows from a result in Ref. 21 [and using the fact that the local holonomy group $\Phi^*(p)$ of M at p, which is a Lie group containing $\Phi'(p)$, as a connected Lie subgroup, has a dimension that satisfies the upper semicontinuity condition: $\{p \in M: \dim \Phi^*(p) \leqslant m\}$ is open in M for each integer m]. With this choice of W it does not follow that the infinitesimal holonomy group resulting from a metric in W is perfect at each $p \in M$ but it does follow that it is perfect at each point of an open, dense subset of M.

ACKNOWLEDGMENTS

Both authors thank Dr. J. R. Pulham, Dr. I. G. Craw, and Dr. A. D. Rendall in Aberdeen for useful discussions.

One of the authors (W. Kay) gratefully acknowledges the award of a Science and Engineering Research Council studentship.

[1] J. F. Schell, J. Math. Phys. **2**, 202 (1961).
[2] J. N. Goldberg and R. P. Kerr, J. Math. Phys. **2**, 327 (1961).
[3] R. P. Kerr and J. N. Goldberg, J. Math. Phys. **2**, 332 (1961)
[4] R. Debever and M. Cahen, Bull. Acad. Belg. Cl. Sci. **47**, 491 (1962).
[5] W. Beiglböck, Z. Phys. **179**, 148 (1964).
[6] L. K. Norris and W. R. Davis, Nuovo Cimento B **53**, 209 (1979).
[7] C. D. Collinson and P. N. Smith, Gen. Relativ. Gravit. **19**, 95 (1987).
[8] G. S. Hall, J. Phys. A **9**, 541 (1976).
[9] G. S. Hall, in *Differential Geometry*, Banach Centre Publications (Polish Scientific, Warsaw, 1984), Vol. 12, pp. 53–73.
[10] G. S. Hall and W. Kay, J. Math. Phys. **29**, 420 (1988).
[11] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963), Vol. I.
[12] R. Shaw, Q. J. Math. (Oxford) **21**, 101 (1970).
[13] G. S. Hall, Gen. Relativ. Gravit. **15**, 581 (1983).
[14] G. S. Hall and C. B. G. McIntosh, Int. J. Theor. Phys. **22**, 469 (1983).
[15] G. S. Hall, Class. Quant. Gravit. **1**, 545 (1984).
[16] W. Kay, Ph.D. thesis, University of Aberdeen, 1987.
[17] D. Kramer, H. Stephani, M. A. H. MacCallum and E. Herlt, *Exact Solutions of Einstein's Field Equations* (V.E.B. Deutscher Verlag der Wissenschaften, Berlin, 1980).
[18] R. K. Sachs, Proc. R. Soc. London **264**, 309 (1961).
[19] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space Time* (Cambridge U.P., Cambridge, 1963).
[20] D. E. Lerner, Commun. Math. Phys. **32**, 19 (1973).
[21] A. Rendall, J. Math Phys. (to appear).

On the GKP and BS constructions of the *c*-boundary

Kuang zhi-quan
Center of Theoretical Physics, CCAST (World Laboratory), Beijing, People's Republic of China and Institute of Mathematics, Academia Sinica, Beijing, People's Republic of China

Liang can-bin
Center of Theoretical Physics, CCAST (World Laboratory), Beijing, People's Republic of China and Department of Physics, Beijing Normal University, Beijing, People's Republic of China

(Received 11 August 1987; accepted for publication 14 October 1987)

Two examples are presented in this paper, the first is unfavorable to the *c*-boundary construction given by Geroch, Kronheimer, and Penrose [Proc. R. Soc. London Ser. A **327**, 545 (1972)] but in favor of that given by Budic and Sachs [J. Math. Phys. **15**, 1302 (1974)], while the second plays an opposite role. The second example is also an example of a causally continuous space-time with a "really big gap," contrary to what was believed in the literature.

I. INTRODUCTION

In order to have a better description of space-time singularities within the framework of classical general relativity, one would like to construct an enlarged topological space $\bar{M}$ interpreted as the space-time manifold M with some singular boundary ∂ attached. Various constructions have been put forward. The constructions of the *b*-boundary[1] and the *g*-boundary[2] have been known to be unsatisfactory.[3,4] The construction known as the *c*-boundary (causal boundary) construction given by Geroch, Kronheimer, and Penrose[5] (GKP) in 1972 makes use only of the causal structure of the space-time and hence has certain merit from the physical point of view. However, as illustrated by its authors, it fails to construct a Hausdorff topological space $\bar{M}$ that is also a causal space in general. To surmount this difficulty, Budic and Sachs (BS) gave an improved definition of the *c*-boundary construction in 1974.[6] They proved that the resulting Hausdorff topological space $\bar{M}$ is also a causal space with causal structure extended from that of the original space-time (M,g) itself, provided that (M,g) is causally continuous (a causal requirement much stronger than distinguishing required by Ref. 5); thus it makes good sense to ask whether signals with speed less than or equal to that of light can be sent between a regular point and an ideal point. We will refer to the *c*-boundary construction given in Ref. 5 as the GKP construction and that given in Ref. 6 as the BS construction. In a recent paper by Kuang, Li, and Liang,[7] it was shown that for some singular exact solutions to Einstein equations the *c*-boundary of the GKP construction is unsatisfactory, for example, the "singular portion" of the *c*-boundary of Taub's plane-symmetric vacuum solution turned out to be a single point, suggesting that it might not be fruitful describing the structure of singularities using the notion of the *c*-boundary defined by GKP. Besides, as will be shown in the next section, there is something else that is also unfavorable to the GKP construction. The fact that these two deficiencies do not exist in the BS construction suggests that the BS construction might be more acceptable. Nevertheless, we will give an example in Sec. III showing that there is also something unfavorable to it, a drawback which is not shared by the GKP construction. Therefore it seems still an open question whether one can construct some improved *c*-boundary free of deficiencies.

II. A SECOND EXAMPLE UNFAVORABLE TO THE GKP CONSTRUCTION

Assuming the reader is familiar with the GKP construction, we present the example as follows.

Let $(\tilde{M},\eta)$ be a three-dimensional Minkowski space-time with Cartesian coordinates (t,x,y) and (M,η) a sub-space-time where $M=\{(t,x,y): y>0\}$. Consider a future directed timelike curve $\gamma\subset M$ with the origin (0,0,0) as its future end point in $\tilde{M}$ and a past directed timelike curve $\lambda\subset M$ with (0,0,0) as past end point in $\tilde{M}$. Note that γ (resp. λ) is future (resp. past) unextendable in M. It is reasonable to require that the TIP (terminal indecomposable past set), $I^-(\gamma,M)$, and the TIF (terminal indecomposable future set), $I^+(\lambda,M)$, be identified in $\bar{M}$, and this is exactly the case according to the BS identification rule. It is, however, not true in the GKP construction. Indeed, there exist two open sets O_1 and O_2 with $I^-(\gamma,M)^*\in O_1$, $I^+(\lambda,M)^*\in O_2$, and $O_1\cap O_2=\varnothing$. To see this, consider the following two subsets of M:

$$A=\{(t,x,y): t>x, y>0\},$$
$$B=\{(t,x,y): t<x, y>0\}.$$

They are, respectively, a TIF and a TIP in M, since there exist some past (resp. future) unextendable timelike curves α (resp. β) in M such that $A=I^+(\alpha,M)$ and $B=I^-(\beta,M)$. For instance, one can take the following curve to be β: $t=t$, $x=t+1/t$, $y=1/t$ $(t>1)$, and dually for α. According to the GKP construction, the following two subsets of the in-

$$A^{\text{ext}}=\{P^*\in\mathscr{M}^{\#}: P\in\hat{\mathscr{M}} \text{ and } P=I^-(S)\Rightarrow I^+(S)\not\subseteq A \text{ for all } S\subseteq M\},$$

$$B^{\text{ext}}=\{F^*\in\mathscr{M}^{\#}: F\in\check{\mathscr{M}} \text{ and } F=I^+(S)\Rightarrow I^-(S)\not\subseteq B \text{ for all } S\subseteq M\}.$$

It is straightforward to check that $I^+(\lambda,M)^*\in B^{\text{ext}}$ and $I^-(\gamma,M)^*\in A^{\text{ext}}$ by showing that any $S\subseteq M$ with $I^+(S,M)=I^+(\lambda,M)$ [resp. $I^-(S,M)=I^-(\gamma,M)$] satisfies

 0022-2488/88/020433-03$02.50

$I^-(S,M) \not\subset B$ [resp. $I^+(S,M) \not\subset A$]. Consequently A^{ext} and B^{ext} can be taken to be the desired O_1 and O_2, respectively. Note, however, that this is not true if we are dealing with $\tilde{M}$ instead of M since the origin (0,0,0) can then be taken as S violating the requirement in the definitions of A^{ext} and B^{ext}.

III. AN EXAMPLE UNFAVORABLE TO THE BS CONSTRUCTION

We first give a brief outline of the essential contents of the BS construction relevant to this paper as follows.

Define binary relations $\succ$ and $\succ\!\!\succ$ on a time-orientable space-time (M,g) as usual. Define concepts IP and IF as in the GKP construction. Denote the power set, the topology, the collections of past sets, future sets, IP's and IF's of (M,g) as $\mathscr{S}, \mathscr{T}, \mathscr{P}, \mathscr{F}, \hat{\mathscr{M}}$, and $\check{\mathscr{M}}$, respectively. Define a map I^-: $\mathscr{S} \to \mathscr{P}$ by $I^-S = \{x \in M: x \ll s \text{ for some } s \in S\}$ $\forall S \in \mathscr{S}$. Define a map $\hat{I}$: $M \to \mathscr{P}$ by $\hat{I}x = I^-\{x\}$ $\forall x \in M$. Define a map $\downarrow$: $\mathscr{T} \to \mathscr{P}$ by $\downarrow U = I^-\{x \in M: x \ll u \ \forall u \in U\} = \text{interior}\{x \in M: x \ll u \ \forall u \in U\}$ $\forall U \in \mathscr{T}$. The maps I^+, $\check{I}$, and $\uparrow$ are defined dually. Define $\succ$ and $\succ\!\!\succ$ on $\hat{\mathscr{M}} \cup \check{\mathscr{M}}$ by Table 2.2 in Ref. 6. For example, if $P, Q \in \hat{\mathscr{M}}$, then $P \succ\!\!\succ Q$ iff $P \cap \uparrow Q \neq \emptyset$. Define an equivalence relation $\sim$ on $\hat{\mathscr{M}} \cup \check{\mathscr{M}}$ as follows: for $A,B \in \hat{\mathscr{M}}$ (or $\check{\mathscr{M}}$), $A \sim B$ iff $A = B$; for $A \in \hat{\mathscr{M}}$, $B \in \check{\mathscr{M}}$, $A \sim B$ iff $A = \downarrow B$ and $B = \uparrow A$. Define the causal completion of (M,g) as $\bar{M} = \hat{\mathscr{M}} \cup \check{\mathscr{M}} / \sim$, then $\succ$ and $\succ\!\!\succ$ are well defined on $\bar{M}$. Define the extended Alexandrov topology $\bar{\mathscr{T}}$ on $\bar{M}$ as the smallest topology on $\bar{M}$ such that for all $c \in \bar{M}$, each of the following four subcollections is open: $I^+\{c\}$, $I^-\{c\}$, $K^+\{c\} \equiv \bar{M} - J^-\{c\}$, $K^-\{c\} \equiv \bar{M} - J^+\{c\}$, where $I^+\{c\} \equiv \{a \in \bar{M}: a \succ\!\!\succ c\}$, and $J^+\{c\} \equiv \{a \in \bar{M}: a \succ c\}$. It was shown that $(\bar{M}, \succ, \succ\!\!\succ, \bar{\mathscr{T}})$ is a causal space with Hausdorff topology and $\hat{I}$: $M \to \bar{M}$ has all the desired properties (e.g., it is a dense imbedding) provided that (M,g) is causally continuous, thus the boundary ∂ is naturally interpreted as the causal boundary of (M,g).

An essential requirement for constructing a causal completion $\bar{M}$ which is both a Hausdorff topological space and a causal space is the causal continuity of (M,g). A space-time (M,g) is said to be causally continuous iff it is both distinguishing and reflective. Here (M,g) is said to be reflective iff $\uparrow \hat{I}x = \check{I}x$ and $\downarrow \check{I}x = \hat{I}x$ $\forall x \in M$. The causal continuity of space-times has been investigated in detail by some authors.[6,8–10] It was pointed out in Ref. 8 that "roughly, a causally continuous space-time... has no really big gaps (gaps of 'dimension' more than 2)" and some statements similar in spirit to it can also be found in the other references quoted. However, we have found a four-dimensional space-time (artificial though) with a "really big gap," i.e., a "gap" of four dimensions which is causally continuous. It is also this space-time to which the application of the BS construction gives some unfavorable result, as will be illustrated shortly.

Although the motivation of the BS construction was to overcome the noncooperation between the Hausdorff topology and the causal structure of the resulting space $\bar{M}$, it turns out that the two defects of the GKP construction mentioned in Secs. I and II are also surmounted. Nonetheless, the following example illustrates that it might have its own drawback.

Consider an $(n+1)$-dimensional Minkowski space-time $(\tilde{M},\eta)$. Denote the Cartesian coordinates of $\tilde{M}$ by $(t,x^1,...,x^n)$. Let $a = (-1,0,...,0) \in \tilde{M}$ and $b = (1,0,...,0) \in \tilde{M}$. By removing a closed subset $R = \text{closure}[I^+(a,\tilde{M}) \cap I^-(b,\tilde{M})]$ of the same dimension from $\tilde{M}$ we get a submanifold $M = \tilde{M} - R$ and a sub-space-time (M,η). Since two space-times $(\tilde{M},\eta)$ and (M,η) will be alternatively dealt with, we will, whenever necessary, add subscripts $\tilde{M}$ or M to the symbols for the relation $\succ\!\!\succ$ and maps $I^-, I^+, \hat{I}, \check{I}, \downarrow$, and $\uparrow$ to clarify the space-time involved. We will also write $\check{I}_{\tilde{M}}a \cap \hat{I}_{\tilde{M}}b$ instead of $I^+(a,\tilde{M}) \cap I^-(b,\tilde{M})$ to be in accordance with the BS notation. It will be proved in the next section that the sub-space-time (M,η) is causally continuous provided that $n > 1$; thus the BS construction is applicable. Let $\bar{M}$ be the causal completion of (M,η). In addition to the infinity portion ∂_i of the c-boundary ∂, there is also some singular portion ∂_s. Obviously, there is a natural correspondence between ∂R and ∂_s, hence one would intuitively expect that near ∂_s the topological structure of $\bar{M}$ should be the same as that of $\tilde{M}$, i.e., the way of "gluing" ∂_S to M should be the same as that of "gluing" ∂R to M. However, the following shows that it is not the case, thus suggesting that there might be something unsatisfactory about the BS construction.

Choose a point $e = (-\frac{1}{2}, -1,0,...,0) \in M$, then $\check{I}_M e \sim \hat{I}_M e$ is a regular point in $\bar{M}$. Let $\gamma \subset M$ be a past inextendible timelike curve which, viewed as a curve in $\tilde{M}$, has b as its past end point, then $I^+_M \gamma$ is an ideal point in $\bar{M}$. Since $\downarrow I^+_M \gamma \cap \check{I}_M e \neq \emptyset$, we have, according to the BS construction, $I^+_M \gamma \succ\!\!\succ \check{I}_M e$ or equivalently $I^+_M \gamma \in I^+\{\check{I}_M e\}$. Consider a point sequence $\{f_i\}$ in M defined by $f_i = (1,1/i,0,...,0)$, then one has a corresponding point sequence $\{F_i\}$ in $\bar{M}$ defined by $F_i = \check{I}_M f_i \in M$. Since $\downarrow \check{I}_M f_i \cap \check{I}_M e = \emptyset$, we have $F_i \notin I^+\{\check{I}_M e\}$ for any i. This, together with the fact that $I^+\{\check{I}_M e\}$ is an open set in the extended Alexandrov topology, implies that $\{F_i\}$ does not converge to $I^+_M \gamma$ in $\bar{M}$. It is, however, obvious that $\{f_i\}$ converges to b in $\tilde{M}$, therefore we conclude that the topology of $\bar{M}$ near $I^+_M \gamma$ is different from that of $\tilde{M}$.

IV. PROOF OF THE CAUSAL CONTINUITY OF $(\tilde{M}-R,\eta)$

Throughout the proof we will use the following notation: for $x \in M$ (resp. $x \in \tilde{M}$) and $S \subseteq M$ (resp. $S \subseteq \tilde{M}$), we write $x \underset{M}{\ll} S$ (resp. $x \underset{\tilde{M}}{\ll} S$) iff $x \underset{M}{\ll} s$ (resp. $x \underset{\tilde{M}}{\ll} s$) $\forall s \in S$. Dual statements (if any) to those in the following lemmas are taken for granted and are not written.

Lemma 1: Let $x,y \in M$ and $\{u_i\}$ be a sequence in M satisfying (1) $\{u_i\} \subset \check{I}x$, and (2) x is a limit point of $\{u_i\}$, then $y \ll \check{I}x$ iff $y \ll \{u_i\}$.

This lemma is true for all chronological space-times, the proof is trivial and is omitted.

To prove the causal continuity of (M,η) is to prove $\uparrow_M \hat{I}_M c = \check{I}c$ and $\downarrow_M \check{I}_M c = \hat{I}_M c$ for all $c \in M$. Since $\check{I}_{\tilde{M}} c \cap R \neq \emptyset$ and $\hat{I}_{\tilde{M}} c \cap R \neq \emptyset$ would imply $c \in R$, we have only three possible cases:

(1) $\check{I}_{\tilde{M}} c \cap R = \emptyset$, $\hat{I}_{\tilde{M}} c \cap R = \emptyset$;

(2) $\check{I}_{\tilde{M}} c \cap R \neq \emptyset$, $\hat{I}_{\tilde{M}} c \cap R = \emptyset$;

(3) $\check{I}_{\tilde{M}} c \cap R = \emptyset$, $\hat{I}_{\tilde{M}} c \cap R \neq \emptyset$.

Lemma 2: $\check{I}_{\tilde{M}}c\cap R=\emptyset$ implies $\check{I}_M c=\check{I}_{\tilde{M}}c$.

Proof: It suffices to show $\check{I}_{\tilde{M}}c\subseteq\check{I}_M c$. For any $x\in\check{I}_{\tilde{M}}c$, the timelike curve connecting c to x must not intersect R or there would be $y\in\check{I}_{\tilde{M}}c\cap R$. Hence $x\in\check{I}_M c$. ■

Lemma 3: $\uparrow_M\hat{I}_M c\subseteq\uparrow_{\tilde{M}}\hat{I}_{\tilde{M}}c$.

Proof: For any $x\in\uparrow_M\hat{I}_M c$, there exists $y\in M$, $x\underset{M}{\gg}y\underset{M}{\gg}\hat{I}_M c\subset\hat{I}_{\tilde{M}}c$. Let $\{u_i\}$ be a sequence in $\hat{I}_M c$ with c as its limit point, then $y\underset{M}{\gg}\{u_i\}$, which implies $y\underset{\tilde{M}}{\gg}\{u_i\}$, hence $y\underset{\tilde{M}}{\gg}\check{I}_{\tilde{M}}c$ and $x\in\uparrow_{\tilde{M}}\hat{I}_{\tilde{M}}c$. Note that Lemma 1 has been used twice. ■

Since we always have $\check{I}_M c\subseteq\uparrow_M\hat{I}_M c$ and $\hat{I}_M c\subseteq\downarrow_M\check{I}_M c$, what remains to be shown is $\downarrow_M\check{I}_M c\subseteq\hat{I}_M c$ and $\uparrow_M\hat{I}_M c\subseteq\hat{I}_M c$. On account of Lemmas 2 and 3 as well as the causal continuity of $(\tilde{M},\eta)$, $\uparrow_M\hat{I}_M c\subseteq\check{I}_M c$ is true for cases (1) and (3), while $\downarrow_M\check{I}_M c\subseteq\hat{I}_M c$ is true for cases (1) and (2). Therefore the essential issue is to prove $\uparrow_M\hat{I}_M c\subseteq\check{I}_M c$ for case (2) since $\downarrow_M\check{I}_M c\subseteq\hat{I}_M c$ for case (3) will then follow dually.

Let $c=(t_c,x_c^1,\ldots,x_c^n)$, then $\check{I}_{\tilde{M}}c\cap R\neq\emptyset$ implies $t_c<0$. Define

$$S_1\equiv\uparrow_M\hat{I}_M c\cap\{(t,x^1,\ldots,x^n):t\leqslant 0\},$$

$$\tilde{S}_1\equiv\uparrow_{\tilde{M}}\hat{I}_{\tilde{M}}c\cap\{(t,x^1,\ldots,x^n):t\leqslant 0\}$$
$$=\check{I}_{\tilde{M}}c\cap\{(t,x^1,\ldots,x^n):t\leqslant 0\},$$

$$S_2\equiv\uparrow_M\hat{I}_M c\cap\{(t,x^1,\ldots,x^n):t>0\},$$

$$\tilde{S}_2\equiv\uparrow_{\tilde{M}}\hat{I}_{\tilde{M}}c\cap\{(t,x^1,\ldots,x^n):t>0\}$$
$$=\check{I}_{\tilde{M}}c\cap\{(t,x^1,\ldots,x^n):t>0\},$$

then $\uparrow_M\hat{I}_M c=S_1\cup S_2$, $S_1\subseteq\tilde{S}_1-R$, $S_2\subseteq\tilde{S}_2-R$. We want to show $S_1\subseteq\check{I}_M c$ and $S_2\subseteq\check{I}_M c$.

Let $p\in S_1\subseteq\tilde{S}_1-R$, then $p\in\check{I}_{\tilde{M}}c$. The timelike curve connecting c to p must not intersect R or there would be $q\in R\cap\hat{I}_{\tilde{M}}p$ which implies $p\in\check{I}_{\tilde{M}}a\cap\{(t,x^1,\ldots,x^n):t\leqslant 0\}\subset R$, thus $p\in\check{I}_M c$.

Let $p=(t_p,x_p^1,\ldots,x_p^n)\in S_2$, then $p\in\text{interior}\{y\in M:y\underset{M}{\gg}\hat{I}_M c\}$ and one can choose $a<t_p$ such that $p'=(t_p-a,x_p^1,\ldots,x_p^n)\in\text{interior}\{y\in M:y\underset{M}{\gg}\hat{I}_M c\}$, i.e., $p'\underset{M}{\gg}\hat{I}_M c$. Let $v_i=(t_c-1/i,x_c^1,\ldots,x_c^n)$, then $\{v_i\}\subset\hat{I}_M c$ and $\{v_i\}$ converge to c. By Lemma 1 we have $p'\underset{M}{\gg}\{v_i\}$, hence there exist timelike curves γ_i in M connecting v_i to p'. Since $t_p-a>0$ and $t_c-1/i<0$, each γ_i must intersect the plane $E\equiv\{(0,x^1,\ldots,x^n)\}$ at some point $q_i\in E\cap M$. The timelike property of γ_i gives

$$(t_c-1/i)^2>(x_{q_i}^1-x_c^1)^2+\cdots+(x_{q_i}^n-x_c^n)^2,$$

while $q_i\in\gamma_i\cap E$ and $\gamma_i\cap R=\emptyset$ yield

$$1<(x_{q_i}^i)^2+\cdots+(x_{q_i}^n)^2.$$

On the other hand, $\lim(t_c-1/i)=t_c$ implies that all q_i's with sufficiently large i are within a compact region of the n-dimensional Euclidean space E, hence there exists a subsequence $\{q_i'\}$ of $\{q_i\}$ such that $\{q_i'\}$ converges to a point $q=(0,x_q^1,\ldots,x_q^n)\in E$ satisfying

$$t_c^2\geqslant(x_q^1-x_c^1)^2+(x_q^n-x_c^n)^2, \tag{1}$$

$$1\leqslant(x_q^1)^2+\cdots+(x_q^n)^2, \tag{2}$$

And $q_i'\in\hat{I}_M p'\subset\hat{I}_{\tilde{M}}p'$ implies $q\in\text{closure}(\hat{I}_{\tilde{M}}p')\subset\hat{I}_{\tilde{M}}p$.

Lemma 4: If there exists $r\in E$ satisfying (a) r is sufficiently close to q so that $r\in\hat{I}_{\tilde{M}}p$, and (b) $r\in(\check{I}_{\tilde{M}}c-R)\cap E=(\check{I}_{\tilde{M}}c\cap E)-(R\cap E)$, then $p\in\check{I}_M c$.

Proof: Since $(\tilde{M},\eta)$ is a Minkowski space-time, $r\in\hat{I}_{\tilde{M}}p$ implies that there exists a timelike geodesic γ connecting r to p. But $r\in E-R$ implies $\gamma\cap R=\emptyset$, hence $p\in\check{I}_M r$. On the other hand, requirement (b) leads to $r\in\tilde{S}_1-R=S_1\subseteq\check{I}_M c$, therefore $p\in\check{I}_M c$. ■

Let $B_0,B_c\subset E$ be open balls centered at $(0,0,\ldots,0)$ and $(0,x_c^1,x_c^n)$ with radii 1 and $|t_c|$, respectively, then the requirement $r\in(\check{I}_{\tilde{M}}c-R)\cap E$ and inequalities (1) and (2) are equivalent to $r\in B_c-\text{closure}(B_0)$ and $q\in\text{closure}(B_c)-B_0$, respectively. Since $B_c\not\subseteq B_0$ or c would be in R, it is clear that one can always find such an r for any q unless $q\in\partial B_c\cap\partial B_0$ and $n=1$. Therefore we conclude that the space-time (M,η) with $n>1$ is causally continuous.

ACKNOWLEDGMENTS

The authors would like to thank Professor Abdus Salam, the International Atomic Energy Agency, and UNESCO for hospitality at the International Centre for Theoretical Physics, Trieste, where part of this work was done.

[1] B. G. Schmidt, Gen. Relativ. Gravit. **1**, 269 (1971).
[2] R. Geroch, J. Math. Phys. **9**, 450 (1968).
[3] R. A. Johnson, J. Math. Phys. **18**, 898 (1977).
[4] R. Geroch, Liang can-bin, and R. M. Wald, J. Math. Phys. **23**, 432 (1982).
[5] R. Geroch, E. H. Kronheimer, and R. Penrose, Proc. R. Soc. London Ser. A **327**, 545 (1972).
[6] R. Budic and R. K. Sachs, J. Math. Phys. **15**, 1302 (1974).
[7] Kuang zhi-quan, Li jian-zeng, and Liang can-bin, Phys. Rev. D **33**, 1533 (1986).
[8] S. W. Hawking and R. K. Sachs, Commun. Math. Phys. **35**, 287 (1974).
[9] U. D. Vyas and G. M. Akolia, Gen. Relativ. Gravit. **18**, 309 (1986).
[10] J. K. Beem, Gen. Relativ. Gravit. **8**, 245 (1977).

Static uniform-density stars must be spherical in general relativity

Lee Lindblom
Department of Physics, Montana State University, Bozeman, Montana 59717

(Received 11 June 1987; accepted for publication 30 September 1987)

In this paper the uniqueness of the static solutions of Einstein's equation that represent isolated uniform-density perfect-fluid stellar models is demonstrated: any static asymptotically flat space-time containing only a uniform-density perfect fluid confined to a spatially compact world tube is necessarily spherically symmetric. This result generalizes to relativistic uniform-density models the well known Newtonian theorem of Carleman and Lichtenstein.

I. INTRODUCTION

The inevitability of spherical symmetry in isolated static (i.e., time independent and nonrotating) fluid stellar models was first demonstrated in the Newtonian theory by Carleman[1] and Lichtenstein.[2,3] To date the analogous result has not been established in general relativity theory. A number of studies of the properties of static relativistic stellar models have been published, however. Some of the more interesting results of these investigations are as follows. Masood-ul-Alam[4] has shown that the topology of the space-times containing these static stellar models must be diffeomorphic to $\mathbf{R}^3\times\mathbf{R}$. Avez,[5,6] Künzle,[7] and Lindblom[8,9] have studied the geometry of this general class of space-times. They established the equivalence of spherical symmetry and a number of other geometrical properties (e.g., spatial conformal flatness) in these space-times. Künzle and Savage[10] showed that the spherical static space-times are isolated in the sense that no continuous family of static fluid space-times exists which contains both spherical and nonspherical space-times.

Recently Masood-ul-Alam[11] explored the implications of the positive mass theorem[12–14] on the geometry of static fluid space-times. He demonstrated that the positive mass theorem could be used to prove the necessity of spherical symmetry in a subset of these space-times that satisfies certain special properties. He assumed that the fluid obeyed a particular equation of state, which having $d\rho/dp<0$ is unfortunately extremely unphysical. He also limited his attention to a subset of the stellar models based on this equation of state which have $\rho\geqslant 0$. Since the spherical models in this subset all have $\rho=0$ at the center of the star, he has implicitly assumed that the central pressure in these (potentially nonspherical) models is never greater than that achieved in the corresponding spherical model. When stated in this way the additional assumption, $\rho\geqslant 0$, seems to me to be an unnatural auxiliary assumption in the context of the particular equation of state considered by him.

In this paper the necessity of spherical symmetry in isolated static uniform-density stellar models is demonstrated. These stellar models have a somewhat more physically acceptable equation of state than the one considered by Masood-ul-Alam. Furthermore, no unnatural auxiliary assumption is necessary in this case. Thus Masood-ul-Alam's[11] recognition of the importance of the positive mass theorem in the study of static space-times is further supported. This work also supersedes portions of Ref. 8 which erroneously claimed to prove the necessity of spherical symmetry in static uniform-density stellar models. (This error has been noted previously in Ref. 9.) Section II of this paper reviews the established properties of static stellar models that are needed in this analysis. Section III presents the proof that spherical symmetry is a necessary property of isolated static uniform-density stellar models in general relativity theory. The method of proof is to perform a particular conformal transformation on the spatial metric which sets the mass to zero and leaves the scalar curvature non-negative. The demonstration that the scalar curvature resulting from this transformation is non-negative requires the use of the divergence identities for static stellar models found in Ref. 8. The positive mass theorem implies that this conformally transformed metric is flat. The desired result follows from the already established equivalence of spatial conformal flatness and spherical symmetry in static fluid space-times.[8]

II. STATIC STELLAR MODELS

In this section some of the basic properties of static perfect-fluid space-times are reviewed. Careful derivations of these results can be found in the literature. The statements in this section are valid for any static perfect-fluid space-time while those in the next section are valid only for uniform-density stellar models.

A static space-time must admit a hypersurface orthogonal timelike Killing vector field, t^a. Let t be a function whose level surfaces are orthogonal to t^a, and let $t^a\partial_a t=1$. The space-time metric can then be represented in the form

$$ds^2=-V^2\,dt^2+g_{ab}\,dx^a\,dx^b, \tag{1}$$

where g_{ab} is the positive definite three-metric of the constant-t surfaces and $0=\partial_t V=\partial_t g_{ab}$. Einstein's equation for such a space-time with a perfect-fluid stress-energy tensor is equivalent to the system of equations

$$D^aD_aV=4\pi V(\rho+3p), \tag{2}$$

$$R_{ab}=V^{-1}D_aD_bV+4\pi(\rho-p)g_{ab}, \tag{3}$$

where D_a and R_{ab} are the three-dimensional covariant derivative and the Ricci curvature tensor associated with g_{ab}, ρ is the total energy density (including rest-mass energy), and p is the pressure of the fluid. To these equations must be added an equation of state: a function $\rho=\rho(p)$ that summarizes the microscopic properties of the particular fluid. This function must be positive and monotonically increasing to be

 0022-2488/88/020327-05$02.50

physically relevant. Associated with Eq. (3) is a Bianchi identity, which is equivalent to Euler's equation for these static fluids,

$$D_a p = -V^{-1}(\rho + p) D_a V. \tag{4}$$

The solutions of Eqs. (2) and (3) that are of interest in this paper are the physically isolated solutions. Thus we only consider solutions in which the support of the pressure is spatially compact, and in which the space-time metric is asymptotically flat in an appropriate sense. We assume that V and g_{ab} are given asymptotically by expressions of the form,

$$V = 1 - m/r + v, \tag{5}$$

$$g_{ab} = (1 + 2m/r)\delta_{ab} + h_{ab}, \tag{6}$$

where δ_{ab} is the standard flat metric on a constant-t surface, the function r is the asymptotic spherical coordinate given by $r^2 = \delta_{ab} x^a x^b$, and the x^a are the Cartesian coordinates associated with δ_{ab} on each constant-t surface. The quantities v and h_{ab} must vanish like r^{-2} as $r \to \infty$, and their first and second derivatives must vanish with successively higher powers of r^{-1}. The constant m is the mass of the star. These asymptotic conditions on the solutions of Eqs. (2) and (3) can be deduced from rather mild asymptotic falloff assumptions.[15]

To avoid the possibility of surface stresses and surface energy densities and thereby violate the assumption that the stress energy in these space-times is purely that of a perfect fluid, some care must be taken to ensure the proper matching conditions at the boundary between the exterior vacuum region and the interior fluid portion of the stellar model (and in addition at any interior surface on which the equation of state is not continuous). From Eq. (4) it follows that p (and consequently ρ) must have level surfaces that coincide with the level surfaces of V. It also follows from Eq. (4) that p must be a continuous function (since V must be continuous) to avoid the existence of surface stresses on the boundary between the interior and exterior of the star. The pressure must vanish, therefore, on this boundary. Let $V = V_s$ be the level surface of V that corresponds to this boundary. We must also impose an appropriate discontinuity in $D_a D_b V$ at this surface if the equation of state is one for which $\rho(0) \neq 0$ [see, e.g., Eq. (2)]. The needed condition is most easily expressed in terms of the function $W \equiv D^a V D_a V$. This function must satisfy the following discontinuity condition[8] on the surface $V = V_s$:

$$[W^{-1} D_a V D^a W] = -8\pi V_s \rho(0), \tag{7}$$

where $[Q]$ represents the discontinuity (exterior minus interior) in the quantity Q on the surface $V = V_s$.

The conformal properties of a three-geometry are expressed in terms of a certain third-rank tensor field R_{abc} defined by

$$R_{abc} = D_c R_{ab} - D_b R_{ac} + \tfrac{1}{4}(g_{ac} D_b R - g_{ab} D_c R), \tag{8}$$

where $R = R_{ab} g^{ab}$. This tensor vanishes if and only if the geometry is conformally flat.[16] Two different expressions for R_{abc} will be useful in the analysis that follows. The first relates R_{abc} to the geometrical properties of the constant-V two-surfaces in static perfect-fluid space-times:

$$R_{abc} R^{abc} = 8V^{-4} W^2 (\psi_{ab} \psi^{ab} + \tfrac{1}{8} W^{-2} \beta^{ab} D_a W D_b W), \tag{9}$$

where ψ_{ab} is the trace-free part of the extrinsic curvature and β_{ab} is the intrinsic metric of the constant-V two-surfaces. If the metric g_{ab} were conformally flat then the left-hand side of Eq. (9) would vanish. Since the metric g_{ab} is positive definite it would follow that

$$\psi_{ab} = \beta^{ab} D_b W = 0, \tag{10}$$

in this case. Avez[5,6] and Künzle[7] have shown that these conditions, Eq. (10), are equivalent to spherical symmetry. Therefore, Eq. (9) establishes the equivalence of spatial conformal flatness and spherical symmetry for static perfect-fluid space-times.[8] Using Eqs. (2) and (3) R_{abc} can also be expressed completely in terms of V, the fluid variables, and their derivatives. An expression of this type that will be useful in the analysis that follows is given by

$$\begin{aligned} \tfrac{1}{4} V^4 W^{-1} R_{abc} R^{abc} &= D^a D_a W - V^{-1} D^a V D_a W \\ &\quad - \tfrac{3}{4} W^{-1} D^a W D_a W + 8\pi W(\rho + p) \\ &\quad + 4\pi V W^{-1}(\rho + 3p) D^a V D_a W \\ &\quad - 16\pi^2 V^2 (\rho + 3p)^2 - 8\pi V D^a V D_a \rho. \end{aligned} \tag{11}$$

One further property of the conformal transformation of three-geometries will be useful. Consider the conformal metric $\bar{g}_{ab} = \psi^4 g_{ab}$. The conformally transformed scalar curvature $\bar{R}$ is related to R by the equation[16]

$$\bar{R} = \psi^{-4}(R - 8\psi^{-1} D^a D_a \psi), \tag{12}$$

where R and D_a are the scalar curvature and covariant derivative associated with g_{ab}.

III. STATIC UNIFORM-DENSITY STARS MUST BE SPHERICAL

The necessity of spherical symmetry in isolated static uniform-density stellar models will be demonstrated by showing that any such model must be spatially conformally flat. To accomplish this an explicit conformal transformation is performed on the metric. The scalar curvature of the conformally transformed metric is shown to be non-negative. The positive mass theorem is then used to demonstrate that the transformed metric is in fact flat.

Consider the conformal transformation $\bar{g}_{ab} = \psi^4 g_{ab}$, where ψ is the following function of V:

$$\psi(V) = \begin{cases} \tfrac{1}{2}(1 + V), & V_s \leqslant V \leqslant 1, \\ \tfrac{1}{2}(1 + V_s)^{3/2}(1 + 3V_s - 2V)^{-1/2}, & 0 < V \leqslant V_s. \end{cases} \tag{13}$$

Note that $\psi(V)$ and its first derivative are continuous at the surface $V = V_s$. Also note that $\psi''(V)$, the second derivative of $\psi(V)$, vanishes for $V_s \leqslant V \leqslant 1$ and is positive for $0 < V \leqslant V_s$ since

$$\psi''(V) = \tfrac{3}{2}(1 + V_s)^{3/2}(1 + 3V_s - 2V)^{-5/2} > 0. \tag{14}$$

The scalar curvature associated with the metric $\bar{g}_{ab}$ can now be computed using Eq. (12) with the result

$$\bar{R} = 8\psi^{-5}\psi''\{W_0(V) - W\}. \tag{15}$$

The function $W_0(V)$ used in Eq. (15) is defined by

$$W_0(V) = \begin{cases} \frac{2}{3}\pi\rho(1-V^2)^4(1-V_s^2)^{-3}, & V_s \leqslant V \leqslant 1, \\ \frac{8}{3}\pi\rho V(3V_s - V) + \frac{2}{3}\pi\rho(1-9V_s^2), & 0 < V \leqslant V_s. \end{cases} \tag{16}$$

To establish Eq. (15) it is necessary to use Eqs. (2) and (3) and the fact that the integral of Eq. (4) for the pressure can be written in the form

$$p = \rho V^{-1}(V_s - V) \tag{17}$$

for uniform-density fluids in the domain $0 < V \leqslant V_s$.

The next step is to demonstrate that the scalar curvature $\bar{R}$ given in Eq. (15) is non-negative. Since ψ and ψ'' are non-negative from Eqs. (13) and (14) it remains only to determine the sign of $W_0(V) - W$. The function $W_0(V)$ is continuous at the surface $V = V_s$, while its first derivative satisfies the following discontinuity condition:

$$[W^{-1}D_a V D^a W_0] = -8\pi V_s \rho. \tag{18}$$

This is precisely the same, for uniform-density stellar models, as the discontinuity condition, Eq. (7), satisfied by the first derivative of W. Consequently the function $W_0(V) - W$ and its first derivative are continuous everywhere including the boundary surface $V = V_s$.

The sign of $W_0(V) - W$ will be determined using two identities and the maximum principle for elliptic differential operators. Using Eqs. (11), (16), and (17) it is straightforward to show that in the interior of the star (i.e., the region $0 < V \leqslant V_s$) the following identity must be satisfied[8]:

$$\begin{aligned} D^a\{V^{-1}D_a(W - W_0)\} &= \tfrac{1}{4}V^3 W^{-1} R_{abc} R^{abc} \\ &\quad + \tfrac{3}{4}V^{-1}W^{-1}D_a(W - W_0)D^a(W - W_0). \end{aligned} \tag{19}$$

The right-hand side of Eq. (19) is non-negative. The left-hand side is an elliptic differential operator acting on the function $W - W_0$. The maximum principle (see, e.g., Ref. 17) states that $W - W_0$ must achieve its maximum value at a boundary point of the domain on which Eq. (19) is valid (i.e., on the surface $V = V_s$ in this case). Furthermore the gradient $D_a(W - W_0)$ must be nonvanishing and directed out of the domain (the interior of the star in this case) at this maximum point unless the function $W - W_0$ is in fact constant.

A similar identity exists in the exterior of the star[11,18] (i.e., the region where $V_s \leqslant V \leqslant 1$):

$$D_a\{V^{-1}D^a Y\} = \frac{V^4 R_{abc} R^{abc} + 3X_a X^a}{4VW(1-V^2)^3}, \tag{20}$$

where X_a and Y are defined by

$$X_a = D_a W + 8VW(1-V^2)^{-1}D_a V, \tag{21}$$

$$Y = (W - W_0)/(1-V^2)^3. \tag{22}$$

The right-hand side of Eq. (20) is also non-negative while the left-hand side is an elliptic differential operator on the function Y. The maximum principle implies that the maximum of Y must occur either on the surface of the star where $V = V_s$ or at infinity where $V = 1$.

Consider first the case where the maximum Y occurs at infinity. The asymptotic falloff conditions [i.e., Eqs. (5) and (6)] imply that W and W_0 go to zero like r^{-4} while $1 - V^2$ vanishes like r^{-1}. Therefore Y vanishes at infinity. If the maximum of Y occurs at infinity then $W - W_0$ is necessarily nonpositive in the exterior of the star from Eq. (22). By continuity and the argument given above for the location of the maximum of $W - W_0$ in the interior of the star, it follows that $W \leqslant W_0$ everywhere in the space-time in this case. (This case was inadvertently overlooked in Ref. 8.) Finally, it follows from Eq. (15) that the conformally transformed scalar curvature is non-negative in this case: $\bar{R} \geqslant 0$.

Consider next the case where the maximum of Y (with respect to the exterior of the star) occurs on the surface of the star, $V = V_s$. In this case $Y \geqslant 0$ at this maximum since $Y = 0$ at infinity. It follows that the maximum of the function $Y(1-V^2)^3$ must occur at the same location as the maximum of Y in this case, since Y is non-negative and the maximum of $(1-V^2)^3$ occurs on the surface of the star $V = V_s$. Therefore the maximum of $W - W_0 = Y(1-V^2)^3$ with respect to both the interior and the exterior regions must occur on the surface of the star in this case. Since $D_a(W - W_0)$ is continuous it must vanish at this maximum point. The gradient of Y at this maximum point is given therefore by $D_a Y = 6VY(1-V^2)^{-1}D_a V$. Since V is larger in the exterior of the star than the interior, this gradient points *into* the exterior region. The maximum principle demands that this gradient points *out of* the exterior region unless Y is constant. Since $Y = 0$ at infinity it follows that Y must vanish everywhere in this case, and consequently $W = W_0$ everywhere as well. Thus the conformally transformed scalar curvature would vanish identically in this case: $\bar{R} = 0$.

To summarize, the conformally transformed scalar curvature $\bar{R}$ is necessarily non-negative in a static asymptotically flat uniform-density fluid stellar model.

To complete the proof of the necessity of spherical symmetry the asymptotic behavior of the conformally transformed metric $\bar{g}_{ab}$ must be determined. The asymptotic expansion of the conformal factor defined in Eq. (13) can be determined by the asymptotic form of V given in Eq. (6): $\psi = 1 - m/2r + \phi$, where ϕ vanishes like r^{-2} as $r \to \infty$. It follows that the conformal metric is given in this limit by $\bar{g}_{ab} = \delta_{ab} + \bar{h}_{ab}$, where $\bar{h}_{ab}$ vanishes like r^{-2}. Thus the mass associated with the metric $\bar{g}_{ab}$ vanishes. The positive mass theorem[12-14] states that any three-geometry having non-negative scalar curvature and zero mass is in fact flat. Therefore the metric $\bar{g}_{ab}$ is flat. The physical spatial metric g_{ab} is consequently conformally flat, and the stellar model is therefore spherical.

ACKNOWLEDGMENTS

I thank Robert Geroch for numerous discussions over the years on the subject of the symmetry of static space-times. I also thank Gregory Burnett and Robert Geroch for carefully reading an early version of this paper.

This work was supported by Grant No. PHY 86-18490 from the National Science Foundation.

[1]T. Carleman, Math. Z. B **3**, 1 (1919).
[2]L. Lichtenstein, Sitzungsber. Preuss. Akad. Wiss. Phys. Math. Kl. **1919**, 1120.
[3]L. Lichtenstein, *Gleichgewichtsfiguren Rotierender Flussigkeiten* (Springer, Berlin, 1933).
[4]A. K. M. Masood-ul-Alam, Commun. Math. Phys. **108**, 193 (1987).
[5]A. Avez, C. R. Acad. Sci. Paris **257**, 3316 (1963).
[6]A. Avez, Ann. Inst. H. Poincaré A **1**, 291 (1964).
[7]H. P. Künzle, Commun. Math. Phys. **20**, 85 (1964).
[8]L. Lindblom, J. Math. Phys. **21**, 1455 (1980).
[9]L. Lindblom, J. Math. Phys. **22**, 1324 (1981).
[10]H. P. Künzle and J. R. Savage, Gen. Relativ. Gravit. **12**, 155 (1980).
[11]A. K. M. Masood-ul-Alam, Class. Quantum Grav. **4**, 625 (1987).
[12]R. Schoen and S. T. Yau, Commun. Math. Phys. **65**, 45 (1979).
[13]E. Witten, Commun. Math. Phys. **80**, 381 (1981).
[14]R. Bartnik, Commun. Pure Appl. Math. **39**, 661 (1986).
[15]R. Beig, Gen. Relativ. Gravit. **12**, 439 (1980).
[16]L. P. Eisenhart, *Riemannian Geometry* (Princeton U. P., Princeton, NJ, 1925), pp. 89–92.
[17]R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1962), Vol. 2, pp. 326–328.
[18]D. C. Robinson, Gen. Relativ. Gravit. **8**, 695 (1977).

Vacuum solutions admitting a geodesic null congruence with shear proportional to expansion

A. H. Kupeli[a)]
Physics Department, University of Alabama at Birmingham, Birmingham, Alabama 35294

(Received 22 July 1986; accepted for publication 14 October 1987)

Algebraically general, nontwisting solutions for the vacuum to vacuum generalized Kerr–Schild (GKS) transformation are obtained. These solutions admit a geodesic null congruence with shear proportional to expansion. In the Newman–Penrose formalism, if l^μ is chosen to be the null vector of the GKS transformation, this property is stated as $\sigma = a\rho$ and $Da = 0$. It is assumed that a is a constant, and the background is chosen as a *pp*-wave solution. For generic values of a, the GKS metrics consist of the Kasner solutions. For $a = \pm(1 \pm \sqrt{2})$, there are solutions with less symmetries including special cases of the Kóta–Perjés and Lukács solutions.

I. INTRODUCTION

In this paper we shall obtain algebraically general, nontwisting generalized Kerr–Schild (GKS) type vacuum solutions using the Newman–Penrose (NP) formalism.[1] We recall that the generalization consists of allowing nonflat backgrounds in the Kerr–Schild metrical anzatz (see Kramer *et al.*,[2] Chap. 28). That is, the GKS metric is of the form $\tilde{g}_{\mu\nu} = g_{\mu\nu} + 2Vl_\mu l_\nu$, where l_μ is a null vector (with respect to both metrics), but $g_{\mu\nu}$ is not necessarily flat. In order to obtain new solutions $\tilde{g}_{\mu\nu}$ one starts with a (known) background solution and, using the Einstein equations for $\tilde{g}_{\mu\nu}$, obtains and solves differential equations for V. This process is called the GKS transformation.

For vacuum to vacuum ($V \to V$) GKS transformations, the null vector (l^μ) of the transformation must be geodesic ($\kappa = 0$).[3] Then the GKS metrics are obtained by solving first-order linear equations.[4,5] When l^μ is shear-free ($\sigma = 0$) the transformation equation are compatible, but the GKS solutions are necessarily algebraically special.[5] This case has been studied in the literature by various authors.[6–9] When l^μ is shearing ($\sigma \neq 0$), the transformation equations give the following constraint on the background:

$$\Psi_0 = [2\sigma/(\rho + \bar{\rho})][\sigma\bar{\sigma} - \rho^2]\,. \tag{1.1}$$

Equation (1.1) is invariant under the GKS transformation. When l^μ is nontwisting ($\rho = \bar{\rho}$), this constraint gives[5]

$$\sigma = a\rho\,, \tag{1.2}$$

where a is a function such that $Da = 0$. Thus for the $V \to V$ transformation with shearing but nontwisting null vectors, the admissible backgrounds and corresponding GKS solutions are both characterized by the existence of a geodesic null congruence whose shear and expansion have the same "radial" dependence. This property eliminates the possibility of obtaining asymptotically flat "radiating" solutions.[1,5]

In Sec. II we shall obtain a type N background solution satisfying (1.2) with $a = \text{const}$. In fact this solution is a (vacuum) *pp* wave (see Kramer *et al.*,[2] Sec. 21.5). In Sec. III, using these backgrounds, we shall obtain algebraically general GKS metrics. These solutions depend on a function V_0 such that $DV_0 = 0$. For generic values of a, V_0 reduces to a constant and these metrics are the Kasner solutions.[10] For $a = \pm(1 \pm \sqrt{2})$, V_0 is either a linear function of x^k, or an exponential function of u and x^k, $k = 3$ or 4 (these coordinates are defined in the beginning of Sec. II). The solutions with linear V_0 are the nontwisting cases of the Kóta–Perjés[11] and Lukács[12] solutions (space-times with geodesic eigenrays and, respectively, a timelike or spacelike Killing vector). The GKS solutions with exponential type V_0 admit only one spacelike Killing vector. By linearity of the $V \to V$ GKS transformation equations, it is also possible to have metrics with arbitrary functions. Also if V_0 is a superposition of linear and exponential type functions, the resulting space-times can be considered as nonstationary vacuum perturbations of the Kasner, or the Kóta–Perjés–Lukács solutions. The results of Sec. II and III are summarized in Table I.

We shall use the coordinates and the tetrad frame defined in Sec. VI of NP.[1] To avoid repetitions, we shall use the quantities therein without redefining, and we shall directly refer to the equations in that section. For example, Eq. (NP6.5) will denote Eq. (6.5) in Sec. VI of NP.

II. BACKGROUND SOLUTIONS

We shall start by a background metrical anzatz that will ensure that l^μ is a shearing nontwisting geodesic null vector. We choose the coordinates $(x^1 = u, x^2 = r, x^3 = x, x^4 = y)$ and the tetrad frame $(l^\mu, n^\mu, m^\mu, \bar{m}^\mu)$ as in the Sec. VI of NP (in particular ρ is real). In addition we assume that the metric functions [see Eqs. (NP6.5-6)] satisfy

$$\omega = U = X^k = 0, \quad \xi^k = \xi^k(r), \quad k = 3,4\,.$$

Then the directional derivatives reduce to

$$D = \partial_r, \quad \Delta = \partial_u, \quad \delta = \xi^k(r)\partial_k, \quad k = 3,4\,. \tag{2.1}$$

TABLE I. Algebraic types for background and GKS solutions. The metric is given by $d\tilde{s}^2 = 2\,du\,dr - [r^{2a^+}\,dx^2 + r^{2a^-}\,dy^2] + 2V_0 r^{-s}\,du^2$.

Background solutions	$V_0 = 0$	Type N
GKS solutions	$V_0 \neq 0$	Flat for $a = \pm 1$ Type D for $a = 0, \pm 3$ Type I otherwise

a) Present address: Department of Mathematics, State University of New York at Stony Brook, Stony Brook, New York 11794.

 0022-2488/88/020440-03$02.50

From the commutation relations [Eqs. (NP6.8)] it can be seen that the only nonzero spin coefficients are ρ and σ. In addition we assume that

$$\sigma = a\rho\,. \tag{2.2}$$

Then without loss of generality we can take a real, since the phase of σ does not depend on r.[13]

Equations (NP6.11a) and (NP6.11b) give

$$\rho = -1/(1+a^2)r\,, \tag{2.3}$$

$$\sigma = -a(1+a^2)r\,, \tag{2.4}$$

$$\Psi_0 = -(1-a^2)/(1+a^2)^2 r^2\,, \tag{2.5}$$

and it remains to solve only [Eq. (NP6.10a)]

$$D\xi^k = \rho\xi^k + \sigma\bar{\xi}^k\,. \tag{2.6}$$

In the following, whenever convenient, we shall use the notation

$$\eta^{\pm} = \eta \pm \bar{\eta}\,, \tag{2.7}$$

where η is either a spin coefficient, Ψ_i, $i = 0,\ldots,4$, or δ. Using (2.3) and (2.4), Eq. (2.6) can be written as

$$D\xi^{k\pm} = -(a^{\pm}/r)\xi^{k\pm}\,, \tag{2.8}$$

where

$$a^{\pm} = (1 \pm a)/(1+a^2)\,. \tag{2.9}$$

Then $\xi^{k\pm}$ can be solved easily as

$$\xi^{k\pm} = \xi_0^{k\pm}/r^{a\pm}\,, \tag{2.10}$$

where $\xi_0^{k\pm}$ are constants. We choose these constants as $\xi_0^{3+} = \sqrt{2}$, $\xi_0^{4-} = i\sqrt{2}$, $\xi_0^{3-} = \xi_0^{4+} = 0$. Then (2.11) reduces to

$$D = \partial_r, \quad \Delta = \partial_u,$$

$$\delta = (1/\sqrt{2})\left[r^{-a^+}\partial_x + i r^{-a^-}\partial_y\right]. \tag{2.11}$$

The background metric is then given by

$$ds^2 = 2\,du\,dr - (r^{2a^+}dx^2 + r^{2a^-}dy^2)\,. \tag{2.12}$$

Since all the spin coefficients except ρ and σ vanish, it can be seen that $n_{\mu;\nu} = 0$. Then the metric (2.12) admits a covariantly constant null vector field. Hence it is a *pp*-wave solution (see Kramer *et al.*,[2] Sec. 21.5).

III. SOLUTION OF THE $V \to V$ GKS TRANSFORMATION EQUATIONS AND THE GKS METRICS

We start by giving a brief description of the GKS transformaton: Let $g^{\mu\nu}$ be a (known) background solution, and $(l^\mu, n^\mu, m^\mu, \bar{m}^\mu)$ the corresponding null tetrad frame. The GKS transformation is defined by the following transformation of the tetrad vectors:

$$l^\mu \to l^\mu, \quad n^\mu \to n^\mu - Vl^\mu, \quad m^\mu \to m^\mu\,. \tag{3.1}$$

Then, the GKS metric has the form

$$\tilde{g}^{\mu\nu} = g^{\mu\nu} - 2Vl^\mu l^\nu\,. \tag{3.2}$$

For the background given in Sec. II, the GKS transformation equations[5] reduce to

$$DV = (1-a^2)\rho V\,, \tag{3.3}$$

$$\tfrac{1}{2}\delta DV + a\rho\bar{\delta}V = 0\,, \tag{3.4}$$

$$\rho\,\Delta V = \delta\bar{\delta}V\,, \tag{3.5}$$

where ρ, σ, and the directional derivatives are defined by Eqs. (2.3), (2.4), and (2.11).

Equation (3.3) is solved as

$$V = V_0 r^{-s}\,, \tag{3.6}$$

where $DV_0 = 0$ and $s = (1-a^2)/(1+a^2)$. Thus the GKS metric has the form

$$d\tilde{s}^2 = 2\,du\,dr - [r^{2a^+}dx^2 + r^{2a^-}dy^2] + 2V_0 r^{-s}du^2\,. \tag{3.7}$$

Using Eqs. (3.3) and (2.11), Eq. (3.4) can be written as

$$(a^2 - 2a - 1)\partial_x V_0\, r^{-a^+} + i(a^2 + 2a - 1)\partial_y V_0\, r^{-a^-} = 0\,, \tag{3.8}$$

where a^+ and a^- are defined by (2.9).

Therefore if $a \neq \pm(1+\sqrt{2})$, then $\partial_x V_0 = \partial_y V_0 = 0$, and Eq. (3.5) gives $\partial_u = 0$. Thus generically V_0 reduces to a constant. In this case the GKS metrics given by Eq. (3.7) admit three Killing vectors $\{\partial_x, \partial_y, \partial_u\}$. The vectors ∂_x and ∂_y are spacelike, and ∂_u is spacelike (timelike) if V_0 is negative (positive). For $V_0 > 0$, these metrics have been explicitly identified with the Kasner solutions[10] [see Eq. (11.5) in Kramer *et al.*[2]].

We recall that the solutions with positive and negative values of a are related by a tetrad rotation. Thus we consider only the special values $a = 1 \pm \sqrt{2}$ [note that if $a = 1 \pm \sqrt{2}$, then $a^+ = \frac{1}{2}$, $a^- = \frac{1}{2}(1 \mp \sqrt{2})$, and $s = \mp\sqrt{2}/2$]. In this case Eq. (3.8) gives $\partial_y V_0 = 0$, and Eq. (3.5) reduces to

$$\partial_u V_0 + (2 \pm \sqrt{2})\partial_{xx} V_0 = 0\,. \tag{3.9}$$

The constant solutions of (3.9) are discussed above. For linear solutions of (3.9) of the form $V_0 = Px$, P constant, the GKS metrics admit two Killing vectors $\{\partial_y, \partial_u\}$ which have the same character as above. These metrics correspond to the nontwisting cases of the Kóta–Perjés ($P > 0$) and Lukács ($P < 0$) solutions.[11,12] In fact, the identification can be explicitly made by taking $Q = 0$ in the Kóta–Perjés–Lukács solutions with $\tau \neq 0$, $Q = \text{const}$. The eigenfunctions of (3.9) are

$$V_{0_\gamma} = e^{-\gamma u + \alpha x}, \quad \alpha^2 = \gamma/(2 \pm \sqrt{2}), \quad \gamma = \text{const}. \tag{3.10}$$

By specifying boundary conditions, one can obtain metrics with arbitrary functions. If V_0 is given by (3.10) the GKS

TABLE II. The dependence of V_0 on a in the GKS solutions.

Generic case ($a \neq 1 \pm \sqrt{2}$)	$V_0 = \text{const}$ (Kasner solutions)
	$V_0 = Px$, P constant ($P > 0$ Kóta–Perjés solutions) ($P < 0$ Lukács solutions)
$a = 1 \pm \sqrt{2}$	
	$V_0 = C_\gamma e^{\gamma u + \alpha x}$ C_γ and γ constants $\alpha^2 = \gamma/(2 \pm \sqrt{2})$

metrics admit only one spacelike Killing vector $\{\partial_y\}$. The results for both the background and the GKS solutions are given in Tables I and II.

For completeness, we give the expressions for the Weyl spinor components (using $\partial_y V_0 = 0$),

$$\tilde{\Psi}_0 = \Psi_0 = -[as/(1+a^2)]r^{-2}, \tag{3.11a}$$

$$\tilde{\Psi}_1 = \Psi_1 = 0, \tag{3.11b}$$

$$\tilde{\Psi}_2 = [s/(1+a^2)]V_0 r^{-2-s}, \tag{3.11c}$$

$$\tilde{\Psi}_3 = [-(3-a^2)/4(1+a^2)]\left[\sqrt{2}\,\partial_x V_0\right] r^{-1-s-a^+}, \tag{3.11d}$$

$$\tilde{\Psi}_4 = [(1-a)/4]\left[2\,\partial_{xx} V_0\, r^{-s-2a^+}\right] - [as/(1+a^2)]V_0^2\, r^{-2-2s}. \tag{3.11e}$$

For $a^2 \mp 2a - 1 \neq 0$, the eigenvalues of the Weyl tensor are $\{-V(1+a)\Psi_0/a, -V(1-a)\Psi_0/a, 2V\Psi_0\}$. The solutions are flat for $a = \pm 1$, type D for $a = 0, \pm 3$, and algebraic general otherwise. For $a^2 \mp 2a - 1 = 0$, the Petrov classification algorithm is used to show that all solutions are algebraically general.

Note added in proof: The complete solution of vacuum metrics satisfying $\sigma = a\rho$ was recently obtained. Point: Generically, in addition to the metrics described here, there is only one solution not of GKS type. Thus the metrics given in Table II constitute all the vacuum GKS solutions with a shearing, nontwisting, null vector.

ACKNOWLEDGMENT

The author would like to thank Professor Gurses for help, encouragement, and valuable suggestions, and the referee for suggesting the final form of the manuscript.

[1] E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962); **4**, 998 (1963).
[2] D. Kramer, H. Stephani, M. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (VEB Deutscher, Berlin, 1980).
[3] A. H. Bilge, and M. Gürses, in *XI International Colloquium on Group Theoretical Methods in Physics* (Springer, Berlin, 1982).
[4] B. C. Xanthopoulos, J. Math. Phys. **19**, 1607 (1978).
[5] A. H. Bilge and M. Gürses, J. Math. Phys. **27**, 1819 (1986).
[6] C. J. Talbot, Commun. Math. Phys. **13**, 45 (1969).
[7] I. Robinson and J. R. Robinson, Int. J. Theor. Phys. **2**, 231 (1969).
[8] L. P. Hugston, Int. J. Theor. Phys. **4**, 267 (1971).
[9] A. Thompson, Tensor **17**, 92 (1966).
[10] E. Kasner, J. Math. **43**, 217 (1921).
[11] J. Kóta and Z. Perjés, J. Math. Phys. **13**, 1655 (1972).
[12] B. Lukacs, Acta Phys. Slovaca **33**, 225 (1983).
[13] M. Carmeli and M. Kaye, Ann. Phys. (NY) **99**, 188 (1976).

General relativistic fluid spheres with nonzero vacuum energy density

William A. Hiscock
Department of Physics, Montana State University, Bozeman, Montana 59717

(Received 1 July 1987; accepted for publication 30 September 1987)

Bounds are developed for the ratios M/R and m/R for fluid spheres in asymptotically de Sitter or anti-de Sitter space-times, where M is the mass of the fluid sphere, and m is the total mass interior to R: M plus the interior vacuum energy. This represents a generalization of the work of Buchdahl to the case of a nonvanishing vacuum energy density. In the asymptotically de Sitter case, it is possible to construct models which have $m/r \to \frac{1}{2}$. Further, it is shown that static fluid spheres can exist in an asymptotically de Sitter space with vacuum energy density ρ_v only if their radius satisfies $R \leqslant (8\pi\rho_v)^{1/2}$, a maximum radius smaller by a factor of $3^{-1/2}$ than the horizon size of the de Sitter space in the absence of a fluid sphere. If the vacuum energy density is negative, then the ratio m/R is shown to be bounded above by the asymptotically flat limit of $\frac{4}{9}$, and the radius of a positive total mass (m) sphere is shown to be bounded above by $R < (2\pi|\rho_v|)^{-1/2}$.

I. INTRODUCTION

The purpose of this paper is to examine the degree of compaction (revealed in the ratio M/R) possible for perfect fluid spheres in general relativity when the vacuum energy density is nonzero, and space-time is thus asymptotically described by the de Sitter or anti-de Sitter metrics[1] rather than being asymptotically flat. The "degree of compaction" is closely related (in the asymptotically flat case) to the surface redshift of a fluid sphere; this analysis is thus a generalization of the work of Buchdahl,[2] which showed that in the asymptotically flat case, any static fluid sphere in general relativity must have $M/R \leqslant \frac{8}{9}$, and hence a surface redshift $z \leqslant 2$.

The main motivation for this study is formal; the vacuum energy density today, if nonzero, is minute and unimportant in the construction of stellar models. While the vacuum energy density may have been much larger in the early universe,[3] conditions there probably precluded the formation of gravitationally condensed objects which could be well modeled by static fluid spheres. If some form of "shadow matter" exists,[4] defined by the property that it interacts with "ordinary" matter only via the gravitational interaction, then it is conceivable that such objects might exist (or might have existed). It is also conceivable that the universe may yet decay into a negative energy density anti-de Sitter state, in which the new low-temperature phenomenological laws might allow the formation of gravitationally condensed objects.

Section II sets out the basic equations and boundary conditions for general relativistic static fluid spheres with nonvanishing vacuum energy density. Since the space-times are not asymptotically flat, the notion of surface redshift is not well defined; thus, instead of bounding the redshift, bounds are determined for the total fluid mass divided by the radius of the sphere (M/R) and also the total mass (fluid plus vacuum energy) divided by the radius of the sphere (m/R). In Sec. III a lower limit on the fluid mass and an associated upper limit on the radius of fluid spheres in a positive vacuum energy density universe are derived and discussed. In Sec. IV Buchdahl's theorem is examined and extended easily to the case of nonzero vacuum energy density; its implications are then examined for the case of a negative vacuum energy density. An Appendix contains an explicit description of the solutions for uniform density fluid spheres with nonzero vacuum energy density.

II. FLUID SPHERES WITH NONZERO VACUUM ENERGY DENSITIES

The basic equations describing a static, spherically symmetric ball of perfect fluid in the presence of a nonzero vacuum energy density are exactly the same as in the asymptotically flat case[5]; the metric may be put into the standard form

$$ds^2 = -e^{2v(r)}\,dt^2 + [1-2m(r)/r]^{-1}\,dr^2 + r^2(d\theta^2 + \sin^2\theta\,d\phi^2)\,, \tag{1}$$

where

$$\frac{dm(r)}{dr} = 4\pi\rho r^2\,, \tag{2}$$

and

$$\frac{dv}{dr} = \frac{m(r)+4\pi r^3 p}{r^2[1-2m(r)/r]}\,, \tag{3}$$

where ρ is the total energy density (fluid plus vacuum), and p is the total pressure (again, including any vacuum contribution which may be present). Conservation of stress energy yields the Oppenheimer–Volkoff equation,

$$\frac{dp}{dr} = -(\rho+p)\frac{m(r)+4\pi r^3 p}{r^2[1-2m(r)/r]}\,. \tag{4}$$

The energy density and pressure may be divided into contributions from the vacuum and from the fluid, respectively,

$$\rho = \rho_v + \rho_f\,, \tag{5}$$

$$p = p_v + p_f\,, \tag{6}$$

where, since the stress-energy tensor of the vacuum has the form $-\rho g_{\alpha\beta}$,

 0022-2488/88/020443-03$02.50

$$p_v = -\rho_v \,. \tag{7}$$

The mass function $m(r)$ may also be divided into contributions from the fluid and vacuum energy,

$$m(r) = M(r) + 4\pi\rho_v r^3/3 \,, \tag{8}$$

where

$$M(r) = \int_0^r 4\pi\rho_f r^2\, dr \,. \tag{9}$$

The difference in the description of general relativistic fluid spheres in the case of nonzero vacuum energy density is in the boundary conditions imposed on Eqs. (2)–(9). The surface of the fluid sphere (denoted by $r = R$) is defined by the condition that $p_f = 0$ there. The metric functions for all $r \geqslant R$ are then given by

$$m(r) = M + 4\pi\rho_v r^3/3 \tag{10}$$

and

$$e^{2v(r)} = 1 - 2m(r)/r = 1 - 2M/r - 8\pi\rho_v r^2/3 \,, \tag{11}$$

where $M = M(R)$ is the total mass of the fluid. The usual boundary conditions at $r = 0$, namely that $m(0) = 0$, and that the pressure be bounded for all $r \neq 0$, will also be imposed. It is also assumed that the fluid density and pressure are non-negative, and that

$$\frac{dp_f}{d\rho_f} \geqslant 0 \,. \tag{12}$$

III. FLUID SPHERES WITH POSITIVE VACUUM ENERGY DENSITY

The Oppenheimer–Volkoff equation [Eq. (4)] may be used to create an interesting lower bound on the mass of a fluid sphere in the presence of a positive vacuum energy density. Evaluating Eq. (4) at the surface of the fluid sphere, where $p_f = 0$, one obtains

$$\frac{dp_f}{dr} = -\rho_f \frac{M(R) - 8\pi\rho_v R^3/3}{R\,[R - 2m(R)]} \,. \tag{13}$$

The quantity $R - 2m(R)$ must be non-negative at the surface of a static fluid sphere; for the moment assume it is in fact positive [the exceptional case where $R = 2m(R)$ will be discussed below]. Since it has been assumed that p_f is non-negative throughout the star, the pressure gradient at the surface, where $p_f = 0$, must also be non-negative, and hence the numerator of the right-hand side of Eq. (12) must be non-negative, i.e.,

$$M(R) \geqslant 8\pi\rho_v R^3/3 \,. \tag{14}$$

The existence of this bound does not depend on the magnitude or sign of ρ_v; the bound, however, is clearly only of consequence if $\rho_v > 0$. In the case of a uniform fluid density, $M(R) = 4\pi\rho_f R^3$, and Eq. (13) simply implies that

$$\rho_f \geqslant 2\rho_v \quad \text{(uniform density)}. \tag{15}$$

Equation (14) together with Eq. (10) imply that

$$m(R)/R \geqslant 4\pi\rho_v R^2, \tag{16}$$

for any fluid sphere in an asymptotically Schwarzschild–de Sitter space-time.

Now consider the possibility of an upper bound to the ratio $m(R)/R$. The requirement that the fluid sphere not extend past its own horizon may be written as

$$m(R)/R \leqslant \tfrac{1}{2} \,. \tag{17}$$

In the asymptotically flat case, better bounds are available than that given by Eq. (17), namely, the bound discovered by Buchdahl,[2] $m(R)/R \leqslant \frac{4}{9}$. When the vacuum energy density is positive, however, it is *always* possible to find fluid spheres which satisfy the equality in Eq. (17); no better bound (such as Buchdahl found) is available or valid. Equality in Eq. (17) may be achieved by the simple uniform density models (e.g., a portion of the Einstein universe) described in the Appendix.

Combining the constraints of Eqs. (16) and (17) yields

$$R \leqslant (8\pi\rho_v)^{-1/2}. \tag{18}$$

This is an upper bound on the radius of any fluid sphere in a space-time with positive vacuum energy density. Note that the maximum radius, given by equality in Eq. (18), is only $3^{-1/2}$ times the radius of the de Sitter space in the absence of the fluid sphere.

IV. FLUID SPHERES WITH NEGATIVE VACUUM ENERGY DENSITY

It is very easy to extend the analysis of Buchdahl[2] to the case of nonzero vacuum energy density; in the case of a negative vacuum energy density it is also fruitful. The details of the proof are unchanged; all that is required is that the boundary conditions be modified to agree with Eqs. (10) and (11). Buchdahl derived the following inequality:

$$m(r)/r \leqslant \tfrac{2}{9}\{1 - 6\pi r^2 p(r) + [1 + 6\pi r^2 p(r)]^{1/2}\} \,. \tag{19}$$

Applying Eq. (19) to the surface of the fluid sphere at $r = R$ gives

$$m(R)/R \leqslant \tfrac{2}{9}[1 + 6\pi\rho_v R^2 + (1 - 6\pi R^2\rho_v)^{1/2}] \,. \tag{20}$$

Equation (20) only holds for values of $\rho_v R^2 \leqslant (8\pi)^{-1}$. Larger values of $\rho_v R^2$ are spurious, and do not satisfy the original inequality which has been squared once to reach the form given in Eq. (20).

The largest value $m(R)/R$ can ever attain, according to Eq. (20), with a negative vacuum energy density, is the asymptotically flat value of $\frac{4}{9}$. This value may be approached for arbitrary ρ_v by simply making R small enough (when the fluid sphere is extremely small, compared to $|\rho_v|^{-1/2}$, then the vacuum energy density cannot have a large effect on its structure). The maximum value of $m(R)/R$ attainable steadily decreases as $|\rho_v| R^2$ increases; fluid spheres become less and less gravitationally compact. At the same time, the ratio M/R increases steadily,

$$M/R \leqslant \tfrac{2}{9}[1 + (1 - 6\pi R^2\rho_v)^{1/2}] \,. \tag{21}$$

The right-hand side of Eq. (21) increases steadily (and without bound) as $|\rho_v| R^2$ is increased. Although the minimum radius of a given fluid mass sphere is decreasing, as shown in Eq. (21), the horizon radius is shrinking even more rapidly as $|\rho_v| R^2$ is increased.

A final constraint which is of interest may be obtained from Eq. (20). If $|\rho_v| R^2$ is too large, then the total mass

associated with a fluid ball $m(R)$ can never be positive. Setting the right-hand side of Eq. (20) equal to zero gives

$$R < (2\pi|\rho_v|)^{-1/2}. \tag{22}$$

Equation (20) must be satisfied if the total mass measured at the surface of the fluid sphere is to be non-negative.

ACKNOWLEDGMENTS

I would like to thank P. Anderson and L. Lindblom for helpful conversations.

This research was supported by the National Science Foundation Grant No. PHY85-05484.

APPENDIX: UNIFORM DENSITY FLUID SPHERES

In this appendix the metric functions for a uniform density fluid sphere in a space-time with nonzero vacuum energy density are explicitly given. The uniform density spheres are described, as in the asymptotically flat case, by the interior Schwarzschild metric,[5,6] since it is the unique solution of the Einstein field equations for a static, spherically symmetric uniform density perfect fluid. The relations between the central pressure, total mass, and radius of the fluid sphere, however, depend on the boundary conditions at $r = R$, and are hence different in the present case from the usual relations which assume asymptotic flatness.[5] Integrating the Einstein equations [Eqs. (2)–(4)] for a fluid with uniform density, $\rho_f = \rho_0 = \text{const}$, with boundary conditions given by Eqs. (5)–(8), one finds that

$$m(r) = 4\pi(\rho_0 + \rho_v)r^3/3, \tag{A1}$$

$$e^{\nu(r)} = \frac{3[1-2m(R)/R]^{1/2} - (1-2\rho_v/\rho_0)[1-2m(r)/r]^{1/2}}{2(1+\rho_v/\rho_0)}, \tag{A2}$$

$$p(r) = (\rho_0+\rho_v)\frac{(1-2\rho_v/\rho_0)[1-2m(r)/r]^{1/2} - [1-2m(R)/R]^{1/2}}{3[1-2m(R)/R]^{1/2} - (1-2\rho_v/\rho_0)[1-2m(r)/r]^{1/2}}. \tag{A3}$$

The central pressure and the radius of the fluid sphere are related by

$$R = [(3/8\pi\rho)\{1-(1-2\rho_v/\rho_0)^2 \times [(\rho+p_c)/(\rho+3p_c)]^2\}]^{1/2}, \tag{A4}$$

where $\rho = \rho_v + \rho_0$. Notice that as $\rho_0 \to 2\rho_v$, $p(r) \to -\rho_v$, a constant, and $e^{2\nu(r)} \to 1 - 2m(R)/R$, also a constant. The interior metric is then the metric of a portion of the Einstein universe,[1] in which pressureless dust is held in equilibrium by the opposing forces of gravitational attraction and the repulsion caused by the nonzero vacuum energy density. This case, with $\rho_0 = 2\rho_v$, is the lowest density uniform stellar model possible when the vacuum energy density is positive, since the fluid pressure is zero everywhere. A uniform density fluid sphere with a smaller density would lack sufficient mass to hold itself together against the repulsive force associated with the vacuum energy density. Since the fluid pressure within the Einstein universe is zero (all pressure coming from the vacuum stress energy), the radius of a uniform density fluid sphere with $\rho_0 = 2\rho_v$ can be chosen arbitrarily, up to and including the radius of the event horizon, $R_H = 2m(R_H) = (8\pi\rho_v)^{-1/2}$.

[1] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-time* (Cambridge U. P., Cambridge, 1973), Chap. 5.
[2] H. A. Buchdahl, Phys. Rev. **116**, 1027 (1959).
[3] See, e.g., R. H. Brandenberger, Rev. Mod. Phys. **57**, 1 (1985).
[4] E. W. Kolb, D. Seckel, and M. S. Turner, Nature **314**, 415 (1985).
[5] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), Chap. 23.
[6] K. Schwarzschild, Sitzungsber. Preuss. Akad. Wiss. Phys. Math. Kl. **1916**, 424.

On the classification of vacuum zero Simon tensor solutions in relativity

J. P. Krisch
Department of Physics, The University of Michigan, Ann Arbor, Michigan 48109

(Received 21 July 1987; accepted for publication 7 October 1987)

The Perjes vector is decomposed on a Frenet basis. Geometric insights into the Perjes classification of zero Simon tensor vacuum solutions to the Einstein field equations are obtained. The behavior of the classification under a Killing vector preserving transform is studied.

I. INTRODUCTION

The metric for a stationary axis-symmetric space can be written as

$$dS^2 = \lambda(dt + \omega_i\, d\chi^i) - (h_{ij}/\lambda) d\chi^i\, d\chi^j. \tag{1}$$

The metric functions can be found by solving the field equations or by transforming from known solutions.

A particularly interesting set of solutions for this metric is the static vacuum metrics with a conformally flat three-space (g_{ij}). These spaces have a zero York tensor Y_{ij},[1]

$$Y_i^{\ j} = 2\epsilon^{jkl}(R_{ik} - \tfrac{1}{4}Rg_{ik})_{;l}, \tag{2}$$

where R_{ik} is calculated for g_{ik}. The perhaps best-known member of this solution set is the Schwarzschild metric with g_{ik} as the metric for the $t = \text{const}$, three-space.[2] A second example is the three-space of timelike Killing trajectories, ξ_μ, of the metric described by Eq. (1). This three-space has metric h_{ij} and will be conformally flat if $Y_i^{\ j}$ due to $h_i^{\ j}$ is zero.

Simon[3] has discussed a complex generalization of the York tensor. The Simon tensor C_{ij} was constructed to characterize the Kerr solution for Eq. (1) in the same way that the York tensor characterizes the Schwarzschild solution. A vanishing Simon tensor, coupled with the requirement of asymptotic flatness, will single out the Kerr solution to the vacuum field equations for the metric (1).[4] Perjes[5] has shown that the condition of zero Simon tensor alone, includes many interesting solutions that can be divided into three classes.

The purpose of this paper is to examine the transformation properties of the zero Simon tensor vacuum solutions. In particular, we will examine the behavior of the Perjes classes under a transformation that preserves the timelike Killing vector ξ_μ of a stationary axis-symmetric space-time.

In the next section we briefly review the Perjes classification of the zero Simon tensor solution. Some new geometric insights into this grouping are obtained. The transformation is discussed in Sec. III.

II. THE CLASSIFICATION OF ZERO SIMON TENSOR VACUUM SOLUTION

A. Perjes classification

The timelike Killing vector associated with the metric (1) has a norm λ and a vector twist ω_μ (Ref. 6),

$$\lambda = \xi_\mu \xi^\mu > 0, \tag{3}$$

$$\omega^\mu = \epsilon^{\mu\nu\rho\sigma}\xi_\nu \xi_{\rho;\sigma}/\sqrt{-g} = 2\overset{D}{\xi}{}^{\mu\nu}\xi_\nu, \tag{4}$$

where $\overset{D}{\xi}{}^{\mu\nu}$ is the dual Killing bivector

$$\overset{D}{\xi}{}^{\mu\nu} = \epsilon^{\mu\nu\rho\sigma}\xi_{\rho;\sigma}/2\sqrt{-g}, \tag{5}$$

and ω^μ describes the rotation of the Killing congruence. This will be made more precise in the next section. This ω^μ is curl-free and can be written as the gradient of a scalar potential, where the gradient is defined in the three-space of Killing trajectories H; h_{ij},

$$\omega_i = -D_i\phi, \tag{6}$$

where i runs over spatial indices and D_i is the covariant derivative in h_{ij}. Here ω_i can be pulled back to the four-space $g_{\mu\nu}$ to create ω_μ.[6] The negative sign is added to conform to Perjes' conventions. An Ernst potential can be formed from the norm and scalar twist

$$\tau = \lambda + i\phi. \tag{7}$$

The Simon tensor is defined in terms of the Ernst potential τ,

$$C_i^{\ j} = (2\epsilon^{lkj}/\lambda^2)\left[\tau_{i,l}\tau_k - h_{il}h^{mn}\tau_{m;[n}\tau_{k]}\right], \tag{8}$$

where $\tau_i = \partial_i\tau$ is the gradient of the scalar potential τ. For static space-times, $\phi = 0$; the Simon tensor is equivalent to the York tensor.

The vacuum spaces with zero Simon tensor can be classified by using the vector $\mathbf{G}$ (Refs. 5 and 7),

$$G_\mu = (1/2\lambda)\tau_\mu = (1/2\lambda)(\lambda_\mu + i\phi_\mu). \tag{9}$$

We use μ as an index of the pulled back functions. Here λ_μ is related to the normal to the surface $\lambda = \text{const}$, and ϕ_μ is related to the twist associated with that trajectory. The three classes of zero Simon tensor vacuum solutions correspond to (1) $\mathbf{G}\cdot\mathbf{G} = 0$, the null class (2) $\mathbf{G}\times\mathbf{G}^* = 0$, the degenerate class, and (3) $\mathbf{G}\times\mathbf{G}^* \neq 0$, the general class. Perjes identified specific solution sets within each class by the behavior of spin coefficients for a triad (l_μ, $\bar{m}_\mu$, m_μ) defined on h_{ij}. The spin coefficient method of Perjes is very useful in identifying the Petrov class of solutions belonging to each set. We found that decomposing $\mathbf{G}$ in terms of a Frenet–Serret tetrad of vectors allowed some new geometric insights into the Perjes classification.

B. Frenet decomposition of G_μ—non-null Killing vector

The groups that Perjes uses to classify the zero Simon tensor solutions, differentiate properties of the Killing bivector $\xi_{\mu;\nu}$ and of the congruence of Killing vectors forming h_{ij}.

 0022-2488/88/020446-03$02.50

The Killing vector and bivector properties are conveniently described in terms of the orthonormal Frenet tetrad $(e^{\mu}_{(0)}, A^{\mu}, B^{\mu}, C^{\mu})$ with

$$-e^{\mu}_{(0)}e_{\mu(0)} = A^{\mu}A_{\mu} = B^{\mu}B_{\mu} = C^{\mu}C_{\mu} = -1. \quad (10)$$

The timelike tetrad member is chosen to lie along the Killing vector

$$e^{\mu}_{(0)} = \xi^{\mu}/\sqrt{\lambda}. \quad (11)$$

The spatial triad $(A^{\mu}, B^{\mu}, C^{\mu})$ are the normal and first and second binormals to the Killing trajectory. The trajectories can be described by three scalars k, τ_1, and τ_2, the curvature and first and second torsions, respectively. These scalars enter the absolute derivatives of the tetrad,

$$\begin{bmatrix} \dot{e}^{\mu}_{(0)} \\ \dot{A}^{\mu} \\ \dot{B}^{\mu} \\ \dot{C}^{\mu} \end{bmatrix} = \begin{bmatrix} 0 & k & 0 & 0 \\ k & 0 & \tau_1 & 0 \\ 0 & -\tau_1 & 0 & \tau_2 \\ 0 & 0 & -\tau_2 & 0 \end{bmatrix} \begin{bmatrix} e^{\mu}_{(0)} \\ A^{\mu} \\ B^{\mu} \\ C^{\mu} \end{bmatrix}, \quad (12)$$

where, for example, $\dot{A}^{\mu} = A^{\mu}{}_{;\nu}e^{\nu}(0)$

The Frenet formalism is ideally suited to the discussion of Killing vectors and bivectors. Normally a timelike vector derivative is written in terms of acceleration, angular speed, expansion, and shear. Because the Killing bivector is antisymmetric, only the three parameters describing acceleration and angular speed are nonzero. These can easily be identified with the Frenet scalars. The bivector expansion is

$$\xi_{\mu;\nu} = k(-\xi_{\mu}A_{\nu} + \xi_{\nu}A_{\mu}) + \sqrt{\lambda}\,\tau_1(A_{\mu}B_{\nu} - A_{\nu}B_{\mu}) + \sqrt{\lambda}\,\tau_2(B_{\mu}C_{\nu} - B_{\nu}C_{\mu}). \quad (13)$$

The parameters k, τ_1, and τ_2 are constant along the Killing trajectory.[8] The "acceleration" and "angular speed" of the Killing vector ξ_{μ} are[8]

$$\dot{\xi}_{\mu} \equiv \xi_{\mu;\nu}\xi^{\nu} = \lambda k A_{\mu},$$
$$\Omega^{\mu} = \tfrac{1}{2}\omega^{\mu} = -\phi^{\mu}/2 = \overset{D}{\xi}_{\mu;\nu}\xi^{\nu} = -\lambda(\tau_1 C^{\mu} + \tau_2 A^{\mu}), \quad (14)$$

Ω^{μ} describes the rotation of the Frenet frame relative to a frame that is Fermi–Walker transported along the trajectory. A static space has both torsions zero. From Eq. (13) we find the bivector norms and products can be parametrized in terms of the Frenet scalars and are given by

$$\xi^{\mu;\nu}\xi_{\mu;\nu} = -2\lambda(k^2 - \tau_1^2 - \tau_2^2),$$
$$\overset{D}{\xi}{}^{\mu;\nu}\xi_{\mu;\nu} = -4\lambda k\tau_2. \quad (15)$$

The vector G can be written in terms of Frenet vectors,

$$G_{\mu} = (1/2\lambda)(\lambda_{\mu} + i\phi_{\mu}) = -kA_{\mu} + i(\tau_1 C_{\mu} + \tau_2 A_{\mu}), \quad (16)$$

G has no component along the first binormal B_{μ}.

The first class of solutions corresponds to $\mathbf{G}\cdot\mathbf{G} = 0$. Using Eq. (16), this is equivalent to

$$k^2 = \tau_1^2 + \tau_2^2, \quad (17)$$
$$k\tau_2 = 0. \quad (18)$$

The first condition requires the Killing bivector have zero norm. The second condition requires that the Killing bivector be orthogonal to its dual, or that the bivector be simple. These are the conditions for the bivector to be null, as pointed out by Perjes.[4] The first or null class then consists of solutions with null Killing bivectors. Simple bivectors are expressible as a single antisymmetric product. Using Eq. (13) we find

$$\xi_{\mu;\nu} = k\sqrt{\lambda}\,[A_{\mu}(e^{(0)}_{\nu} - B_{\nu}) - A_{\nu}(e^{(0)}_{\mu} - B_{\mu})]. \quad (19)$$

The two vectors are A_{μ} and the null vector $e^{(0)}_{\mu} - B_{\mu}$. A possible Perjes triad $(l,m,\bar{m})$ for this class is $(B,(A \pm iC)/\sqrt{2})$.

The spaces with a null bivector are algebraically special.[9] The restriction to non-null Killing vector restricts the space-time to be Petrov III or N.[10,11]

The second class of zero Simon tensor solutions is characterized by $\mathbf{G}\times\mathbf{G}^* = 0$. Using (15), this can be written as

$$(G \times G^*)_{\mu} = i2\tau_1 k B_{\mu}. \quad (20)$$

Normally k is taken nonzero so that for class 2 solutions there is no first torsion. The bivector is not null. For this class G_{μ} can be written

$$G_{\mu} = (-k + i\tau_2)A_{\mu}. \quad (21)$$

This set includes both static $(\phi_{\mu} = 0)$ and stationary $(\phi_{\mu} = +i\tau_2 A_{\mu})$ solutions. It is interesting that the rotation of the Frenet frame relative to the Fermi transported frame lies along the acceleration. A possible triad for this class is $(l,m,\bar{m}) = (A,B \pm iC/\sqrt{2})$.

One could choose $k = 0$ in this class. The G vector is

$$G_{\mu} = i(\tau_1 C_{\mu} + \tau_2 A_{\mu}).$$

This subclass seems to be artificial since the only example we were able to find also had τ_1 and τ_2 zero. We will see it is mathematically nice and so include it for completeness.

The general class is described by Eq. (20). There are no restrictions on any of the Killing scalars.

The Frenet decomposition makes very clear how G behaves from class I to II. In addition it also describes the classwise behavior of the bivector. This behavior is especially interesting in the case when the second torsion, τ_2, is zero. The first class has bivector norm zero and G perpendicular to $l = B$. The second class has two subclasses. In II a, the bivector norm is negative and G is along $l = A$. In II b, the bivector norm is positive and G is along $l = C$. In all three cases, the bivector is orthogonal to its dual. This is summarized in Table I.

TABLE I. Behavior of Perjes vector and Killing bivector norm for zero second torsion.

Class	G	$\xi^{\mu;\nu}\xi_{\mu;\nu}$	l	k	τ_1	τ_2
I $G\times G = 0$	$k(-A + iC)$	0	B	k	k	0
II a $G\times G^* = 0$	$-kA$	$-2\lambda k^2 < 0$	A	k	0	0
II b $G\times G^* = 0$	$i\tau_1 C$	$2\lambda\tau_1^2 > 0$	C	0	τ_1	0

C. Frenet decomposition—null Killing vector

The decomposition discussed in the previous section is only valid for non-null Killing vectors, $\lambda \neq 0$. If we wish to discuss the behavior of G on a Killing horizon, a different set of Frenet tetrads must be used because of differences in defining trajectory parameters.

If the Killing vector is null but has a non-null normal, the appropriate tetrad is[12] described by

$$\begin{bmatrix} \dot{\xi}^\mu \\ \dot{A}^\mu \\ \dot{B}^\mu \\ \dot{C}^\mu \end{bmatrix} = \begin{bmatrix} 0 & k_1 & 0 & 0 \\ k_2 & 0 & k_1 & 0 \\ 0 & k_2 & 0 & k_3 \\ k_3 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \xi^\mu \\ A^\mu \\ B^\mu \\ C^\mu \end{bmatrix}, \tag{22}$$

with $\xi^\mu B_\mu = 1 = -A^\mu A_\mu = -B^\mu C_\mu$, all others zero. In terms of these vectors

$$\begin{aligned} &\xi^\mu = \xi^\mu, \quad \lambda = 0, \quad n^\mu = -k_1 A^\mu, \\ &n^\mu n_\mu \neq 0, \quad \omega^\mu = -k_1 C^\mu, \end{aligned} \tag{23}$$

n^μ is the normal to the $\lambda = 0$ Killing horizon and should be compared to λ^μ in the previous section. The Killing bivector has norm and dual product[12]

$$\xi_{\mu;\nu}\xi^{\mu;\nu} = -4k_1k_2, \quad \xi_{\mu;\nu}\overset{D}{\xi}{}^{\mu;\nu} = -4k_1k_3. \tag{24}$$

For this case the vector G is defined by

$$G^\mu \equiv -k_1(A^\mu - iC^\mu), \tag{25}$$

which is clearly of the Perjes null class with no bivector restriction. A metric example of this case is the Kerr solution. The Killing horizon ($\lambda = 0$) is not coincident with the event horizon ($\lambda^\mu \lambda_\mu = 0$) except on the axis of rotation. On the Kerr Killing horizon, G^μ is a null vector.

If the Killing vector and the normal are both null, as for example on the Schwarzschild horizon, then a single null vector N_μ parametrizes all vectors and we have[12,13]

$$\begin{aligned} &\xi^\mu = N^\mu, \quad \lambda = 0, \quad \lambda^\mu = \epsilon N^\mu, \quad \lambda^\mu \lambda_\mu = 0, \\ &\omega^\mu = \delta N^\mu, \quad \omega^\mu \omega_\mu = 0, \quad G = \gamma N^\mu, \quad G^\mu G_\mu = 0. \end{aligned} \tag{26}$$

For this case the classes merge.

III. TRANSFORMATIONS

The transformations we want to consider are transformations among the three classes of zero Simon tensor solutions which preserve the Killing vector. We consider only non-null Killing vectors. This transformation has been described by Geroch.[6] It generalizes the work of Ehlers[14] and Harrison.[15] The transform is a projective transform on the complex function $i\tau$,

$$i\tau' = (ai\tau + b)/(ci\tau + d), \tag{27}$$

with a, b, c, d as constants. The transform is performed in the space h; h_{ij} of Killing trajectories with the requirement

$$h_{ij}' = h_{ij}. \tag{28}$$

Under the transformation the Killing scalars transform as[16]

$$\begin{aligned} &\phi' = \{[a\phi - b][d - c\phi] - ac\lambda^2\}/[(d - c\phi)^2 + \lambda^2 c^2], \\ &\lambda' = \lambda(ad - bc)/[(d - c\phi)^2 + \lambda^2 c^2]. \end{aligned} \tag{29}$$

Under this transform, the zero Simon tensor is preserved. We wish to find the effect of this transform on the vector G. Take the covariant derivative of Eq. (29) in the Killing space. Pulling back to the metric space one finds

$$\begin{aligned} &\lambda'_\mu = \lambda_\mu \cos\alpha + \phi_\mu \sin\alpha, \\ &\phi'_\mu = \phi_\mu \cos\alpha - \lambda_\mu \sin\alpha, \end{aligned} \tag{30}$$

with

$$\begin{aligned} &\cos\alpha = [(d - c\phi)^2 - c^2\lambda^2]/[(d - c\phi)^2 + c^2\lambda^2], \\ &\sin\alpha = 2\lambda c(d - c\phi)/[(d - c\phi)^2 + c^2\lambda^2]. \end{aligned} \tag{31}$$

Using these vectors to form G we have

$$G'_\mu = (\lambda/\lambda')e^{-i\alpha}G_\mu, \tag{32}$$

where G' differs from G only by a scaled phase factor. Class is clearly preserved by this transformation. For example, this transform will take the Schwarzschild metric into the Taub–Nut[17] space-time. We would expect the Schwarzschild and Taub–NUT spaces to be of the same Perjes class II.

In conclusion, we have decomposed the Perjes G vector onto a Frenet basis and discussed the Frenet scalar relations in each class. We find also that Perjes class is preserved under a transform that preserves the Killing vector.

[1] L. P. Eisenhart, *Riemannian Geometry* (Princeton U. P., Princeton, 1966), p. 92.
[2] A. K. Raychaudhuri and S. R. Maiti, J. Math. Phys. **20**, 245 (1979).
[3] W. Simon, Gen. Relativ. Gravit. **16**, 465 (1984).
[4] Z. Perjes, *Proceedings of the 3rd Quantum Gravity Meeting*, Moscow, 1984, edited by M. A. Markov (World Scientific, Singapore, 1984), p. 446.
[5] Z. Perjes, Int. J. Theor. Phys. **10**, 217 (1974).
[6] R. Geroch, J. Math. Phys. **12**, 918 (1971); **13**, 394 (1972).
[7] B. Lukacs, Z. Perjes, and A. Sebestyen, Gen. Relativ. Gravit. **15**, 511 (1983).
[8] E. Honig, E. L. Schucking, and C. V. Vishveshwara, J. Math. Phys. **15**, 775 (1974).
[9] G. C. Debney, Jr., J. Math Phys. **12**, 1088, 2372 (1971).
[10] C. B. G. McIntosh, Gen. Relativ. Gravit. **7**, 215 (1976).
[11] R. Catenacci, A. Marzuoli, and F. Salmistraro, Gen. Relativ. Gravit. **12**, 575 (1980).
[12] C. C. Dyer and E. Honig, J. Math. Phys. **20**, 1 (1979).
[13] C. C. Dyer and E. Honig, J. Math. Phys. **20**, 409 (1979).
[14] J. Ehlers, in *Les theories relatavistics de la gravitation* (CNRS, Paris, 1959).
[15] B. K. Harrison, J. Math. Phys. **9**, 1744 (1968).
[16] J. P. Krisch, J. Math. Phys. **24**, 2783 (1983).
[17] J. P. Krisch, J. Math. Phys. **22**, 663 (1981).

Bianchi type VI_0 space-times with perfect fluid source

Shri Ram
Department of Applied Mathematics, Institute of Technology, Banaras Hindu University, Varanasi 221005, India

(Received 4 September 1986; accepted for publication 2 September 1987)

An analytic solution to Einstein's field equations is presented for the Bianchi type VI_0 class of models. The energy-momentum tensor is of the perfect fluid type. The solution corresponds to a locally rotationally symmetric and expanding cosmological universe which would give an essentially empty universe for large time. Some kinematical properties of the solution are discussed.

I. INTRODUCTION

The simplest model of the expanding universe is well represented by the Friedmann–Robertson–Walker (FRW) models which are both spatially homogeneous and isotropic. FRW models, in some sense, are good global approximations of the present universe, but it is unreasonable to assume that the regular expansion predicted by these models is also suitable for describing the early stages of the universe. After spatially homogeneous and isotropic models, the simplest cosmologies are spatially homogeneous and anisotropic models. Bianchi spaces[1] play an important role in constructing models of spatially homogeneous cosmologies. Homogeneous cosmological models filled with matter together with specified equations of state have already been widely studied. Recently Bayin and Krisch[2] presented some analytic solutions to Einstein's field equations with perfect fluids for Bianchi type I and III spaces. Using metric solutions parametrized over several Bianchi types is a useful tool in constructing and studying current modern cosmologies since spatially homogeneous and anisotropic Bianchi models are a method somewhere between FRW models and completely nonhomogeneous and anisotropic universes.

We obtain an analytic solution to Einstein's field equation with perfect fluids for Bianchi type VI_0 spaces. The solution represents a locally rotationally symmetric (LRS) cosmological model that expands at large time to a very dilute, essentially empty space for vanishing cosmological constant.

II. FIELD EQUATIONS AND THEIR SOLUTIONS

The field equations in general relativity are

$$R_{\mu\nu} - \tfrac{1}{2} R g_{\mu\nu} + \Lambda g_{\mu\nu} = -8\pi T_{\mu\nu}. \tag{2.1}$$

The energy-momentum tensor for a perfect fluid is

$$T_{\mu\nu} = (\rho + p) V_\mu V_\nu - p g_{\mu\nu}, \quad V_\mu V^\mu = -1. \tag{2.2}$$

The metric of Bianchi type VI_0 is

$$ds^2 = dt^2 - A(t)dx^2 - B(t)e^{-2mx}dy^2 - C(t)e^{2mx}dz^2, \tag{2.3}$$

where A, B, C are cosmic scale functions and m is a nonzero constant.[3] We number the coordinates x,y,z,t as 1,2,3,4, respectively. In comoving coordinates the field equations to be considered are

$$8\pi p = -\tfrac{1}{2}(B''/B) - \tfrac{1}{2}(C''/C) + \tfrac{1}{4}(B'/B)^2 + \tfrac{1}{4}(C'/C)^2 - \tfrac{1}{4}(B'C'/BC) - m^2/A + \Lambda, \tag{2.4}$$

$$8\pi p = -\tfrac{1}{2}(A''/A) - \tfrac{1}{2}(C''/C) + \tfrac{1}{4}(A'/A)^2 + \tfrac{1}{4}(C'/C)^2 - \tfrac{1}{4}(A'C'/AC) + m^2/A + \Lambda, \tag{2.5}$$

$$8\pi p = -\tfrac{1}{2}(A''/A) - \tfrac{1}{2}(B''/B) + \tfrac{1}{4}(A'/A)^2 + \tfrac{1}{4}(B'/B)^2 - \tfrac{1}{4}(A'B'/AB) + m^2/A + \Lambda, \tag{2.6}$$

$$8\pi p = \tfrac{1}{4}(A'B'/AB) + \tfrac{1}{4}(B'C'/BC) + \tfrac{1}{4}(A'C'/AC) - m^2/A - \Lambda, \tag{2.7}$$

$$B'/B - C'/C = 0, \tag{2.8}$$

where a prime denotes a derivative with respect to t.

Equation (2.8), on integration, gives $B = nC$, n being an arbitrary constant. Without loss of generality one can take $n = 1$. Then Eqs. (2.4) and (2.5) give

$$\tfrac{1}{2}(B''/B - A''/A) + \tfrac{1}{4}(A'/A)(A'/A - B'/B) + 2m^2/A = 0, \tag{2.9}$$

which is a single equation in two unknowns. For a complete solution of this equation we require an extra condition. A number of solutions to (2.9) could be generated by assuming solutions for A (or B). Here we obtain a physically realistic solution by assuming

$$A = (at + b)^2, \tag{2.10}$$

a and b being arbitrary constants. Substitution of (2.10) into (2.9) gives

$$(at + b)^2 B'' - a(at + b)B' + 4m^2 B = 0, \tag{2.11}$$

which is a Lagrange linear differential equation. The general solution of (2.11) is

$$B = \alpha(at + b)^{1+\gamma} + \beta(at + b)^{1-\gamma}, \tag{2.12}$$

where α and β are integration constants and

$$\gamma = (a^2 - 4m^2)^{1/2}/a. \tag{2.13}$$

Hence the metric of our solution is

$$ds^2 = dt^2 - (at + b)^2 dx^2 - \{\alpha(at + b)^{1+\gamma} + \beta(at + b)^{1-\gamma}\}(e^{-2mx}dy^2 + e^{2mx}dz^2). \tag{2.14}$$

The pressure and density are

 0022-2488/88/020449-02$02.50

$$8\pi p = -a^2\left\{\frac{\alpha\gamma(1+\gamma)(at+b)^{\gamma-1}-\beta\gamma(1-\gamma)(at+b)^{-\gamma-1}}{\alpha(at+b)^{1+\gamma}+\beta(at+b)^{1-\gamma}}\right\}+\frac{a^2}{4}\left\{\frac{\alpha(1+\gamma)(at+b)^{\gamma}+\beta(1-\gamma)(at+b)^{-\gamma}}{\alpha(at+b)^{1+\gamma}+\beta(at+b)^{1-\gamma}}\right\}^2 - m^2/(at+b)^2+\Lambda, \tag{2.15}$$

$$8\pi\rho = \frac{a^2}{2}\left\{\frac{\alpha(1+\gamma)(at+b)^{\gamma-1}+\beta(1-\gamma)(at+b)^{-\gamma-1}}{\alpha(at+b)^{1+\gamma}+\beta(at+b)^{1-\gamma}}\right\}+\frac{a^2}{4}\left\{\frac{\alpha(1+\gamma)(at+b)^{\gamma}+\beta(1-\gamma)(at+b)^{-\gamma}}{\alpha(at+b)^{1+\gamma}+\beta(at+b)^{1-\gamma}}\right\}^2 - m^2/(at+b)^2-\Lambda, \tag{2.16}$$

III. PHYSICAL AND KINEMATICAL PROPERTIES OF THE MODEL

We discuss the properties of the shear tensor. It has been pointed out by Collins and Wainwright[4] that the shear tensor $\sigma_{\mu\nu}$ plays an important role in general relativistic cosmological and stellar models. The shear tensor arises in the decomposition of four-velocity vector of the fluid, i.e.,

$$V_{\mu;\nu} = -V'_\nu V_\nu + \omega_{\mu\nu} + \sigma_{\mu\nu} + \Theta h_{\mu\nu}/3, \tag{3.1}$$

$$V'_\mu = V_{\mu;\nu}V^\nu, \quad V'_\mu V^\mu = 0, \tag{3.2}$$

$$\omega_{\mu\nu} = V_{[\mu;\nu]} + V'_{[\mu}V_{\nu]}, \quad \omega_{\mu\nu}V^\nu = 0, \tag{3.3}$$

$$h_{\mu\nu} = g_{\mu\nu} + V_\mu V_\nu, \quad h_{\mu\nu}V^\nu = 0, \tag{3.4}$$

$$\sigma_{\mu\nu} = V_{(\mu;\nu)} + V'_{(\mu}V_{\nu)} - \Theta h_{\mu\nu}/3, \quad \sigma_{\mu\nu}V^\nu = 0, \tag{3.5}$$

and $\Theta = V^\mu_{;\mu}$, where V'^μ, $\omega_{\mu\nu}$, Θ, and $\sigma_{\mu\nu}$ are called acceleration, rotation, expansion, and shear, respectively, and a semicolon means a covariant derivative.[5]

For the model (2.14) we find

$$\Theta = a\left\{\frac{\alpha(2+\gamma)(at+b)^{\gamma}+\beta(2-\gamma)(at+b)^{-\gamma}}{\alpha(at+b)^{1+\gamma}+\beta(at+b)^{1-\gamma}}\right\}, \quad V'\mu = 0. \tag{3.6}$$

The shear scalar σ defined by $\sigma^2 = \frac{1}{2}\sigma_{\mu\nu}\sigma^{\mu\nu}$ has the value

$$\sigma = \frac{a}{2\sqrt{3}}\left\{\frac{\alpha(1-\gamma)(at+b)^{\gamma}+\beta(1+\gamma)(at+b)^{-\gamma}}{\alpha(at+b)^{1+\gamma}+\beta(at+b)^{1-\gamma}}\right\}, \tag{3.7}$$

which is nonzero for all values of t $(0<t<\infty)$ and drops to zero at infinite time $(t\to\infty)$. Thus the model is anisotropic for $0<t<\infty$.

As $t\to\infty$, $p(\infty)=\Lambda$ and $\rho(\infty)=-\Lambda$, so $\Lambda=0$ would give at large time an essentially empty universe. The ratio

$$\frac{\sigma}{\Theta} = \frac{1}{2\sqrt{3}}\left\{\frac{\alpha(1-\gamma)(at+b)^{\gamma}+\beta(1+\gamma)(at+b)^{-\gamma}}{\alpha(2+\gamma)(at+b)^{\gamma}+\beta(2-\gamma)(at+b)^{-\gamma}}\right\} \tag{3.8}$$

tends to $(1/2\sqrt{3})(1-\gamma)/(2+\gamma)$ as $t\to\infty$. Thus the shear scalar σ does not tend to zero faster than the expansion. The ratio in (3.8) tends to zero as $t\to\infty$ for $\gamma=1$. But for $\gamma=1$, Eq. (2.13) gives $m=0$ and consequently we obtain the LRS Bianchi type I perfect fluid universe with spatial isotropy at infinite time. For the model (2.14) all the components of the rotation tensor are zero. Hence (2.14) represents an expanding and anisotropic cosmological model in which all of the fluids are acceleration- and rotation-free.

ACKNOWLEDGMENT

The author is grateful to Professor K. P. Singh for regular encouragement and valuable suggestions.

[1] L. Bianchi, Mem. Soc. Ital. Sci. Nat. Mus. Civ. Stor. Nat. Milano **11**, 267 (1978).
[2] S. S. Bayin and J. P. Krisch, J. Math. Phys. **27**, 264 (1986).
[3] M. P. Ryan and L. C. Shepley, *Homogeneous Relativistic Cosmologies* (Princeton U.P., Princeton, NJ, 1975).
[4] C. B. Collins and J. Wainwright, Phys. Rev. D **27**, 1209 (1983).
[5] G. F. R. Ellis, *General Relativity and Cosmology* (Academic, New York, 1971), p. 104.

Correlation inequalities for multicomponent ferromagnets

Kei-ichi Kondo and Yūki Sugiyama
Department of Physics, Faculty of Science, Nagoya University, Chikusa-ku, Nagoya 464, Japan

(Received 19 May 1987; accepted for publication 7 October 1987)

A systematic method to obtain a series of new correlation inequalities for a class of two-component vector spin systems is presented. These correlation inequalities are applied to lattice scalar field models with two-body, anisotropic or isotropic ferromagnetic interactions and interaction potential of even polynomials, especially the $\lambda(|\Phi|^2)^2$ model, which includes the plane rotator model as a special case. Here an external field $\mathbf{h} = (h,H)$, $h \geqslant 0$, $H \geqslant 0$, is present. The possible way to extend our method to the N-component ($N \geqslant 3$) case is also discussed.

I. INTRODUCTION

Correlation inequalities are a very powerful tool in rigorous studies of quantum field theory and statistical mechanics.[1,2] For a set of lattice sites Λ, let $A = \{A_i\}_{i\in\Lambda}$ be a family of non-negative integers, only a finite number of which are nonzero. Define $\varphi^A = \Pi_{i\in\Lambda}\varphi_i^{A_i}$ and $|A| = \Sigma_{i\in\Lambda}A_i$, where the φ_i are real-valued scalar fields. For appropriate "ferromagnetic" expectations $\langle\cdot\rangle$, the first and second GKS inequalities[1] read

$$\langle\varphi^A\rangle \geqslant 0, \qquad \langle\varphi^A;\varphi^B\rangle \equiv \langle\varphi^A\varphi^B\rangle - \langle\varphi^A\rangle\langle\varphi^B\rangle \geqslant 0, \tag{1.1}$$

for *all* A,B.

But, it is unknown, in general, whether similar correlation inequalities hold for twice and more-times-truncated expectations, although they would have many important applications in rigorous studies of quantum field theory and statistical mechanics.[2,3] For nearest neighbor ferromagnetic interactions, e.g., the correlation inequality $\langle\varphi^A;\varphi^B;\varphi^C\rangle \geqslant 0$ (which has *not* been proved) would imply the convexity of the susceptibility and the monotonicity of the specific heat w.r.t. (with respect to) the inverse temperature J in the high temperature region $J < J_c$ for $|A| = |B| = |C| = 2$ (Ref. 3), and the convexity of the magnetization w.r.t. J for $|A| = 1$, $|B| = |C| = 2$ in the presence of the magnetic field. For the twice-truncated expectation, the following correlation inequality (called the new Lebowitz inequality[4]) is known to hold, for *all* A,B,C;

$$\langle\varphi^A;\varphi^B;\varphi^C\rangle \geqslant -2\min[\langle\varphi^A\rangle\langle\varphi^B;\varphi^C\rangle, \langle\varphi^B\rangle\langle\varphi^A;\varphi^C\rangle, \langle\varphi^C\rangle\langle\varphi^A;\varphi^B\rangle], \tag{1.2}$$

where

$$\langle\varphi^A;\varphi^B;\varphi^C\rangle \equiv \langle\varphi^A\varphi^B\varphi^C\rangle - \langle\varphi^A\rangle\langle\varphi^B\varphi^C\rangle - \langle\varphi^B\rangle\langle\varphi^A\varphi^C\rangle - \langle\varphi^C\rangle\langle\varphi^A\varphi^B\rangle + 2\langle\varphi^A\rangle\langle\varphi^B\rangle\langle\varphi^C\rangle. \tag{1.3}$$

Yet, it is incorrect to expect that $\langle\varphi^A;\varphi^B;\varphi^C\rangle \geqslant 0$ holds for all A,B,C. In fact, for $|A| = |B| = |C| = 1$, the GHS inequality holds,[5] namely

$$\langle\varphi_i;\varphi_j;\varphi_k\rangle \leqslant 0, \tag{1.4}$$

which implies the concavity of the magnetization and monotonicity of the susceptibility w.r.t. the magnetic field h. Moreover, for the four-times-truncated expectation, we have the Lebowitz inequality[6]

$$\langle\varphi_i;\varphi_j;\varphi_k;\varphi_l\rangle \leqslant 0 \tag{1.5}$$

in the absence of external magnetic field.

In the earlier work,[7] we presented a series of correlation inequalities for higher-times-truncated expectations. For example, in the presence of an external field, we proved

$$\langle\varphi_i;\varphi_j;\varphi_k;\varphi_l\rangle \geqslant -4\langle\varphi_i;\varphi_j\rangle\langle\varphi_k;\varphi_l\rangle, \tag{1.6a}$$

$$\langle\varphi_i;\varphi_j;\varphi_k;\varphi_l\rangle \leqslant -4\langle\varphi_i\rangle\langle\varphi_j;\varphi_k;\varphi_l\rangle. \tag{1.6b}$$

Note that the inequality (1.6b) reduces to (1.5) in the absence of an external field. Inequalities of this type [i.e., (1.2), (1.6a), and (1.6b)] turn out to be very useful in the rigorous study of the continuum limit and the critical behavior of the broken-symmetry lattice scalar field models.[8] But they are restricted to one-component ferromagnets whose single spin measures belong to the EMN (Ellis–Monroe–Newman) class.[9–12]

The extension of correlation inequalities to multicomponent ferromagnets has been performed by several authors.[13–26] For two-component spin systems where the spin is denoted as $\Phi = (\varphi,\xi)$, $|\Phi| = 1$, Monroe[14] presented the following correlation inequalities of GKS type:

$$\langle\varphi^A;\varphi^B\rangle \geqslant 0, \quad \langle\xi^A;\xi^B\rangle \geqslant 0, \quad \langle\varphi^A;\xi^B\rangle \leqslant 0. \tag{1.7}$$

We would like to obtain explicit correlation inequalities for the three- and four-times- (at least) truncated expectations for multicomponent ferromagnets whose spin variables Φ are allowed to take *unbounded* values. Although the broken-symmetry scalar $\lambda\Phi^4$ theory plays an important role in the Higgs mechanism,[27] it has been recently proved[8] that (under reasonable assumptions) the continuum scalar $\lambda(\varphi^4)_d$ field theory obtained from the corresponding lattice regularized model by taking the continuum limit is trivial in $d > 4$ dimensions, if one adopts the renormalization condition that the vacuum expectation value of the single (renormalized) scalar field remains finite and nonzero in this limit. However, the result is restricted to the one-component model.[8] This work has begun with the motivation of extending the triviality proof of broken-symmetry $\lambda(\varphi^4)_d$ field theory to the two- and more-component cases.

In Sec. II, we present a strategy based on duplicated variables for proving correlation inequalities. In Sec. III, we consider the anisotropic case. Section IV is devoted to the

 0022-2488/88/020451-09$02.50

isotropic case. In this paper, explicit correlation inequalities are derived only for a class of two-component ferromagnets. In the final section, we discuss the possibility of extending our method to N-component ($N \geqslant 3$) ferromagnets. The explicit correlation inequalities involving three- and four-times-truncated expectations are listed in Appendix A for the anisotropic case and in Appendix B for the isotropic case.

II. TWO-COMPONENT MODEL

In this section we consider the two-component scalar field or spin model on a lattice Λ that consists of a set of N sites in d-dimensional space. To each site we associate a two-dimensional vector spin $\Phi = (\varphi,\xi)$. The Hamiltonian is of the form

$$\mathscr{H}_\Gamma(\Phi) = -\sum_{i,j:i<j=1}^{N} (J_{ij}\varphi_i\varphi_j + K_{ij}\xi_i\xi_j) - \sum_{j=1}^{N} (h_j\varphi_j + H_j\xi_j), \tag{2.1}$$

where $\Gamma \equiv \{J_{ij},K_{ij},h_j,H_j\}$ and $J_{ij},K_{ij},h_j,H_j \geqslant 0$ for all i and j. Note that the ferromagnetic interaction strength J_{ij} and the external fields h_j,H_j are made to vary from bond to bond and from site to site, respectively. Let a finite family of real-valued random variables $\mathbf{\Phi} = \{\Phi_j \in \mathbb{R}^2; i = 1,...,N\}$ be distributed by the measure μ on $(\mathbb{R}^2)^N$ given by

$$d\mu_\Gamma(\mathbf{\Phi}) = Z_\Gamma^{-1} \exp[-\mathscr{H}_\Gamma(\mathbf{\Phi})] \prod_{j=1}^{N} d\nu(\varphi_j), \tag{2.2}$$

where Z_Γ is the partition function that guarantees the normalization $\int d\mu_\Gamma(\mathbf{\Phi}) = 1$ and $d\nu(\varphi_j)$ is the single spin measure,

$$d\nu(\Phi) = d\Phi \exp[-V(\Phi)], \quad d\Phi = d\varphi\, d\xi, \tag{2.3}$$

whose explicit form is specified below.

Before proceeding to find new correlation inequalities for two-component vector spin systems, we should explain the connection of our paper with that of Monroe.[14] Monroe considered the spin Φ with unit length $|\Phi| \equiv (\varphi^2+\xi^2)^{1/2} = 1$, whose distribution over the unit circle is given by $f(\Phi)$ assumed to be even, i.e., $f(-\Phi) = f(\Phi)$. Our model allows the spin to be unbounded and, if desired, the fixed-length case is recovered by adopting $V(\Phi) = \lambda(|\Phi|^2-1)^2$ and taking the limit $\lambda \to \infty$. So our method covers a more general class of two-component ferromagnets than those encompassed by Monroe.

Consider the duplicate system whose random variables $\mathbf{\Phi}$ and $\tilde{\mathbf{\Phi}} \equiv (\tilde{\varphi},\tilde{\xi})$ are independently, identically distributed according to μ.

Introducing the vector notation

$$\vec{\varphi} \equiv (\varphi,\tilde{\varphi}), \quad \vec{\xi} \equiv (\xi,\tilde{\xi}), \quad \vec{\Phi} \equiv (\vec{\varphi},\vec{\xi}), \quad \vec{H} \equiv (h,\tilde{h},H,\tilde{H}), \tag{2.4}$$

the sum of the original and duplicated Hamiltonian can be written as

$$\mathscr{H}_\Gamma(\mathbf{\Phi}) + \mathscr{H}_\Gamma(\tilde{\mathbf{\Phi}}) = -\sum_{i,j:i<j}^{N} (J_{ij}\vec{\varphi}_i\cdot\vec{\varphi}_j + K_{ij}\vec{\xi}\cdot\vec{\xi}) - \sum_{j=1}^{N} \vec{H}_j\cdot\vec{\Phi}_j. \tag{2.5}$$

For an orthogonal matrix T, define

$$\vec{X} \equiv T\vec{\Phi}, \tag{2.6}$$

where we have defined the variables

$$\vec{X} \equiv (\vec{x},\vec{y}), \quad \vec{x} \equiv (x,\tilde{x}), \vec{y} \equiv (y,\tilde{y}). \tag{2.7}$$

(I) If T is taken to be the direct sum of two orthogonal matrices A,B such that

$$T = \begin{vmatrix} A & O \\ O & B \end{vmatrix}, \tag{2.8}$$

then we obtain

$$\mathscr{H}_\Gamma(\mathbf{\Phi}) + \mathscr{H}_\Gamma(\tilde{\mathbf{\Phi}}) = -\sum_{i,j:i<j}^{N} (J_{ij}\vec{x}_i\cdot\vec{x}_j + K_{ij}\vec{y}_i\cdot\vec{y}_j) - \sum_{j=1}^{N} (T\vec{H}_j)\cdot\vec{X}_j, \tag{2.9}$$

where $\vec{x} = A\vec{\varphi}$ and $\vec{y} = B\vec{\xi}$.

(II) If the ferromagnetic interaction is of isotropic type: $J_{ij} = K_{ij}$, then, for any orthogonal matrix T,

$$\mathscr{H}_\Gamma(\mathbf{\Phi}) + \mathscr{H}_{\bar{\Gamma}}(\tilde{\mathbf{\Phi}}) = -\sum_{i,j:i<j=1}^{N} J_{ij}\vec{X}_i\cdot\vec{X}_j - \sum_{j=1}^{N} (T\vec{H})\cdot\vec{X}_j. \tag{2.10}$$

For the multi-index $\mathbf{P}(a) = \{P_j(a)\}_{j\in\Lambda}$ (a collection of non-negative integers), define

$$\{(\vec{X})^{(a)}\}^{\mathbf{P}(a)} = \prod_{j=1}^{N} \{(\vec{X}_j)^{(a)}\}^{P_j(a)}, \tag{2.11}$$

where a labels the components of $\vec{\mathbf{X}}$.

Definition 1: Let $\mathbf{\Phi}$ and $\tilde{\mathbf{\Phi}}$ be two independent copies of a random variable $\mathbf{\Phi}$ distributed by ν. Now we define the class $\mathfrak{g}$ of single spin measures by

$$\mathfrak{g} \equiv \left\{ \nu; E_0\left[\prod_{a=1}^{4} \{(\vec{X})^{(a)}\}^{\mathbf{P}(a)}\right] \geqslant 0, \text{ for all } \mathbf{P}(a) \geqslant 0,\ a = 1,...,4 \right\}, \tag{2.12}$$

where we have defined the unnormalized expectation

$$E_0[F(\Phi,\tilde{\Phi})] = \int F(\Phi,\tilde{\Phi}) \prod_{i=1}^{N} d\nu(\Phi_i) \prod_{j=1}^{N} d\nu(\tilde{\Phi}_j). \tag{2.13}$$

Now we look for an orthogonal matrix T that satisfies the conditions (2.12). First, we restrict the potential V to be of the Φ^4 type, namely,

$$V_4(\Phi) = \lambda(\varphi^2+\xi^2)^2 + \mu(\varphi^2+\xi^2), \quad \lambda \geqslant 0, \quad \mu \in \mathbb{R}. \tag{2.14}$$

In general, however, we suppose V is a polynominal of degree D, that is,

$$V_D(\Phi) = \sum_{n=1}^{D} \lambda_{2n}(|\Phi|^2)^n, \quad |\Phi|^2 \equiv \varphi^2+\xi^2. \tag{2.15}$$

Then we define

$$\mathfrak{V}(\Phi,\tilde{\Phi}) \equiv V_D(\Phi) + V_D(\tilde{\Phi}) = \sum_{n=1}^{D} \lambda_{2n}[(|\Phi|^2)^n + (|\tilde{\Phi}|^2)^n], \tag{2.16}$$

which is rewritten in terms of the variable X by substituting the relation $\vec{\Phi} \equiv T^t \vec{X}$ (T^t: transposed matrix of T): $\mathfrak{B}_D(\Phi,\tilde{\Phi}) = W_D(X^{(1)}, X^{(2)}, X^{(3)}, X^{(4)})$, and is decomposed into a sum of (even terms in each $X^{(a)}$) and (odd terms in each $X^{(a)}$). The even terms cause no problem. The odd terms must be ferromagnetic.

Define

$$P_{2n}(\Phi,\tilde{\Phi}) \equiv (|\Phi|^2)^n + (|\tilde{\Phi}|^2)^n. \tag{2.17}$$

Since the orthogonality of T implies

$$P_2(\Phi,\tilde{\Phi}) = (X^{(1)})^2 + (X^{(2)})^2 + (X^{(3)})^2 + (X^{(4)})^2, \tag{2.18}$$

it suffices to check the ferromagnetic character of $P_{2n}(\Phi,\tilde{\Phi})$ for $n \geqslant 2$.

Finally, note that

$$\begin{aligned} &d\nu(\Phi)d\nu(\tilde{\Phi}) \\ &\quad = d\varphi\, d\xi\, d\tilde{\varphi}\, d\tilde{\xi} \exp[-\{V(\varphi,\xi) + V(\tilde{\varphi},\tilde{\xi})\}] \\ &\quad = dx\, dy\, d\tilde{x}\, d\tilde{y} \exp[-W(\vec{x},\vec{y})] \det \left| \frac{\partial(\vec{\varphi},\vec{\xi})}{\partial(\vec{x},\vec{y})} \right| \\ &\quad = dx\, dy\, d\tilde{x}\, d\tilde{y} \exp[-W(\vec{x},\vec{y})] \det T^t. \end{aligned} \tag{2.19}$$

Thus we must check the following points: (I) the orthogonality of T, (II) $\det T^T = \det T > 0$, (III) ferromagnetic character of the odd terms in $W(\vec{x},\vec{y})$, and (IV) $(T\vec{H}_j)^{(a)} \geqslant 0$ for all $a = 1,...,4$ (in the presence of external fields).

The requirements (II) and (III) are sufficient to conclude that $\nu \in \mathfrak{g}$. Therefore, the requirements (I) and (IV) imply that, for all $\mathbf{P}(a) \geqslant 0$ ($a = 1,...,4$),

$$\int d\mu_\Gamma(\Phi) d\mu_\Gamma(\tilde{\Phi}) \prod_{a=1}^{4} \{(T\vec{\Phi})^{(a)}\}^{\mathbf{P}(a)} \geqslant 0, \tag{2.20}$$

provided that $\nu \in \mathfrak{g}$ (see, e.g., Secs. 4.3 and 4.7 of Ref. 2).

For a given set of multi-index $\mathbf{P}(a) \equiv \{P_i(a)\}_{i\in\Lambda}$ ($a = 1,...,4$), the corresponding correlation inequalities are obtained by explicitly expanding the product

$$\begin{aligned} &\prod_{a=1}^{4} \{(T\vec{\Phi})^{(a)}\}^{\mathbf{P}(a)} \\ &\quad = \prod_{i,j,k,l=1}^{N} \{(T\vec{\Phi}_i)^{(1)}\}^{P_i(1)} \{(T\vec{\Phi}_j)^{(2)}\}^{P_j(2)} \\ &\qquad \times \{(T\vec{\Phi}_k)^{(3)}\}^{P^k(3)} \{(T\vec{\Phi}_l)^{(4)}\}^{P_l(4)}, \end{aligned}$$

where

$$\begin{aligned} (T\vec{\Phi}_j)^{(a)} &\equiv \sum_{b=1}^{4} T_{ab} \vec{\Phi}_j{}^{(b)} \\ &= T_{a1}\varphi_j + T_{a2}\tilde{\varphi}_j + T_{a3}\xi_j + T_{a4}\tilde{\xi}_j, \end{aligned} \tag{2.21}$$

and rewriting each term of the result in terms of the normalized expectation

$$\langle(\cdot)\rangle \equiv \int d\mu_\Gamma(\Phi)(\cdot), \tag{2.22}$$

according to the rule that, for $F(\Phi) = \varphi^A \xi^B$ and $G(\tilde{\Phi}) \equiv \tilde{\varphi}^C \tilde{\xi}^D$,

$$\int d\mu_\Gamma(\Phi) d\mu_\Gamma(\tilde{\Phi}) F(\Phi) G(\tilde{\Phi}) = \langle F(\Phi)\rangle \langle G(\Phi)\rangle.$$

Here it should be remarked that, in this procedure, there appear only terms that are a product of at most two expectations, since we have prepared the twofold duplicated system; but see Sec. V. Subsequently, we rewrite them in terms of the truncated expectations; see Appendixes A and B.

Finally, note that the measure $d\mu_\Gamma(\Phi) d\mu_\Gamma(\tilde{\Phi})$ is invariant under the transformation of exchanging the original and duplicated variables

$$\varphi_i \to \tilde{\varphi}_i \text{ and } \xi_j \to \tilde{\xi}_j, \text{ for all } i,j \in \Lambda, \tag{2.23}$$

and, if the external field is zero, under the following independent four transformations:

$$\varphi_i \to -\varphi_i, \quad \xi_j \to -\xi_j, \quad \tilde{\varphi}_k \to -\tilde{\varphi}_k, \quad \tilde{\xi}_l \to -\tilde{\xi}_l. \tag{2.24}$$

III. THE ANISOTROPIC CASE

In this section, we look for an orthogonal matrix satisfying the above conditions (I)–(IV) such that T is of block diagonal form entailing two 2×2 orthogonal matrices. Any 2×2 orthogonal matrix can be written as

$$A = \begin{vmatrix} \alpha & \beta \\ -\beta & \alpha \end{vmatrix} \quad \text{or} \quad B = \begin{vmatrix} \alpha & \beta \\ \beta & -\alpha \end{vmatrix}, \tag{3.1}$$

where $\alpha^2 + \beta^2 = 1$. Here note that $\det A = 1$, but $\det B = -1$.

Then the 4×4 matrix T is obtained as follows.

Case (1):

$$T = \begin{vmatrix} A & 0 \\ 0 & \tilde{A} \end{vmatrix} = \frac{1}{\sqrt{2}} \begin{vmatrix} \alpha & \beta & 0 & 0 \\ -\beta & \alpha & 0 & 0 \\ 0 & 0 & \tilde{\alpha} & \tilde{\beta} \\ 0 & 0 & -\tilde{\beta} & \tilde{\alpha} \end{vmatrix}. \tag{3.2}$$

Indeed, the requirement (II) is satisfied, since

$$\det T^t = \tfrac{1}{4}(\alpha^2 + \beta^2)(\tilde{\alpha}^2 + \tilde{\beta}^2) > 0.$$

By explicit calculations, we obtain

$$\begin{aligned} P_4(\Phi,\tilde{\Phi}) = \tfrac{1}{2}\{&(X^{(1)})^4 + (X^{(2)})^4 + (X^{(3)})^4 + (X^{(4)})^4 \\ &+ 6(X^{(1)})^2(X^{(2)})^2 + 2(X^{(1)})^2(X^{(3)})^2 \\ &+ 2(X^{(1)})^2(X^{(4)})^2 + 2(X^{(2)})^2(X^{(3)})^2 \\ &+ 2(X^{(2)})^2(X^{(4)})^2 + 6(X^{(3)})^2(X^{(4)})^2\} \\ &+ 4\alpha\beta\tilde{\alpha}\tilde{\beta} X^{(1)} X^{(2)} X^{(3)} X^{(4)}. \end{aligned} \tag{3.3}$$

Then the requirement (III) forces us to take

$$\alpha\beta\tilde{\alpha}\tilde{\beta} = -1 (\leqslant 0), \tag{3.4}$$

which is sufficient to conclude that $\nu \in \mathfrak{g}$. Then we have

$$\begin{aligned} T\vec{H} = 1/\sqrt{2}\{&(\alpha+\beta)h, (\alpha-\beta)h, \\ &(\tilde{\alpha}+\tilde{\beta})H, (\tilde{\alpha}-\tilde{\beta})H\}. \end{aligned} \tag{3.5}$$

Let $\alpha,\beta,\tilde{\alpha},\tilde{\beta} = \pm 1$. Then one has

$$\begin{aligned} T\vec{H} = 1/\sqrt{2}\{&(\alpha+\beta)h, (\alpha-\beta)h, \\ &\tilde{\alpha}(1-\alpha\beta)H, \tilde{\alpha}(1+\alpha\beta)H\}. \end{aligned} \tag{3.6}$$

(i) The case of $\alpha\beta = 1$: $T\vec{H} = \sqrt{2}(\alpha h, 0, 0, \tilde{\alpha} H)$. Now the requirement $(T\vec{H})^{(a)} \geqslant 0$ for all $a = 1,...,4$ leads to the final result

$$\alpha = 1, \quad \tilde{\alpha} = 1, \quad \beta = 1, \quad \tilde{\beta} = -1. \tag{3.7}$$

(ii) The case of $\alpha\beta = -1$: in this case $T\vec{H} = \sqrt{2}(0, \alpha h, \tilde{\alpha} H, 0)$. Then we obtain

$$\alpha=1,\quad \tilde{\alpha}=1,\quad \beta=-1,\quad \tilde{\beta}=1. \tag{3.8}$$

Case (2):

$$T=\begin{vmatrix} B & 0 \\ 0 & \tilde{B} \end{vmatrix}=\frac{1}{\sqrt{2}}\begin{vmatrix} \alpha & \beta & 0 & 0 \\ \beta & -\alpha & 0 & 0 \\ 0 & 0 & \tilde{\alpha} & \tilde{\beta} \\ 0 & 0 & \tilde{\beta} & -\tilde{\alpha} \end{vmatrix}. \tag{3.9}$$

Then

$$\det T^t=(\alpha^2+\beta^2)(\tilde{\alpha}^2+\tilde{\beta}^2)/4>0.$$

Also in this case, the odd term has the same form as above. So we can proceed as before and obtain the following results.

(i) The case of $\alpha\beta=1$:

$$\vec{TH}=\sqrt{2}(\alpha h,0,0-\tilde{\alpha}H).$$

Then $\alpha=1,\tilde{\alpha}=-1,\beta=1,\tilde{\beta}=1$.

(ii) The case of $\alpha\beta=-1$: in this case $\vec{TH}=\sqrt{2}(0,-\alpha h,\tilde{\alpha}H,0)$. Then we obtain

$$\alpha=-1,\quad \tilde{\alpha}=1,\quad \beta=1,\quad \tilde{\beta}=1.$$

Case (3):

$$T=\begin{vmatrix} A & 0 \\ 0 & \tilde{B} \end{vmatrix}=\frac{1}{\sqrt{2}}\begin{vmatrix} \alpha & \beta & 0 & 0 \\ -\beta & \alpha & 0 & 0 \\ 0 & 0 & \tilde{\alpha} & \tilde{\beta} \\ 0 & 0 & \tilde{\beta} & -\tilde{\alpha} \end{vmatrix}, \tag{3.10}$$

which is orthogonal, but yields

$$\det T^t=(\alpha^2+\beta^2)(\tilde{\alpha}^2+\tilde{\beta}^2)/4<0.$$

Hence this choice of T contradicts requirement (II).

In this paper, explicit correlation inequalities are obtained for the matrix T case (1) (i):

$$T=\frac{1}{\sqrt{2}}\begin{vmatrix} 1 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 1 \end{vmatrix}. \tag{3.11}$$

Then one finds

$$\int d\mu_\Gamma(\Phi)d\mu_\Gamma(\tilde{\Phi})\prod_{a=1}^{4}\{(T\vec{\Phi})^{(a)}\}^{P(a)}$$
$$=\int d\mu_\Gamma(\Phi)d\mu_\Gamma(\tilde{\Phi})\prod_{i,j,k,l=1}^{N}(\varphi_i+\tilde{\varphi}_i)^{P_i(1)}$$
$$\times(-\varphi_j+\tilde{\varphi}_j)^{P_j(2)}(\xi_k-\tilde{\xi}_k)^{P_k(3)}(\xi_l+\tilde{\xi}_l)^{P_l(4)}. \tag{3.12}$$

Combining this with the invariance of the measure $d\mu_\Gamma(\Phi)d\mu_\Gamma(\tilde{\Phi})$ under the transformation (2.23), nontrivial correlation inequalities are obtained only if

$$\sum_{i=1}^{N}P_j(2)\text{ and }\sum_{k=1}^{N}P_k(3)\text{ are both even or both odd.} \tag{3.13}$$

By performing the transformations (2.24) simultaneously, it is easy to see that all the possible choices of T above produce the same set of correlation inequalities as those obtained by, e.g., (3.11). So all cases are exhausted by considering (3.11).

For the matrix T of (3.2), one has

$$|\Phi|^2=\vec{X}\cdot\vec{X}-2\alpha\beta X^{(1)}X^{(2)}-2\tilde{\alpha}\tilde{\beta}X^{(3)}X^{(4)},$$
$$|\Phi|^2=\vec{X}\cdot\vec{X}+2\alpha\beta X^{(1)}X^{(2)}+2\tilde{\alpha}\tilde{\beta}X^{(3)}X^{(4)}, \tag{3.14}$$

where we have defined the inner product

$$\vec{X}\cdot\vec{X}\equiv(X^{(1)})^2+(X^{(2)})^2+(X^{(3)})^2+(X^{(4)})^2. \tag{3.15}$$

It is not difficult to show that the condition (3.4) is sufficient to guarantee the ferromagnetic character (III) of the polynomial $P_{2n}(\Phi,\tilde{\Phi})$ for all $n\geqslant 2$. So the correlation inequalities presented below hold also for

$$V_D(\Phi)=\sum_{n=1}^{D}\lambda_{2n}(|\Phi|^2)^n,$$
$$\text{if } \lambda_{2n}\geqslant 0\quad (n\geqslant 2)\text{ and }\lambda_2\equiv\mu\in\mathbb{R}. \tag{3.16}$$

We define $N(a)\equiv\Sigma_{i\in\Lambda}P_i(a)$ and the index $N\equiv(N(1),N(2);N(3),N(4))$. Then, for example, we obtain

$$\langle\varphi_i;\xi_j\rangle\leqslant 0, \tag{3.17}$$

for (0,1;1,0), which is a special case of (1.7), and

$$\langle\varphi_i;\xi_j\rangle\geqslant -2\langle\varphi_i\rangle\langle\xi_j\rangle, \tag{3.18}$$

for (1,0;0,1). For more correlation inequalities, see Appendix A. The inequalities (25), (28), and (29) in Ref. 14 correspond to (1,2;0,0), (1,0;2,0), and (0,1;1,1), respectively.

Especially, in the case that $h\equiv 0$ (but $H\geqslant 0$), our inequalities have the following form:

(1,1;1,0) [the special case of (1.7)],

$$\langle\varphi_i\varphi_j;\xi_k\rangle\leqslant 0; \tag{3.19}$$

(0,2;0,1),(2,0;0,1),

$$\langle\varphi_i\varphi_j;\xi_k\rangle\geqslant -2\langle\varphi_i\varphi_j\rangle\langle\xi_k\rangle; \tag{3.20}$$

(0,0;2,1),

$$\langle\xi_i;\xi_j;\xi_k\rangle\geqslant -2\langle\xi_i;\xi_j\rangle\langle\xi_k\rangle; \tag{3.21}$$

(2,2;0,0),

$$\langle\varphi_i;\varphi_j;\varphi_k;\varphi_l\rangle\geqslant -2\langle\varphi_i\varphi_j\rangle\langle\varphi_k\varphi_l\rangle; \tag{3.22}$$

(1,1;1,1),

$$\langle\varphi_i\varphi_j;\xi_k;\xi_l\rangle\leqslant -2\langle\varphi_i\varphi_j;\xi_k\rangle\langle\xi_l\rangle; \tag{3.23}$$

(0,2;2,0),(2,0;2,0),

$$\langle\varphi_i\varphi_j;\xi_k;\xi_l\rangle\geqslant -2\langle\varphi_i\varphi_j\rangle\langle\xi_k;\xi_l\rangle; \tag{3.24}$$

(0,2;0,2),

$$\begin{aligned}\langle\varphi_i\varphi_j;\xi_k;\xi_l\rangle&\geqslant -2\langle\varphi_i\varphi_j;\xi_k\rangle\langle\xi_l\rangle;\\ &\quad -2\langle\varphi_i\varphi_j;\xi_l\rangle\langle\xi_k\rangle\\ &\quad -2\langle\varphi_i\varphi_j\rangle\langle\xi_k;\xi_l\rangle\\ &\quad -4\langle\varphi_i\varphi_j\rangle\langle\xi_k\rangle\langle\xi_l\rangle \qquad (3.25)\\ &\geqslant -2\langle\varphi_i\varphi_j\rangle\langle\xi_k;\xi_l\rangle\\ &\quad -4\langle\varphi_i\varphi_j\rangle\langle\xi_k\rangle\langle\xi_l\rangle; \qquad (3.26)\end{aligned}$$

where we have used (3.19),

(0,0;2,2),

$$\begin{aligned}\langle\xi_i;\xi_j;\xi_k;\xi_l\rangle&\geqslant -2\langle\xi_i;\xi_j;\xi_k\rangle\langle\xi_l\rangle\\ &\quad -2\langle\xi_i;\xi_j;\xi_l\rangle\langle\xi_k\rangle\\ &\quad -2\langle\xi_i;\xi_j\rangle\langle\xi_k;\xi_l\rangle\\ &\quad -4\langle\xi_i;\xi_j\rangle\langle\xi_k\rangle\langle\xi_l\rangle. \qquad (3.27)\end{aligned}$$

IV. THE ISOTROPIC CASE

Consider now an orthogonal matrix T given as the tensor product

$$T=A\otimes\tilde{A}=\frac{1}{2}\begin{vmatrix} \alpha\tilde{\alpha} & \alpha\tilde{\beta} & \beta\tilde{\alpha} & \beta\tilde{\beta} \\ -\alpha\tilde{\beta} & \alpha\tilde{\alpha} & -\beta\tilde{\beta} & \beta\tilde{\alpha} \\ -\beta\tilde{\alpha} & -\beta\tilde{\beta} & \alpha\tilde{\alpha} & \alpha\tilde{\beta} \\ \beta\tilde{\beta} & -\beta\tilde{\alpha} & -\alpha\tilde{\beta} & \alpha\tilde{\alpha} \end{vmatrix}, \quad (4.1)$$

where A and $\tilde{A}$ are the 2×2 orthogonal matrices (3.1). However, an explicit calculation shows that

$$\begin{aligned} P_4(\Phi,\tilde{\Phi}) =& \tfrac{1}{2}\{(X^{(1)})^4+(X^{(2)})^4+(X^{(3)})^4+(X^{(4)})^4 \\ &+6(X^{(1)})^2(X^{(2)})^2+2(X^{(1)})^2(X^{(3)})^2 \\ &+2(X^{(1)})^2(X^{(4)})^2+2(X^{(2)})^2(X^{(3)})^2 \\ &+2(X^{(2)})^2(X^{(4)})^2+6(X^{(3)})^2(X^{(4)})^2\} \\ &+4X^{(1)}X^{(2)}X^{(3)}X^{(4)}. \end{aligned} \quad (4.2)$$

Note that, in this expansion, the quantities $\alpha,\beta,\tilde{\alpha},\tilde{\beta}$ to be specified do not appear anywhere. Moreover, for both $T=A\otimes\tilde{B}$ and $T=B\otimes\tilde{B}$, $P_4(\Phi,\tilde{\Phi})$ has the same form as (4.2). Hence the requirement (III) cannot be satisfied for these 4×4 matrices T of the tensor-product type.

Next we try to look for an orthogonal matrix T of the form

$$T=\alpha\mathbf{1}+\beta S, \quad (4.3)$$

where $\mathbf{1}$ is the unit matrix. Note that

$$\begin{aligned} TT^t &= (\alpha\mathbf{1}+\beta S)(\alpha\mathbf{1}+\beta S^t) \\ &= \alpha^2\mathbf{1}+\alpha\beta(S+S^t)+\beta^2SS^t. \end{aligned}$$

Hence for $TT^t=1$ to hold, it is sufficient that S satisfy the conditions

$$S=-S^t, \quad SS^t=1, \quad \alpha^2+\beta^2=1. \quad (4.4)$$

For example, the 2×2 matrix A is recovered from

$$S=\begin{vmatrix} 0 & 1 \\ -1 & 0 \end{vmatrix}, \quad T=\begin{vmatrix} \alpha & \beta \\ -\beta & \alpha \end{vmatrix}, \quad \alpha^2+\beta^2=1.$$

For the 4×4 matrix, we set

$$T=\frac{1}{2}\begin{vmatrix} a & b & c & d \\ -b & a & e & f \\ -c & -e & a & g \\ -d & -f & -g & a \end{vmatrix}=\frac{1}{2}a\mathbf{1}+S. \quad (4.5)$$

The condition that all the off-diagonal elements of AA^t vanish turns out to be satisfied if we take

$$f=-cde \quad \text{and} \quad g=bde, \quad (4.6)$$

provided that $b,c,d,e,f,g=+1$ or -1.

In this reduction, we obtain

$$\begin{aligned} P_4(\Phi,\tilde{\Phi}) =& \tfrac{1}{2}\{(X^{(1)})^4+(X^{(2)})^4+(X^{(3)})^4+(X^{(4)})^4 \\ &+2(2-abce)(X^{(1)})^2(X^{(2)})^2 \\ &+2(X^{(1)})^2(X^{(3)})^2 \\ &+2(2+abce)(X^{(1)})^2(X^{(4)})^2 \\ &+2(2+abce)(X^{(2)})^2(X^{(3)})^2 \\ &+2(X^{(2)})^2(X^{(4)})^2 \\ &+2(2-abce)(X^{(3)})^2(X^{(4)})^2\} \\ &-4abcdX^{(1)}X^{(2)}X^{(3)}X^{(4)}. \end{aligned} \quad (4.7)$$

Then the requirement (ferromagnetic interaction)

$$abcd=1 \quad (4.8)$$

is sufficient to conclude that $\nu\in\mathrm{g}$. Finally we have

$$T=\frac{1}{2}\begin{vmatrix} a & b & c & abc \\ -b & a & e & -abe \\ -c & -e & a & ace \\ -abc & -abe & -ace & a \end{vmatrix}. \quad (4.9)$$

Note that $\det T^t=1$, and in addition,

$$\begin{aligned} \vec{TH} =& \tfrac{1}{2}[(a+b)h+c(1+ab)H, \\ &(a-b)h+e(1-ab)H, \\ &-(c+e)h+a(1+ce)H, \\ &ab(-c+e)h+a(1-ce)H]. \end{aligned} \quad (4.10)$$

Consider now the case $ab=1$, $ce=1$ for which

$$\begin{aligned} \vec{TH} &= (ah+cH,0,-ch+aH,0) \\ &= a(h+(c/a)H,0-(c/a)h+H,0). \end{aligned}$$

For $ac=1$, $\vec{TH}=a(h+H,0,-h+H,0)$, the requirement (IV) is satisfied if we take $a=1$, $b=1$, $c=1$, $e=1$, provided that $H\geq h\geq0$. For $ac=-1$, $\vec{TH}=a(h-H,0,h+H,0)$, then the requirement (IV) is satisfied if $a=1$, $b=1$, $c=-1$, $e=-1$, provided that $h\geq H\geq0$. Other cases are discussed in a similar manner, and the results are summarized as seen in Table I. Here the requirement (IV) is satisfied for cases from 1–4 in the region $H\geq h\geq0$, and for the cases from 5–8 in the region $h\geq H\geq0$.

We present correlation inequalities for the matrix T (under the condition $H\geq h\geq0$):

$$T=\frac{1}{2}\begin{vmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ -1 & -1 & 1 & 1 \end{vmatrix}. \quad (4.11)$$

By the same argument as that in the previous section, nontrivial correlation inequalities are obtained only if

$$\sum_{j=1}^{N}P_j(2) \text{ and } \sum_{k=1}^{N}P_k(3) \text{ are both even or both odd.} \quad (4.12)$$

As in the anisotropic case, one can check that all the possible

TABLE I. Allowed sets of matrix elements for the matrix (4.5) which satisfies all the requirements (I)–(IV); we assumed that $a=+1$.

	a	b	c	d	e	f	g	ab	ce	$abce$
1	+	+	+	+	+	−	+	+	+	+
2	+	+	+	+	−	+	−	+	−	−
3	+	−	+	−	+	+	+	−	+	−
4	+	−	−	+	+	+	−	−	−	+
5	+	+	−	−	−	+	+	+	+	+
6	+	+	−	−	+	−	−	+	−	−
7	+	−	−	+	−	−	+	−	+	−
8	+	−	+	−	−	−	−	−	−	+

choices of T above produce the same set of correlation inequalities as those obtained by, e.g., (4.11).

By explicit calculations, one may show that the condition (4.8) that guarantees the ferromagnetic character (III) of the polynominal $P_n(\Phi,\bar{\Phi})$, for $n=2,4$, is sufficient to satisfy the requirement (III) for $n=3$, but fails for $n=4$. So the correlation inequalities presented below hold also for the

$$V_6(\Phi)=\eta(|\Phi|^2)^3+\lambda(|\Phi|^2)^2+\mu|\Phi|^2$$

model ($\eta,\lambda\geqslant 0, \mu\in\mathbb{R}$), together with the $\lambda(|\Phi|^2)^2$ model.

For example, we have the following correlation inequalities corresponding to the multi-index $(N(1),\ldots,N(4))$:

(0,1,1,0),

$$\langle\varphi_i;\varphi_j\rangle-\langle\varphi_i;\xi_j\rangle+\langle\xi_i;\varphi_j\rangle-\langle\xi_i;\xi_j\rangle\geqslant 0; \tag{4.13}$$

(1,0,0,1),

$$\begin{aligned}&\langle\varphi_i;\varphi_j\rangle+2\langle\varphi_i\rangle\langle\varphi_j\rangle-\langle\varphi_i;\xi_j\rangle\\&\quad-2\langle\varphi_i\rangle\langle\xi_j\rangle-\langle\xi_i;\varphi_j\rangle-2\langle\xi_i\rangle\langle\varphi_j\rangle\\&\quad+\langle\xi_i;\xi_j\rangle+2\langle\xi_i\rangle\langle\xi_j\rangle\leqslant 0.\end{aligned} \tag{4.14}$$

Further explicit correlation inequalities are presented in Appendix B, for the case of $h\equiv 0$. For the isotropic case, in contrast with the anisotropic case, we can obtain upper bounds on $\langle\xi_i;\xi_j;\xi_k\rangle$. For example, corresponding to the multi-index (0,1,1,1), we have

$$\begin{aligned}\langle\xi_i;\xi_j;\xi_k\rangle\leqslant&-2\langle\xi_i;\xi_j\rangle\langle\xi_k\rangle\\&+\langle\varphi_i\varphi_j;\xi_k\rangle+\langle\varphi_i\varphi_k;\xi_j\rangle\\&-\langle\xi_i;\varphi_j\varphi_k\rangle+2\langle\varphi_i\varphi_j\rangle\langle\xi_k\rangle.\end{aligned} \tag{4.15}$$

Note here that in the right-hand side of this inequality the first three terms are nonpositive and the remaining two are non-negative.

V. ON THE EXTENSIONS TO N-COMPONENT MODELS ($N\geqslant 3$)

In this section we discuss the extension of our method to get more correlation inequalities for N-component ferromagnets ($N\geqslant 3$). Let us consider R-fold replicated systems ($R\geqslant 2$). The case $N=2, R=2$ has been already considered in the previous sections. Let us put $F=N\times R$. In the following we try to find an $F\times F$ ($F>4$) orthogonal matrix T that satisfies the requirement (III). If we could find such a matrix T, all the remaining steps are carried out easily, as exemplified in the preceding sections.

First, we consider the two-component model, but we increase R to obtain correlation inequalities for higher- (at most R)-times-truncated expectations. This might enable us to obtain, e.g., the GHS inequality for two-component ferromagnets: $\langle\xi_i;\xi_j;\xi_k\rangle\leqslant 0$.[21]

(i) The case $N=2, R=3$: For example, consider the orthogonal 6×6 matrix

$$T(6)=\begin{vmatrix}\vartheta^{(1)} & \mathbf{0}\\ \mathbf{0} & \vartheta^{(2)}\end{vmatrix},$$

where $\vartheta^{(a)}$ is a 3×3 orthogonal matrix, e.g.,

$$\vartheta=\begin{vmatrix}1 & 0 & 0\\ 0 & \alpha & \beta\\ 0 & -\beta & \alpha\end{vmatrix},\quad \alpha^2+\beta^2=1.$$

But, by explicit calculations, it turns out that all the 6×6 matrices T of this type do not satisfy the requirement (III).

(ii) The case $N=2, R=4$: A candidate is the following 8×8 matrix:

$$T(8)=\begin{vmatrix}A^{(1)} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & A^{(2)} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & A^{(3)} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & A^{(4)}\end{vmatrix},$$

where each $A^{(a)}$ is a 2×2 orthogonal matrix of the form (3.1). This matrix is obviously orthogonal, but does not satisfy the requirement (III). Furthermore, the orthogonal matrix

$$\mathrm{T}(8)=\begin{vmatrix}T^{(1)} & \mathbf{0}\\ \mathbf{0} & T^{(2)}\end{vmatrix}$$

made from the 4×4 matrix T in (4.9) cannot satisfy the requirement (III), either.

Next we consider the N-component case, $N\geqslant 3$ with $R=2$. A possible (orthogonal) matrix T is obtained as the direct sum

$$T=\begin{vmatrix}A^{(1)} & & & \\ & A^{(2)} & & \\ & & \ddots & \\ & & & A^{(N)}\end{vmatrix},$$

where each $A^{(a)}$ is a 2×2 orthogonal matrix as in (3.1). But a matrix of this type fails to satisfy the requirement (III) for $N=3$ and 4.

The above consideration forces us to take T not of the diagonal form. As a first step, we tried to look for an $F\times F$ ($F\geqslant 5$) orthogonal matrix T with elements $T_{ab}=+1$ or -1. In the range $5\leqslant F\leqslant 7$, we exhausted all the possible cases and obtained the result that there exist no orthogonal matrices T which take values only $+1$ or -1 as their matrix elements. For $F=8$, however there exist such orthogonal matrices. We expect that some of them satisfy all the requirements (I)–(IV). Work to check them is now in progress. We hope that the results can be reported in a subsequent paper. The applications of the new correlation inequalities obtained in this paper to the study of the critical behavior of multicomponent ferromagnets will be presented elsewhere.

APPENDIX A: NEW CORRELATION INEQUALITIES (ANISOTROPIC CASE)

In what follows we enumerate correlation inequalities together with the index $N\equiv(N(1),N(2),N(3),N(4))$:

(0,1;1,0),

$$\langle\varphi_i;\xi_j\rangle\leqslant 0; \tag{A1}$$

(1,0;0,1),

$$\langle\varphi_i;\xi_j\rangle\geqslant -2\langle\varphi_i\rangle\langle\xi_j\rangle; \tag{A2}$$

(1,2;0,0),

$$\langle\varphi_i;\varphi_j;\varphi_k\rangle\geqslant -2\langle\varphi_i\rangle\langle\varphi_j;\varphi_k\rangle; \tag{A3}$$

(1,1;1,0),

$$\langle\varphi_i;\varphi_j;\xi_k\rangle\leqslant -2\langle\varphi_i\rangle\langle\varphi_j;\xi_k\rangle; \tag{A4}$$

(0,2;0,1),

$$\langle\varphi_i;\varphi_j;\xi_k\rangle \geqslant -2\langle\varphi_i;\varphi_j\rangle\langle\xi_k\rangle; \tag{A5}$$

(0,1;1,1)

$$\langle\varphi_i;\xi_j;\xi_k\rangle \leqslant -2\langle\varphi_i;\xi_j\rangle\langle\xi_k\rangle; \tag{A6}$$

(1,0;2,0),

$$\langle\varphi_i;\xi_j;\xi_k\rangle \geqslant -2\langle\varphi_i\rangle\langle\xi_j;\xi_k\rangle; \tag{A7}$$

(0,0;2,1),

$$\langle\xi_i;\xi_j;\xi_k\rangle \geqslant -2\langle\xi_i;\xi_j\rangle\langle\xi_k\rangle; \tag{A8}$$

(2,0;0,1),

$$\begin{aligned}\langle\varphi_i;\varphi_j;\xi_k\rangle \geqslant & -2\langle\varphi_i\rangle\langle\varphi_j;\xi_k\rangle \\ & -2\langle\varphi_i\rangle\langle\varphi_i;\xi_k\rangle \\ & -2\langle\varphi_i;\varphi_j\rangle\langle\xi_k\rangle \\ & -4\langle\varphi_i\rangle\langle\varphi_j\rangle\langle\xi_k\rangle;\end{aligned} \tag{A9}$$

(1,0;0,2),

$$\begin{aligned}\langle\varphi_i;\xi_j;\xi_k\rangle \geqslant & -2\langle\varphi_i\rangle\langle\xi_j;\xi_k\rangle \\ & -2\langle\xi_j\rangle\langle\varphi_i;\xi_k\rangle \\ & -2\langle\varphi_i;\xi_j\rangle\langle\xi_k\rangle \\ & -4\langle\varphi_i\rangle\langle\xi_j\rangle\langle\xi_k\rangle;\end{aligned} \tag{A10}$$

(2,2;0,0),

$$\begin{aligned}\langle\varphi_i;\varphi_j;\varphi_k;\varphi_l\rangle \geqslant & -2\langle\varphi_i\rangle\langle\varphi_j;\varphi_k;\varphi_l\rangle \\ & -2\langle\varphi_j\rangle\langle\varphi_i;\varphi_k;\varphi_l\rangle \\ & -2\langle\varphi_i;\varphi_j\rangle\langle\varphi_k;\varphi_l\rangle \\ & -4\langle\varphi_i\rangle\langle\varphi_j\rangle\langle\varphi_k;\varphi_l\rangle;\end{aligned} \tag{A11}$$

(2,1;1,0),

$$\begin{aligned}\langle\varphi_i;\varphi_j;\varphi_k;\xi_l\rangle \leqslant & -2\langle\varphi_i\rangle\langle\varphi_j;\varphi_k;\xi_l\rangle \\ & -2\langle\varphi_j\rangle\langle\varphi_i;\varphi_k;\xi_l\rangle \\ & -2\langle\varphi_i;\varphi_j\rangle\langle\varphi_k;\xi_l\rangle \\ & -4\langle\varphi_i\rangle\langle\varphi_j\rangle\langle\varphi_k;\xi_l\rangle;\end{aligned} \tag{A12}$$

(0,3;1,0),

$$\begin{aligned}\langle\varphi_i;\varphi_j;\varphi_k\xi_l\rangle \leqslant & -2\langle\varphi_i;\varphi_j\rangle\langle\varphi_k;\xi_l\rangle \\ & -2\langle\varphi_i;\varphi_k\rangle\langle\varphi_j;\xi_l\rangle \\ & -2\langle\varphi_j;\varphi_k\rangle\langle\varphi_i;\xi_l\rangle;\end{aligned} \tag{A13}$$

(1,1;1,1),

$$\begin{aligned}\langle\varphi_i;\varphi_j;\xi_k;\xi_l\rangle \leqslant & -2\langle\varphi_i;\varphi_j;\xi_k\rangle\langle\xi_l\rangle \\ & -2\langle\varphi_i\rangle\langle\varphi_j;\xi_k;\xi_l\rangle \\ & -2\langle\varphi_j;\xi_k\rangle\langle\varphi_i;\xi_l\rangle \\ & -4\langle\varphi_j;\xi_k\rangle\langle\varphi_i\rangle\langle\xi_l\rangle;\end{aligned} \tag{A14}$$

(0,2;2,0),

$$\begin{aligned}\langle\varphi_i;\varphi_j;\xi_k\xi_l\rangle \geqslant & -2\langle\varphi_i;\varphi_j\rangle\langle\xi_k;\xi_l\rangle \\ & -2\langle\varphi_i;\xi_k\rangle\langle\varphi_j;\xi_l\rangle \\ & -2\langle\varphi_j;\xi_l\rangle\langle\varphi_j;\xi_k\rangle;\end{aligned} \tag{A15}$$

(0,2;0,2),

$$\begin{aligned}\langle\varphi_i;\varphi_j;\xi_k;\xi_l\rangle \geqslant & -2\langle\varphi_i;\varphi_j;\xi_k\rangle\langle\xi_l\rangle \\ & -2\langle\varphi_i;\varphi_j;\xi_l\rangle\langle\xi_k\rangle \\ & -2\langle\varphi_i;\varphi_j\rangle\langle\xi_k;\xi_l\rangle \\ & -4\langle\varphi_i;\varphi_j\rangle\langle\xi_k\rangle\langle\xi_l\rangle;\end{aligned} \tag{A16}$$

(2,0;2,0),

$$\begin{aligned}\langle\varphi_i;\varphi_j;\xi_k\xi_l\rangle \geqslant & -2\langle\varphi_i\rangle\langle\varphi_j;\xi_k;\xi_l\rangle \\ & -2\langle\varphi_j\rangle\langle\varphi_i;\xi_k;\xi_l\rangle \\ & -2\langle\varphi_i;\varphi_j\rangle\langle\xi_k;\xi_l\rangle \\ & -4\langle\varphi_i\rangle\langle\varphi_j\rangle\langle\xi_k;\xi_l\rangle;\end{aligned} \tag{A17}$$

(0,1;3,0),

$$\begin{aligned}\langle\varphi_i;\xi_j;\xi_k;\xi_l\rangle \leqslant & -2\langle\varphi_i;\xi_j\rangle\langle\xi_k;\xi_l\rangle \\ & -2\langle\varphi_i;\xi_k\rangle\langle\xi_j;\xi_l\rangle \\ & -2\langle\varphi_i;\xi_l\rangle\langle\xi_j;\xi_k\rangle;\end{aligned} \tag{A18}$$

(0,1;1,2),

$$\begin{aligned}\langle\varphi_i;\xi_j;\xi_k;\xi_l\rangle \leqslant & -2\langle\varphi_i;\xi_j;\xi_k\rangle\langle\xi_l\rangle \\ & -2\langle\varphi_i;\xi_j;\xi_l\rangle\langle\xi_k\rangle \\ & -2\langle\varphi_i;\xi_j\rangle\langle\xi_k;\xi_l\rangle \\ & -4\langle\varphi_i;\xi_j\rangle\langle\xi_k\rangle\langle\xi_l\rangle;\end{aligned} \tag{A19}$$

(0,0;2,2),

$$\begin{aligned}\langle\xi_i;\xi_j;\xi_k;\xi_l\rangle \geqslant & -2\langle\xi_i;\xi_j;\xi_k\rangle\langle\xi_l\rangle \\ & -2\langle\xi_i;\xi_j;\xi_l\rangle\langle\xi_k\rangle \\ & -2\langle\xi_i;\xi_j\rangle\langle\xi_k;\xi_l\rangle \\ & -4\langle\xi_i;\xi_j\rangle\langle\xi_k\rangle\langle\xi_l\rangle.\end{aligned} \tag{A20}$$

APPENDIX B: NEW CORRELATION INEQUALITIES (ISOTROPIC CASE)

The correlation inequalities are obtained for the matrix T (4.11) that allows us to include the external field (h,H), provided that $H \geqslant h \geqslant 0$. However, the full form is rather complicated. So, in the following, we present them in the case of $h \equiv 0$, $H \geqslant 0$:

(0,1,1,0),

$$\langle\varphi_i\varphi_j\rangle \geqslant \langle\xi_i;\xi_j\rangle; \tag{B1}$$

(1,0,0,1),

$$\langle\varphi_i\varphi_j\rangle \leqslant \langle\xi_i;\xi_j\rangle + 2\langle\xi_i\rangle\langle\xi_j\rangle; \tag{B2}$$

(0,1,1,1),

$$\begin{aligned}\langle\xi_i;\xi_j;\xi_k\rangle \leqslant & -2\langle\xi_i;\xi_j\rangle\langle\xi_k\rangle \\ & -\langle\xi_i;\varphi_j;\varphi_k\rangle \\ & +\langle\varphi_i\varphi_j;\xi_k\rangle \\ & +\langle\varphi_i\varphi_k;\xi_j\rangle \\ & +2\langle\varphi_i\varphi_j\rangle\langle\xi_k\rangle;\end{aligned} \tag{B3}$$

(1,1,1,0),

$$\begin{aligned}\langle\xi_i;\xi_j;\xi_k\rangle \leqslant & -2\langle\xi_i\rangle\langle\xi_j;\xi_k\rangle \\ & +\langle\xi_i;\varphi_j\varphi_k\rangle \\ & -\langle\varphi_i\varphi_j;\xi_k\rangle \\ & +\langle\varphi_i\varphi_k;\xi_j\rangle \\ & +2\langle\xi_i\rangle\langle\varphi_j\varphi_k\rangle; \end{aligned} \tag{B4}$$

(1,2,0,0),

$$\begin{aligned}\langle\xi_i;\xi_j;\xi_k\rangle \geqslant & -2\langle\xi_i\rangle\langle\xi_j;\xi_k\rangle \\ & -\langle\xi_i;\varphi_j\varphi_k\rangle \\ & -\langle\varphi_i\varphi_j;\xi_k\rangle \\ & -\langle\varphi_i\varphi_k;\xi_j\rangle \\ & -2\langle\xi_i\rangle\langle\varphi_j\varphi_k\rangle; \end{aligned} \tag{B5}$$

(1,0,2,0),

$$\begin{aligned}\langle\xi_i;\xi_j;\xi_k\rangle \geqslant & -2\langle\xi_i\rangle\langle\xi_j;\xi_k\rangle \\ & -\langle\xi_i;\varphi_j\varphi_k\rangle \\ & +\langle\varphi_i\varphi_j;\xi_k\rangle \\ & +\langle\varphi_i\varphi_k;\xi_j\rangle \\ & -2\langle\xi_i\rangle\langle\varphi_j\varphi_k\rangle; \end{aligned} \tag{B6}$$

(0,2,0,1),

$$\begin{aligned}\langle\xi_i;\xi_j;\xi_k\rangle \geqslant & -2\langle\xi_i;\xi_j\rangle\langle\xi_k\rangle \\ & +\langle\xi_i;\varphi_j\varphi_k\rangle \\ & -\langle\varphi_i\varphi_j;\xi_k\rangle \\ & +\langle\varphi_i\varphi_k;\xi_j\rangle \\ & -2\langle\varphi_i\varphi_j\rangle\langle\xi_k\rangle; \end{aligned} \tag{B7}$$

(0,0,2,1),

$$\begin{aligned}\langle\xi_i;\xi_j;\xi_k\rangle \geqslant & -2\langle\xi_i;\xi_j\rangle\langle\xi_k\rangle \\ & -\langle\xi_i;\varphi_j\varphi_k\rangle \\ & -\langle\varphi_i\varphi_j;\xi_k\rangle \\ & -\langle\varphi_i\varphi_k;\xi_j\rangle \\ & -2\langle\varphi_i\varphi_j\rangle\langle\xi_k\rangle; \end{aligned} \tag{B8}$$

(2,0,0,1),

$$\begin{aligned}\langle\xi_i;\xi_j;\xi_k\rangle \geqslant & -2\langle\xi_i;\xi_j\rangle\langle\xi_k\rangle \\ & -2\langle\xi_i;\xi_k\rangle\langle\xi_j\rangle \\ & -2\langle\xi_j;\xi_k\rangle\langle\xi_i\rangle \\ & +\langle\xi_i;\varphi_j\varphi_k\rangle \\ & +\langle\xi_j;\varphi_i\varphi_k\rangle \\ & -\langle\varphi_i\varphi_j;\xi_k\rangle \\ & -2\langle\varphi_i\varphi_j\rangle\langle\xi_k\rangle \\ & +2\langle\varphi_i\varphi_k\rangle\langle\xi_j\rangle \\ & +2\langle\varphi_j\varphi_k\rangle\langle\xi_i\rangle \\ & -4\langle\xi_i\rangle\langle\xi_j\rangle\langle\xi_k\rangle; \end{aligned} \tag{B9}$$

(1,0,0,2),

$$\begin{aligned}\langle\xi_i;\xi_j;\xi_k\rangle \geqslant & -2\langle\xi_i;\xi_j\rangle\langle\xi_k\rangle \\ & -2\langle\xi_i;\xi_k\rangle\langle\xi_j\rangle \\ & -2\langle\xi_j;\xi_k\rangle\langle\xi_i\rangle \\ & -\langle\xi_i;\varphi_j\varphi_k\rangle \\ & +\langle\xi_j;\varphi_i\varphi_k\rangle \\ & +\langle\varphi_i\varphi_j;\xi_k\rangle \\ & +2\langle\varphi_i\varphi_j\rangle\langle\xi_k\rangle \\ & +2\langle\varphi_i\varphi_k\rangle\langle\xi_j\rangle \\ & -2\langle\varphi_j\varphi_k\rangle\langle\xi_i\rangle \\ & -4\langle\xi_i\rangle\langle\xi_j\rangle\langle\xi_k\rangle; \end{aligned} \tag{B10}$$

(0,0,0,3),

$$\begin{aligned}\langle\xi_i;\xi_j;\xi_k\rangle \geqslant & -2\langle\xi_i;\xi_j\rangle\langle\xi_k\rangle \\ & -2\langle\xi_i;\xi_k\rangle\langle\xi_j\rangle \\ & -2\langle\xi_j;\xi_k\rangle\langle\xi_i\rangle \\ & -\langle\xi_i;\varphi_j\varphi_k\rangle \\ & -\langle\xi_j;\varphi_i\varphi_k\rangle \\ & -\langle\varphi_i\varphi_j;\xi_k\rangle \\ & -2\langle\varphi_i\varphi_j\rangle\langle\xi_k\rangle \\ & -2\langle\varphi_i\varphi_k\rangle\langle\xi_j\rangle \\ & -2\langle\varphi_j\varphi_k\rangle\langle\xi_i\rangle \\ & -4\langle\xi_i\rangle\langle\xi_j\rangle\langle\xi_k\rangle; \end{aligned} \tag{B11}$$

(1,1,1,1),

$$\begin{aligned}&\langle\varphi_i;\varphi_j;\varphi_k;\varphi_l\rangle + \langle\varphi_i;\varphi_j;\xi_k;\xi_l\rangle \\ &\quad -\langle\varphi_i;\xi_j;\varphi_k;\xi_l\rangle - \langle\varphi_i;\xi_j;\xi_k;\varphi_l\rangle \\ &\quad -\langle\xi_i;\varphi_j;\varphi_k;\xi_l\rangle - \langle\xi_i;\varphi_j;\xi_k;\varphi_l\rangle \\ &\quad +\langle\xi_i;\xi_j;\varphi_k;\varphi_l\rangle + \langle\xi_i;\xi_j;\xi_k;\xi_l\rangle \\ &\quad -2\langle\xi_i\rangle\langle\varphi_j;\xi_k;\varphi_l\rangle + 2\langle\xi_i\rangle\langle\xi_j;\varphi_k;\varphi_l\rangle \\ &\quad -2\langle\xi_i\rangle\langle\varphi_j;\varphi_k;\xi_l\rangle + 2\langle\xi_i\rangle\langle\xi_j;\xi_k;\xi_l\rangle \\ &\quad +2\langle\varphi_i;\varphi_i;\xi_k\rangle\langle\xi_l\rangle - 2\langle\varphi_i;\xi_j;\varphi_k\rangle\langle\xi_l\rangle \\ &\quad -2\langle\xi_i;\varphi_j;\varphi_k\rangle\langle\xi_l\rangle + 2\langle\xi_i;\xi_j;\xi_k\rangle\langle\xi_l\rangle \\ &\quad +2\langle\varphi_i;\varphi_l\rangle\langle\varphi_j;\varphi_k\rangle - 2\langle\varphi_i;\varphi_j\rangle\langle\xi_k;\xi_l\rangle \\ &\quad -2\langle\xi_i\rangle\langle\xi_l\rangle\langle\varphi_i;\varphi_k\rangle + 2\langle\xi_i;\xi_l\rangle\langle\xi_j;\xi_k\rangle \\ &\quad -4\langle\xi_i\rangle\langle\varphi_j;\varphi_k\rangle\langle\xi_l\rangle + 4\langle\xi_i\rangle\langle\xi_j;\xi_k\rangle\langle\xi_l\rangle \leqslant 0; \end{aligned} \tag{B12}$$

$(0,2,2,0)$,

$$\begin{aligned}
&\langle\varphi_i;\varphi_j;\varphi_k;\varphi_l\rangle + \langle\varphi_i;\varphi_j;\xi_k;\xi_l\rangle \\
&\quad - \langle\varphi_i;\xi_j;\varphi_k;\xi_l\rangle - \langle\varphi_i;\xi_j;\xi_k;\varphi_l\rangle \\
&\quad - \langle\xi_i;\varphi_j;\varphi_k;\xi_l\rangle - \langle\xi_i;\varphi_j;\xi_k;\varphi_l\rangle \\
&\quad + \langle\xi_i;\xi_j;\varphi_k;\varphi_l\rangle + \langle\xi_i;\xi_j;\xi_k;\xi_l\rangle \\
&\quad + 2\langle\varphi_i;\varphi_j\rangle\langle\varphi_k;\varphi_l\rangle + 2\langle\varphi_i;\varphi_l\rangle\langle\varphi_j;\varphi_k\rangle \\
&\quad + 2\langle\varphi_i;\varphi_k\rangle\langle\varphi_j;\varphi_l\rangle \\
&\quad + 2\langle\varphi_i;\varphi_j\rangle\langle\xi_k;\xi_l\rangle - 2\langle\varphi_i;\varphi_k\rangle\langle\xi_j;\xi_l\rangle \\
&\quad - 2\langle\varphi_i;\varphi_l\rangle\langle\xi_j;\xi_k\rangle - 2\langle\varphi_j;\varphi_k\rangle\langle\xi_i;\xi_l\rangle \\
&\quad - 2\langle\varphi_j;\varphi_l\rangle\langle\xi_i;\xi_k\rangle + 2\langle\varphi_k;\varphi_l\rangle\langle\xi_i;\xi_j\rangle \\
&\quad + 2\langle\xi_i;\xi_j\rangle\langle\xi_k;\xi_l\rangle + 2\langle\xi_i;\xi_k\rangle\langle\xi_j;\xi_l\rangle \\
&\quad + 2\langle\xi_i;\xi_l\rangle\langle\xi_j;\xi_k\rangle \geqslant 0. \qquad \text{(B13)}
\end{aligned}$$

Other cases are omitted.

[1]B. Simon, *The $P(\phi)_2$ Euclidean (Quantum) Field Theory* (Princeton U. P., Princeton, NJ, 1974).
[2]J. Glimm and A. Jaffe, *Quantum Physics: A Functional Integral Point of View* (Springer, New York, 1981).
[3]K.-I. Kondo, Prog. Theor. Phys. **73**, 504 (1985).
[4]A. D. Sokal, J. Stat. Phys. **251**, 25 (1981).
[5]G. S. Sylvester, J. Stat. Phys. **15**, 327 (1976).
[6]J. L. Lebowitz, Commun. Math. Phys. **35**, 87 (1974).
[7]K.-I. Kondo, T. Otofuji, and Y. Sugiyama, J. Stat. Phys. **40**, 563 (1985); K.-I. Kondo, J. Stat. Phys. **46**, 35 (1987).
[8]K.-I. Kondo, unpublished manuscript (1986) and papers to be published in Proceedings of the Physical Society of Japan (in Japanese) and Prog. Theor. Phys.; J. Fröhlich, "Quantum field theory in terms of random walks and random surfaces," in *Progress in Gauge Field Theory*, Cargese Lectures, 1983, edited by G. 't Hooft, A. Jaffe, H. Lehman, P. K. Mitter, I. M. Singer, and R. Stora (Plenum, New York, 1984).
[9]R. S. Ellis and J. L. Monroe, Commun. Math. Phys. **41**, 33 (1975).
[10]R. S. Ellis, J. L. Monroe, and C. M. Newman, Commun. Math. Phys. **46**, 167 (1976).
[11]R. S. Ellis and C. M. Newman, Trans. Am. Math. Soc. **237**, 83 (1978).
[12]R. S. Ellis, C. W. Newman, and M. R. O'Connell, J. Stat. Phys. **26**, 37 (1981).
[13]J. Ginibre, Commun. Math. Phys. **16**, 310 (1970).
[14]J. L. Monroe, J. Math. Phys. **16**, 1809 (1975).
[15]F. Dunlop and C. M. Newman, Commun. Math. Phys. **44**, 223 (1975).
[16]H. Kunz, Ch.-Ed. Pfister, and P.-A. Vuillermot, Phys. Lett. A **54**, 428 (1975).
[17]F. Dunlop, Commun. Math. Phys. **49**, 247 (1976).
[18]J. Bricmont, J. Stat. Phys. **17**, 289 (1977).
[19]J. Bricmont, J. R. Fontaine, and L. J. Landau, Commun. Math. Phys. **56**, 281 (1977).
[20]A. Messager, S. Miracle-Sole, and C. Pfister, Commun. Math. Phys. **58**, 19 (1978).
[21]J. L. Monroe and P. A. Pearce, J. Stat. Phys. **21**, 615 (1979).
[22]G. S. Sylvester, Commun. Math. Phys. **73**, 105 (1980).
[23]B. Simon, J. Stat. Phys. **22**, 491 (1980).
[24]N. Kalus, J. Stat. Phys. **25**, 635 (1981).
[25]D. C. Brydges, J. Fröhlich, and A. D. Sokal, Commun. Math. Phys. **91**, 117 (1983).
[26]F. Dunlop, J. Stat. Phys. **41**, 733 (1985).
[27]E. S. Abers and B. W. Lee, Phys. Rep. **9**, 1 (1973).

On a parametrization method for the semiclassical vacuum problem. I

Mario Castagnino
Instituto de Astronomía y Física del Espacio, Ciudad Universitaria, 1428 Buenos Aires, Argentina and Instituto de Física de Rosario, Av. Pellegrini 250, 2000 Rosario, Argentina

Carlos Laciana
Instituto de Astronomía y Física del Espacio, Ciudad Universitaria, 1428 Buenos Aires, Argentina

(Received 28 October 1986; accepted for publication 26 August 1987)

A formalism to study the semiclassical vacuum problem is introduced. Using this formalism we try to satisfy two competitive natural hypotheses: H1, the vacuum definition must assure the absorption of the infinities, coming from the vacuum expectation value of the energy-momentum tensor of the matter fields, in the bare constants of the gravitational classic action; H2, the vacuum must be the ground state. It is proved that the system of equations yielded by these two hypotheses is, in general, incompatible. With this formalism, the vacuum problem in curved space-time is stated. It is hoped that this formalism could eventually be used to solve the problem.

I. INTRODUCTION

As with quantum electrodynamics with time variable background fields, quantum field theory in curved space-time presents a nontrivial problem for its vacuum definition.[1] This paper is devoted to the introduction of a general formalism to study this problem. Even if we do not find a new solution, we believe that the method introduced sheds light on the main difficulties and could eventually either yield a solution, or rigorously prove that no solution exists.

Let us first review the problem from an historical point of view.[2]

A. Vacuum structure of flat space-time

There is experimental evidence that the vacuum structure in bounded space has several components of a different nature.

The first experimental sign that the vacuum structure was not at all trivial was observed in the last century. If we extract all the gas from a receptacle, we do not obtain an empty space devoid of all physical phenomena, as radiation pressure is still present. This pressure is a function of the temperature and exists in the best vacuum. It can only be eliminated if the absolute temperature vanishes. But even at zero degrees the vacuum is not trivial. In the late 1940's Casimir[3] suggested a set of experiments to see if there is an interaction among two uncharged metallic plates in the vacuum. The experiment showed that a residual attractive force remains, even at zero temperature. This force, known as the Casimir or polarization force, is of a different nature than the normal gas pressure or radiation pressure. It is independent of the temperature, proportional to the plate area and inversely proportional to the fourth power of the distance between the plates. Thus we can only eliminate this residual force moving the plates apart to an infinite distance. Therefore we finally obtain a trivial vacuum when we transform our receptacle in an unbounded space. In fact this place must also be flat to eliminate gravity, too. (From the equivalence principle gravity is equal to acceleration, thus to eliminate gravity means also to work with inertial observers only.) Therefore we obtain a structureless vacuum only if we remove all the matter, the temperature, the curvature, and make the topology trivial, i.e., eliminate the boundaries as well. We shall see how these physical facts play a central role in the quantum vacuum problem.

B. The semiclassical vacuum problem

If we would like to define a partial vacuum, i.e., removing some elements of the above list and leaving the others, we shall have an impure vacuum notion that logically has a nontrivial structure. That is the case of the semiclassical vacuum where all the matter and temperature are eliminated but we leave gravity—as a curved space-time or a noninertial reference system—or a nontrivial topology (as boundaries, point identification, etc.).

Thus it is not surprising that this partial unconventional vacuum would be ill-defined and nontrivial.

In this paper we only study the problem of the vacuum definition in curved unbounded space-time using inertial or geodesic observers. We believe that if the problem could be solved, in this case, the other cases would also be solved by using the same method.

As we know we need a ground state of vacuum to build the Fock space of a quantum field theory. Thus we must define a vacuum in our semiclassical case, and it is necessary to give some criterion to choose this quantum state. An obviously necessary hypothesis, which we shall call the "fundamental hypothesis," is that the vacuum must yield a predictive theory, i.e., finite renormalizable physical quantities. This hypothesis is normally focused on the most important object in a curved space-time theory, the energy-momentum tensor. But there are several methods of renormalization; thus, in order to make the hypothesis precise, we must single out one method or a class of methods and so we must precisely state the following: the vacuum must yield a finite energy-momentum tensor when it is used with the chosen method or the chosen class. We shall choose our renormalization class among those renormalizations satisfying the Wald axioms (this choice being more general than in previous papers

 0022-2488/88/020460-16$02.50

where only the canonical renormalization was taken into account (Refs. 4 and 5). The Wald axioms are the following.

(1) The matrix elements of $T_{\mu\nu}$ between any pairs of orthogonal states coincide with the formal (unrenormalized) expression, which is finite.

(2) In flat Minkowski space-time the renormalized operator T is the normal ordered one, i.e.,

$$T_{\mu\nu}{}^{(\mathrm{ren})} = :T_{\mu\nu}:_{\mathrm{Mink}} .$$

This axiom implies that

$$\lim_{R_{\mu\nu\sigma\delta}\to 0} \langle T_{\mu\nu}\rangle_{(\mathrm{ren})} = \langle 0|:T_{\mu\nu}:_{\mathrm{Mink}}|0\rangle .$$

(3) The expectation value of $T_{\mu\nu}$ is conserved, i.e.,

$$\nabla_\mu \langle T^{\mu\nu}\rangle_{(\mathrm{ren})} = 0 .$$

(4) The causality axiom: For a quantum state defined in the far past (future), $\langle T_{\mu\nu}\rangle_{(\mathrm{ren})}$, at a point p, only depends on the quantum state and the geometry of space-time inside the past (future) null conoide of p. [Originally there was a fifth Wald axiom (cf. Ref. 57) that we do not use (refer to Ref. 6 or Ref. 1 to see why we eliminate this axiom).]

From the Wald axioms we can deduce the different components of $\langle T_{\mu\nu}\rangle_{(\mathrm{ren})}$ (which would reproduce the feature of the vacuum structure of Sec. I A). Precisely, giving two energy momentum tensor operators (renormalized using any method that satisfies the Wald axioms) $T_{\mu\nu}{}^{\mathrm{W}}$ and $T_{\mu\nu}{}^{\mathrm{W}'}$, if $|a\rangle$ and $|b\rangle$ are two orthogonal quantum states then, from axiom 1,

$$\langle a|T_{\mu\nu}{}^{\mathrm{W}}|b\rangle = \langle a|T_{\mu\nu}{}^{\mathrm{W}'}|b\rangle .$$

Calling $\mathscr{O}_{\mu\nu} = T_{\mu\nu}{}^{\mathrm{W}} - T_{\mu\nu}{}^{\mathrm{W}'}$ we have that $\langle a|\mathscr{O}_{\mu\nu}|b\rangle = 0$ for every pair such that $\langle a|b\rangle = 0$. Then calling $|c\pm\rangle = 2^{1/2}(|a\rangle \pm |b\rangle)$ we have

$$\langle c+|\mathscr{O}_{\mu\nu}|c-\rangle = 0 ,$$

thus

$$\langle a|\mathscr{O}_{\mu\nu}|a\rangle = \langle b|\mathscr{O}_{\mu\nu}|b\rangle , \quad \forall |a\rangle, |b\rangle .$$

Therefore $\mathscr{O}_{\mu\nu} = It_{\mu\nu}$, where I is the identity operator and $t_{\mu\nu}$ is a c number. We have reached this conclusion using only Axiom 1. Now it is well known that, from Axiom 4 (causality), it turns out that $T_{\mu\nu}$ is a local term, i.e., a point function (cf. Ref. 4). In fact if we consider the vacuum $|0,\mathrm{in}\rangle$ (i.e., the vacuum $|0,\tau\rangle$ when $\tau\to-\infty$, then $\langle 0,\mathrm{in}|\mathscr{O}_{\mu\nu}|0,\mathrm{in}\rangle = t_{\mu\nu}$ must only depend on the geometry of the interior of the past null conoide of p and if we consider $|0,\mathrm{out}\rangle$ (i.e., $\tau\to+\infty$), $\langle 0,\mathrm{out}|\mathscr{O}_{\mu\nu}|0,\mathrm{out}\rangle = t_{\mu\nu}$ must only depend on the geometry of the interior of the future null conoide of p. Thus $t_{\mu\nu}$ is only a function of the geometry of the interception of the past and future conoides, i.e., the point t, therefore $t_{\mu\nu} = t_{\mu\nu}(p)$.

Let us now suppose that we make a foliation of space-time by spacelike hypersurfaces. Each surface is defined by an arbitrary time parameter τ, and let us suppose that we define somehow a vacuum state $|0,\tau\rangle$, i.e., for each τ. Thus creation and annihilation operators and a normal ordering can be defined on each hypersurface that we shall call $:\ :_\tau$. It can be proved (cf. Appendix A) that a normal orderer $:T_{\mu\nu}:_\tau$ is an operator that satisfies Wald's Axiom 1, thus, every operator $T_{\mu\nu}{}^{\mathrm{W}}$ (obtained by a Wald renormalization) can be written as

$$T_{\mu\nu}{}^{\mathrm{W}} = :T_{\mu\nu}:_\tau + It_{\mu\nu} , \tag{1.1}$$

where $t_{\mu\nu}$ is a c number (but in general it is not a point function). Therefore the vacuum expectation value (VEV) for the vacuum defined on another surface τ will be

$$\langle 0,\tau|T_{\mu\nu}{}^{\mathrm{W}}|0,\tau\rangle = \langle 0,\tau|:T_{\mu\nu}:_\tau|0,\tau\rangle + t_{\mu\nu} . \tag{1.2}$$

In addition, it is easy to prove (cf. Appendix A) that

$$\begin{aligned}&\langle 0,\tau|:T_{\mu\nu}(\tau',\mathbf{x}):_\tau|0,\tau\rangle \\ &\quad = \langle 0,\tau|T_{\mu\nu}(\tau',\mathbf{x})|0,\tau\rangle - \langle 0,\tau'|T_{\mu\nu}(\tau',\mathbf{x})|0,\tau'\rangle .\end{aligned} \tag{1.3}$$

Then if $\tau = \tau'$ we have

$$\langle 0,\tau'|:T_{\mu\nu}(\mathbf{x},\tau'):_\tau|0,\tau'\rangle = 0 ,$$

thus

$$\langle 0,\tau'|\overset{\mathrm{W}}{T}_{\mu\nu}(\tau',\mathbf{x})|0,\tau'\rangle = t_{\mu\nu} .$$

This equation gives the physical interpretation of $t_{\mu\nu}$: it is the VEV of the renormalized operator $\overset{\mathrm{W}}{T}_{\mu\nu}(\tau',\mathbf{x})$ using the vacuum at τ'. Here t_{00} can thus be identified as the energy density, of the vacuum at τ', i.e., the vacuum polarization density.

Thus for every vacuum $|0,\tau\rangle$ and for every renormalization method W that satisfies the Wald axioms we can perform the canonical decomposition (1,2) where the component $t_{\mu\nu}(\tau',\mathbf{x})$ is a c number. Moreover, even if we use as the quantum state of the vacuum state at τ': $|0,\tau'\rangle$, i.e., we consider a state with neither matter nor temperature at τ', we have a nonvanishing component of the energy-momentum tensor $t_{\mu\nu}$ that must be identified with the Casimir or polarization term that remains when there is neither matter nor temperature. This term is originated by a nontrivial geometry or topology and it vanishes in unbounded flat space-time. The definition of a vacuum at τ' is essential to single out this component. The other term in decomposition (1.2), $\langle 0,\tau|:T_{\mu\nu}:_\tau|0,\tau\rangle$, must contain a component due to the presence of matter (particles created by the evolution of the universe) and a component of radiation (created by the same cause). We shall try to single out these two components below. In fact, if we compute $\langle 0,\tau|:T_{\mu\nu}:_\tau|0,\tau\rangle$ we shall find an analytical component that can be expanded in powers of $\omega_{\mathbf{k}}$, the energy in mode $\mathbf{k}$ and a nonanalytical component that vanishes faster than any power of the energy. We shall identify the first component with the matter term and the second one with the radiation term because in a thermal spectrum the densities vanish faster than any power of $\omega_{\mathbf{k}}$. Thus the three elements that we must eliminate in order to obtain a trivial vacuum matter, radiation, and nontrivial geometry or topology, are related with the different canonical components that can be found in the expectation value of the renormalized energy-momentum tensor operator using any renormalization method that satisfies the Wald axioms.

Thus we shall implement our fundamental hypothesis asking that our semiclassical vacuum yield a finite renorma-

lizable VEV of the energy momentum tensor using a Wald renormalization and we shall study the convergence of this VEV studying the convergence of its canonical components separately.

But there is yet another different hypothesis that is seldom used to define the vacuum (see Refs. 7 and 8): at each time the vacuum is the quantum state that minimizes the renormalized VEV of the energy, i.e., the vacuum is the ground state. This is a very old idea that we shall call the "ground state hypothesis." The fundamental hypothesis is completely compelling, because the theory must be predictive. On the contrary the ground state hypothesis, in fact, is only a natural and convenient condition. The aim of this paper is to study how and when we can implement one of the two conditions, using a general Wald renormalization.

C. Organization of the paper

In Secs. II and III we shall analyze the restrictions imposed on the vacuum by the fundamental hypothesis. In Sec. II we shall compute the polarization terms for spin-0 and spin-1 fields and in Sec. III the matter-radiation term.

In Sec. IV we shall study the ground state hypothesis and how this hypothesis allows to introduce the elementary particle notion in our problem. We shall use the following conventions $c=1$, $h=1$, and space-time will have signature $+\ -\ -\ -$.

II. THE FUNDAMENTAL HYPOTHESIS

A. The problem in a globally hyperbolic space-time

The essential feature of every exact science, such as physics, is that its methods must foresee the outcome of the measurement of parameters known as observables. Thus these methods must necessarily yield finite quantities. Therefore, the fundamental hypothesis of our theory, as the one in every physical theory, is that the values predicted by the theory ought to be finite. In quantum field theory in curved space the most important object is the VEV of the energy-momentum tensor operator. Thus we choose this operator as the first one where our fundamental hypothesis must be verified. In this context we shall state our first hypothesis.

H1: The total VEV of the energy-momentum tensor operator must be devoid of divergent terms.

A consequence of H1 is the following. The $\langle T_{\mu\nu}\rangle$ of the matter field can only have divergent terms that can be absorbed by the classical gravitational action via the redefinition of the coupling constants.

It is now convenient to mention a technical point. Frequently a different criterion is used instead of H1: to ask that the expectation value of the symmetric two-point function of the quantum field should have Hadamard form. This is, in fact, a stronger hypothesis than H1. On one hand, it is well known that all the Hadamard vacua yield finite renormalized energy-momentum tensors. On the other hand, in Ref. 9 (p. 188) it is demonstrated that the expectation value of the energy-momentum tensor is a non-Hadamard vacuum could be finite (this is the case of a minimum-energy vacuum, for a massive spin-0 field in a Robertson–Walker universe with coupling $\zeta=\frac{1}{6}$).

The gravitational action that we shall use is

$$S_g=\int\sqrt{|g|}\,dx^4(\alpha_1+\alpha_2R^2+\alpha_3R^2+\alpha_4R^{\alpha\beta}R_{\alpha\beta}+\alpha_5R^{\alpha\beta\gamma\delta}R_{\alpha\beta\gamma\delta})\,. \tag{2.1}$$

The "dressed" constants $\alpha_1,\ldots,\alpha_5$ after renormalization must be finite and must be fixed by experiments (they must satisfy some restriction if we want to reobtain the flat space limit, cf. Ref. 10).

The variation of action (2.1) with respect to $g_{\mu\nu}$ yields the terms

$$\frac{2}{\sqrt{|g|}}\frac{\delta S_g}{\delta g_{\mu\nu}}=\lambda_1g_{\mu\nu}+\lambda_2G_{\mu\nu}+\lambda_3{}^{(1)}H_{\mu\nu}+\lambda_4{}^{(2)}H_{\mu\nu}\,, \tag{2.2}$$

where

$$\lambda_1=-\alpha_1\,,\quad \lambda_2=2\alpha_2\,,$$
$$\lambda_3=2(\alpha_3-\alpha_5)\,,\quad \lambda_4=2(\alpha_1+\alpha_5)\,,$$

and

$$G_{\mu\nu}=R_{\mu\nu}-\tfrac{1}{2}g_{\mu\nu}R\,,$$
$${}^{(1)}H_{\mu\nu}=2R_{;\mu\nu}-2g_{\mu\nu}\Box R-\tfrac{1}{2}g_{\mu\nu}R^2+2RR_{\mu\nu}\,,$$
$${}^{(2)}H_{\mu\nu}=R_{;\mu\nu}-\tfrac{1}{2}g_{\mu\nu}\Box R-\Box R_{\mu\nu}-\tfrac{1}{2}g_{\mu\nu}R^{\beta}R_{\alpha\beta}+2R^{\alpha\beta}R_{\alpha\beta\mu\nu}\,.$$

Every term of (2.2) turns out to be divergenceless.

As a model for matter we shall use a neutral scalar field φ (we shall also use a vector field φ^μ below) and we shall define some matter action $s_m\ [\varphi]$. The total action will be s_g+s_m and the field equation for $g_{\mu\nu}$ is

$$\frac{2}{\sqrt{|g|}}\frac{\delta S_g}{\delta g_{\mu\nu}}=-\frac{2}{\sqrt{|g|}}\frac{\delta S_m}{\delta g_{\mu\nu}}\,.$$

As the lhs is divergenceless so must be the rhs. We shall quantize S_m and leave S_g unquantized because we are dealing with a semiclassical theory. Thus the rhs will not be $T_{\mu\nu}$ but the unrenormalized values of $\langle T_{\mu\nu}\rangle$. We must consider the variation of the total action with respect to φ to compute this quantity, i.e.,

$$\delta S_m/\delta\varphi=0\,, \tag{2.3}$$

the field equation of field φ, and find an orthonormal basis of solution $\{\varphi_{\mathbf{k}}\}\cup\{\varphi^*_{\mathbf{k}}\}$ of this equation, in the inner product:

$$\langle u,v\rangle=i\int_\Sigma(u^*\nabla_\mu v-v\nabla_\mu u^*)d\sigma^\mu\,.$$

We shall work in a globally hyperbolic manifold so Σ is a Cauchy surface and the inner product is invariant if we perform the integration on other Cauchy surfaces. Here the $\mathbf{k}$ are the labels of the basis functions. If we foliate our manifold by Cauchy surfaces, labeled by a time parameter—as in Sec. I, we can choose a different basis at each surface that we shall call $\{\phi^\tau_{\mathbf{k}}\}\cup\{\phi^{\tau*}_{\mathbf{k}}\}$. We can expand the field operator in these bases as

$$\varphi(x)=\int dk^3\,[a^\tau_{\mathbf{k}}\phi^\tau_{\mathbf{k}}(x)+a^{\tau\dagger}_{\mathbf{k}}\phi^{\tau*}_{\mathbf{k}}(x)]\,, \tag{2.4}$$

where $a_{\mathbf{k}}^{\tau}$ and $a_{\mathbf{k}}^{\tau\dagger}$ are annihilation and creation operators, that allows us to define a vacuum at τ, $|0,\tau\rangle$ such that

$$a_{\mathbf{k}}^{\tau}|0,\tau\rangle = 0, \quad \forall \mathbf{k} \tag{2.5}$$

and build a Fock space at τ as

$$|1_{\mathbf{k}},\tau\rangle = a_{\mathbf{k}}^{\tau\dagger}|0,\tau\rangle \text{ , etc.}$$

A Bogoliubov transformation allows us to go from one basis at time τ, $\{\phi_{\mathbf{k}}^{\tau}\}\cup\{\overset{*}{\phi}_{\mathbf{k}}^{\tau}\}$, to another basis at time τ', $\{\phi_{\mathbf{k}}^{\tau'}\}\cup\{\overset{*}{\phi}_{\mathbf{k}}^{\tau'}\}$. As we said in Sec. I the VEV of the energy-momentum tensor can be decomposed as (cf. Appendix A)

$$\langle 0,\tau|T_{\nu}(\tau',\mathbf{x})|0,\tau\rangle = \langle 0,\tau|:T_{\mu\nu}(\tau',\mathbf{x}):_{\tau'}|0,\tau\rangle + \langle 0,\tau'|T_{\mu\nu}(\tau',\mathbf{x})|0,\tau'\rangle \ . \tag{2.6}$$

The first term corresponds to the created matter and radiation. It must be finite and we shall study this requirement in the next section. We shall call the second term the polarization term

$$P_{\mu\nu}(\tau,\mathbf{x}) = \langle 0,\tau|T_{\mu\nu}(\tau,\mathbf{x})|0,\tau\rangle \ .$$

It will contain finite and infinite terms and thus it can be decomposed as

$$P_{\mu\nu} = P_{\mu\nu_{\text{finite}}} + P_{\mu\nu_{\text{div}}} \ .$$

As a consequence of H1 we must have that

$$P_{\mu\nu_{\text{div}}} = \mu_1 g_{\mu\nu} + \mu_2 G_{\mu\nu} + \mu_3 {}^{(1)}H_{\mu\nu} + \mu_4 {}^{(2)}H_{\mu\nu} \ , \tag{2.7}$$

where the μ will be infinite constants that must be absorbed in the bare constants of action (2.1). The terms of $P_{\mu\nu_{\text{finite}}}$ will be finite and have a different form, i.e., they are not proportional to $g_{\mu\nu}$, $G_{\mu\nu}$, ${}^{(1)}H_{\mu\nu}$, nor ${}^{(2)}H_{\mu\nu}$. Thus the decomposition is unique and

$$P_{\mu\nu_{\text{finite}}} = \langle 0,\tau'|T_{\mu\nu}(\tau',\mathbf{x})|0,\tau'\rangle - \mu_1 g_{\mu\nu} - \mu_2 G_{\mu\nu} - \mu_3 {}^{(1)}H_{\mu\nu} - \mu_4 {}^{(2)}H_{\mu\nu} \ . \tag{2.8}$$

Therefore a sufficient condition to satisfy H1 is to choose the vacua $|0,\tau\rangle$ and $|0,\tau'\rangle$ such that (i) $P_{\mu\nu_{\text{div}}}$ would satisfy Eq. (2.8); (ii) $\langle 0,\tau|:T_{\mu\nu}(\tau'):_{\tau'}|0,\tau\rangle \neq \infty$.

Conditions (i) and (ii) are the starting points of our analysis about the way to choose the vacuum. We shall study these conditions in the Robertson–Walker universe using as examples the massive field of spin 0 and spin 1. We shall also see how and when we can impose the ground state and other similar hypotheses.

B. Robertson–Walker universe

In this section we shall study condition (i) in a spatially flat RW universe with metric:

$$ds^2 = dt^2 - a(t)^2(dx^2 + dy^2 + dz^2) \ . \tag{2.9}$$

We shall develop a formalism to solve the problem in the cases of spin 0 and spin 1.

1. Scalar field

We shall study the massive scalar field with an arbitrary coupling. The action is

$$S = \int \sqrt{|g|}\, dx^4(\partial_\mu\varphi\, \partial^\mu\varphi + \xi R\varphi^2 + m^2\varphi^2) \ , \tag{2.10}$$

where $|g|$ is the absolute value of the determinant of the metric tensor. From this action we can obtain the classical energy momentum tensor

$$T_{\mu\nu} = (1/\sqrt{|g|})(\delta S/\delta g^{\mu\nu}) \ . \tag{2.11}$$

The corresponding operators as a function of φ is

$$T_{\mu\nu} = \tfrac{1}{2}(1-2\xi)\{\nabla_\mu\varphi,\nabla_\nu\varphi\} + \tfrac{1}{2}(2\xi-\tfrac{1}{2})g_{\mu\nu}\{\nabla_\lambda\varphi,\nabla^\lambda\varphi\} - \xi\{\nabla_\mu\nabla_\nu\varphi,\varphi\} + [(\xi/2)R_{\mu\nu} + g_{\mu\nu}(\xi R + m^2)(\xi - \tfrac{1}{4})]\{\varphi,\varphi\} \ , \tag{2.12}$$

where $\{\ ,\ \}$ is the anticommutator.

Now we can quantize the field using Eq. (2.4) and we find a basis. We shall obtain such a basis by variable separation writing $\phi_{\mathbf{k}}^{\tau}(x)$ as

$$\phi_{\mathbf{k}}^{\tau}(x) = \frac{1}{(2\pi a)^{3/2}(2\Omega_{\mathbf{k}}^{\tau})^{1/2}} \times \exp i\left(-\int_{t_0}^{t}\Omega_{\mathbf{k}}^{\tau}\, dt' + \mathbf{k}\cdot\mathbf{x}\right) \tag{2.13}$$

(in this way we automatically assure that the basis should be orthonormal), where $\Omega_{\mathbf{k}}^{\tau} = \Omega_{\mathbf{k}}^{\tau}(t)$ is a function of the coordinate time only. Substituting Eq. (2.13) in Eq. (2.12) we can obtain

$$\langle 0,\tau|T_{00}(\tau,\mathbf{x})|0,\tau\rangle = \frac{1}{4}\int\frac{dk^3}{(2\pi a)^3}\frac{1}{\Omega_{\mathbf{k}}^{\tau}(\tau)} \times\left[\Omega_{\mathbf{k}}^{\tau 2}(x) + \omega_{\mathbf{k}}^2 + \left(\frac{9}{4} - 12\xi\right)H^2 - \left(6\xi - \frac{3}{2}\right)H\frac{\dot{\Omega}_{\mathbf{k}}^{\tau}(\tau)}{\Omega_{\mathbf{k}}^{\tau}(\tau)} + \frac{1}{4}\left(\frac{\dot{\Omega}_{\mathbf{k}}^{\tau}(\tau)}{\Omega_{\mathbf{k}}^{\tau}(\tau)}\right)^2\right], \tag{2.14}$$

where $H = \dot{a}/a$ is the Hubble coefficients, $\omega^2 = (m^2 + k^2/a^2)$ is the particle "energy," and $\Omega_{\mathbf{k}}^{\tau}$ is a solution of the equation

$$\frac{1}{2}\frac{\ddot{\Omega}_{\mathbf{k}}^{\tau}}{\Omega_{\mathbf{k}}^{\tau}} - \frac{3}{2}\left(\frac{\dot{\Omega}_{\mathbf{k}}^{\tau}}{\Omega_{\mathbf{k}}^{\tau}}\right)^2 + \Omega_{\mathbf{k}}^{\tau 2} = \omega_{\mathbf{k}}^2 + \left(\xi - \frac{1}{4}\right)R + \frac{3}{4}H^2 \ , \tag{2.15}$$

where R is the scalar of curvature that we can compute as

$$R = 6(2H^2 + \dot{H}) \ .$$

Each different function $\Omega_{\mathbf{k}}^{\tau}$ yields a different basis at time τ and each different basis corresponds to different vacuum $|0,\tau\rangle$. Thus $\Omega_{\mathbf{k}}$, or better the Cauchy data of $\Omega_{\mathbf{k}}$ at τ, defines the vacuum.

Let us briefly review what happens if we have another time τ': $\{\phi_{\mathbf{k}}^{\tau'}\}\cup\{\overset{*}{\phi}_{\mathbf{k}}^{\tau'}\}$. The bases will be related by

$$\phi_{\mathbf{k}}^{\tau'} = \sum_{\mathbf{k}'}(\alpha_{\mathbf{k}\mathbf{k}'}^{\tau'\tau}\,\phi_{\mathbf{k}'}^{\tau} + \beta_{\mathbf{k}\mathbf{k}'}^{\tau'\tau}\,\overset{*}{\phi}_{\mathbf{k}'}^{\tau}) \tag{2.16a}$$

and the annihilation and creation operators by

$$a_{\mathbf{k}}^{\tau'} = \sum_{\mathbf{k}'}(\overset{*}{\alpha}_{\mathbf{k}\mathbf{k}'}^{\tau'\tau}\,a_{\mathbf{k}'}^{\tau} - \overset{*}{\beta}_{\mathbf{k}\mathbf{k}'}^{\tau'\tau}\,a_{\mathbf{k}'}^{\tau\dagger}) \ ,$$
$$a_{\mathbf{k}}^{\tau'\dagger} = \sum_{\mathbf{k}'}(\alpha_{\mathbf{k}\mathbf{k}'}^{\tau'\tau}\,a_{\mathbf{k}'}^{\tau\dagger} - \beta_{\mathbf{k}\mathbf{k}'}^{\tau'\tau}\,a_{\mathbf{k}'}^{\tau}) \ . \tag{2.16b}$$

Due to the orthogonality properties of the bases we must have

$$\sum_{\mathbf{k}''}(\alpha^{\tau'\tau}_{\mathbf{k}\mathbf{k}''}\,\alpha^{*\tau'\tau}_{\mathbf{k}'\mathbf{k}''}-\beta^{\tau'\tau}_{\mathbf{k}\mathbf{k}''}\,\beta^{*\tau'\tau}_{\mathbf{k}'\mathbf{k}''})=\delta_{\mathbf{k}\mathbf{k}'}\,,$$

$$\sum_{\mathbf{k}''}(\alpha^{\tau'\tau}_{\mathbf{k}\mathbf{k}''}\,\beta^{*\tau'\tau}_{\mathbf{k}'\mathbf{k}''}-\beta^{\tau'\tau}_{\mathbf{k}'\mathbf{k}''}\,\alpha^{*\tau'\tau}_{\mathbf{k}'\mathbf{k}''})=0\,.$$

In the new basis we can define a new vacuum $|0,\tau'\rangle$ such that

$$a^{\tau'}_{\mathbf{k}}|0,\tau'\rangle=0,\quad \forall\mathbf{k}\,,$$

and a new Fock space. Both bases depend on their Cauchy data at τ: $\Omega^{\tau}_{\mathbf{k}}$, $\dot{\Omega}^{\tau}_{\mathbf{k}}$ or at τ': $\Omega^{\tau'}_{\mathbf{k}}$, $\dot{\Omega}^{\tau'}_{\mathbf{k}}$. This data must be a function of the local geometry and therefore of local geometric objects such as H^2 or R. We can also use the particle energy, in mode k, $w=(m^2+k^2/a^2)^{1/2}$. Thus if we study the problem only up to the fourth order in the metric derivatives (and this fact also introduces the objects H^4, H^2R, R^2, $H\dot{R}$, and $\ddot{R}$ because all other objects with fourth derivative can be expressed as linear combination of these five) we can write the Cauchy data at τ as

$$\Omega^{\tau}(\tau)=\omega\Big[\gamma_0^2+\gamma_1\frac{H^2}{\omega^2}+\gamma_2\frac{R}{\omega^2}+\gamma_3\frac{H^4}{\omega^4}+\gamma_4\frac{H^2R}{\omega^4}+\gamma_5\frac{R^2}{\omega^4}+\gamma_6\frac{H\dot{R}}{\omega^4}+\gamma_7\frac{\ddot{R}}{\omega^4}+\alpha_1\frac{m^2}{\omega^4}H^2+\alpha_2\frac{m^2R}{\omega^4}+\mathcal{O}\Big(\frac{1}{\omega^6}\Big)\Big]^{1/2},\tag{2.17a}$$

$$\frac{\dot{\Omega}^{\tau}(\tau)}{\Omega^{\tau}(\tau)}=\zeta_1H+\zeta_2\frac{HR}{\omega^2}+\zeta_3\frac{H^3}{\omega^2}+\zeta_4\frac{\ddot{R}}{\omega^2}+\alpha_3\frac{m^2H}{\omega^2}+\mathcal{O}\Big(\frac{1}{\omega^4}\Big)\tag{2.17b}$$

(from now on we shall suppress the label k).

These are the most general expressions with the correct dimensionality and the correct Minkowski limit, the γ, α, and ζ are dimensionless coefficients. Equations (2.17) are the basic equations of our formalism. In fact, changing the coefficients we can obtain all kinds of vacua, and we shall see how the hypotheses of Sec. I determine some coefficients and fix some relation between the others. Furthermore Eqs. (2.17) contain all the terms that yield divergencies in the energy-momentum tensor. We shall see how the hypotheses fix these coefficients in a unique way. Really, in a complete treatment, all coefficients up to orders higher than the fourth, and also the nonanalytical component of (2.17) must be defined. We cannot reach this goal in this paper, we merely give a formalism to study the problem. It might eventually be used to obtain the final solution in the future.

We have taken into account the following facts to write Eq. (2.17).

(a) The Cauchy data (2.17) must contain only even terms in the derivatives to yield divergencies that can be absorbed in the bare constant of Eq. (2.1). Thus these are the only relevant terms.

(b) Furthermore, all terms that yield a divergency that cannot be absorbed in the bare constants must be multiplied by a null coefficient in (2.17) (which is also the case of the odd powers of ω^{-1}).

(c) Terms proportional to ω^{-4} in (2.17a) and to ω^{-2} in (2.17b) are the ones of smaller power in ω that yield ultraviolet divergencies. All terms up to $\mathcal{O}(\omega^{-6})$ in (2.17a) and $\mathcal{O}(\omega^{-4})$ in (2.17b) will produce finite contributions.

(d) The mass m eliminates the presence of infrared divergencies.

(e) For a general treatment terms $\mathcal{O}(\omega^{-6})$ and $\mathcal{O}(\omega^{-4})$ must be explicitly written. We only study the ones that appear in Eq. (2.17) because at this level we are interested in the absorption of infinites only.

Let us now choose the Cauchy data (2.17) among the ones that satisfy condition (1). Thus we must compute $\langle 0,\tau'|T_{\mu\nu}(\tau',\mathbf{x})|0,\tau'\rangle$ as a function of the coefficients of Eq. (2.17) and also the terms in Eq. (2.8) that may be independent of these coefficients γ, α, and ζ must satisfy a system of equations obtained via this computation.

For the RW metric we have

$${}^{(2)}H_{\mu\nu}=\tfrac{1}{3}{}^{(1)}H_{\mu\nu}\,,$$

thus we can define a new constant in Eq. (2.8) as

$$\sigma=\lambda_3+\tfrac{1}{3}\lambda_4\,.$$

Then the 00 components of $P_{\mu\nu}$ are

$$P_{00}=\langle 0,\tau'|T_{00}(\tau',\mathbf{x})|0,\tau'\rangle-\lambda_1+3H^2\lambda_2+\sigma(6H^2R-\tfrac{1}{2}R^2+6H\dot{R})\,,\tag{2.18}$$

and we can make the decomposition

$$\langle 0,\tau'|T_{00}(\tau',\mathbf{x})|0,\tau'\rangle=\langle 0,\tau'|T_{00}(\tau',\mathbf{x})|0,\tau'\rangle_{\text{finite}}+\langle 0,\tau'|T_{00}(\tau',\mathbf{x})|0,\tau'\rangle_{\text{div}}\,.$$

The ultraviolet divergent terms are

$$\langle 0,\tau'|T_{00}(\tau',\mathbf{x})|0,\tau'\rangle_{\text{div}}=\gamma_0\Big(1+\frac{1}{\gamma_0^2}\Big)I_1+\frac{H^2}{\gamma_0}\Big\{\Big[\frac{\gamma_1}{2}\Big(1-\frac{1}{\gamma_0^2}\Big)+\frac{9}{4}-12\xi-\Big(6\xi-\frac{3}{2}\Big)\zeta_1+\frac{1}{4}\zeta_1^2\Big]I_2+\frac{m^2}{2}\Big[\alpha_1\Big(1-\frac{1}{\gamma_0^2}\Big)-2\alpha_3\Big(6\xi-\frac{3}{2}\Big)+\zeta_1\alpha_3\Big]I_3\Big\}$$

$$+\frac{R}{2\gamma_0}\Big[\gamma_2\Big(1-\frac{1}{\gamma_0^2}\Big)I_2+m^2\alpha_2\Big(1-\frac{1}{\gamma_0^2}\Big)I_3\Big]+\frac{H^4}{\gamma_0}\Big[\frac{\gamma_3}{2}\Big(1-\frac{1}{\gamma_0^2}\Big)+\frac{1}{8}\frac{\gamma_1^2}{\gamma_0^2}\Big(\frac{3}{\gamma_0^2}-1\Big)-\frac{\gamma_1}{2\gamma_0^2}\Big(\frac{9}{4}-12\xi\Big)-\Big(6\xi-\frac{3}{2}\Big)\Big(\zeta_3-\frac{\gamma_1\zeta_1}{2\gamma_0^2}\Big)+\frac{\zeta_1\zeta_2}{2}-\frac{\gamma_1\zeta_1^2}{8\gamma_0^2}\Big]I_3$$

$$+\frac{H^2R}{\gamma_0}\left[\frac{\gamma_4}{2}\left(1-\frac{1}{\gamma_0^2}\right)-\frac{\gamma_2}{2\gamma_0^2}\left(\frac{9}{4}-12\xi\right)-\left(6\xi-\frac{3}{2}\right)\left(\zeta_2-\frac{\gamma_2\zeta_1}{2\gamma_0^2}\right)+\frac{\gamma_1\gamma_2}{4\gamma_0^2}\left(\frac{3}{\gamma_0^2}-1\right)\right.$$
$$\left.+\frac{\zeta_1\zeta_2}{2}-\frac{\gamma_2\zeta_1^2}{8\gamma_0}\right]I_3+\frac{R^2}{\gamma_0}\left[\frac{\gamma_5}{2}\left(1-\frac{1}{\gamma_0^2}\right)+\frac{\gamma_2^2}{4\gamma_0^3}\left(\frac{3}{\gamma_0^2}-1\right)\right]I_3$$
$$+\frac{H\dot{R}}{\gamma_0}\left[\frac{\zeta_1\zeta_4}{2}-\left(6\xi-\frac{3}{2}\right)\zeta_4+\frac{\gamma_6}{2}\left(1-\frac{1}{\gamma_0^2}\right)\right]I_3+\frac{\gamma_7\ddot{R}}{2\gamma_0}\left(1-\frac{1}{\gamma_0^2}\right)I_3\,, \tag{2.19}$$

where we have introduced the following divergent integrals:

$$I_1=\frac{1}{8\pi^2}\int_0^\infty p^2(p^2+m^2)^{1/2}dp\,,$$
$$I_2=\frac{1}{8\pi^2}\int_0^\infty \frac{p^2\,dp}{(p^2+m^2)^{1/2}}\,,$$
$$I_3=\frac{1}{8\pi^2}\int_0^\infty \frac{p^2\,dp}{(p^2+m^2)^{3/2}}\,,$$

where $p=k/a$.

In Ref. 7 it is shown that the absorption of the divergencies in the bare constants of the component 00 of the energy momentum tensor is sufficient to assure the same absorption for the other components because all the components are linked by the conservation equation. Therefore the divergencies will be absorbed if

$$\lambda_1=\lambda_1^{(0)}-\gamma_0(1+1/\gamma_0^2)I_1, \tag{2.20a}$$
$$\lambda_2=\lambda_2^{(0)}+\frac{1}{3\gamma_0}\{[(\gamma_1/2)(1-1/\gamma_0^2)+\tfrac{9}{4}-12\xi$$
$$-(6\xi-\tfrac{3}{2})\zeta_1+\tfrac{1}{4}\zeta_1^2]I_2+(m^2/2)[\alpha_1(1-1/\gamma_0^2)$$
$$-2\alpha_3(6\xi-\tfrac{3}{2})+\zeta_1\alpha_3]I_3\}\,, \tag{2.20b}$$
$$\alpha_2(1-1/\gamma_0^2)/\gamma_0=0\,, \tag{2.20c}$$
$$\gamma_2(1-1/\gamma_0^2)/\gamma_0=0\,, \tag{2.20d}$$
$$\frac{1}{\gamma_0}\left[\frac{\gamma_3}{2}\left(1-\frac{1}{\gamma_0^2}\right)+\frac{1}{8}\frac{\gamma_1^2}{\gamma_0^2}\left(\frac{3}{\gamma_0^2}-1\right)\right.$$
$$-\frac{\gamma_1}{2\gamma_0^2}\left(\frac{9}{4}-12\xi\right)-\left(6\xi-\frac{3}{2}\right)\left(\zeta_3-\frac{\gamma_1\zeta_1}{2\gamma_0^2}\right)$$
$$\left.+\frac{\zeta_1\zeta_3}{2}-\frac{\gamma_1\zeta_1^2}{8\gamma_0^2}\right]=0\,, \tag{2.20e}$$
$$\sigma=\sigma^{(0)}+\frac{1}{6\gamma_0}\left[\frac{\gamma_4}{2}\left(1-\frac{1}{\gamma_0^2}\right)\right.$$
$$-\frac{\gamma_2}{2\gamma_0^2}\left(\frac{9}{4}-12\xi\right)-\left(6\xi-\frac{3}{2}\right)\left(\zeta_2-\frac{\gamma_2\zeta_1}{2\gamma_0^2}\right)$$
$$\left.+\frac{\gamma_1\gamma_2}{4\gamma_0^2}\left(\frac{3}{\gamma_0^2}-1\right)+\frac{\zeta_1\zeta_2}{2}-\frac{\gamma_2\zeta_1^2}{8\gamma_0^2}\right]I_3\,, \tag{2.20f}$$
$$\sigma=\sigma^{(0)}-\frac{1}{\gamma_0}\left[\gamma_5\left(1-\frac{1}{\gamma_0^2}\right)+\frac{\gamma_2^2}{4\gamma_0^3}\left(\frac{3}{\gamma_0^2}-1\right)\right]I_3\,, \tag{2.20g}$$
$$\sigma=\sigma^{(0)}+\frac{1}{6\gamma_0}\left[\frac{1}{2}\zeta_1\zeta_4-\left(6\xi-\frac{3}{2}\right)\zeta_4\right.$$
$$\left.+\frac{\gamma_6}{2}\left(1-\frac{1}{\gamma_0^2}\right)\right]I_3\,, \tag{2.20h}$$
$$\gamma_7\left(1-\frac{1}{\gamma_0^2}\right)(\gamma_0)^{-1}=0\,, \tag{2.20i}$$

where $\lambda_1^{(0)}$, $\lambda_2^{(0)}$, and $\sigma^{(0)}$ are the bare constants.

From Eq. (2.20) we can see that condition (i) relates the 15 coefficients of Eq. (2.17) with nine equations. We shall find new restrictions from condition (ii) in the next section.

2. Spin-1 massive field

In this case the action is

$$S=\int\sqrt{|g|}\{-\tfrac{1}{4}F^{\mu\nu}F_{\mu\nu}+\tfrac{1}{2}m^2\varphi^\mu\varphi_\mu\}\,, \tag{2.21}$$

with

$$F_{\mu\nu}=\partial_\mu\varphi_\nu-\partial_\nu\varphi_\mu\,.$$

Following the same procedure as in the previous section we obtain

$$\langle 0,\tau|T_{00}|0,\tau\rangle=\frac{1}{4}\sum_{\sigma=1}^{3}\int\frac{dk^3}{(2\pi a)^3}D_{0_T}\,, \tag{2.22a}$$
$$\langle 0,\tau|T_{jj}|0,\tau\rangle=\frac{1}{2}\int\frac{dk^3}{(2\pi)^3a}D_{0_T}\,,\quad \text{with } j\neq 3\,, \tag{2.22b}$$
$$\langle 0,\tau|T_{33}|0,\tau\rangle=\frac{1}{2}\int\frac{dk^3}{(2\pi)^3a}(D_{3_T}+D_{3_L})\,, \tag{2.22c}$$

where T and L correspond to the transversal and longitudinal modes, and

$$D_{0_\sigma}=\frac{H_\sigma^2}{4\Omega_\sigma}+\frac{H_\sigma}{2}\frac{\dot{\Omega}_\sigma}{\Omega_\sigma^2}+\frac{1}{4\Omega_\sigma}\left(\frac{\dot{\Omega}_\sigma}{\Omega_\sigma}\right)^2+\frac{\omega^2}{\Omega_\sigma}+\Omega_\sigma\,, \tag{2.23a}$$
$$D_{3_T}=\frac{H^2}{4\Omega_T}+\frac{H}{2}\frac{\dot{\Omega}_T}{\Omega_T^2}+\frac{1}{4\Omega_T}\left(\frac{\dot{\Omega}_T}{\Omega_T}\right)^2$$
$$-\frac{m^2}{\Omega_T}+\Omega_T+\frac{k^2}{a^2\Omega_T}\,, \tag{2.23b}$$
$$D_{3_L}=\left[\frac{h^2}{4\Omega_L}+\frac{h}{2}\frac{\dot{\Omega}_L}{\Omega_L^2}+\frac{1}{4\Omega_L}\left(\frac{\dot{\Omega}_L}{\Omega_L}\right)^2\right]$$
$$\times\left(1-\frac{3}{2}\frac{m^2}{\omega^2}\right)-\frac{3}{2}\frac{m^2}{\omega^2}\Omega_L+\frac{m^2}{2\Omega_L}$$
$$+\Omega_L+\frac{k^2}{2a^2\Omega_L}\,, \tag{2.23c}$$

where

$$h=H\left(3-\frac{2m^2}{\omega^2}\right),\quad H_\sigma=\begin{cases}h, & \text{if } \sigma=L\,,\\ H, & \text{if } \sigma=T\,.\end{cases} \tag{2.23d}$$

The function Ω_σ satisfies

$$\Omega_\sigma^2+\frac{1}{2}\left(\frac{\dot\Omega_\sigma}{\Omega_\sigma}\right)^{\cdot}-\frac{1}{4}\left(\frac{\dot\Omega_\sigma}{\Omega_\sigma}\right)^2+\frac{1}{2}H_\sigma+\frac{1}{4}H_\sigma^2=\omega^2. \quad (2.24)$$

Note that for $\xi=\frac{1}{6}$ Eq. (2.15a) has the same functional form as Eq. (2.23a), and Eq. (2.16) the same as Eq. (2.25) if we replace H by h and Ω_σ by Ω_L in the first one.

For the spin-1 field we can also write Eq. (2.18), but we shall do so showing explicitly the polarization modes, to see how the divergencies are absorbed by the bare constants,

$$P_{00}(\tau',\mathbf{x})=2\big[\langle 0,\tau'|T_{00_T}(\tau',\mathbf{x})|0,\tau'\rangle-\lambda^{(0)}_{1_T}+3\lambda^{(0)}_{2_T}H^2+\sigma_T^{(0)}(6H^2R-\tfrac{1}{2}R^2+6H\dot R)\big]$$
$$+\langle 0,\tau'|T_{00_L}(\tau',\mathbf{x})|0,\tau'\rangle-\lambda^{(0)}_{1_L}+3\lambda^{(0)}_{2_L}H^2+\sigma_L^{(0)}(6H^2R-\tfrac{1}{2}R^2+6H\dot R)\,. \quad (2.25)$$

We can see that the equations for the constants of the transversal mode are similar to Eqs. (2.20) with $\xi=\frac{1}{6}$, thus we must study the longitudinal mode only. We begin by computing

$$\langle 0,\tau'|T_{00_L}(\tau',\mathbf{x})|0,\tau'\rangle=\frac{1}{4}\int\frac{dk^3}{(2\pi a)^3}D_{0_L}, \quad (2.26)$$

where D_{0_L} is given by Eq. (2.23a). Now, using the generic Cauchy data introduced in Eq. (2.17) we have

$$\begin{aligned}\langle 0,\tau'|T_{00_L}(\tau',\mathbf{x})|0,\tau'\rangle_{\rm div}&=\gamma_0\left(1+\frac{1}{\gamma_0^2}\right)I_1+\frac{H^2}{2\gamma_0}\left\{\left[\gamma_1\left(1-\frac{1}{\gamma_0^2}\right)+\frac{9}{2}+\frac{\zeta_1}{2}(\zeta_1+6)\right]I_2\right.\\&\left.+m^2\left[\alpha_1\left(1-\frac{1}{\gamma_0^2}\right)+\alpha_3(\zeta_1+3)-2(\zeta_1+3)\right]I_3\right\}+\frac{R}{2\gamma_0}\left(1-\frac{1}{\gamma_0^2}\right)(\gamma_2I_2+m^2\alpha_2I_3)+\frac{H^4}{2\gamma_0}\left[\gamma_3\left(1-\frac{1}{\gamma_0^2}\right)\right.\\&\left.+\frac{\gamma_1^2}{4\gamma_0^2}\left(3-\frac{1}{\gamma_0^2}\right)+\frac{9}{4}\frac{\gamma_1}{\gamma_0^2}+\zeta_3(\zeta_1+3)-\frac{1}{2}\frac{\gamma_1}{\gamma_0^2}\zeta_1(\zeta_1+3)\right]I_3+\frac{R^2}{2\gamma_0}\left[\gamma_5\left(1-\frac{1}{\gamma_0^2}\right)+\frac{\gamma_2^2}{4\gamma_0^2}\left(\frac{3}{\gamma_0^2}-1\right)\right]I_3\\&+\frac{H\dot R}{2\gamma_0}\left[\gamma_4\left(1-\frac{1}{\gamma_0^2}\right)+\zeta_4(\zeta_1+3)\right]I_3+\frac{\gamma_7}{2\gamma_0}\ddot R\left(1-\frac{1}{\gamma_0^2}\right)I_3\,. \quad (2.27)\end{aligned}$$

The only difference between this equation and Eq. (2.19), for the scalar field with conformal coupling, is the term $-(m^2/\gamma_0)(\zeta_1+3)I_3$ that is absent in the scalar case. This fact only modifies λ_2. Thus the dressed constants are

$$\lambda_{1_L}=\lambda^{(0)}_{1_L}-\gamma_0(1-1/\gamma_0^2)I_1\,, \quad (2.28a)$$

$$\lambda_{2_L}=\lambda^{(0)}_{2_L}+\left(\frac{1}{6\gamma_0}\right)\left\{\left[\gamma_1\left(1-\frac{1}{\gamma_0^2}\right)+\frac{\zeta_1^2}{2}+\frac{9}{2}+3\zeta_1\right]I_2\right.$$
$$\left.+m^2\left[\alpha_1\left(1-\frac{1}{\gamma_0^2}\right)+\alpha_3(\zeta_1+3)-2(3+\zeta_1)\right]I_3\right\}, \quad (2.28b)$$

and all the rest of the equation is similar to the one of (2.20) with $\xi=0$.

Thus we have

$$\langle 0,\tau|T_{00_T}(\tau,\mathbf{x})|0,\tau\rangle=\langle 0,\tau|T_{00}(\tau,\mathbf{x})_{\xi=1/6}|0,\tau\rangle_{\rm div}\,,$$
$$\langle 0,\tau|T_{00_L}(\tau,\mathbf{x})|0,\tau\rangle_{\rm div}=\langle 0,\tau|T_{00}(\tau,\mathbf{x})_{\xi=0}|0,\tau\rangle_{\rm div}-(m^2/\gamma_0)H^2(\zeta_1+3)I_3\,. \quad (2.29)$$

Therefore, in the limit $m^2\to 0$, the longitudinal mode has the same divergencies as the scalar mode with minimal coupling.

We shall now analyze the constraints imposed on the vacuum by condition (ii), i.e., the condition that the created number of particles and the created energy must be finite.

III. THE ENERGY AND MOMENTUM OF THE CREATED PARTICLES

The fundamental hypothesis also yields condition (ii); the VEV of the normal ordered energy-momentum tensor operator $\langle 0,\tau|:T_{\mu\nu}(\tau',\mathbf{x}):_{\tau'}|0,\tau\rangle$ must be finite, we can compute this quantity as follows.

(i) We compute the classical expression of the energy-momentum tensor,

$$T_{\mu\nu}=(2/\sqrt{|g|})\delta S_m/\delta g^{\mu\nu}. \quad (3.1)$$

In the cases we are studying, this expression has the form

$$T_{\mu\nu}=\{D_1\varphi^r,D_2\varphi_r\}_{\mu\nu}\,, \quad (3.2)$$

where the index r shall only have one value for a scalar field or four for a vector field and where D_1 and D_2 are linear operators defined in Appendix A. Precisely, linearity is the only property of these operators that we shall use.

(ii) We transform the classical functional $T_{\mu\nu}$ into a quantum operator $\hat T_{\mu\nu}$ (we shall only use the caret when we need to specify the operator nature of $\hat T_{\mu\nu}$),

$$\hat T_{\mu\nu}=T_{\mu\nu}[\hat\varphi\,] \quad (3.3)$$

(we neglect index r if it is not necessary), where $\hat\varphi$ is the quantum field operator, that can be expanded in a basis of solutions defined by its Cauchy data at a time τ' that we shall call $\{\phi_{\mathbf{k}}^{\tau'},\phi_{\mathbf{k}}^{*\tau'}\}$ as in Eq. (2.4).

(iii) We normal order the operator $:\hat T_{\mu\nu}:_{\tau'}$ using the

same definition as in flat space-time, i.e., we put in the first place the creation operators.

(iv) We change the basis to another basis related with time τ: $\{\phi_{\mathbf{k}}^{\tau}\}\cup\{\overset{*}{\phi}_{\mathbf{k}}^{\tau}\}\to\{\phi_{\mathbf{k}}^{\tau}\}\cup\{\overset{*}{\phi}_{\mathbf{k}}^{\tau}\}$ via a Bogoliubov transformation, like (2.16), and we obtain the operator transformation, like (2.16b) (with τ and τ' interchanged) and we obtain

$$:\hat{T}_{\mu\nu}:_{\tau'} = :\hat{T}_{\mu\nu}:_{\tau}[\phi^{\tau},\overset{*}{\phi}^{\tau},a^{\tau},a^{\tau\dagger}]\,. \tag{3.4}$$

(v) We compute the VEV of $:\hat{T}_{\mu\nu}:_{\tau'}$ as

$$\langle 0,\tau|:T_{\mu\nu}:_{\tau'}|0,\tau\rangle = 4\int dk^3\{\mathrm{Re}[\alpha^{\tau\tau'}\overset{*}{\beta}^{\tau\tau'}D_1\phi^{\tau'}\cdot D_2\phi^{\tau'}]_{\mu\nu} + |\beta^{\tau\tau'}|^2\,\mathrm{Re}[D_1\phi^{\tau'}\cdot D_2\overset{*}{\phi}^{\tau'}]_{\mu\nu}\}\,, \tag{3.5}$$

for the isotropic universe that we considered; we can drop one of the k' integrals because the spatial phase produce a delta function.

The second term represents the component corresponding to the particles created between times τ and τ' because

$$\langle 0,\tau|N_{\mathbf{k}}^{\tau}|0,\tau\rangle = |\beta_{\mathbf{k}}^{\tau\tau'}|^2\,,$$

where $N_{\mathbf{k}}^{\tau} = a_{\mathbf{k}}^{\tau\dagger}a_{\mathbf{k}}^{\tau}$ is the particle number operator for mode k at time τ'.

We shall compute Eq. (3.5) for the cases of spin 0 and 1, to identify the divergencies and see how we can eliminate them by choosing an adequate vacuum.

A. Scalar field

We can obtain the field equation of the scalar field taking the variation of the action (2.10):

$$(\Box + m^2 + \xi R)\varphi_{\xi} = 0\,, \tag{3.6}$$

where $\Box = -\nabla_{\mu}\nabla^{\mu}$ and ξ is an arbitrary dimensionless coupling constant that we also use as an index of the field φ_3.

By variable separation in the RW metric (2.9) we can find an orthogonal basis

$$\phi_3 = [1/(2\pi)^{3/2}a]\chi_{\xi}e^{i\mathbf{k}\cdot\mathbf{x}} \tag{3.7}$$

where we have suppressed the index k and introduced the "conformal time" η defined as $\eta = \int(dt/a)$.

The new function χ_{ξ} must satisfy the equation

$$\frac{d^2\chi_{\xi}}{d\eta^2} + k^2p_{\xi}\chi_{\xi} = 0\,, \tag{3.8}$$

with

$$p_{\xi} = 1 + (a^2/k^2)m^2 + (\xi - \tfrac{1}{6})(a^2/k^2)R^2\,.$$

Besides, we can write function χ_{ξ} as

$$\chi_{\xi} = \frac{1}{\sqrt{2V_{\xi}}}\exp\left(-i\int^{\eta}V_{\xi}\,d\eta'\right) \tag{3.9a}$$

with

$$V_{\xi} = a\Omega_{\xi}\,, \tag{3.9b}$$

i.e., in a "WKB way." Function Ω_{ξ} must satisfy Eq. (2.16). Replacing Eq. (3.7) in (2.12) and making a Bogoliubov transformation between $\{\chi_{\xi}^{\tau}\}\cup\{\overset{*}{\chi}_{\xi}^{\tau}\}$ and $\{\chi_{\xi}^{\tau}\}\cup\{\overset{*}{\chi}_{\xi}^{\tau}\}$ we have an equation similar to (3.5),

$$\langle 0,\tau|:T_{00_{\xi}}(\tau'):|0,\tau\rangle = \frac{1}{2}\int\frac{dk^3}{(2\pi a)^3}\{2|\beta^{\tau\tau'}|^2[|\dot{\chi}_{\xi}^{\tau}|^2 + 2\,\mathrm{Re}(\dot{\chi}_{\xi}^{\tau}\overset{*}{\chi}_{\xi}^{\tau})(6\xi-1)H + |\chi_{\xi}^{\tau}|^2[\omega^2 + (1-6\xi)H^2] + 2\,\mathrm{Re}[\alpha\overset{*}{\beta}(\dot{\chi}_{\xi}^{\tau^2} + 2\dot{\chi}_{\xi}^{\tau}\chi_{\xi}^{\tau}(6\xi-1)H + \chi_{\xi}^{\tau^2}(\omega^2 + (1-6\xi)H^2))]]\}\,, \tag{3.10}$$

$$\langle 0,\tau|:T_{ij\xi}(\tau'):_{\tau'}|0,\tau\rangle = 2g_{ij}\int\frac{dk^3}{(2\pi)^3a^2}\left\{|\beta^{\tau\tau'}|^2\left[|\chi^{\tau'}|^2\left(2\left(\xi-\frac{1}{6}\right)\left(\xi R - \frac{3}{2}H^2\right) + \frac{1}{3}\frac{k^2}{a^2} + 2\left(\xi-\frac{1}{4}\right)\omega^2\right) - 2\left(\xi-\frac{1}{4}\right)|\dot{\chi}^{\tau'}|^2 + 2H(\dot{\chi}^{\tau'}\overset{*}{\chi}^{\tau'} + \overset{*}{\chi}^{\tau'}\chi^{\tau'})\left(\xi-\frac{1}{8}\right)\right] + \mathrm{Re}\left[\alpha^{\tau\tau'}\overset{*}{\beta}^{\tau\tau'}\left[\chi^{\tau'^2}\left(2\left(\xi-\frac{1}{6}\right)\left(\xi R - \frac{3}{2}H^2\right) + \frac{1}{3}\frac{k^2}{a^2} + 2\left(3-\frac{1}{4}\right)\omega^2\right) - 2\left(\xi-\frac{1}{4}\right)\dot{\chi}^{\tau'^2} + 4H\left(\xi-\frac{1}{8}\right)\dot{\chi}^{\tau'}\chi^{\tau'}\right]\right]\right\}\,. \tag{3.11}$$

The terms with factor $|\beta^{\tau\tau'}|^2$ in these equations are related with the particle creation, they have a factor $\mathbf{k}$, thus they will be convergent if either $\lim_{k\to\infty}k^3|\beta|^2 = 0$ or $|\beta|^2 \lesssim 1/k^5$ from a certain value of $\mathbf{k}$ onwards. In fact, these are two sufficient conditions for the convergence of the expression.

From Ref. 11 we can see that the terms of Eqs. (3.10) and (3.11) with products like $\alpha\overset{*}{\beta}$ do not introduce divergencies if the terms with $|\beta|^2$ do not. In fact, it is sufficient to expand function χ_{ξ} of Eq. (3.9a) in a WKB series, to take the lowest order of the expansion and to study the terms that grow faster for high energies. Thus the conditions stated for the terms with $|\beta|^2$ are the only conditions to fulfill condition (ii) in the scalar case.

B. Massive vector fields

From action (2.22) we can obtain the field equation

$$\nabla^{\mu}F_{\mu\nu} + m^2\varphi_{\nu} = 0\,. \tag{3.12}$$

In Appendix B we find a basis of solutions of the field equation $\{\phi_{\mathbf{k},\sigma}^{\mu}\}\cup\{\overset{*}{\phi}_{\mathbf{k},\sigma}^{\nu}\}$ where index σ symbolizes the polarization modes, k is the linear momentum, and μ the world index. This basis is orthonormal under the inner product $\langle;\rangle$, i.e.,

$$\langle\phi_{\mathbf{k},\sigma}^{\mu};\phi_{\mathbf{k}',\sigma'}^{\nu}\rangle = i\int_{\Sigma}(\overset{*}{\phi}_{\mathbf{k}}^{\mu}\nabla_{\lambda}\phi_{\mathbf{k}',\mu} - \phi_{\mathbf{k}'}^{\mu}\nabla_{\lambda}\overset{*}{\phi}_{\mathbf{k}\mu})\,d\sigma^{\lambda} = \delta_{\sigma\sigma'}\delta(\mathbf{k}-\mathbf{k}')\,.$$

As in the scalar field case we can expand the field as

$$\phi^{\mu}(x)=\sum_{\sigma}\int dk^{3}\left[a_{\mathbf{k}\sigma}\phi^{\mu}_{\mathbf{k}\sigma}(x)+a^{\dagger}_{\mathbf{k}\sigma}\overset{*}{\phi}{}^{\mu}_{\mathbf{k}\sigma}(x)\right]. \quad (3.13)$$

The basis of solutions of the field equation is

$$\phi^{\mu}_{1}=\frac{1}{(2\pi)^{3/2}}\begin{pmatrix}0\\1\\0\\0\end{pmatrix}\chi_{T}e^{i\mathbf{k}\cdot\mathbf{x}},$$

$$\phi^{\mu}_{2}=\frac{1}{(2\pi)^{3/2}}\begin{pmatrix}0\\0\\1\\0\end{pmatrix}\chi_{T}e^{i\mathbf{k}\cdot\mathbf{x}},$$

$$\phi^{\mu}_{3}=\frac{\omega}{(2\pi)^{3/2}m}\begin{pmatrix}kW/a^{2}\omega^{2}\\0\\0\\1\end{pmatrix}\chi_{L}e^{i\mathbf{k}\cdot\mathbf{x}}, \quad (3.14)$$

where

$$W=i\frac{d}{dt}\ln(\omega\chi_{L}).$$

In fact, we have chosen a basis such that the z axis is parallel to vector $\mathbf{k}$ (see Appendix B). Function χ_{σ} with $\sigma=T,L$ for the transversal or the longitudinal mode, respectively, satisfies an equation similar to (3.13) but with a new p_{σ} in the place of p_{ξ},

$$p_{\sigma}=1+\frac{a^{2}}{k^{2}}m^{2}+\left(\xi_{\sigma}-\frac{1}{6}\right)\frac{a^{2}}{k^{2}}R+\left[\frac{a^{2}m^{2}}{k^{2}\omega^{2}}\left(3H^{2}+\frac{R}{6}\right)-3\frac{a^{2}}{k^{2}}\frac{m^{4}}{\omega^{4}}H^{2}\right]\delta_{\sigma_{L}}, \quad (3.15)$$

where

$$\xi_{\sigma}=\begin{cases}\frac{1}{6}, & \text{if } \sigma=T,\\ 0, & \text{if } \sigma=L.\end{cases}$$

In the first case the equation coincides with the one for the scalar field with conformal coupling.

In order to obtain an orthogonal basis, function χ_{σ} must satisfy

$$\frac{d\overset{*}{\chi}_{\mathbf{k}\sigma}}{d\eta}\chi_{\mathbf{k}\sigma}-\overset{*}{\chi}_{\mathbf{k}\sigma}\frac{d\chi_{\mathbf{k}\sigma}}{d\eta}=i, \quad (3.16)$$

thus we can chose them as

$$\chi^{\tau}_{\sigma}=\frac{1}{\sqrt{2a\Omega^{\tau}_{\sigma}}}\exp\left(-i\int^{t}\Omega^{\tau}_{\sigma}\,dt'\right), \quad (3.17)$$

where function Ω^{τ}_{σ} is a solution of Eq. (2.24). The index makes explicit the time when the Cauchy data of the basis was stated.

With the change of variables $\chi_{L}=(m/\omega)f_{L}$ and $\chi_{T}=f_{T}$ the nonvanishing components of the energy-momentum tensor turn out to be

$$\langle 0,\tau|T_{00}(\tau')|0,\tau\rangle=\frac{1}{2a^{2}(2\pi)^{3}}\int dk^{3}\{2[|\dot{f}^{\tau}_{T}|^{2}+\omega^{2}|f^{\tau}_{T}|^{2}]+\frac{m^{2}}{\omega^{2}}[|\dot{f}^{\tau}_{L}|^{2}+\omega^{2}|f^{\tau}_{L}|^{2}]\}, \quad (3.18a)$$

$$\langle 0,\tau|T_{jj}(\tau')|0,\tau\rangle=\frac{1}{(2\pi)^{3}}\int dk^{3}\{|\dot{f}^{\tau}_{T}|^{2}+\omega^{2}|f^{\tau}_{T}|^{2}\}\ \text{with } j=1,2, \quad (3.18b)$$

$$\langle 0,\tau|T_{33}(\tau')|0,\tau\rangle=\frac{1}{(2\pi)^{3}}\int dk^{3}\left\{|\dot{f}^{\tau}_{T}|^{2}+\left(\frac{k^{2}}{a^{2}}-m^{2}\right)|f^{\tau}_{T}|^{2}+\frac{m^{2}}{\omega^{2}}\left(1-\frac{3}{2}\frac{m^{2}}{\omega^{2}}\right)|\dot{f}^{\tau}_{L}|^{2}+\frac{m^{2}}{2}|f^{\tau}_{L}|^{2}\right\}. \quad (3.18c)$$

Via a Bogoliubov transformation from $\{f^{\tau}\}\cup\{\overset{*}{f}{}^{\tau}\}$ to basis $\{f^{\tau'}\}\cup\{\overset{*}{f}{}^{\tau'}\}$ we can obtain an expression like Eqs. (3.16) and (3.17) for the massive vector field, and also using Eq. (2.6) we can obtain

$$\langle 0,\tau|:T_{00}(\tau'):_{\tau'}|0,\tau\rangle=\frac{1}{a^{2}(2\pi)^{3}}\int dk^{3}\{2|\beta^{\tau\tau'}_{T}(\tau')|^{2}(|\dot{f}^{\tau'}_{T}|+\omega^{2}|f^{\tau'}_{T}|^{2})+2\,\mathrm{Re}[\alpha^{\tau\tau'}_{T}\beta^{\tau\tau'}_{T}(\dot{f}^{\tau'2}_{T}+\omega_{2}f^{\tau'2}_{T})]+\frac{m^{2}}{\omega^{2}}[|\beta^{\tau\tau'}_{L}(\tau')|^{2}(\omega^{2}|f^{\tau'}_{L}|^{2}+|\dot{f}^{\tau'}_{L}|^{2})+\mathrm{Re}[\alpha^{\tau\tau'}_{L}\beta^{\tau\tau'}_{L}(\dot{f}^{\tau'2}_{L}+\omega^{2}f^{\tau'2}_{L})]]\}, \quad (3.19a)$$

$$\langle 0,\tau|:T_{jj}(\tau'):_{\tau'}|0,\tau\rangle=\frac{2}{(2\pi)^{3}}\int dk^{3}\{|\beta^{\tau\tau'}_{T}(\tau')|^{2}(|\dot{f}^{\tau'}_{T}|^{2}+\omega^{2}|f^{\tau'}_{T}|^{2})+\mathrm{Re}[\alpha^{\tau\tau'}_{T}\beta^{\tau\tau'}_{T}(\dot{f}^{\tau'2}_{T}+\omega^{2}f^{\tau'2}_{T})]\}, \quad (3.19b)$$

$$\langle 0,\tau|:T_{33}(\tau'):_{\tau'}|0,\tau\rangle=\frac{2}{(2\pi)^{3}}\int dk^{3}\left\{|\beta^{\tau\tau'}_{T}(\tau')|^{2}\left[|\dot{f}^{\tau'}_{T}|^{2}+\left(\frac{k^{2}}{\alpha^{2}}-m^{2}\right)|f^{\tau'}_{T}|^{2}\right]+|\beta^{\tau\tau'}_{L}(\tau')|^{2}\left[\frac{m^{2}}{\omega^{2}}\left(1-\frac{3}{2}\frac{m^{2}}{\omega^{2}}\right)|\dot{f}^{\tau'}_{L}|^{2}+\frac{m^{2}}{2}|f^{\tau'}_{L}|^{2}\right]+\mathrm{Re}\left\{\alpha^{\tau\tau'}_{T}\overset{*}{\beta}{}^{\tau\tau'}_{T}\left[\dot{f}^{\tau'2}_{T}+\left(\frac{k^{2}}{a^{2}}-m^{2}\right)f^{\tau'2}_{T}\right]+\frac{m^{2}}{2\omega^{2}}\alpha^{\tau\tau'}_{L}\beta^{\tau\tau'}_{L}\left[\left(2-3\frac{m^{2}}{\omega^{2}}\right)\dot{f}^{\tau'2}_{L}+\omega^{2}f^{\tau'2}_{L}\right]\right\}\right\}. \quad (3.19c)$$

As we can see from these equations, for the vector field we obtain the same rule (to assure that we have no divergency), i.e.,

$$\langle 0,\tau|:T_{\mu\nu}(\tau'):_{\tau'}|0,\tau\rangle\neq\infty\Leftrightarrow|\beta^{\tau\tau'}(\tau')|^{2}\lesssim\mathcal{O}(1/k^{5}). \quad (3.20)$$

Thus this is the mathematical form of condition (ii) of the fundamental hypothesis H1. Next, we shall see how this

condition constraints the coefficient introduced in the vacuum definition.

C. Constraint to obtain a finite energy and moment of the created particles

We shall study the problem in a unified way for both spin 0 and 1. We begin by the system of equations

$$\chi_\sigma^\tau = \alpha_\sigma^{\tau\tau'}\chi_\sigma^{\tau'} + \beta_\sigma^{\tau\tau'}\overset{*}{\chi}{}_\sigma^{\tau'}, \tag{3.21a}$$

$$\frac{d\chi_\sigma^\tau}{d\eta} = \alpha_\sigma^{\tau\tau'}\frac{d\chi_\sigma^{\tau'}}{d\eta} + \beta_\sigma^{\tau\tau'}\frac{d\overset{*}{\chi}{}_\sigma^{\tau'}}{d\eta}, \tag{3.21b}$$

where $\sigma = T,L$, because as we have shown we can independently treat both polarization modes and $\sigma = \xi$ for spin 0.

From system (3.21) we can obtain

$$|\beta^{\tau\tau'}(\tau')|^2 = \left|\chi^{\tau'}(\tau')\frac{d\chi^\tau}{d\eta}(\tau') - \chi^\tau(\tau')\frac{d\chi^{\tau'}}{d\eta}(\tau')\right|^2. \tag{3.22}$$

To study the convergence of Eq. (3.22) we use a theorem of Ref. 12 that gives us an upper bound for the particle production.

Theorem: If $k>0, p=p(\mathbf{k},\eta)$ is a positive defined function and $d^2p/d\eta^2$ is a continuous function of η, then Eq. (3.13) has two solutions ψ and $\overset{*}{\psi}$ given by the formula

$$\psi(\mathbf{k},\eta) = \frac{1}{\sqrt{2kp^{1/2}}}\left[\exp\left(ik\int_0^\eta p^{1/2}\,d\eta'\right) + \epsilon(k,\eta)\right], \tag{3.23}$$

with derivatives

$$\frac{d\psi}{d\eta}(\mathbf{k},\eta) = i\sqrt{\frac{k}{2}p^{1/2}}\left\{\exp\left(ik\int_0^\eta p^{1/2}\,d\eta'\right) + \frac{i}{4k}\frac{1}{p^{3/2}}\frac{dp}{d\eta} \times\left[\exp\left(ik\int_0^\eta p^{1/2}\,d\eta'\right) + \epsilon(k,\eta)\right] - i\zeta(k,\eta)\right\},$$

where $\epsilon(k,\eta)$ and $\zeta(k,\eta)$ are functions with their modulus bounded as

$$|\epsilon|,|\zeta| \leqslant \exp(F/k) - 1, \tag{3.24}$$

where

$$F(k,\eta) = \int_0^\eta |p^{-1/4}\frac{d^2}{d\eta^2}p^{-1/4}|d\eta'. \tag{3.25}$$

To use the theorem, p must be positive defined. Then, for high energies, Eq. (3.21) yields

$$k^2/a^2 \geqslant R/6 - m^2. \tag{3.26}$$

Thus we cannot use the theorem at low energies, but this is unimportant as we only want to study the high energy behavior.

So let us use the theorem computing Eq. (3.28). We must take into account that an arbitrary solution of Eq. (3.13) is the linear combination

$$\psi_\sigma(t) = A\chi_\sigma^\tau(t) + B\overset{*}{\chi}{}_\sigma^\tau(t), \tag{3.27a}$$

$$\frac{d\psi_\sigma}{d\eta}(t) = A\frac{d\chi_\sigma^\tau}{d\eta}(t) + B\frac{d\overset{*}{\chi}{}_\sigma^\tau}{d\eta}(t). \tag{3.27b}$$

Also using Eqs. (3.22) and (3.26) and the conjugated expression of (3.27) corresponding to times τ and τ' we obtain

$$\begin{aligned}|\beta_\sigma^{\tau\tau'}(\tau')|^2 &= \left| -\chi_\sigma^\tau(\tau)\chi_\sigma^{\tau'}(\tau')\mathrm{Re}\left[\frac{d\overset{*}{\psi}_\sigma}{d\eta}(\tau)\frac{d\psi_\sigma}{d\eta}(\tau')\right]\right.\\ &+\chi_\sigma^{\tau'}(\tau')\frac{d\chi_\sigma^\tau}{d\eta}(\tau')\mathrm{Re}\left[\overset{*}{\psi}_\sigma(\tau)\frac{d\psi_\sigma}{d\eta}(\tau')\right]\\ &-\frac{d\chi_\sigma^\tau}{d\eta}(\tau)\frac{d\chi_\sigma^{\tau'}}{d\eta}(\tau')\mathrm{Re}[\overset{*}{\psi}_\sigma(\tau)\psi_\sigma(\tau')]\\ &\left.+\chi_\sigma^\tau(\tau)\frac{d\chi_\sigma^{\tau'}}{d\eta}(\tau')\mathrm{Re}\left[\psi_\sigma(\tau')\frac{d\overset{*}{\psi}_\sigma}{d\eta}(\tau)\right]\right|^2\\ &\times\left[\mathrm{Im}\left(\psi_\sigma(\tau)\frac{d\overset{*}{\psi}_\sigma}{d\eta}(\tau)\right)\right]^{-2}.\end{aligned} \tag{3.28}$$

This expression is useful because it relates $|\beta^{\tau\tau'}|^2$ with the Cauchy data of the vacuum we are using. We can compute these data from

$$\chi_\sigma^t(t) = \frac{1}{\sqrt{2a\Omega_\sigma^t(t)}}\exp\left(-i\int_{t_0}^t \Omega_\sigma^t\,dt'\right), \tag{3.29a}$$

$$\frac{d\chi_\sigma^t}{d\eta}(t) = -\frac{a}{2}\left[H + \frac{\dot\Omega_\sigma^t}{\Omega_\sigma^t}(t) + 2i\Omega_\sigma^t(t)\right]\chi_\sigma^t(t), \tag{3.29b}$$

where $\Omega_\sigma^t(t)$ and $\dot\Omega_\sigma^t(t)/\Omega_\sigma^t(t)$ are given by Eq. (2.17).

As we shall analyze the behavior for high energies we can consider that the particles kinetic energy is much bigger than their mass: $k^2/a^2 \gg m^2$. Then we can compute function $F(k,\eta)$ as a power expansion in k^{-1}, and we obtain, at the lowest order, that

$$F(k,\eta) \simeq (1/k^2)|G(\eta)|,$$

$$G(\eta) = \frac{1}{4}\int_0^\eta \left|m^2\frac{d^2a^2}{d\eta^2} + \left(\xi_\sigma - \frac{1}{6}\right)\frac{d^2}{d\eta^2}(a^2R)\right|d\eta',$$

where G is a function of η only. Then for high energies Eq. (3.24) turns out to be

$$|\zeta|,|\epsilon| < |G(\eta)|/k^3 + \mathcal{O}(1/k^5). \tag{3.30}$$

From this equation we can study the high energy behavior of the terms of Eq. (3.23). In fact, from Eqs. (3.23) and (3.24) we can compute the mean value of the created particles $|\beta^{\tau\tau'}|^2$ as a function of the Cauchy data:

$$|\beta_\sigma^{\tau\tau'}(\tau')|^2=\frac{1}{2a(\tau)a(\tau')\Omega_\sigma^\tau(\tau)\Omega_\sigma^{\tau'}(\tau')}\left|\mathrm{Re}\left[\frac{d\overset{*}{\psi}_\sigma}{d\eta}(\tau)\frac{d\psi_\sigma}{d\eta}(\tau')\right]+\frac{a(\tau)}{2}\left[H(\tau)+\frac{\dot\Omega_\sigma^\tau(\tau)}{\Omega_\sigma^\tau(\tau)}+2i\Omega_\sigma^\tau(\tau)\right]\right.$$
$$\times\mathrm{Re}\left[\overset{*}{\psi}_\sigma(\tau)\frac{d\psi_\sigma}{d\eta}(\tau')\right]+\frac{a(\tau')}{2}\left[H(\tau')+\frac{\dot\Omega_\sigma^{\tau'}(\tau')}{\Omega_\sigma^{\tau'}(\tau')}+2i\Omega_\sigma^{\tau'}(\tau')\right]\mathrm{Re}\left[\overset{*}{\psi}_\sigma(\tau')\frac{d\overset{*}{\psi}_\sigma}{d\eta}(\tau)\right]$$
$$+\frac{1}{4}a(\tau)a(\tau')\left[H(\tau)+\frac{\dot\Omega_\sigma^\tau(\tau)}{\Omega_\sigma^\tau(\tau)}+i\,\Omega_\sigma^\tau(\tau)\right]\left[H(\tau')+\frac{\dot\Omega_\sigma^{\tau'}(\tau')}{\Omega_\sigma^{\tau'}(\tau')}+2i\,\Omega_\sigma^{\tau'}(\tau')\right]$$
$$\left.\times\mathrm{Re}[\overset{*}{\psi}_\sigma(\tau)\psi_\sigma(\tau')]\right|^2\left[\mathrm{Im}\left(\psi_\sigma(\tau)\frac{d\overset{*}{\psi}_\sigma}{d\eta}(\tau)\right)\right]^{-2}. \tag{3.31}$$

If we now use Eq. (3.29) in Eq. (3.37) and we only write the lowest powers in k^{-1} we have

$$|\beta_\sigma^{\tau\tau'}(\tau')|^2\simeq\frac{1}{2a(\tau)a(\tau')\Omega_\sigma^\tau(\tau)\Omega_\sigma^{\tau'}(\tau')}\left|\left[1-\frac{a(\tau)a(\tau')\Omega_\sigma^\tau(\tau)\Omega_\sigma^{\tau'}(\tau')}{k^2p^{1/4}(\tau)p^{1/4}(\tau')}\right]\cos k\int_{\eta'}^{\eta}p^{1/2}\,d\eta''\right.$$
$$+\frac{i}{k}\left[a(\tau)\Omega_\sigma^\tau(\tau)-a(\tau')\Omega_\sigma^{\tau'}(\tau')\right]\sin k\int_{\eta'}^{\eta}p^{1/2}\,d\eta''+\frac{1}{2k}\left\{a(\tau)\left[H(\tau)+\frac{\dot\Omega_\sigma^\tau(\tau)}{\Omega_\sigma^\tau(\tau)}\right]-a(\tau')\right.$$
$$\left.\times\left[H(\tau')+\frac{\dot\Omega_\sigma^{\tau'}(\tau')}{\Omega_\sigma^{\tau'}(\tau')}\right]\right\}\sin k\int_{\eta'}^{\eta}p^{1/2}\,d\eta''+\frac{a(\tau)a(\tau')}{4k^2p^{1/4}(\tau)p^{1/4}(\tau')}\left\{\left[H(\tau)+\frac{\dot\Omega_\sigma^\tau(\tau)}{\Omega_\sigma^\tau(\tau)}\right]\left[H(\tau')+\frac{\dot\Omega_\sigma^{\tau'}(\tau')}{\Omega_\sigma^{\tau'}(\tau')}\right]\right.$$
$$+2i\left[H(\tau')+\frac{\dot\Omega_\sigma^{\tau'}(\tau')}{\Omega_\sigma^{\tau'}(\tau')}\right]\Omega_\sigma^\tau(\tau)+2i\left[H(\tau)+\frac{\dot\Omega_\sigma^\tau(\tau)}{\Omega_\sigma^\tau(\tau)}\right]\Omega_\sigma^{\tau'}(\tau')\left.\right\}\cos k\int_{\eta'}^{\eta}p^{1/2}\,d\eta''$$
$$\left.+T\left[|\epsilon(k,\eta)|,|\epsilon(k,\eta')|,|\xi(k,\eta)|,|\xi(k,\eta')|\right]\right|^2\left[\mathrm{Im}\left(\psi_\sigma(\tau)\frac{d\overset{*}{\psi}_\sigma}{d\eta}(\tau)\right)\right]^{-2}. \tag{3.32a}$$

Introducing the new functions $\varphi(t)$ and $\gamma(t)$ as

$$\epsilon(t)=|\epsilon(t)|\exp i\varphi(t),\quad \zeta(t)=|\zeta(t)|\exp i\gamma(t),$$

we obtain

$$T\left[|\epsilon(k,\eta)|,|\epsilon(k,\eta')|,|\,\zeta(k,\eta)|,|\zeta(k,\eta')|\,\right]$$
$$=|\zeta(k,\eta)|\sin\left[\gamma(\eta)-k\int_0^{\eta'}p^{1/2}\,d\eta''\right]$$
$$+|\zeta(k,\eta')\sin\left[\gamma(\eta')-k\int_0^{\eta}p^{1/2}\,d\eta'\right]-\frac{a(\tau)}{2k}\left[H(\tau)+\frac{\dot\Omega_\sigma^\tau(\tau)}{\Omega_\sigma^\tau(\tau)}+2i\Omega_\sigma^\tau(\tau)\right]|\epsilon(k,\eta)|\sin\left(\varphi(\eta)-k\int_0^{\eta'}p^{1/2}\,d\eta''\right)$$
$$-\frac{a(\tau')}{2k}\left[H(\tau')+\frac{\dot\Omega_\sigma^{\tau'}(\tau')}{\Omega_\sigma^{\tau'}(\tau')}+2i\Omega_\sigma^{\tau'}(\tau')\right]|\epsilon(k,\eta')|\sin\left(\varphi(\eta')-k\int_0^{\eta}p^{1/2}\,d\eta''\right)$$
$$+\frac{a(\tau)a(\tau')}{4k^2p^{1/4}(\tau)p^{1/4}(\tau')}\left[H(\tau)+\frac{\dot\Omega_\sigma^\tau(\tau)}{\Omega_\sigma^\tau(\tau)}+2i\Omega_\sigma^\tau(\tau)\right]\left[H(\tau')+\frac{\Delta\Omega_\sigma^{\tau'}(\tau')}{\Omega_\sigma^{\tau'}(\tau')}+2i\Omega_\sigma^{\tau'}(\tau')\right]|\epsilon(k,\eta')|$$
$$\times\cos\left(k\int_0^{\eta}p^{1/2}\,d\eta'-\varphi(\eta')\right)|\epsilon(k,\eta)|\cos\left(k\int_0^{\eta'}p^{1/2}\,d\eta''-\varphi(\eta)\right), \tag{3.32b}$$

where

$$\mathrm{Im}\left[\psi_\sigma(k,\eta)\frac{d\overset{*}{\psi}_\sigma}{d\eta}(k,\eta)\right]=-\frac{1}{2}-\frac{1}{2}|\epsilon(k,\eta)|\cos\left(\varphi-k\int_0^{\eta}p^{1/2}\,d\eta'\right)+\mathcal{O}(1/k^4).$$

We must replace Eq. (2.27) in Eq. (3.32) to analyze the behavior of $|\beta|^2$ for high energies. Then, for the lowest order in powers of k^{-1} (which we shall call the zeroth order) of the Cauchy data, Eq. (3.20) yields the constraints

$$1-\frac{a(\tau)a(\tau')\Omega_\sigma^{\tau(0)}(\tau)\Omega_\sigma^{\tau'(0)}(\tau')}{k^2p^{1/4(0)}(\tau)p^{1/4(0)}(\tau')}, \tag{3.33a}$$

$$a(\tau)\Omega_\sigma^{\tau(0)}(\tau)-a(\tau')\Omega_\sigma^{\tau'(0)}(\tau')=0, \tag{3.33b}$$

$$H(\tau)+\left(\frac{\dot\Omega_\sigma^\tau(\tau)}{\Omega_\sigma^\tau(\tau)}\right)^{(0)}=0, \tag{3.33c}$$

and a similar equation for τ'.

The restriction in the parameters of Eq. (2.17) is then

$$1-\gamma_0^\tau\gamma_0^{\tau'}=0, \tag{3.34a}$$

$$\gamma_0^\tau-\gamma_0^{\tau'}=0, \tag{3.34b}$$

$$1+\zeta_1^\tau=0\,, \tag{3.34c}$$

$$1+\zeta_1^{\tau'}=0\,. \tag{3.34d}$$

Thus these equations are the conditions that must be satisfied to fulfill point (ii) of hypothesis H1. They may be written as

$$\gamma_0^{t^2}=1\,, \tag{3.34e}$$

$$\zeta_1^t=-1,\quad \forall t\,, \tag{3.34f}$$

i.e., at the lowest order Ω_k^2 must coincide with ω_k^2 and $\dot\Omega_k/\Omega_k$ with $-H$ [cf. Eq. (2.17)]. In the next section we shall analyze the physical consequences of (3.34).

Finally we can remark that when Eq. (3.34) is fulfilled $\beta^{\tau\tau'}$ is bound by

$$|\beta_{\tau\tau'}(\tau')|^2\leqslant|G(\eta)|^2/k^6+\mathcal{O}(1/k^8) \tag{3.35}$$

(see Ref. 13) and thus the VEV of Eq. (3.25) turns out to be finite.

IV. THE VACUUM AND THE PARTICLE MODEL: THE GROUND STATE HYPOTHESIS H2

A very natural criterion, and also a natural generalization of one of the most important properties of flat space vacua, is to state that the VEV of the energy must be a minimum: i.e., the ground state hypothesis H2. We shall compute the restrictions that the parameters of Eq. (2.17) must fulfill in order that H2 should be valid, as well as the relation between H1 and H2. As we shall see, H2 will coincide with the diagonalization of the Hamiltonian (see also Refs. 13 and 14). The VEV of the metric Hamiltonian $\mathscr{H}$ can be defined in RW universes from the 00 component of the energy-momentum tensor as

$$\mathscr{H}=\int a^3\,d\mathbf{x}^3\langle 0,\tau|T_{00}|0,\tau\rangle\,. \tag{4.1}$$

(For a definition in more general cases see Ref. 15.)

We shall use a notation that allows us to simultaneously analyze the scalar field with arbitrary coupling and the massive vector field. Thus for the RW metric (2.9) we can write the VEV of the energy for the scalar field or for each mode of the vector field as

$$\mathscr{H}_\sigma=\frac{v}{4(2\pi a)^3}\int dk^3\left\{\frac{\omega^2}{\Omega_\sigma^\tau}+\Omega_\sigma^\tau+\frac{1}{2}H_\sigma\frac{\dot\Omega_\sigma^\tau}{\Omega_\sigma^{\tau\,2}}\lambda_\sigma+\frac{1}{4}\frac{\dot\Omega_\sigma^{\tau^2}}{\Omega_\sigma^{\tau^2}}+\frac{1}{4}\gamma_\sigma\frac{H_\sigma^2}{\Omega_\sigma^\tau}\right\}. \tag{4.2}$$

For the scalar field we shall have $\mathscr{H}=\mathscr{H}_\xi$ and for the massive vector field we shall have $\mathscr{H}=2\mathscr{H}_T+\mathscr{H}_L$. This last decomposition can be used to minimize each mode separately. The meaning of each quantity in (4.2) is shown in Table I.

In Eq. (4.2), $v=\int a^3\,dx^3$ is an infinite constant, because we are working in a spatially flat universe. We neglect this constant because what we shall really minimize is the energy density.

Thus we shall derivate $\mathscr{H}_\sigma$ with respect to its Cauchy data $\Omega_\sigma^\tau(t)$ and $\dot\Omega_\sigma^\tau(t)$ at $t=\tau$, and compute these data to obtain vanishing derivatives.

If we call D_{0_σ} the integrand of Eq. (4.2) we can see that

$$\frac{\partial D_{0_\sigma}}{\partial(\partial\Omega_\sigma^\tau/\partial k_j)}=\frac{\partial D_{0_\sigma}}{\partial(\partial\dot\Omega_\sigma^\tau/\partial k_j)}=0\,,\quad j=1,2,3,$$

then

$$\frac{\delta\mathscr{H}_\sigma}{\delta\Omega_\sigma^\tau}=\frac{\partial D_{0_\sigma}}{\partial\Omega_\sigma^\tau},\quad \frac{\delta\mathscr{H}_\sigma}{\delta\dot\Omega_\sigma^\tau}=\frac{\partial D_{0_\sigma}}{\partial\dot\Omega_\sigma^\tau}.$$

Then an extremum must satisfy

$$\left.\frac{\delta\mathscr{H}_\sigma}{\delta\Omega_\sigma^\tau}\right|_\tau=\left.\frac{\delta\mathscr{H}_\sigma}{\delta\dot\Omega_\sigma^\tau}\right|_\tau=0 \tag{4.3}$$

or

$$\left.\frac{\partial D_{0_\sigma}}{\partial\Omega_\sigma^\tau}\right|_\tau=\left.\frac{\partial D_{0_\sigma}}{\partial\dot\Omega_\sigma^\tau}\right|_\tau=0\,.$$

From these equations we obtain

$$\Omega_\sigma^{\tau^2}(\tau)=\tfrac{1}{4}H_\sigma^2(\tau)(\mu_\sigma-\rho_\sigma^2)+\omega^2\,, \tag{4.4a}$$

$$\dot\Omega_\sigma^\tau(\tau)/\Omega_\sigma^\tau(\tau)=-\rho_\sigma H_\sigma\,. \tag{4.4b}$$

These are the constraints that assure the fulfillment of H2. They completely fix the Cauchy data. Now we can study the relation of H2 with the conditions coming from H1. In fact if (3.40f) and (4.4b) are fulfilled simultaneously we must have

TABLE I. Definitions of quantities found in Eq. (4.2).

σ	Field	Ω_σ	H	μ_σ	ρ_σ
T	transversal mode of the vector massive field; it coincides with the scalar field with conformal coupling ($\xi=1/6$)	it satisfies Eq. (2.24)	$H=\dot a/a$ Hubble coefficient	1	1
L	longitudinal mode of the massive vector field		$h=H(3-2m^2/\omega^2)$	1	1
ζ	scalar field with arbitrary coupling	it satisfies Eq. (2.15)	H	$3(3-16\xi)$	$3(1-4\xi)$

$$\rho_\sigma H_\sigma = H + \mathcal{O}(H^2)\,. \tag{4.5}$$

In the scalar case this relation only holds for $\xi = \frac{1}{6}$. It is also valid for the transversal mode of the massive vector case, but for the longitudinal one, we have

$$\rho_\sigma H_\sigma = (3 - 2m^2/\omega^2)H = H + \mathcal{O}(H^2)\,. \tag{4.6}$$

Thus both hypotheses can be satisfied only if $H = 0$. This problem with the massive spin-1 case has already been noticed by other authors (see Ref. 14).

We conclude that we have a good vacuum in the scalar case if $\xi = \frac{1}{6}$. In this case we can satisfy hypotheses H1 (ii) and H2. H1 (i) can also be satisfied in this case (see Ref. 9). Thus we have a satisfactory vacuum that fulfills both hypotheses H1 and H2. But in all other cases, the scalar field with $\xi \neq \frac{1}{6}$ and the massive vector field, we can only fulfill both hypotheses if $H = 0$ at least at the time when the vacuum is defined.

But let us continue our analysis of hypothesis H2, taking second derivatives of $\mathscr{H}_\sigma$ to see the role played by the coupling constant ξ. If we call

$$A = \frac{\partial^2 D_{0_J}}{\partial \Omega^2}, \quad B = \frac{\partial^2 D_{0_\sigma}}{\partial \Omega\, \partial \dot{\Omega}}, \quad C = \frac{\partial^2 D_{0_J}}{\partial \dot{\Omega}^2},$$

the necessary condition for the existence of an extremum is that

$$\Delta = AC - B^2 > 0\,. \tag{4.7}$$

Moreover if $A > 0$ we have a minimum and if $A < 0$ we have a maximum.

With our notation we obtain

$$\Delta = \Omega^{-4}; \tag{4.8}$$

thus because Ω is a real function, Δ is always positive and it satisfies Eq. (4.7); so we have an extremum.

Moreover, from (4.4) for the scalar field is $\Omega^2 = \omega^2 + 6\xi(1 - 6\xi)H^2$, therefore we must have $\xi < \frac{1}{6}$, to avoid imaginary values of Ω. If we choose the positive root of Ω, we have a minimum, because

$$A = [2\omega^2 + \tfrac{9}{2}H^2(1 - \tfrac{16}{3}\xi)]\Omega^{-3}\,, \tag{4.9}$$

then $A > 0$.

For the massive spin-1 case we can use the same equation to study the extremum, making $\xi = \frac{1}{6}$ for the transversal and longitudinal modes, but changing in the last case H by h (defined in Table I). Then

$$\Delta_{1/6} = \Delta_T = \Delta_L\,,$$

but

$$A_{1/6} = A_T \neq A_L\,, \quad A_L = 2(\omega^2 + h^2/4)\Omega^{-3}\,.$$

Then also for the longitudinal case, we have a minimum $\mathscr{H}_L$, if we choose the positive root of Ω_L.

Let us now study the physical meaning of the Hamiltonian minimization. As we have shown H2 is not always valid and we would like to know what we lose in that case. In flat space-time in the expectation value of the energy of a many-particle state a term $\Sigma_{\mathbf{k}} \omega_{\mathbf{k}} N_{\mathbf{k}}$ appears, where $\omega_{\mathbf{k}}$ is the energy of mode $\mathbf{k}$ and $N_{\mathbf{k}}$ is the particle number in this mode. Thus total energy turns out to be the sum of the energy of each particle. We eventually lose this interpretation because H2 could not be valid in a curved space-time. In fact, let us suppose that we want to retain the particle interpretation in curved space, then we must have

$$\langle 0,\tau | :T_{00}(\tau'):_{\tau'} | 0,\tau\rangle = \int \frac{dk^3}{(2\pi a)^3}\, \Omega_{\mathbf{k}}^{\tau'}(\tau') |\beta_{\mathbf{k}}^{\tau\tau'}(\tau')|^2\,, \tag{4.10}$$

where $\Omega_{\mathbf{k}}^{\tau'}(\tau')$ is some unknown coefficient that will be interpretated as the energy of the particle with k mode. Let us see what happens in the scalar case for arbitrary coupling. If we impose condition (4.10) to Eq. (3.10) we obtain

$$|\dot{\chi}_\xi^{(\tau')}(\tau')|^2 + 2\,\mathrm{Re}(\dot{\chi}_\xi^{\tau'}(\tau')\overset{*}{\chi}_\xi^{\tau'}(\tau'))(6\xi - 1)H + |\chi_\xi^{\tau'}(\tau')|^2[\omega^2 + (1 - 6\xi)H^2] = \Omega^{\tau'}(\tau')/a\,, \tag{4.11a}$$

$$[\dot{\chi}_\xi^{\tau'}(\tau') + \chi_\xi^{\tau'}(\tau')(6\xi - 1)H\,]^2 + \omega^2 \chi_\xi^{\tau'2}(\tau') = 0\,. \tag{4.11b}$$

With no loss of generality we can have

$$\chi_\xi^{\tau'}(\tau') = \frac{1}{\sqrt{2a\Omega^{\tau'}(\tau')}} \exp\left(i \int_{\tau_0}^{\tau'} \Omega^{\tau'}(t)dt\right), \tag{4.12a}$$

$$\dot{\chi}_\xi^{\tau'}(\tau') = i\psi_\xi(\tau')\chi_\xi^{\tau'}(\tau')\,. \tag{4.12b}$$

Then from Eq. (4.11b) we obtain

$$\psi_\xi(\tau')\omega + i(6\xi - 1)H(\tau')\,. \tag{4.13a}$$

Then written in terms of $\Omega^{\tau'}(\tau')$, and taking into account Eq. (4.12a), it is

$$\dot{\Omega}^{\tau'}(\tau')/\Omega^{\tau'}(\tau') = -3(1 - 4\xi)H(\tau')\,, \tag{4.13b}$$

where the factor that multiplies H is, in fact, ρ_ξ that we have previously defined. Replacing Eqs. (4.12) and (4.13) in Eq. (4.11a) we obtain

$$\Omega^{\tau'2}(\tau') = 6\xi(1 - 6\xi)H^2(\tau') + \omega^2(\tau')\,. \tag{4.13c}$$

Equations (4.13b) and (4.13c) are identical to Eq. (4.4) obtained from H2. Moreover the polarization term has the same form as in flat space-time:

$$\langle 0,\tau' | T_{00}(\tau',\mathbf{x}) | 0,\tau'\rangle = \frac{1}{2}\int \frac{dk^3}{(2\pi a)^3}\, \Omega^{\tau'}(\tau')\,. \tag{4.14}$$

Then, from Eqs. (2.6), (4.10), and (4.14) we have

$$\langle 0,\tau | T_{00}(\tau',\mathbf{x}) | 0,\tau\rangle = \frac{1}{2}\int \frac{dk^3}{(2\pi a)^3}\, \Omega_{\mathbf{k}}^{\tau'}(\tau') + \int \frac{dk^3}{(2\pi a)^3}\, \Omega_{\mathbf{k}}^{\tau'}(\tau') |\beta_{\mathbf{k}}^{\tau\tau'}(\tau')|^2\,. \tag{4.15}$$

We can identify $\Omega_{\mathbf{k}}^{\tau'}(\tau')$ as the natural generalization of the energy in mode $\mathbf{k}$.

For the massive spin-1 case the analogous equation to Eq. (4.10) is

$$\langle 0,\tau | :T_{00}(\tau',\mathbf{x}):_{\tau'} | 0,\tau\rangle = \sum_\sigma \int \frac{dk^3}{(2\pi a)^3}\, \Omega_\sigma^{\tau'}(\tau') |\beta_\sigma^{\tau\tau'}(\tau')|^2\,. \tag{4.16}$$

We reach the same result if we repeat the computation in this case. Then the last solution is valid if H2 holds and then we can reobtain the interpretation that the created energy is the sum of the energy of each particle.

All these facts show that H2 is a very natural hypothesis. If we try to make a theory without this hypothesis we shall have to deal with a theory with no particle interpretation.

Going back to the technical point of Sec. II A about the Hadamard vacua, we can ask the following question: Can the failure, in general, of the energy-minimization scheme to satisfy their finiteness criterion H1 be interpretated as saying that this scheme produces a two-point function of non-Hadamard form? The answer is no, because there is at least one example (cf. Ref. 9) where the minimization scheme yields a non-Hadamard vacuum with a finite renormalized energy-momentum tensor. We do not know if this is only a very peculiar case or if it really shows a general feature of the problem.

V. CONCLUSIONS

We have introduced a parametrization method that allows us to present the vacuum problem in curved space-time, simply and precisely.

Two natural but competitive hypotheses are introduced: H1, the fundamental hypothesis; and H2, the ground state hypothesis. We have shown that to fulfill these hypotheses the coefficients of Eqs. (2.17) must satisfy certain constraints. All of the problem is embodied in these coefficients and these constraints.

We have shown that the set of the two hypotheses H1 and H2 yields a set of incompatible constraints in the general case. This system of equations is only compatible in the scalar case when $\xi = \frac{1}{6}$ or in all cases when $H = 0$. We have also shown that both hypotheses are natural and necessary, thus the vacuum problem is stated for the general case.

We shall try to solve the problem or at least make a deeper study of it, using our formalism, in a forthcoming paper.

APPENDIX A: PROPERTIES OF THE OPERATOR : :$_\tau$

The classical expression for the energy-momentum tensor, valid for different matter fields, is

$$T_{\mu\nu} = \{D_1\varphi^\rho, D_2\varphi_\rho\}_{\mu\nu} \tag{A1}$$

(where the ρ index may in general be scalar, vectorial, or tensorial), and $\{\ ,\ \}$ is the anticommuter.

The operators D_1 and D_2 are linear differential operators that act on the φ field. For example, in the spin-0 case with arbitrary coupling ξ, they are

$$D_1\varphi = \{(\tfrac{1}{2} - \xi)\nabla_\mu, (\xi - \tfrac{1}{4})g_{\mu\nu}\nabla^\sigma, - \xi, \xi g_{\mu\nu}, \tfrac{1}{2}[\xi(R_{\mu\nu} - (R/2)\,g_{\mu\nu}) + \tfrac{1}{2}m^2 g_{\mu\nu}]\}\varphi\,, \tag{A2}$$

$$D_2\varphi = \{\nabla_\nu, \nabla_\sigma, \nabla_\nu\nabla_\mu, \Box, 1\}\varphi\,, \tag{A3}$$

with

$$\Box = -\nabla^\sigma\nabla_\sigma\,,$$

and, for the spin-1 field case, we have

$$D_1\varphi^\rho = \{H_{\mu\nu}{}^{\sigma\zeta\lambda\gamma}g_{\gamma\rho}\nabla_\xi, H_{\mu\nu}{}^{\sigma\zeta\lambda\gamma}g_{\zeta\rho}\nabla_\gamma, -H_{\mu\nu}{}^{\sigma\zeta\lambda\gamma}g_{\zeta\rho}\nabla_\gamma, -H_{\mu\nu}{}^{\sigma\zeta\lambda\gamma}g_{\gamma\rho}\nabla_\xi, m^2 M_{\mu\nu}{}^{\sigma\zeta}g_{\sigma\rho}\}\varphi^\rho\,, \tag{A4}$$

$$D_2\varphi_\rho = \{\delta_\lambda{}^\rho\nabla_\sigma, \delta_\sigma{}^\rho\nabla_\lambda, -\delta_\lambda{}^\rho\nabla_\sigma, -\delta_\sigma{}^\rho\nabla_\lambda, \delta_\zeta{}^\rho\}\varphi_\rho\,, \tag{A5}$$

with

$$H_{\mu\nu}{}^{\sigma\zeta\lambda\gamma} = \tfrac{1}{2}(\tfrac{1}{4}g_{\mu\nu}g^{\sigma\zeta} - g_\mu{}^\zeta g_\nu{}^\sigma)g^{\gamma\delta},$$
$$M_{\mu\nu}{}^{\sigma\zeta} = \tfrac{1}{2}(g_\mu{}^\sigma g_\nu{}^\zeta - \tfrac{1}{2}g_{\mu\nu}g^{\sigma\zeta})\,.$$

We drop the indices ρ, μ, ν because they are irrelevant for the demonstration that we shall make below.

From Eq. (A1) we can obtain the VEV replacing the quantized field

$$\varphi = \int dk^3(a_{\mathbf{k}}^\tau\phi_{\mathbf{k}}^\tau + a_{\mathbf{k}}^{\tau\dagger}\overset{*}{\phi}{}_{\mathbf{k}}^\tau) \tag{A6}$$

with

$$a_{\mathbf{k}}^\tau|0,\tau\rangle = 0,\quad a_{\mathbf{k}}^{\tau\dagger}|0,\tau\rangle = |1_{\mathbf{k}},\tau\rangle,\ \text{etc.},\quad \forall\mathbf{k}\,.$$

By the substitution in (A1) we set

$$\langle 0,\tau|T(\tau')|0,\tau\rangle = 2\int dk^3\,\mathrm{Re}[D_1\phi_{\mathbf{k}}^\tau(\tau')\cdot D_2\overset{*}{\phi}{}_{\mathbf{k}}^\tau(\tau')]\,. \tag{A7}$$

We can relate Eq. (A7) with the particle creation by means of Bogoliubov's transformation that goes from the basis $\{\phi^{\tau'}, \overset{*}{\phi}{}^{\tau'}\}$ to the basis $\{\phi^\tau, \overset{*}{\phi}{}^\tau\}$ (we omit the indices k and k'); and, for simplicity, we suppose that

$$\phi^\tau = \alpha^{\tau\tau'}\phi^{\tau'} + \beta^{\tau\tau'}\overset{*}{\phi}{}^{\tau'}\,, \tag{A8}$$

where $\alpha^{\tau\tau'}, \beta^{\tau\tau'}$ are matrices with elements $\alpha_{\mathbf{k}\mathbf{k}'}^{\tau\tau'}$ and $\beta_{\mathbf{k},\mathbf{k}'}^{\tau\tau'}$.

Replacing (A8) in (A7) we obtained

$$\langle 0,\tau|T(\tau')|0,\tau\rangle - \langle 0,\tau'|T(\tau')|0,\tau'\rangle = 4\int dk^3\{|\beta^{\tau\tau'}(\tau')|^2\,\mathrm{Re}[D_1\overset{*}{\phi}{}^{\tau'}(\tau')\cdot D_2\phi^{\tau'}(\tau')] + \mathrm{Re}[\alpha^{\tau\tau'}(\tau')\overset{*}{\beta}{}^{\tau\tau'}(\tau')D_1\phi^{\tau'}(\tau')\cdot D_2\phi^{\tau'}(\tau')]\}\,. \tag{A9}$$

We shall normal order Eq. (A9). We express the field φ in terms of the basis $\{\phi^{\tau'}, \overset{*}{\phi}{}^{\tau'}\}$, then

$$\varphi = \int dk^3(a_{\mathbf{k}}^{\tau'}\phi_{\mathbf{k}}^{\tau'}) + a_{\mathbf{k}}^{\tau'\dagger}\overset{*}{\phi}{}_{\mathbf{k}}^{\tau'}) \tag{A10}$$

with the corresponding operators defined by

$$a_{\mathbf{k}}^{\tau'}|0,\tau'\rangle = 0 \text{ and } a_{\mathbf{k}}^{\tau'\dagger}|0,\tau'\rangle = |1_{\mathbf{k}},\tau'\rangle,\quad \forall\mathbf{k}\,.$$

Replacing expression (A10) in (A1) and using the definition of normal ordering for the primed operators, dropping also the operators that produced orthogonal states, we have that the normal ordered-momentum operator is

$$:T:_{\tau'} = 2\int dk^3(a_{\mathbf{k}}^{\tau'}a_{\mathbf{k}}^{\tau'}D_2\phi_{\mathbf{k}}^{\tau'}\cdot D_1\phi_{\mathbf{k}}^{\tau'} + a_{\mathbf{k}}^{\tau'\dagger}a_{\mathbf{k}}^{\tau'\dagger}D_1\overset{*}{\phi}{}_{\mathbf{k}}^{\tau'}\cdot D_2\overset{*}{\phi}{}_{\mathbf{k}}^{\tau'} + 2a_{\mathbf{k}}^{\tau'\dagger}a_{\mathbf{k}}^{\tau'}\,\mathrm{Re}[D_1\phi_{\mathbf{k}}^{\tau'}\cdot D_2\overset{*}{\phi}{}_{\mathbf{k}}^{\tau'}])\,. \tag{A11}$$

Taking into account the transformation among the operators $\{a^\tau, a^{\tau\dagger}\}$ and $\{a^{\tau'}, a^{\tau'\dagger}\}$,

$$a^{\tau'} = \alpha^{\tau\tau'}a^\tau + \overset{*}{\beta}{}^{\tau\tau'}a^{\tau\dagger},\quad a^{\tau'\dagger} = \beta^{\tau\tau'}a^\tau + \overset{*}{\alpha}{}^{\tau\tau'}a^{\tau\dagger}, \tag{A12}$$

replacing (A12) in (A11) we obtain

$$\langle 0,\tau|:T:_{\tau}|0,\tau\rangle = 4\int dk^3\{|\beta^{\tau\tau'}|^2\,\mathrm{Re}[D_1\phi^{\tau}\cdot D_2\overset{*}{\phi}{}^{\tau'}] + \mathrm{Re}[\alpha\overset{*}{\beta}D_1\phi^{\tau}\cdot D_2\phi^{\tau'}]\}, \tag{A13}$$

then by comparison between (A13) and (A9) we obtain

$$\langle 0,\tau|:T(\tau'):_{\tau}|0,\tau\rangle = \langle 0,\tau|T(\tau')|0,\tau\rangle - \langle 0,\tau'|T(\tau')|0,\tau'\rangle . \tag{A14}$$

Now we shall show that $:T:_{\tau}$ is an operator that satisfies the first three Wald axioms.

(1) The off-diagonal elements must coincide with the expression without normalization.

Avoiding the unnecessary indices, the operator without normalization has the form

$$\begin{aligned} T = \int dk^3\,dk'^3[&a_{\mathbf{k}}^{\tau}a_{\mathbf{k}}^{\tau}(D_1\phi_{\mathbf{k}}^{\tau}\cdot D_2\phi_{\mathbf{k}}^{\tau} + D_2\phi_{\mathbf{k}}^{\tau}\cdot D_1\phi_{\mathbf{k}}^{\tau}) \\ &+ a_{\mathbf{k}}^{\tau}a_{\mathbf{k}'}^{\tau'\dagger}(D_1\phi_{\mathbf{k}}^{\tau}\cdot D_2\overset{*}{\phi}{}_{\mathbf{k}'}^{\tau'} + D_2\phi_{\mathbf{k}}^{\tau}\cdot D_1\overset{*}{\phi}{}_{\mathbf{k}'}^{\tau'}) \\ &+ a_{\mathbf{k}}^{\tau'\dagger}a_{\mathbf{k}'}^{\tau}(D_1\overset{*}{\phi}{}_{\mathbf{k}}^{\tau'}\cdot D_2\phi_{\mathbf{k}'}^{\tau} + D_2\overset{*}{\phi}{}_{\mathbf{k}}^{\tau'}\cdot D_1\phi_{\mathbf{k}}^{\tau}) \\ &+ a_{\mathbf{k}}^{\tau'\dagger}a_{\mathbf{k}'}^{\tau'\dagger}(D_1\overset{*}{\phi}{}_{\mathbf{k}}^{\tau'}\cdot D_2\overset{*}{\phi}{}_{\mathbf{k}'}^{\tau'} + D_2\overset{*}{\phi}{}_{\mathbf{k}'}^{\tau'}\cdot D_1\overset{*}{\phi}{}_{\mathbf{k}'}^{\tau'})] . \end{aligned} \tag{A15}$$

Let $\langle a|b\rangle = 0$; then we see that

$$\langle a|[a_{\mathbf{k}}^{\tau},a_{\mathbf{k}'}^{\tau'\dagger}]|b\rangle = \delta_{\mathbf{k}\mathbf{k}'}\langle a|b\rangle = 0$$

so

$$\langle a|a_{\mathbf{k}}^{\tau}a_{\mathbf{k}'}^{\tau'\dagger}|b\rangle = \langle a|a_{\mathbf{k}}^{\tau'\dagger}a_{\mathbf{k}'}^{\tau}|b\rangle . \tag{A16}$$

Thus from Eq. (A15)

$$\langle a|T|b\rangle = \langle a|:T:_{\tau}|b\rangle , \tag{A17}$$

which proves Axiom 1.

(2) In the Minkowskian limit we must reobtain the normal ordering of the flat space-time. This is true because $:\ :_{\tau}$ is the natural generalization of the normal ordering.

(3) The normal ordering must give us a conserved current. We can show this property using the expression (A14):

$$\nabla_{\mu}\langle 0,\tau|:T^{\mu}{}_{\nu}:_{\tau}|0,\tau\rangle = \langle 0,\tau|\nabla_{\mu}T^{\mu}{}_{\nu}|0,\tau\rangle - \langle 0,\tau'|\nabla_{\mu}T^{\mu}{}_{\nu}|0,\tau'\rangle = 0 . \tag{A18}$$

(4) Causality: In general, normal order does not satisfy the causality axiom.

APPENDIX B: CALCULATION OF A BASIS OF SOLUTIONS FOR SPIN 1

The massive spin-1 field equation in a Robertson–Walker metric, called Proca's equation, is

$$\partial^{\mu}\partial_{\mu}\phi_{\nu} - \partial_{\nu}\partial^{\mu}\phi_{\mu} + H(\partial_0\phi_{\nu} - \partial_{\nu}\phi_0) + m^2\phi_{\nu} = 0 \tag{B1}$$

(see Ref. 16).

We can separate in temporal and spatial variables by the following ansatz: We propose

$$\phi_{\mathbf{k}\mu} = f_{\mathbf{k}\mu}(t)e^{i\mathbf{k}\cdot\mathbf{x}} , \tag{B2}$$

where the $f_{\mathbf{k}\mu}$ are four vectorial functions of the time and the linear momentum $\mathbf{k}/a$. Replacing Eq. (B2) into Eq. (B1), we obtain

$$\omega_{\mathbf{k}}^2 f_{\mathbf{k}0} + (i/a^2)\mathbf{k}\cdot\mathbf{f}_{\mathbf{k}} = 0 , \tag{B3a}$$

$$\ddot{\mathbf{f}}_{\mathbf{k}} + H\dot{\mathbf{f}}_{\mathbf{k}} + \omega^2\mathbf{f}_{\mathbf{k}} - \mathbf{k}[(1/a^2)\mathbf{k}\cdot\mathbf{f}_{\mathbf{k}} + i(\dot{f}_{\mathbf{k}0} + Hf_{\mathbf{k}0})] = 0 , \tag{B3b}$$

where the $\mathbf{f}_{\mathbf{k}}$ are the spatial components of $f_{\mathbf{k}_\mu}$ and $\omega_{\mathbf{k}} = (k^2/a^2 + m^2)^{1/2}$ is the energy of the particle in the flat limit. From Eq. (B3) it is clear that we only need three vectors for the basis. Each element of the basis is identified by the subindex $\sigma = 1,2,3$. For the separation in transversal and longitudinal modes of polarization, we propose

$$\mathbf{f}_{\mathbf{k}\sigma}(t) = \mathbf{C}_{\mathbf{k}\sigma}f_{\mathbf{k}\sigma}(t) , \tag{B4}$$

where the time dependence is only in $f_{\mathbf{k}\sigma}$ and the trivectorial behavior in $\mathbf{C}_{\mathbf{k}\sigma}$.

For the decoupling of (B3), we choose one of the vectors parallel to $\mathbf{k}$ and the others two orthogonal to $\mathbf{k}$, i.e.,

$$\mathbf{C}_{\mathbf{k}3}\|\mathbf{k} \quad \text{and} \quad \mathbf{C}_{\mathbf{k}1},\mathbf{C}_{\mathbf{k}2}\perp\mathbf{k} . \tag{B5}$$

Replacing (B4) into (B3) we obtain

$$\mathbf{C}_{\mathbf{k}\sigma}(\ddot{f}_{\mathbf{k}\sigma} + H\dot{f}_{\mathbf{k}\sigma} + \omega^2 f_{\mathbf{k}\sigma}) - \frac{\mathbf{k}(\mathbf{k}\cdot\mathbf{C}_{\mathbf{k}\sigma})}{a^2\omega^2}\left[\ddot{f}_{\mathbf{k}\sigma} - \left(H + 2\frac{\dot{\omega}}{\omega}\right)\dot{f}_{\mathbf{k}\sigma} + \omega^2 f_{\mathbf{k}\sigma}\right] = 0 . \tag{B6}$$

Using (B5), Eqs. (B6) are equivalent to

$$\ddot{f}_{\mathbf{k}\sigma} + H_{\sigma}\dot{f}_{\mathbf{k}\sigma} + \omega_{\mathbf{k}}^2 f_{\mathbf{k}\sigma} = 0 , \tag{B7}$$

where $f_1 = f_2 \equiv f_{\sigma}$ (transversal polarization modes) with $H_T = H$ and $f_3 = f_L$ with $H_L = H(3 - 2m^2/\omega^2)$.

If we change the variables as

$$f_T = \frac{1}{(2\pi)^{3/2}}\chi_T , \quad f_L = \frac{\omega_{\mathbf{k}}}{(2\pi)^{3/2}m}\chi_L ,$$

we obtain

$$\frac{d^2\chi_{\sigma}}{d\eta^2} + k^2 p_{\sigma}\chi_{\sigma} = 0 , \tag{B8}$$

with p_{σ} given by Eq. (3.15).

The orthogonalization of the basis is done under the inner product

$$\langle\varphi^{\mu};\psi^{\nu}\rangle = i\int_{\Sigma}(\overset{*}{\varphi}{}^{\mu}\nabla_{\lambda}\psi_{\mu} - \psi^{\mu}\nabla_{\lambda}\overset{*}{\varphi}{}_{\mu})d\sigma^{\lambda} ,$$

with the requirement

$$\langle\phi_{\mathbf{k}\sigma}^{\mu};\phi_{\mathbf{k}'\sigma'}^{\nu}\rangle = \delta(\mathbf{k} - \mathbf{k}')\delta_{\sigma\sigma'} .$$

The result is the basis $\{\phi_{\mathbf{k}\sigma}^{\mu},\overset{*}{\phi}{}_{\mathbf{k}\sigma}^{\mu}\}$ with

$$\phi_{\mathbf{k}1}^{\mu}(\chi) = \frac{1}{(2\pi)^{3/2}\sqrt{k_1^2 + k_2^2}}\begin{pmatrix} 0 \\ k_2 \\ -k_1 \\ 0 \end{pmatrix}\chi_T e^{i\mathbf{k}\cdot\mathbf{x}} , \tag{B9a}$$

$$\phi^{\mu}_{k2}(\chi) = \frac{k_3}{(2\pi)^{3/2}\sqrt{k_1^2 + k_2^2}}$$

$$\times \begin{bmatrix} 0 \\ k_1 \\ k_2 \\ -(k_1^2 + k_2^2)/k_3 \end{bmatrix} \chi_T e^{i\mathbf{k}\cdot\mathbf{x}}, \tag{B9b}$$

$$\phi^{\mu}_{k3}(\chi) = \frac{\omega}{(2\pi)^{3/2} mk} \begin{pmatrix} k^2 W/a^2\omega^2 \\ k_1 \\ k_2 \\ k_3 \end{pmatrix} \chi_L e^{i\mathbf{k}\cdot\mathbf{x}}, \tag{B9c}$$

and

$$W = i\frac{d}{dt}\ln\left(\frac{\omega}{m}\chi_L\right),$$

where the function χ_σ is a solution of Eq. (3.16) (orthogonality) with the functional form (3.17). Due to the fact that the universe is spatially flat, there is an invariance under rotations in $\mathbb{R}^3$, so therefore we can obtain a simpler basis doing two successive rotations. Thus we finally have Eq. (3.14).

[1] N. D. Birell, and P. C. W. Davies, *Quantum Fields in Curved Space* (Cambridge U. P., London, 1982).
[2] T. H. Boyer, Sci. Am. **1985** (8), 56 (1985).
[3] H. B. G. Casimir, Proc. Kon. Ned. Akad. Wet. **51**, 793 (1948).
[4] R. M. Wald, Commun. Math. Phys. **45**, 9 (1975).
[5] R. M. Wald, Ann. Phys. **110**, 472 (1978).
[6] G. T. Horowitz, Phys. Rev. D **21**, 1445 (1980).
[7] M. Castagnino, Gen. Relativ. Gravit. **15**, 1149 (1983).
[8] M. Castagnino and F. D. Mazzitelli, Phys. Rev. D **31**, 742 (1985).
[9] M. Castagnino, D. D. Harari, and C. A. Nuñez, J. Math. Phys. **28**, 184 (1987).
[10] S. A. Fulling and L. Parker Ann. Phys. (NY) **87**, 176 (1974).
[11] G. Mamaev, V. M. Mostepanenko, and A. A. Starobinskii, Sov. Phys. JETP **13**, 823 (1976).
[12] F. W. J. Oliver, *Asymptotics and Special Functions* (Academic, New York, 1974).
[13] M. Castagnino, L. Chimento, and C. Laciana, Prog. Theor. Phys. **75**, 1142 (1986).
[14] A. A. Grib, and A. V. Nesteruk, Gen. Relativ. Gravit. **15**, 31 (1983).
[15] M. Castagnino, and R. Ferraro, Phys. Rev. D **34**, 497 (1986).
[16] M. Castagnino, L. Chimento, and C. Laciana, Prog. Theor. Phys. **71**, 309 (1984).

Infinite number of conserved quantities and extended conformal algebra in the Thirring model

Kazuhiko Inoue, Kazuhiko Odaka, and Minoru Omote
Institute of Physics, University of Tsukuba, Ibaraki 305, Japan

(Received 21 May 1987; accepted for publication 7 October 1987)

It is shown that the Thirring model has an infinite number of local conserved quantities, explicit forms of which are presented. These quantities are shown to be expressed in terms of scattering parameters. It will be shown that in this model there exists an extended symmetry algebra that includes the Virasoro algebra as its subalgebra.

I. INTRODUCTION

The inverse scattering method[1] has provided a unified framework for the investigation of classical completely integrable field theories. By this method we can systematically obtain an infinite number of conserved quantities that correspond to action variables of the system.

Based on these successes the quantum versions of the inverse scattering problem [the quantum inverse scattering method (QISM)] had been proposed in order to study completely integrable quantum field theories,[2] especially the nonlinear Schrödinger model. The QISM supports the results obtained by the Bethe ansatz approach to these models. It also gives creation and annihilation operators of the Bethe states.

In spite of these interesting results, the formulation of the QISM includes some troubles associated with the definition of local products of field operators in the quantum field theory. As is well known the naive definition of the local product of the field operator, which has been used in the QISM and in the Bethe ansatz approach, leads to contradictory results. For example, in the Bethe ansatz approach to the fermionic model the definition is inconsistent[3] since it does not give the nontrivial Schwinger terms of current commutation relations.[4] For another example it was shown that in the nonlinear Schrödinger model the eigenvalues of quantum operators corresponding to the action variables are divergent and then that we cannot obtain a quantum version of these conserved quantities.[5]

The problems of defining the local products of operators in quantum field theory have been studied perturbatively in various models,[6] but nonperturbatively only in the Thirring model. In the quantum theory of the Thirring model[7] it is shown that the local product of field operators, current operators in this case, is defined nonperturbatively and that this model can be solved completely in a well-defined Hilbert space.

These results suggest that in the Thirring model there is a hint of how to define operator products in the formulation of the inverse scattering problem. From this point of view, we think it is worthwhile to rediscuss the quantum theory of the Thirring model in terms of the inverse scattering method. To establish the quantum inverse scattering method in this model may help us to study other completely integrable quantum field theories in the framework of the QISM.

In the model, unfortunately, the classical version of the inverse scattering method has not been well discussed yet, and no one has found an infinite number of conserved quantities that are necessary to specify all states completely in the quantum theory.

The purpose of this paper is to formulate the classical inverse scattering method in the model. In Sec. II we show that this model has the Lax pair from which we can obtain explicit forms of the infinite number of conserved quantities. In Sec. III the scattering problem will be studied with this Lax pair and the conserved quantities are expressed in terms of the scattering parameters. The expression suggests that the scattering parameters may be related directly to the asymptotic field in the quantum theory. Finally, in Sec. IV we will show that the model has an extended symmetry algebra constructed from a large class of conserved quantities and that the algebra includes, as its subalgebra, the usual conformal algebra.[8]

The quantum theory of the Thirring model in terms of the scattering parameters will be discussed in a separate paper.[9]

II. THE LAX PAIR AND INFINITE NUMBER OF CONSERVED QUANTITIES

The Lagrangian density of the Thirring model[10] is given by

$$L = i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - (g/2)\bar{\psi}\gamma_{\mu}\psi\bar{\psi}\gamma^{\mu}\psi\,, \tag{2.1}$$

where $\bar{\psi} = \psi^{\dagger}\gamma^{0}$, $(x^{0},x^{1}) = (t,x)$, and the γ matrices are given by

$$\gamma^{0} = \gamma_{0} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \gamma^{1} = -\gamma_{1} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \gamma^{5} = \gamma^{0}\gamma^{1}\,. \tag{2.2}$$

In the classical theory, $\psi^{\dagger}$ and ψ fields satisfy the anticommutation relations

$$[\psi_{i}(xt),\psi_{j}(x't')]_{+} = [\psi_{i}^{\dagger}(xt),\psi_{j}^{\dagger}(x't')]_{+} = [\psi_{i}(xt),\psi_{j}^{\dagger}(x't')]_{+} = 0\,, \tag{2.3}$$

so they are odd elements of the Grassmann algebra.

In this model we find the Lax pair, the consistency condition of which gives the field equation. In a previous paper,[11] we have shown that the Lax pair of the system with fermionic fields and its scattering problem can be well formulated in the superspace (t,x,θ), where θ is the Grassman number $\theta^{2} = 0$.[12]

 0022-2488/88/020476-05$02.50

In the superspace formulation the Lax pair of the Thirring model is given by

$$\begin{aligned} i\partial_x \Phi^{(i)}(t,x,\theta) &= L^{(i)}\Phi^{(i)}(t,x,\theta) \\ i\partial_t \Phi^{(i)}(t,x,\theta) &= M^{(i)}\Phi^{(i)}(t,x,\theta) \end{aligned} \quad (i=1,2)\,, \tag{2.4}$$

where $\Phi^{(i)}(t,x,\theta)$ is a superfield and is expanded into the component fields $\xi^{(i)}(t,x)$ and $\eta^{(i)}(t,x)$ as

$$\Phi^{(i)}(t,x,\theta) = \xi^{(i)}(t,x) + \theta\eta^{(i)}(t,x)\,. \tag{2.5}$$

In (2.4) $L^{(i)}$ and $M^{(i)}$ are operators in the superspace and are defined by

$$\begin{aligned} L^{(i)} &= L_1^{(i)} + L_2^{(i)}\partial_\theta + \theta L_3^{(i)} + \theta L_4^{(i)}\partial_\theta \\ M^{(i)} &= M_1^{(i)} + M_2^{(i)}\partial_\theta + \theta M_3^{(i)} + \theta M_4^{(i)}\partial_\theta \end{aligned} \quad (i=1,2)\,, \tag{2.6}$$

where

$$\begin{aligned} &L_1^{(1)} = \lambda^2 + (g/2)\psi_2^\dagger\psi_2, \quad L_2^{(1)} = \lambda\psi_1\,, \\ &L_3^{(1)} = \lambda\psi_1^\dagger, \quad L_4^{(1)} = -2L_1^{(1)}\,, \\ &L_1^{(2)} = \lambda^2 - (g/2)\psi_1^\dagger\psi_1, \quad L_2^{(2)} = \lambda\psi_2\,, \\ &L_3^{(2)} = \lambda\psi_2^\dagger, \quad L_4^{(2)} = -2L_1^{(2)}\,, \end{aligned} \tag{2.7a}$$

and

$$\begin{aligned} &M_1^{(1)} = -\lambda^2 + (g/2)\psi_2^\dagger\psi_2, \quad M_2^{(1)} = -\lambda\psi_1\,, \\ &M_3^{(1)} = -\lambda\psi_1^\dagger, \quad M_4^{(1)} = -2M_1^{(1)}\,, \\ &M_1^{(2)} = \lambda^2 + (g/2)\psi_1^\dagger\psi_1, \quad M_2^{(2)} = \lambda\psi_2\,, \\ &M_3^{(2)} = \lambda\psi_2^\dagger, \quad M_4^{(2)} = -2M_1^{(2)}\,. \end{aligned} \tag{2.7b}$$

Here it should be noticed that component fields $\xi(t,x)$ and $\eta(t,x)$ have a character different from that of the Grassmann algebra. In the following we will assume that $\xi^{(i)}$ is an even element and $\eta^{(i)}$ is an odd element.

The integrability condition of $\xi^{(i)}$,

$$\partial_t(\partial_x \log\xi^{(i)}) - \partial_x(\partial_t \log\xi^{(i)}) = 0 \tag{2.8}$$

(which is equivalent to the field equation), leads to the equation of continuity

$$\partial_t(L_1^{(i)} + L_2^{(i)}\Omega^{(i)}) - \partial_x(M_1^{(i)} + M_2^{(i)}\Omega^{(i)}) = 0\,, \tag{2.9}$$

where $\Omega^{(i)}$ is defined by

$$\Omega^{(i)} = \eta^{(i)}/\xi^{(i)}\,, \tag{2.10}$$

and it is called the pseudopotential. Then conserved quantities $Q^{(i)}$ of the Thirring model can be defined by

$$Q^{(i)} = \int dx\{L_1^{(i)} + L_2^{(i)}\Omega^{(i)}\}\,. \tag{2.11}$$

From (2.4) we can further show that $\Omega^{(i)}$ satisfies

$$i\partial_x\Omega^{(i)} = L_3^{(i)} + L_4^{(i)}\Omega^{(i)}\,, \tag{2.12}$$

where the properties $(\Omega^{(i)})^2 = 0$ have been used. In completely integrable models with Bose fields, as is well known, pseudopotentials satisfy the Riccati-type equations that have nonlinear terms. In the Thirring model, on the other hand, the equation of the pseudopotentials is the linear one (2.12) because of the properties $(\Omega^{(i)})^2 = 0$, so we can find explicit expressions of all conserved quantities.

Now let us assume that $\Omega^{(i)}$ can be expanded into a power series of λ^{-1} such as

$$\Omega^{(i)} = \sum_{n=0}^{\infty}\omega_n^{(i)}\lambda^{-2n-1} + \sum_{n=0}^{\infty}\tilde{\omega}_n^{(i)}\lambda^{-2n}\,. \tag{2.13}$$

Substituting (2.13) into (2.12), we obtain the following recurrence formulas:

$$\begin{aligned} &\omega_0^{(1)} = \tfrac{1}{2}\psi_1^\dagger, \quad \omega_n^{(1)} = -\tfrac{1}{2}(i\partial_x + g\psi_2^\dagger\psi_2)\omega_{n-1}^{(1)}\,, \\ &\tilde{\omega}_0^{(1)} = 0, \quad \tilde{\omega}_n^{(1)} = -\tfrac{1}{2}(i\partial_x + g\psi_2^\dagger\psi_2)\tilde{\omega}_{n-1}^{(1)}\,, \end{aligned} \tag{2.14a}$$

$$\begin{aligned} &\omega_0^{(2)} = \tfrac{1}{2}\psi_2^\dagger, \quad \omega_n^{(2)} = -\tfrac{1}{2}(i\partial_x - g\psi_1^\dagger\psi_1)\omega_{n-1}^{(2)}\,, \\ &\tilde{\omega}_0^{(2)} = 0, \quad \tilde{\omega}_n^{(2)} = -\tfrac{1}{2}(i\partial_x - g\psi_1^\dagger\psi_1)\tilde{\omega}_{n-1}^{(2)}\,, \end{aligned} \quad n = 1,2,\ldots\,. \tag{2.14b}$$

The solutions of these recurrence formulas are

$$\begin{aligned} &\omega_n^{(1)} = -(-2)^{-n-1}(i\partial_x + g\psi_2^\dagger\psi_2)^n\psi_1^\dagger\,, \\ &\tilde{\omega}_n^{(1)} = 0, \end{aligned} \tag{2.15a}$$

and

$$\begin{aligned} &\omega_n^{(2)} = -(-2)^{-n-1}(i\partial_x - g\psi_1^\dagger\psi_1)^n\psi_2^\dagger\,, \\ &\tilde{\omega}_n^{(2)} = 0\,. \end{aligned} \tag{2.15b}$$

From (2.7), (2.11), (2.13), and (2.15) it is shown that the conserved quantities $Q^{(i)}$ are written in the power series of λ^{-1}, and that each coefficient of λ^{-2n} in $Q^{(i)}$ becomes a conserved quantity. Then we obtain an infinite number of conserved quantities:

$$I_n^{(1)} = \int_{-\infty}^{\infty} dx\,\psi_1^\dagger(i\overleftarrow{\partial}_x + g\psi_2^\dagger\psi_2)^n\psi_1\,, \tag{2.16a}$$

and

$$I_n^{(2)} = \int_{-\infty}^{\infty} dx\,\psi_2^\dagger(i\overleftarrow{\partial}_x - g\psi_1^\dagger\psi_1)^n\psi_2\,. \tag{2.16b}$$

Moreover conserved quantities are shown to satisfy the Poisson brackets

$$\{I_n^{(i)}, I_m^{(j)}\} = 0, \quad i,j = 1,2, \quad n,m = 1,2,\ldots\,. \tag{2.17}$$

Some conserved quantities among the $I_n^{(i)}$ have well-known physical meanings,

$$N = I_0^{(1)} + I_0^{(2)} = \int_{-\infty}^{\infty}\bar{\psi}\gamma^0\psi\,dx\,, \tag{2.18a}$$

$$N_A = I_0^{(1)} - I_0^{(2)} = \int_{-\infty}^{\infty}\bar{\psi}\gamma^0\gamma^5\psi\,dx\,, \tag{2.18b}$$

$$\begin{aligned} H &= I_1^{(1)} - I_1^{(2)} \\ &= \int_{-\infty}^{\infty}\left\{-i\bar{\psi}\gamma^1\partial_x\psi + \left(\frac{g}{2}\right)\bar{\psi}\gamma^\mu\psi\bar{\psi}\gamma_\mu\psi\right\}dx\,, \end{aligned} \tag{2.18c}$$

and

$$P = I_1^{(1)} + I_1^{(2)} = \int_{-\infty}^{\infty}\{-i\bar{\psi}\gamma^0\partial_x\psi\}dx\,, \tag{2.18d}$$

where N, N_A, H, and P are the fermion number, chiral

charge, Hamiltonian, and momentum of the system, respectively.

III. SCATTERING PARAMETERS AND CONSERVED QUANTITIES

In this section we will investigate the scattering problem of (2.4). Assuming that ψ and $\psi^\dagger$ vanish at $x = \pm\infty$, we will define the Jost functions $g_1^{(i)}(t,x,\theta,\lambda)$, $g_2^{(i)}(t,x,\theta,\lambda)$, $f_1^{(i)}(t,x,\theta,\lambda)$, and $f_2^{(i)}(t,x,\theta,\lambda)$ of solutions for (2.4) with the following boundary conditions:

$$x\to-\infty,\quad g_1^{(i)}(t,x,\theta,\lambda)\to e^{-i\lambda^2 x},\quad g_2^{(i)}(t,x,\theta,\lambda)\to\theta e^{i\lambda^2 x}, \tag{3.1a}$$

and

$$x\to\infty,\quad f_1^{(i)}(t,x,\theta,\lambda)\to e^{-i\lambda^2 x},\quad f_2^{(i)}(t,x,\theta,\lambda)\to\theta e^{i\lambda^2 x}. \tag{3.1b}$$

From the boundary conditions (3.1) we see that the Jost functions $g_1^{(i)}(t,x,\theta,\lambda)$ and $f_1^{(i)}(t,x,\theta,\lambda)$ [$g_2^{(i)}(t,x,\theta,\lambda)$ and $f_2^{(i)}(t,x,\theta,\lambda)$] have even (odd) characters.

Since $f_1^{(i)}$ and $f_2^{(i)}$ constitute a complete set of solutions for the scattering problem, $g_1^{(i)}$ can be expressed in terms of linear combinations of $f_1^{(i)}$ and $f_2^{(i)}$:

$$g_1^{(i)}(t,x,\theta,\lambda)=f_1^{(i)}(t,x,\theta,\lambda)a^{(i)}(\lambda)+f_2^{(i)}(t,x,\theta,\lambda)b^{(i)}(\lambda), \tag{3.2}$$

where the coefficients $a^{(i)}(\lambda)$ and $b^{(i)}(\lambda)$ are called scattering parameters. It can be shown that these scattering parameters do not depend on x. The scattering parameter $a^{(i)}(\lambda)(b^{(i)}(\lambda))$ is an even (odd) element of the Grassmann algebra.

They have the following properties: (i) the normalization condition,

$$a^{(i)}(\lambda)^* a^{(i)}(\lambda)+b^{(i)}(\lambda)^* b^{(i)}(\lambda)=1; \tag{3.3a}$$

(ii) analyticity, $a^{(i)}(\lambda)$ is analytic in the region with $\mathrm{Im}\,\lambda^2\geqslant 0$;

$$\text{(iii)}\quad a^{(i)}(\lambda)=a^{(i)}(-\lambda); \tag{3.3b}$$

and (iv) time dependence,

$$\frac{d}{dt}a^{(i)}(\lambda)=0. \tag{3.3c}$$

The property (iv) suggests that $a^{(i)}(\lambda)$ can be expressed in terms of conserved quantities $I_n^{(i)}$ given by (2.16) and vice versa. In the following we will present these expressions.

From the boundary condition (3.1) and (3.2), $\log a^{(i)}(\lambda)$ is expressed as follows:

$$\begin{aligned}\log a^{(i)}(\lambda)&=\lim_{x\to\infty}\log\{(g_1^{(i)}(t,x,\theta,\lambda))_e e^{i\lambda^2 x}\}\\&=-i\int_{-\infty}^{\infty} i\,\partial_x\log\{(g_1^{(i)}(t,x,\theta,\lambda))_e e^{i\lambda^2 x}\}dx,\end{aligned} \tag{3.4}$$

where $(g_1^{(i)})_e$ means the component with even character given by

$$(g_1^{(i)}(x,t,\theta,\lambda))_e=\int d\theta\,\theta g_1^{(i)}(x,t,\theta,\lambda). \tag{3.5}$$

Since $(g_1^{(i)})_e$ satisfies the integrability condition (2.8), we can rewrite (3.4), by tracing the same procedure from (2.8) to (2.15), in the form

$$\begin{aligned}\log a^{(i)}(\lambda)&=-i\int_{-\infty}^{\infty}(L_1^{(i)}-\lambda^2)dx-i\sum_{n=0}^{\infty}(-2)^{-n-1}\lambda^{-2n}I_n^{(i)}\\&=\begin{cases}-\dfrac{i}{2}(I_0^{(1)}+gI_0^{(2)})-i\displaystyle\sum_{n=0}^{\infty}(-2)^{-n-1}\lambda^{-2n}I_n^{(1)}, & i=1,\\ -\dfrac{i}{2}(I_0^{(2)}-gI_0^{(1)})-i\displaystyle\sum_{n=0}^{\infty}(-2)^{-n-1}\lambda^{-2n}I_n^{(2)}, & i=2.\end{cases}\end{aligned} \tag{3.6}$$

Next, we will express $I_n^{(i)}$ in terms of the scattering parameters conversely. For the sake of convenience, new scattering parameters $\tilde a^{(i)}(\lambda)$ are introduced by

$$\begin{aligned}\tilde a^{(1)}(\lambda)&=\exp[(i/2)(I_0^{(1)}+gI_0^{(2)})]a^{(1)}(\lambda),\\ \tilde a^{(2)}(\lambda)&=\exp[(i/2)(I_0^{(1)}-gI_0^{(2)})]a^{(2)}(\lambda).\end{aligned} \tag{3.7}$$

It is easily shown that $\tilde a^{(i)}(\lambda)$ satisfies

$$\tilde a^{(i)}(\lambda)^*\tilde a^{(i)}(\lambda)+b^{(i)}(\lambda)^* b^{(i)}(\lambda)=1, \tag{3.8}$$

and, from (3.6) and (3.7),

$$\lim_{\substack{|\lambda|\to 0\\ \mathrm{Im}\,\lambda^2>0}}\tilde a^{(1)}(\lambda)=1. \tag{3.9}$$

Since $\tilde a^{(i)}(\lambda)$ as well as $a^{(i)}(\lambda)$ is analytic in the region with $\mathrm{Im}\,\lambda^2\geqslant 0$, we find, with the help of the dispersion relation,

$$\log\tilde a^{(i)}(\lambda^2)=\frac{1}{i\pi}\int_{-\infty}^{\infty}\frac{\log|\tilde a^{(i)}(k^2)|}{k^2-\lambda^2-i0}dk^2, \tag{3.10}$$

for real λ^2. By expanding the right-hand side of (3.10) into the power series of λ^{-2}, we have

$$\log\tilde a^{(i)}(\lambda^2)=\sum_{n=0}^{\infty}\frac{1}{\lambda^{2n}}C_n^{(i)}, \tag{3.11}$$

where $C_n^{(i)}$ is given by

$$C_n^{(i)}=-\frac{1}{i\pi}\int_{-\infty}^{\infty}k^{2n-2}\log|\tilde a^{(i)}(k^2)|dk^2,\quad n\geqslant 1. \tag{3.12}$$

Then, from (3.8), (3.11), and (3.12) we arrive at the expression

$$I_n^{(i)}=-\frac{(-2)^{n+1}}{\pi}\int_{-\infty}^{\infty}k^{2n-2}\log|\tilde a^{(i)}(k^2)|dk^2. \tag{3.13}$$

Moreover, since the normalization condition leads to

$$\begin{aligned}|\tilde a^{(i)}(\lambda^2)|&=\sqrt{1-b^{(i)}(\lambda)^* b^{(i)}(\lambda)}\\&=(1-\tfrac{1}{2}b^{(i)}(\lambda)^* b^{i}(\lambda)),\end{aligned} \tag{3.14}$$

it is shown that

$$\log|\tilde{a}^{(i)}(\lambda^2)| = \log(1 - \tfrac{1}{2}b^{(i)}(\lambda)^* b^{(i)}(\lambda)) = -\tfrac{1}{2}b^{(i)}(\lambda)^* b^{(i)}(\lambda)\,, \tag{3.15}$$

where we have used the relation $b^{(i)}(\lambda)^2 = b^{*(i)}(\lambda)^2 = 0$. Here new scattering parameters are defined by $\tilde{b}^{(i)}(\lambda) = \lambda^{-1}b^{(i)}(\lambda)$. Then we get an interesting result that the infinite number of conserved quantities $I_n^{(i)}$ can be represented in terms of the scattering parameters $\tilde{b}^{(i)}(\lambda)$ with odd character;

$$I_n^{(i)} = -\frac{(-2)^n}{\pi}\int_{-\infty}^{\infty} k^{2n}\tilde{b}^{(i)}(k)^*\tilde{b}^{(i)}(k)dk^2\,. \tag{3.16}$$

Expression (3.16) suggests that the scattering parameters $\tilde{b}^{(i)}(\lambda)$ and $\tilde{b}^{(i)\dagger}(\lambda)$ play the role of fundamental operators in the quantum theory of the Thirring model. We will discuss these problems in a separate paper.

IV. EXTENDED CONFORMAL ALGEBRA

It is well known that in two-dimensional conformal invariant field theories there exists the symmetric energy-momentum tensors θ^{--} and θ^{++} that satisfy

$$\partial_+\theta^{++} = 0, \quad \partial_-\theta^{--} = 0\,, \tag{4.1}$$

where $x^{\pm} = t \pm x$, $\partial_{\pm} = \partial/\partial x^{\pm}$, and

$$\theta^{++} = \tfrac{1}{2}(\theta^{00} + \theta^{01} + \theta^{10} + \theta^{11})\,,$$
$$\theta^{--} = \tfrac{1}{2}(\theta^{00} - \theta^{01} - \theta^{10} + \theta^{11})\,. \tag{4.2}$$

Equations (4.1) mean that θ^{++} (θ^{--}) is a function of x^- (x^+) only.

Furthermore, it can be shown that the moments of these energy-momentum tensors, defined by

$$L_{1-n}^{(+)} = \int (x^-)^n\theta^{++}(x^-)dx\,,$$
$$L_{1-n}^{(-)} = \int (x^+)^n\theta^{--}(x^+)dx\,, \tag{4.3}$$

are constants of motion and that they constitute the Virasoro algebra without the central charge

$$\{L_n^{(\pm)}, L_m^{(\pm)}\} = (m-n)L_{n+m}^{(\pm)}\,,$$
$$\{L_n^{(+)}, L_m^{(-)}\} = 0\,. \tag{4.4}$$

Here we have to notice that Eqs. (4.1) and (4.4) can result from the conformal invariance, and that these equations are shared by many conformal invariant field theories in two dimensions.

The energy-momentum tensor θ^{++}, θ^{--} of the Thirring model is given by

$$\theta^{++} = i\psi_1^\dagger \overleftrightarrow{\partial}_-\psi_1, \quad \theta^{--} = i\psi_2^\dagger \overleftrightarrow{\partial}_+\psi_2\,, \tag{4.5}$$

and the moments of θ^{++} and θ^{--} are obtained by substituting (4.5) into (4.3). Since the Thirring model is invariant under conformal transformations, θ^{++} and θ^{--}, given by (4.5), satisfy the continuity equations (4.1). Then their moments L_n^+ and L_n^- constitute an infinite number of conserved quantities and satisfy the Virasoro algebra.

On the other hand, in Sec. II we have shown that there exists another set of the infinite number of conserved quantities $I_n^{(i)}$ that satisfy

$$\{I_n^{(i)}, I_m^{(j)}\} = 0\,. \tag{4.6}$$

Thus we find that the Thirring model has two kinds of sets constructed of conserved quantities $\{L_n;\ n = 0, \pm 1, \pm 2,...\}$ and $\{I_n^{(i)};\ n = 0,1,2,...\}$, the algebras of which are given by (4.4) and (4.6), respectively.

In the following we will show that in this model there is an extended symmetry algebra that includes (4.4) and (4.6) as its subalgebras. In order to do so let us return to the continuity equation (2.9). Making use of (2.12) and (2.13) the coefficients in the power series of λ^{-1} for (2.9) are shown to be the equations

$$\partial_+\theta_n^{(1)} = 0, \quad \partial_-\theta_n^{(2)} = 0, \quad n = 0,1,2,..., \tag{4.7}$$

where

$$\theta_n^{(1)} = \psi_1^\dagger(i\overleftarrow{\partial}_x + g\psi_2^\dagger\psi_2)^n\psi_1\,,$$
$$\theta_n^{(2)} = \psi_2^\dagger(i\overleftarrow{\partial}_x - g\psi_1\psi_1)^n\psi_2\,. \tag{4.8}$$

Equations (4.7) mean that $\theta_n^1(\theta_n^2)$ is the function of x^- (x^+) only.

Furthermore, the field equations of ψ and $\psi^\dagger$ yield

$$-i\psi_1^\dagger\overleftarrow{\partial}_- = \psi_1^\dagger(i\overleftarrow{\partial}_x + g\psi_2^\dagger\psi_2)$$

and (4.9)

$$i\psi_2^\dagger\overleftarrow{\partial}_+ = \psi_2^\dagger(i\overleftarrow{\partial}_x - g\psi_1^\dagger\psi_1)\,.$$

Using the method of mathematical induction, we obtain, from (4.9),

$$\psi_1^\dagger(-i\overleftarrow{\partial}_-)^n = \psi_1^\dagger(i\overleftarrow{\partial}_x + g\psi_2\psi_2)^n$$

and (4.10)

$$\psi_2^\dagger(i\overleftarrow{\partial}_+)^n = \psi_2^\dagger(i\overleftarrow{\partial}_x - g\psi_1^\dagger\psi_1)^n\,.$$

Hence substituting (4.10) into (4.8), θ_n^1 and θ_n^2 can be expressed in the simple forms

$$\theta_n^{(1)} = \psi_1^\dagger(-i\overleftarrow{\partial}_-)^n\psi_1, \quad \theta_n^{(2)} = \psi_2^\dagger(i\overleftarrow{\partial}_+)^n\psi_2\,. \tag{4.11}$$

It is obvious that $\theta_1^{(1)}$ and $\theta_1^{(2)}$ reduce to θ^{++} and θ^{--}, respectively:

$$\theta_1^{(1)} = -\theta^{++}, \quad \theta_1^{(2)} = \theta^{--}\,. \tag{4.12}$$

Thus we have shown that in the Thirring model there are two sets of generalized densities $\{\theta_n^{(1)}(x^-)\}$ and $\{\theta_n^{(2)}(x^+)\}$, $n = 0,1,2,...$, which are associated with the complete integrability.

Now we will introduce a higher-rank tensor density given by

$$\theta^{\mu\nu\lambda\cdots\sigma} = \bar{\psi}\gamma^\mu\partial\!\!\!/\gamma^\nu\partial\!\!\!/\gamma^\lambda\partial\!\!\!/\cdots\partial\!\!\!/\gamma^\sigma\psi, \quad \partial\!\!\!/ = \gamma^\alpha\partial_\alpha\,, \tag{4.13}$$

which is symmetric with respect to any pair of suffixes and is traceless $\theta^{\mu\lambda\cdots}{}_{\mu\cdots\sigma} = 0$. It is easy to see that this tensor satisfies the continuity equation

$$\partial_\mu\theta^{\alpha\beta\cdots\mu\cdots\sigma} = 0\,, \tag{4.14}$$

and that $\theta_n^{(1)}$ and $\theta_n^{(2)}$ are given in terms of linear combinations of $\theta^{\mu\nu\cdots\lambda}$ such as

$$\theta^{+++\cdots+} = \bar{\psi}\gamma^+\partial\!\!\!/\gamma^+\partial\!\!\!/\gamma^+\cdots\partial\!\!\!/\gamma^+\psi,$$
$$\theta^{---\cdots-} = \bar{\psi}\gamma^-\partial\!\!\!/\gamma^-\partial\!\!\!/\cdots\partial\!\!\!/\gamma^-\psi\,, \tag{4.15}$$

where $\gamma^{\pm}=\gamma^0\pm\gamma^1$.

Next let us define moments of generalized densities $\theta_n^{(1)}(x^-)$ and $\theta_n^{(2)}(x^+)$ by

$$M_{n,1-l}^{(1)}=\int dx^-(x^-)^l\theta_n^{(1)}(x^+)\,,$$
$$M_{n,1-l}^{(2)}=\int dx^+(x^+)^l\theta_n^{(2)}(x^+)\,. \tag{4.16}$$

These moments include $L_l^{\pm}$ and I_n as its subsets,

$$L_l^{+}=-M_{1,l}^{(1)},\quad L_l^{-}=M_{1,l}^{(2)} \tag{4.17}$$

and

$$I_n^{(1)}=M_{n,1}^{(1)},\quad I_n^{(2)}+M_{n,1}^{(2)}\,. \tag{4.18}$$

Moreover, it can be shown that $M_{n,1-l}^{(i)}$ $(i=1,2)$ are constants of motions and that they satisfy the Poisson brackets

$$\begin{aligned}\{M_{n,l}^{(i)},M_{m,k}^{(j)}\}=\delta_{ij}\Bigg[&-i\sum_{r=0}^{m}(-1)^{jr}(i)^r\\&\times\binom{m}{r}\binom{1-l}{r}r!M_{n+m-r,k+l+r-1}^{(j)}\\&+i\sum_{r=0}^{n}(-1)^{jr}(i)^r\binom{n}{r}\binom{1-k}{r}r!\\&\times M_{n+m-r,k+l+r-1}^{(j)}\Bigg],\end{aligned} \tag{4.19}$$

where we have used the canonical Poisson brackets

$$\begin{aligned}&\{\psi_i(x),\psi_j(x')\}=0,\quad\{\psi_i^{\dagger}(x),\Psi_j^{\dagger}(x')\}=0\,,\\&\{\psi_i^{\dagger}(x),\psi_j(x')\}=-i\delta_{ij}\delta(x-x')\,.\end{aligned} \tag{4.20}$$

Then the sets of $\{M_{n,1-l}^{(i)}\}$ constitute the extended algebra that we call the extended conformal algebra. The Virasoro algebra (4.4) and the algebra (4.6) are the subalgebra of the extended conformal algebra.

Whereas the conformal algebra is associated with the space-time symmetry, it is not obvious what the physical meaning of the extended conformal algebra is. In spite of the obscurity for its physical meanings, we have to stress that the existence of the extended algebra is the characteristic feature resulting from the fact that the Thirring model has properties of both conformal invariance and complete integrability.

[1]For example, A. C. Newell, *Solitons in Mathematics and Physics* (SIAM, Philadelphia, 1985).

[2]H. B. Thacker, Rev. Mod. Phys. **53**, 253 (1982); P. P. Kulish and E. K. Sklyanin, *Proceedings of the Tvärminne Symposium on Integrable Quantum field Theories*, 1981, edited by J. Hieterinta and C. Montonen (Springer, New York, 1982); L. D. Faddeev, *Les Houches XXXIX*, 1982, edited by J. Zuber and R. Stora (North-Holland, Amsterdam, 1984).

[3]K. Odaka and S. Tokitake, J. Phys. Soc. Jpn. **56**, 3062 (1987); K. Johnson, Nuovo Cimento **20**, 773 (1961).

[4]T. Goto and T. Imamura, Prog. Theor. Phys. **14**, 396 (1955); J. Schwinger, Phys. Rev. Lett. **3**, 296 (1959).

[5]M. Omote, M. Sakagami, R. Sasaki, and I. Yamanaka, Phys. Rev. D **35**, 2423 (1987); T. Itoh (private communication).

[6]W. Zimmermann, *Lectures on Elementary Particles and Quantum Field Theory*, edited by S. Deser, M. Grisaru, and H. Pendleton (MIT, Cambridge, 1970).

[7]B. Klaiber, *Lectures in Theoretical Physics*, edited by A. Barut and W. Britten (Gordon and Breach, New York, 1968), Vol. XA, and references cited therein.

[8]S. Ferrara, R. Gatto, and A. F. Grillo, Nuovo Cimento A **12**, 959 (1972).

[9]K. Inoue, T. Itoh, K. Odaka, and M. Omote, in preparation.

[10]W. E. Thirring, Ann. Phys. (NY) **3**, 91 (1958).

[11]M. Omote and K. Inoue, Prog. Theor. Phys. **72**, 641 (1984); Phys. Lett. B **147**, 317 (1984).

[12]For the case that ψ and $\psi^{\dagger}$ are regarded as complex functions, the Lax pair had been found by M. Wadati, J. Phys. Soc. Jpn. **52**, 1084 (1983).

Spinors in periodic self-dual gauge field backgrounds

A. Chakrabarti
Centre de Physique Théorique de l'Ecole Polytechnique, LP N° 14 du Centre National du la Recherche Scientifique, 91128-Palaiseau Cedex, France

(Received 3 June 1987; accepted for publication 14 October 1987)

Normalizable zero modes of the Dirac operator are constructed for a class of self-dual, periodic SU(2) gauge field backgrounds characterized by two independent integer invariants. The integers are $(S_T/8\pi^2)$, where S_T is the action over one period T, and the asymptotic winding number (q) in R_3, the solutions reducing to static "monopoles" for large spatial distances independently of the time. The spinor solutions are obtained for the simplest class of the hierarchy presented in Chakrabarti [Phys. Rev. D **35**, 696 (1987)], corresponding to $q = 1$ and $S_T = 8\pi^2 \cdot 2n$ $(n = 1,2,3,...)$. The full number of zero modes for such backgrounds is $((S_T/8\pi^2) - q) = (2n - 1)$. They are all constructed explicitly. It is shown how these results can be obtained through a simple scaling limit by starting with special classes of instantons with finite action over R_4. A derivation of S_T is also given.

I. INTRODUCTION

The zero modes of Dirac spinors (for isospin $\frac{1}{2}$) are constructed for periodic self-dual Yang–Mills fields characterized by *two* integer topological indices.

Let us first recall the significance of such zero modes for standard instanton backgrounds with *one* index only. They have been studied by various authors. There are important reasons for doing that. In calculating quantum fluctuations their effect has to be separated out. Otherwise the fermionic determinant vanishes. The pioneering work of 't Hooft[1] studies this aspect for one instanton. Results and references for multi-instanton backgrounds can be found in Ref. 2. The structure of the instanton fields themselves can be studied via the spinors. The inverse construction of the ADHM solutions using spinors[3] furnishes a proof of the completeness of such instanton solutions. Linear classical fluctuations around instantons and their parameter space can be related[4] to Dirac spinors of isospin 1. Yang–Mill fields can be formulated[5] in terms of multispinors, helpful in the study of the zero modes in question. (Anomalies and spinor zero modes have been studied[6] also for other types of backgrounds in different dimensions for different types of manifolds and boundary conditions.)

The index of the instanton background fixes the number of spinor zero modes. Thus for an index n and isospin $\frac{1}{2}$ one has exactly n Dirac zero modes. *What happens to spinors when the background is characterized by two topological integers?* We discuss and illustrate this for a simple case through explicit construction. Both indices are found to play a role in fixing the number of zero modes. The next step should be to study carefully the implications of the new situation concerning anomalies, fluctuations, and reconstruction of the gauge fields in terms of the spinor solutions. I intend to study these aspects elsewhere.

In a previous paper[7] I have presented self-dual, periodic SU(2) gauge field solutions. A method was indicated for constructing a canonical hierarchy. Apart from periodicity and finite action over one period the technique leads to the following special feature. For large spatial distances (asymptotically in R_3 and for any time) the solutions reduce to Euclidean versions of static Bogomolny–Prasad–Sommerfield (BPS) monopoles. The iteration of poles[7] increases, at each step, this asymptotic charge (say q) by unity. The construction gives *for each value of* q,

$$q = 1,2,3,...,$$

an *infinite sequence* of periodic solutions. Different members of each sequence have different types of pulsating core but reduce to the same static form asymptotically. The simplest type is briefly described in the Appendix. For this class (S_T denoting the action over one period T),

$$q = 1, \quad S_T = 8\pi^2 \cdot 2n \quad (n = 1,2,...). \tag{1.1}$$

Section II gives explicitly the normalizable zero modes of the Dirac operator for such a background. It turns out that

$$\text{number of zero modes} = 2n - 1 = S_T/8\pi^2 - 1. \tag{1.2}$$

I have learned from Hitchin[8] that, for the above-mentioned asymptotic conditions, one should have more generally (for arbitrary integer q)

$$\text{index of the Dirac operator} = S_T/8\pi^2 - q. \tag{1.3}$$

The second term $(-q)$ comes from the boundary $S_2 \times S_1$ as the three-distance $r \to \infty$. Thus the explicit construction of Sec. II succeeds in illustrating (for $q = 1$) the combined effect of the *two independent integer invariants* of such gauge field configurations. The two invariants were already discussed by Gross, Pisarski, and Yaffe.[9] Generalization of the construction of the Appendix following Ref. 7 will yield periodic fields with

$$q = 1,2,..., \quad (S_T/8\pi^2) = q \cdot 2n \quad (n = 1,2,...) \tag{1.4}$$

[see the remarks following (A11) concerning a modified choice of periodic functions]. But here spinor solutions are obtained only for $q = 1$. Even this would have been difficult in the absence of the very convenient forms[10] (2.18) and (2.19). The rest of Sec. II is devoted to the delicate task of selecting, in (2.19), suitable f's for a given g to assure regularity and normalizability.

In a series of papers (cited for example, in Ref. 7) I have shown how multicharged BPS monopoles can be obtained very simply as scaling limits of sequences of instantons. This

 0022-2488/88/020481-06$02.50

is again possible for the periodic sequences (1.4) and for the corresponding spinor zero modes. This is briefly indicated in Sec. III.

In the following sections I have tried to emphasize how the relation (1.2) arises inevitably from our constructions.

II. ZERO MODES OF THE DIRAC OPERATOR

The gauge field background is given by

$$A_t = (\partial_r \zeta)\sigma_3/2, \quad A_r = -(\partial_t \zeta)\sigma_3/2,$$
$$A_\theta = -e^{\zeta}\sigma_2/2, \quad A_\varphi = \sin\theta\, e^{\zeta}\sigma_1/2 - \cos\theta\,\sigma_3/2, \tag{2.1}$$

where

$$e^{\zeta} = r(1-g\bar{g})^{-1}\left(\frac{dg}{dz}\frac{d\bar{g}}{d\bar{z}}\right)^{1/2} \quad (2z = r+it) \tag{2.2}$$

and $g(z)$ is a holomorphic function.

Choosing

$$g(z) = \prod_{j=1}^{n}\left(\frac{a_j + e^{-2kz}}{\bar{a}_j + e^{2kz}}\right) \quad (k>0, \quad |a_j|<1), \tag{2.3}$$

one gets a periodic self-dual solution with a finite action per period T ($= 2\pi/k$), namely,

$$S_T = 4\pi T(2nk) = 8\pi^2\cdot 2n \quad (0\leqslant t\leqslant T) \tag{2.4}$$

(see the Appendix). This is to be compared with the choice (Witten[11])

$$g = \prod_{j=1}^{n}\left(\frac{bj-2z}{\bar{b}_j+2z}\right) \quad (b_j+\bar{b}_j>0), \tag{2.5}$$

giving a finite action over the entire R_4 ($t\in[-\infty,\infty]$),

$$S = 8\pi^2(n-1). \tag{2.6}$$

The choice (2.3) can be generalized to include factors with different periods, if they are integer multiples of a basic one. Only the simple form (2.3) will be considered here. And since then one can go over, through a rescaling, to the case $k=1$, only this last case will be essentially our concern hereafter. [Note that in (2.4) S_T is independent of k.]

Consider the zero-mass Dirac equation in the background (2.1). It is convenient to use the spherical Euclidean coordinates from the beginning. Let

$$ds^2 = dt^2 + dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\varphi^2). \tag{2.7}$$

Define

$$\hat{\gamma}_t = \gamma_0, \quad \hat{\gamma}_r = \gamma_3, \quad \hat{\gamma}_\theta = r\gamma_1, \quad \hat{\gamma}_\varphi = r\sin\theta\,\gamma_2, \tag{2.8}$$

where

$$\gamma_0 = \begin{vmatrix} & \sigma_0 \\ \sigma_0 & \end{vmatrix}, \quad \gamma_j = \begin{vmatrix} & -i\sigma_j \\ i\sigma_j & \end{vmatrix} \quad (j=1,2,3),$$

and

$$B_\mu = (i/4)[\hat{\gamma}^\nu, \partial_\nu \hat{\gamma}_\mu],$$

giving

$$B_t = 0 = B_r, \quad B_\theta = (i/2)\gamma_3\gamma_1,$$
$$B_\varphi = (i/2)(\sin\theta\,\gamma_3 + \cos\theta\,\gamma_1)\gamma_2. \tag{2.9}$$

The isospin-$\frac{1}{2}$ Dirac equation is

$$\hat{\gamma}^\mu(i\partial_\mu - B_\mu - A_\mu)\Psi = 0. \tag{2.10}$$

Here A_μ corresponds to (2.1) and B_μ supplies the spin connections corresponding to (2.7). Our gauge fields are self-dual with definitions

$$\epsilon_{tr\theta\varphi} = 1, \quad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + i[A_\mu, A_\nu], \tag{2.11}$$

so that defining

$$\Psi = \begin{vmatrix} \Psi_U \\ \Psi_L \end{vmatrix}, \tag{2.12}$$

with upper (U) and lower (L) helicity indices, only Ψ_U has normalizable zero modes.

Let

$$\Psi_U = \begin{vmatrix} \Psi_+ \\ \Psi_- \end{vmatrix}, \tag{2.13}$$

where $\pm$ are isospin indices ($I_3 = \pm\frac{1}{2}$).

Defining

$$\Psi_+ = r^{-1}e^{\zeta/2}\begin{vmatrix} 0 \\ h_1(r,t) \end{vmatrix}, \quad \Psi_- = r^{-1}e^{\zeta/2}\begin{vmatrix} h_2(r,t) \\ 0 \end{vmatrix}, \tag{2.14}$$

the Dirac equation can be shown finally to reduce to

$$\partial_{\bar{z}} h_1 = -(1-g\bar{g})^{-1}\left(\frac{dg}{dz}\frac{d\bar{g}}{d\bar{z}}\right)^{1/2} h_2,$$
$$\partial_z h_2 = -(1-g\bar{g})^{-1}\left(\frac{dg}{dz}\frac{d\bar{g}}{d\bar{z}}\right)^{1/2} h_1. \tag{2.15}$$

Defining

$$\tilde{h}_1 = \left(\frac{dg}{dz}\right)^{-1/2} h_1, \quad \tilde{h}_2 = \left(\frac{d\bar{g}}{d\bar{z}}\right)^{-1/2} h_2, \tag{2.16}$$

and changing variables from $(z,\bar{z})$ to $(g,\bar{g})$,

$$\partial_{\bar{g}}\tilde{h}_1 = -(1-g\bar{g})^{-1}\tilde{h}_2, \quad \partial_g \tilde{h}_2 = -(1-g\bar{g})^{-1}\tilde{h}_1. \tag{2.17}$$

The following general solutions have been obtained[10]:

$$\tilde{h}_1 = \frac{df}{dg} + (1-g\bar{g})^{-1}(f\bar{g} - \bar{f}),$$
$$\tilde{h}_2 = \frac{d\bar{f}}{d\bar{g}} + (1-g\bar{g})^{-1}(\bar{f}g - f), \tag{2.18}$$

where f is $f(g)$, a holomorphic function and $\bar{f}(\bar{g})$ is the complex conjugate.

The function f has to be chosen suitably to get regular, normalizable solutions of Ψ. Going back to the variables $(z,\bar{z})$, from (2.16) and (2.18),

$$h_1 = \left(\frac{dg}{dz}\right)^{-1/2}\left\{\frac{df}{dz} + (1-g\bar{g})^{-1}(f\bar{g}-\bar{f})\frac{dg}{dz}\right\},$$
$$h_2 = \left(\frac{d\bar{g}}{d\bar{z}}\right)^{-1/2}\left\{\frac{d\bar{f}}{d\bar{z}} + (1-g\bar{g})^{-1}(\bar{f}g - f)\frac{d\bar{g}}{f\bar{z}}\right\}. \tag{2.19}$$

Define

$$G = (1-g\bar{g})(1-g)^{-1}(1-\bar{g})^{-1}$$
$$= \tfrac{1}{2}\{(1+g)(1-g)^{-1} + (1+\bar{g})(1-\bar{g})^{-1}\}, \tag{2.20}$$

the phase

$$e^{i\beta} = (1-g)(1-\bar{g})^{-1}\left(\frac{dg}{dz}\right)^{-1/2}\left(\frac{d\bar{g}}{d\bar{z}}\right)^{1/2}, \tag{2.21}$$

and set

$$f(z) = (1 - g(z))H(z), \tag{2.22}$$

where $H(z)$ remains to be chosen suitably. Then in (2.14), using (2.2), (2.19), (2.20), and (2.21), one can show that

$$\begin{aligned} r^{-1}e^{\xi/2}h_1 &= \sqrt{2}e^{i\beta/2}(G/r)^{1/2}\,\partial_z((H+\bar{H})/G), \\ r^{-1}e^{\xi/2}h_2 &= \sqrt{2}e^{-i\beta/2}(G/r)^{1/2}\,\partial_{\bar{z}}((H+\bar{H})/G). \end{aligned} \tag{2.23}$$

How should one choose $H(z)$? Consider first the well-known case (2.5). Let the roots of

$$1 - g = 0, \tag{2.24}$$

which are all purely imaginary when nonzero, be $z = z_j$, $j = 1,2,...,m$, $m = (n-1)$ for even n, $m = n$ for odd n. Then choosing

$$H = H_j = (\text{const})/(z - z_j) \quad (j = 1,2,...,m), \tag{2.25}$$

one gets solutions equivalent to the known set of normalizable zero modes[12] for the 't Hooft or Jackiw–Nohl–Rebbi gauge. [The phase β in (2.21) and (2.23) is the effect of a gauge transformation connecting Witten's solution to their 't Hooft or JNR forms. To go over from Witten's to the "string gauge" of (2.1) one needs a further transformation by $e^{-i\varphi\sigma_3/2}\,e^{-i\theta\sigma_2/2}$.] In fact [remembering that in (2.24) and (2.25) $z_j = -\bar{z}_j$],

$$(z+\bar{z})^{-1}G \simeq \left(1 + \sum_{j=1}^{n-1} \lambda_j^2 |z - z_j|^{-2}\right), \quad \text{for even } n, \tag{2.26}$$

$$\approx \sum_{j=1}^{n} \lambda_j^2 |z - z_j|^{-2}, \quad \text{for odd } n, \tag{2.27}$$

where the λ's are real functions of the b's. This makes the above-mentioned relations evident on comparing with Ref. 12.

Let us try to construct *periodic* spinor solutions for g given by (2.3) bearing the following points in mind.

(1) H must be periodic in t with a period $T = 2\pi/k$ like the gauge field.

(2) As $r \to 0$, $(H + \bar{H})$ should $\to 0$ at least as r to avoid a singularity.

(3) Asymptotic behavior of H as $r \to \infty$ should be compatible with normalizability of Ψ.

There should not, of course, be undesirable properties of $(H + \bar{H})/G$ elsewhere in the (r,t) half-plane $(0 \leqslant t \leqslant 2\pi)$. Define

$$\begin{aligned} &U = \tanh z = \tanh 1/2(r + it), \\ &b_j = (1 + a_j)(1 - a_j)^{-1}, \text{ when } b_j + \bar{b}_j > 0 \text{ if } |a_j| < 1. \end{aligned} \tag{2.28}$$

Then, rejecting a possible overall constant phase [which disappears in (2.2)] one has from (2.3) (for $k = 1$)

$$g = \left\{\prod_{j=1}^{n} (b_j - U)(\bar{b}_j + U)^{-1}\right\}(1 - U)^n (1 + U)^{-n}. \tag{2.29}$$

The equation $1 - g = 0$ turns out to have $(2n-1)$ roots (zero or purely imaginary), which we denote by

$$U = ic_i \quad (i = 1,...,2n-1). \tag{2.30}$$

Then, from (2.20), one can show that

$$G = (\text{const})(U + \bar{U}) \times \left\{1 + \sum_{i=1}^{2n-1} \lambda_i^2 (U - ic_i)^{-2}(\bar{U} + ic_i)^{-2}\right\}, \tag{2.31}$$

where the λ's are real functions of the b's. For the simplest case, namely

$$g = (a + e^{-2z})(a + e^{2z})^{-1} \quad (0 < a = \bar{a} < 1), \tag{2.32}$$

$$G = (U + \bar{U})(1 + b)^{-1}(1 + b/U\bar{U}) \quad [b = (1+a)(1-a)^{-1}]. \tag{2.33}$$

Comparing with (2.25) and (2.26) one sees that the choice, in (2.23),

$$H_i \approx 1/(U - ic_i) \quad (i = 1,...,2n-1) \tag{2.34}$$

gives regular normalizable solutions. *This gives* $(2n-1)$ *solutions*, while from (2.4)

$$S_T = 8\pi^2 2n.$$

This is to be compared to (2.5), (2.6), and (2.25), where a similar procedure gives $(n-1)$ solutions for

$$S = 8\pi^2(n-1).$$

Setting

$$H + \bar{H} = U + \bar{U}, \tag{2.35}$$

one also gets a good solution. But this is not linearly independent from the set (2.34) since

$$(U + \bar{U}) + \sum_{i=1}^{2n-1} \lambda_i^2 (H_i + \bar{H}_i) \simeq G. \tag{2.36}$$

So, here we have

$$\text{number of zero modes} = \{\text{action}/(8\pi^2)\} - 1. \tag{2.37}$$

This corresponds to the case $q = 1$ of (1.3).

The task of finding the roots $(c_i,\ i = 1,2,...,2n-1)$ of the polynomial in U(i.e., $g = 1$) has to be tackled separately for each case. They all correspond to $r = 0$, $U = i \tan \frac{1}{2} t = ic_i$. There is no simple general prescription for the general case (2.29). Particular values of b_j can make the problem manageable. For arbitrary n one has a simple case for $b_j = b$ $(j = 1,2,...,n)$,

$$(b - U)(\bar{b} + U)^{-1}(1 - U)(1 + U)^{-1} = e^{i(m/n)2\pi} \quad (m = 0,1,...,n-1), \tag{2.38}$$

the nth root of unity.

III. A GENERALIZATION: PERIODIC SOLUTIONS AS SCALING LIMITS

A subclass of 't Hooft instantons can be displayed in coordinates which permit a derivation of the periodic solutions as simple scaling limits. One can then compare the respective spinor solutions in such backgrounds. With this aim the formalism is briefly presented. (See Sec. V of Ref. 7 and my related previous papers cited in Ref. 7.)

Let

$$\begin{aligned} &r + it = \tanh \tfrac{1}{2}(\rho + i\tau), \quad r \in [0, \infty], \\ &t \in [-\infty, \infty]; \quad \rho \in [0, \infty], \quad \tau \in [-\pi, \pi]. \end{aligned} \tag{3.1}$$

Then

$$ds^2 = dt^2 + dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\varphi^2)$$
$$= (\cosh\rho + \cos\tau)^{-2}\{d\tau^2 + d\rho^2 + \sinh^2\rho(d\theta^2 + \sin^2\theta\, d\varphi^2)\}. \quad (3.2)$$

Set

$$A_\tau = (\partial_\rho\eta)\sigma_3/2, \quad A_\rho = -(\partial_\tau\eta)\sigma_3/2,$$
$$A_\theta = -e^\eta\sigma_2/2, \quad A_\varphi = \sin\theta\, e^\eta\sigma_1/2 - \cos\theta\,\sigma_3/2, \quad (3.3)$$

where

$$e^\eta = \sinh\rho(1 - g\bar{g})^{-1}\left(\frac{dg}{d\omega}\frac{d\bar{g}}{d\omega}\right)^{1/2}$$
$$[\omega = \tfrac{1}{2}(\rho + i\tau), \bar{\omega} = \tfrac{1}{2}(\rho - i\tau)], \quad (3.4)$$

and

$$g = \prod_{j=1}^{n}\left(\frac{a_j + e^{-\alpha}(\rho + i\tau)}{\bar{a}_j + e^{\alpha}(\rho + i\tau)}\right) \quad (|a_j| < 1, \quad 2n\alpha > 1). \quad (3.5)$$

[Here α cannot be absorbed by a rescaling like k of (2.3), keeping (3.4) intact due to the presence in (3.4) of the factor $\sinh\rho$, instead of just ρ, as compared to r in (2.2). The total action will depend on α, whereas S_T of (2.4) does not depend on k.] Integrating over φ, θ, ρ, and τ ($\tau\in[0,2\pi]$) gives now the *total* action over the entire R_4 ($t\in[-\infty,\infty]$, $r\in[0,\infty]$). In terms of (r,t) one will be integrating a solution *nonperiodic* in t over R_4. The *total* action is

$$S = 4\pi\int_0^{2\pi} d\tau\int_0^{\infty} d\rho\,\partial\omega\,\partial\bar{\omega}\left[-\ln\frac{\sinh\rho}{1 - g\bar{g}} + \frac{1}{2}e^{2\eta}\right]$$
$$(\partial\omega\,\partial\bar{\omega} = \partial_\rho^2 + \partial_\tau^2). \quad (3.6)$$

Steps quite analogous to those in the Appendix give

$$S = 8\pi^2(\alpha\cdot 2n - 1) \quad (\text{for } 2n\alpha > 1). \quad (3.7)$$

Comparing with (A11) note, apart from the factor α, *the subtraction of 1*. In the contour integral obtained from (3.6) [in a way analogous to the discussion leading to (A6)],

$$-\partial_\rho \ln\sinh\rho = -\coth\rho \to -1 \quad \text{as } \rho\to\infty,$$

and does not vanish like $(\partial_r \ln r)_{r\to\infty}$. Hence the difference.

Consider now the zero modes of a Dirac spinor in such a background. The conformal factor in (3.2) gives a factor $(\cosh\rho + \cos\tau)^{3/2}$ for spinors. Apart from this one can proceed as in Sec. II with a few differences. Replacing the arguments r,t by ρ,τ, respectively, and the *factors* r in (2.8), (2.14), and (2.23) by $\sinh\rho$, one obtains [in place of (2.23)]

$$(\sinh\rho)^{-1}e^{\eta/2}h_1$$
$$= \sqrt{2}e^{i\beta/2}G^{1/2}(\sinh\rho)^{-1/2}\partial_\omega\{(H + \bar{H})/G\}$$
$$(\partial_\omega = \partial_\rho - i\partial_\tau) \quad (3.8)$$

and its complex conjugate for h_2. Here $H = H(\rho + i\tau)$ and so on. In what follows α will be restricted to be an *integer*.

Expressing g in terms of (r,t) one obtains, from (3.5), for $(1/r)G$ the well-known generating functions of the 't Hooft gauge. Taking a simple example ($n = 1$, $\alpha = 2$),

$$g = (a + e^{-2(\rho + i\tau)})(a + e^{2(\rho + i\tau)})^{-1} \quad (0 < a < 1), \quad (3.9)$$

$$G = \frac{1 - g\bar{g}}{(1 - g)(1 - \bar{g})} \simeq r\left[1 + \frac{1}{r^2 + t^2} + \frac{2}{b}\frac{1}{r^2 + (t-1)^2} + \frac{2}{b}\frac{1}{r^2 + (t+1)^2}\right]. \quad (3.10)$$

Taking another example, for $\alpha = 1$, $n = 2$ in (3.5) and $a_j = \bar{a}_j$, $b_j = (1 + a_j)(1 - a_j)^{-1}$ the poles of G are at

$$r = 0, \quad t = 0, \quad \pm((b_1 + b_2 + 2b_2b_2)/(b_1 + b_2 + 2)). \quad (3.11)$$

For the general case (3.5) leads, consistently with (3.7), to

$$\text{number of poles} = (\alpha\cdot 2n - 1). \quad (3.12)$$

Hence, choosing H in (3.8) as

$$H = H_j = \{\tanh\tfrac{1}{2}(\rho + i\tau) - ic_j\}^{-1} = (r + it - ic_j)^{-1}, \quad (3.13)$$

where the c_j's represent the roots of

$$1 - g = 0 \quad (r = 0, \quad t = c_j),$$

one gets the full number of normalizable spinor zero modes of Grossman.[12]

Now define

$$\alpha\rho = r', \quad \alpha\tau = t' \quad (r'\in[0,\infty], \quad t'\in[-\alpha\pi,\alpha\pi]) \quad (3.14)$$

when

$$(1/\alpha)A_\rho = A_{r'}, \quad (1/\alpha)A_\tau = A_{t'},$$

and let $\alpha\to\infty$, when $t'\in[-\infty,\infty]$. One gets a periodic solution (for $A_{t'}, A_{r'}, A_\theta, A_\varphi$) of the type studied in Sec. II with

$$g = \prod_{j=2}^{n}\left(\frac{a_j + e^{-(r' + it')}}{\bar{a}_j + e^{(r' + it')}}\right). \quad (3.15)$$

The total action (3.7) remains invariant under such a rescaling and diverges with α. The action over *one period* ($= 2\pi$) is now

$$S_T = 8\pi^2\cdot 2n \quad (3.16)$$

and corresponds to the scaling limit of (3.7),

$$S_T = \lim_{\alpha\to\infty}(S/\alpha) = \lim_{\alpha\to\infty} 8\pi^2(2n - 1/\alpha) = 8\pi^2\cdot 2n. \quad (3.17)$$

In terms of $V = \tanh\frac{1}{2}(r' + it')$, for the limiting case,

$$G \simeq (V + \bar{V})\left[1 + \sum_{k=1}^{2n-1}\lambda_k^2\{(V - ic_k')(\bar{V} + ic_k')\}^{-1}\right]. \quad (3.18)$$

With now

$$H_k = (\text{const})(V - ic_k')^{-1}, \quad (3.19)$$

one gets $(2n - 1)$ spinor zero modes and *not* $2n$. This discrepancy corresponds to the fact that the last term (-1) of (3.7) has disappeared in the limit (3.17) since $1/\alpha\to 0$. Thus the relation (1.2) is automatically reproduced in the limit.

ACKNOWLEDGMENTS

I am indebted to P. Forgács, N. J. Hitchin, and Z. Horváth for help.

APPENDIX: ACTION OVER ONE PERIOD

The gauge potentials are [see (2.1)]

$$A_t = (\partial_r \zeta)\sigma_3/2,\ A_r = -(\partial_t \zeta)\sigma_3/2, \tag{A1}$$
$$A_\theta = -e^{\zeta}\sigma_2/2,\ A_\varphi = \sin\theta\, e^{\zeta}\sigma_2/2 - \cos\theta\,\sigma_3/2,$$

where

$$e^{\zeta} = r(1-g\bar{g})^{-1}\left(\frac{dg}{dz}\frac{d\bar{g}}{d\bar{z}}\right)^{1/2}, \tag{A2}$$

with

$$z = \tfrac{1}{2}(r+it),\quad \bar{z} = \tfrac{1}{2}(r-it),$$

and

$$g(z) = \prod_{j=1}^{n}(a_j + e^{-2z})(\bar{a}_j + e^{2z})^{-1} \quad (|a_j|<1). \tag{A3}$$

Note the absence of a factor e^{-2z} as compared to Sec. II of Ref. 7. See the remarks following (A11).

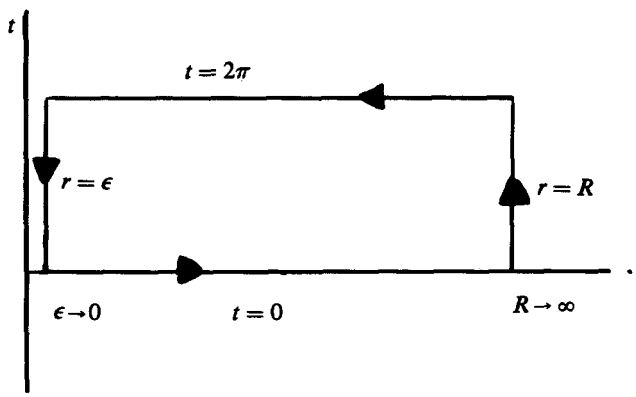
The action over one period T (here $T=2\pi$) reduces, after angular integrations, to

$$S_T = 4\pi\int_{t=0}^{t=T=2\pi} dt \int_{r=0}^{\infty} dr\, S_d, \tag{A4}$$

where

$$S_d = \partial_z\partial_{\bar{z}}(-\ln[r/(1-g\bar{g})] + \tfrac{1}{2}e^{2\zeta}) \equiv \partial_z\partial_{\bar{z}}\Omega. \tag{A5}$$

Consider the domain $0\leqslant t\leqslant 2\pi$, $\epsilon\leqslant r\leqslant R$:



Here S_d is regular inside, so using Stoke's theorem

$$4\pi\int_0^{2\pi} dt \int_{\epsilon}^{R} dr\, S_d$$
$$= 4\pi\int_{\epsilon}^{R} dr[\partial_t\Omega]_{t=0}^{t=2\pi} - 4\pi\int_0^{2\pi} dt\,[\partial_r\Omega]_{r=\epsilon}^{r=R}. \tag{A6}$$

Here S_T is obtained as $\epsilon\to 0$ and $R\to\infty$. The first integral vanishes due to periodicity. In the second only the first term of Ω contributes in the final limit.

As $r\to 0$, $g\bar{g}\to 1-2r\delta+2r^2\delta^2$, where

$$\delta = \sum_{j=1}^{n}\{(1+a_j e^{it})^{-1} + (1+\bar{a}_j e^{-it})^{-1}\}$$
$$= 2\sum_{j=1}^{n}\frac{1+|a_j|\cos(t+\delta_j)}{1+|a_j|^2+2|a_j|\cos(t+\delta_j)}$$
$$(a_j = |a_j|e^{i\delta_j}, |a_j|<1). \tag{A7}$$

Hence

$$-\partial_r \ln\{r(1-g\bar{g})^{-1}\}_{r\to 0} \to \delta. \tag{A8}$$

As $r\to\infty$, $g\bar{g}\to 0$ exponentially.

Hence using (for $0<a<1$)

$$\int_0^{2\pi}\frac{dt}{1+a^2+2a\cos t} = 2\int_0^{\pi}\frac{dt}{1+a^2+2a\cos t} = \frac{\pi}{1-a^2},$$

one obtains

$$S_T = 8\pi^2\cdot 2n \quad (T=2\pi). \tag{A9}$$

For

$$g = \prod_{j=1}^{n}\left(\frac{a_j + e^{-2kz}}{\bar{a}_j + e^{2kz}}\right) \quad (k>0), \tag{A10}$$

integrating over

$$t\in[0, T=2\pi/k\,],$$

one again obtains

$$S_T = 4\pi(2\pi/k)2nk = 8\pi^2\cdot 2n \quad (T=2\pi/k). \tag{A11}$$

We have implicitly modified f of (2.21) of Ref. 7 to

$$f = e^{2z}\prod_{j=1}^{n}(a_j + e^{-2kz})(\bar{a}_j + e^{2kz})^{-1}$$

giving (A10) for $g = e^{-2z}f$. This makes S_T independent of k and the correspondence to (4.1) of Ref. 7 even closer. Analogous modifications can also be made in choosing the f's in Secs. III–V of Ref. 7.

In the limit all

$$a_j = 0,\quad g(z) = e^{-4nkz},$$

and

$$e^{\zeta} = 4nkre^{-2nkr}/(1-e^{-4nkr}) = 2nkr/\sinh 2nkr. \tag{A12}$$

Now (A1) corresponds to a rescaled version of the *P-S* monopole. Since S_d is static, for (A12), the t integration in (A4) gives a factor $T = 2\pi/k$, so that

$$S_T = 4\pi\frac{2\pi}{k}\int_{r=0}^{\infty} S_d = \frac{8\pi^2}{k}(2nk) = 8\pi^2 2n. \tag{A13}$$

Thus (A11) is continuous in the limit $a_j = 0$.

Note that G in (2.32) has the pole structure typical of the 't Hooft representation [see (2.26)] in terms of U ($=\tanh Z$) and *not* in terms of z. But it is from the harmonic (in four dimensions)

$$\Sigma \simeq r^{-1}G = (2r)^{-1}\{(1+g)(1-g)^{-1} + (1+\bar{g})(1-\bar{g})^{-1}\}$$
$$\simeq \frac{U+\bar{U}}{z+\bar{z}}\left[1 + \sum_{i=1}^{2n-1}\lambda_i^2\{(U+ic_i)(\bar{U}-ic_i)\}^{-1}\right], \tag{A14}$$

satisfying

$$\Box\,\Sigma = (\partial_t^2 + \partial_r^2 + (2/r)\partial_r)\Sigma = 0 \tag{A15}$$

(except for δ-function singularities at poles), that one can obtain[13] [after a gauge transformation involving the phase β of (2.22)] the 't Hooft form

$$A_\mu = \sigma_{\mu\nu}\,\partial_\nu \ln \Sigma. \tag{A16}$$

The extra factor in (A14),

$$(U+\bar{U})/(z+\bar{z}) = r^{-1}\sinh r(\cosh z \cosh \bar{z})^{-1} \tag{A17}$$

is crucial. In calculating the action given by (A5) and (A6), in terms of

$$U = u_1 + iu_2, \quad (u_1 \in [0,\infty], \quad u_2 \in [-\infty,\infty]),$$

$$\int dt \int dr\, \partial_z \partial_{\bar{z}} \left[-\ln \frac{r}{1-g\bar{g}} \right]$$

$$= \int\int dt\, dr\, \partial_z \partial_{\bar{z}} \left[\ln\left(\frac{U+\bar{U}}{r}\right) - \ln \frac{U+\bar{U}}{1-g\bar{g}} \right]. \tag{A18}$$

The second term can be written as

$$\int_0^\infty du_1 \int_{-\infty}^{\infty} du_2\, \partial_U \partial_{\bar{U}} \left[-\ln \frac{U+\bar{U}}{1-g\bar{g}} \right] \tag{A19}$$

and gives an action [typical of the second factor of (A14)]

$$S' = 8\pi^2(2n-1). \tag{A20}$$

The first term gives

$$S'' = 8\pi^2 \int_{\epsilon\to 0}^{R\to\infty} dr \left(\frac{1}{r^2} - \frac{1}{\sinh^2 r} \right)$$

or

$$S'' = 8\pi^2 [\coth r - 1/r]_0^\infty = 8\pi^2. \tag{A21}$$

Thus

$$S = S' + S'' = 8\pi^2 2n. \tag{A22}$$

In Sec. II the $(2n-1)$ spinor zero modes correspond to the poles of G' where

$$G \simeq (U+\bar{U})G',$$

$$G' = 1 + \sum_{i=1}^{2n-1} \lambda_i^2 \{(U-ic_i)(\bar{U}+ic_i)\}^{-1}. \tag{A23}$$

It has been noted that the choice

$$H + \bar{H} = U + \bar{U},$$

which, one might have thought, provides a supplementary solution corresponding to the extra factor in G, does not, in fact, give a linearly independent new solution [Eq. (2.36)].

I have not considered the zeros of dg/dZ, which are more complicated.[11,13] The zeros of $(1-g)$ lead directly to 't Hooft or JNR representations corresponding to the structure (A23). As $r\to\infty$, dg/dz has a zero of infinite order due to the factor e^{-2nkz}. This corresponds to the factor $\ln(\sinh r/r)$ giving infinite action over R_4 ($t\in[-\infty,\infty]$). But we are integrating over a period 2π. Hence the finite result.

[1] G. 't Hooft, Phys. Rev. D **14**, 3432 (1976); **18**, 2199(E) (1978).
[2] B. Berg and J. Stehr, Nucl. Phys. B **175**, 293 (1980).
[3] E. Corrigan and P. Goddard, Ann. Phys. (NY) **154**, 253 (1984).
[4] L. S. Brown, R. D. Carlitz, and C. Lee, Phys. Rev. D **16**, 417 (1977).
[5] R. Jaciw and C. Rebbi, Phys. Rev. D **16**, 1052 (1977).
[6] D. Boyanovsky and R. Blankenbecler, Phys. Rev. D **31**, 3234 (1985); M. Hirayama, *ibid.* **33**, 1079 (1986); M. Hirayama and N. Sugimasa, *ibid.* **35**, 600 (1987); J. R. Schmidt and A. M. Bincer, *ibid.* **35**, 3995 (1987); P. Forgács, L. O'Raifeartaigh, and A. Wipf, "Scattering theory, U(1) anomaly and index theorems for compact and non-compact manifolds," DIAS-STP-86-44.
[7] A. Chakrabarti, Phys. Rev. D **35**, 696 (1987).
[8] N. J. Hitchin (private communication).
[9] D. J. Gross, R. D. Pisarski, and L. G. Yaffe, Rev. Mod. Phys. **53**, 43 (1981). For a restricted class, with the winding number fixing the periodic action, see N. Christ and R. Jaciw, Phys. Lett. B **91**, 228 (1980).
[10] The solutions (2.18) and (2.19) were essentially obtained by Z. Horváth and P. Forgács during a visit of the present author to Budapest. This was done in the context of Witten-type background (2.5).
[11] E. Witten, Phys. Rev. Lett. **38**, 121 (1977).
[12] B. Grossman, Phys. Lett. A **61**, 86 (1977).
[13] C. H. Lo, K. S. Narain, S. Y. Park, and K. C. Wali, Phys. Rev. D **20**, 3316 (1979).

Driven three-state model and its analytic solutions

C. E. Carroll[a] and F. T. Hioe
Department of Physics, St. John Fisher College, Rochester, New York 14618

(Received 23 December 1986; accepted for publication 30 September 1987)

For an atom or molecule in which two transitions are driven by laser beams, a three-state model is used. Transitions among the three states are caused only by the oscillating electric fields of the laser beams. The amplitudes and detunings of the two laser beams, which are various functions of the time, appear in the Schrödinger equation for the atom or molecule. In certain cases, the Schrödinger equation can be solved analytically, to find transition probabilities and the probability of no transition. This is done by using Clausen's special function, or by assuming that the sum of the two detunings is zero at all times. Conditions for complete transfer of the population from the ground state to an excited state are obtained from the analytic solutions.

I. INTRODUCTION

The excitation of an atom or molecule by multiphoton processes plays an important part in quantum optics, and this suggests study of simple models of an atom driven by laser beams. Here, we use a three-state model for an atom or molecule in which two transitions are driven by two laser beams, possibly derived from the same laser. Calculations involving all three states are necessary when the two laser beams act simultaneously, as in the cases we consider. The quantum dynamics of this three-state model is more complicated than that of the two-state model,[1] partly because the detunings of the laser beams driving the two transitions can be arbitrary functions of the time. In this paper, a few of the many possible cases are treated analytically.

We treat the oscillating electric fields of the two laser beams as classical external fields that appear in the equation of motion of the three-state system. Also, we neglect the relaxation terms that appear in the Bloch equations[2] and their generalization to three states. Hence, we can use the Schrödinger equation as the equation of motion. The external oscillating fields, which appear in the time-dependent Schrödinger equation, drive transitions between states labeled by consecutive integers. We use the electric-dipole approximation, which implies that transitions between states 1 and 3 cannot be driven. The Hamiltonian for this model is a 3×3 matrix that appears in the Schrödinger equation. This Hamiltonian and Schrödinger equation are simplified by using the rotating-wave approximation and the transformation associated with it[3]; this transformation eliminates all the optical-frequency terms. If the amplitudes and detunings of the two oscillating fields are independent of time, so is the Hamiltonian matrix; solution of the Schrödinger equation is straightforward in this case. Since optical pulses of short duration are often used, and time-dependent detunings could be used, this paper treats applied oscillating fields having time-dependent amplitudes and detunings. The changes in the amplitudes and detunings are supposed to be slow compared to the optical frequencies; they appear as slow changes in the Hamiltonian matrix. Clausen's function[4] can be used to solve the Schrödinger equation in some time-dependent cases, specified below. On the other hand, two dynamic symmetries that can be imposed on this three-state model were found earlier,[5] by using alternative sets of SU(3) generators. Use of Gell-Mann's set of SU(3) generators[6] is suggested by the assumption of "two-photon resonance," meaning that the detunings of the two laser beams add up to zero at all times. Our application of Clausen's function to this model depends on the weaker assumption that the sum of the two detunings is zero at the beginning of the two concurrent optical pulses, or at the end. Throughout this paper, we assume that the amplitudes of the two applied oscillating fields have a ratio independent of time. This should not be difficult to arrange if the two oscillating fields are derived from the same laser. It may be convenient to use two oscillating fields of fixed frequency, and to use the Stark or Zeeman effect to obtain time-dependent energy levels in the atom or molecule, which give time-dependent detunings. It seems possible that the calculations presented below can be tested by feasible experiments.

Although the Schrödinger equation for the three-state model can certainly be integrated numerically, analytic solutions are much more convenient for some purposes. For example, we shall set the occupation probabilities of two of the three states equal to zero at the initial time and also at the final time, and find conditions for complete return of the occupation probability to the initial state. Similar calculations will give conditions for complete transfer of the occupation probability from the ground state to the highest state.

We write the Schrödinger equation and describe transformations of it in Sec. II, which is a preface to the analytic solutions. The cases in which interesting results can be obtained from Clausen's function are described in Sec. III. Section IV treats the case in which the sum of the two detunings is always zero. The details of the calculations are given in the six Appendices.

II. SCHRÖDINGER EQUATION AND ITS TRANSFORMATIONS

In this section, we describe the three-state model of an atom or molecule and its simplification by the electric-dipole and rotating-wave approximations. The resulting Schrödinger equation is a system of three coupled first-order differential equations. It can be simplified by the transforma-

[a] Present address: Department of Physics, Auburn University, Auburn, Alabama 36849.

 0022-2488/88/020487-23$02.50

tion of Einwohner, Wong, and Garrison,[3] and then reduced to single uncoupled differential equations. The uncoupled differential equations, which are relegated to Appendices A and B, can sometimes be solved by use of known special functions.

Since the oscillating electric fields are treated as classical external fields, the three-state atom or molecule is described by a wave function with three complex components, each of which is the probability amplitude for finding the atom or molecule in a specific state. Transitions between these states are caused only by the applied oscillating fields, and we neglect the effect of the atom or molecule on the applied fields. In H, the 3×3 Hamiltonian matrix for this model, the diagonal elements are E_1, E_2, and E_3, the energies of the three states. Each off-diagonal element is proportional to an applied oscillating field times the corresponding transition dipole moment. The 1–3 transition is not driven and cannot be driven, because of Laporte's rule. This is to say that matrix elements in two corners vanish:

$$H_{13}=H_{31}=0. \tag{1}$$

The Schrödinger equation that had been described is greatly simplified by the rotating-wave approximation, which is neglect of counterrotating terms appearing in nonzero off-diagonal elements of the Hamiltonian matrix. This is closely related to the assumption that each oscillating electric field drives only its own transition. The result of these two approximations is that the four nonzero off-diagonal elements of H are proportional to $\exp(i\omega_{12}t)$, $\exp(-i\omega_{12}t)$, $\exp(i\omega_{23}t)$, and $\exp(-i\omega_{23}t)$, where ω_{12} and ω_{23} are the frequencies of the applied fields, and t is the time.

The optical-frequency terms can now be eliminated from the Hamiltonian and Schrödinger equation by using the time-dependent unitary transformation that is associated with the rotating-wave approximation. This transformation was formulated, for an arbitrary number of states, by Einwohner, Wong, and Garrison.[3] Since the unitary matrix used here is diagonal, the transformation does not affect the numbering of our three states, nor (1). The transformation is arranged to remove the factors of $\exp(\pm i\omega_{12}t)$ and $\exp(\pm i\omega_{23}t)$ from the off-diagonal elements of H. Since the transformation is time dependent, it changes the diagonal elements of H, replacing the energies E_1, E_2, and E_3 by linear combinations of the detunings Δ_1 and Δ_2, which are more relevant to this problem. Assuming that $\hbar=1$, the transformation puts the Schrödinger equation into the form

$$i\frac{d}{dt}\begin{pmatrix}a_1\\a_2\\a_3\end{pmatrix}=\begin{pmatrix}(2\Delta_1+\Delta_2)/3 & -\tfrac{1}{2}\Omega_1 & 0\\ -\tfrac{1}{2}\Omega_1 & (-\Delta_1+\Delta_2)/3 & -\tfrac{1}{2}\Omega_2\\ 0 & -\tfrac{1}{2}\Omega_2 & -(\Delta_1+2\Delta_2)/3\end{pmatrix}\begin{pmatrix}a_1\\a_2\\a_3\end{pmatrix}. \tag{2}$$

Here a_1, a_2, and a_3 are the probability amplitudes of the three states, and their absolute squares are the time-dependent occupation probabilities. Furthermore, Ω_1 and Ω_2 are the so-called Rabi frequencies; the products of field amplitudes and transition dipole moments are $\frac{1}{2}\Omega_1$ and $\frac{1}{2}\Omega_2$. The oscillation frequency calculated by Rabi[7] is equal to Ω_1 only if $\Delta_1=\Omega_2=0$. In general, Δ_1, Δ_2, Ω_1, and Ω_2 depend on t.

The signs of Δ_1 and Δ_2 depend on a convention[8] that should be mentioned. We do not assume anything about the signs of E_1-E_2 and E_2-E_3, the energy differences; any one of the three states could be the ground state. We assume that ω_{12} and ω_{23} have the same signs as E_1-E_2 and E_2-E_3, respectively. The measured frequencies are $|\omega_{12}|$ and $|\omega_{23}|$. The two detunings are defined as

$$\Delta_1=E_1-E_2-\omega_{12}$$

and

$$\Delta_2=E_2-E_3-\omega_{23}.$$

These differences are time dependent, and can change sign.

The Hamiltonian matrix appearing in (2) has vanishing trace, because of a convenient choice in the transformation described by Einwohner, Garrison, and Wong.[3] They determine only the differences of the coefficients they call γ_α; an arbitrary function of t could be added to every one of them. In other words, an arbitrary time-dependent phase transformation could be used. The vanishing trace of H is convenient for many calculations. However, in the case of two-photon resonance at all times, $\Delta_1+\Delta_2=0$ always holds, and we can arrange to obtain

$$i\frac{d}{dt}\begin{pmatrix}a_1\\a_2\\a_3\end{pmatrix}=\begin{pmatrix}0 & -\tfrac{1}{2}\Omega_1 & 0\\ -\tfrac{1}{2}\Omega_1 & -\Delta_1 & -\tfrac{1}{2}\Omega_2\\ 0 & -\tfrac{1}{2}\Omega_2 & 0\end{pmatrix}\begin{pmatrix}a_1\\a_2\\a_3\end{pmatrix} \tag{3}$$

instead of (2). The zeros at each corner of the Hamiltonian matrix are a simple and desirable feature.

Analytic solutions of (2) or (3) can be found by using various special functions. Before writing any of them, it is expedient to introduce a new independent variable, called z or $z(t)$. This nondecreasing function of t ranges from 0 to 1 as t increases from $-\infty$ to $+\infty$. This arbitrary function of t is independent of the main calculation, and it allows one analytic solution to be applied to a variety of pulse shapes. In an early treatment of the two-state problem, using the hyperbolic-secant pulse shape, Rosen and Zener[9] used

$$z(t)=\tfrac{1}{2}[1+\tanh(\pi t/\tau)], \tag{4}$$

where τ is the time constant. Several other examples of functions $z(t)$ are given in a recent paper[10] on the two-state problem. In order to change the independent variable from t to z, we put $\Delta_1=\delta_1(z)dz/dt$, $\Delta_2=\delta_2(z)dz/dt$, $\Omega_1=\omega_1(z)dz/dt$, and $\Omega_2=\omega_2(z)dz/dt$ into (2) and (3). If references to infinitely long time intervals are undesirable, we can make the amplitudes of the two optical pulses vanish outside a finite time interval, by making $z(t)$ constant when $|t|$ is sufficiently large.

This change of variable transforms the Schrödinger equations into

$$i\frac{d}{dz}\begin{pmatrix}a_1\\a_2\\a_3\end{pmatrix}=\begin{pmatrix}(2\delta_1+\delta_2)/3 & -\frac{1}{2}\omega_1 & 0\\ -\frac{1}{2}\omega_1 & (-\delta_1+\delta_2)/3 & -\frac{1}{2}\omega_2\\ 0 & -\frac{1}{2}\omega_2 & -(\delta_1+2\delta_2)/3\end{pmatrix}\begin{pmatrix}a_1\\a_2\\a_3\end{pmatrix} \tag{5}$$

and

$$i\frac{d}{dz}\begin{pmatrix}a_1\\a_2\\a_3\end{pmatrix}=\begin{pmatrix}0 & -\frac{1}{2}\omega_1 & 0\\ -\frac{1}{2}\omega_1 & -\delta_1 & -\frac{1}{2}\omega_2\\ 0 & -\frac{1}{2}\omega_2 & 0\end{pmatrix}\begin{pmatrix}a_1\\a_2\\a_3\end{pmatrix}, \tag{6}$$

which are systems of coupled first-order differential equations. We plan to choose $\delta_1(z)$, $\delta_2(z)$, $\omega_1(z)$, and $\omega_2(z)$ so that analytic solutions can be written down.

We should mention the simple case in which δ_1, δ_2, ω_1, and ω_2 are constants. Solution of (5) begins with solution of a cubic equation, to get the eigenvalues of the Hamiltonian matrix, unless eigenvalues or eigenvectors can be obtained by inspection. In the special case treated by Cook and Shore,[11] one of the eigenvalues is zero and the other two are easily found; this simplifies the remainder of the calculation. In the case of two-photon resonance at all times, constant values of $\delta_1 = -\delta_2$, ω_1, and ω_2 lead to similar simplifications.

If δ_1, δ_2, ω_1, and ω_2 depend on z, it seems necessary to uncouple the differential equations in order to find and apply appropriate special functions. Uncoupled third-order differential equations for $a_1(z)$ and $a_3(z)$ can be found from (5); this is done in Appendix A. Comparison with the third-order differential equation written by Clausen[4] is postponed; another third-order differential equation could be used. To solve (6), we need only solve a second-order differential equation.

This section has described our three-state model for an atom or molecule, and the time-dependent Schrödinger equation has been written explicitly. Solutions for $a_1(z)$, $a_2(z)$, and $a_3(z)$ can be obtained by using an appropriate special function. Finally, the time-dependent occupation probabilities are the absolute squares of a_1, a_2, and a_3.

III. APPLICATION OF CLAUSEN'S FUNCTION

Analytic solutions of the time-dependent Schrödinger equation can be written, by using Clausen's function, in some special cases that are described in this section. Further specialization gives cases of complete transfer of the occupation probability from one state to another, and of complete return of the occupation probability to the initial state. These results have been briefly described elsewhere.[12]

We should mention that Clausen[4] introduced the series

$$1+\frac{\alpha'}{1!}\frac{\beta'}{\gamma'}\frac{\delta'}{\epsilon'}x + \frac{\alpha'(\alpha'+1)}{2!}\frac{\beta'(\beta'+1)}{\gamma'(\gamma'+1)}\frac{\delta'(\delta'+1)}{\epsilon'(\epsilon'+1)}x^2+\cdots, \tag{7}$$

which defines a function of x. Also, Clausen gave the third-order differential equation satisfied by this function of x. This function is

$$F(\alpha',\beta',\delta';\gamma',\epsilon';x),$$

in the semicolon notation that Pochhammer introduced much later.[13]

We have transformed the Schrödinger equation into the form (5), and indicated that $a_1(z)$ and $a_3(z)$ satisfy third-order differential equations. Comparison of these differential equations with Clausen's equation is best done indirectly, using the mathematics outlined in Appendix B.

To get from the general model described in Sec. II to specific results for certain cases, we impose successive restrictions on the model. First, we require the differential equations for $a_1(z)$ and $a_3(z)$ to have only three singular points, all of which are regular, because Clausen's equation has this property. Second, we require the differential equations for $a_1(z)$ and $a_3(z)$ to have solutions expressible in terms of Clausen's function. Third, we require that Clausen's function can be evaluated in terms of gamma functions or simpler functions, at least at the beginning and end of the two concurrent optical pulses. Fourth, we may require complete transfer of the occupation probability to another state, or complete return of the occupation probability to the initial state.

These requirements lead to calculations given in Appendices A–E. The main points will be described here.

The four functions of z that appear in (5) are chosen as

$$\begin{aligned}\omega_1(z) &= (\alpha_1/\pi)[z(1-z)]^{-1/2},\\ \omega_2(z) &= (\alpha_2/\pi)[z(1-z)]^{-1/2},\\ \delta_1(z) &= \beta_1/z+\gamma_1/(1-z),\\ \delta_2(a) &= \beta_2/z+\gamma_2/(1-z).\end{aligned} \tag{8}$$

Here, $z(t)$ is the new independent variable introduced in Sec. II. The resulting differential equations for $a_1(z)$ and $a_3(z)$ are given in Appendix A. The only singular points are at $z=0$, $z=1$, and $z=\infty$; each is a regular singular point. Singular points off the real z axis have been used recently,[10] but the possibility is not pursued in this section. We have assumed that z increases from 0 to 1 as t increases from $-\infty$ to $+\infty$; (4) is an example. If we use (4), the forms of the two optical pulses are given by

$$\Omega_1=(\alpha_1/\tau)\operatorname{sech}(\pi t/\tau),\quad \Omega_2=(\alpha_2/\tau)\operatorname{sech}(\pi t/\tau),$$
$$\Delta_1=(\pi/\tau)[\beta_1+\gamma_1-(\beta_1-\gamma_1)\tanh(\pi t/\tau)],$$

and

$$\Delta_2=(\pi/\tau)[\beta_2+\gamma_2-(\beta_2-\gamma_2)\tanh(\pi t/\tau)].$$

The six real parameters of this model are

$$\alpha_1,\alpha_2,\beta_1,\beta_2,\gamma_1,\gamma_2. \tag{9}$$

The dimensional areas[14] of the two optical pulses are

$$\int_{-\infty}^{\infty}\Omega_1(t)dt=\alpha_1,\quad \int_{-\infty}^{\infty}\Omega_2(t)dt=\alpha_2.$$

The detuning parameters β_1 and β_2 are proportional to the detunings at large negative times, corresponding to the beginning of the two concurrent optical pulses. The detuning

TABLE I. Conditions for obtaining final occupation probabilities in closed form, using Clausen's series. The parameters (9) of the three-state model appear here. If Clausen's series terminates, the integer n, which is defined by (E2), appears.

Initial occupation probabilities	Conditions on pulse parameters	Final occupation probabilities
arbitrary	$\gamma_1+\gamma_2=0,\quad \gamma_1=-\frac{1}{2}(\beta_1-\beta_2)$ $\alpha_1^2-\alpha_2^2=2\pi^2(\beta_1^2-\beta_2^2)$	results in Table II
arbitrary	$\beta_1+\beta_2=0,\quad \beta_1=-\frac{1}{2}(\gamma_1-\gamma_2)$ $\alpha_1^2-\alpha_2^2=2\pi^2(\gamma_1^2-\gamma_2^2)$	results in Table III
1 0 0	$\gamma_1+\gamma_2=0,\quad n>0$ $\alpha_1^2=4\pi^2n^2[1+(\beta_1-\gamma_1)/(\beta_1+\beta_2)\ ]$ $0<\alpha_1<2\pi n$ $\alpha_2^2=-4\pi^2(\beta_1-\gamma_1)[\beta_1+\beta_2+n^2/(\beta_1+\beta_2)\ ]$	example in Table IV
0 1 0	$\gamma_1+\gamma_2=0,\quad n\geq 0$ $(\beta_1+\gamma_1)(\beta_2-\gamma_1)>0$ $\alpha_1^2=4\pi^2[\beta_1^2+(n+\frac{1}{2})^2](\beta_2-\gamma_1)/(\beta_1+\beta_2)$ $\alpha_2^2=4\pi^2[\beta_2^2+(n+\frac{1}{2})^2](\beta_1+\gamma_1)/(\beta_1+\beta_2)$	examples in Table IV
0 0 1	$\gamma_1+\gamma_2=0,\quad n>0$ $\alpha_1^2=-4\pi^2(\beta_2+\gamma_1)[\beta_1+\beta_2+n^2/(\beta_1+\beta_2)\ ]$ $\alpha_2^2=4\pi^2n^2[1+(\beta_2+\gamma_1)/(\beta_1+\beta_2)\ ]$ $0<\alpha_2<2\pi n$	example in Table IV

parameters γ_1 and γ_2 are proportional to the detunings at the end of the two pulses. These six parameters satisfy various equations that express the conditions we impose on the model. The requirement that solutions of (5) can be written in terms of Clausen's function is treated in Appendix B. The result is simply

$$\gamma_1+\gamma_2=0 \tag{10}$$

or

$$\beta_1+\beta_2=0. \tag{11}$$

We may call these the cases of two-photon resonance at the final time and of two-photon resonance at the initial time. They are treated in the following two subsections. The solutions of (5) are written explicitly in Appendix C, assuming that (10) holds. We emphasize that

$$\beta_1+\beta_2=\gamma_1+\gamma_2=0 \tag{12}$$

is not a case in which Clausen's function is applicable. This is the previously studied case[5] of two-photon resonance, and it is treated in Sec. IV.

The requirement that occupation probabilities can finally be expressed in simple terms, despite the appearance of Clausen's series in all the wave functions, leads to two more equations connecting the six parameters of the model. The resulting conditions for various cases are listed in Table I. If the series (7) terminates, its evaluation is straightforward, and the number of terms in the series is approximately the same as the integer n that appears in the corresponding formulas. If the series (7) does not terminate, it converges when $|x|<1$; but numerical summation of the series has not been attempted. Evaluation of the infinite series at $x=1$ is used to obtain the results shown in Tables II and III. In some of our calculations, the infinite series diverges at $x=1$; this causes no difficulty, because the behavior of Clausen's function in this case is derived in Appendix D. When the infinite series converges at $x=1$, we need an applicable summation formula. Among the known formulas (Appendix D), only

TABLE II. Transition probabilities, and probabilities of no transition, for the first case listed in Table I. We define $\Phi=[\frac{1}{4}(\alpha_1^2+\alpha_2^2)-\frac{1}{2}\pi^2(\beta_1^2+\beta_2^2)]^{1/2}$; this angle is real or pure imaginary.

Initial occupation probabilities	Final occupation probabilities
1	$\dfrac{[\cosh(\pi\beta_1)+\cos\Phi][\cosh(\pi\beta_2)+\cos\Phi]}{4\cosh(\pi\beta_1)\cosh[\frac{1}{2}\pi(\beta_1+\beta_2)]\cosh[\frac{1}{2}\pi(\beta_1-\beta_2)]}$
0	$\dfrac{[\cosh(\pi\beta_1)+\cos\Phi][\cosh(\pi\beta_1)-\cos\Phi]}{2\cosh(\pi\beta_1)\cosh[\frac{1}{2}\pi(\beta_1+\beta_2)]\cosh[\frac{1}{2}\pi(\beta_1-\beta_2)]}$
0	$\dfrac{[\cosh(\pi\beta_1)-\cos\Phi][\cosh(\pi\beta_2)-\cos\Phi]}{4\cosh(\pi\beta_1)\cosh[\frac{1}{2}\pi(\beta_1+\beta_2)]\cosh[\frac{1}{2}\pi(\beta_1-\beta_2)]}$
0	$\dfrac{[\cosh(\pi\beta_1)-\cos\Phi][\cosh(\pi\beta_2)+\cos\Phi]}{2\cosh(\pi\beta_1)\cosh(\pi\beta_2)}$
1	$\dfrac{\cos^2\Phi}{\cosh(\pi\beta_1)\cosh(\pi\beta_2)}$
0	$\dfrac{[\cosh(\pi\beta_1)+\cos\Phi][\cosh(\pi\beta_2)-\cos\Phi]}{2\cosh(\pi\beta_1)\cosh(\pi\beta_2)}$
0	$\dfrac{[\cosh(\pi\beta_1)-\cos\Phi][\cosh(\pi\beta_2)-\cos\Phi]}{4\cosh(\pi\beta_2)\cosh[\frac{1}{2}\pi(\beta_1+\beta_2)]\cosh[\frac{1}{2}\pi(\beta_1-\beta_2)]}$
0	$\dfrac{[\cosh(\pi\beta_2)+\cos\Phi][\cosh(\pi\beta_2)-\cos\Phi]}{2\cosh(\pi\beta_2)\cosh[\frac{1}{2}\pi(\beta_1+\beta_2)]\cosh[\frac{1}{2}\pi(\beta_1-\beta_2)]}$
1	$\dfrac{[\cosh(\pi\beta_1)+\cos\Phi][\cosh(\pi\beta_2)+\cos\Phi]}{4\cosh(\pi\beta_2)\cosh[\frac{1}{2}\pi(\beta_1+\beta_2)]\cosh[\frac{1}{2}\pi(\beta_1-\beta_2)]}$

TABLE III. Transition probabilities, and probabilities of no transition, for the second case listed in Table I. We define $\Phi = [\frac{1}{4}(\alpha_1^2 + \alpha_2^2) - \frac{1}{2}\pi^2(\gamma_1^2 + \gamma_2^2)]^{1/2}$; this angle is real or pure imaginary.

Initial occupation probabilities	Final occupation probabilities
1	$\dfrac{[\cosh(\pi\gamma_1) + \cos\Phi][\cosh(\pi\gamma_2) + \cos\Phi]}{4\cosh(\pi\gamma_1)\cosh[\frac{1}{2}\pi(\gamma_1+\gamma_2)]\cosh[\frac{1}{2}\pi(\gamma_1-\gamma_2)]}$
0	$\dfrac{[\cosh(\pi\gamma_1) - \cos\Phi][\cosh(\pi\gamma_2) + \cos\Phi]}{2\cosh(\pi\gamma_1)\cosh(\pi\gamma_2)}$
0	$\dfrac{[\cosh(\pi\gamma_1) - \cos\Phi][\cosh(\pi\gamma_2) - \cos\Phi]}{4\cosh(\pi\gamma_2)\cosh[\frac{1}{2}\pi(\gamma_1+\gamma_2)]\cosh[\frac{1}{2}\pi(\gamma_1-\gamma_2)]}$
0	$\dfrac{[\cosh(\pi\gamma_1) + \cos\Phi][\cosh(\pi\gamma_1) - \cos\Phi]}{2\cosh(\pi\gamma_1)\cosh[\frac{1}{2}\pi(\gamma_1+\gamma_2)]\cosh[\frac{1}{2}\pi(\gamma_1-\gamma_2)]}$
1	$\dfrac{\cos^2\Phi}{\cosh(\pi\gamma_1)\cosh(\pi\gamma_2)}$
0	$\dfrac{[\cosh(\pi\gamma_2) + \cos\Phi][\cosh(\pi\gamma_2) - \cos\Phi]}{2\cosh(\pi\gamma_2)\cosh[\frac{1}{2}\pi(\gamma_1+\gamma_2)]\cosh[\frac{1}{2}\pi(\gamma_1-\gamma_2)]}$
0	$\dfrac{[\cosh(\pi\gamma_1) - \cos\Phi][\cosh(\pi\gamma_2) - \cos\Phi]}{4\cosh(\pi\gamma_1)\cosh[\frac{1}{2}\pi(\gamma_1+\gamma_2)]\cosh[\frac{1}{2}\pi(\gamma_1-\gamma_2)]}$
0	$\dfrac{[\cosh(\pi\gamma_1) + \cos\Phi][\cosh(\pi\gamma_2) - \cos\Phi]}{2\cosh(\pi\gamma_1)\cosh(\pi\gamma_2)}$
1	$\dfrac{[\cosh(\pi\gamma_1) + \cos\Phi][\cosh(\pi\gamma_2) + \cos\Phi]}{4\cosh(\pi\gamma_2)\cosh[\frac{1}{2}\pi(\gamma_1+\gamma_2)]\cosh[\frac{1}{2}\pi(\gamma_1-\gamma_2)]}$

those named for Watson[15] and Whipple[16] are applicable to our calculations; they are used to obtain Tables II and III.

The requirement of complete transfer of the occupation probability to another state, or complete return to the initial state, leads to further requirements on the parameters (9), in addition to those listed in Table I. It also leads to use of a time-reversal transformation. Before we consider complete transfer and complete return, the calculations based on (10) and (11) will be described separately.

A. Two-photon resonance at the final time

Here, we assume that (10) holds, but not (11). The sum of the two detunings vanishes at the end of the two optical pulses, but not at all times. After substituting (8) into (5), we can write solutions in terms of Clausen's function; this is done in Appendix C. The connection between (7) and (5) is simply

$$x = z. \tag{13}$$

The natural initial condition is that one of the three states is occupied at the beginning of the two optical pulses, and the other two are unoccupied. Since Clausen's function is equal to unity at $x = z = 0$, such an initial condition is easy to satisfy. There are three such initial conditions, and they are satisfied by three independent solutions of (5), given in Appendix C.

The end of the two optical pulses corresponds to $x = z = 1$, and the difficulty of evaluating (7) at $x = 1$ has been mentioned. The final occupation probability of state 2 is the limit of $|a_2|^2$ as $x = z \to 1$, and this can always be evaluated. Under the conditions listed in the first section of Table I, the other two final occupation probabilities can be calculated, in the way described in Appendix E. The results are given in Table II.

Evaluating the final occupation probabilities is much more straightforward if Clausen's series terminates wherev-

TABLE IV. Transition probabilities, and probabilities of no transition, for the last three cases listed in Table I. The integer n, which appears in Table I, here is given specific values. We assume $\gamma_1 + \gamma_2 = 0$; see Table I for applicable inequalities and for α_1 and α_2, the pulse areas.

Initial occupation probabilities	Integer n	Final occupation probabilities
1		$\dfrac{4\beta_1^2(\beta_1+\beta_2)^2 + (3\beta_1+\beta_2-2\gamma_1)^2}{(4\beta_1^2+1)(\beta_1+\beta_2)^2}$
0	1	0
0		$\dfrac{-4(\beta_1-\gamma_1)(2\beta_1+\beta_2-\gamma_1)}{(4\beta_1^2+1)(\beta_1+\beta_2)^2}$
0		$(\beta_2-\gamma_1)/(\beta_1+\beta_2)$
1	0	0
0		$(\beta_1+\gamma_1)/(\beta_1+\beta_2)$
0		$\dfrac{(\beta_2-\gamma_1)[(\beta_1^2+\frac{1}{4})(\beta_2^2+\frac{1}{4}) + 2\beta_1(\beta_1-\beta_2) + 2(3\beta_1-\beta_2)\gamma_1 + 4\gamma_1^2]}{(\beta_1+\beta_2)(\beta_1^2+\frac{1}{4})(\beta_2^2+\frac{1}{4})}$
1	1	0
0		$\dfrac{(\beta_1+\gamma_1)[(\beta_1^2+\frac{1}{4})(\beta_2^2+\frac{1}{4}) - 2\beta_2(\beta_1-\beta_2) + 2(\beta_1-3\beta_2)\gamma_1 + 4\gamma_1^2]}{(\beta_1+\beta_2)(\beta_1^2+\frac{1}{4})(\beta_2^2+\frac{1}{4})}$
0		$\dfrac{-4(\beta_2+\gamma_1)(\beta_1+2\beta_2+\gamma_1)}{(4\beta_2^2+1)(\beta_1+\beta_2)^2}$
0	1	0
1		$\dfrac{4\beta_2^2(\beta_1+\beta_2)^2 + (\beta_1+3\beta_2+2\gamma_1)^2}{(4\beta_2^2+1)(\beta_1+\beta_2)^2}$

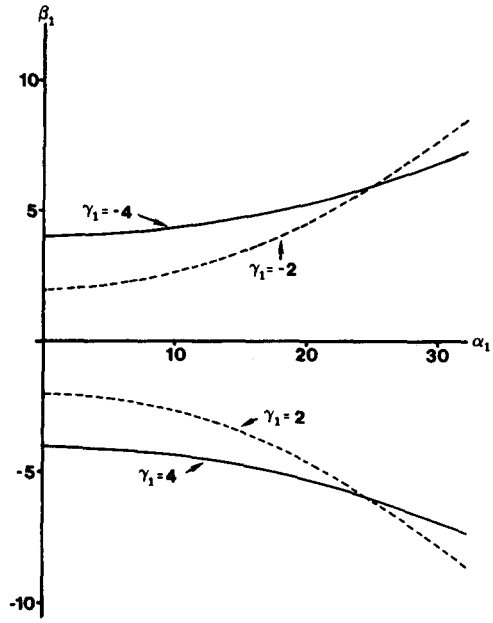


FIG. 1. Region of the α_1-β_1 plane corresponding to the first section of Table I lies above a lower parabola if γ_1 is positive. If γ_1 is negative, it lies below an upper parabola. It is required that $\gamma_2 = -\gamma_1$. For any point in the allowed region, α_2 and β_2 can easily be calculated.

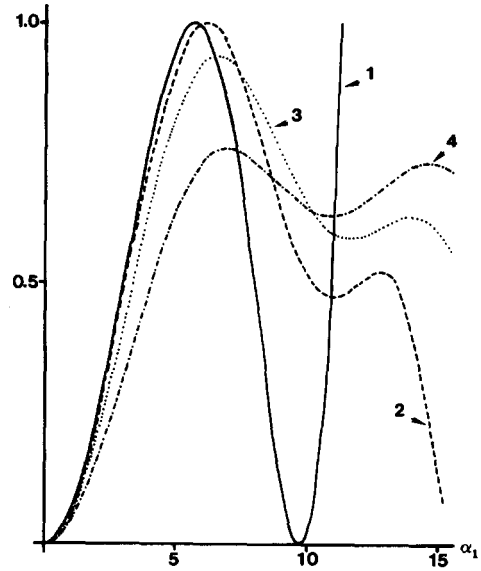


FIG. 3. Probability that the concurrent optical pulses cause transitions from state 2 to state 1. We assume $\beta_1 = 0.988$, $\beta_2 = 0.835n - 0.582$, and $\gamma_2 = -\gamma_1$. The values of α_2 and γ_1 are such that all series in the wave function terminate. Values of n are used to label the four curves. Each curve ends where $\alpha_2 = 0$, but large values of α_1 are not shown.

er it appears. The conditions for termination depend on which of the three initial conditions is used; they are listed in the last three sections of Table I. The final occupation probability for state 2 is zero, and the other two final occupation probabilities are rational functions of the parameters β_1, β_2, and γ_1. The simplest examples are listed in Table IV, and the results for $n = 2$ could be worked out.

Table I emphasizes that we impose three conditions on the parameters (9); (10) is the first of the three conditions. Hence, if we do not count the integer n as a parameter, each section of Table I represents a three-parameter solvable model. One might treat α_1, β_1, and γ_1 as the three independent parameters, and use Table I to calculate α_2, β_2, and γ_2 from them. It seems worthwhile to describe the inequalities restricting these three independent parameters, and to show how some of the transition probabilities depend on α_1.

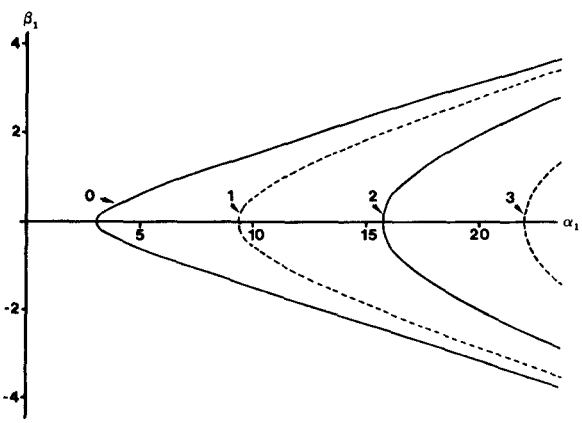


FIG. 2. Region of the α_1-β_1 plane corresponding to the penultimate section of Table I lies to the left of a hyperbola that depends on n. However, the line $\beta_1 + \gamma_1 = 0$ (not shown) is excluded from the allowed region. As n increases, the hyperbola recedes to infinity.

Figure 1 shows the region of the α_1-β_1 plane corresponding to the first section of Table I; the atom or molecule is initially in any one of its three states. Now, suppose that the atom or molecule is initially in state 2, and that Clausen's series terminates. This corresponds to the penultimate section of Table I, and the allowed region of the α_1-β_1 plane is shown in Fig. 2, for small values of n. The final occupation probability of state 1 is shown in Fig. 3. If the atom or mole-

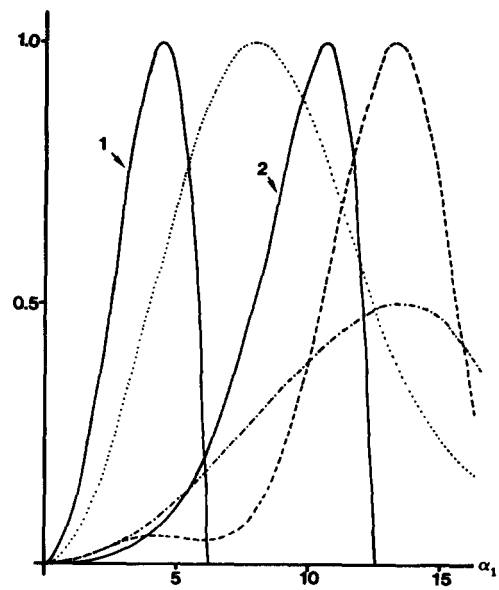


FIG. 4. Probability that the concurrent optical pulses cause transitions from state 1 to state 3. We fix β_1 and β_2, let $\gamma_2 = -\gamma_1$, and adjust α_2 and γ_1 so that all series in the wave function terminate. For curve 1, $n = 1$ and $\beta_1 = 0$; every nonzero value of β_2 gives the same curve 1. For curve 2, $n = 2$, $\beta_1 = 0.625$, and $\beta_2 = -1.276$ are used. For the remaining three curves, $n = 3$. Dashed curve: $\beta_1 = 0$ and $\beta_2 = -1.276$; dotted curve: $\beta_1 = 1.26$ and $\beta_2 = 0$; dash–dotted curve: $\beta_1 = 1.26$ and $\beta_2 = -1.276$. Each curve ends where $\alpha_2 = 0$.

cule is initially in state 1, then $0<\alpha_1<2\pi n$ and $\beta_1\neq\gamma_1$ are required for termination of the series; this corresponds to the third section of Table I. The final occupation probability of state 3 is shown in Fig. 4. Further curves could be obtained by using other values of β_1 and β_2, or by using other sections of Table I. The second section of Table I corresponds to the following topic.

B. Two-photon resonance at the initial time

Here, we assume that (11) holds, but not (10). The sum of the two detunings vanishes at the beginning of the two concurrent optical pulses, but not at all times. Solutions of (5) can be written in terms of Clausen's function; the connection between (5) and (7) is

$$x=1-z. \tag{14}$$

A time-reversal transformation that connects (10) and (11) can be constructed, even though the Hamiltonian used here is time dependent. This transformation is specified in Appendix C, and is used to avoid writing the solutions of (5) explicitly.

The initial condition used is that only one of the three states is occupied at the beginning of the two optical pulses, which corresponds to $x=1$. Since the Clausen functions are difficult to evaluate at $x=1$, in general, we satisfy this initial condition only in a special case. The parameters (9) are required to satisfy the conditions listed in the second section of Table I, so that the formulas of Watson[15] and Whipple[16] can be applied. The nine resulting asymptotic forms as $x\to 1$ and $z\to 0$, which involve numerous gamma functions, are used in the following way.

The three solutions of (5) that we used in the previous subsection satisfied simple initial conditions. Using the time-reversal transformation, we find solutions of the present problem that simplify at the final time, or as $x\to 0$. For each such solution, two occupation probabilities vanish at the end of the two concurrent optical pulses, and the occupation probability that approaches unity comes from a probability amplitude that is asymptotic to a known power of x. Using the formulas of Watson and Whipple to find the asymptotic form of each wave function as $x\to 1$, we can find the S matrix. Each of the known solutions of (5) gives one column of the inverse of the S matrix. From the three independent solutions, we obtain a 3×3 matrix. After checking that this matrix is unitary, we invert it. The resulting S matrix gives the transition probabilities, and probabilities of no transition, that are shown in Table III.

The transition probabilities that appear in Table III are notably different from those shown in Table II. However, the probabilities of no transition are quite similar to those shown in Table II; the only difference is exchange of β_1, β_2 with γ_1, γ_2. We could have anticipated this result; a time-reversal argument connects the probabilities of no transition shown in the two tables. The transition probabilities themselves are not obtainable from any such general argument.

The calculation leading to Table III is possible because the conditions for using the formulas of Watson and Whipple are independent of initial conditions and final conditions. This independence, which is emphasized in Table I, allows us to find three independent solutions of (5), and to calculate the whole S matrix.

C. Complete transfer and complete return

Analytic solutions of the Schrödinger equation are quite useful if we seek conditions for complete transfer of the occupation probability from the initial state to any desired final state, or for complete return to the initial state. In particular, in a system with a finite number of states, we can ask for complete inversion, which takes the atomic or molecular population from the ground state to the highest state. In the three-state model we use, the order of the three states on the energy scale is arbitrary, and any case of complete transfer could be complete inversion. The parameters (9) appear in the time-dependent Hamiltonian, and they satisfy certain conditions if we demand complete transfer or complete return. These conditions are added to those listed in Table I, with the result that our examples of complete transfer and complete return have only one adjustable parameter that varies continuously.

Conditions for complete transfer or complete return can be obtained, after the initial condition is satisfied, by setting two components of the wave function equal to zero at the final time, or by setting two of the final occupation probabilities equal to zero. Although it seems simpler, in principle, to work with the known wave functions, we begin with an inspection of the final occupation probabilities listed in Tables II and III. The simpler and more general way to find the cases of complete transfer and complete return will be given afterwards. First, we assume that the atom or molecule is initially in state 2, and we use Tables II and III. Complete transfer to another state is evidently not possible. Complete return to state 2 can be obtained only in the simple case of vanishing detunings, which is not considered in this paper. Thus we now assume that the atom or molecule is initially in state 1. The first condition for complete transfer or complete return is $\beta_1=0$ or $\gamma_1=0$ or $\gamma_2=0$. We are led to cases with resonant driving of one transition at the initial time and two-photon resonance at the final time, or vice versa. Complete transfer from state 1 to state 2 is not possible. For complete transfer from state 1 to state 3, the second condition is that Φ, the angle in Tables II and III, is an odd multiple of π. For complete return to state 1, Φ must be a positive even multiple of π; the case of $\Phi=0$ is excluded, because we now have $\Phi=2^{-1/2}\alpha_1$. If the atom or molecule is initially in state 3, similar simple results are obtained. Notice that Φ is a multiple of π in all these cases; this implies that Clausen's series terminates, wherever it appears in the wave function. Termination of the series implies that the final or initial occupation probability for state 2 vanishes, depending on whether (10) or (11) holds. Termination also means that we have applied Watson's formula to a finite series; this is the case actually treated in Watson's brief paper.[15]

We now find more general conditions for complete transfer or complete return, by starting with the assumption that Clausen's series always terminates. No cases that can be obtained from Table I are lost by this assumption, for we have just seen that the first two sections of Table I lead to terminating series. The conditions for complete return or

TABLE V. Conditions for complete return or complete transfer, obtained by assuming that (10) holds and Clausen's series terminates. The conditions written here are to be added to the inequalities, and formulas for α_1 and α_2, that are given in Table I.

Initial occupation probabilities	Final occupation probabilities	Condition for this
1	1	$F(1+n,1-n,1-i(2\beta_1+\beta_2-\gamma_1);$
0	0	$\frac{3}{2}-i\beta_1,2-i(\beta_1+\beta_2);1)=0$
0	0	
1	0	$F(n,-n,-i(2\beta_1+\beta_2-\gamma_1);$
0	0	$\frac{1}{2}-i\beta_1,-i(\beta_1+\beta_2);1)=0$
0	1	
0	1	$F(-n,n+1,\frac{1}{2}+i(\beta_1-\beta_2+\gamma_1);$
1	0	$\frac{1}{2}+i\beta_1,\frac{3}{2}-i\beta_2;1)=0$
0	0	
0	1	$F(n,-n,i(\beta_1+2\beta_2+\gamma_1);$
0	0	$\frac{1}{2}+i\beta_2,i(\beta_1+\beta_2);1)=0$
1	0	

complete transfer will be algebraic equations in the parameters (9), obtained by requiring two components of the wave function to vanish at the final time. Furthermore, the time-reversal transformation is more useful in this context than in the previous subsection. The initial and final conditions are that two components of the wave function vanish at the initial time and that two components, possibly the same two, vanish at the final time. Hence time reversal connects a case of complete transfer or return with another case of complete transfer or return. We shall use time reversal to treat the case of two-photon resonance at the initial time, after considering the case of two-photon resonance at the final time.

1. General conditions, when (10) holds

Here, we find the conditions for complete transfer and complete return, assuming that (10) holds, and that Clausen's series always terminates. As mentioned above, the final occupation probability for state 2 vanishes automatically. Hence, the condition for complete transfer or complete return is that one of Clausen's function vanishes at $x=1$, which corresponds to the final time. Such conditions are listed in Table V, for three cases of complete transfer and one case of complete return. The integer n appears, because terminating series are used. Since both real and imaginary parts of Clausen's function vanish at $x=1$, Table V gives two conditions additional to those appearing in the appropriate section of Table I.

The simplest examples are obtained by using small integers n. Although $n=0$ appears in Tables I and IV, the resulting cases of complete transfer are the trivial cases of $\alpha_1=0$ and $\alpha_2=0$. We reject these cases, to which Table V does not apply. Simple conditions for complete transfer from state 2 to state 1 are obtained by setting $n=1$ in the third section of Table V:

$$\beta_1\beta_2=\tfrac{1}{4},\quad \gamma_1=\tfrac{1}{4}(-\beta_1+3\beta_2),\quad \gamma_2=-\gamma_1,$$
$$\alpha_1^2=\pi^2(\beta_1^2+\tfrac{9}{4}),\quad \alpha_2^2=3\pi^2(\beta_2^2+\tfrac{9}{4}).$$

These conditions cannot be obtained from Table II or Table III. Simple conditions for complete return to state 1 are obtained by setting $n=2$ in the first section of Table V:

$$\beta_1=0,\quad \gamma_1=\tfrac{1}{2}\beta_2,\quad \gamma_2=-\gamma_1,$$
$$\alpha_1^2=8\pi^2,\quad \alpha_2^2=2\pi^2(\beta_2^2+4).$$

The same conditions can be obtained from Table II; this was done above, but not so explicitly. This specific example shows that there is an overlap between the first and third sections of Table I; it occurs because Watson's formula is applicable to finite series. By using even integers n in the first section of Table V, we should be able to recover the simple conditions for complete return that were obtained above from Table II. However, the algebra becomes more difficult as n increases.

Other simple examples can be obtained from Table V, but the formulas will not be shown explicitly. Simple conditions for complete transfer from state 1 to state 3, or vice versa, are obtained by setting $n=1$ in the appropriate sections of Table V. These conditions agree with those found above from Table II. Use of odd integers n should, in princi-

TABLE VI. Examples of complete return and complete transfer, obtained from the conditions shown in Table V. The integer n is specified here, and the time-reversal transformation is used to obtain the last line.

Initial occupation probabilities	Final occupation probabilities	n	α_1	α_2	β_1	β_2	γ_1	γ_2
1	1							
0	0	3	17.3401	9.5634	0.2180	2.2455	0.5967	−0.5967
0	0							
1	0							
0	0	2	10.6690	6.9826	0.6250	−1.2760	0.4433	−0.4433
0	1							
0	1							
1	0	2	6.1206	15.9668	0.9880	1.0880	0.8154	−0.8154
0	0							
1	0							
0	0	2	9.8424	7.9085	0.9670	−0.9670	0.5643	−0.8154
0	1							

ple, give all the cases of complete transfer that were obtained from Table II. We emphasize that the conditions obtained from Table V are more general than the simple conditions obtained from Table II, and that there are several cases in which five simple equations connecting the parameters (9) can be obtained from Table V. We show some numerical examples in Table VI. The parameters used in Figs. 3 and 4 are such that two of the specific examples in Table VI are the points of complete transfer on the $n = 2$ curves. Numerical exploration has been used to arrange that, in Fig. 4, points of complete transfer appear on two of the $n = 3$ curves.

2. General conditions, when (11) holds

The conditions for complete transfer or complete return, in the case of two-photon resonance at the initial time, are obtained by applying time reversal to the above calculations. Time reversal obviously interchanges the numbers in the first two colums of Table V. Equation (8) indicates that interchange of β_1, β_2, with γ_1, γ_2 is necessary. No other changes are actually needed in Table V. The appropriate formulas for α_1^2 and α_2^2, which involve the integer n, must be derived from the time-reversal transformation specified in Appendix C; they are not available in Table I. A specific example of complete transfer appears as the last line in Table VI.

This section has described specific cases in which the Schrödinger equation specified by (5) and (8) is solvable by use of (7), Clausen's series. We impose successive conditions on the parameters (9), and we finally obtain complete transfer or complete return, along with termination of Clausen's series.

IV. TWO-PHOTON RESONANCE AT ALL TIMES

The three-state model that leads to (2) can motivate a study of the algebra of 3×3 matrices, and this algebra simplifies greatly in two cases found earlier.[5] The two cases could also be derived by looking for simplifications in the differential equations derived from (2). If the two detunings are always equal and $\Omega_1(t) = \Omega_2(t)$ always holds, the third-order differential equations can be reduced to second order by Appell's transformation[17]; this case will not be described here. If the two detunings always add up to zero, we have the coupled differential equations shown as (3) or (6). We shall use the simplifying assumption that $\Omega_1(t)/\Omega_2(t)$ is independent of t. This leads to "coherent trapping" of atoms in one state, which Gray, Whitley, and Stroud[18] demonstrated by using a beam of Na atoms. The effect appeared somewhat earlier, in the calculations of Arimondo and Orriols.[19] The conditions for coherent trapping are satisfied by the model used in Sec. III when (12) holds, but this limiting case is not treated in Sec. III. Analytic solutions of this and some other special cases will be given below, after some general properties of (3) and (6) are mentioned.

A. General properties

The simplest calculation[18] that shows coherent trapping is construction of the constant solution of (3). Since we assume Ω_1/Ω_2 is constant,

$$a_1 = \Omega_2(\Omega_1^2 + \Omega_2^2)^{-1/2}, \quad a_2 = 0, \qquad a_3 = -\Omega_1(\Omega_1^2 + \Omega_2^2)^{-1/2} \tag{15}$$

is this solution; we have normalized it. This represents a quantum state whose occupation probability can neither increase nor decrease, so long as we use the assumptions emphasized above.

Further transformation of the Schrödinger equation could be used to separate (15) from the other two solutions, so that one obtains a two-state model. We may indicate the form of solutions of (3), without writing such a transformation. If state 1 is occupied initially with unit probability, then the probability amplitudes for states 2 and 3 vanish at $t = 0$, the initial time, and

$$a_1 = \Omega_2^2/(\Omega_1^2 + \Omega_2^2) + [\Omega_1^2/(\Omega_1^2 + \Omega_2^2)]f(t),$$
$$a_2 = g(t),$$

and

TABLE VII. Transition probabilities, and probabilities of no transition, for two-photon resonance at all times, with constant functions $\omega_1(z) = \alpha_1$, $\omega_2(z) = \alpha_2$, and $\delta_1(z) = \delta_1$. The angle Φ is defined by (16).

Initial occupation probabilities	Final occupation probabilities
1	$\frac{\alpha_2^4}{(\alpha_1^2+\alpha_2^2)^2} + \frac{\alpha_1^2\alpha_2^2}{(\alpha_1^2+\alpha_2^2)^2}\left[\left(1+\frac{\delta_1}{\Phi}\right)\cos\left(\frac{\Phi-\delta_1}{2}\right) + \left(1-\frac{\delta_1}{\Phi}\right)\cos\left(\frac{\Phi+\delta_1}{2}\right)\right] + \frac{\alpha_1^4}{2(\alpha_1^2+\alpha_2^2)\Phi^2}\left(\frac{\alpha_1^2+\alpha_2^2+2\delta_1^2}{\alpha_1^2+\alpha_2^2} + \cos\Phi\right)$
0	$\frac{1}{2}(\alpha_1^2/\Phi^2)(1-\cos\Phi)$
0	$\frac{\alpha_1^2\alpha_2^2}{(\alpha_1^2+\alpha_2^2)^2}\left[1-\left(1+\frac{\delta_1}{\Phi}\right)\cos\left(\frac{\Phi-\delta_1}{2}\right) - \left(1-\frac{\delta_1}{\Phi}\right)\cos\left(\frac{\Phi+\delta_1}{2}\right)\right] + \frac{\alpha_1^2\alpha_2^2}{2(\alpha_1^2+\alpha_2^2)\Phi^2}\left(\frac{\alpha_1^2+\alpha_2^2+2\delta_1^2}{\alpha_1^2+\alpha_2^2} + \cos\Phi\right)$
0	$\frac{1}{2}(\alpha_1^2/\Phi^2)(1-\cos\Phi)$
1	$[\alpha_1^2+\alpha_2^2+2\delta_1^2+(\alpha_1^2+\alpha_2^2)\cos\Phi]/(2\Phi^2)$
0	$\frac{1}{2}(\alpha_2^2/\Phi^2)(1-\cos\Phi)$
0	$\frac{\alpha_1^2\alpha_2^2}{(\alpha_1^2+\alpha_2^2)^2}\left[1-\left(1+\frac{\delta_1}{\Phi}\right)\cos\left(\frac{\Phi-\delta_1}{2}\right) - \left(1-\frac{\delta_1}{\Phi}\right)\cos\left(\frac{\Phi+\delta_1}{2}\right)\right] + \frac{\alpha_1^2\alpha_2^2}{2(\alpha_1^2+\alpha_2^2)\Phi^2}\left(\frac{\alpha_1^2+\alpha_2^2+2\delta_1^2}{\alpha_1^2+\alpha_2^2} + \cos\Phi\right)$
0	$\frac{1}{2}(\alpha_2^2/\Phi^2)(1-\cos\Phi)$
1	$\frac{\alpha_1^4}{(\alpha_1^2+\alpha_2^2)^2} + \frac{\alpha_1^2\alpha_2^2}{(\alpha_1^2+\alpha_2^2)^2}\left[\left(1+\frac{\delta_1}{\Phi}\right)\cos\left(\frac{\Phi-\delta_1}{2}\right) + \left(1-\frac{\delta_1}{\Phi}\right)\cos\left(\frac{\Phi+\delta_1}{2}\right)\right] + \frac{\alpha_2^4}{2(\alpha_1^2+\alpha_2^2)\Phi^2}\left(\frac{\alpha_1^2+\alpha_2^2+2\delta_1^2}{\alpha_1^2+\alpha_2^2} + \cos\Phi\right)$

$a_3 = -\Omega_1\Omega_2/(\Omega_1^2+\Omega_2^2) + [\Omega_1\Omega_2/(\Omega_1^2+\Omega_2^2)]f(t)$.

Here, $f(t)$ and $g(t)$ are undetermined functions, which may have complex values; they satisfy $f(0)=1$ and $g(0)=0$. For complete transfer of the occupation probability from state 1 to state 2, there must be some time t such that $f(t)$ is real and positive and negative. Since this is impossible, complete transfer from state 1 to state 2 is ruled out. Similarly, complete transfer from state 3 to state 2 can be ruled out. Similar calculations show that complete transfer from state 2 to state 1 or state 3 can be ruled out. An explicit consideration of the differential equations can be used to rule out complete transfer and complete return for a wide class of pulse shapes. These calculations are not given here, but Robinson[20] gives some relevant calculations.

B. Analytic solutions

Analytic solutions for three special cases are given here. We increase the generality of the solutions by using (6) in place of (3). The functions ω_1, ω_2, and $\delta_1 = -\delta_2$, which appear in (6), are to be chosen.

In the simplest case, these three functions are constant, and we obtain a generalization of the Rabi solution.[7] We may assume that $z(t)$ increases from 0 to 1 as t increases from $-\infty$ to $+\infty$. The two dimensionless pulse areas are

$$\alpha_1 = \int_0^1 \omega_1\,dz = \omega_1, \quad \alpha_2 = \int_0^1 \omega_2\,dz = \omega_2.$$

We assume that α_1 and α_2 are positive. The solutions of (6) are briefly described in Appendix F, and the resulting transition probabilities are listed in Table VII, where the angle

$$\Phi = (\alpha_1^2+\alpha_2^2+\delta_1^2)^{1/2} \tag{16}$$

appears. Since complete transfer from state 1 or state 3 to state 2 has been ruled out, Table VII shows that a necessary condition for complete transfer or complete return is that Φ is a multiple of 2π. In fact, complete return to state 2 is obtained whenever Φ is a multiple of 2π. For complete return to state 1, Φ and δ_1 and $(\Phi+\delta_1)/2$ must all be multiples of 2π. For complete transfer from state 1 to state 3, $\alpha_1=\alpha_2$ is necessary, and the other conditions can be derived from Table VII.

We shall also treat two cases in which solutions of (6) can be written in terms of hypergeometric functions. We assume that

TABLE VIII. Transition probabilities, and probabilities of no transition, for two-photon resonance at all times, with functions $\omega_1(z)$, $\omega_2(z)$, and $\delta_1(z)$ given by (17). We define $\Phi = \{[\alpha_1^2+\alpha_2^2+\pi^2(\beta^2-\gamma^2)]/8 + \frac{1}{8}[(\alpha_1^2+\alpha_2^2)^2 + 2\pi^2(\alpha_1^2+\alpha_2^2)(\beta^2-\gamma^2)+\pi^4(\beta^2+\gamma^2)^2]^{1/2}\}^{1/2}$ and $X = \mathrm{sgn}(\beta\gamma)\{-[\alpha_1^2+\alpha_2^2+\pi^2(\beta^2-\gamma^2)]/8 + \frac{1}{8}[(\alpha_1^2+\alpha_2^2)^2+2\pi^2(\alpha_1^2+\alpha_2^2)(\beta^2-\gamma^2)+\pi^4(\beta^2+\gamma^2)^2]^{1/2}\}^{1/2}$.

Initial occupation probabilities	Final occupation probabilities
1	$\dfrac{\alpha_2^4}{(\alpha_1^2+\alpha_2^2)^2} + \dfrac{2\alpha_1^2\alpha_2^2[\cos(\Phi-\frac{1}{2}\pi\beta)\sinh(X+\frac{1}{2}\pi\gamma) - \cos(\Phi+\frac{1}{2}\pi\beta)\sinh(X-\frac{1}{2}\pi\gamma)]}{(\alpha_1^2+\alpha_2^2)^2\sinh(\pi\gamma)} + \dfrac{\alpha_1^4}{(\alpha_1^2+\alpha_2^2)^2}\dfrac{\cosh(\pi\gamma)[\cos(2\Phi)+\cosh(2X)]-1-\cos(2\Phi)\cosh(2X)}{\sinh^2(\pi\gamma)}$
0	$\dfrac{\alpha_1^2[\cosh(\pi\gamma)-\cos(2\Phi)][\cosh(\pi\gamma)-\cosh(2X)]}{(\alpha_1^2+\alpha_2^2)\sinh^2(\pi\gamma)}$
0	$\dfrac{\alpha_1^2\alpha_2^2}{(\alpha_1^2+\alpha_2^2)^2}\left\{1+\dfrac{2\cos(\Phi+\frac{1}{2}\pi\beta)\sinh(X-\frac{1}{2}\pi\gamma)-2\cos(\Phi-\frac{1}{2}\pi\beta)\sinh(X+\frac{1}{2}\pi\gamma)}{\sinh(\pi\gamma)} + \dfrac{\cosh(\pi\gamma)[\cos(2\Phi)+\cosh(2X)]-1-\cos(2\Phi)\cosh(2X)}{\sinh^2(\pi\gamma)}\right\}$
0	$\dfrac{\alpha_1^2[\cosh(\pi\gamma)-\cos(2\Phi)][\cosh(\pi\gamma)-\cosh(2X)]}{(\alpha_1^2+\alpha_2^2)\sinh^2(\pi\gamma)}$
1	$\dfrac{\cosh(\pi\gamma)[\cos(2\Phi)+\cosh(2X)]-1-\cos(2\Phi)\cosh(2X)}{\sinh^2(\pi\gamma)}$
0	$\dfrac{\alpha_2^2[\cosh(\pi\gamma)-\cos(2\Phi)][\cosh(\pi\gamma)-\cosh(2X)]}{(\alpha_1^2+\alpha_2^2)\sinh^2(\pi\gamma)}$
0	$\dfrac{\alpha_1^2\alpha_2^2}{(\alpha_1^2+\alpha_2^2)^2}\left\{1+\dfrac{2\cos(\Phi+\frac{1}{2}\pi\beta)\sinh(X-\frac{1}{2}\pi\gamma)-2\cos(\Phi-\frac{1}{2}\pi\beta)\sinh(X+\frac{1}{2}\pi\gamma)}{\sinh(\pi\gamma)} + \dfrac{\cosh(\pi\gamma)[\cos(2\Phi)+\cosh(2X)]-1-\cos(2\Phi)\cosh(2X)}{\sinh^2(\pi\gamma)}\right\}$
0	$\dfrac{\alpha_2^2[\cosh(\pi\gamma)-\cos(2\Phi)][\cosh(\pi\gamma)-\cosh(2X)]}{(\alpha_1^2+\alpha_2^2)\sinh^2(\pi\gamma)}$
1	$\dfrac{\alpha_1^4}{(\alpha_1^2+\alpha_2^2)^2} + \dfrac{2\alpha_1^2\alpha_2^2[\cos(\Phi-\frac{1}{2}\pi\beta)\sinh(X+\frac{1}{2}\pi\gamma) - \cos(\Phi+\frac{1}{2}\pi\beta)\sinh(X-\frac{1}{2}\pi\gamma)]}{(\alpha_1^2+\alpha_2^2)^2\sinh(\pi\gamma)} + \dfrac{\alpha_2^4}{(\alpha_1^2+\alpha_2^2)^2}\dfrac{\cosh(\pi\gamma)[\cos(2\Phi)+\cosh(2X)]-1-\cos(2\Phi)\cosh(2X)}{\sinh^2(\pi\gamma)}$

$$\omega_1(z)/\omega_2(z) = \Omega_1(t)/\Omega_2(t)$$

is constant. Hence, we can divide

$$\left[\frac{d}{dz} - i\delta_1(z)\right]a_2 = \frac{1}{2}i\omega_1(z)a_1 + \frac{1}{2}i\omega_2(z)a_3$$

by $\omega_1(z)$ or $\omega_2(z)$, differentiate, and obtain a second-order differential equation for $a_2(z)$. We now assume that the singular points of this differential equation are at $z = i, z = -i$, and $z = \infty$, as in recent work on the two-state problem.[10] As t increases from $-\infty$ to $+\infty$, $z(t)$ runs along the real axis from $-\infty$ to $+\infty$. To obtain hypergeometric functions, we assume

$$\omega_1(z) = \alpha_1/\pi(z^2+1), \quad \omega_2(z) = \alpha_2/\pi(z^2+1),$$

and

$$\delta_1(z) = (\beta + \gamma z)/(z^2+1). \tag{17}$$

The dimensionless areas are α_1 and α_2, which are positive. Solutions of (6) are written in Appendix F, and the resulting transition probabilities are listed in Table VIII. The necessary condition for complete transfer or complete return is that $\gamma = 0$, and $\gamma = 0$ implies that $\omega_1(z)$, $\omega_2(z)$, and $\delta_1(z)$ are all proportional to $(z^2+1)^{-1}$. By using $z' = \arctan(z)$ as a new independent variable, we return to the simple case treated in the previous paragraph. Hence, for (17), the cases of complete transfer and complete return have already been treated.

In the earlier applications of the hypergeometric function to the two-state problem,[9,21] the arbitrary function that we call $z(t)$ is the argument of the hypergeometric function, and it increases from 0 to 1 as t increases from $-\infty$ to $+\infty$. To obtain the corresponding solution of (6), we derive a second-order differential equation in the way indicated above, and put the singular points at $z = 0$, $z = 1$, and $z = \infty$. Let

$$\omega_1(z) = (\alpha_1/\pi)[z(1-z)]^{-1/2},$$

$$\omega_2(z) = (\alpha_2/\pi)[z(1-z)]^{-1/2},$$

and

$$\delta(z) = \beta/z + \gamma/(1-z). \tag{18}$$

Although this is a special case of (8), Clausen's function cannot be applied here. Solutions of (6) are written in Appendix F, in terms of hypergeometric functions. To find the limits of $a_1(z)$ and $a_3(z)$ as $z \to 1$, which corresponds to the end of the two optical pulses, we use the Gaussian formula[22] for the sum of the hypergeometric series at $z = 1$. The resulting gamma functions do not disappear when we calculate the transition probabilities; see Table IX, where the notation

$$\Phi = \tfrac{1}{2}[\alpha_1^2 + \alpha_2^2 - \pi^2(\beta-\gamma)^2]^{1/2} \tag{19}$$

is used.

The special cases in which evaluation of gamma functions is not necessary for computation of these transition probabilities should be mentioned. One of them is a limiting case of the problem treated in Sec. III. Recall that the conditions listed in Table I lead to final occupation probabilities that are elementary functions of the parameters (9). In the first section of Table I, $\gamma_1 + \gamma_2 = 0$ holds, and we now let $\beta_1 + \beta_2 \to 0$. This limiting process gives two-photon resonance at all times. It also gives

$$\alpha_1 = \alpha_2, \quad \beta_1 + \gamma_1 = 0, \quad \beta_2 + \gamma_2 = 0, \tag{20}$$

so that the two amplitude functions, $\Omega_1(t)$ and $\Omega_2(t)$, are equal. Furthermore, if $\Omega_1(t) = \Omega_2(t)$ is an even function, $\Delta_1(t) = -\Delta_2(t)$ is odd. From the second section of Table I, we obtain (20) by a similar limiting process. This examination of Table I leads us to set $\alpha_1 = \alpha_2$ and $\beta + \gamma = 0$ in Table IX, which simplifies the formulas for all nine final occupation probabilities. The same simplified formulas are obtained from Tables II and III, by using (20). The angles Φ, defined in Tables II and III and in (19), are all the same in this limit. We conclude that the nine final occupation probabilities for the case (20) are independent of whether this case is treated directly or as a limit. However, the calculations in Appendices C and E can hardly be made applicable to this limiting case.

There is perhaps another limiting case that we could derive from Sec. III. Assuming that Clausen's series always terminates, we could seek the limit of two-photon resonance at all times. We have not done this.

Finally, we should mention the conditions for complete transfer and complete return when (18) is used. For complete return of the occupation probability to state 2, the conditions are that $\beta = \gamma$ and that Φ is a multiple of π; see Table IX. For complete return to state 1, or complete transfer from state 1 to state 3, these two conditions are necessary but not sufficient. They make the hypergeometric series terminate, so that a table similar to Table V could be constructed. From $\beta = \gamma$, we find that, if $\Omega_1(t)$ and $\Omega_2(t)$ are even functions,

TABLE IX. Transition probabilities, and probabilities of no transition, for two-photon resonance at all times, with functions $\omega_1(z)$, $\omega_2(z)$, and $\delta_1(z)$ given by (18). The angle Φ, defined by (19), can be real or pure imaginary. We define R as the real part of $[\alpha_1^2\alpha_2^2/(\alpha_1^2+\alpha_2^2)^2]\,\Gamma(\frac{1}{2} - i\beta)\Gamma(\frac{1}{2} - i\gamma)/\Gamma[\frac{1}{2} - \frac{1}{2}i(\beta+\gamma) + \Phi/\pi]\Gamma[\frac{1}{2} - \frac{1}{2}i(\beta+\gamma) - \Phi/\pi]$.

Initial occupation probabilities	Final occupation probabilities
1	$\dfrac{\alpha_2^4}{(\alpha_1^2+\alpha_2^2)^2} + \dfrac{\alpha_1^4[\cosh(\pi\beta+\pi\gamma)+\cos(2\Phi)]}{2(\alpha_1^2+\alpha_2^2)^2\cosh(\pi\beta)\cosh(\pi\gamma)} + 2R$
0	$\dfrac{\alpha_1^2[\cosh(\pi\beta-\pi\gamma)-\cos(2\Phi)]}{2(\alpha_1^2+\alpha_2^2)\cosh(\pi\beta)\cosh(\pi\gamma)}$
0	$\dfrac{\alpha_1^2\alpha_2^2}{(\alpha_1^2+\alpha_2^2)^2}\left[1 + \dfrac{\cosh(\pi\beta+\pi\gamma)+\cos(2\Phi)}{2\cosh(\pi\beta)\cosh(\pi\gamma)}\right] - 2R$
0	$\dfrac{\alpha_1^2[\cosh(\pi\beta-\pi\gamma)-\cos(2\Phi)]}{2(\alpha_1^2+\alpha_2^2)\cosh(\pi\beta)\cosh(\pi\gamma)}$
1	$\dfrac{\cosh(\pi\beta+\pi\gamma)+\cos(2\Phi)}{2\cosh(\pi\beta)\cosh(\pi\gamma)}$
0	$\dfrac{\alpha_2^2[\cosh(\pi\beta-\pi\gamma)-\cos(2\Phi)]}{2(\alpha_1^2+\alpha_2^2)\cosh(\pi\beta)\cosh(\pi\gamma)}$
0	$\dfrac{\alpha_1^2\alpha_2^2}{(\alpha_1^2+\alpha_2^2)^2}\left[1 + \dfrac{\cosh(\pi\beta+\pi\gamma)+\cos(2\Phi)}{2\cosh(\pi\beta)\cosh(\pi\gamma)}\right] - 2R$
0	$\dfrac{\alpha_2^2[\cosh(\pi\beta-\pi\gamma)-\cos(2\Phi)]}{2(\alpha_1^2+\alpha_2^2)\cosh(\pi\beta)\cosh(\pi\gamma)}$
1	$\dfrac{\alpha_1^4}{(\alpha_1^2+\alpha_2^2)^2} + \dfrac{\alpha_2^4[\cosh(\pi\beta+\pi\gamma)+\cos(2\Phi)]}{2(\alpha_1^2+\alpha_2^2)^2\cosh(\pi\beta)\cosh(\pi\gamma)} + 2R$

$\Delta_1(t) = -\Delta_2(t)$ is also an even function.

This section has treated the simple case of two-photon resonance at all times, which is closely connected with the phenomenon of coherent trapping and with calculations on the two-state problem. One special case of two-photon resonance at all times appears as a simple limiting case of our calculations in Secs. III A and III B.

V. CONCLUSION

The cases of the three-state problem that we have treated here show a variety of atomic or molecular responses to two optical pulses that overlap in time. Complete transfer to another state, or complete inversion, is obtainable in many cases that were not covered in previous calculations. The results suggest that experiments with two laser beams can explore an interesting area of quantum dynamics that lies beyond earlier calculations.

ACKNOWLEDGMENTS

We thank Professor R. Askey for the proof that appears in Appendix D.

This research was partially supported by the Division of Chemical Sciences, Office of Basic Energy Sciences, U.S. Department of Energy, under Grant No. DE-FG02-84-ER13243.

APPENDIX A: THIRD-ORDER DIFFERENTIAL EQUATIONS

In order to apply Clausen's function or another special function to our three-state model, it is convenient to have third-order ordinary differential equations in one dependent variable. We can derive such equations from (5). Elimination of $a_2(z)$ and $a_3(z)$ gives

$$\left[\frac{d}{dz}-\frac{i(\delta_1+2\delta_2)}{3}\right]\frac{1}{\omega_2}\left[\frac{d}{dz}-\frac{i(\delta_1-\delta_2)}{3}\right]\frac{1}{\omega_1}\left[\frac{d}{dz}+\frac{i(2\delta_1+\delta_2)}{3}\right]a_1$$
$$+\frac{1}{4}\left[\frac{d}{dz}-\frac{i(\delta_1+2\delta_2)}{3}\right]\left(\frac{\omega_1}{\omega_2}\right)a_1+\frac{1}{4}\left(\frac{\omega_2}{\omega_1}\right)\left[\frac{d}{dz}+\frac{i(2\delta_1+\delta_2)}{3}\right]a_1=0, \tag{A1}$$

where δ_1, δ_2, ω_1, ω_2 are functions of z. Elimination of $a_1(z)$ and $a_3(z)$ is not so straightforward, and is not needed for our calculations. Elimination of $a_1(z)$ and $a_2(z)$ gives

$$\left[\frac{d}{dz}+\frac{i(2\delta_1+\delta_2)}{3}\right]\frac{1}{\omega_1}\left[\frac{d}{dz}-\frac{i(\delta_1-\delta_2)}{3}\right]\frac{1}{\omega_2}\left[\frac{d}{dz}-\frac{i(\delta_1+2\delta_2)}{3}\right]a_3$$
$$+\frac{1}{4}\left[\frac{d}{dz}+\frac{i(2\delta_1+\delta_2)}{3}\right]\left(\frac{\omega_2}{\omega_1}\right)a_3+\frac{1}{4}\left(\frac{\omega_1}{\omega_2}\right)\left[\frac{d}{dz}-\frac{i(\delta_1+2\delta_2)}{3}\right]a_3=0. \tag{A2}$$

The derivatives of ω_1 and ω_2 occur in (A1) and (A2). We shall write these equations in terms of

$$L_1=\frac{d}{dz}\ln\omega_1, \quad L_2=\frac{d}{dz}\ln\omega_2,$$

and their first derivatives with respect to z. We arrange to have L_1, L_2, and their derivatives appear in coefficients that stand to the left of the three differential operators that appear in square brackets in (A1) and (A2). Next, elimination of these differential operators gives

$$\frac{d^3a_1}{dz^3}-(2L_1+L_2)\frac{d^2a_1}{dz^2}+\left\{L_1^2+L_1L_2-\frac{dL_1}{dz}-\frac{i\delta_1}{3}(2L_1+L_2)-\frac{i\delta_2}{3}(L_1+2L_2)+i\left[\frac{d}{dz}(\delta_1+\delta_2)\right]\right.$$
$$\left.+\frac{\delta_1^2+\delta_1\delta_2+\delta_2^2}{3}+\frac{(\omega_1)^2+(\omega_2)^2}{4}\right\}\frac{da_1}{dz}+\left\{\frac{i}{3}(2\delta_1+\delta_2)\left(L_1^2+L_1L_2-\frac{dL_1}{dz}\right)-\frac{i}{3}\left[\frac{d}{dz}(2\delta_1+\delta_2)\right](2L_1+L_2)\right.$$
$$-\frac{1}{9}(2\delta_1+\delta_2)\left[(2\delta_1+\delta_2)L_1+(\delta_1-\delta_2)L_2\right]+\frac{i}{3}\left[\frac{d^2}{dz^2}(2\delta_1+\delta_2)\right]+\frac{1}{3}(2\delta_1+\delta_2)\frac{d\delta_1}{dz}$$
$$\left.-\frac{i}{27}(2\delta_1+\delta_2)(\delta_1-\delta_2)(\delta_1+2\delta_2)-\frac{i\delta_1}{12}\left[(\omega_1)^2-2(\omega_2)^2\right]-\frac{i\delta_2}{12}\left[2(\omega_1)^2-(\omega_2)^2\right]+\frac{1}{4}(\omega_1)^2(L_1-L_2)\right\}a_1=0 \tag{A3}$$

and

$$\frac{d^3a_3}{dz^3}-(L_1+2L_2)\frac{d^2a_3}{dz^2}+\left\{L_1L_2+L_2^2-\frac{dL_2}{dz}+\frac{i\delta_1}{3}(2L_1+L_2)+\frac{i\delta_2}{3}(L_1+2L_2)-i\left[\frac{d}{dz}(\delta_1+\delta_2)\right]\right.$$
$$\left.+\frac{\delta_1^2+\delta_1\delta_2+\delta_2^2}{3}+\frac{(\omega_1)^2+(\omega_2)^2}{4}\right\}\frac{da_3}{dz}+\left\{-\frac{i}{3}(\delta_1+2\delta_2)\left(L_1L_2+L_2^2-\frac{dL_2}{dz}\right)\right.$$
$$+\frac{i}{3}\left[\frac{d}{dz}(\delta_1+2\delta_2)\right](L_1+2L_2)+\frac{1}{9}(\delta_1+2\delta_2)\left[(\delta_1-\delta_2)L_1-(\delta_1+2\delta_2)L_2\right]-\frac{i}{3}\left[\frac{d^2}{dz^2}(\delta_1+2\delta_2)\right]$$
$$+\frac{1}{3}(\delta_1+2\delta_2)\frac{d\delta_2}{dz}-\frac{i}{27}(2\delta_1+\delta_2)(\delta_1-\delta_2)(\delta_1+2\delta_2)-\frac{i\delta_1}{12}\left[(\omega_1)^2-2(\omega_2)^2\right]$$

$$-\frac{i\delta_2}{12}[2(\omega_1)^2-(\omega_2)^2]-\frac{1}{4}(\omega_2)^2(L_1-L_2)\Big\}a_3=0. \tag{A4}$$

In these equations, the coefficients of d^2a_1/da^2 and d^2a_3/dz^2 are real when z is real, and do not involve the detunings. This simplification is a consequence of the vanishing trace of the Hamiltonian matrix in (2).

We plan to choose the functions $\delta_1, \delta_2, \omega_1$, and ω_2 so that (A3) and (A4) can be treated analytically. In this paper, we use the choice (8), which gives

$$\begin{aligned}
&\frac{d^3a_1}{dz^3}+\left[\frac{(3/2)}{z}+\frac{(3/2)}{z-1}\right]\frac{d^2a_1}{dz^2}+\left\{\frac{(\beta_1^2+\beta_1\beta_2+\beta_2^2)/3-i(\beta_1+\beta_2)/2}{z^2}\right.\\
&\quad+\left[1-\frac{\alpha_1^2+\alpha_2^2}{4\pi^2}-\frac{2\beta_1\gamma_1+\beta_1\gamma_2+\beta_2\gamma_1+2\beta_2\gamma_2}{3}+\frac{i(\beta_1+\beta_2-\gamma_1-\gamma_2)}{2}\right]\frac{1}{z(z-1)}\\
&\quad\left.+\frac{(\gamma_1^2+\gamma_1\gamma_2+\gamma_2^2)/3+i(\gamma_1+\gamma_2)/2}{(z-1)^2}\right\}\frac{da_1}{dz}+\frac{1}{3}\left\{(2\beta_1+\beta_2)\,\frac{(-\beta_1+i)/2-i(\beta_1-\beta_2)(\beta_1+2\beta_2)/9}{z^3}\right.\\
&\quad+\left[\frac{(\beta_1-i)(2\beta_1+\beta_2)+\beta_1\gamma_2-\beta_2\gamma_1}{2}+\frac{i(2\beta_1^2+2\beta_1\beta_2-\beta_2^2)\gamma_1}{3}+\frac{i(\beta_1^2-2\beta_1\beta_2-2\beta_2^2)\gamma_2}{3}\right.\\
&\quad\left.+\frac{i\alpha_1^2}{4\pi^2}(\beta_1+2\beta_2)-\frac{i\alpha_2^2}{4\pi^2}(2\beta_1+\beta_2)\right]\frac{1}{z^2(z-1)}+\left[\frac{(\gamma_1+i)(2\gamma_1+\gamma_2)-(\beta_1\gamma_2-\beta_2\gamma_1)}{2}-\frac{i\beta_1(2\gamma_1^2+2\gamma_1\gamma_2-\gamma_2^2)}{3}\right.\\
&\quad\left.-\frac{i\beta_2(\gamma_1^2-2\gamma_1\gamma_2-2\gamma_2^2)}{3}-\frac{i\alpha_1^2}{4\pi^2}(\gamma_1+2\gamma_2)+\frac{i\alpha_2^2}{4\pi^2}(2\gamma_1+\gamma_2)\right]\frac{1}{z(z-1)^2}\\
&\quad\left.+(2\gamma_1+\gamma_2)\,\frac{-(\gamma_1+i)/2+i(\gamma_1-\gamma_2)(\gamma_1+2\gamma_2)/9}{(z-1)^3}\right\}a_1=0
\end{aligned} \tag{A5}$$

and

$$\begin{aligned}
&\frac{d^3a_3}{dz^3}+\left[\frac{(3/2)}{z}+\frac{(3/2)}{z-1}\right]\frac{d^2a_3}{dz^2}+\left\{\frac{(\beta_1^2+\beta_1\beta_2+\beta_2^2)/3+i(\beta_1+\beta_2)/2}{z^2}\right.\\
&\quad+\left[1-\frac{\alpha_1^2+\alpha_2^2}{4\pi^2}-\frac{2\beta_1\gamma_1+\beta_1\gamma_2+\beta_2\gamma_1+2\beta_2\gamma_2}{3}-\frac{i(\beta_1+\beta_2-\gamma_1-\gamma_2)}{2}\right]\frac{1}{z(z-1)}\\
&\quad\left.+\frac{(\gamma_1^2+\gamma_1\gamma_2+\gamma_2^2)/3-i(\gamma_1+\gamma_2)/2}{(z-1)^2}\right\}\frac{da_3}{dz}+\frac{1}{3}\left\{(\beta_1+2\beta_2)\frac{-(\beta_2+i)/2-i(\beta_1-\beta_2)(2\beta_1+\beta_2)/9}{z^3}\right.\\
&\quad+\left[\frac{(\beta_2+i)(\beta_1+2\beta_2)-(\beta_1\gamma_2-\beta_2\gamma_1)}{2}+\frac{i(2\beta_1^2+2\beta_1\beta_2-\beta_2^2)\gamma_1}{3}+\frac{i(\beta_1^2-2\beta_1\beta_2-2\beta_2^2)\gamma_2}{3}\right.\\
&\quad\left.+\frac{i\alpha_1^2}{4\pi^2}(\beta_1+2\beta_2)-\frac{i\alpha_2^2}{4\pi^2}(2\beta_1+\beta_2)\right]\frac{1}{z^2(z-1)}+\left[\frac{(\gamma_2-i)(\gamma_1+2\gamma_2)+\beta_1\gamma_2-\beta_2\gamma_1}{2}\right.\\
&\quad\left.-\frac{i\beta_1(2\gamma_1^2+2\gamma_1\gamma_2-\gamma_2^2)}{3}-\frac{i\beta_2(\gamma_1^2-2\gamma_1\gamma_2-2\gamma_2^2)}{3}-\frac{i\alpha_1^2}{4\pi^2}(\gamma_1+2\gamma_2)+\frac{i\alpha_2^2}{4\pi^2}(2\gamma_1+\gamma_2)\right]\frac{1}{z(z-1)^2}\\
&\quad\left.+(\gamma_1+2\gamma_2)\,\frac{(-\gamma_2+i)/2+i(2\gamma_1+\gamma_2)(\gamma_1-\gamma_2)/9}{(z-1)^3}\right\}a_3=0.
\end{aligned} \tag{A6}$$

These two differential equations belong to the class discussed in the following Appendix.

APPENDIX B: THREE REGULAR SINGULAR POINTS

Here, we consider homogeneous linear differential equations with only three singular points, all of which are regular. We emphasize third-order differential equations. Examples are (A5) and (A6) and

$$\begin{aligned}
&\frac{d^3F}{dx^3}+\frac{[(3+\alpha'+\beta'+\delta')x-(1+\gamma'+\epsilon')]}{x(x-1)}\frac{d^2F}{dx^2}\\
&\quad+\frac{[(1+\alpha'+\beta'+\delta'+\alpha'\beta'+\alpha'\delta'+\beta'\delta')x-\gamma'\epsilon']}{x^2(x-1)}\frac{dF}{dx}+\frac{\alpha'\beta'\delta'}{x^2(x-1)}F=0,
\end{aligned} \tag{B1}$$

which is the differential equation satisfied by (7). The mathematical theory will tell us when solutions of (A5) and (A6) can be written in terms of Clausen's function, and will remove the need for laborious transformation of (A5) and (A6) into the form (B1), or vice versa.

We may mention the second-order differential equation with three singular points, all regular. Its general form is

$$\frac{d^2y}{dz^2}+\left(\frac{1-\alpha-\alpha'}{z-a}+\frac{1-\beta-\beta'}{z-b}+\frac{1-\gamma-\gamma'}{z-c}\right)\frac{dy}{dz}$$
$$+\frac{1}{(z-a)(z-b)(z-c)}\left[\frac{\alpha\alpha'(a-b)(a-c)}{z-a}+\frac{\beta\beta'(b-c)(b-a)}{z-b}+\frac{\gamma\gamma'(c-a)(c-b)}{z-c}\right]y=0. \tag{B2}$$

This is called the Riemann–Papperitz equation, because Riemann discussed it,[23] and Papperitz later wrote it explicitly.[24] In (B2), the three singular points are a, b, and c. The six exponents are $\alpha, \alpha', \beta, \beta', \gamma$, and γ'. These nine complex parameters are arbitrary except for the relation

$$\alpha+\alpha'+\beta+\beta'+\gamma+\gamma'=1,$$

and the requirement that a, b, and c are distinct points. The exponents α', β', and γ' should not be confused with the parameters in (7). Furthermore, a, b, and c need not all be finite, and it is often convenient to use the limit of (B2) as $c\to\infty$. We emphasize two simple features of (B2): The coefficients of y and dy/dz are completely determined by the locations of the singular points and the exponents at each singular point. Also, solutions of (B2) can always be written in terms of hypergeometric functions. Properties corresponding to these two features are absent in the case of third-order equations, to which we now turn.

The third-order equation with only three singular points, all of which are regular, has the form

$$\frac{d^3y}{dz^3}+\left(\frac{3-\alpha-\alpha'-\alpha''}{z-a}+\frac{3-\beta-\beta'-\beta''}{z-b}+\frac{3-\gamma-\gamma'-\gamma''}{z-c}\right)\frac{d^2y}{dz^2}+\left(\cdots\right)\frac{dy}{dz}+\left(\cdots\right)y=0,$$

where the singular points are a, b, and c; the coefficients of y and dy/dz are rational functions that we need not find, because we shall consider only the case in which c is infinite. A change of the variable called z has the same effect as letting $c\to\infty$. In the early work of Fuchs[25] and other treatments[26] of the theory of differential equations in which all singular points are regular, one singular point is put at infinity. Fuchs[25] found the sum of all the exponents, among other results. In our case,

$$\alpha+\alpha'+\alpha''+\beta+\beta'+\beta''+\gamma+\gamma'+\gamma''=3. \tag{B3}$$

In this paper, the singular points are at 0, 1, and ∞. Hence the general form of the differential equation is

$$\frac{d^3y}{dz^3}+\left(\frac{3-\alpha-\alpha'-\alpha''}{z}+\frac{3-\beta-\beta'-\beta''}{z-1}\right)\frac{d^2y}{dz^2}$$
$$+\left[\frac{M_1}{z^2}+\frac{M_2}{z(z-1)}+\frac{M_3}{(z-1)^2}\right]\frac{dy}{dz}+\left[\frac{N_1}{z^3}+\frac{N_2}{z^2(z-1)}+\frac{N_3}{z(z-1)^2}+\frac{N_4}{(z-1)^3}\right]y=0. \tag{B4}$$

Equations (A5) and (A6) have this form, and we could write (B1) in this form. Given any differential equation of the form (B4), (B3) can be used to find $\gamma+\gamma'+\gamma''$ easily. Further calculation gives α, α', and α'' in terms of M_1 and N_1, and gives β, β', β'', and so forth. We could calculate the nine exponents, and then find M_1, M_2, M_3, N_1, N_2+N_3, and N_4 in terms of the nine exponents. We emphasize that the exponents do not determine N_2 nor N_3 separately. The exponents do not fully describe the form of the differential equation, and they cannot fully determine any solution. We shall use the exponents and other information to decide when solutions of (B4) can be written in terms of Clausen's function. We shall need the exponents of (A5), (A6), and (B1); they are listed in Table X.

Clausen's differential equation, written as (B1), has a peculiarity that appears in Table X: Two of the exponents at $x=1$ differ by unity. This difference cannot be changed by any simple transformation of dependent or independent variables, and it distinguishes one of the singular points of (B1) from the other two. We do not have the full symmetry under interchange of singular points that appeared in (B2), and this leads to fewer useful formulas for Clausen's function than for the hypergeometric function. In order to apply Clausen's function to (A5) and (A6), we must have a difference of unity between two exponents at $z=0$ or at $z=1$ or at $z=\infty$. The third alternative leads to complications mentioned at the end of this Appendix. The first two alternatives lead to (14) and (13) as possible relations between x and z. Then, Table X gives (11) and (10), which are necessary conditions for use of Clausen's function. We still have to show that either condition is sufficient, or that it allows us to write solutions of (A5) and (A6) in terms of Clausen's function.

TABLE X. Exponents of (A5), (A6), and (B1), the third-order differential equations for $a_1(z)$, $a_3(z)$, and Clausen's function. The definition of ρ_1, ρ_2, and ρ_3 is given by (C1) and (C2).

Singular point	Three exponents at the singular point
Equation (A5):	
$z=0$	$-i(2\beta_1+\beta_2)/3,\frac{1}{2}+i(\beta_1-\beta_2)/3,1+i(\beta_1+2\beta_2)/3$
$z=1$	$i(2\gamma_1+\gamma_2)/3,\frac{1}{2}-i(\gamma_1-\gamma_2)/3,1-i(\gamma_1+2\gamma_2)/3$
$z=\infty$	$i\rho_1,i\rho_2,i\rho_3$
Equation (A6):	
$z=0$	$i(\beta_1+2\beta_2)/3,\frac{1}{2}+i(\beta_1-\beta_2)/3,1-i(2\beta_1+\beta_2)/3$
$z=1$	$-i(\gamma_1+2\gamma_2)/3,\frac{1}{2}-i(\gamma_1-\gamma_2)/3,1+i(2\gamma_1+\gamma_2)/3$
$z=\infty$	$i\rho_1,i\rho_2,i\rho_3$
Equation (B1):	
$x=0$	$0,1-\gamma',1-\epsilon'$
$x=1$	$0,1,\gamma'+\epsilon'-\alpha'-\beta'-\delta'$
$x=\infty$	α',β',δ'

In using Table X to derive (10) and (11), we have tacitly assumed that $\alpha'+\beta'+\delta'-\gamma'-\epsilon'$ is not an integer. If all the differences of exponents at $x=1$ of (B1) are integers, we have irrelevant complications. A similar assumption applies to the exponents at $x=0$: We assume that there are no integers among γ', ϵ', and $\gamma'-\epsilon'$.

The logarithm of $x-1$ could appear, in principle, when we seek solutions of (B1). We can abbreviate our calculations by considering whether the logarithm does appear, although the long account given by Forsyth[26] indicates that the question is sometimes complicated. Since $\alpha'+\beta'+\delta'-\gamma'-\epsilon'$ is not an integer, squares and higher powers of $\ln(x-1)$ cannot appear. Clausen's equation, (B1), has a solution of the form

$$F=A_0+A_1(x-1)+B_1(x-1)\ln(x-1)+A_2(x-1)^2+B_2(x-1)^2\ln(x-1)+\cdots. \tag{B5}$$

Substitution into (B1) gives linear equations to determine A_0, A_1, B_1, and so forth. A simple calculation gives

$$B_1=0. \tag{B6}$$

Since the difference of exponents at $x=1$ is unity rather than a larger integer, we obtain

$$B_2=0,\quad B_3=0,\ldots,$$

without further calculations.[26] Since $\alpha'+\beta'+\delta'-\gamma'-\epsilon'$ is not an integer, the solution of (B1) that has the form

$$(x-1)^{\gamma'+\epsilon'-\alpha'-\beta'-\delta'}[C_0+C_1(x-1)+C_2(x-1)^2+\cdots]$$

is also free of logarithms. Hence, the logarithm of $x-1$ does not appear in solutions of (B1). This disappearance of the logarithm is the feature that distinguishes Clausen's differential equation from other differential equations with the same singular points and the same exponents.

We now ask whether logarithms appear in any solution of (B4), assuming that the difference of two exponents is unity. The exponents do not determine N_2 and N_3 separately, but a requirement that the logarithms disappear will determine N_2 and N_3, and allow solutions to be written in terms of Clausen's function. To state the plan for solving (B4) precisely, we must specify which singular point gives exponents that differ by unity. Suppose that two of the exponents of (B4) at $z=1$ differ by unity, and that $\ln(z-1)$ does not appear when we find solutions of (B4) in ascending powers of $z-1$. Then, (B4) has a solution of the form

$$z^{\lambda}(z-1)^{\mu}F(\alpha',\beta',\delta';\gamma',\epsilon';z). \tag{B7}$$

Similar calculations can start with another singular point. Suppose that two of the exponents of (B4) at $z=0$ differ by unity, and that $\ln z$ does not appear when we find solutions of (B4) in ascending powers of z. Then, (B4) has a solution of the form

$$z^{\mu}(z-1)^{\lambda}F(\alpha',\beta',\delta';\gamma',\epsilon';1-z). \tag{B8}$$

Three independent solutions in terms of Clausen's function can be obtained if the conditions given above are satisfied. We assume, in the example leading to (B7), that one of the differences of exponents at $z=1$ is unity, and that no differences of exponents at $z=0$ are integers. Hence, it is possible to find three solutions in ascending powers of z. Each has the form (B7), where λ is one of the exponents at $z=0$. Since there are three different exponents at $z=0$, we obtain three independent solutions of the form (B7). Furthermore, the other parameters in (B7) can be obtained from the exponents of (B4). Since multiplication by z^{λ} shifts the three exponents at $z=0$, the three exponents of (B4) at $z=0$ must be λ, $\lambda+1-\gamma'$, and $\lambda+1-\epsilon'$. This determines γ' and ϵ'. Since multiplication by $(z-1)^{\mu}$ shifts the exponents at $z=1$, two of the exponents of (B4) at $z=1$ are μ and $\mu+1$; the third exponent is $\mu+\gamma'+\epsilon'-\alpha'-\beta'-\delta'$. Finally, the exponents of (B4) at $z=\infty$ are $\alpha'-\lambda-\mu$, $\beta'-\lambda-\mu$, and $\delta'-\lambda-\mu$. In this way, we determine each parameter in (B7), and also $\gamma'+\epsilon'-\alpha'-\beta'-\delta'$; the redundant condition is consistent because of (B3).

In this scheme for finding solutions in terms of Clausen's function, after deciding whether they exist, the search for terms in $\ln(x-1)$ is an essential step. This search will be simplified before we compute the solutions of (A5) and (A6). We consider the possible solutions of (B4), assuming that μ and $\mu+1$ are two of the exponents at $x=1$. The third exponent differs from μ by a number that is not an integer. Hence, one of the solutions has the form

$$(x-1)^{\mu}[A_0+A_1(x-1)+B_1(x-1)\ln(x-1)+\cdots], \tag{B9}$$

which is the generalization of (B5). We have to decide whether (B6) holds. The general condition for having all solutions free of logarithms was given by Frobenius.[27] However, we can use the simple method of Forsyth,[26] which is suggested by one of Cayley's papers.[28] Instead of using (B9), we substitute

$$(x-1)^{\mu}[A_0+A_1(x-1)+A_2(x-1)^2+\cdots]$$

into (B4). This gives homogeneous linear equations to determine A_0, A_1, A_2, and so forth. Suppose that the first two equations have the form

$$0\cdot A_0=0 \tag{B10a}$$

and

$$0\cdot A_0+0\cdot A_1=0, \tag{B10b}$$

so that A_0 and A_1 are not determined. Since $\mu+2,\mu+3,\mu+4,\ldots$ are not among the exponents, $A_2, A_3, A_4,\ldots$ are determined in terms of A_0 and A_1, which are two arbitrary constants. In this way, we obtain two independent solutions of (B4), in ascending powers of $x-1$. Both are free of logarithms, and the third solution is also free of logarithms. This implies that there are solutions of the form (B7) or (B8). We should emphasize that, if the equations for A_0 and A_1 do not have the form (B10), further calculation[26] is needed to determine whether logarithms appear or this simple method fails. However, we can apply this simple method to (A5) and (A6). We have required two exponents to differ by unity, and have found that this leads to (10) or (11). We write tentative solutions in ascending powers of $z-1$ or z, and require (B10) to hold. Since (B10a) is merely a check on the

calculation of the exponents, the condition for use of Clausen's function is (B10b). This condition is satisfied automatically when (10) or (11) holds. However, this condition is not satisfied automatically if $z=\infty$ corresponds to $x=1$, the singular point where two exponents differ by unity. This difficult case is not considered further.

This Appendix has explained the conditions under which solutions of a third-order differential equation can be written in terms of Clausen's function. Assuming that there are three singular points and that they are regular, we have nine exponents, eight of which can vary independently. Since only seven parameters appear in (B7) and (B8), the appearance of one condition on the exponents seems natural. We have emphasized that the exponents do not determine the form of the equation, and that disappearance of the logarithm is a condition for use of Clausen's function. The method of writing three independent solutions, when these two conditions are satisfied, has been given. Solutions of (A5) and (A6) are written in the following Appendix.

APPENDIX C: SOLUTIONS IN TERMS OF CLAUSEN'S FUNCTION

We can now proceed to construct solutions of (5), using the choice (8). The differential equations for $a_1(z)$ and $a_3(z)$ are (A5) and (A6), which can be solved without lengthy calculations by using the mathematical theory given in Appendix B. After finding general solutions for $a_1(z)$ and $a_3(z)$, differentiation formulas are used to find $a_2(z)$, the other component. If the arbitrary constants in the general solutions of (A5) and (A6) are adjusted to correspond, these two calculations of $a_2(z)$ will agree, and we shall have the general solution of (5), which is essentially the Schrödinger equation. However, we shall avoid very long formulas by writing three independent solutions of (5); a linear combination of them can be written out, if desired.

Our first step is to find the nine exponents of (A5) and (A6), which are not shown quite explicitly in Table X. For either differential equation, the three exponents at $z=\infty$ are the three roots of

$$\zeta^3+\left[\frac{(\beta_1-\gamma_1)^2+(\beta_1-\gamma_1)(\beta_2-\gamma_2)+(\beta_2-\gamma_2)^2}{3}-\frac{\alpha_1^2+\alpha_2^2}{4\pi^2}\right]\zeta$$
$$+\frac{i}{3}\left\{\frac{[2(\beta_1-\gamma_1)+(\beta_2-\gamma_2)][(\beta_1-\gamma_1)-(\beta_2-\gamma_2)][(\beta_1-\gamma_1)+2(\beta_2-\gamma_2)]}{9}\right.$$
$$\left.-\frac{\alpha_1^2}{4\pi^2}[(\beta_1-\gamma_1)+2(\beta_2-\gamma_2)]+\frac{\alpha_2^2}{4\pi^2}[2(\beta_1-\gamma_1)+(\beta_2-\gamma_2)]\right\}=0. \tag{C1}$$

The sum of the three roots is zero, because we have chosen to have vanishing trace of the Hamiltonian matrix in (2). If the three roots are not all pure imaginary, then one of them is pure imaginary and the real parts of the other two add up to zero. This suggests that we write the three roots of (C1) as

$$i\rho_1,\ i\rho_2,\ i\rho_3. \tag{C2}$$

Then,

$$\rho_1+\rho_2+\rho_3=0. \tag{C3}$$

Either the numbers ρ_1,ρ_2,ρ_3 are all real or two of them are complex conjugates. This determination of ρ_1,ρ_2, and ρ_3 makes Table X more explicit.

Two differentiation formulas for Clausen's function will be used here. A simple calculation shows that

$$(\alpha'\beta'\delta'/\gamma'\epsilon')\ F(\alpha'+1,\beta'+1,\delta'+1;\gamma'+1,\epsilon'+1;x)$$

is the derivative of (7). We shall also need

$$\left(x\frac{d}{dx}+\gamma'-1\right)F(\alpha',\beta',\delta';\gamma',\epsilon';x)=(\gamma'-1)\,F(\alpha',\beta',\delta';\gamma'-1,\epsilon';x).$$

This formula is given by Rainville,[29] along with a similar formula obtained by interchange of γ' and ϵ'.

Solutions of the Schrödinger equation can now be written explicitly, using (13) to connect (7) with (5), (A5) and (A6). We have derived (10) from (13), and have mentioned that (B10) is satisfied automatically. We could use (10) to rewrite (C1) in a shorter form. The exponents at $x=z=1$ give $\mu=i\gamma_1/3$ and

$$\alpha'+\beta'+\delta'-\gamma'-\epsilon'=i\gamma_1-\tfrac{1}{2}. \tag{C4}$$

The solutions of (5) follow.

Suppose that $|a_1(z)|\to 1$ as $z\to 0+$, and that the initial occupation probabilities of states 2 and 3 vanish. This initial condition and (5) are satisfied by

$$a_1=z^{-i(2\beta_1+\beta_2)/3}(1-z)^{i\gamma_1/3}F(i(\rho_1+\sigma_1),i(\rho_2+\sigma_1),i(\rho_3+\sigma_1);\tfrac{1}{2}-i\beta_1,-i(\beta_1+\beta_2);z),$$
$$a_2=\frac{2\pi(\rho_1+\sigma_1)(\rho_2+\sigma_1)(\rho_3+\sigma_1)}{\alpha_1(\beta_1+\frac{1}{2}i)(\beta_1+\beta_2)}z^{1/2-i(2\beta_1+\beta_2)/3}(1-z)^{1/2+i\gamma_1/3}$$
$$\times F(1+i(\rho_1+\sigma_1),1+i(\rho_2+\sigma_1),1+i(\rho_3+\sigma_1);\tfrac{3}{2}-i\beta_1,1-i(\beta_1+\beta_2);z),$$

and

$$a_3=\frac{-\alpha_2(\rho_1+\sigma_1)(\rho_2+\sigma_1)(\rho_3+\sigma_1)}{\alpha_1(\beta_1+\frac{1}{2}i)(\beta_1+\beta_2)(\beta_1+\beta_2+i)}z^{1-i(2\beta_1+\beta_2)/3}(1-z)^{i\gamma_1/3}$$
$$\times F(1+i(\rho_1+\sigma_1),1+i(\rho_2+\sigma_1),1+i(\rho_3+\sigma_1);\tfrac{3}{2}-i\beta_1,2-i(\beta_1+\beta_2);z),$$

where

$$\sigma_1=(-2\beta_1-\beta_2+\gamma_1)/3 \tag{C5}$$

is found by calculating γ' and ϵ' and then using (C3) and (C4), or by using the three exponents at $z=\infty$.

Suppose that the occupation probabilities of states 1 and 3 vanish at the initial time, and $|a_2(z)|\to 1$ and $z\to 0+$. This initial condition and (5) are satisfied by

$$a_1=[\alpha_1/\pi(2\beta_1-i)]z^{1/2+i(\beta_1-\beta_2)/3}(1-z)^{i\gamma_1/3}F(\tfrac{1}{2}+i(\rho_1+\sigma_2),\tfrac{1}{2}+i(\rho_2+\sigma_2),\tfrac{1}{2}+i(\rho_3+\sigma_2);\tfrac{1}{2}-i\beta_2,\tfrac{3}{2}+i\beta_1;z),$$
$$a_2=z^{i(\beta_1-\beta_2)/3}(1-z)^{1/2+i\gamma_1/3}F(\tfrac{1}{2}+i(\rho_1+\sigma_2),\tfrac{1}{2}+i(\rho_2+\sigma_2),\tfrac{1}{2}+i(\rho_3+\sigma_2);\tfrac{1}{2}+i\beta_1,\tfrac{1}{2}-i\beta_2;z),$$

and

$$a_3=[-\alpha_2/\pi(2\beta_2+i)]z^{\frac{1}{2}+i(\beta_1-\beta_2)/3}(1-z)^{i\gamma_1/3}F(\tfrac{1}{2}+i(\rho_1+\sigma_2),\tfrac{1}{2}+i(\rho_2+\sigma_2),\tfrac{1}{2}+i(\rho_3+\sigma_2);\tfrac{1}{2}+i\beta_1,\tfrac{3}{2}-i\beta_2;z),$$

where $\sigma_2=(\beta_1-\beta_2+\gamma_1)/3$.

Suppose the occupation probabilities of states 1 and 2 vanish at the initial time, and $|a_3(z)|\to 1$. This initial condition and (5) are satisfied by

$$a_1=\frac{\alpha_1(\rho_1+\sigma_3)(\rho_2+\sigma_3)(\rho_3+\sigma_3)}{\alpha_2(\beta_2-\frac{1}{2}i)(\beta_1+\beta_2)(\beta_1+\beta_2-i)}z^{1+i(\beta_1+2\beta_2)/3}(1-z)^{i\gamma_1/3}F(1+i(\rho_1+\sigma_3),$$
$$1+i(\rho_2+\sigma_3),1+i(\rho_3+\sigma_3);\tfrac{3}{2}+i\beta_2,2+i(\beta_1+\beta_2);z),$$
$$a_2=\frac{2\pi(\rho_1+\sigma_3)(\rho_2+\sigma_3)(\rho_3+\sigma_3)}{\alpha_2(\beta_2-\frac{1}{2}i)(\beta_1+\beta_2)}z^{1/2+i(\beta_1+2\beta_2)/3}(1-z)^{1/2+i\gamma_1/3}F(1+i(\rho_1+\sigma_3),$$
$$1+i(\rho_2+\sigma_3),1+i(\rho_3+\sigma_3);\tfrac{3}{2}+i\beta_2,1+i(\beta_1+\beta_2);z),$$

and

$$a_3=z^{i(\beta_1+2\beta_2)/3}(1-z)^{i\gamma_1/3}F(i(\rho_1+\sigma_3),i(\rho_2+\sigma_3),i(\rho_3+\sigma_3);\tfrac{1}{2}+i\beta_2,i(\beta_1+\beta_2);z),$$

where $\sigma_3=(\beta_1+2\beta_2+\gamma_1)/3$.

These three independent solutions were obtained by using (C4) and other results to construct $a_1(z)$ and $a_3(z)$, before finding $a_2(z)$. In each formula for $a_2(z)$, Clausen's five parameters satisfy

$$\alpha'+\beta'+\delta'-\gamma'-\epsilon'=i\gamma_1+\tfrac{1}{2}. \tag{C6}$$

Thus, the real part of $\alpha'+\beta'+\delta'-\gamma'-\epsilon'$ is $\pm\frac{1}{2}$; it is entirely independent of the parameters of the model, and of the initial conditions.

We should also treat the case of two-photon resonance at the initial time, in which (11) and (14) hold. The exponents at $z=0$ give $\mu=-i\beta_1/3$, and (B10) is satisfied automatically. We can use (11) to write (C1) in a shorter form. Time reversal can be used to simplify the remaining calculations. The usual connection between time reversal and the Schrödinger equation leads us to write down the complex conjugates of the above three solutions and interchange z and $1-z$. Since the Hamiltonian is time dependent, we must also interchange β_1, β_2 and γ_1, γ_2. These operations leave (8), (C1), (C2), and (C3) unchanged, but (11) is now used instead of (10). This time-reversal transformation has the effect of replacing σ_1, σ_2, and σ_3 by $\sigma_1'=(-\beta_1+2\gamma_1+\gamma_2)/3$, $\sigma_2'=-(\beta_1+\gamma_1-\gamma_2)/3$, and $\sigma_3'=-(\beta_1+\gamma_1+2\gamma_2)/3$.

For the cases of two-photon resonance at the initial time and two-photon resonance at the final time, we now have three independent solutions of the Schrödinger equation. A minor difficulty with these solutions is that we have to solve (C1) to obtain the roots (C2). The main difficulty is finding the behavior of Clausen's function near $x=1$, and it is considered in the following Appendix.

APPENDIX D: SUMMATION OF CLAUSEN'S SERIES

We now seek the sum of Clausen's series, (7), assuming that x is equal to unity or nearly equal to unity. The series may terminate, but finite series are not considered in this Appendix. For infinite series, the first question is that of convergence at $x=1$. A test given by Weierstrass[30] gives the result we need: If the real part of $\alpha'+\beta'+\delta'-\gamma'-\epsilon'$ is positive, then (7) diverges at $x=1$; if the real part is negative, then (7) converges at $x=1$. We need not consider the difficult case in which the real part is zero, but a recent paper[31] treats the case of $\alpha'+\beta'+\delta'-\gamma'-\epsilon'=0$.

In the present application, the value of $\alpha'+\beta'+\delta'-\gamma'-\epsilon'$ is given by (C4) or (C6), which lead to convergence and divergence at $x=1$. We shall deal with both cases.

1. Asymptotic form when series diverges

Suppose that the real part of $\alpha'+\beta'+\delta'-\gamma'-\epsilon'$ is positive, so that (7) diverges when $x=1$. However, (7) converges when $1-x$ is small and positive, and $F(\alpha',\beta',\delta';\gamma',\epsilon';x)$ is asymptotic to

$$\frac{\Gamma(\gamma')\Gamma(\epsilon')\Gamma(\alpha'+\beta'+\delta'-\gamma'-\epsilon')}{\Gamma(\alpha')\Gamma(\beta')\Gamma(\delta')} \times(1-x)^{\gamma'+\epsilon'-\alpha'-\beta'-\delta'}. \tag{D1}$$

This result can be used to find the limit of $|a_2(z)|^2$; formulas are given in Appendix C. To prove it, we write

$$F(\alpha',\beta',\delta';\gamma',\epsilon';x) = \frac{\Gamma(\gamma')\Gamma(\epsilon')}{\Gamma(\alpha')\Gamma(\beta')\Gamma(\delta')} \times \sum_{n=0}^{\infty} \frac{\Gamma(\alpha'+n)\Gamma(\beta'+n)\Gamma(\delta'+n)}{\Gamma(\gamma'+n)\Gamma(\epsilon'+n)} \frac{x^n}{n!}.$$

Assuming that x and $1-x$ are positive, we write the coefficient of $\Gamma(\gamma')\Gamma(\epsilon')/\Gamma(\alpha')\Gamma(\beta')\Gamma(\delta')$ as

$$\sum_{n=0}^{\infty} \frac{\Gamma(\alpha'+\beta'+\delta'-\gamma'-\epsilon'+n)}{n!} x^n + \sum_{n=0}^{\infty} \left[\frac{\Gamma(\alpha'+n)\Gamma(\beta'+n)\Gamma(\delta'+n)}{\Gamma(\gamma'+n)\Gamma(\epsilon'+n)} - \Gamma(\alpha'+\beta'+\delta'-\gamma'-\epsilon'+n) \right] \frac{x^n}{n!}.$$

Here, the first series is proportional to the binomial series. The second series converges at $x=1$, or is negligible for our purpose. To prove this last assertion, we use Stirling's formula to show that the nth terms in the first and second series are asymptotic to

$$n^{(\alpha'+\beta'+\delta'-\gamma'-\epsilon'-1)}x^n$$

and

$$(\text{constant})\, n^{(\alpha'+\beta'+\delta'-\gamma'-\epsilon'-2)}x^n,$$

respectively. Finally, we sum the binomial series and obtain (D1).

2. Sum of series at $x=1$

Finding the sum of Clausen's series at $x=1$, when it converges, is an old mathematical problem,[32] and we do not expect to find any general formula. Four summation formulas for special cases should be mentioned. Saalchütz[33] gives a simple formula for $F(\alpha',\beta',\delta';\gamma';\epsilon';1)$ that is valid if the series terminates and $\alpha'+\beta'+\delta'-\gamma'-\epsilon'=-1$. Clearly, we cannot use his formula. Dixon[34] evaluated

$$F(\alpha',\beta',\delta';1+\alpha'-\beta',1+\alpha'-\delta';1),$$

assuming that the series converges. The sum is

$$\frac{\Gamma(1+\tfrac{1}{2}\alpha')\Gamma(1+\tfrac{1}{2}\alpha'-\beta'-\delta')\Gamma(1+\alpha'-\beta')\Gamma(1+\alpha'-\delta')}{\Gamma(1+\tfrac{1}{2}\alpha'-\beta')\Gamma(1+\tfrac{1}{2}\alpha'-\delta')\Gamma(1+\alpha')\Gamma(1+\alpha'-\beta'-\delta')}.$$

This series can appear in Appendix C only if (12) holds, which would give vanishing denominators. If (12) holds, see Sec. IV; we do not use Dixon's formula.

The formula

$$F(\alpha',\beta',\delta';\tfrac{1}{2}(1+\alpha'+\beta'),2\delta';1) = \frac{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2}+\delta')\Gamma[\frac{1}{2}(1+\alpha'+\beta')]\Gamma[\frac{1}{2}(1-\alpha'-\beta')+\delta']}{\Gamma(\frac{1}{2}+\frac{1}{2}\alpha')\Gamma(\frac{1}{2}+\frac{1}{2}\beta')\Gamma(\frac{1}{2}-\frac{1}{2}\alpha'+\delta')\Gamma(\frac{1}{2}-\frac{1}{2}\beta'+\delta')},$$

which is named for Watson,[15] is valid if the series converges. The proof was given by Whipple,[16] who also showed that

$$F(\alpha',1-\alpha',\delta';\gamma',1+2\delta'-\gamma';1) = \frac{2^{1-2\delta'}\pi\Gamma(\gamma')\Gamma(1+2\delta'-\gamma')}{\Gamma(\frac{1}{2}\alpha'+\frac{1}{2}\gamma')\Gamma(\frac{1}{2}-\frac{1}{2}\alpha'+\frac{1}{2}\gamma')\Gamma(\frac{1}{2}+\delta'-\frac{1}{2}\gamma'+\frac{1}{2}\alpha')\Gamma(1+\delta'-\frac{1}{2}\gamma'-\frac{1}{2}\alpha')},$$

holds whenever the series converges. The first two sections of Table I list cases in which the formulas of Watson and Whipple can be used together, to simplify the results in Appendix C; details are given in Appendix E. We should mention that Watson's formula or Whipple's formula can sometimes be applied when Clausen's series terminates. This leads to some overlap of different sections of Table I; a specific example is given in Sec. III C.

APPENDIX E: CALCULATION OF FINAL OCCUPATION PROBABILITIES

After writing solutions of the Schrödinger equation in terms of Clausen's function, we seek cases in which the final occupation probabilities can be written in terms of elementary functions. This leads to various sets of three equations to determine the six parameters (9) in terms of three of them. The restrictions on the parameters (9) are listed in Table I, and the resulting final occupation probabilities are given in Tables II, III, and IV. This Appendix sketches the calculations that lead from Appendix C to these tables.

In the case of two-photon resonance at the final time, (10) holds, and three independent solutions of the Schrödinger equation are written explicitly in Appendix C. The final occupation probability for state 2 can always be found, by using (D1). The final occupation probabilities for states 1 and 3 can be calculated from the formulas of Watson and Whipple, provided that

$$2(\beta_1-\beta_2+\gamma_1)/3=\rho_3, \quad \gamma_1/3+\rho_3=(\beta_1-\beta_2)/6.$$

These two equations are valid for any initial condition, and they give

$$\gamma_1=-\tfrac{1}{2}(\beta_1-\beta_2), \quad \rho_3=(\beta_1-\beta_2)/3. \tag{E1}$$

As one of the roots is now prescribed, solution of (C1) is simple. We find that ρ_1 and ρ_2 are $-(\beta_1-\beta_2)/6 \pm i\Phi/\pi$, where Φ is the angle defined in Table II; (C1) also gives $\alpha_1^2-\alpha_2^2=2\pi^2(\beta_1^2-\beta_2^2)$. These calculations give the first section of Table I. To derive the results listed in Table II, we consider the three initial conditions separately.

If state 1 is initially occupied, and the initial occupation

probabilities of states 2 and 3 are zero, we use the first solution written explicitly in Appendix C. Using (E1), we find

$$(\sigma_1+\rho_1)(\sigma_1+\rho_2)(\sigma_1+\rho_3) = -\alpha_1^2(\beta_1+\beta_2)/(2\pi)^2,$$

where σ_1 is given by (C5). This result simplifies the explicit solution. Using (D1) and Watson's formula, we find the asymptotic form of this explicit solution, which is listed in Table XI. Taking absolute squares gives the results listed in Table II.

If state 2 is initially occupied, and the initial occupation probabilities of state 1 and 3 are zero, we use the second solution given explicitly in Appendix C. Using (D1) and Whipple's formula, we find the results listed in Table XI.

If state 3 is initially occupied, we derive

$$(\sigma_3+\rho_1)(\sigma_3+\rho_2)(\sigma_3+\rho_3) = \alpha_2^2(\beta_1+\beta_2)/(2\pi)^2$$

from (E1), and use this result to simplify the explicit solution. Using (D1) and Watson's formula, we find the results listed in Table XI.

We should also describe the simple cases in which Clausen's series terminates wherever it appears in the wave function. We assume two-photon resonance at the final time, so that the solutions written explicitly in Appendix C can be used. Table I emphasizes that the three initial conditions give different formulas for α_1^2 and α_2^2; they are derived in the following way.

If state 1 is initially occupied, the conditions for terminating series in Appendix C are

$$\rho_1 = -\sigma_1 + in, \quad \rho_2 = -\sigma_1 - in, \quad \rho_3 = 2\sigma_1. \quad \text{(E2a)}$$

Here, n is any integer. Since (C1) and the explicit solution are unaffected by interchange of ρ_1 and ρ_2, we may assume $n \geqslant 0$. If $n = 0$, (C1) gives $\alpha_1^2(\beta_1+\beta_2) = 0$. Since the cases of $\alpha_1 = 0$ and $\beta_1+\beta_2 = 0$ do not belong here, we assume $n > 0$. Then, (C1) gives

$$n[(\alpha_1^2+\alpha_2^2)/(2\pi)^2 + (\beta_1+\beta_2)(\beta_1-\gamma_1) - n^2] - i[\alpha_1^2(\beta_1+\beta_2)/(2\pi)^2 + 3n^2\sigma_1] = 0.$$

The real and imaginary parts of this result give formulas listed in Table I. Two of the parameters in (7) are now integers, and two such series appear in Table V. Setting $n = 1$ gives the simple example appearing in Table IV.

If state 2 is initially occupied, the conditions for terminating series in Appendix C are

$$\rho_1 = -\sigma_2 + i(n+\tfrac{1}{2}), \quad \rho_2 = -\sigma_2 - i(n+\tfrac{1}{2}), \quad \rho_3 = 2\sigma_2, \quad \text{(E2b)}$$

where n is any integer. Since ρ_1 and ρ_2 could be interchanged, we assume $n \geqslant 0$. From (C1), we find

$$(n+\tfrac{1}{2})[(\alpha_1^2+\alpha_2^2)/(2\pi)^2 - \beta_1\beta_2 + (\beta_1-\beta_2)\gamma_1 - (n+\tfrac{1}{2})^2] - i[(\alpha_1^2\beta_2 - \alpha_2^2\beta_1)/(2\pi)^2 + \beta_1\beta_2\gamma_1 + 3\sigma_2(n+\tfrac{1}{2})^2] = 0,$$

which gives two of the formulas listed in Table I. Two of the parameters in (7) are now integers, and one such series appears in Table V. Setting $n = 0$ and 1 gives the two examples appearing in Table IV.

If state 3 is initially occupied, the conditions for terminating series in Appendix C are

$$\rho_1 = -\sigma_3 + in, \quad \rho_2 = -\sigma_3 - in, \quad \rho_3 = 2\sigma_3. \quad \text{(E2c)}$$

We may assume that n is a positive integer, and use (C1) to find

TABLE XI. Asymptotic forms of wave functions as $z \to 1$, from below. The wave functions for the case of two-photon resonance at the final time are given explicitly in Appendix C, and the simplifying assumptions used here are listed in the first section of Table I. The angle Φ is defined in Table II.

Initial occupation probabilities	Asymptotic form of wave function
1	$a_1 \sim \dfrac{\pi^{1/2}\Gamma(\frac{1}{2}-i\beta_1)\Gamma(\frac{1}{2}-i\gamma_1)\Gamma[\frac{1}{2}-\frac{1}{2}i(\beta_1+\beta_2)](1-z)^{i\gamma_1/3}}{\Gamma(\frac{1}{2}-\frac{1}{2}i\beta_1+\Phi/2\pi)\Gamma(\frac{1}{2}-\frac{1}{2}i\beta_1-\Phi/2\pi)\Gamma(\frac{1}{2}-\frac{1}{2}i\beta_2+\Phi/2\pi)\Gamma(\frac{1}{2}-\frac{1}{2}i\beta_2-\Phi/2\pi)}$
0	$a_2 \sim \dfrac{2^{-i(\beta_1+\beta_2)}i\alpha_1\Gamma(\frac{1}{2}-i\beta_1)\Gamma(\frac{1}{2}+i\gamma_1)\Gamma[\frac{1}{2}-\frac{1}{2}i(\beta_1+\beta_2)](1-z)^{-2i\gamma_1/3}}{2\pi^{3/2}\Gamma(1-i\beta_1+\Phi/\pi)\Gamma(1-i\beta_1-\Phi/\pi)}$
0	$a_3 \sim -\dfrac{\alpha_1\alpha_2\Gamma(\frac{1}{2}-i\beta_1)\Gamma(\frac{1}{2}-i\gamma_1)\Gamma[\frac{1}{2}-\frac{1}{2}i(\beta_1+\beta_2)](1-z)^{i\gamma_1/3}}{8\pi^{3/2}\Gamma(1-\frac{1}{2}i\beta_1+\Phi/2\pi)\Gamma(1-\frac{1}{2}i\beta_1-\Phi/2\pi)\Gamma(1-\frac{1}{2}i\beta_2+\Phi/2\pi)\Gamma(1-\frac{1}{2}i\beta_2-\Phi/2\pi)}$
0	$a_1 \sim \dfrac{2^{2i\gamma_1}i\alpha_1\Gamma(\frac{1}{2}+i\beta_1)\Gamma(\frac{1}{2}-i\beta_2)(1-z)^{i\gamma_1/3}}{2\Gamma(1+\frac{1}{2}i\beta_1+\Phi/2\pi)\Gamma(1+\frac{1}{2}i\beta_1-\Phi/2\pi)\Gamma(\frac{1}{2}-\frac{1}{2}i\beta_2+\Phi/2\pi)\Gamma(\frac{1}{2}-\frac{1}{2}i\beta_2-\Phi/2\pi)}$
1	$a_2 \sim \dfrac{\Gamma(\frac{1}{2}+i\beta_1)\Gamma(\frac{1}{2}+i\beta_2)\Gamma(\frac{1}{2}+i\gamma_1)}{\Gamma(\frac{1}{2}+\Phi/\pi)\Gamma(\frac{1}{2}-\Phi/\pi)\Gamma(\frac{1}{2}-i\gamma_1)}(1-z)^{-2i\gamma_1/3}$
0	$a_3 \sim \dfrac{2^{2i\gamma_1}i\alpha_2\Gamma(\frac{1}{2}+i\beta_1)\Gamma(\frac{1}{2}-i\beta_2)(1-z)^{i\gamma_1/3}}{2\Gamma(\frac{1}{2}+\frac{1}{2}i\beta_1+\Phi/2\pi)\Gamma(\frac{1}{2}+\frac{1}{2}i\beta_1-\Phi/2\pi)\Gamma(1-\frac{1}{2}i\beta_2+\Phi/2\pi)\Gamma(1-\frac{1}{2}i\beta_2-\Phi/2\pi)}$
0	$a_1 \sim -\dfrac{\alpha_1\alpha_2\Gamma(\frac{1}{2}+i\beta_2)\Gamma(\frac{1}{2}-i\gamma_1)\Gamma[\frac{1}{2}+\frac{1}{2}i(\beta_1+\beta_2)](1-z)^{i\gamma_1/3}}{8\pi^{3/2}\Gamma(1+\frac{1}{2}i\beta_1+\Phi/2\pi)\Gamma(1+\frac{1}{2}i\beta_1-\Phi/2\pi)\Gamma(1+\frac{1}{2}i\beta_2+\Phi/2\pi)\Gamma(1+\frac{1}{2}i\beta_2-\Phi/2\pi)}$
0	$a_2 \sim \dfrac{2^{i(\beta_1+\beta_2)}i\alpha_2\Gamma(\frac{1}{2}+i\beta_2)\Gamma(\frac{1}{2}+i\gamma_1)\Gamma[\frac{1}{2}+\frac{1}{2}i(\beta_1+\beta_2)](1-z)^{-2i\gamma_1/3}}{2\pi^{3/2}\Gamma(1+i\beta_2+\Phi/\pi)\Gamma(1+i\beta_2-\Phi/\pi)}$
1	$a_3 \sim \dfrac{\pi^{1/2}\Gamma(\frac{1}{2}+i\beta_2)\Gamma(\frac{1}{2}-i\gamma_1)\Gamma[\frac{1}{2}+\frac{1}{2}i(\beta_1+\beta_2)](1-z)^{i\gamma_1/3}}{\Gamma(\frac{1}{2}+\frac{1}{2}i\beta_1+\Phi/2\pi)\Gamma(\frac{1}{2}+\frac{1}{2}i\beta_1-\Phi/2\pi)\Gamma(\frac{1}{2}+\frac{1}{2}i\beta_2+\Phi/2\pi)\Gamma(\frac{1}{2}+\frac{1}{2}i\beta_2-\Phi/2\pi)}$

$$n[(\alpha_1^2+\alpha_2^2)/(2\pi)^2+(\beta_1+\beta_2)(\beta_2+\gamma_1)-n^2]$$
$$+i[\alpha_2^2(\beta_1+\beta_2)/(2\pi)^2-3n^2\sigma_3]=0.$$

The real and imaginary parts give two formulas listed in Table I, and the simplest example is shown in Table IV.

Finally, we shall describe the case of two-photon resonance at the initial time, without writing solutions of (5) explicitly. We assume that (11) holds. We obtain three independent solutions of (5) in the way indicated at the end of Appendix C. Each component of the wave function contains Clausen's series, and the cases of terminating series are mentioned in Sec. III C. Here, we assume that the series do not terminate, and look for the behavior of the general solution of (5) as $z\to 0$ and as $z\to 1$. We easily find the behavior of the general solution at early times, or as $z\to 0$:

$$a_1\sim A_1 z^{-i\beta_1/3},\quad a_2\sim A_2 z^{2i\beta_1/3},\quad a_3\sim A_3 z^{-i\beta_1/3}.$$

Here, A_1, A_2, and A_3 are three constants that can be calculated, in principle. At late times, or as $z\to 1$, we have

$$a_1\sim B_1(1-z)^{i(2\gamma_1+\gamma_2)/3},$$
$$a_2\sim B_2(1-z)^{-i(\gamma_1-\gamma_2)/3},$$
$$a_3\sim B_3(1-z)^{-i(\gamma_1+2\gamma_2)/3}.$$

Here, B_1, B_2, and B_3 are three constants appearing in the general solution of (5); indeed, they are the coefficients of the three independent solutions mentioned above. To find A_2 in terms of B_1, B_2, and B_3, we use (D1). To find A_1 and A_3, we apply the formulas of Watson and Whipple. This is possible only if

$$\rho_3+2(\gamma_1-\gamma_2+\beta_1)/3=0,\quad \rho_3+(\gamma_1-\gamma_2)/6=\beta_1/3.$$

These two conditions give

$$\beta_1=-\tfrac{1}{2}(\gamma_1-\gamma_2),\quad \rho_3=-(\gamma_1-\gamma_2)/3.$$

Then, (C1) gives $\alpha_1^2-\alpha_2^2=2\pi^2(\gamma_1^2-\gamma_2^2)$ and the values of ρ_1 and ρ_2, which are $(\gamma_1-\gamma_2)/6\pm i\Phi/\pi$. Here, Φ is the angle defined in Table III. We can now let $z\to 0$ in the three independent solutions of (5). The formulas of Watson and Whipple are used to find

$$A_1=\frac{\pi^{1/2}\Gamma(\frac{1}{2}+i\beta_1)\Gamma(\frac{1}{2}+i\gamma_1)\Gamma[\frac{1}{2}+\frac{1}{2}i(\gamma_1+\gamma_2)]}{\Gamma(\frac{1}{2}+\frac{1}{2}i\gamma_1+\Phi/2\pi)\Gamma(\frac{1}{2}+\frac{1}{2}i\gamma_1-\Phi/2\pi)\Gamma(\frac{1}{2}+\frac{1}{2}i\gamma_2+\Phi/2\pi)\Gamma(\frac{1}{2}+\frac{1}{2}i\gamma_2-\Phi/2\pi)}B_1$$
$$-\frac{2^{i(\gamma_1-\gamma_2)}i\alpha_1\Gamma(\frac{1}{2}-i\gamma_1)\Gamma(\frac{1}{2}+i\gamma_2)}{2\Gamma(1-\frac{1}{2}i\gamma_1+\Phi/2\pi)\Gamma(1-\frac{1}{2}i\gamma_1-\Phi/2\pi)\Gamma(\frac{1}{2}+\frac{1}{2}i\gamma_2+\Phi/2\pi)\Gamma(\frac{1}{2}+\frac{1}{2}i\gamma_2-\Phi/2\pi)}B_2$$
$$-\frac{\alpha_1\alpha_2\Gamma(\frac{1}{2}+i\beta_1)\Gamma(\frac{1}{2}-i\gamma_2)\Gamma[\frac{1}{2}-\frac{1}{2}i(\gamma_1+\gamma_2)]}{8\pi^3\Gamma(1-\frac{1}{2}i\gamma_1+\Phi/2\pi)\Gamma(1-\frac{1}{2}i\gamma_1-\Phi/2\pi)\Gamma(1-\frac{1}{2}i\gamma_2+\Phi/2\pi)\Gamma(1-\frac{1}{2}i\gamma_2-\Phi/2\pi)}B_3$$

and

$$A_2=-\frac{2^{i(\gamma_1+\gamma_2)}i\alpha_1\Gamma(\frac{1}{2}-i\beta_1)\Gamma(\frac{1}{2}+i\gamma_1)\Gamma[\frac{1}{2}+\frac{1}{2}i(\gamma_1+\gamma_2)]}{2\pi^{3/2}\Gamma(1+i\gamma_1+\Phi/\pi)\Gamma(1+i\gamma_1-\Phi/\pi)}B_1+\frac{\Gamma(\frac{1}{2}-i\beta_1)\Gamma(\frac{1}{2}-i\gamma_1)\Gamma(\frac{1}{2}+i\gamma_2)}{\Gamma(\frac{1}{2}+i\beta_1)\Gamma(\frac{1}{2}+\Phi/\pi)\Gamma(\frac{1}{2}-\Phi/\pi)}B_2$$
$$-\frac{2^{-i(\gamma_1+\gamma_2)}i\alpha_2\Gamma(\frac{1}{2}-i\beta_1)\Gamma(\frac{1}{2}-i\gamma_2)\Gamma[\frac{1}{2}-\frac{1}{2}i(\gamma_1+\gamma_2)]}{2\pi^{3/2}\Gamma(1-i\gamma_2+\Phi/\pi)\Gamma(1-i\gamma_2-\Phi/\pi)}B_3$$

and

$$A_3=-\frac{\alpha_1\alpha_2\Gamma(\frac{1}{2}+i\beta_1)\Gamma(\frac{1}{2}+i\gamma_1)\Gamma[\frac{1}{2}+\frac{1}{2}i(\gamma_1+\gamma_2)]}{8\pi^3\Gamma(1+\frac{1}{2}i\gamma_1+\Phi/2\pi)\Gamma(1+\frac{1}{2}i\gamma_1-\Phi/2\pi)\Gamma(1+\frac{1}{2}i\gamma_2+\Phi/2\pi)\Gamma(1+\frac{1}{2}i\gamma_2-\Phi/2\pi)}B_1$$
$$-\frac{2^{i(\gamma_1-\gamma_2)}i\alpha_2\Gamma(\frac{1}{2}-i\gamma_1)\Gamma(\frac{1}{2}+i\gamma_2)}{2\Gamma(\frac{1}{2}-\frac{1}{2}i\gamma_1+\Phi/2\pi)\Gamma(\frac{1}{2}-\frac{1}{2}i\gamma_1-\Phi/2\pi)\Gamma(1+\frac{1}{2}i\gamma_2+\Phi/2\pi)\Gamma(1+\frac{1}{2}i\gamma_2-\Phi/2\pi)}B_2$$
$$+\frac{\pi^{1/2}\Gamma(\frac{1}{2}+i\beta_1)\Gamma(\frac{1}{2}-i\gamma_2)\Gamma[\frac{1}{2}-\frac{1}{2}i(\gamma_1+\gamma_2)]}{\Gamma(\frac{1}{2}-\frac{1}{2}i\gamma_1+\Phi/2\pi)\Gamma(\frac{1}{2}-\frac{1}{2}i\gamma_1-\Phi/2\pi)\Gamma(\frac{1}{2}-\frac{1}{2}i\gamma_2+\Phi/2\pi)\Gamma(\frac{1}{2}-\frac{1}{2}i\gamma_2-\Phi/2\pi)}B_3.$$

We can verify that the A vector is a unitary matrix times the B vector. Since the matrix is unitary, matrix inversion is easy, and it leads to Table III.

APPENDIX F: CALCULATIONS FOR TWO-PHOTON RESONANCE

The assumptions of two-photon resonance at all times and of constant ratio of the two Rabi frequencies simplify both the Schrödinger equation for the three-state problem and its solutions. This Appendix indicates how we construct analytic solutions of (6), which is equivalent to the Schrödinger equation. We can find the behavior of the wave functions at the end of the two concurrent optical pulses, and use this behavior to obtain Tables VII–IX. This behavior also allows us to find easily the conditions for complete transfer or complete return, which can be difficult to derive from the three tables.

In the simplest case, we assume that $\omega_1(z)$, $\omega_2(z)$, and $\delta_1(z)$ are constant functions, and use (15) and (16) in writing solutions of (6). We also use the functions $\exp(\frac{1}{2}i\delta_1 z)\times\sin(\frac{1}{2}\Phi z)$ and $\exp(\frac{1}{2}i\delta_1 z)[\cos(\frac{1}{2}\Phi z)\pm(i\delta_1/\Phi)\times\sin(\frac{1}{2}\Phi z)]$, where the ambiguous sign depends on the initial conditions used. These calculations lead to Table VII.

To apply the hypergeometric function, we derive a second-order differential equation, as indicated in Sec. IV B, and require it to have the form of (B2), the Riemann–Pap-

peritz equation. We let $c \to \infty$. We now put the other singular points at $\pm i$, by using (17). Calculation of the parameters $\alpha, \alpha', \beta, \beta', \gamma, \gamma'$ of Riemann and Papperitz leads to two square roots of complex quantities; let

$$r = \tfrac{1}{4}[(\alpha_1^2 + \alpha_2^2)/\pi^2 + (\beta + i\gamma)^2]^{1/2},$$
$$s = \tfrac{1}{4}[(\alpha_1^2 + \alpha_2^2)/\pi^2 + (\beta - i\gamma)^2]^{1/2}.$$

Clearly, $\pi(r+s) = \Phi$ is real and $\pi(r-s) = iX$ is imaginary; the real quantities Φ and X appear in Table VIII. Solutions of the Riemann–Papperitz equation can be written in terms of hypergeometric functions of

$$(z-i)/(z+i)$$

and complex powers of this ratio. To prevent ambiguity or discontinuous solutions, we specify that the logarithm of this ratio increases from 0 to $2\pi i$ as t and z increase from $-\infty$ to $+\infty$. To abbreviate the formulas for $a_1(z)$ and $a_3(z)$, we let

$$f(\pm r,z) = \left(\frac{z+i}{2i}\right)^{i\gamma}\left[\pm 2rF\left(-\frac{i\gamma}{2} \pm r + s, -\frac{i\gamma}{2} \pm r - s; \pm 2r; \frac{z-i}{z+i}\right)\right.$$
$$\left. - \left(\frac{\beta + i\gamma}{4} \pm r\right)F\left(-\frac{i\gamma}{2} \pm r + s, -\frac{i\gamma}{2} \pm r - s; 1 \pm 2r; \frac{z-i}{z+i}\right)\right].$$

Despite the first factor, this function $f(\pm r,z)$ approaches a definite limit as $z \to \pm\infty$; the limit is

$$\frac{\Gamma(1 \pm 2r)\Gamma(-i\gamma)}{\Gamma(-\frac{1}{2}i\gamma \pm r + s)\Gamma(-\frac{1}{2}i\gamma \pm r - s)}.$$

If state 1 is initially occupied, then $a_2(z)$ and $a_3(z)$ must vanish as $z \to -\infty$; we use (15) to write

$$a_1 = \frac{\alpha_2^2}{\alpha_1^2 + \alpha_2^2} - \frac{\alpha_1^2\Gamma(1+i\gamma)}{\alpha_1^2 + \alpha_2^2}\left(\frac{z-i}{z+i}\right)^{(\beta+i\gamma)/4}\left[\frac{\Gamma(2r)f(-r,z)[(z-i)/(z+i)]^{-r}}{\Gamma(1+\frac{1}{2}i\gamma + r + s)\Gamma(1+\frac{1}{2}i\gamma + r - s)}\right.$$
$$\left. + \frac{\Gamma(-2r)f(r,z)[(z-i)/(z+i)]^{r}}{\Gamma(1+\frac{1}{2}i\gamma - r + s)\Gamma(1+\frac{1}{2}i\gamma - r - s)}\right],$$
$$a_2 = -\frac{\alpha_1\Gamma(1+i\gamma)}{4\pi}\left(\frac{z+i}{2i}\right)^{i\gamma}\left(\frac{z-i}{z+i}\right)^{(\beta+i\gamma)/4}$$
$$\times\left[\frac{\Gamma(-2r)[(z-i)/(z+i)]^{r}F(-\frac{1}{2}i\gamma + r + s, -\frac{1}{2}i\gamma + r - s; 1 + 2r; (z-i)/(z+i))}{\Gamma(1+\frac{1}{2}i\gamma - r + s)\Gamma(1+\frac{1}{2}i\gamma - r - s)}\right.$$
$$\left. + \frac{\Gamma(2r)[(z-i)/(z+i)]^{-r}F(-\frac{1}{2}i\gamma - r + s, -\frac{1}{2}i\gamma - r - s; 1 - 2r; (z-i)/(z+i))}{\Gamma(1+\frac{1}{2}i\gamma + r + s)\Gamma(1+\frac{1}{2}i\gamma + r - s)}\right],$$

and

$$a_3 = \frac{\alpha_1\alpha_2}{\alpha_1^2 + \alpha_2^2}\left\{-1 - \Gamma(1+i\gamma)\left(\frac{z-i}{z+i}\right)^{(\beta+i\gamma)/4}\left[\frac{\Gamma(2r)f(-r,z)[(z-i)/(z+i)]^{-r}}{\Gamma(1+\frac{1}{2}i\gamma + r + s)\Gamma(1+\frac{1}{2}i\gamma + r - s)}\right]\right.$$
$$\left. + \frac{\Gamma(-2r)f(r,z)[(z-i)/(z+i)]^{r}}{\Gamma(1+\frac{1}{2}i\gamma - r + s)\Gamma(1+\frac{1}{2}i\gamma - r - s)}\right]\right\}.$$

To show that (6) is satisfied, we use

$$\left(\frac{d}{dx} + \frac{c-1}{x}\right)F(a,b;c;x) = \frac{c-1}{x}F(a,b;c-1;x)$$

and

$$\left(\frac{d}{dx} + \frac{a+b-c}{x-1}\right)F(a,b;c;x) = \frac{(c-a)(c-b)}{c(1-x)}F(a,b;c+1;x),$$

which are derived from two of the differentiation formulas[35] for $F(a,b;c;x)$. As $z \to -\infty$, $a_1(z) \to 1$. As $z \to +\infty$,

$$a_1 \to \frac{\alpha_2^2}{\alpha_1^2 + \alpha_2^2} + \frac{\alpha_1^2}{\alpha_1^2 + \alpha_2^2}\frac{\exp(\frac{1}{2}i\pi\beta)}{\sinh(\pi\gamma)}\left[\exp\left(\frac{\pi\gamma}{2}\right)\cos(2\pi r) - \exp\left(-\frac{\pi\gamma}{2}\right)\cos(2\pi s)\right]$$

and

$$a_3 \to \frac{\alpha_1\alpha_2}{\alpha_1^2 + \alpha_2^2}\left\{-1 + \frac{\exp(\frac{1}{2}i\pi\beta)}{\sinh(\pi\gamma)}\left[\exp\left(\frac{\pi\gamma}{2}\right)\cos(2\pi r) - \exp\left(-\frac{\pi\gamma}{2}\right)\cos(2\pi s)\right]\right\},$$

but $a_2(z)$ does not approach a limit, unless $\gamma = 0$. We find that

$$a_2 \sim \left(\frac{z}{2}\right)^{i\gamma}\frac{(\frac{1}{2}i\alpha_1)\exp(\frac{1}{2}i\pi\beta)[\Gamma(1+i\gamma)]^2}{\Gamma(1+\frac{1}{2}i\gamma + r + s)\Gamma(1+\frac{1}{2}i\gamma + r - s)\Gamma(1+\frac{1}{2}i\gamma - r + s)\Gamma(1+\frac{1}{2}i\gamma - r - s)}$$

as $z\to+\infty$. Complete return to the initial state, or complete transfer from state 1 to state 3, requires that $a_2(z)\to 0$ as $z\to+\infty$. This is possible only if $\gamma=0$, because the gamma function has poles only on the real axis.

If we use (17) and a different set of initial conditions, the solution of (6) is

$$a_1=\frac{4\pi\alpha_1}{\alpha_1^2+\alpha_2^2}\exp\left(\frac{\pi\gamma}{2}\right)\Gamma(-i\gamma)\left[\frac{\Gamma(-2r)[(z-i)/(z+i)]^{r+(\beta+i\gamma)/4}f(r,z)}{\Gamma(-\frac{1}{2}i\gamma-r+s)\Gamma(-\frac{1}{2}i\gamma-r-s)}\right.$$
$$\left.+\frac{\Gamma(2r)[(z-i)/(z+i)]^{-r+(\beta+i\gamma)/4}f(-r,z)}{\Gamma(-\frac{1}{2}i\gamma+r+s)\Gamma(-\frac{1}{2}i\gamma+r-s)}\right],$$
$$a_2=\exp\left(\frac{\pi\gamma}{2}\right)\Gamma(-i\gamma)\left(\frac{z+i}{2i}\right)^{i\gamma}\left(\frac{z-i}{z+i}\right)^{(\beta+i\gamma)/4}$$
$$\times\left[\frac{\Gamma(-2r)[(z-i)/(z+i)]^{r}F(-i\gamma/2+r+s,-i\gamma/2+r-s;1+2r;(z-i)/(z+i))}{\Gamma(-\frac{1}{2}i\gamma-r+s)\Gamma(-\frac{1}{2}i\gamma-r-s)}\right.$$
$$\left.+\frac{\Gamma(2r)[(z-i)/(z+i)]^{-r}F(-i\gamma/2-r+s,-i\gamma/2-r-s;1-2r;(z-i)/(z+i))}{\Gamma(-\frac{1}{2}i\gamma+r+s)\Gamma(-\frac{1}{2}i\gamma+r-s)}\right],$$

and

$$a_3=\frac{4\pi\alpha_2}{\alpha_1^2+\alpha_2^2}\exp\left(\frac{\pi\gamma}{2}\right)\Gamma(-i\gamma)\left[\frac{\Gamma(-2r)[(z-i)/(z+i)]^{r+(\beta+i\gamma)/4}f(r,z)}{\Gamma(-\frac{1}{2}i\gamma-r+s)\Gamma(-\frac{1}{2}i\gamma-r-s)}\right.$$
$$\left.+\frac{\Gamma(2r)[(z-i)/(z+i)]^{-r+(\beta+i\gamma)/4}f(-r,z)}{\Gamma(-\frac{1}{2}i\gamma+r+s)\Gamma(-\frac{1}{2}i\gamma+r-s)}\right].$$

As $z\to-\infty$, $a_1(z)\to 0$, $|a_2(z)|\to 1$, and $a_3(z)\to 0$. Hence, the three-state system is in state 2 at the initial time. As $z\to+\infty$,

$$a_1\to\frac{-8\pi^2 i\alpha_1}{\alpha_1^2+\alpha_2^2}\frac{\exp(\frac{1}{2}i\pi\beta)[\Gamma(-i\gamma)]^2}{\Gamma(-\frac{1}{2}i\gamma+r+s)\Gamma(-\frac{1}{2}i\gamma+r-s)\Gamma(-\frac{1}{2}i\gamma-r+s)\Gamma(-\frac{1}{2}i\gamma-r-s)},$$
$$a_2\sim\left(\frac{z}{2}\right)^{i\gamma}\frac{\exp(\frac{1}{2}i\pi\beta)}{\sinh(\pi\gamma)}\left[\exp\left(\frac{\pi\gamma}{2}\right)\cos(2\pi s)-\exp\left(-\frac{\pi\gamma}{2}\right)\cos(2\pi r)\right],$$

and

$$a_3\to\frac{-8\pi i\alpha_2}{\alpha_1^2+\alpha_2^2}\frac{\exp(\frac{1}{2}i\pi\beta)[\Gamma(-i\gamma)]^2}{\Gamma(-\frac{1}{2}i\gamma+r+s)\Gamma(-\frac{1}{2}i\gamma+r-s)\Gamma(-\frac{1}{2}i\gamma-r+s)\Gamma(-\frac{1}{2}i\gamma-r-s)}.$$

Complete return of the occupation probability to state 2 at the final time occurs if the limits of a_1 and a_3 are zero. The gamma functions that appear here show that this is possible only if $\gamma=0$ or $r-s=\pm\frac{1}{2}i\gamma$; we reject the trivial possibility that $\alpha_1=\alpha_2=0$. Further calculations shows that $r-s=\pm\frac{1}{2}i\gamma$ implies $\alpha_1=\alpha_2=0$. Hence, $\gamma=0$ is the necessary condition for complete transfer or complete return, as we asserted in Sec. IV B. Another solution of (6), such that $a_1(z)\to 0$, $a_2(z)\to 0$, and $a_3(z)\to 1$ as $z\to-\infty$, is needed; it will not be written out. The final occupation probabilities are given in Table VIII.

Finally, we put the singular points of the Riemann–Papperitz equation at 0, 1, and ∞, by using (18). Calculation of the parameters $\alpha,\alpha',\beta,\beta',\gamma$, and γ' of Riemann and Papperitz involves solution of a quadratic equation; the resulting square root appears in (19). Suppose the three-state system is in state 1 at the initial time. We demand that $a_1(z)\to 1$, $a_2(z)\to 0$, and $a_3(z)\to 0$ as $z\to 0$. The solution is

$$a_1=\frac{\alpha_2^2}{\alpha_1^2+\alpha_2^2}+\frac{\alpha_1^2}{\alpha_1^2+\alpha_2^2}F\left(-\frac{i(\beta-\gamma)}{2}+\frac{\Phi}{\pi},-\frac{i(\beta-\gamma)}{2}-\frac{\Phi}{\pi};\frac{1}{2}-i\beta;z\right),$$

$$a_2=\frac{-\alpha_1}{\pi(2\beta+i)}z^{1/2}(1-z)^{1/2}F\left(1-\frac{i(\beta-\gamma)}{2}+\frac{\Phi}{\pi},1-\frac{i(\beta-\gamma)}{2}-\frac{\Phi}{\pi};\frac{3}{2}-i\beta;z\right),$$

and

$$a_3=\frac{\alpha_1\alpha_2}{\alpha_1^2+\alpha_2^2}\left[-1+F\left(-\frac{i(\beta-\gamma)}{2}+\frac{\Phi}{\pi},-\frac{i(\beta-\gamma)}{2}-\frac{\Phi}{\pi};\frac{1}{2}-i\beta;z\right)\right].$$

To verify that (6) is satisfied, we use

$$\left(\frac{d}{dz}+\frac{c-1}{z}+\frac{a+b-c}{z-1}\right)F(a,b;c;z)=\frac{c-1}{z(1-z)}F(a-1,b-1;c-1;z),\tag{F1}$$

and the simple differentiation formula given by Gauss.[36] Equation (F1) is derived from the last of the differentiation formulas given by Oberhettinger.[35] The limits as $z\to 1$ and $t\to+\infty$ can be found from the Gaussian summation formula[22] for the hypergeometric series:

$$a_1 \to \frac{\alpha_2^2}{\alpha_1^2+\alpha_2^2} + \frac{\alpha_1^2}{\alpha_1^2+\alpha_2^2} \frac{\Gamma(\frac{1}{2}-i\beta)\Gamma(\frac{1}{2}-i\gamma)}{\Gamma[\frac{1}{2}-\frac{1}{2}i(\beta+\gamma)+\Phi/\pi]\Gamma[\frac{1}{2}-\frac{1}{2}i(\beta+\gamma)-\Phi/\pi]}$$

and

$$a_3 \to \frac{\alpha_1\alpha_2}{\alpha_1^2+\alpha_2^2}\left\{-1+\frac{\Gamma(\frac{1}{2}-i\beta)\Gamma(\frac{1}{2}-i\gamma)}{\Gamma[\frac{1}{2}-\frac{1}{2}i(\beta+\gamma)+\Phi/\pi]\Gamma[\frac{1}{2}-\frac{1}{2}i(\beta+\gamma)-\Phi/\pi]}\right\}.$$

The gamma functions go away if we assume $\beta+\gamma=0$, and this case is considered in Sec. IV B. The asymptotic form of $a_2(z)$ as $z\to 1$ and $t\to+\infty$ is

$$a_2 \sim \frac{i\alpha_1}{2\pi} \frac{\Gamma(\frac{1}{2}-i\beta)\Gamma(\frac{1}{2}+i\gamma)(1-z)^{-i\gamma}}{\Gamma[1-\frac{1}{2}i(\beta-\gamma)+\Phi/\pi]\Gamma[1-\frac{1}{2}i(\beta-\gamma)-\Phi/\pi]}.$$

The gamma functions disappear when we calculate the final occupation probability, or the limit of $|a_2|^2$. This limit is given in Table IX.

If the three-state system is in state 2 at the initial time and (18) is used, we assert that

$$a_1 = [\alpha_1/\pi(2\beta-i)]z^{1/2+i\beta}F(\tfrac{1}{2}+\tfrac{1}{2}i(\beta+\gamma)+\Phi/\pi,\tfrac{1}{2}+\tfrac{1}{2}i(\beta+\gamma)-\Phi/\pi;\tfrac{3}{2}+i\beta;z),$$

$$a_2 = z^{i\beta}(1-z)^{1/2}F(\tfrac{1}{2}+\tfrac{1}{2}i(\beta+\gamma)+\Phi/\pi,\tfrac{1}{2}+\tfrac{1}{2}i(\beta+\gamma)-\Phi/\pi;\tfrac{1}{2}+i\beta;z),$$

$$a_3 = [\alpha_2/\pi(2\beta-i)]z^{1/2+i\beta}F(\tfrac{1}{2}+\tfrac{1}{2}i(\beta+\gamma)+\Phi/\pi,\tfrac{1}{2}+\tfrac{1}{2}i(\beta+\gamma)-\Phi/\pi;\tfrac{3}{2}+i\beta;z)$$

is the appropriate solution of (6). As $z\to 0$ and $t\to-\infty$, $a_1(z)\to 0$, $|a_2(z)|\to 1$, and $a_3(z)\to 0$. As $z\to 1$ and $t\to+\infty$,

$$a_1 \to \frac{i\alpha_1}{2\pi} \frac{\Gamma(\frac{1}{2}+i\beta)\Gamma(\frac{1}{2}-i\gamma)}{\Gamma[1+\frac{1}{2}i(\beta-\gamma)+\Phi/\pi]\Gamma[1+\frac{1}{2}i(\beta-\gamma)-\Phi/\pi]},$$

$$a_2 \sim \frac{\Gamma(\frac{1}{2}+i\beta)\Gamma(\frac{1}{2}+i\gamma)(1-z)^{-i\gamma}}{\Gamma[\frac{1}{2}+\frac{1}{2}i(\beta+\gamma)+\Phi/\pi]\Gamma[\frac{1}{2}+\frac{1}{2}i(\beta+\gamma)-\Phi/\pi]},$$

and

$$a_3 \to \frac{i\alpha_2}{2\pi} \frac{\Gamma(\frac{1}{2}+i\beta)\Gamma(\frac{1}{2}-i\gamma)}{\Gamma[1+\frac{1}{2}i(\beta-\gamma)+\Phi/\pi]\Gamma[1+\frac{1}{2}i(\beta-\gamma)-\Phi/\pi]}.$$

The resulting final occupation probabilities are given in Table IX.

Finally, we perform similar calculations, using (18), for a three-state system that is in state 3 at the initial time.

[1]L. Allen and J. H. Eberly, *Optical Resonance and Two-Level Atoms* (Wiley, New York, 1975).
[2]F. Bloch, Phys. Rev. **70**, 460 (1946).
[3]T. H. Einwohner, J. Wong, and J. C. Garrison, Phys. Rev. A **14**, 1452 (1976).
[4]T. Clausen, J. Reine Angew. Math. **3**, 89 (1828).
[5]F. T. Hioe, Phys. Rev. A **28**, 879 (1983).
[6]M. Gell-Mann, California Institute of Technology Synchrotron Laboratory Report CTSL-20, reprinted in M. Gell-Mann and Y. Ne'eman, *The Eightfold Way* (Benjamin, New York, 1964).
[7]I. I. Rabi, Phys. Rev. **51**, 652 (1937).
[8]F. T. Hioe, Phys. Rev. A **26**, 1466 (1982).
[9]N. Rosen and C. Zener, Phys. Rev. **40**, 502 (1932).
[10]C. E. Carroll and F. T. Hioe, J. Phys. A **19**, 3579 (1986).
[11]R. J. Cook and B. W. Shore, Phys. Rev. A **20**, 539 (1979).
[12]C. E. Carroll and F. T. Hioe, Phys. Rev. A **36**, 724 (1987).
[13]L. Pochhammer, J. Reine Angew. Mat. **102**, 76 (1888).
[14]S. L. Mc Call and E. L. Hahn, Phys. Rev. **183**, 457 (1969).
[15]G. N. Watson, Proc. London Math. Soc. **23**, xiii–xv (1925); the Roman numerals indicate that this paper was actually read at a meeting.
[16]F. J. W. Whipple, Proc. London Math. Soc. **23**, 104 (1925).
[17]M. Appell, C. R. Acad. Sci. **91**, 211 (1880).
[18]H. R. Gray, R. M. Whitley, and C. R. Stroud, Opt. Lett. **3**, 218 (1978).
[19]E. Arimondo and G. Orriols, Lett. Nuovo Cimento **17**, 333 (1976).
[20]E. J. Robinson, Phys. Rev. A **24**, 2239 (1981); **29**, 1665 (1984).
[21]A. Bambini and P. R. Berman, Phys. Rev. A **23**, 2496 (1981); F. T. Hioe, Phys. Rev. A **30**, 2100 (1984).
[22]C. F. Gauss, *Werke* (Königlichen Gesellschaft der Wissenschaften, Göttingen, 1866), Vol. 3, p. 147 (reprinted by Georg Olms, Hildesheim, 1981).
[23]G. F. B. Riemann, *Bernhard Riemann's Gesammelte Mathematische Werke und Wissenschaftlicher Nachlass* (Tuebner, Leipzig, 1892), pp. 67–85.
[24]E. Papperitz, Math. Ann. **25**, 212 (1885).
[25]L. Fuchs, J. Reine Angew. Math. **66**, 121 (1866).
[26]A. R. Forsyth, *Theory of Differential Equations* (Dover, New York, 1959), Part 3, Chaps. 3 and 4.
[27]G. Frobenius, J. Reine Angew. Math. **76**, 214 (1873).
[28]A. Cayley, *The Collected Mathematical Papers* (Cambridge U. P., Cambridge, 1895), Vol. 8, pp. 458–462.
[29]E. D. Rainville, Bull. Am. Math. Soc. **51**, 714 (1945); or E. D. Rainville, *Special Functions* (Macmillan, New York, 1960), page 82.
[30]K. Weierstrass, *Mathematische Werke* (Mayer und Müller, Berlin, 1894), Vol. 1, p. 185 (reprinted by Johnson Reprint, New York, 1967).
[31]R. J. Evans and D. Stanton, SIAM J. Math. Anal. **15**, 1010 (1984).
[32]A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 1, Chap. 4; L. J. Slater, *Generalized Hypergeometric Functions* (Cambridge U. P., Cambridge, 1966), Chaps. 2 and 4.
[33]L. Saalshütz, Z. Math. Phys. **35**, 186 (1890). Secondary sources are more accessible; see Ref. 32.
[34]A. C. Dixon, Proc. London Math. Soc. **35**, 284 (1903).
[35]F. Oberhettinger, in *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. Stegun (National Bureau of Standards, Washington, DC, 1964), p. 557.
[36]See Ref. 22, p. 126.

Correlated states and collective transition operators for multilevel atomic systems. I

Jagdish Rai
Department of Physics and Astronomy, University of Alabama, Tuscaloosa, Alabama 35487

C. L. Mehta
Department of Physics, Indian Institute of Technology, Delhi, New Delhi 110016, India

N. Mukunda
Center for Theoretical Studies, Indian Institute of Science, Bangalore 560012, India

(Received 7 April 1987; accepted for publication 14 October 1987)

The correlated states of an assembly of N multilevel atoms interacting with resonant radiation fields are determined. On the basis of the duality between the permutation group and the linear group, one can construct the internal states for an assembly of N atoms, with n levels each. The possible states are determined by various allowed Young diagrams, corresponding to various representations of the permutation group S_N. The collective transition operators are obtained as the generators of the group SU(n).

I. INTRODUCTION

In this paper we define the correlated states of an assembly of N multilevel atoms interacting with common radiation fields. It was first pointed out by Dicke[1] that all the atoms are interacting with common radiation fields and hence cannot be treated as independent. In the usual treatment of spontaneous radiation by a gas, the radiation process is calculated as though separate atoms radiate independently of each other. To justify this assumption it might be argued that, as a result of the large distance between atoms and subsequent weak interactions, the probability of a given atom emitting a photon should be independent of the states of other atoms. This model is wrong in principle and many of the results obtained from it are incorrect.

A simple example will be used to illustrate the inadequacy of this description. Assume that a spin-$\frac{1}{2}$ particle is placed in a uniform magnetic field in the higher energy of the two spin states (i.e., antiparallel to the magnetic field). In due course the spin-$\frac{1}{2}$ particle will spontaneously radiate a photon via a magnetic dipole transition and flip to the lower energy state. The probability of finding the spin in its upper energy state falls exponentially to zero.[2,3] If, now, a spin-$\frac{1}{2}$ in its ground state (parallel to the magnetic field) is placed near the first antiparallel spin at a distance small compared with a radiation wavelength but large compared with a particle wavelength, the radiation process would be unaffected if the spins were treated independently. Actually the radiation process would be strongly affected. The initial transition probability would be the same as before but the probability of finding an excited spin-$\frac{1}{2}$ would fall exponentially to $\frac{1}{2}$ rather than to zero. Thus the presence of the unexcited spin doubles the radiation rate. Thus emission for an atomic system is a cooperative process involving in a collective mode all the atoms of the system. In this collective mode, an "order" appears in the system which can be defined by the buildup of correlations between the dipoles belonging to different atoms. This correlation can be used to define the states of the atomic system and has a connection with the permutational symmetry of the particles because the radiation coupling is symmetrical with respect to exchange of any two atoms in the system. This important hypothesis naturally follows from the fact that the atoms are supposed to be confined in a volume small compared to λ: the emission or absorption of a photon of wavelength λ cannot—according to Heisenberg uncertainty—be assigned to a specific atom. We shall call these collective modes of the atomic system "correlated states." These correlated states can be classified on the basis of permutational symmetry as the radiation coupling has this property.

With the advent of lasers in recent years, it has become possible to achieve the excitation of correlated states[4] of atomic radiating systems with the subsequent emission of spontaneous radiation.[5] In this work we generalize these considerations to a multilevel system of radiators where magnetic or electric dipole transitions can take place among the levels. In this treatment the atomic system as a whole will be considered as a single quantum mechanical system. The problem will be one of finding those energy states that represent the correlated motions in the system. The spontaneous emission of radiation takes place by transitions from such states. The collective transition operator will describe the transitions among such levels. The correlated states for the two-level atomic system are angular momentum states.[1] The collective transition operators for this system are the angular momentum operators. Our main purpose in this paper is to describe the correlated states for multilevel atoms. We shall obtain the correlated states as the basis states of the irreducible representation of the group SU(n), for the n-level system. [The group[6] SU(n) is defined as the group of all $n \times n$ unitary matrices with determinant $= 1$.] The collective transition operators are obtained as the generators of this group. This is a generalization of the two-level problem where the angular momentum group was SU(2). The two-level problem was indirectly formulated by Dicke[1] using the spinor representation of SU(2), and by taking direct products of this representation, the correlated states were obtained.

 0022-2488/88/020510-06$02.50

II. ATOMIC SYSTEM AND STATES

The interaction system (radiating gas) consists of N identical atoms. Each atom has n-nondegenerate states among which transitions can take place. In our later calculation we shall also assume that the walls of the container are transparent to the radiation field. We also assume that the collisions do not affect the internal state of the atoms. The transitions under question take place only between nondegenerate states. We shall restrict our study to the electric dipole transition through the electromagnetic field. The resonant dipole transitions play the important role provided that $\rho\lambda_i^3 \gg 1$, where ρ is the number density of atoms and the λ_i's are the wavelengths of the transitions.

The wave function for the atomic system may be written conveniently in a representation diagonal in the center of mass coordinates and the internal coordinates. That is, the total Hamiltonian consists of two parts,

$$H = H_0 + H_i, \tag{1}$$

where H_0 contains translational and rotational energies while H_i contains the internal energy related to the state of excitation of the molecules. We also assume that H_0 and H_i commute, as translation and rotational motion do not affect the internal excitation state of the molecules. The eigenfunction for the above Hamiltonian may be written as

$$\psi^{(N)} = U_g^{(N)}(\mathbf{r}_1,\mathbf{r}_2,...,\mathbf{r}_n) \times \chi^{(N)}(i_1,i_2,...,i_N). \tag{2}$$

Here $U_g^{(N)}$ is the part of the wave function describing the center of mass coordinates $\mathbf{r}_1,\mathbf{r}_2,...,\mathbf{r}_N$ and satisfies

$$H_0 U_g^{(N)} = E_g U_g^{(N)}, \tag{3}$$

where E_g is the energy due to motion. The part

$$\chi^{(N)} = \chi(i_1,i_2,...,i_N) \tag{4}$$

is the internal energy wave function or what we shall call the "spin state" of the system. The jth particle is in the i_jth state $(1 \leqslant i_j \leqslant n,\ 1 \leqslant j \leqslant N)$. We are mainly interested in the χ part of the wave function ψ as U_g does not change during the interaction of radiation with system. There are some constraints imposed by the symmetry properties of the overall wave function. The total wave function for indistinguishable particles must be either symmetric (for Bose molecules) or antisymmetric (for Fermi molecules). The restrictions imposed by such symmetry are discussed in detail in a later paper.

The total number of possible states of the type χ is n^N. But this classification is not very useful as transitions occurring in the whole assembly may connect a state of the above type with many other states. Moreover these states have no definite symmetry in the exchange of two particles. We are interested in obtaining states that will have the following properties.

(1) Transitions in the atomic system connect one state with the minimum number of other states.

(2) States have definite permutational symmetry.

The main aim of this paper is the group theoretical classification of states. Basically we have to reduce a space of n^N dimensions. For this it is necessary to consider in detail the representations of the symmetric group S_N of all $N!$ permutations of N symbols. This problem in the abstract sense has been solved by Weyl.[7] We shall use these results for our classification of correlated states.

The single n-level atom states span the vector space V. Therefore the total space spanned by N-atoms assembly is $V^{\otimes N}$. The structure of $V^{\otimes N}$ as $S_N \times \mathrm{GL}(V)$ module is given by[8]

$$V^{\otimes N} \underset{S_N \times \mathrm{GL}(V)}{\cong} \coprod_{\lambda \vdash N} S^\lambda \otimes \square^\lambda V, \tag{5}$$

where λ is a partition of N treated as a nonincreasing ordered K-tuple of positive integers with sum N. Also

$$\square^\lambda V = 0, \quad \text{if number of parts of } \lambda > n. \tag{6}$$

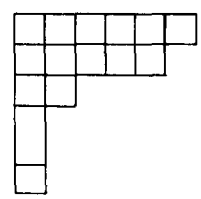
In Eq. (5) $\cong$ denotes isomorphism, S^λ is a representation of S_N, and $\square^\lambda$ is a representation of $\mathrm{GL}(V)$. Equation (5) is the main equation of this paper. Hence it is possible to choose a basis of $V^{\otimes N}$ whose elements transform simultaneously as would a basis for an irreducible representation of $\mathrm{GL}(V)$ and a basis for an irreducible representation of S_N.

Therefore the correlated states of the assembly are taken to be the basis states of the various irreducible representations occurring in the direct product space $V^{\otimes N}$. We denote these states by

$$\phi_{\{\}}^{(N)}, \tag{7}$$

and they are obtained as linear combinations of the earlier states $\chi(i_1,i_2,...,i_N)$ as illustrated in the examples below. The transition from these states to others in the scheme are described with the help of the collective transition operators described in the next section of this paper. We shall also give a scheme for generating ϕ's from the χ's in a later paper with the help of permutation operators.

It is interesting to note that the frequency of $S^\lambda V$ is equal to the dimensionality of $\square^\lambda V$ and the frequency of $\square^\lambda V$ is equal to the dimensionality of $S^\lambda V$ (Schur's theorem[9]). A simple formula[10] for obtaining the dimensionality of $\square^\lambda V$ is given here. Take the corresponding Young's diagram for this representation:

$\square^\lambda V =$  . (8)

Now start filling this diagram from the first corner (left top) with number n and increase by one as you go right and decrease by one as you go down. Filling all the squares in this way one then obtains the following diagram:

$$\begin{array}{ccccc} n & n+1 & n+2 & \cdots \\ n-1 & n & n+1 & \cdots \\ n-2 & n-1 & n & \cdots \\ \vdots & \vdots & \vdots & \end{array} . \tag{9}$$

Then the dimensionality of $\square^\lambda V$ is given by the product of all the factors in the above diagram (say π) multiplied by $\deg \chi^{(\lambda)}$ and divided by $N!$, i.e.,

$$\dim \square^\lambda V = (\deg \chi^\lambda / N!) \times \pi. \tag{10}$$

We illustrate this procedure in the following examples.

These examples also show the correlated states in these cases.

Example A: (Two particles, $N=2$). The general equation (5) becomes

$$V\otimes V\cong \mathrm{sym}^2 V\oplus \Lambda^2 V$$

$$\cong_{S_2\times GL(V)} [(\mathrm{triv})\otimes \mathrm{sym}^2 V]\oplus[(\mathrm{sgn})\otimes\Lambda^2 V], \tag{11}$$

where the two representations of S_2 are denoted by (triv) and (sgn). The trivial representation maps the two elements to $+1$ and is of dimension 1. The (sgn) representation maps permutation e to $+1$ and (12) to -1 and is also of dimension 1. $\mathrm{Sym}^2 V$ is the representation by symmetric functions and is of dimension $n(n+1)/2$. Here V is the representation by antisymmetric functions and is of dimension $n(n-1)/2$. To start with we had n^2 states of the type

$$\chi(i,j);\quad 1\leqslant i,j\leqslant n. \tag{12}$$

There are $n(n+1)/2$ possible symmetric functions of the type $[\chi(i,j)+\chi(j,i)]$ and $n(n-1)/2$ of the type $[\chi(i,j)-\chi(j,i)]$. The correlated states therefore are

$$\begin{aligned}\phi_S^{(2)} &= (1/\sqrt{2})\{\chi(i,j)+\chi(j,i)\},\\ \phi_a^{(2)} &= (1/\sqrt{2})\{\chi(i,j)-\chi(j,i)\}.\end{aligned} \tag{13}$$

Example B: (Three particles, $N=3$). the structure of $V^{\otimes 3}$ as $S_3\times GL(V)$ module is given by

$$V^{\otimes 3}\cong_{S_3\times GL(V)} [\mathrm{triv}\otimes \mathrm{sym}^3 V]\oplus[\rho\otimes \square V]\oplus[\mathrm{sgn}\otimes\Lambda^3 V]. \tag{14}$$

The three irreducible representations of S_3 are shown in Table I. In all there are n^3 functions which break up into $n(n+1)(n+2)/6$ symmetric, $n(n^2-1)/3$ of the type $\square$, and $n(n-1)(n-2)/6$ antisymmetric functions. As before these functions define the correlated states of a three-particle system.

Having achieved the representations of $GL(V)$ in the tensor breakup of $V^{\otimes N}$ we are able to give the molecular states of definite symmetry. One more trick due to Weyl called the "unitarian trick" completes the classification. A representation is irreducible with respect to $GL(V)$ if and only if it is irreducible with respect to $SU(n)$. The proof follows rather easily from the fact that over complex numbers C all operators are linear combinations of Hermitian operators. Therefore we classify atomic states to be the basis functions of the irreducible representation of the unitary group $SU(n)$. Now we illustrate the two- and three-level cases here.

Example C: (Two-level case.) The internal energy coordinates take two values denoted by $+$ (upper state) and $-$ (for lower state). The spin states are

$$\chi^{(N)} = |+\ -\ -\ +\cdots\rangle. \tag{15}$$

The correlated states for two-particle systems are

$$\begin{aligned}\phi_{22}^{(2)} &= |+\ +\rangle,\\ \phi_{10}^{(2)} &= (1/\sqrt{2})\{|+\ -\rangle+|-\ +\rangle\},\\ \phi_{1-1}^{(2)} &= |-\ -\rangle;\end{aligned} \tag{16}$$

TABLE I. Character table for S_3 in example B.

e	(12),(23),(13)	(123),(132)
+1	+1	+1
+1	−1	+1
+1	0	−2

and

$$\phi_{00}^{(2)} = (1/\sqrt{2})\{|+\ -\rangle-|-\ +\rangle\}.$$

(Three-level case, $N=3$.) The three atom states are

$$\begin{aligned}\phi_{3/2\,3/2}^{(3)} &= |+\ +\ +\rangle,\\ \phi_{3/2\,1/2}^{(3)} &= (1/\sqrt{3})\{|+\ +\ -\rangle\\ &\qquad +|+\ -\ +\rangle+|-\ +\ +\rangle\},\\ \phi_{3/2\,-1/2}^{(3)} &= (1/\sqrt{3})\{|-\ -\ +\rangle\\ &\qquad +|-\ +\ -\rangle+|+\ -\ -\rangle\},\\ \phi_{3/2\,-3/2}^{(3)} &+|-\ -\ -\rangle,\end{aligned} \tag{17}$$

and twofold degenerate states are (corresponding to $\square$)

$$\begin{aligned}\phi_{1/2\,-1/2}^{(3)} &= (1/\sqrt{6})\{2|-\ +\ +\rangle\\ &\qquad -|+\ +\ -\rangle-|+\ -\ +\rangle\},\\ \phi_{1/2\,-1/2}^{(3)} &= (1/\sqrt{6})\{|-\ +\ -\rangle\\ &\qquad +|-\ -\ +\rangle-2|+\ -\ -\rangle\},\end{aligned} \tag{18}$$

and

$$\begin{aligned}\phi_{1/2\,1/2}^{(3)} &= (1/\sqrt{2})\{|+\ +\ -\rangle-|+\ -\ +\rangle\},\\ \phi_{1/2\,-1/2}^{(3)} &= (1/\sqrt{2})\{|-\ +\ -\rangle-|-\ -\ +\rangle\}.\end{aligned} \tag{19}$$

III. n-LEVEL ATOMIC TRANSITION OPERATORS

In the previous section we have defined the collective states of the atomic system as the basis states of the irreducible representation of the group $SU(n)$. Now we shall describe the operators which cause transitions among these states. The action of an operator $\hat{R}_{\alpha\beta}^{(j)}$ can be described as

$$\hat{R}_{\alpha\beta}^{(j)}\phi_{\{\alpha_j\}}^{\{N\}} = \delta_{\alpha_j}^{\alpha}\phi_{\{\alpha_1,\ldots,\alpha_{j-1},\beta,\ldots,\alpha_N\}}^{\{N\}} - (1/n)\delta_{\beta}^{\alpha}\phi_{\{\ldots,\alpha_p,\ldots\}}^{\{N\}}, \tag{20}$$

where $j=1,2,\ldots,N$ and $\alpha,\beta=1,2,\ldots,n$. It is clear that $\hat{R}_{\alpha\beta}^{(j)}$ satisfy the traceless condition,

$$\sum_{\alpha=1}^{n}\hat{R}_{\alpha\alpha}^{(j)} = 0. \tag{21}$$

In analogy with the Dicke operators of a two-level system we may also define the collective transition operators for the total system as

$$\hat{R}_{\alpha\beta} = \sum_{j=1}^{N}\hat{R}_{\alpha\beta}^{(j)}. \tag{22}$$

The operators $\hat{R}$ obey the commutation relations,

$$[\hat{R}_{\alpha\beta},\hat{R}_{\mu\gamma}] = \delta_{\beta\mu}\hat{R}_{\alpha\gamma} - \delta_{\alpha\gamma}\hat{R}_{\mu\beta}. \tag{23}$$

These commutation relations[6] are identical to those satisfied by the generators of the group $SU(n)$. Therefore the $\hat{R}$'s

generate the group SU(n). Out of all $\hat{R}_{ii}$ ($i = 1,2,...,n$), there are $n-1$ linearly independent diagonal generators because of the tracelessness condition given in Eq. (21). We can define the $n-1$ diagonal generators as

$$\begin{aligned}&\hat{H}_1 = \hat{R}_{11} - \hat{R}_{22},\\ &\hat{H}_2 = \hat{R}_{11} + \hat{R}_{22} - 2\hat{R}_{33},\\ &\vdots \qquad\qquad \vdots\\ &\hat{H}_{n-1} = \hat{R}_{11} + \hat{R}_{22} + \cdots - (n-1)\hat{R}_{nn}.\end{aligned} \tag{24}$$

This completes our classification for the states and the transition operators. We now give some more examples to illustrate the concept of transition operators.

A. Two-level case

The individual transition operators are Pauli spin matrices. In our notation

$$\hat{R}^{(j)}_{12} = \hat{\sigma}^{(j)}_{+}, \quad \hat{R}^{(j)}_{21} = \hat{\sigma}^{(j)}_{-},$$

and collective two-level operators are

$$\hat{R}_{12} = \hat{J}_{+}, \quad \hat{R}_{21} = \hat{J}_{-}, \tag{25}$$

with the diagonal generators

$$\hat{\sigma}^{(j)}_3 = \tfrac{1}{2}(\hat{R}^{(j)}_{11} - \hat{R}^{(j)}_{22})$$

and

$$\hat{J}_3 = \tfrac{1}{2}(\hat{R}_{11} - \hat{R}_{22}). \tag{26}$$

The angular momentum eigenstates are the correlated states of the two-level atom. They can also be taken as the simultaneous eigenstates of $\hat{J}^2 = \frac{1}{2}(\hat{J}_+\hat{J}_- + \hat{J}_-\hat{J}_+) + \hat{J}_3^2$ and $\hat{J}_3$ denoted by $\phi^{(N)}_{rm}$ which obey

$$\hat{J}^2\phi^{(N)}_{rm} = r(r+1)\phi^{(N)}_{rm},$$

and

$$\hat{J}_3\phi^{(N)}_{rm} = m\phi^{(N)}_{rm}. \tag{27}$$

The states $\phi^{(N)}_{rm}$ are exactly the states defined in our earlier Example C of two-level states. Now let us see the action of transition operators on these states. We have from elementary angular momentum theory

$$\begin{aligned}&\hat{J}_+\phi^{(N)}_{rm} = \{j(j+1) - m(m+1)\}^{1/2}\phi^{(N)}_{r,m+1},\\ &\hat{J}_-\phi^{(N)}_{rm} = \{j(j+1) - m(m-1)\}^{1/2}\phi^{(N)}_{r,m-1}.\end{aligned} \tag{28}$$

From simple quantum mechanics the transition probabilities are given by[1]

$$I = I_0(j+m)(j-m+1). \tag{29}$$

B. Three-level case

The spin states are of the form (3^N in number)

$$\chi^{(N)} = \chi(i_1,i_2,i_3).$$

The correlated states as usual are the linear combination of the above spin states. The individual transition operators are related to the Gell-Mann matrices[11] defined as

$$\begin{aligned}&\hat{\lambda}_1 = \begin{pmatrix}0&1&0\\1&0&0\\0&0&0\end{pmatrix}, \quad \hat{\lambda}_2 = \begin{pmatrix}0&-i&0\\i&0&0\\0&0&0\end{pmatrix},\\ &\hat{\lambda}_3 = \begin{pmatrix}1&0&0\\0&-1&0\\0&0&0\end{pmatrix}, \quad \hat{\lambda}_4 = \begin{pmatrix}0&0&1\\0&0&0\\1&0&0\end{pmatrix},\\ &\hat{\lambda}_5 = \begin{pmatrix}0&0&-i\\0&0&0\\i&0&0\end{pmatrix}, \quad \hat{\lambda}_6 = \begin{pmatrix}0&0&1\\0&0&1\\0&1&0\end{pmatrix},\\ &\hat{\lambda}_7 = \begin{pmatrix}0&0&0\\0&0&-i\\0&i&0\end{pmatrix}, \quad \hat{\lambda}_8 = \frac{1}{\sqrt{3}}\begin{pmatrix}1&0&0\\0&1&0\\0&0&-2\end{pmatrix}.\end{aligned} \tag{30}$$

It is easy to verify that the analog of Eq. (23) is

$$[\hat{\lambda}_i, \hat{\lambda}_j] = 2if_{ijk}\hat{\lambda}_k, \quad 1 \leqslant i,j,k \leqslant 8, \tag{31}$$

where f_{ijk} is fully antisymmetric, so that it changes sign when any two of its indices are interchanged. The only nonvanishing value of f_{ijk} are permutations of the following:

$$\begin{aligned}&f_{123} = 1, \quad f_{147} = f_{165} = f_{246} = f_{257} = f_{345} = f_{376} = \tfrac{1}{2},\\ &f_{458} = f_{678} = \sqrt{3}/2.\end{aligned} \tag{32}$$

We further define

$$\hat{F}_i = \hat{\lambda}_i/2, \quad [\hat{F}_i, \hat{F}_j] = if_{ijk}\hat{F}_k. \tag{33}$$

The transition operators R are obtained as the linear combinations of the $\hat{F}$'s. A suitable choice is

$$\begin{aligned}\hat{T}_{\pm} &= \hat{F}_1 \pm i\hat{F}_2,\\ \hat{U}_{\pm} &= \hat{F}_6 \pm i\hat{F}_7,\\ \hat{V}_{\pm} &= \hat{F}_4 + i\hat{F}_5.\end{aligned} \tag{34}$$

IV. ATOMIC STATE LABELING

Thus we have proved that the correlated states of the n-level atomic system are the basis of the irreducible representations of the group SU(n). We would like to label each state belonging to a definite representation of the group SU(n). Now we shall solve this problem with the help of the transition operators, i.e., the generators of the group, according to the scheme proposed by Baird and Biedenharn.[12,13]

The number of diagonal generators of the group (also called the rank) is equal to $(n-1)$. Also there exist $(n-1)$ operators called Casimir invariants which commute with the generators of this group. The eigenvalues of these invariants could be used to label a particular representation. The question arises, how many labels are needed to specify state vectors within the irreducible representation? Since the object is to specify the representation matrices, it is clear that we shall need as many "labels" as there are parameters, i.e., one needs as many operators on the group as there are parameters. For SU(n) the problem is to label uniquely the (n^2-1) parameters. We have $(n-1)$ invariant operators which lead to $(n-1)$ labels. Of the remaining $n^2 - 1 - (n-1) = n(n-1)$ labels, we know that diagonal operators furnish $2(n-1)$ labels, the factor 2 entering since the (ij)th matrix element has $\hat{H}_i$ labels for both the ith state and the jth state (or since $\hat{H}_i$ are not invariant operators every matrix ele-

ment $\langle m'|\cdots|m\rangle$ carries two elements for each $\hat{H}_i$). Hence we still need $\frac{1}{2}n(n-1)-2(n-1)=\frac{1}{2}(n-1)(n-2)$ independent operators commuting with each other and with all the diagonal generators in order to label the states uniquely. (The factor $\frac{1}{2}$ occurs for the same reason that each $\hat{H}_i$ defines two labels.)

For example, SU(2) has three generators

$$\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix},\quad \begin{pmatrix}0 & -i\\ i & 0\end{pmatrix},\quad \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}, \tag{35}$$

and

$$\hat{I}_2^{(2)}=3\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}, \tag{36}$$

where $\hat{I}_2^{(2)}$ is the Casimir operator and $\hat{H}_1=\left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$ is the diagonal operator. No further labeling operators are required since $\frac{1}{2}(n-1)(n-2)=0$ for $n=2$. This is the familiar angular momentum example, which is uniquely labeled by $\hat{J}^2=\hat{I}_2^{(2)}$ and $\hat{H}_1(\hat{J}^2$ and $\hat{J}_z)$ as is well known.

Consider now the group SU(3) which has eight generators. In the notation of Gell-Mann, these are the $\hat{F}_i$'s with $i=1,2,\ldots,8$. In Okubo's notation[14] these are the $\hat{A}_i^j$'s $(i,j=1,2,3)$ with the condition $\hat{A}_i^i=0$. The two Casimir invariants have been shown by de Swart[15] to be

$$\hat{I}_3^{(2)}=\sum_{i=1}^{8}\hat{F}_i^2=\frac{1}{2}\hat{A}_i^j\hat{A}_j^i,$$

and

$$\hat{I}_3^{(3)}=\tfrac{1}{2}(\hat{A}_i^j\hat{A}_j^k\hat{A}_k^i+\hat{A}_i^j\hat{A}_k^i\hat{A}_j^k),$$

where

$$(\hat{A}_i^k)_{\mu\nu}=\delta_{\mu\kappa}\delta_{i\nu}-(1/n)\delta_{\mu\nu}\sigma_{ik}. \tag{37}$$

If $\phi(p,q)$ is an eigenstate of an irreducible representation (IR)$D(p,q)$ of SU(3) then it has been shown by de Swart[15] that

$$\begin{aligned}&\hat{I}_3^{(2)}\phi(p,q)=[\tfrac{1}{3}(p^2+pq+q^2)+p+q]\phi(p,q),\\ &\hat{I}_3^{(3)}\phi(p,q)\\ &\quad=[\tfrac{1}{9}(p-q)(2p+q+3)(p+2q+3)]\phi(p,q).\end{aligned} \tag{38}$$

It is now clear that the eigenstate of $\hat{I}_3^{(2)}$ is the same for $D(p,q)$ and $D(q,p)$ which are inequivalent contragradient representations when $p\neq q$. So we need the invariant operator $\hat{I}_3^{(3)}$, which has the property

$$\hat{I}_3^{(3)}\phi(p,q)=-I_3^{(3)}\phi(p,q), \tag{39}$$

and thereby distinguishes between $D(p,q)$ and its conjugate $D(q,p)$. SU(3) has two diagonal "magnetic" operators which in Gell-Mann notation are H_1 and H_2, and

$$H_1=\begin{pmatrix}1 & 0 & 0\\ 0 & -1 & 0\\ 0 & 0 & 0\end{pmatrix},\quad H_2=\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & -2\end{pmatrix}, \tag{40}$$

since $n=3$. In this case $(n-1)(n-2)/2=1$, which means that we need one more operator (in addition to $\hat{I}_3^{(2)}$, $\hat{I}_3^{(3)}$, H_1, and H_2) which commutes with the other four operators to complete the labeling. The Casimir invariant for SU(2) is a suitable operator to complete the designation of the states in SU(3). (See Table II.)

TABLE II. State labeling operators.

Group	Casimir invariants to specify an IR	Diagonal operators to specify uniquely a state within an IR
SU(2)	$\hat{I}_2^{(2)}$	H_1
SU(3)	$I_3^{(2)},I_3^{(3)}$	$H_1,H_2,$ $I_2^{(2)}$
SU(4)	$I_4^{(2)},I_4^{(3)},I_4^{(4)}$	H_1,H_2,H_3 $I_3^{(2)},I_3^{(3)},I_2^{(2)}$
SU(5)	$\hat{I}_5^{(2)},\hat{I}_5^{(3)},\hat{I}_5^{(4)},\hat{I}_5^{(5)}$	$\hat{H}_1,\hat{H}_2,\hat{H}_3,\hat{H}_4;$ $\hat{I}_4^{(2)},\hat{I}_4^{(3)},\hat{I}_4^{(4)};$ $\hat{I}_3^{(2)},\hat{I}_3^{(3)}$ $\hat{I}_2^{(2)}$

The general result for the SU(n) labeling problem is quite immediate. For the SU(n) case, the labeling problem involved by the canonical factorization is

$$\mathrm{SU}(n)\supset\mathrm{U}(1)\otimes\mathrm{SU}(n-1). \tag{41}$$

Here U(1) is the one-parameter Abelian subgroup generated by a linear combination of the $(n-1)$ operators; SU($n-1$) is a subgroup of SU(n). Each of the generators of the SU($n-1$) subgroup must commute with the generator of U(1) in order to define the direct product.

In other words, every state within an IR of SU(n) is characterized by all quantum numbers necessary to characterize a state of SU($n-1$) plus a quantum number due to a linear operator. The "state labeling problem" is equivalent to the problem of finding the complete set of "commuting operators" for the group.

In effect this labeling scheme assigns to every state vector of representation of SU(n) the labeling

$$|I_n^{(2)},I_n^{(3)},\ldots,I_n^{(n)};I_{n-1}^{(2)},\ldots,I_{n-1}^{(n-1)};\ldots I_2^{(2)};H_2\rangle, \tag{42}$$

where by $I_k^{(l)}$ we mean the first power Casimir invariant of the group SU(k) and the H are the diagonal operators of SU(n). The eigenvalues of all these operators have to be specified to characterize a state vector of an IR of SU(n). Biedenharn has proved that the chain decomposition of SU(n) into SU(1)$\times$SU($n-1$) establishes a series of labels exactly sufficient to label every state of all IR uniquely. This completes the classification of states and operators for the multilevel system. In a later paper we discuss the symmetry properties of these states and calculate the matrix elements for transitions among these states.

ACKNOWLEDGMENTS

This work was supported in part by a grant from the Alabama Research Institute.

[1]R. H. Dicke, Phys. Rev. **93**, 99 (1954).
[2]W. Heitler, *The Quantum Theory of Radiation* (Oxford U. P., London, 1954).
[3]E. P. Wigner and V. Weiskopf, Z. Phys. **63**, 54 (1930).
[4]S. Liberman and P. Pillet, Phys. Rev. Lett. **41**, 1237 (1978).

[5]Ph. Cahuzak, H. Sontag, and P. E. Toschek, Opt. Commun. **31**, 37 (1979).
[6]D. B. Lichtenberg, *Unitary Symmetry and Elementary Particles* (Academic, New York, 1970).
[7]H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover, New York, 1931).
[8]H. Weyl, *The Classical Groups* (Princeton U. P., Princeton, NJ, 1946).
[9]I. Schur, Ph.D. dissertation, Berlin, 1901.
[10]H. Boerner, *Representation of Groups* (North-Holland, Amsterdam, 1955).
[11]L. I. Schiff, *Quantum Mechanics* (McGraw-Hill, New York, 1968).
[12]G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 436 (1962).
[13]G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1962).
[14]S. Okubo, Prog. Theor. Phys. (Kyoto) **27**, 949 (1962).
[15]J. J. de Swart, Rev. Mod. Phys. **35**, 916 (1963).

Space-group theory for a nonperiodic icosahedral quasilattice

Peter Kramer
Institut für Theoretische Physik der Universität, Tübingen, Federal Republic of Germany

(Received 22 July 1987; accepted for publication 7 October 1987)

The hypercubic lattice Y is considered in Euclidean space $\mathbb{E}^6$. Embedding of the icosahedral group $A(5)$ determines a well-known noncrystallographic action of $A(5)$ on an irreducible subspace $\mathbb{E}^3$. From boundaries of Y and its metrical dual Y^*, new convex polytopes, called *Klötze*, are constructed in $\mathbb{E}^6$. The *Klötze* tile $\mathbb{E}^6$ periodically, but not face-to-face, and provide a new fundamental domain under the hypercubic translation group. All boundaries of a *Klotz* are parallel or perpendicular to the subspace $\mathbb{E}^3$. The intersection of the lattice Y with $\mathbb{E}^3$ determines a nonperiodic, face-to-face rhombohedral tiling of $\mathbb{E}^3$. A quasicrystal model is introduced in $\mathbb{E}^6$ and $\mathbb{E}^3$. The model has stable density supported on the rhombohedral tiles. Its Fourier transform yields sharp diffraction maxima and reduces to integrals in $\mathbb{E}^3$ over the representative tiles and their boundaries.

I. INTRODUCTION

From the beginning, point and space groups have played an important part in the theory of nonperiodic order in condensed matter physics. The paradigm of the Penrose[1] pattern in the plane $\mathbb{E}^2$ is linked to fivefold rotations, and the icosahedral quasilattices in $\mathbb{E}^3$ (Refs. 2–7) are related to the icosahedral group $A(5)$.

Space groups provide[3] a connection between a periodic lattice Y with space group P, translation group T, and point group G in $\mathbb{E}^n$, and a projection of Y to a subspace $\mathbb{E}^m$, $m<n$: If a subgroup $H<G$ has an irreducible subspace $\mathbb{E}^m$ of $\mathbb{E}^n$ where it acts noncrystallographically, there can be no translation vector from T in $\mathbb{E}^m$, and a projection of Y to $\mathbb{E}^m$ provides a nonperiodic object. For given H with a noncrystallographic action on $\mathbb{E}^m$, the sequence $H<G<P$ can often be obtained by induction.[7] Suppose that the noncrystallographic representation contains a one-dimensional representation of a subgroup $L<H$. Then the representation of H induced from this representation of L has the form of a permutation matrix with signs and allows for an embedding of H into the hyperoctahedral group $\Omega(n)$, where $n=|H|/|L|$. Since $\Omega(n)$ is the point group of the hypercubic space group $(T,\Omega(n))$, one obtains the sequence

$$L<H<\Omega(n)<(T,\Omega(n)). \tag{1}$$

The Penrose case is obtained by the choice $H=C(5)$ with the noncrystallographic action generated from a fivefold rotation in the plane. Induction from $L=I$, the trivial subgroup, gives the sequence $I<C(5)<\Omega(5)<(T,\Omega(5))$. The icosahedral quasilattice of $\mathbb{E}^3$ with $H=A(5)$ and the noncrystallographic representation $[31^2_+]$,[7] the point symmetry group of the icosahedron, is induced from the dihedral subgroup $D(5)$ and gives the sequence $D(5)<A(5)<\Omega(6)<(T,\Omega(6))$. Other icosahedral quasilattices result from the subgroups $D(3)$ and $D(2)$.

In spite of this framework from space groups, fundamental questions with respect to the theory of quasilattices are still open: How is the atomic density supported on the quasilattices, what is the role played by the tiles in this respect, and how can a Fourier transform with sharp diffraction maxima be derived from this atomic density?

A new approach to these questions, based on the space group action in $\mathbb{E}^n$, is developed in Refs. 8 and 9. It employs the following three notions.

(i) A *metrical dual* Y^* to the lattice Y is constructed.

(ii) New polytopes in $\mathbb{E}^n$ called *Klötze* are constructed from boundaries of Y and Y^*. A representative set of *Klötze* provides a *new fundamental domain for the translation group* acting on $\mathbb{E}^n$. Their boundaries are arranged perpendicular or parallel to the subspace $\mathbb{E}^m$. The *intersection* of $\mathbb{E}^m$ with the *periodic tiling* of $\mathbb{E}^n$ *by the Klötze* yields a *nonperiodic tiling* of $\mathbb{E}^m$.

(iii) On the representative set of *Klötze*, the *density* may be restricted so that it is *independent of the coordinates perpendicular to* $\mathbb{E}^m$. On the intersection with $\mathbb{E}^m$, the density then becomes *stable on the tiles*. This restricted model is called the *quasicrystal model*. The *Fourier transform of the density on the intersection* is reduced to *integrals over the representative tiles*.

In the present paper this new approach is implemented for the hypercubic lattice Y in $\mathbb{E}^6$ and its projection to the subspace $\mathbb{E}^3$ invariant under the representation $[31^2_+]$ of $A(5)$.

II. THE HYPERCUBIC LATTICE IN $\mathbb{E}^6$

In this section we fix the notation for the hypercubic lattice Y in $\mathbb{E}^6$. Let $\mathbf{b}_1,\ldots,\mathbf{b}_6$ denote an orthonormal basis and introduce the hypercubic translation group

$$T=\left\{\mathbf{b}\,\middle|\,\mathbf{b}=\sum_{i=1}^{6} n_i b_i,\ n_j\in\mathbb{Z}\right\}. \tag{2}$$

Definition 2.1: The point group of the hypercubic lattice is the hyperoctahedral group $\Omega(6)$. Its elements g are determined by six numbers $\epsilon_i=\pm 1$ and by a permutation $r\in S(6)$. The defining representation is

$$g\to D_{ij}(g)=\epsilon_i\delta_{i,r(j)},\quad i,j=1,\ldots,6. \tag{3}$$

Now choose the hypercube $h(6)$ as the unit cell and consider its boundaries of dimension $p=0,\ldots,5$ which we call *p-boundaries* (compare Ref. 5).

Definition 2.2: The hypercubic lattice Y in $\mathbb{E}^6$ has the first unit cell

 0022-2488/88/020516-09$02.50

$$h(6)=\left\{\mathbf{y}\,\middle|\,\mathbf{y}=\frac{1}{2}\sum_{i=1}^{6}\lambda_i\mathbf{b}_i\right\} \tag{4}$$

and p-boundaries determined by elements g of $\Omega(6)$,

$$h(p;g)=\left\{y\,\middle|\,y=\frac{1}{2}\sum_{i=1}^{p}\lambda_{r(i)}\epsilon_{r(i)}\mathbf{b}_{r(i)}+\frac{1}{2}\sum_{j=p+1}^{6}\epsilon_{r(j)}\mathbf{b}_{r(j)}\right\}, \tag{5}$$

where all parameters λ range as $-1\leqslant\lambda\leqslant 1$.

The description of the boundaries by elements from $\Omega(6)$ is convenient as a shorthand notation. We shall later restrict the point group to the icosahedral group $A(5)$. Then the boundaries should in principle be described by use of elements from the space group $(T, A(5))$, but we shall keep the notation of Definition 2.2 for simplicity. If the first unit cell is translated to all lattice positions in $\mathbb{E}^6$, we obtain a geometric object characterized by boundaries of dimensions $p=0,\ldots,6$. This object will be referred to as the lattice Y.

The translation group T acts on $\mathbb{E}^6$ and yields a decomposition into *orbits*. The first unit cell is, with appropriate restrictions at the boundaries, a *transversal* or *fundamental domain FD* for the action of the translation group T. There are other possible choices of the fundamental domain, and in Sec. V a new fundamental domain will be constructed.

The point group $\Omega(6)$ has the icosahedral group $A(5)$ as a subgroup (compare Refs. 3–5). The subgroup embedding can be interpreted by inducing a six-dimensional representation of $A(5)$ from the nontrivial one-dimensional representation of the dihedral subgroup $D(5)$ (compare Ref. 7). The embedding of $A(5)$ is specified in Table I. In what follows we shall use the space group $(T, A(5))$ whose elements are the products of hypercubic translations with point group elements from $A(5)$. The induced representation subduces the two three-dimensional irreducible representations $[31^2_+]$ and $[31^2_-]$ of $A(5)$. The reducing matrix is given in Table II and is taken from Ref. 6. The two representations determine two subspaces of $\mathbb{E}^6$ which will be denoted as $\mathbb{E}^3_1$ and $\mathbb{E}^3_2$. The projections of the basis $\mathbf{b}_i$ to these subspaces are given by the top and by the bottom part of the columns of this matrix.

TABLE I. Elements of the hyperoctahedral group $\Omega(6)$. The permutation r is given as the map from the top to the bottom line, the sign $\varepsilon=\pm 1$ is marked by a number without or with a bar on top. The elements g_m are of order m and belong to $A(5)$. The elements g_3' and g_2' generate $D(3)$, g_5 and g_2' generate $D(5)$, g_2' and g_2'' generate $D(2)$. Elements g_3 and g_3' are related by $g_3'=g_5^{-1}g_3g_5$. The elements l_α, l_β, and l_γ do not belong to $A(5)$ but are used in the description of boundaries. The element w maps the vectors used in Ref. 5 into the vectors used here.

$$g_5=\begin{vmatrix}1&2&4&\bar5&6&3\\1&4&\bar5&6&3&2\end{vmatrix}$$

$$g_3=\begin{vmatrix}1&2&3&5&6&4\\2&3&1&6&4&5\end{vmatrix}\qquad g_3'=\begin{vmatrix}1&3&6&\bar4&\bar5&2\\3&6&1&\bar5&2&\bar4\end{vmatrix}$$

$$g_2'=\begin{vmatrix}1&3&6&\bar4&\bar5&2\\\bar1&\bar6&\bar3&4&\bar2&5\end{vmatrix}\qquad g_2''=\begin{vmatrix}1&3&6&\bar4&\bar5&2\\\bar4&6&3&1&5&2\end{vmatrix}$$

$$l_\alpha=\begin{vmatrix}1&2&3&4&5&6\\1&3&6&\bar4&\bar5&2\end{vmatrix}\qquad l_\beta=\begin{vmatrix}1&2&3&4&5&6\\\bar4&\bar5&2&1&3&6\end{vmatrix}$$

$$l_\gamma=\begin{vmatrix}1&2&3&4&5&6\\3&6&\bar4&\bar5&2&1\end{vmatrix}\qquad w=\begin{vmatrix}1&2&3&4&5&6\\2&4&\bar5&6&3&1\end{vmatrix}$$

$$l_\delta=\begin{vmatrix}1&2&3&4&5&6\\1&\bar5&2&4&3&6\end{vmatrix}$$

III. THE METRICAL DUAL Y^* OF THE LATTICE Y

In this section we construct a geometrical object dual to Y in a topological and in a metrical sense.

Definition 3.1: Given a p-boundary $h(p;g)$ of the first unit cell according to Definition 2.1, define its dual by

$$h^*(6-p;g)=\left\{y\,\middle|\,y=\frac{1}{2}\sum_{j=p+1}^{6}\lambda_{r(j)}\epsilon_{r(j)}\mathbf{b}_{r(j)}+\frac{1}{2}\sum_{j=p+1}^{6}\epsilon_{r(j)}\mathbf{b}_{r(j)}\right\}, \tag{6}$$

for $p=0,\ldots,5$ and $h^*(0)=\mathbf{0}$.

Note that dual pairs of boundaries have the same center. Now we give the following definition.

Definition 3.2: The dual Y^* to Y is the geometric object in $\mathbb{E}^6$ obtained by application of all translations from T to the dual boundaries for $p=0,\ldots,6$ given in Definition 3.1.

As observed in Ref. 9, dual boundaries have complementary dimension, and the relation of Y and Y^* resembles the relations of dual *Euclidean cell complexes* in the sense of algebraic topology.[10] Moreover, dual boundaries referred to their common center are spanned by orthogonal sets of vectors and hence are dual in a *metrical sense*. Clearly in the present case the dual Y^* taken by itself is a hypercubic lattice as Y with a shifted origin.

IV. *KLÖTZE* IN $\mathbb{E}^6$

In this section we introduce the notion of *Klötze* in the hypercubic lattice Y of $\mathbb{E}^6$ and study their properties. The Euclidean space $\mathbb{E}^6$ will be decomposed into the orthogonal subspaces of dimension 3 which carry the irreducible representations $[31^2_+]$ and $[31^2_-]$, respectively, of the icosahedral group $A(5)$. Denote these two subspaces by the subscripts 1, 2 so that

$$\mathbb{E}^6\to\mathbb{E}^3_1+\mathbb{E}^3_2,\quad \mathbb{E}^3_1\perp\mathbb{E}^3_2. \tag{7}$$

TABLE II. The matrix representation of $A(5)$ of dimension 6 in $\mathbb{E}^6$ is reduced into the irreducible representations $|31^2_+|$ and $|31^2_-|$, respectively, by the matrix m with entries $c=\cos\alpha$, $s=\sin\alpha$, $\tan 2\alpha=\frac{1}{2}$. The columns of this matrix give the components of the six basis vectors of the hypercubic lattice. The projections of these vectors to the irreducible subspaces $\mathbb{E}^3_1$ and $\mathbb{E}^3_2$, respectively, are the first and the last three entries of these columns. The vectors are labeled by their column number.

$$m=\sqrt{\tfrac{1}{2}}\begin{vmatrix}0&c&s&0&c&-s\\s&0&c&-s&0&c\\c&s&0&c&-s&0\\0&-s&c&0&-s&-c\\c&0&-s&-c&0&-s\\-s&c&0&-s&-c&0\end{vmatrix}$$

The corresponding projections of vectors will be denoted by the same subscripts 1,2,

$$\mathbf{y} = \mathbf{y}_1 + \mathbf{y}_2, \quad \mathbf{y}_1 \perp \mathbf{y}_2. \tag{8}$$

For vectors with a subscript index, the projection will be indicated by the second subscript.

Definition 4.1: Given a pair of dual three-boundaries $h(3;g)$ and $h^*(3;g)$ according to Definition 3.1, define the *Klotz* $\mathrm{kl}(3+3;g)$ as the six-dimensional polytope,

$$\mathrm{kl}(3+3;g) = \{\mathbf{x} | \mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2,\ \mathbf{x}_1 \in h_1^*(3;g),\ \mathbf{x}_2 \in h_2(3;g)\} \tag{9}$$

or in explicit form

$$\mathbf{x} = \frac{1}{2}\sum_{j=4}^{6} (\epsilon_{r(j)}\mathbf{b}_{r(j)}) + \frac{1}{2}\sum_{j=4}^{6} (\lambda_{r(j)}\epsilon_{r(j)}b_{r(j)})_1 + \frac{1}{2}\sum_{i=1}^{3} (\lambda_{r(i)}\epsilon_{r(i)}b_{r(i)})_2, \quad -1 \leqslant \lambda_k \leqslant 1. \tag{10}$$

A *Klotz* is completely determined by the projections $h_1^*(3;g)$ and $h_2(3;g)$; we call these projections the *one-* and *two-chart* of the *Klotz*. Any *Klotz* is a bounded and convex polytope in $\mathbb{E}^3$. Two different *Klötze* overlap if and only if they have common points both in their one- and two-charts, respectively.

The boundaries of a *Klotz* are determined by the extremal values of a subset of its parameters λ. In particular we need the following proposition.

Proposition 4.2: A five-boundary of a *Klotz* $\mathrm{kl}(3+3;g)$ is determined by the values $\lambda_k = \pm 1$ of a single parameter λ. A *Klotz* is bounded by 12 five-boundaries.

The five-boundaries have a simple description in terms of their charts: If $k = r(4), r(5), r(6)$, the one-chart is a rhombus face and the two-chart a rhombohedron, and vice versa for $k = r(1), r(2), r(3)$. In $\mathbb{E}^6$, the five-boundaries appear as bounded parts of hyperplanes.

Definition 4.3: The two representative *Klötze* are given by

$$\begin{aligned}
&\mathrm{kl}_\alpha = \mathrm{kl}(3+3;c_1 l_\alpha):\\
&\mathbf{y} = \tfrac{1}{2}(-\mathbf{b}_4 - \mathbf{b}_5 + \mathbf{b}_2)\\
&\quad + \tfrac{1}{2}(-\lambda_4\mathbf{b}_4 - \lambda_5\mathbf{b}_5 + \lambda_2\mathbf{b}_2)_1\\
&\quad + \tfrac{1}{2}(\lambda_1\mathbf{b}_1 + \lambda_3\mathbf{b}_3 + \lambda_6\mathbf{b}_6)_2,\\
&\mathrm{kl}_\beta = \mathrm{kl}(3+3;c_1 l_\beta):\\
&\mathbf{y} = \tfrac{1}{2}(\mathbf{b}_1 + \mathbf{b}_3 + \mathbf{b}_6)\\
&\quad + \tfrac{1}{2}(\lambda_1\mathbf{b}_1 + \lambda_3\mathbf{b}_3 + \lambda_6\mathbf{b}_6)_1\\
&\quad + \tfrac{1}{2}(-\lambda_4\mathbf{b}_4 - \lambda_5\mathbf{b}_5 + \lambda_2\mathbf{b}_2)_2,
\end{aligned} \tag{11}$$

where $c_1 = e$ and l_α, l_β are elements of $\Omega(6)$ given in Table I.

In the one-chart the projections are a thin and a thick rhombohedron, respectively. We shall use the indices α and β to denote these two *Klötze* and their projections.

In Ref. 5 the boundaries of the first unit cell of the hypercubic lattice were classified with respect to their point and space group symmetry. The notations are related in Table I. For three-boundaries it was shown in Ref. 5 that the space group generates at the centers dihedral point groups $D(3)$. For the three-boundaries $h(3;c_1 l_\alpha)$ and $h^*(3;c_1 l_\beta)$ appearing in Definition 4.3, these groups are generated by the space group elements

TABLE III. The generators of the coset $A(5)/D(3)$ have the form $c_j = (g_5)^\mu (g_2'')^\nu$, with $\mu = j-1$, $\nu = 0$ for $j = 1,\ldots,5$ and $\mu = j-6$, $\nu = 1$ for $j = 6,\ldots,10$. Their expressions as elements of $\Omega(6)$ are given in the notation of Table I.

$$c_1 = \begin{vmatrix} 1 & 3 & 6 & \bar4 & \bar5 & 2 \\ 1 & 3 & 6 & \bar4 & \bar5 & 2 \end{vmatrix} \qquad c_6 = \begin{vmatrix} 1 & 3 & 6 & \bar4 & \bar5 & 2 \\ \bar4 & 6 & 3 & 1 & 5 & \bar2 \end{vmatrix}$$

$$c_2 = \begin{vmatrix} 1 & 3 & 6 & \bar4 & \bar5 & 2 \\ 1 & 2 & 3 & 5 & 6 & 4 \end{vmatrix} \qquad c_7 = \begin{vmatrix} 1 & 3 & 6 & \bar4 & \bar5 & 2 \\ 5 & 3 & 2 & 1 & \bar6 & \bar4 \end{vmatrix}$$

$$c_3 = \begin{vmatrix} 1 & 3 & 6 & \bar4 & \bar5 & 2 \\ 1 & 4 & 2 & \bar6 & 3 & \bar5 \end{vmatrix} \qquad c_8 = \begin{vmatrix} 1 & 3 & 6 & \bar4 & \bar5 & 2 \\ \bar6 & 2 & 4 & 1 & \bar3 & 5 \end{vmatrix}$$

$$c_4 = \begin{vmatrix} 1 & 3 & 6 & \bar4 & \bar5 & 2 \\ 1 & \bar5 & 4 & \bar3 & 2 & 6 \end{vmatrix} \qquad c_9 = \begin{vmatrix} 1 & 3 & 6 & \bar4 & \bar5 & 2 \\ \bar3 & 4 & \bar5 & 1 & \bar2 & \bar6 \end{vmatrix}$$

$$c_5 = \begin{vmatrix} 1 & 3 & 6 & \bar4 & \bar5 & 2 \\ 1 & 6 & \bar5 & \bar2 & 4 & 3 \end{vmatrix} \qquad c_{10} = \begin{vmatrix} 1 & 3 & 6 & \bar4 & \bar5 & 2 \\ \bar2 & \bar5 & 6 & 1 & \bar4 & \bar3 \end{vmatrix}$$

$$\begin{aligned}
&D_\alpha(3):\ (0,g_3'),(-\mathbf{b}_4 - \mathbf{b}_5 + \mathbf{b}_2, g_2'),\\
&D_\beta(3):\ (0,g_3'),(\mathbf{b}_1 + \mathbf{b}_3 + \mathbf{b}_6, g_2').
\end{aligned} \tag{12}$$

These dihedral groups transform both the three-boundaries and their duals into themselves. From this property one finds the following proposition.

Proposition 4.4: The two representative *Klötze* given in Definition 4.3 are transformed into themselves by the two dihedral groups $D_\alpha(3)$ and $D_\beta(3)$, respectively.

Since the dihedral groups are subgroups of the space group, we must require that points carried into one another inside of the *Klotz* be equivalent. Let $D(3) < A(5)$ be the dihedral group at the origin and consider the coset $A(5)/D(3)$.

Definition 4.5: The representative set of 20 *Klötze* is given by

$$\mathrm{kl}(3+3;c_i l_\alpha), \quad \mathrm{kl}(3+3;c_i l_\beta), \quad i = 1,\ldots,10. \tag{13}$$

where the group elements c_i generate the coset $A(5)/D(3)$. (See Table III.)

Note that in Definition 4.5 we employ the dihedral group $D(3)$ that refers to the origin. The set will be shown in Sec. V to form a fundamental domain for the translation group. We shall need several properties of this set.

Proposition 4.6: Any pair of *Klötze* from the representative set has an intersection of dimension less than 6.

Proof: We use the charts for the *Klötze*. From the construction one finds

$$\begin{aligned}
&\dim((\mathrm{kl}(3+3;g)) \cap (\mathrm{kl}(3+3;g')))\\
&\quad = \dim((h_1^*(3;g)) \cap (h_1^*(3;g')))\\
&\qquad + \dim((h_2(3;g)) \cap (h_2(3;g'))).
\end{aligned} \tag{14}$$

Therefore it suffices to show that at least in one of the two charts the dimension of the intersection is less than 3. The positions of the projected boundaries $h_2(3;g)$ are given in Ref. 5. The projected boundaries $h_1^*(3;g)$ are spanned by triples of vectors from the origin. In detail one finds

$i \neq j$:

$$\dim(h_1^*(3;c_i l_\beta) \cap h_1^*(3;c_j l_\beta)) \leqslant 2,$$
$$\dim(h_2(3;c_i l_\alpha) \cap h_2(3;c_j l_\alpha)) \leqslant 1,$$
$$\dim(h_2(3;c_i l_\alpha) \cap h_2(3;c_j l_\beta)) = 0, \quad (15)$$
$$\dim(h_1^*(3;c_i l_\alpha) \cap h_1^*(3;c_i l_\beta)) = 3,$$
$$\dim(h_2(3;c_i l_\alpha) \cap h_2(3;c_i l_\beta)) = 0.$$

Now it is of interest to consider the volumes of the *Klötze*.

Proposition 4.7: The *Klötze* of the representative set have the volume

$i = 1,\ldots,10$:

$$\text{vol}(\text{kl}(3+3;c_i l_\alpha)) = \tfrac{1}{8}(1 - 2\theta^3 - 3\theta^2),$$
$$\text{vol}(\text{kl}(3+3;c_i l_\beta)) = \tfrac{1}{8}(1 + 2\theta^3 - 3\theta^2),$$
$$\theta = \sqrt{1/5}, \quad (16)$$

and the total volume of the representative set is

$$\text{vol}\Big(\bigcup_{i=1}^{10} (\text{kl}(3+3;c_i l_\alpha)) \cup (\text{kl}(3+3;c_i l_\beta))\Big) = 1. \quad (17)$$

Proof: The volume of a *Klotz* is the product of the volumes of its projections. The projections are pairs of thin or thick rhombohedra with volume

$$v_{\alpha,\beta} = \sqrt{\tfrac{1}{8}}\,(1 \mp 2\theta^3 - 3\theta^2)^{1/2}, \quad (18)$$

respectively. The total volume of the representative set is then

$$\text{vol} = 10\,\tfrac{1}{8}\,(2 - 6\theta^2) = 1. \quad (19)$$

Turn now in detail to the five-boundaries of the *Klötze*. It will be shown that each one is covered by parts of precisely three five-boundaries from neighboring *Klötze*. Consider the first representative *Klotz* $\text{kl}(3+3;c_1 l_\alpha)$ and its five-boundary described by $\lambda_4 = 1$. There are four possible positions of rhombohedra which cover $h_1^*(3;c_1 l_\alpha)$ from the outside with this face as the intersection. The covering rhombohedra can all be written as translated copies from projections of the representative set. Construct the center positions of these rhombohedra and the projections of their duals. Now we claim the following proposition.

Proposition 4.8: The translated duals to the covering rhombohedra intersect at most in two-boundaries in $\mathbb{E}_2^3$. Three of them dissect the dual $h_2(3;c_1 l_\alpha$ completely into three convex parts. Similar results hold for $h_2(3;c_1 l_\beta)$.

Proof: The translated covering rhombohedra are given in Table IV for selected faces of the rhombohedra in $\mathbb{E}_1^3$ and $\mathbb{E}_2^3$. (See also Table V.) The three covering rhombohedra intersect pairwise in three two-boundaries. These two-boundaries in turn dissect the rhombohedron which forms the dual to the covered one. There are essentially two types of dissections corresponding to the two rhombohedra α, β. They are shown in Figs. 1 and 2. Once these results have been obtained for representative faces, they can be generalized to all faces by use of the dihedral group $D(3)$ acting on the covered rhombohedron and on its covering. □

TABLE IV. Covering of five-boundaries for representative *Klötze* by three five-boundaries from other *Klötze* translated by the vector b.

Covered five-boundary one-chart	Covered five-boundary two-chart	Covering five-boundary one-chart	Covering five-boundary two-chart	Vector b	Dissected chart	Dissecting charts 1	Dissecting charts 2	Dissecting charts 3
				$\text{kl}(3+3;c_1 l_\alpha)$				
$h_1^*(3;c_1 l_\alpha)$	$h_2(3;c_1 l_\alpha)$	$h_1^*(3;c_6 l_\alpha)$	$h_2(3;c_6 l_\alpha)$	$b_2 - b_4 - b_5$	$h_2(3;c_1 l_\alpha)$		$h_2(3;c_6 l_\alpha)$	$h_2(3;c_6 l_\alpha)$
$\lambda_4 = 1$		$\lambda_1 = -1$					$\lambda_6 = 1$	$\lambda_3 = 1$
		$h_1^*(3;c_{10} l_\beta)$	$h_2(3;c_{10} l_\beta)$	$b_2 - b_4$		$h_2(3;c_{10} l_\beta)$	$h_2(3;c_{10} l_\beta)$	
		$\lambda_6 = -1$				$\lambda_3 = -1$	$\lambda_1 = 1$	
		$h_1^*(3;c_7 l_\beta)$	$h_2(3;c_7 l_\beta)$	$-b_4 - b_5$		$h_2(3;c_7 l_\beta)$		$h_2(3;c_7 l_\beta)$
		$\lambda_3 = -1$				$\lambda_6 = -1$		$\lambda_1 = 1$
$h_1^*(3;c_1 l_\alpha)$	$h_2(3;c_1 l_\alpha)$	$h_1^*(3;c_6 l_\alpha)$	$h_2(3;c_6 l_\alpha)$	$b_2 - b_5$	$h_1^*(3;c_1 l_\alpha)$		$h_1^*(3;c_6 l_\alpha)$	$h_1^*(c_6 l_\alpha)$
	$\lambda_1 = 1$		$\lambda_4 = 1$				$\lambda_2 = 1$	$\lambda_5 = 1$
		$h_1^*(3;c_4 l_\beta)$	$h_2(3;c_4 l_\beta)$	$-b_4$		$h_1^*(3;c_4 l_\beta)$	$h_1^*(3;c_4 l_\beta)$	
			$\lambda_2 = 1$			$\lambda_5 = -1$	$\lambda_4 = 1$	
		$h_1^*(3;c_3 l_\beta)$	$h_2(3;c_3 l_\beta)$	$-b_4$		$h_1^*(3;c_3 l_\beta)$		$h_1^*(3;c_3 l_\beta)$
			$\lambda_5 = 1$			$\lambda_2 = -1$		$\lambda_4 = 1$
				$\text{kl}(3+3;c_1 l_\beta)$				
$h_1^*(3;c_1 l_\beta)$	$h_2(3;c_1 l_\beta)$	$h_1^*(3;c_6 l_\beta)$	$h_2(3;c_6 l_\beta)$	$b_1 + b_4$	$h_2(3;c_1 l_\beta)$		$h_2(3;c_6 l_\beta)$	$h_2(3;c_6 l_\beta)$
$\lambda_1 = 1$		$\lambda_4 = 1$					$\lambda_5 = -1$	$\lambda_2 = -1$
		$h_1^*(3;c_3 l_\alpha)$	$h_2(3;c_3 l_\alpha)$	$b_1 + b_6$		$h_2(3;c_3 l_\alpha)$	$h_2(3;c_3 l_\alpha)$	
		$\lambda_5 = -1$				$\lambda_2 = 1$	$\lambda_4 = 1$	
		$h_1^*(3;c_4 l_\alpha)$	$h_2(3;c_4 l_\alpha)$	$b_1 + b_3$		$h_2(3;c_4 l_\alpha)$		$h_2(3;c_4 l_\alpha)$
		$\lambda_2 = -1$				$\lambda_5 = 1$		$\lambda_4 = 1$
$h_1^*(3;c_1 l_\beta)$	$h_2(3;c_1 l_\beta)$	$h_1^*(3;c_6 l_\beta)$	$h_2(3;c_6 l_\beta)$	b_1	$h_1^*(3;c_1 l_\beta)$		$h_1^*(3;c_6 l_\beta)$	$h_1^*(3;c_6 l_\beta)$
	$\lambda_4 = 1$		$\lambda_1 = -1$				$\lambda_6 = -1$	$\lambda_3 = -1$
		$h_1^*(3;c_{10} l_\alpha)$	$h_2(3;c_{10} l_\alpha)$	b_3		$h_1^*(3;c_{10} l_\alpha)$	$h_1^*(3;c_{10} l_\alpha)$	
			$\lambda_6 = 1$			$\lambda_3 = 1$	$\lambda_1 = 1$	
		$h_1^*(3;c_7 l_\alpha)$	$h_2(3;c_7 l_\alpha)$	b_6		$h_1^*(3;c_7 l_\alpha)$		$h_1^*(3;c_7 l_\alpha)$
			$\lambda_3 = 1$			$\lambda_6 = 1$		$\lambda_1 = 1$

TABLE V. Matrix representations of dimension 3, denoted by D_1^3 and D_2^3, of the generators and some elements of $A(5)$ in the irreducible subspaces E_1^3 and E_2^3, respectively.

Generator g	$D_1^3(g)$	$D_2^3(g)$
g_5	$\begin{vmatrix} \Phi^{-1}/2 & \Phi/2 & -\frac{1}{2} \\ -\Phi/2 & \frac{1}{2} & \Phi^{-1}/2 \\ \frac{1}{2} & \Phi^{-1}/2 & \Phi/2 \end{vmatrix}$	$\begin{vmatrix} -\Phi/2 & -\Phi^{-}/2 & -\frac{1}{2} \\ \Phi^{-1}/2 & \frac{1}{2} & -\Phi/2 \\ \frac{1}{2} & -\Phi/2 & -\Phi^{-1}/2 \end{vmatrix}$
g_3	$\begin{vmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix}$	$\begin{vmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix}$
g_2'	$\begin{vmatrix} 1 & 0 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{vmatrix}$	$\begin{vmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{vmatrix}$
g_2''	$\begin{vmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{vmatrix}$	$\begin{vmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{vmatrix}$

Now we apply these results to the five-boundaries of *Klötze*.

Proposition 4.9: Any five-boundary of a *Klotz* is covered by parts of precisely three five-boundaries from three other translated *Klötze*. The covering *Klötze* intersect with the first *Klotz* and with one another at most in five-boundaries.

Proof: Assume that the chart for the five-boundary of the first *Klotz* is a rhombus face in $\mathbb{E}_1^3$ and a rhombohedron in $\mathbb{E}_2^3$. The three covering *Klötze* as given in Table IV are separated from the first *Klotz* by the hyperplane which extends the rhombus face into $\mathbb{E}^6$. Their two-charts according to Table IV are separated by three rhombus faces and dissect the two-chart of the first five-boundary completely. This means that they exhaust all points in the charts of this five-boundary. The result is extended from the representative cases treated in Table IV to all five-boundaries.

V. A NEW FUNDAMENTAL DOMAIN FOR THE HYPERCUBIC LATTICE

In this section we proceed from local to global properties of the *Klötze* in $\mathbb{E}^3$. We start with the following proposition.

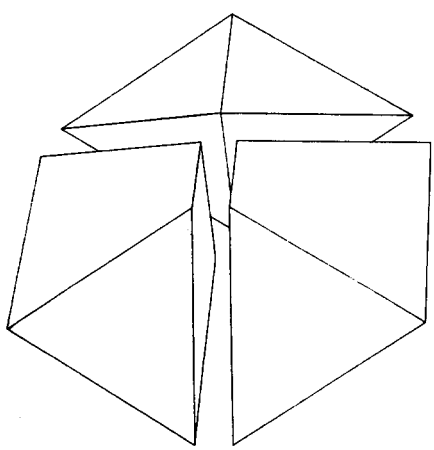


FIG. 1. Dissection of the thin rhombohedron. For a fixed face, there are two thick and one thin rhombohedron which share the face and intersect with the given rhombohedron. These three rhombohedra intersect with one another in three faces, and these three faces dissect the given rhombohedron into three convex parts shown in the figure. This dissection implies for the *Klötze* in $\mathbb{E}^6$ the covering of a five-dimensional boundary by parts of three five-boundaries from other *Klötze*.

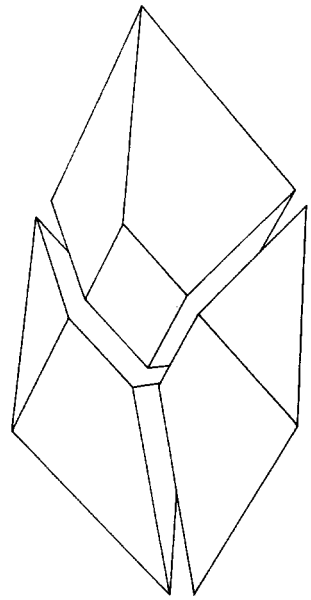


FIG. 2. Dissection of the thick rhombohedron. For a fixed face, there are two thin and one thick rhombohedron which share the face and intersect with the given rhombohedron. Shown is the dissection into three convex parts, with similar implications as given for Fig. 1.

Proposition 5.1: Suppose that $\mathbb{E}^6$ is covered with the bounded union of a finite set of convex polytopes of total volume $V>0$ obtained by translation of the set to all points of the hypercubic lattice Y. If $V=1$ and if the translated polytopes have intersections of dimension 6, then there is an uncovered part in the unit hypercube.

Proof: Let f_0 denote the uncovered part of the unit hypercube and f_i its part which has an i-fold intersection with translated polytopes. Since the set of polytopes is bounded, there is a maximal number $i=m$. Then

$$1=f_0+\sum_{i=1}^{m} f_i. \tag{20}$$

Since there is one set of translated polytopes of volume V per lattice point, we must have

$$V=\sum_{i=1}^{m} if_i \tag{21}$$

and so we get

$$f_0=1-V+\sum_{i=1}^{m}(i-1)f_i. \tag{22}$$

For $V=1$ this yields

$$f_0=\sum_{i=1}^{m}(i-1)f_i, \tag{23}$$

which implies $f_0>0$ if there are parts of dimension 6 of the hypercube with i-fold intersections, $i>1$. □

Proposition 5.2: The union of the representative set of 20 *Klötze* given in Definition 4.5 forms a new fundamental domain on $\mathbb{E}^6$ under the action of the translation group T.

Proof: The representative set is bounded and has total volume $V=1$. The covering of $\mathbb{E}^6$ by its translated copies fulfills the assumptions made in Proposition 5.1, if it is assumed that the covering has intersections of dimension 6. According to this proposition there must be uncovered parts in the first unit hypercube. Since the *Klötze* which form the set are bounded by hyperplanes, the uncovered parts must be bounded in part by five-boundaries from *Klötze*. Then there must be outer points close to these five-boundaries which do

not belong to other *Klötze*. But in Proposition 4.9 it was shown that any five-boundary of a *Klotz* is completely covered by parts of five-boundaries from three other *Klötze*. So one arrives at a contradiction with the result that there can be no uncovered parts and no intersections of dimension 6. The remaining intersections between *Klötze* of dimension less than 6 can be removed by appropriate prescriptions at the boundaries. □

We remark here that the space filling of $\mathbf{E}^6$ by *Klötze* is *not face-to-face* since three neighboring *Klötze* are required to cover a five-boundary. This feature appears already for *Klötze* in spaces of lower dimension.[8,9]

Turn now to the action of the space group $(T, A(5))$. By use of point transformations from the coset $A(5)/D(3)$, the 20 representative *Klötze* can be related to the two representatives of Definition 4.3. These representatives must have $D(3)$ point symmetry.

Proposition 5.3: The two representative *Klötze* $\mathrm{kl}(3+3;c_1 l_\alpha)$ and $\mathrm{kl}(3+3;c_1 l_\beta)$ form a new fundamental domain with respect to the space group $(T, A(5))$ acting on $\mathbf{E}^6$.

VI. THE TWO-CELL QUASICRYSTAL MODEL IN $\mathbf{E}^3$

In this section we study the intersection of the subspace $\mathbf{E}_1^3$ with the lattice Y. We take advantage of the new fundamental domain derived in Sec. V.

Proposition 6.1: The intersection of the subspace $\mathbf{E}_1^3$ with the lattice Y described by translated *Klötze* is an infinite space filling pattern formed from two rhombohedra each in ten different orientations. This pattern is nonperiodic as no translation vector is in the subspace.

Consider now a periodic function f on Y and its value on the intersection. This function can be specified on the two representative *Klötze*. The values of the periodic function on the rhombohedral intersections are dependent on the point $\mathbf{x}_2$ which characterizes the intersection and so *the values of the function f are not stable on the rhombohedra*. Now we *restrict* the values of the periodic function.

Definition 6.2: Let a periodic function f be defined on the *Klötze*. Restrict f by the requirement that, within any *Klotz*,

$$f(\mathbf{x}') = f(\mathbf{x}) \quad \text{if } \mathbf{x}' - \mathbf{x} = \mathbf{x}_2. \tag{24}$$

This requirement ensures that the value of f be *stable on the rhombohedral intersections of the Klötze with* $\mathbf{E}_1^3$. The rhombohedral tiling of $\mathbf{E}_1^3$ with this stable function is called the *two-cell quasicrystal model.*

Note that the rhombohedral tiling in $\mathbf{E}_1^3$ is face-to-face, in contrast to the *Klotz* tiling of $\mathbf{E}^6$.

VII. THE FOURIER TRANSFORM FOR THE QUASICRYSTAL MODEL

In this section we express the Fourier transform of the density f for the two-cell quasicrystal model through the values supported on the representative *Klötze*. We follow the reasoning presented in Ref. 9. First we describe the intersection of a general periodic function with the help of the *cut function*

$$v(\mathbf{y}_1,\mathbf{y}_2) = \delta^3(\mathbf{y}_2 - \mathbf{c}_2), \tag{25}$$

and get, by use of the convolution theorem, the following proposition.

Proposition 7.1: Let the Fourier transform for the function $f(\mathbf{y}_1,\mathbf{c}_2)$ be defined by

$$\tilde{f}(\mathbf{k}_1) = (2\pi)^{-3} \int_{\mathbf{E}^3} d^3\mathbf{y}_1 \, f(\mathbf{y}_1,\mathbf{c}_2)\exp(-i\mathbf{k}_1\cdot\mathbf{y}_1). \tag{26}$$

This Fourier transform is given by

$$\tilde{f}(\mathbf{k}_1) = (2\pi)^3 F(fv)(\mathbf{k}_1,0)$$

$$= \sum_{\mathbf{k}^R \in T^R} \delta^3(\mathbf{k}_1 - \mathbf{k}_1^R)\exp(i\mathbf{k}_2^R\cdot\mathbf{c}_2)a(\mathbf{k}_1^R,\mathbf{k}_2^R), \tag{27}$$

where $\mathbf{k}^R$ is a vector from the reciprocal lattice T^R, F stands for the Fourier transform in $\mathbf{E}^6$, and the symbol a denotes the coefficients of the Fourier series of f in $\mathbf{E}^6$ given by

$$a(\mathbf{k}_1^R,\mathbf{k}_2^R)$$

$$= \int_{\mathrm{FD}} d^3\mathbf{x}_1 \, d^3\mathbf{x}_2 f(\mathbf{x}_1,\mathbf{x}_2)\exp(-i(\mathbf{k}_1^R\cdot\mathbf{x}_1 + \mathbf{k}_2^R\cdot\mathbf{x}_2)). \tag{28}$$

Now we restrict the function f to the two-cell quasicrystal model of Definition 6.2. This can be done by giving two functions f_α and f_β on the two representative *Klötze* which depend only on $\mathbf{x}_1$. We shall introduce the parameters λ as integration variables and define

$$\begin{aligned} f_\alpha(\mathbf{x}_1) &= \hat{f}_\alpha(\lambda_{r(4)},\lambda_{r(5)},\lambda_{r(6)}), \\ f_\beta(\mathbf{x}_1) &= \hat{f}_\beta(\lambda_{r(1)},\lambda_{r(2)},\lambda_{r(3)}). \end{aligned} \tag{29}$$

The symmetry group $D(3)$ of the *Klötze* requires that these functions obey, for $\xi = \alpha,\beta$,

$$\hat{f}_\xi(\lambda,\lambda',\lambda'') = \hat{f}_\xi(\lambda',\lambda''\lambda) = \hat{f}_\xi(-\lambda,-\lambda'',-\lambda'). \tag{30}$$

Definition 7.2: For $\xi = \alpha,\beta$ define auxiliary functions

$$\begin{aligned} d_\xi(\mathbf{k}_1,\mathbf{k}_2,g) &= v_\xi^2 \mathscr{L}(\tfrac{1}{2}\mathbf{k}_2\cdot\mathbf{b}_{r(1)2})\mathscr{L}(\tfrac{1}{2}\mathbf{k}_2\cdot\mathbf{b}_{r(2)2})\mathscr{L}(\tfrac{1}{2}\mathbf{k}_2\cdot\mathbf{b}_{r(3)2})\exp(-i\tfrac{1}{2}\mathbf{k}\cdot(\epsilon_{r(4)}\mathbf{b}_{r(4)} + \epsilon_{r(5)}\mathbf{b}_{r(5)} + \epsilon_{r(6)}\mathbf{b}_{r(6)}))(\tfrac{1}{2})^3 \\ &\times \int_{-1}^{+1}\int_{-1}^{+1}\int_{-1}^{+1} d\lambda_{r(4)}\, d\lambda_{r(5)}\, d\lambda_{r(6)}\, \hat{f}_\xi(\lambda_{r(4)}\lambda_{r(5)}\lambda_{r(6)}) \\ &\times \exp(-i\tfrac{1}{2}\mathbf{k}_1\cdot(\lambda_{r(4)}\epsilon_{r(4)}\mathbf{b}_{r(4)1} + \lambda_{r(5)}\epsilon_{r(5)}\mathbf{b}_{r(5)1} + \lambda_{r(6)}\epsilon_{r(6)}\mathbf{b}_{r(6)1})). \end{aligned} \tag{31}$$

The coefficients v_α, v_β are the volumes of the two rhombohedra given after Proposition 4.7. The letter $\mathscr{L}$ denotes the function $z^{-1}\sin z$. The product of the three functions $\mathscr{L}$ is the Fourier transform of a rhombohedron in $\mathbf{E}_2^3$.

Proposition 7.3: The Fourier transform for the two-cell icosahedral quasicrystal model is given by the expression 7.1

with the coefficients of the Fourier series given by

$$a(\mathbf{k}_1^R,\mathbf{k}_2^R) = \sum_{i=1}^{10} d_\alpha(\mathbf{k}_1^R,\mathbf{k}_2^R,c_i l_\alpha) + \sum_{i=1}^{10} d_\beta(\mathbf{k}_1^R,\mathbf{k}_2^R,c_i l_\beta). \tag{32}$$

The sum runs over the generators of the coset $A(5)/D(3)$ given in Table III.

Proof: Since the function f is independent of $\mathbf{x}_2$, the integration over this variable can be performed to yield the factors given in terms of the function $\mathscr{L}$. The sum extends over the full set of 20 representative *Klötze*; it is equivalent to an integration over the full new fundamental domain. □

VIII. *KLOTZ* BOUNDARIES AND THEIR FOURIER TRANSFORM

The results given in Sec. VII for the Fourier transform in the quasicrystal model apply to a density f supported by interior points of the *Klötze*. As in periodic crystallography in general, the boundaries of cells require a special analysis in view of their different stability group and orbit structure. In addition, the quasicrystal model should be extended to these boundaries. We shall restrict attention to a first class of boundaries of dimension 3, 4, and 5 in $\mathbf{E}^6$ with the property that their one-chart is a vertex, an edge, or a face of the metrical dual Y^* projected to $\mathbf{E}_\parallel^3$ of the form $h_\parallel^*(0)$, $h_\parallel^*(1;g)$, or $h_\parallel^*(2;g)$. As shown in Ref. 5, the stability groups of these p-boundaries of Y^* are $A(5)$, $D(5)$, and $D(2)$ for $p = 0, 1, 2$, respectively. From the point of view of the *Klötze* $\mathrm{kl}(3+3;\tilde{g})$, these boundaries occur in the one-charts $h_\parallel^*(3;\tilde{g})$ as vertices, edges, or faces while their two-charts are full rhombohedra $h_2(3;\tilde{g})$. The *union of these boundaries* admits a description in terms of new *Klotz* polytopes in $\mathbf{E}^6$. With this property, the *Klotz* boundaries of the first class admit an extension of the quasicrystal model.

Proposition 8.1: The first class of boundaries for the *Klötze* $\mathrm{kl}(3+3;\tilde{g})$ has representatives in the form of *Klotz* polytopes of dimension $p+3$ in $\mathbf{E}^6$ given by

$$\mathrm{kl}(p+3;g) = \{\mathbf{y} | \mathbf{y} = \mathbf{y}_1 + \mathbf{y}_2,\ \mathbf{y}_1 \in h_\parallel^*(p;g),\ \mathbf{y}_2 \in h_2(6-p;g)\}. \tag{33}$$

For $p = 0,1,2$ and $g = e, l_\gamma, l_\delta$ (compare Table VI), these *Klotz* polytopes have the stability groups $A(5)$, $D(5)$, and $D(2)$ referred to their centers. The full set of boundaries in $\mathbf{E}^6$ is obtained by first acting with the coset generators $c_i' \in A(5)/D(5)$ or $c_i'' \in A(5)/D(2)$ for $p = 1, 2$, then acting with all translations $b \in T$ on the representatives.

Proof: Start from the set of representative *Klötze* of Definition 4.5 and first consider their one-chart. Choose a fixed boundary of dimension $p = 0, 1, 2$ which is the projection of a p-boundary $h^*(p;g)$ of Y^*. Then determine a set of *Klötze* which, upon translation with a vector $b \in T$, have the following properties: (a) the set shares the boundaries $h^*(p;g)$ and $h^*(0)$ of Y^*; and (b) the set shares a set of interior points depending on a parameter $\epsilon > 0$ which in the limit $\epsilon \to 0$, determines in the one-chart a general point on the chosen boundary. Property (a) implies in the dual Y^* the following property, compare Ref. 5: Since the set of three-boundaries from Y^* share the p-boundary $h^*(p;g)$, the corresponding set of dual three-boundaries is contained in the dual boundary $h(6-p;g)$ of Y. It follows that the two-charts of the set must also be contained in the projection $h_2(6-p;g)$. The *Klötze* are part of a fundamental domain, and therefore property (b) enforces their two-charts to have no intersection of dimension 3. It is then easy to show that these two-charts fill up completely the projection $h_2(6-p;g)$. In the limit $\epsilon \to 0$ one obtains for the collection of boundary points the one-chart $h_\parallel^*(p;g)$ and the two-chart $h_2(6-p;g)$. Now

TABLE VI. *Klötze* $\mathrm{kl}(3+3;\tilde{g})$ shifted by translation vectors $b(\tilde{g})$ with boundaries of dimension $p+3$, $p = 0,1,2$, which contribute to the new *Klotz* polytopes $\mathrm{kl}(0+3)$, $\mathrm{kl}(1+3;l_\gamma)$, and $\mathrm{kl}(2+3;l_\delta)$. The latter polytopes determine representatives of the *Klotz* boundaries and have the stability groups $A(5)$, $D(5)$, and $D(2)$, respectively.

$\tilde{g}$	$p=0$		$p=1$		$p=2$	
	b	λ	b	λ	b	λ
$c_1 l_\alpha$	0	$\lambda_4 = \lambda_5 = \lambda_2 = -1$				
$c_2 l_\alpha$	0	$\lambda_5 = \lambda_6 = \lambda_4 = -1$				
$c_3 l_\alpha$	0	$\lambda_6 = \lambda_3 = \lambda_5 = -1$			b_6	$\lambda_5 = -1$
$c_4 l_\alpha$	0	$\lambda_3 = \lambda_2 = \lambda_6 = -1$			b_3	$\lambda_2 = -1$
$c_5 l_\alpha$	0	$\lambda_2 = \lambda_4 = \lambda_3 = -1$				
$c_6 l_\alpha$	0	$\lambda_1 = \lambda_5 = \lambda_2 = -1$	0	$\lambda_5 = \lambda_2 = -1$		
$c_7 l_\alpha$	0	$\lambda_1 = \lambda_6 = \lambda_4 = -1$	0	$\lambda_6 = \lambda_4 = -1$		
$c_8 l_\alpha$	$b_3 - b_5$	$-\lambda_1 = \lambda_3 = \lambda_5 = 1$	$b_3 - b_5$	$\lambda_3 = \lambda_5 = 1$		
$c_9 l_\alpha$	$b_2 + b_6$	$-\lambda_1 = \lambda_2 = \lambda_6 = 1$	$b_2 + b_6$	$\lambda_2 = \lambda_6 = 1$		
$c_{10} l_\alpha$	b_3	$\lambda_1 = \lambda_4 = -\lambda_3 = -1$	b_3	$\lambda_4 = -\lambda_3 = -1$		
$c_1 l_\beta$	0	$\lambda_1 = \lambda_3 = \lambda_6 = -1$	0	$\lambda_3 = \lambda_6 = -1$	0	$\lambda_1 = -1$
$c_2 l_\beta$	$-b_2$	$\lambda_1 = \lambda_2 = \lambda_3 = -1$	$-b_2$	$\lambda_2 = \lambda_3 = 1$		
$c_3 l_\beta$	$-b_4$	$\lambda_1 = \lambda_4 = \lambda_2 = -1$	$-b_4$	$-\lambda_4 = \lambda_2 = -1$		
$c_4 k_\beta$	$-b_4 + b_5$	$-\lambda_1 = \lambda_5 = \lambda_4 = 1$	$-b_4 + b_5$	$\lambda_5 = \lambda_4 = 1$		
$c_5 l_\beta$	b_5	$\lambda_1 = \lambda_6 = -\lambda_5 = -1$	b_5	$\lambda_6 = -\lambda_5 = -1$		
$c_6 l_\beta$	b_4	$-\lambda_4 = \lambda_6 = \lambda_3 = -1$			b_4	$\lambda_4 = 1$
$c_7 l_\beta$	$-b_5$	$-\lambda_5 = \lambda_3 = \lambda_2 = -1$				
$c_8 l_\beta$	b_6	$-\lambda_6 = \lambda_2 = \lambda_4 = -1$				
$c_9 l_\beta$	b_3	$-\lambda_3 = \lambda_4 = \lambda_5 = -1$				
$c_{10} l_\beta$	b_2	$-\lambda_2 = \lambda_5 = \lambda_6 = -1$				

we construct the sets with the properties (a), (b). For $p=0$, 1, 2, we choose (compare Table VI),

$$
\begin{aligned}
&p=0:\ h^*(0)\ \text{independent of } g,\\
&\quad \mathbf{x}(\epsilon)=\epsilon(\mu_1\mathbf{b}_1+\mu_3\mathbf{b}_3+\mu_6\mathbf{b}_6),\\
&\quad 1>\mu_1>\mu_3>\mu_6>0,\quad 1>\epsilon>0;\\
&p=1:\ g=l_\gamma,\\
&\quad \mathbf{x}(\epsilon)=\mu_1\mathbf{b}_1+\epsilon(\mu_3\mathbf{b}_3+\mu_6\mathbf{b}_6),\\
&\quad 1>\mu_1>\mu_3>\mu_6>0,\quad 1>\epsilon>0;\\
&p=2:\ g=l_\delta,\\
&\quad \mathbf{x}(\epsilon)=\epsilon\mu_1\mathbf{b}_1+\mu_3\mathbf{b}_3+\mu_6\mathbf{b}_6,\\
&\quad 1>\mu_3>\mu_6>\mu_1>0,\quad 1>\epsilon>0.
\end{aligned}
\tag{34}
$$

In Table VI we give the *Klötze*, the translation vectors $\mathbf{b}$, and the parameters λ for the boundaries. In the limit $\epsilon\to 0$, the two-charts of the set fill up the projections $h_2(6-p;g)$. □

Now we extend the quasicrystal model described in Definition 6.2 to the *Klotz* boundaries of the first class: We require that the density f on the boundaries be *independent of* $\mathbf{x}_2$. For the Fourier transform this means that the integration with respect to $\mathbf{x}_2$ can be performed and yields a factor which is the Fourier transform of the two-chart $h_2(6-p;g)$ for $p=0,1,2$. We proceed as in Sec. VII and use the representatives and the coset generators.

Definition 8.2: For the *Klotz* polytopes $\mathrm{kl}(p+3;g)$ and for a density f introduce the following auxiliary functions:

$$
\begin{aligned}
&p=0,\\
&\quad h_1^*(0)=\{\mathbf{x}_1|\mathbf{x}_1=0\},\\
&\quad h_2(6)=\left\{\mathbf{x}_2|\mathbf{x}_2=\frac{1}{2}\sum_{i=1}^{6}\lambda_i\mathbf{b}_{i2}\right\},\\
&\quad f(\mathbf{x}_1)=q\delta(\mathbf{x}_1),\\
&\quad d(\mathbf{k}_1,\mathbf{k}_2)=q\int_{h_2(6)}d^3\mathbf{x}_2\exp(-i\mathbf{k}_2\cdot\mathbf{x}_2);\\
&p=1,\\
&\quad h_1^*(1;g')=\{\mathbf{x}_1|\mathbf{x}_1=\tfrac{1}{2}(1+\lambda_{r(6)})\epsilon_{r(6)}\mathbf{b}_{r(6)1}\},\\
&\quad h_2(5;g')=\left\{\mathbf{x}_2|\mathbf{x}_2=\frac{1}{2}\sum_{i=1}^{5}\lambda_{r(i)}\epsilon_{r(i)}\mathbf{b}_{r(i)2}+\frac{1}{2}\epsilon_{r(6)}\mathbf{b}_{r(6)2}\right\},\\
&\quad f(\mathbf{x}_1)=\hat f(\lambda_{r(6)})=\hat f(-\lambda_{r(6)}),\\
&\quad d(\mathbf{k}_1,\mathbf{k}_2,g')=\int_{h_2(1;g')}d^3\mathbf{x}_2\exp(-i\mathbf{k}_2\cdot\mathbf{x}_2)\\
&\qquad\qquad\times\frac{1}{2}\int_{-1}^{1}d\lambda_{r(6)}\hat f(\lambda_{r(6)})\\
&\qquad\qquad\times\exp\left(-i\frac{1}{2}\mathbf{k}_1\cdot\lambda_{r(6)}\epsilon_{r(6)}\mathbf{b}_{r(6)1}\right);\\
&p=2,\\
&\quad h_1^*(2;g'')=\left\{\mathbf{x}_1|\mathbf{x}_1=\tfrac{1}{2}\sum_{j=5}^{6}(1+\lambda_{r(j)})\epsilon_{r(j)}\mathbf{b}_{r(j)1}\right\},
\end{aligned}
\tag{35}
$$

$$
\begin{aligned}
&\quad h_2(4;g'')=\left\{\mathbf{x}_2|\mathbf{x}_2=\frac{1}{2}\sum_{i=1}^{4}\lambda_{r(i)}\epsilon_{r(i)}\mathbf{b}_{r(i)2}+\frac{1}{2}\sum_{j=5}^{6}\epsilon_{r(j)}\mathbf{b}_{r(j)2}\right\},\\
&\quad f(\mathbf{x}_1)=\hat f(\lambda_{r(5)},\lambda_{r(6)})=\hat f(-\lambda_{r(5)},-\lambda_{r(6)})=\hat f(\lambda_{r(6)},\lambda_{r(5)}),\\
&\quad d(\mathbf{k}_1,\mathbf{k}_2,g'')=\int_{h_2(4;g'')}d^3\mathbf{x}_2\exp(-i\mathbf{k}_2\cdot\mathbf{x}_2)\left(\frac{1}{2}\right)^2\\
&\qquad\qquad\times\int_{-1}^{+1}\int_{-1}^{+1}d\lambda_{r(5)}\,d\lambda_{r(6)}\hat f(\lambda_{r(5)},\lambda_{r(6)})\\
&\qquad\qquad\times\exp\left(-i\frac{1}{2}\mathbf{k}_1\cdot\sum_{j=5}^{6}(1+\lambda_{r(j)})\right.\\
&\qquad\qquad\left.\times\epsilon_{r(j)}\mathbf{b}_{r(j)1}\right).
\end{aligned}
$$

The functions $\hat f$ are introduced to incorporate the required symmetry of the boundaries under $D(5)$ and $D(2)$, respectively.

Proposition 8.1: The Fourier transform of the quasicrystal model for a density f on a klotz boundary as described in Proposition 8.1 is the same sum over reciprocal lattice points as given in Proposition 7.1 but with the Fourier coefficients replaced by

$$
\begin{aligned}
&p=0,\\
&\quad a(\mathbf{k}_1^R,\mathbf{k}_2^R)=d(\mathbf{k}_1^R,\mathbf{k}_2^R);\\
&p=1,\\
&\quad a(\mathbf{k}_1^R,\mathbf{k}_2^R)=\sum_{i=1}^{6}d(\mathbf{k}_1^R,\mathbf{k}_2^R,c_i'l_\gamma);\\
&p=2,\\
&\quad a(\mathbf{k}_1^R,\mathbf{k}_2^R)=\sum_{i=1}^{15}d(\mathbf{k}_1^R,\mathbf{k}_2^R,c_i''l_\delta).
\end{aligned}
\tag{36}
$$

The point group elements c_i' and c_i'' are the generators of the cosets $A(5)/D(5)$ given in Table VII and $A(5)/D(2)$ given in Table VIII. The tables are arranged so that the set of numbers $r(1),\ldots,r(6)$ and the signs $\epsilon_{r(1)},\ldots,\epsilon_{r(6)}$ appear in the bottom row of the group elements c_i',c_i''.

The integration over $\mathbf{x}_2$ in these expressions yields the characteristic function of the polytopes $h_2(p;g)$. These polytopes are well-known zonohedra with $p(p-1)$ rhombus faces which were related to the present quasilattice in Ref. 4.

IX. CONCLUSION

The quasicrystal model proposed here is characterized by an interplay between nonperiodic tiling, diffraction prop-

TABLE VII. The generators of the coset $A(5)/D(5)$.

$c_1'=$	3	6	$\bar4$	$\bar5$	2	1	$c_4'=$	3	6	$\bar4$	$\bar5$	2	1
	3	6	$\bar4$	$\bar5$	2	1		2	1	3	$\bar5$	6	4
$c_2'=$	3	6	$\bar4$	$\bar5$	2	1	$c_5'=$	3	6	$\bar4$	$\bar5$	2	1
	6	1	$\bar5$	2	$\bar4$	3		4	1	2	6	$\bar3$	$\bar5$
$c_3'=$	3	6	$\bar4$	$\bar5$	2	1	$c_6'=$	3	6	$\bar4$	$\bar5$	2	1
	3	1	6	4	5	2		5	1	4	3	$\bar2$	6

TABLE VIII. The generators of the coset $A(5)/D(2)$.

$$c_1'' = \begin{vmatrix} 1 & \bar{5} & 2 & 4 & 3 & 6 \\ 1 & \bar{5} & 2 & 4 & 3 & 6 \end{vmatrix} \quad c_6'' = \begin{vmatrix} 1 & \bar{5} & 2 & 4 & 3 & 6 \\ 3 & 2 & \bar{4} & 5 & 6 & 1 \end{vmatrix} \quad c_{11}'' = \begin{vmatrix} 1 & \bar{5} & 2 & 4 & 3 & 6 \\ 6 & \bar{2} & 1 & \bar{5} & 3 & \bar{4} \end{vmatrix}$$

$$c_2'' = \begin{vmatrix} 1 & \bar{5} & 2 & 4 & 3 & 6 \\ 1 & 6 & 4 & \bar{5} & 2 & 3 \end{vmatrix} \quad c_7'' = \begin{vmatrix} 1 & \bar{5} & 2 & 4 & 3 & 6 \\ 2 & 4 & 5 & \bar{6} & 3 & 1 \end{vmatrix} \quad c_{12}'' = \begin{vmatrix} 1 & \bar{5} & 2 & 4 & 3 & 6 \\ 3 & \bar{4} & 1 & 6 & 2 & 5 \end{vmatrix}$$

$$c_3'' = \begin{vmatrix} 1 & \bar{5} & 2 & 4 & 3 & 6 \\ 1 & 3 & \bar{5} & 6 & 4 & 2 \end{vmatrix} \quad c_8'' = \begin{vmatrix} 1 & \bar{5} & 2 & 4 & 3 & 6 \\ 4 & \bar{5} & \bar{6} & \bar{3} & 2 & 1 \end{vmatrix} \quad c_{13}'' = \begin{vmatrix} 1 & \bar{5} & 2 & 4 & 3 & 6 \\ 2 & 5 & 1 & 3 & 4 & \bar{6} \end{vmatrix}$$

$$c_4'' = \begin{vmatrix} 1 & \bar{5} & 2 & 4 & 3 & 6 \\ 1 & 2 & 6 & 3 & \bar{5} & 4 \end{vmatrix} \quad c_9'' = \begin{vmatrix} 1 & \bar{5} & 2 & 4 & 3 & 6 \\ \bar{5} & 6 & \bar{3} & \bar{2} & 4 & 1 \end{vmatrix} \quad c_{14}'' = \begin{vmatrix} 1 & \bar{5} & 2 & 4 & 3 & 6 \\ 4 & \bar{6} & 1 & 2 & \bar{5} & \bar{3} \end{vmatrix}$$

$$c_5'' = \begin{vmatrix} 1 & \bar{5} & 2 & 4 & 3 & 6 \\ 1 & 4 & 3 & 2 & 6 & \bar{5} \end{vmatrix} \quad c_{10}'' = \begin{vmatrix} 1 & \bar{5} & 2 & 4 & 3 & 6 \\ 6 & 3 & \bar{2} & \bar{4} & \bar{5} & 1 \end{vmatrix} \quad c_{15}'' = \begin{vmatrix} 1 & \bar{5} & 2 & 4 & 3 & 6 \\ \bar{5} & \bar{3} & 1 & 4 & 6 & \bar{2} \end{vmatrix}$$

erties, and atomic positions. These concepts have an independent history.

As for the tiles, Kepler in 1611 (Ref. 11) introduced the triacontahedron bounded by 30 rhombus faces and discussed it in relation to the morphology of crystals. Kowalewski in 1938 (Ref. 12) derived this polytope as the projection of the hypercube from $\mathbb{E}^6$ and decomposed it into ten thin and ten thick rhombohedra. The other members of this family of zonohedra with $l(p) = p(p-1), p = 3,...,6$ rhombus faces were found by Fedorov in 1885 and Bilinski in 1961 (compare Coxeter[13]). Mackay[14,15] proposed these two rhombohedra as tiles for a nonperiodic pattern in $\mathbb{E}^3$, and Kramer and Neri[3] constructed the nonperiodic infinite tiling. Dubost *et al.*[16] in 1986 have prepared large quasicrystals with the morphology of Kepler's triacontahedron.

The class of functions with a Fourier transform supported on a discrete set of points was shown by Bohr[17] to be almost periodic. Functions of this type can be obtained by restricting periodic functions in $\mathbb{E}^n$ to an intersection with $\mathbb{E}^m$, $m < n$. This property has been used by Janner and Janssen[18,19] to describe incommensurate structures. If a general density on Y in $\mathbb{E}^6$ is subject to the symmetry of the space group $(T, A(5))$, it will be almost repeated on $\mathbb{E}^3$ whenever the intersection between Y and $\mathbb{E}^3$ is almost repeated. The Fourier transform in the subspace will then depend on a full six-dimensional density and cannot be stable on tiles. For an illuminating discussion of these points we refer to Bak.[20] To understand the physics of quasicrystals, it becomes essential to distinguish different models and to test their implications on the atomic level.

The particular diffraction properties of the icosahedral quasilattice obtained from $\mathbb{E}^6$ were studied by Gähler and Rhyner,[21] Levine and Steinhardt,[22] Elser,[23] Kramer,[24] Duneau and Katz,[25,26] and others. As a result, the diffraction pattern is determined by the characteristic function of the hypercube projected to $\mathbb{E}_2^3$. In the present model this corresponds to the boundary with the one-chart $h_1^*(0)$ and the two-chart $h_2(6)$ of Proposition 8.3. To the authors knowledge, closed expressions for all other atomic positions discussed here have not appeared in the literature.

The present quasicrystal model implies a nonperiodic tiling of $\mathbb{E}^3$ which is almost repeated along with the almost repetition of intersections between Y and $\mathbb{E}^3$. The atomic density is exactly repeated on the two tiles. This model comes closest to periodic order but requires more than a single cell. In the simpler version of dimension 3 and 2, the model[9] has been shown to reduce to periodic order for commensurate intersections.

ACKNOWLEDGMENT

This work was supported by DFG.

[1]R. Penrose, Bull. Inst. Math. Appl. **10**, 266 (1974).
[2]P. Kramer, Acta Crystallogr. A **38**, 257 (1982).
[3]P. Kramer and R. Neri, Acta Crystallogr. A **40**, 580 (1984).
[4]P. Kramer, Z. Naturforsch. **40a**, 775 (1985).
[5]P. Kramer, Z. Naturforsch. **41a**, 897 (1986).
[6]P. Kramer, Acta Crystallogr. A **43**, 486 (1987).
[7]R. W. Haase, L. Kramer, P. Kramer and H. Lalvani, Acta Crystallogr. A **43**, 574 (1987).
[8]P. Kramer, Mod. Phys. Lett. B **1**, 7 (1987).
[9]P. Kramer, Int. J. Mod. Phys. B **1**, 145 (1987).
[10]W. A. Efremowitsch, "Grundbegriffe der Topologie," in *Enzycl. d. Elementarmathematik* (Deutscher Verlag d. Wissenschaften, Berlin, 1971); *Encycl. Dict. of Math., Transl.*, 244, 820, edited by N. Sugakkai (MIT Press, Cambridge, 1977).
[11]J. Kepler, "Strena seu de nive sexangula," in *Gesammelte Werke Bd. IV*, edited by M. Caspar and F. Hammer (C. H. Beck, München, 1941).
[12]G. Kowalewski, *Der Keplersche Körper und andere Bauspiele* (Koehlers Antiquarium, Leipzig, 1938).
[13]H. S. M. Coxeter, *Regular Polytopes* (Macmillan, New York, 1963).
[14]A. L. Mackay, Sov. Phys. Crystallogr. **26**, 517 (1981).
[15]A. L. Mackay, Physica A **114**, 609 (1982).
[16]B. Dubost, J.-M. Lang, M. Tanaka, P. Sainfort, and M. Audier, Nature **324**, 48 (1986).
[17]H. Bohr, *Fastperiodische Funktionen* (Springer, Berlin, 1932).
[18]A. Janner and T. Janssen, Physica A **99**, 47 (1979).
[19]T. Janssen and A. Janner, Physica A **126**, 163 (1984).
[20]P. Bak, "Icosahedral incommensurate crystals," in *Scaling Phenomena in Disordered Systems*, edited by R. Pynn and A. Skjeltorp (Plenum, New York, 1985).
[21]F. Gähler and J. Rhyner, J. Phys. Gen. **19**, 267 (1986).
[22]D. Levine and P. J. Steinhardt, Phys. Rev. Lett. **53**, 2477 (1984).
[23]V. Elser, Acta Crystallogr. A **42**, 36 (1986).
[24]P. Kramer, Phys. Lett. A **111**, 133 (1985).
[25]M. Duneau and A. Katz, Phys. Rev. Lett. **54**, 2688 (1985).
[26]A. Katz and M. Duneau, J. Phys. **47**, 181 (1986).

Erratum: Symmetries of static, spherically symmetric space-times [J. Math. Phys. 28, 1019 (1987)]

Ashfaque H. Bokhari and Asghar Qadir
Department of Physics, University of Texas, Austin, Texas 78712

(Received 27 October 1987; accepted for publication 28 October 1987)

Professor H. Stephani (Jena) has pointed out a counterexample to the theorem given in the above-mentioned paper, namely the static Einstein universe, which has seven independent Killing vectors (see Ref. 1).

[1]D. Kramer, H. Stephani, M. A. H. McCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (Cambridge U. P., Cambridge, 1980).

 0022-2488/88/020525-01$02.50