Representations of SL(2,*R*) in a Hilbert space of analytic functions and a class of associated integral transforms

Debabrata Basu
Department of Physics, Indian Institute of Technology, Kharagpur 721302, West Bengal, India

(Received 4 April 1988; accepted for publication 27 July 1988)

It is shown that the boson operators of SL(2,R) realized as hyperdifferential operators in Bargmann's Hilbert space of analytic functions yield, on exponentiation, a parametrized continuum of integral transforms. Each value of the group parameters yields an integral transform pair. For the metaplectic representation the resulting integral transform is essentially the mapping of the Moshinsky–Quesne transform in Bargmann's Hilbert space $B(C)$. The formula for the inversion of this transform is obtained simply by replacing the group element by its inverse. The corresponding Hilbert space for arbitrary representations of the discrete series is $B(C_2)$, where C_2 is the two-dimensional complex Euclidean space. To carry out the reduction of $B(C_2)$ into the eigenspaces $B_k(C)$ ($k=\frac{1}{2},1,\frac{3}{2},\ldots$) of irreducible representations of the positive discrete class, the complex polar coordinates ($z_1 = z\cos\phi$, $z_2 = z\sin\phi$) in C_2 are introduced. The "reduced Bargmann space" $B_k(C)$ has many interesting features. The elements of $B_k(C)$ are entire functions of the complex "radius" z analytic in the upper half-plane. In contrast to the Gaussian measure in $B(C_2)$, the integration measure in the scalar product in $B_k(C)$ contains a modified Bessel function of the second kind. The principal vector in $B_k(C)$, on the other hand, is a modified Bessel function of the first kind. The resulting integral transform maps $B_k(C)$ onto itself and the integral kernel is the product of an exponential and a modified Bessel function of the first kind. The inversion formula for this transform is obtained again by replacing the group element by its inverse.

I. INTRODUCTION

Bargmann's Hilbert space[1] $B(C)$ consists of entire analytic functions $f(z)$, $z\in C$, having a finite norm according to the scalar product

$$(f,g) = \int \overline{f(z)}g(z)d\mu(z), \tag{1.1a}$$

where $d\mu(z)$ is the Gaussian measure

$$d\mu(z) = (e^{-|z|^2}/\pi)d^2z, \quad d^2z = dx\,dy, \quad z = x+iy. \tag{1.1b}$$

The isomorphic mapping of $B(C)$ onto the conventional quantum mechanical $L^2(R)$ Hilbert space is given by Bargmann's integral transform. In a previous paper[2] we have shown that the Bargmann transform constitutes an integral transform pair within the Heisenberg–Weyl group. This Hilbert space was used by Bargmann[3] later for the analysis of the rotation group. His method is closely related to Schwinger's boson realization[4] of the angular momentum operators.

The object of this paper is to show that the use of the Hilbert space $B(C)$ as the carrier space of unitary irreducible representations (UIR's) of SL(2,R) leads to a parametrized continuum of integral transforms mapping $B(C)$ onto itself. Each value of the group parameters yields an integral transform pair. Similar integral transforms in $L^2(R)$ Hilbert space were obtained by Moshinsky,[5] Wolf,[6] and co-workers among others[7] in their investigations on the role of canonical transformations in quantum mechanics.

The first step towards the stated objective is an explicit transcription of the well known Hermitian boson operators of SL(2,R) (Ref. 8) in Bargmann's Hilbert space $B(C)$. In contrast to the canonical realization of Gel'fand *et al.*[9] and Bargmann[10] in which the group acts transitively as a group of point transformations in a function space, the group action in this construction is an integral transform in $B(C)$. The integral kernel for the inversion of this transform is obtained simply by replacing the group element by its inverse. Our method of exponentiation of the generators is based on an adaption of the discussion of Barut and Raczka[11] on the theory of "heat equation" on a Lie group and analytic vectors. We first factorize the unitary operator of the representation into an appropriate Baker–Campbell–Hausdorff formula by using a theorem due to Wilcox.[12] The successive application of the operator factors on an element of $B(C)$ yields the integral transform pair.

We first consider the metaplectic representation[13] ($D_{\text{met}} = D_{1/4}\oplus D_{3/4}$) of SL(2,$R$). The resulting integral kernel is an exponential function and the integral transform is essentially the mapping of the Moshinsky–Quesne transform[14] in Bargmann's Hilbert space of analytic functions. We next discuss a few simple transform pairs for special values of the group parameters. The Plancherel formula for this transform is obtained easily from the unitarity of the representation. It should, however, be pointed out that an integral transform in $B(C)$ may be mapped onto a point transformation in $L^2(R)$ and vice versa. For example, the Fourier transform in $L^2(R)$ corresponds to a point transformation in $B(C)$.

We next proceed to perform a parallel analysis for arbitrary representations of the discrete series. For this case the Hilbert space $B(C_2)$ consists of entire analytic functions $f(z_1,z_2)$ of two complex variables z_1 and z_2. To carry out the reduction of $B(C_2)$ into the eigenspaces $B_k(C)$

 0022-2488/89/010001-08$02.50

($k=\frac{1}{2},1,\frac{3}{2},\ldots$) of the UIR's of the positive discrete class, we introduce the "complex polar coordinates" $z_1 = z\cos\phi$, $z_2 = z\sin\phi$, where the "radius" z, and the "angle" ϕ are both complex variables. The "reduced Bargmann space" $B_k(C)$ shares many features of the Bargmann space $B(C)$ discussed above. The elements $f(z)$ of $B_k(C)$ are entire functions analytic in the upper half-plane ($\operatorname{Im} z>0$) whose behavior near the origin is of the form

$$f(z)\approx \text{const}\, z^{2k-1}. \tag{1.2}$$

The scalar product in $B_k(C)$ is of the form

$$(f,g)=\int_{\operatorname{Im} z>0} \overline{f(z)}g(z)d\lambda(z). \tag{1.3}$$

The measure $d\lambda(z)$ which replaces the Gaussian measure (1.1b) in $B(C)$ is given by

$$d\lambda(z)=(2/\pi)|z|^2K_{2k-1}(|z|^2)d^2z, \tag{1.4}$$

where $K_\nu(z)$ stands for the modified Bessel function of the second kind and the integral extends over the half-plane $\operatorname{Im} z>0$. A complete orthonormal set in $B_k(C)$ is given by

$$u_n=\frac{2^{-k-n+1/2}z^{2k-1+2n}}{[(2k+n-1)!n!]^{1/2}},\quad n=0,1,2,\ldots. \tag{1.5}$$

The principal vector or the reproducing kernel in $B_k(C)$ is a modified Bessel function of the first kind,

$$K(z,\xi)=\overline{e_z(\xi)}=I_{2k-1}(z\bar{\xi}). \tag{1.6}$$

The resulting integral transform maps $B_k(C)$ onto itself and the integral kernel contains, in addition to an exponential factor, a modified Bessel function of the first kind. The inversion formula for the transform follows, as before, by replacing the group element by its inverse. The Plancherel formula for the transform pair once again is essentially the statement of the unitarity of the representation. The corresponding reduction for the principal series is entirely different and will be considered elsewhere.

II. FUNDAMENTAL FORMULAS AND THE METAPLECTIC REPRESENTATION

We first briefly describe some basic properties of Bargmann's Hilbert space $B(C)$ that will be needed in the subsequent development. The elements of $B(C)$ are entire analytic functions $f(z)$ having a finite norm according to the scalar product (1.1). The scalar product satisfies

$$(zf,g)=\left(f,\frac{\partial g}{\partial z}\right). \tag{2.1}$$

A complete orthonormal set in $B(C)$ is givn by the powers,

$$u_n(z)=z^n/\sqrt{n!},\quad n=0,1,2,\ldots. \tag{2.2}$$

We now introduce the principal vectors e_z that are bounded linear functionals in $B(C)$ satisfying

$$f(z)=(e_z,f). \tag{2.3}$$

The explicit form of the principal vectors is given by

$$e_z(\xi)=e^{\bar{z}\xi},$$

so that Eq. (2.3) reads

$$f(z)=\int e^{z\bar{\xi}}f(\xi)d\mu(\xi). \tag{2.4}$$

In a previous paper[2] we have shown that Eq. (2.4) in conjunction with finiteness of the norm $\|f\|$ ensures the analyticity and entireness of $f(z)$.

The group $SL(2,R)$ consists of real unimodular matrices

$$g=\begin{pmatrix} a & b\\ c & d\end{pmatrix},\quad \det g=ad-bc=1, \tag{2.5a}$$

and is isomorphic to the group $SU(1,1)$ of pseudounitary unimodular matrices

$$u=\begin{pmatrix} \alpha & \beta\\ \bar{\beta} & \bar{\alpha}\end{pmatrix},\quad \det u=|\alpha|^2-|\beta|^2=1. \tag{2.5b}$$

The parameters of $g\in SL(2,R)$ can be related to the parameters of $u\in SU(1,1)$ through

$$u=\begin{pmatrix} \alpha & \beta\\ \bar{\beta} & \bar{\alpha}\end{pmatrix} =\frac{1}{2}\begin{pmatrix} a+d+i(b-c) & b+c+i(a-d)\\ b+c-i(a-d) & a+d-i(b-c)\end{pmatrix}. \tag{2.6}$$

The Lie algebra of $SL(2,R)\sim SU(1,1)\sim Sp(2R)$ is defined by the commutation relation

$$[J_1,J_2]=-iJ_3,\quad [J_2,J_3]=iJ_1,\quad [J_3,J_1]=iJ_2. \tag{2.7}$$

In the fundamental representation (2.5b)

$$J_3=\sigma_3/2,\quad J_1=i\sigma_1/2,\quad J_2=i\sigma_2/2.$$

To construct a unitary representation of the group in $B(C)$, we introduce the following formal solution of the commutation relation:

$$\begin{aligned} J_1&=\frac{1}{4}\left(z^2+\frac{d^2}{dz^2}\right),\\ J_2&=-\frac{i}{4}\left(z^2-\frac{d^2}{dz^2}\right),\\ J_3&=\frac{1}{2}\left(z\frac{d}{dz}+\frac{1}{2}\right). \end{aligned} \tag{2.8}$$

Equation (2.1) ensures that the operators (2.8) are Hermitian under the scalar production (1.1). The representation generated by the operators (2.8) is the direct sum

$$D_{1/4}\oplus D_{3/4},$$

which is the so-called metaplectic representation. The subset of the orthonormal vectors (2.2) for even n ($n=0,2,4,\ldots$) corresponds to $D_{1/4}$ and that for odd n ($n=1,3,5,\ldots$) to $D_{3/4}$.

A finite element of the group is obtained by exponentiating the operators (2.8). For this we introduce the Euler angle parametrization

$$\begin{pmatrix} \alpha & \beta\\ \bar{\beta} & \bar{\alpha}\end{pmatrix} =\begin{pmatrix} e^{i\theta'/2} & 0\\ 0 & e^{-i\theta'/2}\end{pmatrix}\begin{pmatrix} \cosh(\tau/2) & -\sinh(\tau/2)\\ -\sinh(\tau/2) & \cosh(\tau/2)\end{pmatrix} \times\begin{pmatrix} e^{i\theta/2} & 0\\ 0 & e^{-i\theta/2}\end{pmatrix},$$

$$0\leqslant\theta'\leqslant 2\pi,\quad 0\leqslant\theta\leqslant 4\pi,\quad 0\leqslant\tau<\infty, \tag{2.9a}$$

so that

$$\alpha = e^{i(\theta' + \theta)/2}\cosh(\tau/2),$$
$$\beta = -e^{i(\theta' - \theta)/2}\sinh(\tau/2). \tag{2.9b}$$

An arbitrary element of the group according to this parametrization is given by

$$T_u = e^{i\theta' J_3} e^{i\tau J_1} e^{i\theta J_3}. \tag{2.10}$$

We shall show that the action of the operator T_u on an arbitrary element $f(z) \in B(C)$ is an ingetral transform.

Since J_3 is a first-order operator, the action of $\exp(i\theta J_3)$ is simple:

$$f_\theta(z) = e^{i\theta J_3} f(z) = e^{i\theta/4} f(e^{i\theta/2} z). \tag{2.11}$$

To obtain the action of $\exp(i\tau J_1)$ we first proceed to obtain a Baker–Campbell–Hausdorff formula for exponentials of operators of the type

$$e^{\alpha(d^2/dz^2) + \beta z^2} \tag{2.12}$$

by using a theorem due to Wilcox.[12]

Let P and Q be any two operators satisyfing the commutation relation

$$[P,Q] = cI, \tag{2.13}$$

where c is a complex number. Thus P and Q may be the annihilation and creation operators, momentum and coordinate operators, etc. Then Wilcox's theorem[12] states that

$$e^{\alpha P^2 + \beta Q^2 + \gamma QP} = [Je^{c\gamma}]^{-1/2} N[e^{AP^2 + BQ^2 + GQP}], \tag{2.14}$$

where N stands for the normal ordering operator which, acting on $f(P,Q)$, moves all the P's to the right of the Q's and

$$\begin{aligned} &\alpha^{-1}A = \beta^{-1}B = (\lambda J)^{-1}\sinh\lambda, \\ &G = c^{-1}(J^{-1} - 1), \\ &J = \cosh\lambda - \rho\gamma\sinh\lambda, \quad \rho \equiv \lambda^{-1}c, \\ &\quad \lambda = c[\gamma^2 - 4\alpha\beta]^{1/2}. \end{aligned} \tag{2.15}$$

To apply this formula to our operator we first set $\alpha = \beta = 0$, then $A = B = 0, \lambda = c\gamma, J = e^{-c\gamma}$ so that

$$\gamma = (1/c)\ln(cG + 1). \tag{2.16}$$

This immediately yields

$$N[e^{GQP}] = e^{(1/c)\ln(cG+1)QP}. \tag{2.17}$$

If we now set $\gamma = 0$ in Wilcox's formula (2.14) we obtain

$$\begin{aligned} e^{\alpha P^2 + \beta Q^2} &= J^{-1/2} e^{BQ^2} N[e^{GQP}] e^{AP^2} \\ &= J^{-1/2} e^{BQ^2} e^{(1/c)\ln(cG+1)QP} e^{AP^2}. \end{aligned} \tag{2.18}$$

Setting $\alpha = \beta = i\tau/4, Q = z$, and $P = d/dz$, so that $c = 1$, we finally obtain

$$\begin{aligned} e^{i\tau J_1} = &\left(\cosh\frac{\tau}{2}\right)^{-1/2} \exp\left(\frac{i}{2}\tanh\frac{\tau}{2} z^2\right) \\ &\times \exp\left(\ln \operatorname{sech}\frac{\tau}{2} z\frac{d}{dz}\right) \exp\left(\frac{i}{2}\tanh\frac{\tau}{2}\frac{d^2}{dz^2}\right). \end{aligned} \tag{2.19}$$

To determine the action of the above operator on $f(e^{i\theta/2}z)$ we use the fundamental property of the principal vectors as given by Eq. (2.4),

$$f(e^{i\theta/2}z) = \int e^{ze^{i\theta/2}\bar{\xi}} f(\xi)\, d\mu(\xi). \tag{2.20}$$

We first operate Eq. (2.20) with the second-order operator appearing on the extreme right of Eq. (2.19). Thus

$$\begin{aligned} &\exp\left(\frac{i}{2}\tanh\frac{\tau}{2}\frac{d^2}{dz^2}\right) f(e^{i\theta/2}z) \\ &\quad = \int \exp\left(\frac{i}{2} e^{i\theta}\tanh\frac{\tau}{2}\bar{\xi}^2 + ze^{i\theta/2}\bar{\xi}\right) f(\xi)\, d\mu(\xi). \end{aligned} \tag{2.21a}$$

Applying the remaining factors successively we have

$$\begin{aligned} &e^{i\tau J_1} f(e^{i\theta/2}z) \\ &\quad = \left(\cosh\frac{\tau}{2}\right)^{-1/2} \int \exp\left[\frac{i}{2}\tanh\frac{\tau}{2}(z^2 + e^{i\theta}\bar{\xi}^2)\right. \\ &\qquad \left. + \operatorname{sech}\frac{\tau}{2} e^{i\theta/2} z\bar{\xi}\right] f(\xi)\, d\mu(\xi). \end{aligned} \tag{2.21b}$$

Applying once again the operator $\exp(i\theta' J_3)$ on both sides of Eq. (2.21b), using Eqs. (2.11) and (2.9b), and setting

$$[T_u f](z) = g_u(z), \tag{2.22}$$

we obtain

$$\begin{aligned} g_u(z) = &(\bar{\alpha})^{-1/2} \\ &\times \int \exp\left(\frac{-i}{2\bar{\alpha}}(\beta z^2 + \bar{\beta}\bar{\xi}^2 + 2iz\bar{\xi})\right) f(\xi)\, d\mu(\xi), \end{aligned} \tag{2.23}$$

which is an integral transform mapping $B(C)$ onto itself.

The formula for the inversion of the transform follows immediately by noting

$$f(\xi) = [T_{u^{-1}} g_u](\xi), \tag{2.24a}$$

since

$$u^{-1} = \begin{pmatrix} \bar{\alpha} & -\beta \\ -\bar{\beta} & \alpha \end{pmatrix}. \tag{2.24b}$$

We immediately obtain from (2.23) and (2.24)

$$\begin{aligned} f(\xi) = &(\alpha)^{-1/2} \\ &\times \int \exp\left(\frac{i}{2\alpha}(\beta\xi^2 + \bar{\beta}\bar{z}^2 - 2i\xi\bar{z})\right) g_u(z)\, d\mu(z). \end{aligned} \tag{2.25}$$

Equations (2.23) and (2.25) constitute an integral transform pair for each allowed value of the group parameters α and β.

We now consider some simple special cases. The transform pair for

$$u = \begin{pmatrix} \sqrt{2} & 1 \\ 1 & \sqrt{2} \end{pmatrix}$$

is given by

$$g(z) = 2^{-1/4} \times \int \exp\left(\frac{-i}{2\sqrt{2}}(z^2 + \bar{\xi}^2 + 2iz\bar{\xi})\right) f(\xi) d\mu(\xi),$$

$$f(\xi) = 2^{-1/4} \int \exp\left(\frac{i}{2\sqrt{2}}(\xi^2 + \bar{z}^2 - 2i\xi\bar{z})\right) g(z) d\mu(z). \tag{2.26}$$

Similarly for

$$u = \begin{pmatrix} \sqrt{2} & i \\ -i & \sqrt{2} \end{pmatrix},$$

the transform pair is given by

$$g(z) = 2^{-1/4} \int e^{(1/2\sqrt{2})(z^2 - \bar{\xi}^2 + 2z\bar{\xi})} f(\xi) d\mu(\xi),$$
$$f(\xi) = 2^{-1/4} \int e^{(1/2\sqrt{2})(\bar{z}^2 - \xi^2 + 2\xi\bar{z})} g(z) d\mu(z). \tag{2.27}$$

The Plancherel formula for the transform pair is obtained from the unitarity of the representation

$$(f_1, f_2) = (T_u f_1, T_u f_2) = (g_{1u}, g_{2u}),$$

which yields

$$\int \overline{f_1(\xi)} f_2(\xi) d\mu(\xi) = \int \overline{g_{1u}(z)} g_{2u}(z) d\mu(z). \tag{2.28}$$

To establish the connection of the integral transform with the canonical transformation of Moshinsky and Quesne[5] we start from

$$T_u z f = (\bar{\alpha})^{-1/2} \int \exp\left(\frac{-i}{2\bar{\alpha}}(\beta z^2 + \bar{\beta}\bar{\xi}^2 + 2iz\bar{\xi})\right) \times \xi f(\xi) d\mu(\xi). \tag{2.29}$$

Using the analyticity of $f(\xi)$ and integrating by parts we obtain

$$T_u z f = \alpha z T_u f - i\bar{\beta} \frac{\partial}{\partial z} T_u f. \tag{2.30}$$

This is equivalent to the operator condition

$$T_u z T_u^{-1} = \alpha z - i\bar{\beta} \frac{\partial}{\partial z}. \tag{2.31}$$

In a similar manner

$$T_u \frac{\partial}{\partial z} T_u^{-1} = i\beta z + \bar{\alpha} \frac{\partial}{\partial z}. \tag{2.32}$$

If we now introduce the Fock–Bargmann representation of the coordinate and momentum operators

$$Q = \frac{1}{\sqrt{2}}\left(z + \frac{\partial}{\partial z}\right), \quad P = \frac{i}{\sqrt{2}}\left(z - \frac{\partial}{\partial z}\right), \tag{2.33}$$

and use Eqs. (2.6) connecting the SU(1,1) and $SL(2,R)$ parameters, we obtain

$$Q' = T_u Q T_u^{-1} = dQ + bP, \tag{2.34a}$$

$$P' = T_u P T_u^{-1} = cQ + aP. \tag{2.34b}$$

To get the mapping of the integral transform pair [Eqs. (2.23) and (2.25)] in $L^2(R)$ space we introduce the Bargmann transform

$$v_g(x) = \pi^{-1/4} \times \int \exp\left(-\frac{1}{2}(\bar{z}^2 + x^2) + \sqrt{2}\bar{z}x\right) g_u(z) d\mu(z). \tag{2.35}$$

In the left-hand side (lhs) we have replaced subscript $u \in SU(1,1)$ by its $SL(2,R)$ image g because the final result takes a simple form in terms of the $SL(2,R)$ parameters.

Substituting Eq. (2.23) in Eq. (2.35) we obtain

$$v_g(x) = \pi^{-1/4}(\bar{\alpha})^{-1/2} \times \int \exp\left(-\frac{i\bar{\beta}\bar{\xi}^2}{2\alpha} - \frac{x^2}{2}\right) I(\bar{\xi}, x) f(\xi) d\mu(\xi), \tag{2.36}$$

where

$$I(\bar{\xi}, x) = \int \exp\left(\frac{1}{2}\gamma z^2 + az + \frac{1}{2}\bar{\delta}\bar{z}^2 + \bar{b}\bar{z}\right) d\mu(z),$$
$$\gamma = -i\beta/\bar{\alpha}, \quad a = \bar{\xi}/\bar{\alpha}, \quad \bar{\delta} = -1, \quad \bar{b} = \sqrt{2}x. \tag{2.37}$$

The above integral has been evaluated by Bargmann[1] and the result is

$$I(\bar{\xi}, x) = [\bar{\alpha}/(\bar{\alpha} - i\beta)]^{1/2} \times \exp[(1/(\bar{\alpha} - i\beta))(-\bar{\xi}^2/2\bar{\alpha} - i\beta x^2 + \sqrt{2}\bar{\xi}x)]. \tag{2.38}$$

We now replace the function $f(\xi)$ appearing in Eq. (2.36) by

$$f(\xi) = \pi^{-1/4} \int \exp\left(-\frac{1}{2}(\xi^2 + y^2) + \sqrt{2}\xi y\right) u(y) dy. \tag{2.39}$$

Carrying out the ξ integration and simplifying the result we have

$$v_g(x) = \frac{\exp[i \operatorname{sgn} b(\pi/4)]}{\sqrt{2\pi|b|}} \times \int \exp\left(\frac{1}{2ib}(dx^2 - 2xy + ay^2)\right) u(y) dy. \tag{2.40}$$

The inversion formula for the transform follows from the corresponding formula in $B(C)$, namely, Eq. (2.25), or from the requirement

$$u(y) = [T_{g^{-1}} v_g](y), \tag{2.41}$$

and is given by

$$u(y) = \frac{\exp[-i \operatorname{sgn} b(\pi/4)]}{\sqrt{2\pi|b|}} \times \int \exp\left(-\frac{1}{2ib}(dx^2 - 2xy + ay^2)\right) v_g(x) dx. \tag{2.42}$$

Equations (2.40) and (2.42) constitute the one-dimensional version of the Moshinsky–Quesne transform. The integral trnsform pair [Eqs. (2.23) and (2.25)] is, therefore, a realization of the linear canonical transformation in the Bargmann space. An integral transform in $L^2(R)$ may, however, be mapped onto a point transformation in $B(C)$, and vice versa. For example the Fourier transform obtained by setting $a = d = 0$ and $b = -c = 1$ corresponds to

$$g_u(z) = f(iz),$$

which is a point transformation in $B(C)$.

III. THE DISCRETE REPRESENTATIONS D_k^+

A. The infinitesimal operators

To obtain arbitrary representations of the discrete class we consider in place of Eqs. (2.8) the following set of Hermitian generators:

$$\begin{aligned} J_1 &= \frac{1}{4}\left(z_1^2 + z_2^2 + \frac{\partial^2}{\partial z_1^2} + \frac{\partial^2}{\partial z_2^2}\right), \\ J_2 &= -\frac{i}{4}\left(z_1^2 + z_2^2 - \frac{\partial^2}{\partial z_1^2} - \frac{\partial^2}{\partial z_2^2}\right), \\ J_3 &= \frac{1}{2}\left(z_1\frac{\partial}{\partial z_1} + z_2\frac{\partial}{\partial z_2} + 1\right). \end{aligned} \tag{3.1}$$

The representation D of SL$(2,R)$ generated by the above operators is reducible and is a direct sum of all the UIR's belonging to the positive discrete series of representations

$$D = \sum_{\substack{\oplus \\ k = 1/2, 1, \ldots}} D_k^+ . \tag{3.2}$$

The generators of D_k^- are obviously obtained by replacing the operators J_3 and J_1 by $-J_3$ and $-J_1$, respectively.

If we now introduce the Hermitian operator

$$K = \frac{1}{2} - \frac{i}{2}\left(z_1\frac{\partial}{\partial z_2} - z_2\frac{\partial}{\partial z_1}\right), \tag{3.3}$$

the Casimir operator becomes a function of K:

$$J_1^2 + J_2^2 - J_3^2 = K(1 - K). \tag{3.4}$$

B. The reduced Bargmann space

The subspace $B_k(C)$ of the representation space $B(C_2)$, in which the operator K is a number, will be called the reduced Bargmann space. The form (3.3) of the operator K suggests that we introduce the polar coordinates

$$z_1 = z\cos\phi, \quad z_2 = z\sin\phi, \tag{3.5}$$

where the radius z and the angle ϕ are both complex numbers:

$$\begin{aligned} &z = |z|e^{i\arg z}, \quad 0 \leqslant \arg z \leqslant \pi, \\ &\phi = \phi_1 + i\phi_2, \quad 0 \leqslant \phi_1 \leqslant 2\pi, \quad -\infty < \phi_2 < \infty. \end{aligned} \tag{3.6}$$

The operator K and the generators J_1, J_2, J_3 are now given by

$$K = \frac{1}{2} - \frac{i}{2}\frac{\partial}{\partial\phi}, \tag{3.7}$$

$$\begin{aligned} J_1 &= \frac{1}{4}\left[z^2 + \frac{\partial^2}{\partial z^2} + \frac{1}{z}\frac{\partial}{\partial z} - \frac{(2K-1)^2}{z^2}\right], \\ J_2 &= \frac{i}{4}\left[z^2 - \frac{\partial^2}{\partial z^2} - \frac{1}{z}\frac{\partial}{\partial z} + \frac{(2K-1)^2}{z^2}\right], \\ J_3 &= \frac{1}{2}\left[z\frac{\partial}{\partial z} + 1\right]. \end{aligned} \tag{3.8}$$

Since K is diagonal in the subspace $B_k(C)$ of the UIR's D_k^+, we have

$$f(z_1,z_2) = e^{i(2k-1)\phi} f(z), \tag{3.9}$$

where $f(z)$ is an analytic function regular in the upper half-plane $0 \leqslant \arg z \leqslant \pi$.

To obtain the scalar product in $B_k(C)$ we start from the scalar product in $B(C_2)$,

$$(f,g) = \int \overline{f(z_1,z_2)}\, g(z_1,z_2)\, d\mu(z_1)\, d\mu(z_2), \tag{3.10}$$

where the integral extends over C_2. Using the transformations (3.5) we obtain after some calculations

$$d^2z_1\, d^2z_2 = |z|^2\, d^2z\, d\phi_1\, d\phi_2, \tag{3.11a}$$

$$e^{-|z_1|^2 - |z_2|^2} = e^{-|z|^2\cosh 2\phi_2}. \tag{3.11b}$$

Thus

$$\begin{aligned} (f,g) = \frac{2}{\pi}\delta_{kk'} &\int_{\mathrm{Im}\, z>0} \overline{f(z)}\, g(z)\, |z|^2\, d^2z \\ &\times \int_{-\infty}^{\infty} \exp[-(2k-1)2\phi_2 - |z|^2\cosh 2\phi_2]\, d\phi_2. \end{aligned} \tag{3.12}$$

Using the standard integral representation[15] of the modified Bessel function of the second kind,

$$K_n(x) = \frac{1}{2}\int_{-\infty}^{\infty} e^{-vn - x\cosh v}\, dv, \tag{3.13}$$

we obtain the scalar product in $B_k(C)$,

$$(f,g) = \int_{\mathrm{Im}\, z>0} \overline{f(z)}\, g(z)\, d\lambda(z), \tag{3.14}$$

where

$$d\lambda(z) = (2/\pi)|z|^2 K_{2k-1}(|z|^2)\, d^2z. \tag{3.15}$$

We now introduce the principal vectors e_z that are bounded linear functionals in $B_k(C)$ satisfying

$$f(z) = (e_z, f). \tag{3.16}$$

To find the explicit form of e_z in $B_k(C)$ we start from the two-dimensional version of Eq. (2.4),

$$f(z_1,z_2) = \int e^{z_1\bar{\xi}_1 + z_2\bar{\xi}_2} f(\xi_1,\xi_2)\, d\mu(\xi_1)\, d\mu(\xi_2). \tag{3.17}$$

We now introduce the polar coordinates in both (z_1,z_2) and (ξ_1,ξ_2) and restrict ourselves to functions of the form (3.9). Thus

$$\begin{aligned} e^{i(2k-1)\phi} f(z) = \frac{1}{\pi^2}\int_{\mathrm{Im}\,\xi>0} &\exp[z\bar{\xi}\cos(\phi - \bar{\Psi}) \\ &+ i(2k-1)\Psi - |\xi|^2\cosh 2\Psi_2] \\ &\times f(\xi)\, |\xi|^2\, d^2\xi\, d\Psi_1\, d\Psi_2, \end{aligned} \tag{3.18}$$

where ξ and Ψ are defined by

$$\begin{aligned}
&\xi_1 = \xi \cos\Psi, \quad \xi_2 = \xi \sin \Psi,\\
&\xi = |\xi| e^{i \arg \xi}, \quad 0 \leqslant \arg \xi \leqslant \pi,\\
&\Psi = \Psi_1 + i\Psi_2, \quad 0 \leqslant \Psi_1 \leqslant 2\pi, \quad -\infty < \Psi_2 < \infty,\\
&f(\xi_1,\xi_2) = \exp[i(2k-1)\Psi] f(\xi).
\end{aligned} \tag{3.19}$$

We now note that Eq. (3.18) can be written in the form

$$e^{i(2k-1)\phi} f(z) = \frac{1}{\pi^2} \int_{\operatorname{Im} \xi > 0} d^2\xi\, |\xi|^2 f(\xi) I(z,\bar{\xi},\xi,\phi), \tag{3.20}$$

where

$$\begin{aligned}
I(z,\bar{\xi},\xi,\phi) &= \int_{\Psi_2=-\infty}^{\infty} \int_{\Psi_1=0}^{2\pi} \exp[z\bar{\xi}\cos(\phi - \bar{\Psi}) + i(2k-1)\Psi]\\
&\quad \times \exp(-|\xi|^2 \cosh 2\Psi_2) d\Psi_1\, d\Psi_2.
\end{aligned} \tag{3.21}$$

By a simple change of variables the above integral can be recast in the form

$$\begin{aligned}
&\exp\left[i(2k-1)\left(\phi + \frac{\pi}{2}\right)\right]\\
&\quad \times \frac{1}{2} \int d\alpha_2 \exp[-(2k-1)\alpha_2 - |\xi|^2 \cosh \alpha_2]\\
&\quad \times \int_0^{2\pi} d\alpha_1 \exp[z\bar{\xi} \sin \alpha - i(2k-1)\alpha],
\end{aligned} \tag{3.22}$$

where α_1 and α_2 are the real and imaginary parts of the complex number α,

$$\alpha = \alpha_1 + i\alpha_2, \quad 0 \leqslant \alpha_1 \leqslant 2\pi, \quad -\infty < \alpha_2 < \infty. \tag{3.23}$$

We first rewrite the α_1 integral as an integral over a circle s of radius $\rho = e^{-\alpha_2}$ centered at the origin. Thus writing J for the α_1 integral we obtain

$$J = \frac{1}{i} \int_s du\, e^{(\eta/2)(u - u^{-1})} u^{-(2k-1)-1}, \tag{3.24a}$$

where

$$\eta = -iz\bar{\xi}. \tag{3.24b}$$

The integral appearing in the rhs of Eq. (3.24a) is the standard contour integral representation of the Bessel function[16] and we have

$$J = 2\pi e^{-i(2k-1)\pi/2} I_{2k-1}(z\bar{\xi}), \tag{3.25}$$

where $I_\mu(z)$ stands for the modified Bessel function of the first kind.

Using Eq. (3.13), the α_2 integration can now be easily carried out and we have

$$I(z,\bar{\xi},\xi,\phi) = 2\pi e^{i(2k-1)\phi} I_{2k-1}(z\bar{\xi}) K_{2k-1}(|\xi|^2). \tag{3.26}$$

Substituting Eq. (3.26) in Eq. (3.20) we immediately obtain

$$f(z) = \int_{\operatorname{Im} \xi > 0} I_{2k-1}(z\bar{\xi}) f(\xi) d\lambda(\xi). \tag{3.27}$$

The principal vector in $B_k(C)$ defined by Eq. (3.16) is therefore given by

$$e_z(\xi) = I_{2k-1}(\bar{z}\xi). \tag{3.28}$$

We shall now show that the elements $f(z)$ of $B_k(C)$ are entire analytic functions whose behavior near the origin is of the form

$$f(z) \approx \text{const}\; z^{2k-1}. \tag{3.29}$$

To prove this we start by noting that a complete orthonormal set in $B_k(C)$ is given by the powers

$$u_n(z) = \frac{2^{-k-n+1/2} z^{2k-1+2n}}{[(2k+n-1)!n!]^{1/2}}, \quad n = 0,1,2,\ldots . \tag{3.30}$$

The orthonormality can be easily verified:

$$\begin{aligned}
(u_n,u_m) &= \frac{2^{-2k-n-m+1}}{[(2k+n-1)!(2k+m-1)!n!m!]^{1/2}}\\
&\quad \times \int_{\operatorname{Im} z>0} \bar{z}^{2k-1+2n} z^{2k-1+2m}\, d\lambda(z).
\end{aligned}$$

Setting $z = re^{i\theta}$, $0 \leqslant \theta \leqslant \pi$, $0 \leqslant r < \infty$, $r^2 = x$, we obtain

$$\begin{aligned}
(u_n,u_m) &= \delta_{nm} \frac{2^{-2k-2n+1}}{(2k+n-1)!n!}\\
&\quad \times \int_0^\infty x^{2k+2n} K_{2k-1}(x)\, dx.
\end{aligned}$$

The integral appearing above can be evaluated from the formula[17]

$$\int_0^\infty x^\mu K_\nu(x) dx = 2^{\mu-1} \Gamma\left(\frac{1+\mu+\nu}{2}\right) \Gamma\left(\frac{1+\mu-\nu}{2}\right),$$
$$\operatorname{Re}(\mu + 1 \pm \nu) > 0, \tag{3.31}$$

which immediately yields

$$(u_n,u_m) = \delta_{nm}.$$

The completeness can be ensured by noting that

$$\begin{aligned}
\sum_n (f,u_n)(u_n,g) &= \int_{\operatorname{Im} z>0} d\lambda(z) \overline{f(z)}\\
&\quad \times \left[\int_{\operatorname{Im} \xi>0} \sum_n u_n(z)\, \overline{u_n(\xi)}\right] g(\xi) d\lambda(\xi).
\end{aligned} \tag{3.32}$$

Using the explicit form of the orthonormal vectors u_n as given by Eq. (3.30) we obtain

$$\sum_n u_n(z)\, \overline{u_n(\xi)} = I_{2k-1}(z\bar{\xi}). \tag{3.33}$$

Hence the ξ integral in Eq. (3.32) reads

$$\int_{\operatorname{Im} \xi > 0} I_{2k-1}(z\bar{\xi}) g(\xi) d\lambda(\xi),$$

which is equal to $g(z)$ by Eq. (3.27). Thus

$$\sum_n (f,u_n)(u_n,g) = (f,g) \tag{3.34}$$

and the orthonormal set (3.30) is complete.

To prove that $f(z)$ is an entire function whose behavior near the origin is given by (3.29), we expand the reproducing kernel $I_{2k-1}(z\bar{\xi})$ in Eq. (3.27) in a power series

$$I_{2k-1}(z\bar{\xi}) = \sum_{n=0}^\infty \frac{1}{(2k+n-1)!n!} \left(\frac{z\bar{\xi}}{2}\right)^{2k-1+2n} \tag{3.35}$$

and define

$$a_n = \frac{2^{-2k-2n+1}}{(2k+n-1)!n!} \int \bar{\xi}^{2k-1+2n} f(\xi) d\lambda(\xi). \quad (3.36)$$

Thus Eq. (3.27) yields

$$f(z) = \sum a_n z^{2k-1+2n}. \quad (3.37)$$

We shall show that the radius of convergence of this power series is infinite by requiring the norm $\|f\|$ to be finite. We use the scalar product (3.14) in $B_k(C)$ to calculate the square norm $\|f\|^2$:

$$\|f\|^2 = (f,f) = \sum_{n=0}^{\infty} |a_n|^2 2^{2k+2n-1}(2k+n-1)!n!. \quad (3.38)$$

If $\|f\|$ is to be finite the above series must be absolutely convergent and we have

$$\lim_{n\to\infty} |a_{n+1}/a_n|^2 4n^2 \leqslant 1,$$

i.e.,

$$\lim_{n\to\infty} |a_{n+1}/a_n| \leqslant \frac{1}{2n}. \quad (3.39)$$

Thus writing v_n for the nth term of the power series (3.37), we have

$$\lim_{m\to\infty} \left|\frac{v_{n+1}}{v_n}\right| = \lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| |z|^2 \leqslant \frac{|z|^2}{2n} \to 0,$$

no matter how large $|z|$ is.

The radius of convergence of the Taylor expansion (3.37) is therefore infinite and the analytic function represented by it is an entire function. The behavior of (3.29) near zero is also obvious from (3.32).

We conclude this subsection by giving the explicit forms of the generators of the group in $B_k(C)$:

$$J_1 = \frac{1}{4}\left[z^2 + \frac{d^2}{dz^2} + \frac{1}{z}\frac{d}{dz} - \frac{(2k-1)^2}{z^2}\right],$$

$$J_2 = -\frac{i}{4}\left[z^2 - \frac{d^2}{dz^2} - \frac{1}{z}\frac{d}{dz} + \frac{(2k-1)^2}{z^2}\right], \quad (3.40)$$

$$J_3 = \frac{1}{2}\left[z\frac{d}{dz} + 1\right].$$

C. Finite element of the group and the associated integral transform

To find the action of the finite element,

$$T_u = e^{i\theta' J_3} e^{i\tau J_1} e^{i\theta J_3}, \quad (3.41)$$

of the group on an element $f(z) \in B_k(C)$, we start from

$$e^{i\theta J_3} f(z_1,z_2) = e^{i\theta/2} f_\theta(z_1,z_2), \quad (3.42)$$

where

$$f_\theta(z_1,z_2) = f(z_1 e^{i\theta/2}, z_2 e^{i\theta/2}) = \int e^{(z_1\bar{\xi}_1 + z_2\bar{\xi}_2)e^{i\theta/2}} f(\xi_1,\xi_2) d\mu(\xi_1) d\mu(\xi_2). \quad (3.43)$$

To determine the action of $\exp(i\tau J_1)$ on $f_\theta(z_1,z_2)$, we use the Wilcox decomposition

$$\begin{aligned}
&\exp(i\tau/4)\,[z_1^2 + z_2^2 + \partial^2/\partial z_1^2 + \partial^2/\partial z_2^2] \\
&= \left(\operatorname{sech}\frac{\tau}{2}\right) \exp\left(\frac{i}{2}\tanh\frac{\tau}{2}(z_1^2+z_2^2)\right) \\
&\quad \times \exp\left[\ln \operatorname{sech}\frac{\tau}{2}\left(z_1\frac{\partial}{\partial z_1} + z_2\frac{\partial}{\partial z_2}\right)\right] \\
&\quad \times \exp\left[\frac{i}{2}\tanh\frac{\tau}{2}\left(\frac{\partial^2}{\partial z_1^2} + \frac{\partial^2}{\partial z_2^2}\right)\right].
\end{aligned} \quad (3.44)$$

We first operate both sides of Eq. (3.43) with the second-order operator appearing on the extreme right of Eq. (3.44). Thus

$$\exp\left[\frac{i}{2}\tanh\left(\frac{\tau}{2}\right)\left(\frac{\partial^2}{\partial z_1^2} + \frac{\partial^2}{\partial z_2^2}\right)\right] f_\theta(z_1,z_2) = \int \exp\left(\frac{i}{2} e^{i\theta}\tanh\frac{\tau}{2}(\bar{\xi}_1^2 + \bar{\xi}_2^2) + (z_1\bar{\xi}_1 + z_2\bar{\xi}_2)e^{i\theta/2}\right) f(\xi_1,\xi_2) d\mu(\xi_1) d\mu(\xi_2). \quad (3.45)$$

Following the previous subsection we now introduce the polar coordinates and restrict ourselves to $B_k(C)$. Then the rhs of Eq. (3.45) becomes

$$\frac{1}{\pi^2}\int_{\operatorname{Im}\xi>0} \exp\left(\frac{i}{2}e^{i\theta}\tanh\frac{\tau}{2}\bar{\xi}^2\right) \times I(z,\bar{\xi},\xi,\phi;\theta) f(\xi)|\xi|^2 d^2\xi, \quad (3.46)$$

where

$$I(z,\bar{\xi},\xi,\phi;\theta) = \int_{\Psi_2=-\infty}^{\infty}\int_{\Psi_1=0}^{2\pi} \exp[z\bar{\xi}e^{i\theta/2}\cos(\phi - \bar{\Psi}) + i(2k-1)\Psi - |\xi|^2\cosh 2\Psi_2]\,d\Psi_1\,d\Psi_2. \quad (3.47)$$

The above integral is of the same form as $I(z,\bar{\xi},\xi,\phi)$ appearing in Eq. (3.21) with z replaced by $ze^{i\theta/2}$ so that

$$I(z,\bar{\xi},\xi,\phi;\theta) = I(ze^{i\theta/2},\bar{\xi},\xi,\phi) = 2\pi e^{i(2k-1)\phi} I_{2k-1}(e^{i\theta/2}z\bar{\xi}) K_{2k-1}(|\xi|^2). \quad (3.48)$$

Using the above result and applying the remaining operator factors in Eq. (3.44) successively we obtain

$$e^{i\tau J_1}f_\theta(z_1,z_2) = e^{i(2k-1)\phi}\left(\operatorname{sech}\frac{\tau}{2}\right) \times \int_{\operatorname{Im}\xi>0} \exp\left(\frac{i}{2}\tan\frac{\tau}{2}(z^2 + e^{i\theta}\bar{\xi}^2)\right) \times I_{2k-1}\left(e^{i\theta/2}\operatorname{sech}\frac{\tau}{2}z\bar{\xi}\right)f(\xi)d\lambda(\xi). \tag{3.49}$$

We now apply $\exp(i\theta' J_3)$ on both sides of the above equation, which yields

$$e^{i\theta' J_3}e^{i\tau J_1}f_\theta = e^{i(2k-1)\phi}e^{i\theta'/2}\operatorname{sech}\frac{\tau}{2} \times \int_{\operatorname{Im}\xi>0}\exp\left(\frac{i}{2}\tanh\frac{\tau}{2}(e^{i\theta'}z^2 + e^{i\theta}\bar{\xi}^2)\right) \times I_{2k-1}\left(e^{i(\theta+\theta')/2}\operatorname{sech}\frac{\tau}{2}z\bar{\xi}\right)f(\xi)d\lambda(\xi). \tag{3.50}$$

Since in $B_k(C)$ the function $f(z_1,z_2)$ is of the form (3.9), we obtain, after omitting the common factor $\exp[i(2k-1)\phi]$ from both sides of Eq. (3.50),

$$g_u(z) = [T_u f](z) = \frac{1}{\bar{\alpha}}\int_{\operatorname{Im}\xi>0} e^{(-i/2\bar{\alpha})(\beta z^2 + \bar{\beta}\bar{\xi}^2)} \times I_{2k-1}\left(\frac{z\bar{\xi}}{\alpha}\right)f(\xi)d\lambda(\xi). \tag{3.51}$$

The inversion formula for this transform once again follows from

$$f(\xi) = [T_{u^{-1}}g_u](\xi),$$

where u^{-1} is given by Eq. (2.24b). This immediately yields

$$f(\xi) = \frac{1}{\alpha}\int_{\operatorname{Im} z>0} e^{(i/2\alpha)(\beta\xi^2 + \bar{\beta}\bar{z}^2)}I_{2k-1}\left(\frac{\xi\bar{z}}{\alpha}\right)g_u(z)d\lambda(z). \tag{3.52}$$

Equations (3.51) and (3.52) constitute a parametrized continuum of integral transform pairs in $B_k(C)$. Each value of the group parameters yields an integral transform pair. The Plancherel formula for this transform is essentially a statement of the unitarity of the representation

$$(g_{1u},g_{2u}) = (T_u f_1, T_u f_2) = (f_1,f_2)$$

and can be written as

$$\int_{\operatorname{Im} z>0}\overline{g_{1u}(z)}g_{2u}(z)d\lambda(z) = \int_{\operatorname{Im}\xi>0}\overline{f_1(\xi)}f_2(\xi)d\lambda(\xi). \tag{3.53}$$

To test the correctness of our result we check the behavior of the transform at the identity, i.e., at $\alpha = 1, \beta = 0$. The rhs of Eq. (3.51) then becomes

$$\int_{\operatorname{Im}\xi>0} I_{2k-1}(z\bar{\xi})f(\xi)d\lambda(\xi).$$

By Eq. (3.27) the above integral is $f(z)$, which is the desired result.

ACKNOWLEDGMENT

The author wishes to thank Professor G. P. Sastry for many helpful discussions.

[1] V. Bargmann, Commun. Pure Appl. Math. **14**, 187 (1961); **20**, 1 (1967).
[2] D. Basu and T. K. Kar, J. Math. Phys. **29**, 1854 (1988).
[3] V. Bargmann, Rev. Mod. Phys. **34**, 829 (1962).
[4] J. Schwinger, in *Quantum Theory of Angular Momentum*, edited by L. C. Biedenharn and H. van Dam (Academic, New York, 1965).
[5] M. Moshinsky and C. Quesne, J. Math. Phys. **12**, 1772, 1780 (1971); M. Moshinsky, T. H. Seligman, and K. B. Wolf, *ibid.* **13**, 901 (1972); M. Moshinsky, SIAM J. Appl. Math. **25**, 193 (1973); P. Kramer, M. Moshinsky, and T. H. Seligman, in *Group Theory and Its Applications*, edited by E. M. Loebl (Academic, New York, 1975).
[6] K. B. Wolf, J. Math. Phys. **15**, 1295, 2102 (1974); **17**, 601 (1976); *Integral Transforms in Science and Engineering* (Plenum, New York, 1979); C. P. Boyer and K. B. Wolf, J. Math. Phys. **16**, 1493, 2215 (1975); K. B. Wolf, *ibid.* **21**, 680 (1980); D. Basu and K. B. Wolf, *ibid.* **23**, 189 (1982); **24**, 478 (1983).
[7] N. Mukunda and B. Radhakrishnan, J. Math. Phys. **14**, 254 (1973); **15**, 1320, 1332, 1643, 1656 (1974); D. Basu, *ibid.* **19**, 1667 (1978); D. Basu and D. Mitra, *ibid.* **21**, 636 (1980); **22**, 946 (1981).
[8] W. J. Holman and L. C. Biedenharn, Ann. Phys. (NY) **39**, 1 (1966); Y. Nambu, Phys. Rev. **160**, 1171 (1967); A. O. Barut and A. Bohm, J. Math. Phys. **11**, 2938 (1970).
[9] I. M. Gel'fand, M. I. Graev, and N. Ya Vilenkin, *Generalized Functions* (Academic, New York, 1966), Vol. 5.
[10] V. Bargmann, Ann. Math. **48**, 568 (1947).
[11] A. O. Barut and R. Raczka, *Theory of Group Representations and Applications* (PWN Polish Scientific, Warsaw, 1980), p. 358.
[12] R. M. Wilcox, J. Math. Phys. **8**, 962 (1967).
[13] S. Sternberg and J. A. Wolf, Trans. Am. Math. Soc. **238**, 1 (1978).
[14] See Moshinsky and Quesne in Ref. 5.
[15] *Higher Transcendental Functions*, Bateman Manuscript Project, edited by A. Erdelyi (McGraw-Hill, New York, 1953), Vol. II, p. 25.
[16] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge U.P., Cambridge, 1969), Chap. XVII.
[17] See Ref. 15, p. 51.

Generating relations for reducing matrices. IV. Subduced representations

J. N. Kotzev and M. K. Peev
Department of Solid State Physics, University of Sofia, A. Ivanov Blvd., 5, BG-1126 Sofia, Bulgaria
M. I. Aroyo[a)]
Institut für Kristallographie, Universität Karlsruhe, Kaiserstrasse 12, D-75 Karlsruhe, West Germany
R. Dirl and P. Kasperkovitz
Institut für Theoretische Physik, Technische Universitat, Karlsplatz 13, A-1040 Wien, Austria

(Received 7 January 1988; accepted for publication 31 August 1988)

A subduced representation $D_A \downarrow G_B$ is a (co)representation of the group G_B obtained by restricting the (co)representation D_A of a supergroup G_A to the elements of G_B. A systematic and efficient method for calculating the matrices needed for the decomposition of $D_A \downarrow G_B$ into irreducible constituents is discussed in this paper where the auxiliary group approach developed previously [J. Math. Phys. **27**, 37, 2236 (1986); **28**, 1947 (1987)] is adopted to the present problem.

I. INTRODUCTION

In this paper we continue our discussion of properties of matrices that reduce a given representation of a group or a well-defined set of such representations into irreducible constituents. In the previous papers of this series[1–3] (hereafter denoted by I, II, and III) we introduced auxiliary groups of transformations of representations (reps) or corepresentations (coreps) and established corresponding transformations of the reducing matrices. The auxiliary groups were then used to reduce and systematize the calculation of these matrices.

In this paper we address ourselves to the following problem: let D_A^k be an irreducible (co)representation [(co)irrep] of the group G_A, which is assumed to be finite or compact continuous, and $D_A^{k\downarrow}$ be the (co)rep of a (topological) subgroup G_B obtained by restricting D_A^k to G_B,

$$D_A^{k\downarrow} = D_A^k \downarrow G_B = \{D_A^k(g), \text{ for all } g \in G_B\} \tag{1}$$

This subduced (co)rep is, in general, a reducible one and can be transformed into a direct sum of (co)irreps of G_B by a suitable unitary transformation,

$$D_A^{k\downarrow}(g) S^{k^{(g)}} = S^k \Big[\bigoplus_s E(k\downarrow|s) \otimes D_B^s(g) \Big]. \tag{2}$$

In this equation $D_B^s(g)$ are the matrices of (co)irreps of G_B, $E(d)$ is the unit matrix of dimension d, and $(k\downarrow|s)$ is the number of times D_B^s is contained in $D_A^{k\downarrow}$ (subduction or branching multiplicity). As usual[1–3] the upper index (g) has the following meaning: $M^{(g)} = M$ for matrices transforming ordinary reps; $M^{(g)} = M$ for $g \in H$ and $M^{(g)} = M^*$ for $g \in G(H) \setminus H$ in the case of coreps of magnetic groups $G(H)$. The matrix S^k is the so-called subducing matrix. We are especially interested in an effective calculation of this matrix and other subducing matrices related to it.

Subducing matrices occur in various physical problems. For instance, in crystal field theory the effect of the environment on an electron bounded to a fixed ion is represented by a perturbation that breaks the spherical symmetry of the free atom to a point group symmetry. The corresponding subducing matrices are then used to pass from the wave functions of the free atom to a symmetry adapted basis that simplifies the eigenvalue problem of the full Hamiltonian (see Refs. 4 and 5 and the references therein). The subducing matrices are also needed when considering the morphic effects on infrared and Raman spectra[6–7] in the phase transition theory,[8] and for constructing the most general spin Hamiltonians in the paramagnetic resonance theory.[9] Apart from this, subducing matrices play an essential role in the Racah lemma which will be discussed in a future publication.

In order to apply our approach[1–3] to the problem at hand we have to discuss the relations between the auxiliary groups of the group G_A and its subgroup G_B. This is done in Sec. II, while transformation properties of subducing matrices are discussed in Sec. III. The emerging scheme is compared with the results of our previous papers in Sec. IV. Finally an example treated by the method proposed here is given in Sec. V.

II. RELATIONS BETWEEN THE AUXILIARY GROUPS OF THE GROUP G_A AND ITS SUBGROUP G_B

The goal of this section is to make use of the information contained in both auxiliary groups[1,2]

$$\mathbf{Q}_A = \mathrm{ASS}(G_A) \circledS (\mathrm{AUT}(G_A) \times \mathrm{CON}(G_A)), \tag{3}$$

and

$$\mathbf{Q}_B = \mathrm{ASS}(G_B) \circledS (\mathrm{AUT}(G_B) \times \mathrm{CON}(G_B)). \tag{4}$$

These groups consist of the following transformations of the matrices forming the (co)reps of the group G_A and G_B, respectively.

(i) Associations: $\mathrm{ASS}(G) = \{a_j\}$, where $(a_j D)(g) = D^j(g) \times D(g)$ and D^j is a one-dimensional (co)rep of G.

(ii) Automorphisms: $\mathrm{AUT}(G) = \{b\}$, where $(bD) \times (g) = D(\beta^{-1}(g))$, $\beta \in \mathrm{Aut}\, G$ (group of all automorphisms of G).

(iii) Complex conjugation: $\mathrm{CON}(G) = \{e,c\}$, $(cD)(g) = D^*(g)$.

In general, there does not exist a simple relation between these two groups. However, we can restrict the transformations of $\mathbf{Q}_A$ such that each of them may be considered as a

[a)] Permanent address: Department of Solid State Physics, University of Sofia, A. Ivanov Blvd., 5, BG-1126 Sofia, Bulgaria.

0022-2488/89/010009-09$02.50

transformation of a (co)rep of G_B subduced from a (co)rep of G_A. More precisely, we define a subgroup $Q_A \subset \mathbf{Q}_A$ and a homomorphism ϕ from Q_A onto a subgroup $Q_B \subset \mathbf{Q}_B$, such that the auxiliary groups Q_A and Q_B can be used to establish generating and symmetry relations for subducing matrices.

To define Q_A we first consider the case of ordinary reps, turning to coreps afterwards. Each element $q_A \in \mathbf{Q}_A$ transforms a given rep D_A of G_A into the rep $q_A D_A$ which, if restricted to G_B, is also a rep of G_B. However, the essential point is that the two reps $D_A^{\downarrow}$ and $(q_A D_A)^{\downarrow}$ are not linked by a transformation $q_B \in \mathbf{Q}_B$ in general. This problem arises especially for automorphisms contained in the subgroup $\mathrm{AUT}(G_A) \subset \mathbf{Q}_A$. There are no problems with complex conjugation and with the associations contained in $\mathbf{Q}_A$. For if D_A^j is a one-dimensional irrep of G_A so is the subduced rep

$$D_A^{j\downarrow} = D_A^j \downarrow G_B = \{D_A^j(g),\ g \in G_B\}. \tag{5}$$

Moreover if

$$D_A^{j_1}(g) \otimes D_A^{j_2}(g) = D_A^{j_3}(g),$$

for all $g \in G_A$, then this obviously holds true also if G_A is restricted to G_B. In order to get a similar simple relationship between the automorphisms $\beta_A \in \mathrm{Aut}\ G_A$ and $\beta_B \in \mathrm{Aut}\ G_B$ we have to restrict the automorphisms of G_A to those that leave G_B invariant:

$$\mathrm{Aut}\ G_{A(B)} = \{\beta_A | \beta_A(G_B) = G_B;\ \beta_A \in \mathrm{Aut}\ G_A\}. \tag{6}$$

Note that this subset forms a subgroup of Aut G_A. This leads to the definition of a subgroup of $\mathbf{Q}_A$, namely,

$$Q_A = \mathrm{ASS}(G_A) \circledS (\mathrm{AUT}(G_A)_{(B)} \times \mathrm{CON}(G_A)). \tag{7}$$

In this definition $\mathrm{AUT}(G_A)_{(B)}$ consist of all transformations that correspond to the automorphisms $\beta_A \in \mathrm{Aut}\ G_{A(B)}$.

Next we transfer the action of the auxiliary group Q_A from the reps D_A of G_A to the subduced reps $D_A^{\downarrow}$ of G_B by means of the following definition:

$$(\phi(q_A) D_A^{\downarrow})(g) = (q_A D_A)^{\downarrow}(g), \quad q_A \in Q_A, \quad g \in G_B. \tag{8}$$

The meaning of the rhs is completely determined by Eqs. (7) and (1) with D_A substituted for the (co)irrep D_A^k. Therefore, the lhs defines transformations $\phi(q_A)$ of the subduced rep $D_A^{\downarrow}$. Considered as a set of matrices, the subduced reps may be only a proper subset of the reps of G_B, so that $\phi(q_A)$ is not an element of the auxiliary group $\mathbf{Q}_B$ whose transformations are defined for all reps of G_B (see I, Sec. II B). But in the following we consider subduced reps only so that we can ignore this fact and write

$$\phi(q_A) = q_B \in \mathbf{Q}_B.$$

It follows from the definition (8) that

$$\phi(q_A)\phi(q'_A) D_A^{\downarrow} = (q_A q'_A D_A)^{\downarrow}. \tag{9}$$

The mapping ϕ is therefore a homomorphism from Q_A onto a subgroup of $\mathbf{Q}_B$,

$$Q_B = \phi(Q_A) \subset \mathbf{Q}_B. \tag{10}$$

The kernel of this homomorphism consists of all transformations $q_A \in Q_A$ that leave all subduced reps $D_A^{\downarrow}$ invariant,

$$q_A \in \mathrm{Ker}\ \phi, \quad \text{if } \phi(q_A) D_A^{\downarrow} = D_A^{\downarrow}. \tag{11}$$

This kernel contains all associations of $\mathbf{Q}_A$ corresponding to irreps D_A^j that subduce the identical irrep of G_B. Moreover, it also contains the transformations $B_A \in \mathrm{AUT}(G_A)$ corresponding to automorphisms $\beta_A \in \mathrm{Aut}\ G_A$ that leave each element g of G_B unchanged $[\beta_A(g) = g$, for all $g \in G_B]$.

The complex conjugation c_A belongs to Ker ϕ if, and only if, every rep $D_A \downarrow G_B$ obtained from some rep D_A of G_A consists of real matrices only. This cannot happen for coreps of magnetic groups $G_A(H_A)$ if they are restricted to magnetic subgroups $(G_B \not\subset H_A)$. For even if $D_A \downarrow G_B$ is real it is always possible to pass over from the corep D_A to an equivalent corep D'_A differing only in a (nonreal) phase factor common to all antiunitary elements. If $G_B \subset H_A$ the situation is the same as for ordinary reps where it can be shown that the reality of all the reps $D_A \downarrow G_B$ implies $D_A(g) = E$ for all $g \in G_B$, i.e., $G_B = \{e_A\} \cong C_1$. Since subduction is trivial in that case we always assume in the following that the complex conjugation does not belong to the kernel of the homomorphism.

If G_B is a characteristic subgroup of G_A, i.e., a group left invariant by all automorphisms of G_A, then obviously $Q_A = \mathbf{Q}_A$. In applications this situation is not so rare as one might think at first glance. For instance, all crystallographic groups are solvable[10] so that one finds in these examples not only one characteristic subgroup, but chains of such subgroups.

For magnetic groups $G(H) = H \cup H a_0$ one has to take into account the peculiarities of coreps and the special conventions for corresponding auxiliary groups.[2,3] In Paper III we have chosen one-dimensional coreps satisfying the convention $D^j(a_0) = 1$, for one fixed element $a_0 \in G(H) \setminus H$. Now restricting $G_A(H_A)$ to $G_B(H_B)$ we may find that $a_{0,A} \notin G_B(H_B)$. That should not worry us if $G_B \subset H_A$. In the case where $G_B \not\subset H_A$ we know that the element $a_{0,B}$ used to fix the one-dimensional coirreps of D_B^j and all coirreps of type III (cf. Sec. II A of Paper III) is also an element of $G_A(H_A) \setminus H_A$. Therefore we can redefine the one-dimensional coirreps D_A^j and also the coirreps of type III in such a way that $a_{0,B}$ is now considered as $a_{0,A}$. The remaining coirreps of $G_A(H_A)$ need not be redefined. This new convention guarantees that all coirreps of $G_A(H_A)$ that subduce only one single coirrep of $G_B(H_B)$ lead to coirreps that are of the desired form. Apart from these modifications one has to define Q_A in such a way that it contains only those automorphisms of $G_A(H_A)$ (see Ref. 2) that leave both $G_B(H_B)$ and H_B invariant.

In the following we need several subgroups of the auxiliary groups Q_A and Q_B and the corresponding coset decompositions. One class of subgroups are the groups Q_A^k defined for each (co)irrep D_A^k by

$$Q_A^k = \{q_A | q_A D_A^k \sim D_A^k,\ q_A \in Q_A\}. \tag{12}$$

The coset decomposition of Q_A with respect to Q_A^k is

$$\begin{aligned} &Q_A = q_{A,l_1}^{(k)} Q_A^k \cup q_{A,l_2}^{(k)} Q_A^k \cup \cdots, \\ &R_A^k = \{q_{A,l_1}^{(k)}, q_{A,l_2}^{(k)}, \cdots\} = \text{fixed set of coset representatives}, \\ &q_{A,l_1}^k = q_{A,0} = \text{identity transformation}. \end{aligned} \tag{13}$$

The (co)irreps of G_A may be then decomposed into disjoint Q_A classes,

$$[k]_A = \{l | l = q_A k,\ q_A \in Q_A\}. \tag{14}$$

As in our previous papers $l = qk$ is a shorthand notation for $D^l \sim qD^k$. Likewise we assume that for each class $[k]_A$ a representative D^k_A has been chosen and that a "standard set" of inequivalent (co)irreps contained in $[k]_A$ is fixed by the convention

$$D^l_A = q^{(k)}_{A,l} D^k_A, \quad q^{(k)}_{A,l} \in R^k_A . \tag{15}$$

The action of the transformations $q_A \in Q_A$ on these coirreps is given by the following equations (cf. Sec. II A of Paper III):

$$q_A D^k_A(g) = U^k_A(q_A)^\dagger D^k_A(g) U^k_A(q_A)^{(g)}, \quad q_A \in Q^k_A , \tag{16}$$

$$q_A D^l_A(g) = U^{l',l}_A(q_A)^\dagger D^{l'}_A(g) U^{l',l}_A(q_A)^{(g)}, \quad q_A \in Q_A , \tag{17}$$

$$U^{l',l}_A(q_A) = q^{(k)}_{A,l'} U^k_A(q'_A) ,$$
$$q'_A = [q^{(k)}_{A,l'}]^{-1} q_A q^{(k)}_{A,l} , \tag{18}$$
$$q'_A \in Q^k_A, \quad q^{(k)}_{A,l}, q^{(k)}_{A,l'} \in R^k_A .$$

The symbol qM used for all matrices M that are not (co)rep matrices is defined by

$$qM = \begin{cases} M^*, & \text{if } q \text{ contains the complex conjugation,} \\ M, & \text{otherwise} . \end{cases} \tag{19}$$

It follows from these equations that all the transformations relating standard (co)irreps may be generated from matrices $U^k_A(q_A)$, q_A ranging over a set of generators of Q^k_A. But it should be kept in mind that the generating matrices $U^k_A(q_A)$ are not uniquely defined by Eq. (16) (see Papers I and III) so that some convention is needed to fix them.

In the same way the subgroups Q^s_B and the corresponding matrices $U^s_B(q_B)$ are defined for a standard set of (co)irreps D^s_B of G_B. In complete analogy with Eqs. (16)–(18) one can construct the matrices $U^{t',t}_B(q_B)$ for all elements $q_B \in Q_B$ which will be needed later on.

III. TRANSFORMATION PROPERTIES OF SUBDUCING MATRICES

In this section we discuss the properties of matrices S^l subducing standard (co)irreps D^t_B of G_B from standard (co)irreps D^l_A of G_A,

$$D^{l\downarrow}_A(g) S^{l(g)} = S^l \Big[\oplus_t E(l\downarrow|t) \otimes D^t_B(g) \Big], \quad g \in G_B . \tag{20}$$

As it is evident from Eq. (20) the rows of S^l are labeled by the row index of D^l_A, $\lambda = 1,...,n_l = \dim D^l_A$. The columns of S^l are labeled by (t,m,ϑ):

$$D^t_B = \text{(co)irrep of } G_B \text{ with row index}$$
$$\vartheta = 1,...,n_t = \dim D^t_B \tag{21}$$
$$m = 1,...,(l\downarrow|t) ,$$
$$(l\downarrow|t) = \text{multiplicity of } D^t_B \text{ in } D^{l\downarrow}_A . \tag{22}$$

In the case of coirreps of Wigner canonical form a slightly modified labeling is more convenient [see Eq. (45) of Paper III].

Equation (20) fixes the matrices S^l only up to a left factor $M^{l\downarrow\dagger}_A$, where $M^{l\downarrow}_A$ is a unitary matrix of the commuting algebra of $D^{l\downarrow}_A$, and a unitary right factor $M^{l\downarrow}_B$ that belongs to the commuting algebra of the direct sum $[\oplus_t E(l\downarrow|t) \otimes D^t_B]$. It should be noted that $\{M^l_A\}$, the commuting algebra of D^l_A, is a subalgebra of $\{M^{l\downarrow}_A\}$, the commuting algebra of $D^{l\downarrow}_A$. As has been pointed out in Secs. II A and II C of Paper III, the whole freedom in S^l is already contained in one set of the factors, i.e., either in the unitary matrices $M^{l\downarrow}_A$ or $M^{l\downarrow}_B$. This is the consequence of the fact that for each matrix $M^{l\downarrow}_A$ there exists a matrix $M^{l\downarrow}_B = S^{l\dagger} M^{l\downarrow}_A S^l$ and vice versa so that the two commuting algebras are isomorphic. Therefore in the following we shall express the freedom in S^l only by appropriate right factors $M^{l\downarrow}_B$:

$$\Big[\oplus_t E(l\downarrow|t) \otimes D^t_B(g) \Big] M^{l\downarrow(g)}_B = M^{l\downarrow}_B \Big[\oplus_t E(l\downarrow|t) \otimes D^t_B(g) \Big], \quad \text{for all } g \in G_B , \tag{23}$$

$$M^{l\downarrow}_B = \oplus_t M^{l\downarrow,t}_B . \tag{24}$$

The matrices $M^{l\downarrow,t}_B$ depend on the multiplicity $(l\downarrow|t)$ and the type of the (co)irreps D^t_B [cf. Sec. III of Paper II or Eq. (49) of Paper III].

The subducing matrix S^l, or parts of it, may be related to other matrices of this kind by applying a number of transformations on both sides of its defining equation (20). These transformations are[1–3] (i) multiplication with one-dimensional (co)irreps D^j_B, (ii) substitution of g by $\beta^{-1}(g)$, and (iii) complex conjugation. Using in (20) the transformation properties of the (co)reps and the homomorphism ϕ [Eq. (8)] we finally obtain

$$D^{l'\downarrow}_A(g)\{U^{l',l}_A(q_A)(q_B S^l) Z^{l\downarrow}_B(q_B)^\dagger\}^{(g)} = \{U^{l',l}_A(q_A)(q_B S^l) Z^{l\downarrow}_B(q_B)^\dagger\} \Big[\oplus_t E(l\downarrow|t) \otimes D^{t'}_B(g) \Big] , \tag{25}$$

where

$$Z^{l\downarrow}_B(q_B) = \oplus_t E(l\downarrow|t) \otimes U^{t',t}_B(q_B) \tag{26}$$

and

$$q_B = \phi(q_A) . \tag{27}$$

Moreover in Eq. (25)

$$D^{l'\downarrow}_A \sim q_A D^{l\downarrow}_A, \quad D^{t'}_B \sim q_B D^t_B . \tag{28}$$

The matrix in curly brackets decomposes $D^{l'\downarrow}_A$ into the direct sum $\oplus_t E(l\downarrow|t) \otimes D^{t'}_B$ and is therefore a subducing matrix of the same kind as $S^{l'}$. Equation (25) also states implicitly that

$$(l'\downarrow|t') = (q_A l\downarrow|q_B t) = (l\downarrow|t) . \tag{29}$$

The subducing matrix in curly brackets has to be multiplied from the right with a permutational matrix $P(q_A)$ if we want all subducing matrices to decompose the (co)reps $D^{l\downarrow}_A$ into direct sums where the constituents D^t_B appear in a given lexicographical order,

$$P_{tm\vartheta,t'm'\vartheta'}(q_A) = \delta_{t,q_A t'} \delta_{m,m'} \delta_{\vartheta,\vartheta'} . \tag{30}$$

Because of Eq. (25) every matrix $S^{l'}$ that decomposes $D^{l'\downarrow}_A$ into a direct sum in lexicographical order is related to the subducing matrix S^l by an equation of the form

$$S^{l'} M^{l'\downarrow}_B(q_A) = U^{l',l}_A(q_A)(q_B S^l) Z^{l\downarrow}_B(q_B)^\dagger P(q_A) . \tag{31}$$

Assuming a sufficient number of columns of the matrix S^l to be known, we can use Eq. (31) to generate the remaining columns and other reducing matrices $S^{l'}$ by specifying appropriate transformations q_A and the corresponding matrices $M_B^{l'l}(q_A)$. We do this in such a way that each new column or matrix is defined by one equation only so that no inconsistencies can arise in the definition of these quantities.

In the generating relations of the first kind the matrices S^l, $l\in[k]_A$, are related to the matrix S^k by

$$S^l = (q_B S^k) Z_B^{k\downarrow}(q_B)^\dagger P(q_A) . \tag{32}$$

In this equation $q_B = \phi(q_A)$, $q_A = q_{A,l}^{(k)} \in R_A^k$, and the matrices $Z_B^{k\downarrow}$ and P are given by Eqs. (26) and (30), respectively. Comparison of (32) and (31) shows that here the matrix $M_B^{l\downarrow}(q_A)$ has been chosen as the unit matrix.

To reduce the calculation of the matrix S^k to that of some of its columns, we have to introduce additional subgroups of Q_A and Q_B that are shown in the following diagram:

$$\begin{array}{llllll} G_A\colon & \mathbf{Q}_A \supset & Q_A \supset & Q_A^k \supset & Q_A^{k,t} & \\ & & \downarrow\phi & \downarrow\phi & \downarrow\phi & \\ G_B\colon & \mathbf{Q}_B \supset & Q_B \supset & Q_B^{k\downarrow} \supset & Q_B^{k\downarrow,t} . & \end{array} \tag{33}$$

The groups appearing in (33) can be defined by means of the subgroups Q_A^k, the homomorphism ϕ, already introduced in Sec. II, and some more conventions. First of all the group $Q_B^{k\downarrow}$ is defined by

$$Q_B^{k\downarrow} = \phi(Q_A^k) . \tag{34}$$

The transformations of this group define a partition of the standard (co)irreps of G_B into disjoint $Q_B^{k\downarrow}$ classes,

$$[t]_B^{k\downarrow} = \{t' | t' = q_B t,\ q_B \in Q_B^{k\downarrow}\} . \tag{35}$$

It follows from the definition of $Q_B^{k\downarrow}$ and Q_A^k that $D_A^{k\downarrow}$ either contains all members of such a class with the same multiplicity or none of them. We assume that representatives D_B^t of the classes (35) have been chosen and used to define the groups $Q_B^{k\downarrow,t}$ and $Q_A^{k,t}$,

$$Q_B^{k\downarrow,t} = \{q_B | q_B D_B^t \sim D_B^t,\ q_B \in Q_B^{k\downarrow}\} , \tag{36}$$

$$Q_A^{k,t} = \phi^{-1}(Q_B^{k\downarrow,t}) . \tag{37}$$

Moreover we have to specify coset representatives for the decomposition of $Q_B^{k\downarrow}$ with respect to those subgroups $Q_B^{k\downarrow,t}$ for which $(k\downarrow|t)\neq 0$,

$$\begin{aligned} Q_B^{k\downarrow} &= q_{B,t_1}^{(k\downarrow,t)} Q_B^{k\downarrow,t} \cup q_{B,t_2}^{(k\downarrow,t)} Q_B^{k\downarrow,t} \cup \cdots , \\ R_B^{k\downarrow,t} &= \{q_{B,t_1}^{(k\downarrow,t)}, q_{B,t_2}^{(k\downarrow,t)}, \cdots\} \\ &= \text{fixed set of coset representatives,} \\ q_{B,t_1}^{(k\downarrow,t)} &= q_{B,0} = \text{identity transformation} . \end{aligned} \tag{38}$$

If the transformations q_A are limited to the group Q_A^k, Eq. (31) becomes

$$S^k M_B^{k\downarrow}(q_A) = U^k(q_A)(q_B S^k) Z_B^{k\downarrow}(q_B)^\dagger P(q_A) . \tag{39}$$

We now split the matrices appearing on both sides of this equation into rectangular blocks,

$$\begin{aligned} S_{t'}^k &= \text{rectangular matrix consisting of all columns} \\ &\quad \text{of } S^k \text{ with fixed index } t' \\ &\quad [m = 1,\ldots,(k\downarrow|t'),\ \vartheta = 1,\ldots,n_{t'}] . \end{aligned} \tag{40}$$

Because of (39) these matrices are related by

$$S_{t'}^k = U_A^k(q_A)(q_B S_t^k) Z_{B,t}^{k\downarrow}(q_B)^\dagger , \tag{41}$$

where

$$Z_{B,t}^{k\downarrow}(q_B) = E(k\downarrow|t) \otimes U_B^{t',t}(q_B) . \tag{42}$$

In Eq. (41), $t'\in[t]_B^{k\downarrow}$, the irrep D_B^t is the representative of this class, and $q_B = q_{B,t'}^{(k\downarrow,t)} \in R_B^{k\downarrow,t}$. The transformation q_A can be any inverse image of q_B, but once this element of Q_A has been fixed the block $S_{t'}^k$ is uniquely determined. This is due to the fact that the constituent $M_B^{k\downarrow,t'}(q_A)$ of the matrix $M_B^{k\downarrow}(q_A)$, which in general appears on the lhs of Eq. (41), is chosen as unit matrix. We call these relations linking rectangular blocks of one matrix S^k generating relations of the second kind.

To resolve the multiplicity problem in group theoretical terms, at least partially, we finally restrict the operations q_A to the subgroup $Q_A^{k,t}$ defined in Eq. (37),

$$S_t^k M_B^{k\downarrow,t}(q_A) = U_A^k(q_A)(q_B S_t^k)[E(k\downarrow|t) \otimes U_B^{t,t}(q_B)]^\dagger . \tag{43}$$

If we split the already calculated block S_t^k into $(k\downarrow|t)$ subblocks S_{tm}^k,

$$\begin{aligned} S_{tm}^k &= \text{rectangular matrix consisting of all} \\ &\quad \text{columns of } S_A^k \text{ with fixed indices } tm , \\ &\quad [\vartheta = 1,\ldots,n_t] , \end{aligned} \tag{44}$$

then the rhs of (43) may be considered as an action of an operator $T(q_A)$ on the "vector" S_{tm}^k,

$$T(q_A) S_{tm}^k = U_A^k(q_A)(q_B S_{tm}^k) U_B^{t,t}(q_B)^\dagger . \tag{45}$$

As discussed in our previous papers (see especially Sec. II C of Paper III) the operators $T(q_A)$ turn out to be either linear or antilinear if the matrices $U_A^k(q_A)$ and $U_B^{t,t}(q_B)$ are properly chosen. In that case the matrices $M_B^{k\downarrow,t}(q_A)$ appearing on the lhs of Eq. (43) form (co)reps of the operator group $\tilde{Q}_A = \{T(q_A)\}$ generated by $T(q_A)$, q_A ranging over a set of generators of $Q_A^{k,t}$. By a linear transformation these (co)reps may always be brought into block diagonal form. The new subblocks $\bar{S}_{tm}^k$ obtained this way can then be labeled by the labels of the occurring (co)irreps of $\tilde{Q}_A$; if some (co)irrep occurs more than once additional labels, not related to the auxiliary group, are needed.

IV. COMPARISON OF THE PRESENT APPROACH WITH THE RESULTS OF THE PREVIOUS PAPERS

In Papers I and II we discussed how to find generating relations and how to reduce the multiplicity problem for a fixed reducible (co)rep of a given group. If we would choose this group to be G_B and ignore the origin of the (co)rep $D_A^{k\downarrow}$ we could simply proceed as described there. In such an approach the transformations relating different subblocks of the reducing matrix follow then from a group

$$\mathbf{Q}_B^{k\downarrow} = \{q_B | q_B D_A^{k\downarrow} \sim D_A^{k\downarrow};\ q_B \in \mathbf{Q}_B\} . \tag{46}$$

In general, the group $\mathbf{Q}_B^{k\downarrow}$ is an extension of the group $Q_B^{k\downarrow}$ [Eq. (34)] used in the previous section to establish generating relations of the second kind and for a partial solution of the multiplicity problem. Therefore if we restrict $\mathbf{Q}_B^{k_1}$ to a subgroup

$$\mathbf{Q}_B^{k_1,t} = \{q_B | q_B D^t_B \sim D^t_B,\ q_B \in Q_B^{k_1}\}, \qquad (47)$$

this group is also an extension of $Q_B^{k_1,t}$ and should therefore be more suited to solving the multiplicity problem. As regards generating relations (of the second kind), the approach of Papers I and II should also be at least as effective as the one discussed in the previous section. For the number of these relations depends on the index of $\mathbf{Q}_B^{k_1,t}$ in $\mathbf{Q}_B^{k_1}$ or the index of $Q_B^{k_1,t}$ in $Q_B^{k_1}$, respectively, and the former is equal or greater than the latter. This follows from the fact that in the transition from the first pair of groups to the second one all transformations q_B are eliminated that are not homomorphic images of transformations $q_A \in Q_A$. The two indices coincide if G_B is a normal subgroup of G_A. For according to Clifford's theorem[11] the coset representatives of $\mathbf{Q}_B^{k_1,t}$ with respect to $\mathbf{Q}_B^{k_1}$ can be chosen as inner automorphisms of G_A that are outer automorphisms of G_B. These transformations are also elements of $Q_B^{k_1}$ as follows from the definition of the group Q_A and of the homomorphism ϕ; this implies the coincidence of the two indices.

Up to now the already existng scheme seems to be at least as good as that proposed in the present paper. However, the present scheme also contains generating relations of the first kind, relating different subducing matrices, which have not been considered in I and II. We now extend the approach of Papers I and II to also include this kind of generating relations. The number of these relations is given by the number of (co)irreps D^l_A occurring in the definition of the set

$$\mathbf{R}_{B(A)}^{k_1} = \{q_B | q_B D_A^{k_1} \sim D_A^{l_1},$$
$$\text{for some (co)irrep } D_A^l \text{ of } G_A,\ q_B \in \mathbf{R}_B^{k_1}\}. \qquad (48)$$

Here the set $\mathbf{R}_B^{k_1}$ is a fixed set of coset representatives in the decomposition of $\mathbf{Q}_B$ with respect to $\mathbf{Q}_B^{k_1}$. Note that for a given D_A^k and q_B there may be several (co)irreps D_A^l such that $D_A^{l_1} \sim q_B D_A^{k_1}$. For each pair (l,k) we then have to find a unitary matrix $U_B^{l_1,k_1}(q_B)$ that transforms $q_B D_A^{k_1}$ into $D_A^{l_1}$,

$$q_B D_A^{k_1}(g) = U_B^{l_1,k_1}(q_B)^\dagger D_A^{l_1}(g) U_B^{l_1,k_1}(q_B)^{(g)}. \qquad (49)$$

To formulate the generating relations we also need the matrices $U_B^{t',t}(q_B)$ occurring on the rhs of

$$q_B D^t_B(g) = U^{t',t}(q_B)^\dagger D^{t'}_B(g) U_B^{t',t}(q_B)^{(g)}, \qquad (50)$$

where q_B ranges over the set $\mathbf{R}_{B(A)}^{k_1}$. All the matrices $U_B^{t',t}(q_B)$ can be easily constructed in the same way as the matrices $U_A^{l',l}(q_A)$ were constructed in Sec. II [cf. (16)–(18)]. The generating relations of the first kind then read

$$S^l = U_B^{l_1,k_1}(q_B)(q_B S_A^k) Z_B^{k_1}(q_B)^\dagger P(q_B), \qquad (51)$$

where the matrices $Z_B^{k_1}(q_B)$ and $P(q_B)$ are defined by Eqs. (26) and (30), with q_A replaced by q_B in (30) and q_B now ranging over the set $\mathbf{R}_{B(A)}^{k_1}$. However, generating relations of the form (51) are not very useful as long as there does not exist a systematic method to calculate the matrices $U_B^{l_1,k_1}(q_B)$. In fact, calculating this matrix may be as difficult as a straightforward calculation of the subducing matrix S^l itself.

The second point we want to emphasize in this section is that the approach of the present paper may be viewed as generalization of the results of Paper III. This can be seen by choosing the supergroup G_A as a direct product,

$$G_A = G \times G \times \cdots \times G, \qquad (52)$$

and G_B as its diagonal subgroup,

$$\text{Diag}\ G_A \cong G, \qquad (53)$$

with elements $(g,g,...,g)$, $g\in G$. In this case $Q_B = \mathbf{Q}_B \cong Q$, where Q is one of the two auxiliary groups used in Paper III. The second group, denoted by $\mathbf{Q}$ there, is isomorphic to a subgroup of the group Q_A for the present choice of G_A. The elements of Q_A, which have no counterpart in $\mathbf{Q}$, correspond exactly to those automorphisms that preserve the diagonal elements but cannot be expressed as products of the form $(b,b,...,b)$ times a permutation. If such automorphisms are found, it is possible to extend the calculation of Clebsch–Gordan coefficients as described in Paper III by following the scheme proposed here.

V. SUBDUCING MATRICES FOR THE SPACE GROUP CHAIN *Pm3m*⊃*P23*

The following example is used to illustrate the present approach. We adopt the same notation and conventions as in III, but for convenience we recall some of the definitions to make the paper self-contained.

Here we consider the space groups

$$G_A = O_h^1 = Pm3m, \quad G_B = T^1 = P23, \qquad (54)$$

which are symmorphic and whose translational subgroups coincide. For details concerning these space groups the reader is referred to Refs. 4 and 11.

Next we state the general form of "standard" irreps for symmorphic space groups. By definition these standard irreps are determined by induction[4] out of the one-dimensional irreps of the translation group. Let

$$G = T \circledS P \qquad (55)$$

be an arbitrary symmorphic space group G whose translation group is denoted by T and its point group by P. Then the standard irreps of this group take the following form:

$$D^N_{\underline{R},\underline{R}'}(R|\mathbf{t})$$
$$= \Delta^{\mathbf{q}}(\underline{R}, R\underline{R}')\exp(-i\underline{R}\mathbf{q}\cdot\mathbf{t}) D^n(\underline{R}^{-1}R\underline{R}'), \qquad (56)$$

where we used in part a matrix notation. The symbols $D^n(R')$, $R' \in P(\mathbf{q})$ denote matrix irreps of the little cogroup $P(\mathbf{q})$[4]. In detail our notation has the following meaning:

$N = (\mathbf{q},n)^\dagger G = \mathbf{q}n$, standard G-irrep label,

$q \in \Delta\mathrm{BZ}(g)$,

representation domain of the Brillouin zone $\mathrm{BZ}(G)$,

$n \in \mathscr{A}(\mathbf{q})$, set of $P(\mathbf{q})$irrep labels,

$\underline{R}, \underline{R}' \in \underline{P}(\mathbf{q})$, fixed set of coset representatives (CR's) for the decomposition of P with respect to $P(\mathbf{q})$,

$\mathscr{A}(G) = \{\mathbf{q}n\}$, a set of standard irrep labels. $\qquad$ (57)

Finally, the symbol $\Delta^{\mathbf{q}}$ is defined as

TABLE I. $q_A K$ table for $\mathbf{q} = \mathbf{G}$.

q_A \ Gk	$G1^+$	$G2^+$	$G3^+$	$G4^+$	$G5^+$	$G1^-$	$G2^-$	$G3^-$	$G4^-$	$G5^-$
$G1^\pm$	$G1^\pm$	$G2^\pm$	$G3^\pm$	$G4^\pm$	$G5^\pm$	$G1^\mp$	$G2^\mp$	$G3^\mp$	$G4^\mp$	$G5^\mp$
$G2^\pm$	$G2^\pm$	$G1^\pm$	$G3^\pm$	$G5^\pm$	$G4^\pm$	$G2^\mp$	$G1^\mp$	$G3^\mp$	$G5^\mp$	$G4^\mp$
$R1^\pm$	$R1^\pm$	$R2^\pm$	$R3^\pm$	$R4^\pm$	$R5^\pm$	$R1^\mp$	$R2^\mp$	$R3^\mp$	$R4^\mp$	$R5^\mp$
$R2^\pm$	$R2^\pm$	$R1^\pm$	$R3^\pm$	$R5^\pm$	$R4^\pm$	$R2^\mp$	$R1^\mp$	$R3^\mp$	$R5^\mp$	$R4^\mp$
b_3	$G1^+$	$G2^+$	$G3^+$	$G4^+$	$G5^+$	$G1^-$	$G2^-$	$G3^-$	$G4^-$	$G5^-$
c	$G1^+$	$G2^+$	$G3^+$	$G4^-$	$G5^+$	$G1^-$	$G2^-$	$G3^-$	$G4^-$	$G5^-$

TABLE II. $q_A K$ table for $\mathbf{q} = \mathbf{R}$.

q_A \ Rk	$R1^+$	$R2^+$	$R3^+$	$R4^+$	$R5^+$	$R1^-$	$R2^-$	$R3^-$	$R4^-$	$R5^-$
$G1^\pm$	$R1^\pm$	$R2^\pm$	$R3^\pm$	$R4^\pm$	$R5^\pm$	$R1^\mp$	$R2^\mp$	$R3^\mp$	$R4^\mp$	$R5^\mp$
$G2^\pm$	$R2^\pm$	$R1^\pm$	$R3^\pm$	$R5^\pm$	$R4^\pm$	$R2^\mp$	$R1^\mp$	$R3^\mp$	$R5^\mp$	$R4^\mp$
$R1^\pm$	$G1^\pm$	$G2^\pm$	$G3^\pm$	$G4^\pm$	$G5^\pm$	$G1^\mp$	$G2^\mp$	$G3^\mp$	$G4^\mp$	$G5^\mp$
$R2^\pm$	$G2^\pm$	$G1^\pm$	$G3^\pm$	$G5^\pm$	$G4^\pm$	$G2^\mp$	$G1^\mp$	$G3^\mp$	$G5^\mp$	$G4^\mp$
b_3	$R2^-$	$R1^-$	$R3^-$	$R5^-$	$R4^-$	$R2^+$	$R1^+$	$R3^+$	$R5^+$	$R4^+$
c	$R1^+$	$R2^+$	$R3^+$	$R4^+$	$R5^+$	$R1^-$	$R2^-$	$R3^-$	$R4^-$	$R5^-$

TABLE III. $q_A K$ table for $\mathbf{q} = \mathbf{X}$.

q_A \ Xk	$X1^+$	$X2^+$	$X3^+$	$X4^+$	$X5^+$	$X1^-$	$X2^-$	$X3^-$	$X4^-$	$X5^-$
$G1^\pm$	$X1^\pm$	$X2^\pm$	$X3^\pm$	$X4^\pm$	$X5^\pm$	$X1^\mp$	$X2^\mp$	$X3^\mp$	$X4^\mp$	$X5^\mp$
$G2^\pm$	$X2^\pm$	$X1^\pm$	$X4^\pm$	$X3^\pm$	$X5^\pm$	$X2^\mp$	$X1^\mp$	$X4^\mp$	$X3^\mp$	$X5^\mp$
$R1^\pm$	$M1^\pm$	$M2^\pm$	$M3^\pm$	$M4^\pm$	$M5^\pm$	$M1^\mp$	$M2^\mp$	$M3^\mp$	$M4^\mp$	$M5^\mp$
$R2^\pm$	$M2^\pm$	$M1^\pm$	$M4^\pm$	$M3^\pm$	$M5^\pm$	$M2^\mp$	$M1^\mp$	$M4^\mp$	$M3^\mp$	$M5^\mp$
b_3	$X3^-$	$X4^-$	$X1^-$	$X2^-$	$X5^-$	$X3^+$	$X4^+$	$X1^+$	$X2^+$	$X5^+$
c	$X1^+$	$X2^+$	$X3^+$	$X4^+$	$X5^+$	$X1^-$	$X2^-$	$X3^-$	$X4^-$	$X5^-$

TABLE IV. $q_A K$ table for $\mathbf{q} = \mathbf{M}$.

q_A \ Mk	$M1^+$	$M2^+$	$M3^+$	$M4^+$	$M5^+$	$M1^-$	$M2^-$	$M3^-$	$M4^-$	$M5^-$
$G1^\pm$	$M1^\pm$	$M2^\pm$	$M3^\pm$	$M4^\pm$	$M5^\pm$	$M1^\mp$	$M2^\mp$	$M3^\mp$	$M4^\mp$	$M5^\mp$
$G2^\pm$	$M2^\pm$	$M1^\pm$	$M4^\pm$	$M3^\pm$	$M5^\pm$	$M2^\mp$	$M1^\mp$	$M4^\mp$	$M3^\mp$	$M5^\mp$
$R1^\pm$	$X1^\pm$	$X2^\pm$	$X3^\pm$	$X4^\pm$	$X5^\pm$	$X1^\mp$	$X2^\mp$	$X3^\mp$	$X4^\mp$	$X5^\mp$
$R2^\pm$	$X2^\pm$	$X1^\pm$	$X4^\pm$	$X3^\pm$	$X5^\pm$	$X2^\mp$	$X1^\mp$	$X4^\mp$	$X3^\mp$	$X5^\mp$
b_3	$M4^+$	$M3^+$	$M2^+$	$M1^+$	$M5^+$	$M4^-$	$M3^-$	$M2^-$	$M1^-$	$M5^-$
c	$M1^+$	$M2^+$	$M3^+$	$M4^+$	$M5^+$	$M1^-$	$M2^-$	$M3^-$	$M4^-$	$M5^-$

$$\Delta^{\mathbf{q}}(\underline{R},\underline{R}R') \begin{cases} = 1\,, & \text{if } \underline{R}^{-1}R\underline{R}'\in P(\mathbf{q})\,, \\ = 0\,, & \text{otherwise}\,. \end{cases} \tag{58}$$

Specifying these general formulas and notations to the present example we denote G_A-irrep labels by $K = \mathbf{q}k\in\mathscr{A}(G_A)$ and G_B-irrep labels by $S = \mathbf{q}s\in\mathscr{A}(G_B)$, respectively. The irreps of G_A and G_B are obtained from (56) by inserting the corresponding entities.

As in III we restrict our considerations to the G_A-irreps that are assigned to the high symmetry points G, R, X, M, but discuss only single-valued irreps. We have the following little cogroups:

$$\begin{aligned} &P_A(\mathbf{G}) = P_A(\mathbf{R}) = O_h = m3m\,, \\ &P_A(\mathbf{X}) = D_{4h}^{(y)} = 4_y/mmm\,, \\ &P_A(\mathbf{M}) = D_{4h}^{(z)} = 4_z/mmm\,, \end{aligned} \tag{59}$$

where the corresponding sets of CR's are chosen as

$$\begin{aligned} &\underline{P}_A(\mathbf{G}) = \underline{P}_A(\mathbf{R}) = \{E\}\,, \\ &\underline{P}_A(\mathbf{X}) = \underline{P}_A(\mathbf{M}) = \{E, C_{31}^-, C_{31}^+\}\,. \end{aligned} \tag{60}$$

The first step in our approach is to determine the auxiliary group $\mathbf{Q}_A$ for the group G_A. To obtain the group of associations we inspect the corresponding tables in Ref. 12 (p. 634) and Ref. 13 (p. 374). The group of automorphisms of $Pm3m$ is given in Ref. 14. Therefore we have

$$\mathbf{Q}_A \cong D_{2h} \ltimes (Im3m \times C_2)\,, \tag{61}$$

where the group of associations is

$$D_{2h} = \{G1^+, G2^+, R1^+, R2^+, G1^-, G2^-, R1^-, R2^-\}\,. \tag{62}$$

We use a decomposition of the automorphism group $Im3m$, which clearly shows its relation to the group $G_A = Pm3m$ and $G_B = P23$,

$$\begin{aligned} Im3m &= (E\,|\mathbf{0})Pm3m \cup (E\,|\mathbf{B})Pm3m\,, \\ Pm3m &= (E\,|\mathbf{0})P23 \cup (C_{2b}|\mathbf{0})P23 \\ &\quad \cup (I\,|\mathbf{0})P23 \cup (\sigma_{db}|\mathbf{0})P23\,. \end{aligned} \tag{63}$$

We consider only the following automorphisms:

$$b_1 \leftrightarrow (C_{2B}|\mathbf{0})\,, \quad b_2 \leftrightarrow (I\,|\mathbf{0})\,, \quad b_3 \leftrightarrow (E\,|\mathbf{B})\,, \tag{64}$$

because the eight outer automorphisms of G_B can be generated from these ones. Now one can easily verify that $P23$ is a normal subgroup of $\mathrm{Aut}(Pm3m)$ and therefore

$$Q_A = \mathbf{Q}_A\,. \tag{65}$$

In order to determine the Q_A classes [cf. (14)] we have to derive the $q_A K$ tables. (See Tables I–IV.) For that purpose we have to inspect the KP tables of Ref. 13 if q_A is an association. The mapping of G_A irreps generated by the automorphism β_3 are given in Ref. 15.

From Tables I–IV one can readily deduce the following Q_A classes:

$$\begin{aligned} &[G1^+] = \{G1^+, G2^+, R1^+, R2^+, G1^-, G2^-, R1^-, R2^-\}\,, \\ &[G3^+] = \{G3^+, R3^+, G3^-, R3^-\}\,, \\ &[G4^+] = \{G4^+, G5^+, R4^+, R5^+, G4^-, G5^-, R4^-, R5^-\}\,, \\ &[X1^+] = \{X1^+, X2^+, X3^+, X4^+, M1^+, M2^+, M3^+, M4^+, X1^-, X2^-, X3^-, X4^-, M1^-, M2^-, M3^-, M4^-\}\,, \\ &[X5^+] = \{X5^+, M5^+, X5^-, M5^-\}\,, \end{aligned} \tag{66}$$

The next task is to determine the various groups Q_A^K [cf. Eq. (12)] that leave the corresponding class representatives invariant. Again taking into account the $q_A K$ tables and Eq. (61) we arrive at

$$\begin{aligned}
Q_A^{G1'} &= \mathrm{AUT}(G_A)\times \mathrm{CON}(G_A)\,,\\
Q_A^{G3'} &= \{G1^+,G2^+\}\times \mathrm{AUT}(G_A)\times \mathrm{CON}(G_A)\,,\\
Q_A^{G4'} &= Q_A^{G1'}\,,\\
Q_A^{X1'} &= Pm3m\times \mathrm{CON}(G_A)\,,\\
Q_A^{X5'} &= (G1^+,b_0,c_0)Q_A^{X1'}\cup(G2^+,b_0,c_0)Q_A^{X1'}\cup(G1^-,b_3,c_0)Q_A^{X1'}\cup(G2^-,b_3,c_0)Q_A^{X1'}\,.
\end{aligned}\tag{67}$$

Now we have to fix the corresponding sets of CR's. This turns out to be a nontrivial task if one wants to obtain simple generating relations. To achieve this goal we choose

$$\begin{aligned}
R_A^{G1^+} &= R_A^{G4^+} = \{G1^+,G2^+,R1^+,R2^+,G1^-,G2^-,R1^-,R2^-\} = \mathrm{ASS}(G_A)\,,\\
R_A^{G3^+} &= \{G1^+,R1^+,G1^-,R1^-\}\,,\\
R_A^{X1^+} &= \{(G1^+,b_0,c_0),(G2^+,b_0,c_0),(G1^-,b_3,c_0),(G2^-,b_3,c_0),(R1^+,b_0,c_0),(R2^+,b_0,c_0),\\
&\quad (R1^-,b_3,c_0),(R2^-,b_3,c_0),(G1^-,b_0,c_0),(G2^-,b_0,c_0),(G1^+,b_3,c_0),(G2^+,b_3,c_0),\\
&\quad (R1^-,b_0,c_0),(R2^-,b_0,c_0),(R1^+,b_3,c_0),(R2^+,b_3,c_0)\}\,,\\
R_A^{X5^+} &= \{(G1^+,b_0,c_0),(R1^+,b_0,c_0),(G1^+,b_3,c_0),(R1^+,b_3,c_0)\}\,.
\end{aligned}\tag{68}$$

Next we have to determine the matrices $U_A^K(q_A)$ for $q_A\in Q_A^K$. We list only those matrices that are needed later on.

$$\begin{aligned}
K=G1^+:&\ U_A^K(b_3)=U_A^K(c)=1\,;\\
K=G3^+:&\ U_A^K(c)=\begin{pmatrix}0&1\\1&0\end{pmatrix},\\
&\ U_A^K(G2^+)=\begin{pmatrix}1&0\\0&-1\end{pmatrix},\\
&\ U_A^K(b_3)=\begin{pmatrix}1&0\\0&1\end{pmatrix}\equiv E(2)\,;\\
K=G4^+:&\ U_A^K(b_3)=U_A^K(c)=E(3)\,;\\
K=X1^+:&\ U_A^K(c)=E(3)\,;\\
K=X5^+:&\ U_A^K(c)=E(6)\,,\\
&\ U_A^K((G2^+,b_0,c_0))=E(3)\times\begin{pmatrix}1&0\\0&-1\end{pmatrix},\\
&\ U_A^K((G1^-,b_3,c_0))=E(3)\times\begin{pmatrix}0&-1\\1&0\end{pmatrix},\\
&\ U_A^K((G2^-,b_3,c_0))=E(3)\times\begin{pmatrix}0&1\\1&0\end{pmatrix}.
\end{aligned}\tag{69}$$

The matrices that are assigned to inner automorphisms of G_A can be constructed as proposed in Paper III [cf. Eqs. (3.32) and (3.33)].

In general, the group Q_B is a subgroup of $\mathbf{Q}_B$ [cf. Eqs. (4) and (10)]. The structure of the latter is

$$\mathbf{Q}_B=(C_3\times C_2')\circledS Im3m\times C_2\,. \tag{70}$$

In order to determine the structure of Q_B it is necessary to derive the compatibility relations of the considered irreps of G_A if they are subduced to G_B:

G_A:	qk	$G1^\pm$	$G2^\pm$	$G3^\pm$	$G4^\pm$	$G5^\pm$
G_B:	qs	$G1$	$G1$	$G2+G3$	$G4$	$G4$

G_A:	qk	$R1^\pm$	$R2^\pm$	$R3^\pm$	$R4^\pm$	$R5^\pm$
G_B:	qs	$R1$	$R1$	$R2+R3$	$R4$	$R4$

G_A:	qk	$X1^\pm$	$X2^\pm$	$X3^\pm$	$X4^\pm$	$X5^\pm$
G_B:	qs	$X1$	$X1$	$X4$	$X4$	$X2+X3$

G_A:	qk	$M1^\pm$	$M2^\pm$	$M3^\pm$	$M4^\pm$	$M5^\pm$
G_B:	qs	$M1$	$M1$	$M2$	$M2$	$M3+M4$

(71)

From these relations we deduce

$$Q_B\cong C_2'\circledS(Im3m\times C_2)\,, \tag{72}$$

where C_2' is generated by $R1$. It is worth noting that Q_B is a proper subgroup of $\mathbf{Q}_B$. Hence the Q_B classes differ from the $\mathbf{Q}_B$ classes, i.e., they may be split into subsets of the latter. We find the following Q_B classes inspecting the $q_B S$ tables of Ref. 3:

$$\begin{aligned}
[G1]&=\{G1,R1\}\,,\\
[G2]&=\{G2,G3,R2,R3\}\,,\\
[G4]&=\{G4,R4\}\,,\\
[X1]&=\{X1,X4,M1,M2\}\,,\\
[X2]&=\{X2,X3,M3,M4\}\,.
\end{aligned}\tag{73}$$

In particular, $[G1]\cup[G2]$ coincides with $[G1]$ of Ref. 3 [cf. Eq. (3.26)] [note that in (73) and in (66) the subscripts B and A are omitted].

Next we determine the subgroups Q_B^S of Q_B,

$$\begin{aligned}
Q_B^{G1} &= Q_B^{G4}=\mathrm{AUT}(G_B)\times \mathrm{CON}(G_B)\cong Im3m\times C_2\,,\\
Q_B^{G2} &= (G1,b_0,c_0)Im3\cup(G1,b_1,c)Im3\,,\\
Q_B^{X1} &= Pm3m\times \mathrm{CON}(G_B)\cong Pm3m\times C_2\,,
\end{aligned}$$

$$Q_B^{X2} = \{(E|\mathbf{0})P23 \cup (I|\mathbf{0})P23 \cup (C_{2b}|\mathbf{B})P23 \cup (\sigma_{db}|\mathbf{B})P23\} \times \mathrm{CON}(G_B) \cong Pm3n \times C_2 . \tag{74}$$

The corresponding sets of CR's are chosen as

$$\begin{aligned} R_B^{G1} &= R_B^{G4} = \{(G1,b_0,c_0),(R1,b_0,c_0)\} \cong \mathrm{ASS}(G_B) , \\ R_B^{G2} &= \{(G1,b_0,c_0),(G1,b_0,c), (R1,b_0,c_0),(R1,b_0,c)\} , \\ R_B^{X1} &= R_B^{X2} = \{(G1,b_0,c_0),(G1,b_3,c_0) (R1,b_0,c_0),(R1,b_3,c_0)\} . \end{aligned} \tag{75}$$

The matrices $U_B^S(q_B)$ are given in III [cf. Eqs. (3.34)–(3.38)].

The subgroups Q_B^{K1} [cf. Eq. (34)] that are the homomorphic images of Q_A^K are the following:

$$\begin{aligned} &Q_B^{G1^+1} = Q_B^{G3'1} = Q_B^{G4'1} = Q_B^{X5^+1} \cong Im3m \times C_2 , \\ &Q_B^{X1'1} \cong Pm3m \times C_2 . \end{aligned} \tag{76}$$

Because of

$$\begin{aligned} &Q_B^{G1} = Q_B^{G1'1} , \\ &Q_B^{G4} = Q_B^{G4'1} , \\ &Q_B^{X1} = Q_B^{X1'1} , \end{aligned} \tag{77}$$

it is obvious that there do not exist generating relations of the second kind for the subducing matrices $S^{G1'}, S^{G4'}, S^{X1'}$, whereas for the remaining cases relations of that kind may be expected. Therefore we only derive the Q_B^{K1} classes [cf. Eq. (35)] for the nontrivial cases. However, we give only these classes of G_B irreps that are subduced from Q_A class representatives,

$$\begin{aligned} &[G2]^{G3'1} = \{G2,G3\} , \\ &[X2]^{X5^+1} = \{X2,X3\} \end{aligned} \tag{78}$$

The corresponding $Q_B^{K1,T}$ groups [cf. Eq. (36)] are

$$\begin{aligned} &Q_B^{G3'1,G2} = Q_B^{G2} , \\ &Q_B^{X5'1,X2} = Q_B^{X2} \cong Pm3n \times C_2 . \end{aligned} \tag{79}$$

Their inverse images $Q_A^{K,T} = \phi^{-1}(Q_B^{K1,T})$ turn out to be

$$\begin{aligned} Q_A^{G3',G2} &= \phi^{-1}(Q_B^{G3'1,G2}) \\ &\cong (G1^+,b_0,c_0)Im3 \cup (G1^+,b_1,c)Im3 \\ &\quad \cup (G2^+,b_0,c_0)Im3 \cup (G2^+,b_1,c)Im3, \\ Q_A^{X5',X2} &= \phi^{-1}(Q_B^{X5'1,X2}) \\ &\cong (G1^+,b_0,c_0)[Pm3 \times C_2] \cup (G2^+,b_0,c_0) \\ &\quad \times [Pm3 \times C_2] \cup (G1^-,b_3,C_0)[Pm3 \times C_2] \\ &\quad \cup (G2^-,b_3,c_0)[Pm3 \times C_2] . \end{aligned} \tag{80}$$

Then we take the following sets of CR's $R_B^{K1,T}$:

$$\begin{aligned} &R_B^{G3'1,G2} = \{(G1,b_0,c_0),(G1,b_0,c)\} , \\ &R_B^{X5'1,X2} = \{(G1,b_0,c_0),(G1,b_3,c_0)\} , \end{aligned} \tag{81}$$

and choose the inverse images as

$$\begin{aligned} &\phi^{-1}(R_B^{G3'1,G2}) = \{(G1^+,b_0,c_0),(G1^+,b_0,c)\} , \\ &\phi^{-1}(R_B^{X51,X2}) = \{(G1^+,b_0,c_0),(G1^-,b_3,c_0)\} . \end{aligned} \tag{82}$$

Now we are in the position to derive the generating relations of the first and of the second kind. Starting with generating relations of the first kind [cf. Eq. (32)] we conclude from Eqs. (68) and (75) the following relations:

$$S^{G1'} = S^{G2'} = S^{R1'} = S^{R2'} = S^{G1} = S^{G2} = S^{R1} = S^{R2} , \tag{83}$$

$$S^{G3'} = S^{R3'} = S^{G3} = S^{R3} , \tag{84}$$

$$S^{G4'} = S^{G5'} = S^{R4'} = S^{R5'} = S^{G4} = S^{G5} = S^{R4} = S^{R5} , \tag{85}$$

$$\begin{aligned} S^{X1'} = S^{X2'} = S^{X3'} = S^{X4'} &= S^{M1'} = S^{M2'} = S^{M3'} = S^{M4'} \\ &= S^{X1} = S^{X2} = S^{X3} = S^{X4} \\ &= S^{M1} = S^{M2} = S^{M3} = S^{M4} , \end{aligned} \tag{86}$$

$$S^{X5'} = S^{X5} = S^{M5'} = S^{M5} . \tag{87}$$

This implies that the corresponding matrices, $U_A^{K,L}(q_A)$, $Z_B^K(q_B)$ with $q_B = \phi(q_A)$, and the permutational matrices, are chosen as unit matrices. By similar arguments we arrive at the following nontrivial generating relations of the second kind:

$$S_{G3}^{G3'} = U_A^{G3'}((G1^+,b_0,c))(S_{G2}^{G3'})^* = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} (S_{G2}^{G3'})^* \tag{88}$$

$$\begin{aligned} S_{X3}^{X5^+} &= U_A^{X5'}((G1^-,b_3,c_0)) S_{X2}^{X5'} \\ &= \left[E(3) \times \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \right] S_{X2}^{X5'} . \end{aligned} \tag{89}$$

Here we have used Eqs. (69), (81), and (82).

Finally we have to compute (parts of) the subducing matrices for the class representatives. Direct calculations yield

$$\begin{aligned} &S^{G1^+} = 1 , \quad S_{G2}^{G3^+} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \Rightarrow S^{G3^+} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \\ &S^{G4^+} = E(3) , \quad S^{X1'} = E(3) , \end{aligned} \tag{90}$$

and

$$S_{X2}^{X5^+} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\Rightarrow S^{X5^+} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} .$$

It is clear from the compatibility relations (71) that in the present example all the branching multiplicities are equal to one. Hence there is no multiplicity problem that has to be solved.

ACKNOWLEDGMENTS

This paper is a part of the scientific and technical cooperation between Austria and Bulgaria.

Financial support of research visits in Vienna (J.N.K. and M.K.P) and in Sofia (R.D. and P.K.) is gratefully acknowledged. M.I.A. acknowledges a fellowship by A. v. Humboldt-Stiftung.

[1]R. Dirl, P. Kasperkovitz, M. I Aroyo, J. N. Kotzev, and M. Angelova-Tjurkedjieva, J. Math. Phys. **27**, 37 (1986).

[2]M. I. Aroyo, J. N. Kotzev, M. Angelova-Tjurkedjieva, R. Dirl, and P. Kasperkovitz, J. Math. Phys. **27**, 2236 (1986).

[3]R. Dirl, P. Kasperkovitz, M. I. Aroyo, J. N. Kotzev, and B. L. Davies, J. Math. Phys. **28**, 1947 (1987).

[4]C. J. Bradley and A. P. Crackwell, *The Mathematical Theory of Symmetry in Solids* (Oxford U.P., Oxford, 1972).

[5]A. M. Leushin, *Tables of Functions Transforming by the Equivalent Representations of the Crystallographic Point Groups* (Nauka, Moscow, 1968) (in Russian).

[6]M. Matsumoto and H. Kanamory, Mem. Fac. Indust. Arts Kyoto Tech. Univ. Sci. Tech. **26**, 29 (1977).

[7]J. L. Birman, *Theory of Crystal Space Groups and Infra-Red and Raman Lattice Processes of Insulating Crystals in Handbuch der Physik*, edited by S. Flüge, Bd. XXV/2b (Springer, Berlin, 1974).

[8]L. D. Landau and E. M. Lifshitz, *Statistical Physics, Part 1*, Course of Theoretical Physics (Pergamon, Oxford, 1958), Vol. 5, Chap. XIV.

[9]A. Abraham and B. Bleaney, *Electronic Paramagnetic Resonance of Transition Ions* (Clarendon, Oxford, 1970).

[10]Yu. M. Gufan, *Structure Phase Transitions* (Nauka, Moscow, 1982), p. 159 (in Russian).

[11]L. Jansen and M. Boon, *Theory of Finite Groups. Applications in Physics* (North-Holland, Amsterdam, 1967), Chap. VI, p. 146.

[12]A. P. Cracknell, B. L. Davies, S. C. Miller, and W. F. Love, *Kronecker Product Tables* (Plenum, New York, 1979), Vol. 1.

[13]B. L. Davies and A. P. Cracknell, *Kronecker Product Tables* (Plenum, New York, 1979), Vol. 2.

[14]H. Burzlaff and H. Zimmermann, Z. Krist. **153**, 151 (1980).

[15]B. L. Davies (private communication).

Character calculations for semisimple Lie algebras

L. Vretare
Department of Mathematics, Lund Institute of Technology, Box 118, S-221 00 Lund, Sweden

(Received 7 June 1988, accepted for publication 7 September 1988)

A recursive method for calculation of characters of semisimple Lie algebras is outlined. By slight modifications algorithms are obtained for the computation of weight multiplicities, Kronecker products, and branching rules, as well as symmetrized Kronecker powers.

I. INTRODUCTION

Recent work on characters of semisimple Lie algebras has been concentrated on methods that avoid summation over root systems and Weyl groups.[1–6] However, the purpose of the present paper is to propose a simple algorithm for recursive calculation of characters using summation over the orbits under the Weyl group of the fundamental weights. Such orbits are, in general, much smaller than the Weyl group, which makes calculation on a PC possible even if the rank is high. The efficiency of new methods is often measured by calculations for the Lie algebra E_8. Here is a list of the results for E_8 obtained by the author from an implementation in TURBO PASCAL on a PC: weight multiplicities for 135 representations, Kronecker products of any two of the 12 lowest nontrivial representations, Kronecker products of one of the two lowest and one of the 40 lowest nontrivial representations, symmetrized squares of 11, cubes of 5, and fourth power of 2 nontrivial representations, and branching rules $E_8 \downarrow D_8$ for 46 and $E_8 \downarrow A_8$ for 18 nontrivial representations.

A check of the literature shows that these results reach far beyond earlier computations. Some errors in earlier computations of branching rules were also discovered.

Properties of weights and characters used in the paper are collected in Sec. II. Sections III–VI treat weight multiplicities, Kronecker products, branching rules, and symmetrized Kronecker powers, respectively. A detailed description of the method is found in Sec. III while Secs. IV–VI contain modifications of the method. In each section we give an example of the calculations for the Lie algebra C_3.

II. NOTATION AND RECURRENCE FORMULA

Let $\alpha_1 \cdots \alpha_l$ denote the simple roots and $\mu_1 \cdots \mu_l$ the fundamental weights of a semisimple Lie algebra. The characters, or equivalently the irreducible finite-dimensional representations, are labeled by the dominant weights

$$\Lambda = \sum_{i=1}^{l} n_i \mu_i,$$

where all n_i are non-negative integers. Moreover, the weight system is the set of all integral linear combinations of the fundamental weights. The weight system is partially ordered by

$$\Lambda_1 < \Lambda_2 \quad \text{if } \Lambda_2 - \Lambda_1 \quad \text{is a sum of positive roots.}$$

Note that all weights ν belonging to the representation with dominant weight Λ satisfy

$$\nu \leqslant \Lambda.$$

When we use the words "lower than" and "higher than" we will always refer to a total ordering denoted by $<$, which is compatible with the partial ordering, i.e.,

$$\Lambda_1 < \Lambda_2 \Rightarrow \Lambda_1 < \Lambda_2.$$

The most natural example of such a total ordering is lexicographic ordering with respect to the simple roots.

Let Λ be any weight (not necessarily a dominant one) and $W\Lambda$ the orbit of Λ under the Weyl group W. For convenience we also denote by $W\Lambda$ the factor group W/W_Λ, where W_Λ is the subgroup of W leaving Λ fixed. Put

$$\Delta_\Lambda = \sum_{S \in W\Lambda} \exp(S\Lambda),$$

$$\xi_\Lambda = \sum_{S \in W} \det S \exp(S(\Lambda + \rho)),$$

and

$$\chi_\Lambda = \xi_{\Lambda+\rho}/\xi_\rho,$$

where

$$\rho = \sum_{i=1}^{l} \mu_i.$$

If Λ is dominant then χ_Λ is, of course, the corresponding character given by Weyl's formula.

Before stating the recurrence formula we collect some useful properties:

$$\Delta_\Lambda = \Delta_{S\Lambda}, \quad S \in W, \tag{1}$$

$$\chi_\Lambda = \det S \, \chi_{S(\Lambda+\rho)-\rho}, \quad S \in W, \tag{2}$$

$$\Delta_\nu \chi_\Lambda = \sum_{S \in W\nu} \chi_{\Lambda + S\nu}, \tag{3}$$

if $\Lambda + \rho$ belongs to the wall of a Weyl chamber then $\chi_\Lambda = 0$, (4)

if $\Lambda + \rho$ belongs to the interior of a Weyl chamber then there is a $S \in W$ such that $S(\Lambda + \rho) - \rho$ is dominant. (5)

Recurrence formula: If Λ and ν are dominant then

$$\chi_{\Lambda+\nu} = \Delta_\nu \chi_\Lambda - \sum_{S \in W\nu, S \neq \mathrm{Id}} \chi_{\Lambda + S\nu}. \tag{6}$$

Moreover, either there is a $T \in W$ such that $T(\Lambda + S\nu + \rho) - \rho$ is dominant and lower than $\Lambda + \nu$, or $\chi_{\Lambda + S\nu} = 0$.

III. WEIGHT MULTIPLICITIES

The multiplicities of the weights of χ_Λ are the positive integers c_ν in the expansion

 0022-2488/89/010018-05$02.50

$$\chi_\Lambda = \sum c_\nu \Delta_\nu.$$

Such expansions are well suited for implementation as lists in TURBO PASCAL. In that case the calculation of characters may be considered as an operation like addition, multiplication, etc. We start with an instructive example.

Consider the Lie algebra C_3 for which

$$\alpha_1 = (1,-1,0),\quad \alpha_2 = (0,1,-1),\quad \alpha_3 = (0,0,2),$$
$$\mu_1 = (1,0,0),\quad \mu_2 = (1,1,0),\quad \mu_3 = (1,1,1),$$

and

$$\rho = \mu_1 + \mu_2 + \mu_3 = (3,2,1).$$

The weights Λ have the form

$$\Lambda = n_1\mu_1 + n_2\mu_2 + n_3\mu_3 = (m_1,m_2,m_3),$$

where

$$\begin{aligned} m_1 &= n_1 + n_2 + n_3,\\ m_2 &= \phantom{n_1 +{}} n_2 + n_3,\\ m_3 &= \phantom{n_1 + n_2 +{}} n_3. \end{aligned}$$

Furthermore Λ is dominant if $m_1 \geqslant m_2 \geqslant m_3 \geqslant 0$. The Weyl group is the group of permutations that change signs of arbitrarily many variables. Its order is 48. Now let us compute $\chi_{(2,1,0)}$ from the recurrence formula (6) assuming that all lower characters are known:

$$\begin{aligned} \chi_{(2,1,0)} &= \Delta_{(1,0,0)}\chi_{(1,1,0)} - \chi_{(1,2,0)} - \chi_{(1,1,1)}\\ &\quad - \chi_{(0,1,0)} - \chi_{(1,0,0)} - \chi_{(1,1,-1)}. \end{aligned}$$

By (2) and (4) we have

$$\chi_{(1,2,0)} = \det S\,\chi_{S(4,4,1)-(3,2,1)} = 0,$$

since (4,4,1) remains unchanged when the two fours are transposed. Similar arguments yield

$$\chi_{(0,1,0)} = \chi_{(1,1,-1)} = 0.$$

Inserting the expressions for $\chi_{(1,1,0)}$, $\chi_{(1,1,1)}$, and $\chi_{(1,0,0)}$, which are known by assumption, we obtain

$$\begin{aligned} \chi_{(2,1,0)} &= \Delta_{(1,0,0)}(\Delta_{(1,1,0)} + 2\Delta_{(0,0,0)})\\ &\quad - (\Delta_{(1,1,1)} + \Delta_{(1,0,0)}) - \Delta_{(1,0,0)}. \end{aligned}$$

Finally the product is found by direct multiplication:

$$\Delta_{(1,0,0)}\Delta_{(1,1,0)} = \Delta_{(2,1,0)} + 3\Delta_{(1,1,1)} + 4\Delta_{(1,0,0)}.$$

Thus

$$\chi_{(2,1,0)} = \Delta_{(2,1,0)} + 2\Delta_{(1,1,1)} + 4\Delta_{(1,0,0)}.$$

Remark: To illustrate (5) consider the following example:

$$\begin{aligned} \chi_{(-4,2,1)} &= \det S\,\chi_{S(-1,4,2)-(3,2,1)}\\ &= -\chi_{(4,2,1)-(3,2,1)} = -\chi_{(1,0,0)}. \end{aligned}$$

In the general case we proceed as outlined in the example above. Thus to compute χ_Λ when all lower characters are known we first find the lowest fundamental weight μ_i such that $\Lambda - \mu_i$ is dominant and then use the recurrence formula (6):

$$\chi_\Lambda = \Delta_{\mu_i}\chi_{\Lambda-\mu_i} - \sum_{S\in W\mu_i, S\neq I} \chi_{\Lambda-\mu_i-S\mu_i}.$$

To handle the sum, the values of $S\mu_i$ for all $S \in W$ must be known. Then $\chi_{\Lambda-\mu_i+S\mu_i}$ is easily obtained from (2), (4), and (5) and known expressions for lower characters. For the evaluation of the remaining term we insert the known expression for $\chi_{\Lambda-\mu_i}$ and carry out multiplications of the type

$$\Delta_\mu\Delta_\Lambda = \sum a_\nu \Delta_\nu.$$

Of course, all such products are saved for later use but the first time they occur we proceed as follows.

Compute

$$\Delta_\mu \exp\Lambda = \sum_{S\in W\mu} \exp(S\mu + \Lambda)$$

and identify all $\exp(S\mu + \Lambda)$ that are terms of the same Δ_ν:

$$\Delta_\mu \exp\Lambda = \sum b_\nu \text{ terms of } \Delta_\nu.$$

Now because of the Weyl group symmetry (1) we also have

$$\Delta_\mu \exp S\Lambda = \sum b_\nu \text{ terms of } \Delta_\nu$$

and hence

$$\Delta_\mu\Delta_\Lambda = |W\Lambda| \sum b_\nu \text{ terms of } \Delta_\nu,$$

where $|W\Lambda|$ denotes the order of $W\Lambda$. Since $|W\nu|$ terms of Δ_ν constitute one Δ_ν we get

$$\Delta_\mu\Delta_\Lambda = \sum b_\nu \frac{|W\Lambda|}{|W\nu|}\Delta_\nu.$$

Thus for the multiplication of Δ_μ and Δ_Λ we also need the values of $|W\nu|$ for ν lower than $\Lambda + \mu$. Note, however, that since the leading coefficient must be equal to 1,

$$b_{\Lambda+\mu}|W\Lambda|/|W(\Lambda+\mu)| = 1,$$

the value of $|W(\Lambda+\mu)|$ is obtained as

$$|W(\Lambda+\mu)| = b_{\Lambda+\mu}|W\Lambda|.$$

Consider again the Lie algebra C_3 and the product $\Delta_{(1,0,0)}\Delta_{(1,1,0)}$ in the example above. From

$$\begin{aligned} &\Delta_{(1,0,0)}\exp(1,1,0)\\ &\quad = \exp(2,1,0) + \exp(1,2,0) + \exp(1,1,1)\\ &\qquad + \exp(0,1,0) + \exp(1,0,0) + \exp(1,1,-1)\\ &\quad = 2 \text{ terms of } \Delta_{(2,1,0)} + 2 \text{ terms of } \Delta_{(1,1,1)}\\ &\qquad + 2 \text{ terms of } \Delta_{(1,0,0)}, \end{aligned}$$

we conclude that

$$\begin{aligned} &\Delta_{(1,0,0)}\Delta_{(1,1,0)}\\ &\quad = 2\frac{|W(1,1,0)|}{|W(2,1,0)|}\Delta_{(2,1,0)}\\ &\qquad + 2\frac{|W(1,1,0)|}{|W(1,1,1)|}\Delta_{(1,1,1)} + 2\frac{|W(1,1,0)|}{|W(1,0,0)|}\Delta_{(1,0,0)}. \end{aligned}$$

Earlier calculations show that $|W(1,1,0)| = 12$, $|W(1,1,1)| = 8$, and $|W(1,0,0)| = 6$. Consequently

$$|W(2,1,0)| = 2|W(1,1,0)| = 24$$

and

$\Delta_{(1,0,0)}\Delta_{(1,1,0)} = \Delta_{(2,1,0)} + 3\Delta_{(1,1,1)} + 4\Delta_{(1,0,0)}$.

Summary of the method:

(1) Preparations specific for the Lie algebra under consideration.

(a) Write a procedure called "high" that for a given weight ν finds the highest one, if there is any, of $S(\nu+\rho)-\rho$, where $S \in W$ [cf. (4) and (5)].

(b) Generate an ordering of the dominant weights.

(c) Generate $S\mu_i$ for all S and all i.

(2) The recursive step calculating χ_Λ.

(a) Find the lowest μ_i such that $\Lambda - \mu_i$ is dominant.

(b) For each $S\mu_i$ use the procedure "high" to find the earlier saved character $\chi_{\Lambda-\mu_i+S\mu_i}$ and its sign (or zero) [cf. (2) and (4)].

(c) Compute $\Delta_{\mu_i}\chi_{\Lambda-\mu_i}$ and $|W\Lambda|$. Use earlier results for $\chi_{\Lambda-\mu_i}$ and $\Delta_{\mu_i}\Delta_\nu$. Save $|W\Lambda|$ and all new products $\Delta_{\mu_i}\Delta_\nu$.

(d) Save χ_Λ.

Of course the speed of one recursion depends on the chosen μ_i or, more precisely, on $|W\mu_i|$. For example, some relevant orders for E_8 and C_{10} are

$$E_8:\quad |W| = 696\,729\,600,\quad \min|W\mu_i| = 240,$$
$$\max|W\mu_i| = 483\,840,$$
$$C_{10}:\quad |W| = 3715\,891\,200,\quad \min|W\mu_i| = 20,$$
$$\max|W\mu_i| = 15\,360.$$

IV. KRONECKER PRODUCTS

Resolving the Kronecker product of two irreducible representations is equivalent to finding the positive integers c_ν in

$$\chi_{\Lambda_1}\chi_{\Lambda_2} = \sum c_\nu \chi_\nu.$$

This will be done recursively over Λ_2 by use of the recurrence formula (6):

$$\chi_{\Lambda_1}\chi_{\Lambda_2} = \chi_{\Lambda_1}\Big(\Delta_\mu \chi_{\Lambda_2-\mu} - \sum \chi_{\Lambda_2-\mu+S\mu}\Big)$$
$$= \Delta_\mu \chi_{\Lambda_1}\chi_{\Lambda_2-\mu} - \sum \chi_{\Lambda_1}\chi_{\Lambda_2-\mu+S\mu}.$$

Here $\chi_{\Lambda_1}\chi_{\Lambda_2-\mu+S\mu}$ and $\chi_{\Lambda_1}\chi_{\Lambda_2-\mu}$ are either zero or computed earlier. The multiplication of Δ_μ can be simplified according to (3):

$$\Delta_\mu \chi_{\Lambda_1}\chi_{\Lambda_2-\mu} = \Delta_\mu \sum a_\nu \chi_\nu = \sum a_\nu \chi_{\nu+S\mu}.$$

Once more we use C_3 as an example:

$$\chi_{(2,2,2)}\chi_{(2,1,0)}$$
$$= \chi_{(2,2,2)}(\Delta_{(1,0,0)}\chi_{(1,1,0)} - \chi_{(1,1,1)} - \chi_{(1,0,0)})$$
$$= \Delta_{(1,0,0)}\chi_{(2,2,2)}\chi_{(1,1,0)} - \chi_{(2,2,2)}\chi_{(1,1,1)}$$
$$- \chi_{(2,2,2)}\chi_{(1,0,0)}$$
$$= \Delta_{(1,0,0)}(\chi_{(3,3,2)} + \chi_{(3,2,1)} + \chi_{(2,1,1)})$$
$$- (\chi_{(3,3,3)} + \chi_{(3,3,1)} + \chi_{(3,1,1)}$$
$$+ \chi_{(1,1,1)}) - (\chi_{(3,2,2)} + \chi_{(2,2,1)}).$$

Working out the three remaining products,

$$\Delta_{(1,0,0)}\chi_{(3,3,2)}$$
$$= \chi_{(4,3,2)} + \chi_{(3,3,3)} + \chi_{(3,2,2)} + \chi_{(3,3,1)},$$
$$\Delta_{(1,0,0)}\chi_{(3,2,1)}$$
$$= \chi_{(4,2,1)} + \chi_{(2,2,1)} + \chi_{(3,3,1)} + \chi_{(3,1,1)}$$
$$+ \chi_{(3,2,2)} + \chi_{(3,2,0)},$$

and

$$\Delta_{(1,0,0)}\chi_{(2,1,1)} = \chi_{(3,1,1)} + \chi_{(1,1,1)} + \chi_{(2,2,1)} + \chi_{(2,1,0)},$$

we obtain

$$\chi_{(2,2,2)}\chi_{(2,1,0)}$$
$$= \chi_{(4,3,2)} + \chi_{(4,2,1)} + \chi_{(3,3,1)} + \chi_{(3,2,2)}$$
$$+ \chi_{(3,2,0)} + \chi_{(3,1,1)} + \chi_{(2,2,1)} + \chi_{(2,1,0)}.$$

Summary: To calculate Kronecker products use the computer program for weight multiplicities but replace the procedure for multiplication

$$\Delta_\mu \Delta_\Lambda = \sum a_\nu \Delta_\nu$$

by

$$\Delta_\mu \chi_\Lambda = \sum b_\nu \chi_\nu.$$

V. BRANCHING RULES

Under restriction to a subalgebra the characters $\tilde{\chi}_\Lambda$ can be further decomposed as

$$\tilde{\chi}_\Lambda = \sum c_\nu \psi_\nu,$$

where the ψ_ν denote the characters of the subalgebra and $\tilde{\chi}_\Lambda$ is the restriction of χ_Λ to the subalgebra. To compute such branching rules we take the restriction of the recurrence formula (6):

$$\tilde{\chi}_\Lambda = \tilde{\Delta}_{\mu_i}\tilde{\chi}_{\Lambda-\mu_i} - \sum_{S\in W_{\mu_i}, S\neq I} \tilde{\chi}_{\Lambda-\mu_i+S\mu_i}.$$

As before $\tilde{\chi}_{\Lambda-\mu_i+S\mu_i}$ is handled by the procedure "high" and earlier results for lower characters. Also $\tilde{\chi}_{\Lambda-\mu_i}$ is assumed to be known:

$$\tilde{\chi}_{\Lambda-\mu_i} = \sum a_\nu \psi_\nu.$$

This time the multiplication will be worked out on subalgebra level. First we express $\tilde{\Delta}_{\mu_i}$ in terms of weights λ of the subalgebra:

$$\tilde{\Delta}_{\mu_i} = \sum_\lambda \exp \lambda.$$

Next by use of property (3) for the subalgebra we get

$$\tilde{\Delta}_{\mu_i}\tilde{\chi}_{\Lambda-\mu_i} = \sum_{\nu,\lambda} a_\nu \psi_{\nu+\lambda}.$$

Finally $\psi_{\nu+\lambda}$ is calculated by a second procedure "high" valid for the subalgebra.

The linear mapping of the weight system of the algebra into the one of the subalgebra is found in the following way.[7,8]

Consider the restriction C_3 to A_2. The simple roots of A_2 are

$$\beta_1 = (1,-1,0),\quad \beta_2 = (0,1,-1),$$

and the fundamental weights

$$\lambda_1 = (1,0,0) - \tfrac{1}{3}(1,1,1),\quad \lambda_2 = (1,1,0) - \tfrac{2}{3}(1,1,1).$$

Moreover, the half sum of the positive roots is

$$\sigma = \lambda_1 + \lambda_2 = (1,0,-1).$$

The Weyl group is the group of permutations. Because of nonintegral weights we prefer to use the fundamental weights as a basis for labeling the characters of A_2. The embedding of A_2 into C_3 is defined by[8]

$$\tilde{\chi}_{(1,0,0)} = \psi_{(1,0)} + \psi_{(0,1)}.$$

Let $M_1,\ldots,M_6$ be the weights of $\psi_{(1,0)} + \psi_{(0,1)}$ listed in decreasing order. The ordering is assumed to be lexicographic ordering with respect to β_1 and β_2. Thus

$$M_1 = (1,0),\quad M_2 = (0,1),\quad M_3 = (1,-1),$$
$$M_4 = (-1,1),\quad M_5 = (0,-1),\quad M_6 = (-1,0).$$

The desired linear mapping is now given by

$$\mu_1 = (1,0,0) \to M_1 = (1,0),$$
$$\mu_2 = (1,1,0) \to M_1 + M_2 = (1,1),$$
$$\mu_3 = (1,1,1) \to M_1 + M_2 + M_3 = (2,0).$$

For example, let us calculate $\tilde{\chi}_{(1,1,1)}$. The restricted recurrence formula reads

$$\tilde{\chi}_{(1,1,1)} = \tilde{\Delta}_{(1,1,1)} - \sum_{S\in W(1,1,1),\ S\neq I} \tilde{\chi}_{S(1,1,1)}.$$

Here

$$\sum_{S\in W(1,1,1),\ S\neq I} \tilde{\chi}_{S(1,1,1)} = -\tilde{\chi}_{(1,0,0)} = -\psi_{(1,0)} - \psi_{(0,1)}$$

and

$$\begin{aligned}\Delta_{(1,1,1)} &= \exp(1,1,1) + \exp(-1,1,1) + \exp(1,-1,1)\\ &\quad + \exp(1,1,-1)\\ &\quad + \exp(1,-1,-1) + \exp(-1,1,-1)\\ &\quad + \exp(-1,-1,1) + \exp(-1,-1,-1)\\ &\to \exp(2,0) + \exp(0,0) + \exp(2,-2)\\ &\quad + \exp(0,2) + \exp(0,0) + \exp(-2,2)\\ &\quad + \exp(0,-2) + \exp(-2,0).\end{aligned}$$

Hence

$$\begin{aligned}\tilde{\Delta}_{(1,1,1)} &= \psi_{(2,0)} + \psi_{(0,0)} + \psi_{(2,-2)}\\ &\quad + \psi_{(0,2)} + \psi_{(0,0)} + \psi_{(-2,2)} + \psi_{(0,-2)} + \psi_{(-2,0)}\\ &= \psi_{(2,0)} + \psi_{(0,2)} + 2\psi_{(0,0)} - \psi_{(1,0)} - \psi_{(0,1)}\end{aligned}$$

and

$$\tilde{\chi}_{(1,1,1)} = \psi_{(2,0)} + \psi_{(0,2)} + 2\psi_{(0,0)}.$$

Summary:

(1) Preparations.

(a) Write the procedures "high," one for each algebra.

(b) Generate orderings of the dominant weights, one for each algebra.

(c) Generate $S\mu_i$ for all S and i.

(d) Find the linear mapping between the two weight systems.

(2) The computation scheme is the same as the Kronecker products.

The following errors were found in the literature. In Ref. 9, p. 423, the representation $(4^3 3^5)$ of E_8 contains no $\{3^2 2^3\}$ but one $\{32^3\}$ of A_8. In Ref. 9, p. 423, the representation (543^6) of E_8 contains no $\{3^2 2^4 1\}$ but two $\{32^4 1\}$ of A_8. In Ref. 10, p. 3432, the representation $[2_1 1_7]$ of E_8 contains each of the representations (2_7) and (2_8) of D_8 once.

VI. SYMMETRIZED KRONECKER POWERS

Let λ be a partition of an integer and $\{\lambda\}$ the corresponding Schur function.[11] For example, the Schur functions of order 2 and 3 are

$$\{2\} = \tfrac{1}{2}(S_1^2 + S_2),$$
$$\{1^2\} = \tfrac{1}{2}(S_1^2 - S_2),$$
$$\{3\} = \tfrac{1}{6}(S_1^3 + 3S_1S_2 + 2S_3),$$
$$\{21\} = \tfrac{1}{6}(2S_1^3 - 2S_3),$$
$$\{1^3\} = \tfrac{1}{6}(S_1^3 - 3S_1S_2 + 2S_3).$$

The plethysm of a character χ_Λ corresponding to the symmetry class λ is obtained from $\{\lambda\}$ if we replace S_j by $\chi_\Lambda(jg)$ for all j. Thus

$$\chi_\Lambda \otimes \{2\} = \tfrac{1}{2}((\chi_\Lambda(g))^2 + \chi_\Lambda(2g)),$$
$$\chi_\Lambda \otimes \{1^2\} = \tfrac{1}{2}((\chi_\Lambda(g))^2 - \chi_\Lambda(2g)),$$

and so on. We now show how the decomposition of a plethysm,

$$\chi_\Lambda \otimes \{\lambda\} = \sum c_\nu \chi_\nu,$$

can be obtained by a combination of the computer programs for weight multiplicities and Kronecker products together with one more application of (3).

To find the decomposition of $\chi_\Lambda(jg)$ we use the weight multiplicities for χ_Λ

$$\chi_\Lambda(g) = \sum a_\nu \Delta_\nu(g).$$

Then obviously

$$\chi_\Lambda(jg) = \sum a_\nu \Delta_\nu(jg) = \sum a_\nu \Delta_{j\nu}(g).$$

Now $\Delta_{j\nu}$ can be expressed by (3) as

$$\Delta_{j\nu} = \sum_{S\in W\nu} \det S\, \chi_{jS\nu}$$

and then as a character sum by the procedure "high." Note, however, that to do this it will not be sufficient to generate $S\mu_i$ as usual. In fact, it is necessary to generate $S\nu$ for all weights ν belonging to Λ and all $S\in W\nu$. To complete the calculation use subsequent Kronecker multiplication as described in Sec. III.

As a final example consider the Lie algebra C_3 and the plethysms

$$\chi_{(1,0,0)} \otimes \{1^2\} \quad \text{and} \quad \chi_{(1,0,0)} \otimes \{2\}.$$

Weight multiplicity calculations show that

$$\chi_{(1,0,0)} = \Delta_{(1,0,0)},$$

from which it follows that

$$\chi_{(1,0,0)}(2g) = \Delta_{(2,0,0)}(g).$$

Hence

$$\begin{aligned}\chi_{(1,0,0)}(2g) &= \Delta_{(2,0,0)} \\ &= \sum \chi_{2S(1,0,0)} \\ &= \chi_{(2,0,0)} + \chi_{(-2,0,0)} + \chi_{(0,2,0)} \\ &\quad + \chi_{(0,-2,0)} + \chi_{(0,0,2)} + \chi_{(0,0,-2)} \\ &= \chi_{(2,0,0)} - \chi_{(1,1,0)} - \chi_{(0,0,0)}.\end{aligned}$$

Kronecker product calculations show that

$$(\chi_{(1,0,0)})^2 = \chi_{(2,0,0)} + \chi_{(1,1,0)} + \chi_{(0,0,0)}.$$

This results in

$$\chi_{(1,0,0)} \otimes \{2\} = \chi_{(2,0,0)}$$

and

$$\chi_{(1,0,0)} \otimes \{1^2\} = \chi_{(1,1,0)} + \chi_{(0,0,0)}.$$

ACKNOWLEDGMENT

This work has been supported by the Swedish National Science Research Council.

[1] R. V. Moody and J. Patera, Bull Am. Math. Soc. **7**, 237 (1982).
[2] R. V. Moody and J. Patera, SIAM J. Alg. Disc. Meth. **5**, 359 (1984).
[3] W. G. McKay and J. Patera, *Tables of Dimensions, Indices, and Branching Rules for Representations of Simple Lie Algebras* (Dekker, New York, 1981).
[4] W. G. McKay, R. V. Moody, and J. Patera, CMS Conf. Proc. **5**, 227 (1986).
[5] J. McKay, J. Patera, and R. T. Sharp, J. Math. Phys. **22**, 2770 (1981).
[6] S. Okubo, J. Math. Phys. **26**, 2127 (1985).
[7] E. B. Dynkin, Transl. Am. Math. Soc. Ser. 2 **6**, 111 (1957).
[8] A. Navon and J. Patera, J. Math. Phys. **8**, 489 (1967).
[9] B. G. Wybourne, Aust J. Phys. **32**, 417 (1979).
[10] G. Bélanger, J. Phys. A **16**, 3421 (1983).
[11] D. E. Littlewood, *The Theory of Group Characters* (Clarendon, Oxford, 1940).

Singular anharmonicities and the analytic continued fractions

M. Znojil
Institute of Nuclear Physics, Czechoslovak Academy of Sciences, 250 68 Řež near Prague, Czechoslovakia

(Received 9 September 1986; accepted for publication 13 July 1988)

The potential $V(r) = \mu^2 r^2 + \lambda^2 r^{-4}$ is investigated and its bound states are constructed by a generalized Hill-determinant method. It is shown that the binding energy E and another free parameter $\varkappa$ (in the underlying Laurent-series representation of the wave function) may be obtained from a coupled pair of Hill-determinant equations.

I. INTRODUCTION

The singularly anharmonic potential

$$V(r) = \mu^2 r^2 + \lambda^2 r^{-4} \tag{1.1}$$

is a superposition of the two forces such that in both the extreme cases $\mu = 0$ and $\lambda = 0$, the complete three-dimensional Schrödinger equation becomes solvable exactly.[1] Nevertheless, an infinitesimal transition to $\mu \neq 0$ and $\lambda \neq 0$ represents such a drastic change of the interaction that a perturbation theory fails to give any estimates: the coupling $\mu \neq 0$ introduces a confinement and $\lambda \neq 0$ also represents a singular perturbation. Methodically, such a situation simulates the difficulties encountered, e.g., in the contemporary field theory.[2]

Our present intention is to develop a consistent expansion and solution method. Technically, we are inspired by the paper of Singh *et al.*[3] where a sextic anharmonicity has been treated by the nonperturbative means, based on the use of analytic continued fractions. We shall arrive here at a similar "non-numerical" solution of the eigenvalue problem, which may be interpreted as a "natural" resummation of the divergent perturbative expansions.

Phenomenologically, the form of our potential (1.1) represents an independent motivation for an interest in the corresponding radial Schrödinger equation

$$\left(-\frac{d^2}{dr^2} + \frac{l(l+1)}{r^2} + \mu^2 r^2 + \frac{\lambda^2}{r^4} - E\right)\Psi(r) = 0, \quad l = 0,1,\dots . \tag{1.2}$$

Indeed, the harmonic component of the force is a common approximation verified in the various realistic situations. From a physical point of view, the effects of a finite size or correlations of the interacting objects are often important just at the short distances (cf. the phenomenological nucleon–nucleon interactions, etc.). Thus, in a purely phenomenological sense, a singular repulsion seems to be a very useful form of an anharmonicity.

II. NUMERICAL METHODS

A naive treatment of $\lambda^2 r^{-4}$ as a small perturbation leads immediately to contradictions: for $l = 0$, the first nontrivial correction to the energy (i.e., the matrix element) diverges as an integral in the origin,

$$\langle n|V|n\rangle \sim \int_\epsilon^\infty \langle r|n\rangle^2 r^{-4} r^2\,dr \sim \frac{1}{\epsilon}, \quad \epsilon \to 0.$$

A perturbative interpretation of μ also leads to difficulties: the $\mu = 0$ case does not possess the discrete spectrum and may be characterized by an abrupt disappearance of the confinement.

In a purely numerical framework, the differential equation (1.2) is singular in the origin: the initial boundary conditions at $r = 0$ are ambiguous and a nontrivial modification of the standard techniques[1,4] must be employed.

The more reliable approach may be expected to follow from the various types of the variational *Ansätze*. For a suitably modified standard basis (cf., e.g., Ref. 5), an *a priori* slow rate of convergence with respect to an increasing cutoff dimension may be accelerated by the various techniques (cf., e.g., Ref. 6). Alternatively, an application of the Lanczos algorithm[7] may permit one to choose an "optimal" initial guess of $|0\rangle$ and to construct the basis $\{|n\rangle\}$ and improve the results in a systematic way.

After a brief inspection of the differential equation (1.2), we may arrive at the estimates

$$\psi(r) \sim \exp(-\lambda r^{-1}), \quad r \approx 0, \tag{2.1}$$

and

$$\psi(r) \sim \exp(-\tfrac{1}{2}\mu r^2), \quad r \to \infty, \tag{2.2}$$

of the respective threshold and asymptotic physical behavior of the normalizable bound-state solutions. In this way, the "natural" Lanczos initial state $\langle r|0\rangle$ has the form

$$\langle r|0\rangle = r^{\varkappa} \exp(-\lambda r^{-1} - \tfrac{1}{2}\mu r^2), \tag{2.3}$$

where $\varkappa$ is some new free parameter.

The formalism of Lanczos is based on a subsequent orthogonalization of the states,[8] $|n\rangle \sim H|n-1\rangle + \cdots$. After an appropriate variational truncation, it leads to the approximate wave functions of the type $\Sigma_n \langle r|n\rangle \times c_n$, with

$$\langle r|n\rangle \sim r^{\text{a constant}} \times \langle r|0\rangle \times \text{a polynomial}. \tag{2.4}$$

This will be employed below as a methodical inspiration.

III. HILL DETERMINANTS AND THE ANALYTIC CONTINUED FRACTIONS

Mathematically, the main shortcoming of the variational estimates (2.4) lies in an unclear role played by the trial parameter $\varkappa$. Indeed, in the case of the regular potentials, this is not a free parameter ($\varkappa = l + 1$ specifies solutions regular in the origin). Here, we shall interpret Eq. (2.4) as a tentative *Ansatz* and analyze its consequences in a non-numerical manner.

 0022-2488/89/010023-05$02.50

For the sake of simplicity we shall omit the parity-violating factor $\exp(-\lambda r^{-1})$ and start from the formula

$$\psi(r) = \exp\left(-\frac{1}{2}\mu r^2\right) \sum_{n=-\infty}^{\infty} p_n r^{2n+\varkappa}. \tag{3.1}$$

Its insertion in the Schrödinger equation (1.2) converts this ordinary differential equation of second order into the relations

$$\begin{aligned} &A_{n+1} p_n + B_{n+1} p_{n+1} + C_{n+1} p_{n+2} = 0, \\ &A_{n+1} = (4n + 2\varkappa + 1)\mu - E, \\ &B_{n+1} = -(2n+\varkappa+1)(2n+\varkappa+2) + l(l+1), \\ &C_{n+1} = \lambda^2, \quad n = \ldots, -1,0,1,\ldots . \end{aligned} \tag{3.2}$$

They may be treated as recurrences or as a difference equation of the second order. Of course, our first problem lies now in a rigorous determination of the suitable initial or boundary conditions pertaining to Eq. (3.2).

In the first step of our considerations, an intuitive "variational" truncation of the doubly infinite system of equations (3.2) may be introduced,

$$\det \mathscr{H}_{-M,N} = 0,$$

$$\mathscr{H}_{-M,N} = \begin{pmatrix} B_{-M} & C_{-M} & & & \\ & \ddots & & & \\ & A_{k-M} & B_{k-M} & C_{k-M} & \\ & \ddots & & & \\ & & A_N & B_N & \end{pmatrix},$$

$$M,N \gg 1. \tag{3.3}$$

Later on (in Sec. IV) this equation will become a part of our rigorous construction. Here, let us start its analysis in the spirit of Ref. 3, i.e., postulating it as a source of the "quasivariational" energy approximants of an *a priori* unknown validity.

When we put

$$f_n = p_n / p_{n-1}, \quad N \geqslant n > 0, \tag{3.4}$$

and

$$g_m = p_{-m} / p_{1-m}, \quad M \geqslant m > 0, \tag{3.5}$$

in Eq. (3.2), we may interpret the quantities f_n and g_m as the finite continued-fractional approximants.[9] Indeed, the formal initial values $f_{N+1} = 0$ and $g_{M+1} = 0$ are to be employed in the corresponding recurrences (3.2), i.e.,

$$f_n = -A_n / (B_n + C_n f_{n+1}), \quad n = 1,2,\ldots,N, \quad N \to \infty, \tag{3.6}$$

and

$$\begin{aligned} &g_m = -C_{-m} / (B_{-m} + A_{-m} g_{m+1}), \\ &m = 1,2,\ldots,M, \quad M \to \infty. \end{aligned} \tag{3.7}$$

In this way, an infinite-dimensional limit $M,N \to \infty$ of the Hill–determinant condition (3.3) becomes equivalent to the continued-fractional condition

$$A_0 g_1 + B_0 + C_0 f_1 = 0. \tag{3.8}$$

The proof of this statement is easy—the triplet of Eqs. (3.6)–(3.8) is precisely equivalent to (is a mere transcription of) the original set (3.2). Moreover, we have also the following mathematical result at our disposal.

Lemma 1: Continued fractions f_n and g_m are convergent.

Proof: For $n \gg 1$, an almost n-independent mapping $f_{n+1} \to f_n$ is defined by Eq. (3.6),

$$\begin{aligned} f_n = &\mu / (n_0 - \lambda^2 f_{n+1} / (4n_0)) \\ &+ \text{corrections}, \quad n = O(n_0) \gg 1. \end{aligned} \tag{3.9}$$

In accord with Fig. 1, it has a simple geometric interpretation. The sequence $f_{n_0}, f_{n_0 - 1}, \ldots$ will accumulate near a stable fixed point $\varphi = \varphi(n)$ defined by the quadratic equation

$$\varphi(n) = \mu / (n - \lambda^2 \varphi(n)/(4n)), \quad n = O(n_0) \gg 1. \tag{3.10}$$

Thus, from an arbitrary initial value of $f_{n_1} = O(1)$ at some sufficiently large index $n_1 \gg n_0$, we get

$$f_n = \mu / n + \text{corrections}, \quad n = O(n_0) \gg 1. \tag{3.11}$$

Mutatis mutandis, the asymptotic form

$$\begin{aligned} g_m = &\lambda^2 / (4m_0^2 + 4\mu m_0 g_{m+1}) \\ &+ \text{corrections}, \quad m = O(m_0) \gg 1, \end{aligned} \tag{3.12}$$

of the "conjugate" mapping (3.7) leads to an analog of Eq. (3.11),

$$g_m = \lambda^2 / (4m^2) + \text{corrections}, \quad m = O(m_0) \gg 1. \tag{3.13}$$

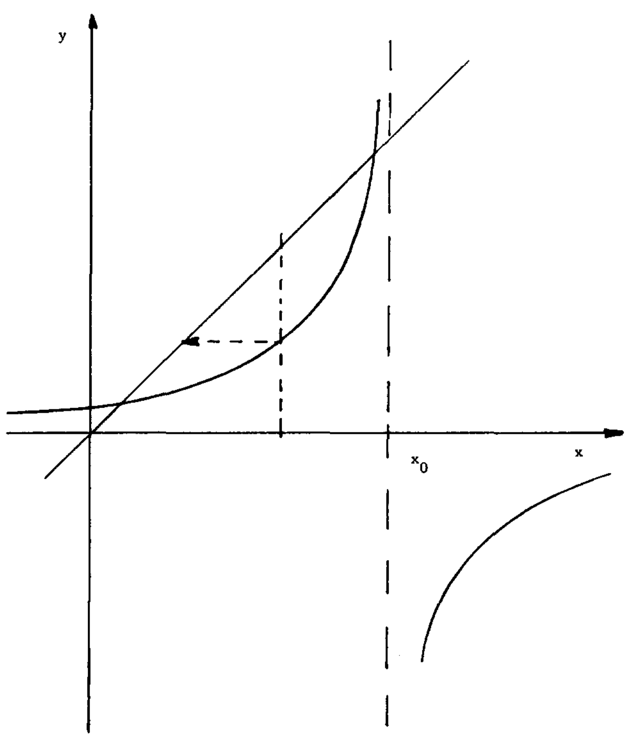

FIG. 1. Geometric proof of the continued-fractional convergence. We may read $y = f_n$ or g_m, $x = f_{n+1}$ or g_{m+1} and $x_0 = 4n^2/\lambda^2$ or m/μ, respectively.

Obviously, the unique values of f_n and g_m will be obtained in the infinite-dimensional limit $m,n \to \infty$. Q.E.D.

We may conclude that the continued-fractional condition (3.8) extends the Hill-determinant requirement (3.3) to the limit $M,N \to \infty$. In this sense, it represents a tentative analytic specification of energies, analogous to the sextic anharmonic-oscillator conjecture of Singh *et al.*[3] In the latter case, a partial proof[10] and partial disproof[11] of the conjecture became available during the further development. In the forthcoming sections, a similar reinterpretation and a rigorous background will also be given to the present "eigenvalue" condition (3.8).

IV. A RIGOROUS INTERPRETATION OF THE HILL-DETERMINANT ZEROS

A. The asymptotic behavior of the difference Schrödinger equation

Let us consider the Schrödinger equation (3.2) in the $|n| \gg 1$ asymptotic region where it acquires a particularly simple form,

$$\mu p_n - (n+1)p_{n+1} + (\lambda^2/4n)p_{n+2} = 0, \quad |n| \gg 1. \tag{4.1}$$

First, let us choose positive $n \gg 1$ and notice that an apparently "dominant" part of Eq. (4.1),

$$\mu \tilde{p}_n - (n+1)\tilde{p}_{n+1} = 0, \quad n \gg 1, \tag{4.2}$$

has an exact solution

$$\tilde{p}_n = (\mu^n/n!)p_0. \tag{4.3}$$

In the next step, we may change the variables. Denote

$$p_n = (\mu^n/n!)\, q_n^{(+)}, \quad n > 0, \tag{4.4}$$

and convert full Eq. (4.1) into an equivalent difference equation

$$q_{n+1}^{(+)} - q_n^{(+)} = [\lambda^2\mu/4n(n+1)(n+2)]\, q_{n+2}^{(+)}. \tag{4.5}$$

Here, we may try to decompose $q_n^{(+)}$ into a Taylor series

$$q_{n+k}^{(+)} = q_n^{(+)} + k\frac{d}{dn}q_n^{(+)} + \cdots,$$

and obtain an approximate differential equation in the lowest nontrivial order,

$$\frac{d}{dn}\tilde{q}_n^{(+)} = \frac{\lambda^2\mu}{4n^3}\tilde{q}_n^{(+)}. \tag{4.6}$$

Its solution

$$q_n^{(+)} \approx \tilde{q}_n^{(+)} = \exp(-\lambda^2\mu/8n^2), \quad n \gg 1, \tag{4.7}$$

improves the leading-order estimate (4.3) and shows the extremely weak n dependence of the product $p_n \cdot n!\mu^{-n}$ for large indices $n \gg 1$.

Let us now alternatively take Eq. (4.1) as a recurrent definition of $p_{n+2} = p_{n+2}(p_{n+1}, p_n)$. Obviously, its latter argument (the contribution of p_n) may also become negligible. In this setting, the relation

$$-(n+1)\hat{\hat{p}}_{n+1} + (\lambda^2/4n)\hat{\hat{p}}_{n+2} = 0, \quad n \gg 1, \tag{4.8}$$

is a counterpart to Eq. (4.2). Let us modify the change of variables (4.4),

$$p_n = \lambda^{-2n}4^n(n-1)!(n-2)!\hat{q}_n^{(+)}, \quad n > 1, \tag{4.9}$$

and replace Eq. (4.5) by a new reparametrization of Eq. (4.1),

$$\hat{q}_{n+2}^{(+)} - \hat{q}_{n+1}^{(+)} = -[\mu\lambda^2/4n(n^2-1)]\,\hat{q}_n^{(+)}. \tag{4.10}$$

The latter equation may again be analyzed in the same manner as before. We decompose $\hat{q}_n^{(+)}$ into a Taylor series and arrive at the second, independent asymptotic solution of our three equivalent difference Eqs. (4.1), (4.5), or (4.10), with

$$\hat{q}_n^{(+)} \approx \hat{\hat{q}}_n^{(+)} = \exp(\lambda^2\mu/8n^2), \quad n \gg 1. \tag{4.11}$$

Summarizing the whole procedure, we may write the general form of coefficients p_n $(n \gg 1)$ in the form of the superposition $c_1\tilde{p}_n + c_2\hat{\hat{p}}_n$,

$$p_n = c_1\frac{\mu^n}{n!}\exp\left(-\frac{\lambda^2\mu}{8n^2}\right) + c_2\frac{4^n(n-1)!(n-2)!}{\lambda^{2n}} \times\exp\left(+\frac{\lambda^2\mu}{8n^2}\right), \quad n \gg 1. \tag{4.12}$$

B. Convergence of the *Ansatz* $\psi(r)$

In the $n \ll -1$ asymptotic domain of indices, it is easy to repeat all the manipulations of Sec. IV A. Indeed, the modified pair of *Ansätze* (4.4) and (4.9),

$$p_{-m} = \frac{(-1)^m(m-1)!}{\mu^m}q_m^{(-)} = \frac{\lambda^{2m}}{4^m m!(m+1)!}\hat{q}_m^{(-)}, \quad m > 0, \tag{4.13}$$

leads to the two new (equivalent) forms of Eq. (4.1) with negative subscripts,

$$q_m^{(-)} - q_{m-1}^{(-)} = [\mu\lambda^2/4m(m-1)(m-2)]\,q_{m-2}^{(-)},$$
$$\hat{q}_{m-1}^{(-)} - \hat{q}_{m-2}^{(-)} = -[\mu\lambda^2/4m(m^2-1)]\,\hat{q}_m^{(-)}. \tag{4.14}$$

The negative-index analog of Eq. (4.12) will read

$$p_{-m} = c_3\frac{(-1)^m(m-1)!}{\mu^m}\exp\left(-\frac{\lambda^2\mu}{8m^2}\right) + c_4\frac{\lambda^{2m}}{4^m m!(m+1)!}\exp\left(+\frac{\lambda^2\mu}{8m^2}\right), \quad m \gg 1. \tag{4.15}$$

Thus we arrive at the following result.

Lemma 2: A necessary and sufficient condition of convergence of the doubly infinite Laurent series (3.1) may be formulated as a restriction $c_2 = c_3 = 0$ imposed on the asymptotics (4.12) and (4.15),

$$p_n \approx \begin{cases} \mu^n/n!, & n \gg 1, \\ (\lambda/2)^{2m}/m!(m+1)!, & m = -n \gg 1. \end{cases} \tag{4.16}$$

Proof: It is trivial and follows from an immediate application of the standard convergence criteria to (3.1), with

$$\frac{p_n}{p_{n-1}} = f_n = \begin{cases} O(n^2), & c_2 \neq 0, \\ \mu/n + \text{corrections}, & c_2 = 0, \end{cases}$$

$$n \gg 1, \tag{4.17}$$

and

$$\frac{p_n}{p_{n+1}} = g_m = \begin{cases} O(|n|), & c_3 \neq 0, \\ \lambda^2/4n^2 + \text{corrections}, & c_3 = 0, \end{cases}$$

$$n \ll -1. \tag{4.18}$$

This completes the proof.

An understanding of Eq. (4.16) is a key point of the present construction: Our formal solution (3.1) of the differential Eq. (1.2) exists if and only if the ambiguous solution of Eq. (3.2) is made unique by means of the "mathematical" boundary conditions (4.16).

We may compare (4.17), (4.18), (3.11), and (3.13) and see that the mathematical conditions are satisfied if and only if we pick up the specific, continued-fractional direction of recurrences (3.2) [rewritten as Eqs. (3.6)–(3.8)]. Thus we may conclude that the continued-fractional restriction (3.8) and prescriptions

$$p_n = \begin{cases} p_0 \prod_{k=1}^{n} f_k, & n > 0, \\ p_0 \prod_{k=1}^{m} g_k, & n = -m < 0, \end{cases} \tag{4.19}$$

define a convergent (not necessarily normalizable) solution $\psi(r)$ in (3.1) of our radial Schrödinger Eq. (1.2).

V. A PHYSICAL SPECIFICATION OF THE PAIR OF PARAMETERS $\varkappa$ AND E

We have seen that the Hill-determinant condition (3.3) or (3.8) is a necessary and sufficient condition of applicability (convergence) of the power-series expansions (3.1). The analytic continued-fractional form of this equation is quickly convergent [cf. Eqs. (3.11) and (3.13)] and compatible with a truncation of $\mathscr{H}_{-M,N}$ at rather small dimensions.

Now, on a background of our numerical tests, we conjecture that in a vicinity of the physical energies $E = E_n^{(\text{phys})}$, $n = 0,1,\ldots$, the continued-fractional equation (3.8) will normally have two roots $\varkappa = \varkappa_1(E)$ and $\varkappa = \varkappa_2(E)$ and that the two corresponding solutions $\psi^{(\varkappa)}(r)$ (3.1) of Eq. (1.2) will be linearly independent. For the particular choices of the couplings we have made, the roots $\varkappa_1$ and $\varkappa_2$ always formed a complex conjugate pair.

We may here recall that for $\mu = 0$, a similar rule also holds: the Hill-determinant conditon of convergence of the Laurent series (3.1) gives the imaginary pair of roots $\varkappa_{1,2}$.[4] In particular, for $r \gg 1$ and $\mu \doteq 0$ where wave functions closely resemble the freely traveling waves, a number of formal analogies with the Mathieu functions[4] still may be recovered in the present case: We omit any mathematical details of this type here.

In a vicinity of an exact binding energy $E_n^{(\text{phys})}$, the wave function acquires a new node (zero) in such a way that[12] the number m of these nodes $\tilde{r}_1, \tilde{r}_2, \ldots, \tilde{r}_m$ is equal to n for $E_{n-1}^{(\text{phys})} < E < E_n^{(\text{phys})}$ and $n = 0,1,\ldots$. In the regular-potential case, merely the limiting transition $r \to 0$ is simplified. Thus we may consider a general solution of Eq. (1.2),

$$\psi(r) = d_1 \psi^{(\varkappa_1)}(r) + d_2 \psi^{(\varkappa_2)}(r), \quad d_1^2 + d_2^2 = 1; \tag{5.1}$$

and (a) requiring that $\psi(r)$ in (5.1) satisfies the $r \to 0$ boundary conditions, specify the so-called regular solution[1]

$$\psi_R(r) = d_{1R} \psi^{(\varkappa_1)}(r) + d_{2R} \psi^{(\varkappa_2)}(r),$$

$$d_{iR} = \lim_{\tilde{r}_1 \to 0} d_i(\tilde{r}_1), \quad i = 1,2,$$

$$d_1(\tilde{r}_1) \psi^{(\varkappa_1)}(\tilde{r}_1) = -d_2(\tilde{r}_1) \psi^{(\varkappa_2)}(\tilde{r}_1); \tag{5.2}$$

(b) requiring that $\psi(r)$ in (5.1) satisfies the $r \to \infty$ boundary condition, specify another unphysical, so-called Jost solution[1]

$$\psi_J(r) = d_{1J} \psi^{(\varkappa_1)}(r) + d_{2J} \psi^{(\varkappa_2)}(r),$$

$$d_{iJ} = \lim_{\tilde{r}_m \to \infty} d_i(\tilde{r}_m), \quad i = 1,2; \tag{5.3}$$

and (c) matching the logarithmic derivatives of (5.2) and (5.3) at an arbitrary $r \in (0, \infty)$, we may determine $E = E^{(\text{phys})}$ and the physical solution $\psi^{(\text{phys})}(r) = \psi_R(r) = \psi_J(r)$.

As well as in the other cases solvable by means of the convergent power-series expansions, the matching condition is most easily formulated either at $r \approx \infty$ or at $r \approx 0$. In both cases, one of the solution [ψ_R or ψ_J, i.e., (5.2) or (5.3)] is needed and the respective relation

$$\psi_{R,J}(r) \approx 0 \tag{5.4}$$

may be interpreted as an appearance of a new node in $\psi(r)$. It is taking place slightly above the exact energy level, so that an efficient and reliable numerical algorithm results.

In both the regular and Jost cases, the physical boundary conditions may be understood as a simultaneous change of sign of $\psi(r)$ in (5.1) at a pair of points $r \approx 0$ and $r \approx \infty$. Then, the related conditions are

$$\begin{aligned} \tilde{d}_1 \psi^{(\varkappa_1)}(r_0) + \tilde{d}_2 \psi^{(\varkappa_2)}(r_0) &= 0, \quad r_0 \ll 1, \\ \tilde{d}_1 \psi^{(\varkappa_1)}(r_1) + \tilde{d}_2 \psi^{(\varkappa_2)}(r_1) &= 0, \quad r_1 \gg 1, \end{aligned} \tag{5.5}$$

and specify the unknowns. The mathematical Hill-determinant requirement (3.8) must be complemented by the physical, two-dimensional eigenvalue condition

$$\det\begin{pmatrix} \psi^{(\varkappa_1)}(r_0) & \psi^{(\varkappa_2)}(r_0) \\ \psi^{(\varkappa_1)}(r_1) & \psi^{(\varkappa_2)}(r_1) \end{pmatrix} = 0, \quad r_0 \ll 1, \quad r_1 \gg 1. \tag{5.6}$$

This is our final result.

VI. NUMERICAL TESTS

Our potential (1.1) is a smooth function of r and also the threshold and asymptotic behavior of the wave function is well known [cf. Eqs. (2.1) and (2.2), respectively]. As a consequence, the radial Schrödinger equation (1.2) may easily be solved, say, by the standard Runge–Kutta method.[13] A sample of the resulting energies may be found here in Table I. They are compatible also with the variational results of Ref. 5 where $E_0 = 4.031\,97$ has been obtained for the same force.

Even in the purely numerical context, our present con-

TABLE I. The first two "exact" energies as obtained by the Runge–Kutta method for $\mu^2 = 1$ and $\lambda^2 = 0.4$.

Matching points		Energies	
r_0	r_1	$E^{(0)}$	$E^{(1)}$
0.045	3.5	4.0321 9	8.342 757
0.040	4.0	4.0319 785	8.316 366 0
0.035	4.5	0.0319 7166	8.314 624 24
0.030	5.0	4.0319 7139	8.314 565 23

TABLE II. The energies and $\varkappa$'s obtained by the present method.

Matching points		Roots of Eq. (3.8)		Energies
r_0	r_1	Re $\varkappa_{1,2}$	Im $\varkappa_{1,2}$	E
0.045	3.5	0.5	$\pm$ 0.606 102	4.032 19
0.040	4.0	0.5	$\pm$ 0.606 083 7	4.031 978 4
0.035	4.5	0.5	$\pm$ 0.606 083 08	4.031 971 69
0.030	5.0	0.5	$\pm$ 0.606 083 10	4.031 971 34
0.045	3.5	0.5	$\pm$ 0.858 110	8.342 755
0.040	4.0	0.5	$\pm$ 0.857 011	8.316 365 4
0.035	4.5	0.5	$\pm$ 0.856 938 7	8.314 624 04
0.030	5.0	0.5	$\pm$ 0.856 936 31	8.314 565 38

struction of wave functions is "almost non-numerical" and may be understood as a specific generalization of the Mathieu functions[4] (where similar *Ansätze* are used). Of course, the determination of energies remains purely numerical in practice. Still, the purely analytic and everywhere convergent character of ψ's may remain useful. Without going into detail, we would like to point out only that an improvement of the matching conditions (5.5) (with derivatives) becomes extremely simple now (a more detailed analysis of this technicality will be described elsewhere[14]).

Methodically, an important feature of our present example may be seen in the presence of the two free parameters (E and $\varkappa$). This forces us to complement the single Hill-determinant condition (as used, e.g., in Ref. 3) by an additional equation [cf. (5.6)]. Of course, similar additional conditions may be expected to appear also in the more complicated systems. In a way, our present analysis indicates their possible treatment by the same techniques. Indeed, the computer implementation of our approach remains extremely easy: The continued fractions converge quickly. In fact, there is no need to use the acceleration of convergence via Eqs. (3.12) or (3.13). A full double-precision compatibility between the $M,N = 100$ and 800 results has been observed in our tests.

In Table II, a sample of results of the present method deviates from the "exact" results of Table I by the rounding errors only. This confirms that the implementation [and, in particular, a search for the complex roots of Eqs. (3.8)] of our prescription is easy and does not lead to any problems in principle. The tests confirm the good convergence properties derived above by purely algebraic means.

VII. DISCUSSION

At each energy value E, our Laurent-series *Ansatz* becomes convergent (solves the Schrödinger equation) for a pair of parameters $\varkappa_1$ and $\varkappa_2$ specified by the continued-fractional condition. The physical binding energies then become determined, in a more or less standard way, from the condition $\|\psi\| < \infty$.

Methodically, our superposition of the two exactly solvable potentials finds a surprisingly natural perturbative interpretation for small μ rather than for small λ. The "anharmonicity" λr^{-4} preserves its nonperturbative character even for small λ.

In the harmonic-oscillator $\lambda \to 0$ limit, and irrelevance of the $n < -1$ part of $\psi(r)$ $(g_1 \approx 0)$ gives $\varkappa(\varkappa - 1) \approx l(l+1)$, i.e., $\varkappa_1 \approx l + 1$ and $\varkappa_2 \approx -l$. In the same approximation, we then get $\tilde{d}_1 \approx 1$ and $\tilde{d}_2 \approx 0$, i.e., $\psi^{(\varkappa_1)}(r_0) \approx 0$ in our secular equation. This leads to the correct power-series termination requirement and harmonic oscillator spectrum $E \approx \mu(4N + 2\varkappa_1 + 1)$ as it should.

Our example clarifies an overall structure of the so-called Hill-determinant eigenvalue method (cf. Ref. 11 and references given therein). It may give both the physical and unphysical energies and solutions in general, depending on the structure of the underlying *Ansatz*. Our present analysis illustrates this and how the corresponding proofs may be based on an asymptotic solution of the related difference equations.

[1]R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1982).
[2]C. Itzykson and J. B. Zuber, *Quantum Field Theory* (McGraw-Hill, New York, 1980).
[3]V. Singh, S. N. Biswas, and K. Datta, Phys. Rev. D **18**, 1901 (1978).
[4]H. Bateman and A. Erdelyi, *Higher Transcendental Functions III* (McGraw-Hill, New York, 1955).
[5]M. Znojil, Phys. Lett. A **101**, 66 (1984).
[6]M. Znojil, J. Math. Phys. **25**, 2979 (1984).
[7]C. Lanczos, J. Res. Natl. Bur. Stand. **45**, 255 (1950).
[8]J. H. Wilkinson, *The Algebraic Eigenvalue Problem* (Clarendon, Oxford, 1965).
[9]H. S. Wall, *Analytic Theory of Continued Fractions* (Van Nostrand, New York, 1943).
[10]M. Znojil, Phys. Rev. D **26**, 3730 (1982).
[11]R. N. Chaudhuri, Phys. Rev. D **31**, 2687 (1985).
[12]E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956).
[13]W. E. Milne, *Numerical Solutions of Differential Equations* (Wiley, New York, 1953).
[14]M. Znojil, "The analytic continued fractions and bound states in the potential $\mu^2 r^2 + \lambda^2(1/r)^4$," Univ. Kaiserslautern preprint, 1988.

Universality and integrability of the nonlinear evolution PDE's describing *N*-wave interactions

F. Calogero
Dipartimento di Fisica, Università di Roma "La Sapienza," 00185 Roma, Italy and Istituto Nazionale di Fisica Nucleare, Sezione di Roma, Roma, Italy

(Received 17 June 1988; accepted for publication 17 August 1988)

The universality of the equations describing N-wave interactions is demonstrated by deriving them from a very large class of nonlinear evolution equations (essentially all those whose linear part is dispersive). Various forms of these equations are displayed. The fact that these "universal" nonlinear evolution equations obtain, by an appropriate asymptotic limit, from such a large class of nonlinear evolution equations, suggests that they should be integrable; since for this it is sufficient that the large class from which they are obtainable contain just one integrable equation. This expectation is validated in several cases, by deriving the equations from known integrable equations. In this manner an explanation may be provided of the (already known) integrable nature of certain equations; and new integrable equations may be obtained. Both S-integrable and C-integrable equations are discussed, namely both equations integrable via an appropriate spectral transform and solvable via an appropriate change of variables. In this paper the treatment is limited to equations in $1+1$ dimensions.

I. INTRODUCTION

The fact that certain nonlinear evolution PDE's of wide applicative relevance—such as, for instance, the nonlinear Schrödinger equation, the Burgers equation, the Korteweg–de Vries (KdV) equation, and some of their variants—are "integrable" (namely, endowed with an exceptionally simple mathematical structure; for a more detailed discussion of the meaning of integrability, see below) has appeared for a long time as a puzzling miracle; perhaps a confirmation of Galileo's intuition, that "Questo grandissimo libro che continuamente ci sta aperto innanzi agli occhi (io dico l'universo) ⋯ è scritto in lingua matematica" ["This great book that stands continuously open before our eyes (I mean the universe)...is written in mathematical language"].[1] Recently a less metaphysical explanation of this fact has been put forward.[2–4] It is based on the observation that the equations in question (and in particular, the nonlinear Schrödinger equation and some of its variants), have a "universal" character, inasmuch as they may be obtained from very large classes of nonlinear evolution equations by a procedure that is asymptotically exact in the limit of weak nonlinearity. Because this limiting procedure is, in many circumstances, just the appropriate one to evince weakly nonlinear effects, the universal model equations obtained in this manner show up in many, disparate, applicative contexts; they are widely applicable. Because this procedure, which amounts to an exact asymptotic limit, generally preserves integrability, these universal model equations are likely to be integrable; since for this to happen it is sufficient that the very large class of evolution equations from which they are obtainable contain just one integrable equation. Indeed, while the fact that an arbitrarily given equation turns out to be integrable should be considered an exceptional event, the fact that a very large class of equations contains at least one integrable specimen may be considered normal, i.e., by no means exceptional. Hence a universal model equation that is obtainable, via a limiting procedure, from (all!) the equations of a large class, is likely to be integrable, provided the limiting procedure preserves integrability (as it generally happens for a correct asymptotic limit). Let us moreover note that this argument may also be run backwards; if a universal model equation, obtainable via a limiting procedure that preserves integrability from all the equations of a large class, turns out *not* to be integrable, then *none* of the equations contained in the large class is integrable; hence this approach also yields *necessary conditions for integrability* of wide and straightforward applicability.[5]

In previous papers,[2–6] these ideas were developed for model equations of "nonlinear Schrödinger type," that obtain from the class of nonlinear evolution equations whose linear part is dispersive and whose linear part is, in some sense (see below), analytic. These universal model equations of nonlinear Schrödinger type emerge naturally from the investigation of a solution, of any nonlinear equation of this class, that is "small" (so that nonlinear effects are "weak") and is "close" to a solution of the *linear* part of the equation representing a *single dispersive wave*. If nonlinear effects were completely neglected, the amplitude of such a solution would be constant (independent of space and time). To evince the effects due to the (weak) nonlinearity, it is convenient to follow the system with the group velocity characteristic of the ("carrier") dispersive wave under consideration, and to introduce appropriate "coarse-grained" and "slow" variables to account for the space and time variation of the amplitude; it is then found that, in these variables, the amplitude generally evolves according to an equation belonging to a (small) group of universal evolution PDE's of nonlinear Schrödinger type.[2–4]

The purpose and scope of this paper is to discuss an analogous approach, in which one takes as point of departure the same class of nonlinear evolution equations (with dispersive linear part and "analytic" nonlinear part), but focuses on a solution that, while being small (just as in the previous case; so that one is again considering a regime of

 0022-2488/89/010028-13$02.50

weak nonlinearity), is close to a superposition of *several different dispersive waves* (solutions of the linear part of the equation) having *different* group velocities (the case with different dispersive waves having the *same* group velocity is instead analogous to that discussed above, since in such a case it is still possible to analyze the problem in a reference frame that moves with the common group velocity, thereby obtaining again equations of nonlinear Schrödinger type).[6] As we show below, in such a case the equations that emerge naturally, to account for the evolution, in appropriately coarse-grained and slow variables, of the N amplitudes of the N dispersive (carrier) waves are (of course) just the standard equations describing N-wave interactions. Typical examples of these equations, as we show below, are the "nonresonant N-wave interaction"

$$\left(\frac{\partial}{\partial\tau}+v_j\frac{\partial}{\partial\xi}\right)\Psi_j(\xi,\tau)$$
$$=i\Psi_j(\xi,\tau)\sum_{l=1}^{N}\alpha_{jl}|\Psi_l(\xi,\tau)|^2,\quad j=1,2,...,N, \qquad (1.1)$$

and the "resonant three-wave interaction"

$$\left(\frac{\partial}{\partial\tau}+v_j\frac{\partial}{\partial\xi}\right)\Psi_j(\xi,\tau)$$
$$=\alpha_j[\Psi_{j+1}(\xi,\tau)]^*[\Psi_{j+2}(\xi,\tau)]^*,$$
$$j=1,2,3,\quad j\equiv j+3. \qquad (1.2)$$

Other instances are reported below; still others can be easily obtained once the technique that yields them is understood. In this paper we focus mainly on the introduction of this technique, rather than on an exhaustive treatment of all nonlinear equations of N-wave interaction type obtainable in this manner.

As implied by the preceding discussion, it is justified to expect that the model equations obtained in this manner be both *widely applicable and integrable*. The first expectation is of course fulfilled by the emergence of the equations describing N-wave interactions, whose applicative relevance is well known. The second expectation is also fulfilled; and in this connection it is useful to recall the heuristic concepts of "C-integrability," i.e., integrability by an appropriate *change of variables* (allowing generally to construct explicit nontrivial solutions and even to solve the Cauchy problem, just by quadratures), and "S-integrability," i.e., integrability via the *spectral transform* (or *inverse scattering*) technique (see, for instance, Ref. 7). For instance, as we show below, the nonresonant N-wave interaction (1.1) is C-integrable if the imaginary part of the constant matrix α_{jl} appearing in its right-hand side (rhs) is either diagonal or proportional to the difference v_j-v_l [see (A2),(A3a),(A3b) below]; while the S-integrability of (2.1), provided the constants α_j have the same phase [mod(π); in which case by rescaling the dependent variables one can replace them by just signs], is also well known; and is indeed "explained" by the results given below, since in Sec. IV we obtain (1.2), with all three constants α_j purely imaginary, from an S-integrable equation.

II. THE ASYMPTOTIC EXPANSION

Our starting point is the general nonlinear evolution PDE

$$Du=F[u]. \qquad (2.1)$$

We assume the dependent variable $u\equiv u(x,t)$ to be real. The left-hand side of (2.1) is the *linear* part of this equation, which is assumed to be *dispersive*. For definiteness, we hereafter assume that the linear differential operator D has one of the following two forms:

$$D=\frac{\partial}{\partial t}+\sum_{l=0}^{L}(-)^l a_l\frac{\partial^{2l+1}}{\partial x^{2l+1}}\quad\text{(odd case)}, \qquad (2.2a)$$

$$D=\frac{\partial^2}{\partial t^2}+\sum_{l=0}^{L}(-)^l b_l\frac{\partial^{2l}}{\partial x^{2l}}\quad\text{(even case)}. \qquad (2.2b)$$

The quantities a_l and b_l are *real* constants. As is clear from the following, there would be no difficulty to treat more general cases (with higher t derivatives, mixed x and t derivatives, or integral operators).

The linear part of (2.1), namely the equation

$$Du=0, \qquad (2.3)$$

admits as a solution the *dispersive wave*

$$u(x,t)=a\exp\{i[kx-\omega(k)t]\}+\text{c.c.}, \qquad (2.4)$$

where a is a (generally complex) constant and

$$\omega(k)=\sum_{l=0}^{L}a_l k^{2l+1}\quad\textit{(odd case)}, \qquad (2.5a)$$

$$\omega^2(k)=\sum_{l=0}^{L}b_l k^{2l}\quad\textit{(even case)}. \qquad (2.5b)$$

Hereafter we consider *real* values of k; the corresponding $\omega(k)$ is also generally *real* in the *odd case*, see (2.5a); and we limit our consideration in the following to values of the parameters b_l and k such that $\omega(k)$ is also *real*, whenever we treat the *even case* [see (2.5b)]. It is of course just the reality of k and $\omega(k)$ that characterizes (2.4) as a *dispersive wave*. Let us also recall that, to the dispersive wave (2.4), is associated the *group velocity*

$$v(k)=\frac{d\omega(k)}{dk}, \qquad (2.6)$$

namely

$$v(k)=\sum_{l=0}^{L}(2l+1)a_l k^{2l}\quad\textit{(odd case)}, \qquad (2.6')$$

$$v(k)=\sum_{l=1}^{L}lb_l\frac{k^{2l-1}}{\omega(k)}\quad\textit{(even case)}. \qquad (2.6'')$$

The right-hand side of (2.1) represents the *nonlinear* part of this PDE; $F[u]$ indicates an (assumedly given) nonlinear function of $u(x,t)$ and its derivatives. For reasons that will be apparent below we also introduce the notation

$$F[u]=\frac{\partial^q\widetilde{F}[u]}{\partial x^q}, \qquad (2.7a)$$

to facilitate the treatment of the cases when the rhs of (2.1) is a derivative of order q (hence q in the following is a nonnegative integer, whose value can be conveniently adjusted to treat interesting examples). We moreover assume $\widetilde{F}[u]$ to be *analytic*, in the following sense: for small ϵ,

$$\widetilde{F}[\epsilon u]=\sum_{m=2}^{M}\epsilon^m\widetilde{F}^{(m)}[u]+o(\epsilon^M), \qquad (2.7b)$$

where $\widetilde{F}^{(m)}[u]$ is a *homogeneous polynomial of degree m* in

$u(x,t)$ and its derivatives, and M is a (small) positive integer (whose minimal value for the validity of the following results will be obvious in each case; generally $M=2$ or $M=3$ will do, see below). Note that the sum in the right-hand side of this equation starts from $m=2$; this reflects the *nonlinear* character of the right-hand side of (2.1).

For definiteness we write

$$\tilde{F}^{(m)}[u] = \sum_{l_1=0}\sum_{l_2=l_1}\cdots\sum_{l_m=l_{m-1}} c^{(m)}_{l_1 l_2\cdots l_m} u^{(l_1)}u^{(l_2)}u^{(l_m)}. \tag{2.7c}$$

Here and below we use the notation

$$u^{(l)} \equiv \frac{d^l u(x,t)}{dx^l}. \tag{2.8}$$

Clearly this notation identifies uniquely the *real* constant $c^{(m)}_{l_1 l_2\cdots l_m}$ as the coefficient of the monomial $u^{(l_1)}u^{(l_2)}\cdots u^{(l_m)}$ (with $l_1 \leqslant l_2 \leqslant \cdots \leqslant l_m$).

The absence of time derivatives in the right-hand side of (2.1) that is implied by this notation [see (2.7a)–(2.7c)] is merely to simplify the *notation*; as is clear from the following developments, their eventual inclusion in the treatment would present no difficulty.

To illustrate this notation, let us display some examples, which are also useful for future reference (see below). The equation

$$u_t - a_1 u_{xxx} + a_2 u_{xxxxx} = -6a_1 u u_x + 10 a_2 (u u_{xxx} + 2u_x u_{xx} - 3u^2 u_x), \tag{2.9a}$$

$$u_t - a_1 u_{xxx} + a_2 u_{xxxxx} = [-3a_1 u^2 + 5a_2 (u_x{}^2 + 2u u_{xx} - 2u^3)]_x, \tag{2.9b}$$

corresponds to (2.1) with (2.2a) and (2.7a)–(2.7c), with

$$L=2,\quad a_0=0,\quad q=1,\quad M=3,\quad c^{(2)}_{00}=-3a_1, \\ c^{(2)}_{11}=5a_2,\quad c^{(2)}_{02}=10a_2,\quad c^{(3)}_{000}=-10a_2, \tag{2.9c}$$

and all other constants vanishing; the equation

$$u_t - a_1 u_{xxx} + a_2 u_{xxxxx} = 2s\{a_1 u^3 - a_2[5(u^2 u_{xx} + u u_x{}^2) - 3s u^5]\}_x, \\ s=\pm, \tag{2.10a}$$

corresponds to (2.1) with (2.2a) and (2.7a)–(2.7c), with

$$L=2,\quad a_0=0,\quad q=1,\quad c^{(3)}_{000}=2sa_1,\quad c^{(3)}_{002}=10sa_2, \\ c^{(3)}_{011}=10sa_2,\quad c^{(5)}_{00000}=-6a_2,\quad s=\pm, \tag{2.10b}$$

and all other constants vanishing; the equation

$$u_{tt} - u_{xx} + s\sin u = 0,\quad s=\pm, \tag{2.11a}$$

corresponds to (2.1) with (2.2b) and (2.7a)–(2.7c), with

$$L=1,\quad b_0=s,\quad b_1=1,\quad q=0,\quad M=\infty, \\ c^{(2l+1)}_{00\cdots 0}=(-)^{l+1}s/(2l+1)!,\quad l=1,2,\ldots, \tag{2.11b}$$

and all other constants vanishing; the equations

$$u_{tt} - u_{xx} - u_{xxxx} = 2(u u_{xx} + u_x{}^2), \tag{2.12a}$$

$$u_{tt} - u_{xx} - u_{xxxx} = (u^2)_{xx} \tag{2.12b}$$

corresponds to (2.1) with (2.2b) and (2.7a)–(2.7c), with

$$L=2,\quad b_0=0,\quad b_1=1,\quad b_2=-1, \\ M=2,\quad q=2,\quad c^{(2)}_{00}=1, \tag{2.12c}$$

and all other contants vanishing; the equation

$$u_t - u_{xxx} = (3u u_x + u^3)_x \tag{2.13a}$$

corresponds to (2.1) with (2.2a) and (2.7a)–(2.7c), with

$$L=1,\quad a_0=0,\quad a_1=1,\quad q=1, \\ c^{(2)}_{01}=3,\quad c^{(3)}_{000}=1, \tag{2.13b}$$

and all other constants vanishing; the equation

$$u_t - u_{xxx} = 3u^2 u_{xx} + 9u u_x{}^2 + 3u^4 u_x \tag{2.14a}$$

corresponds to (2.1) with (2.2a) and (2.7a)–(2.7c), with

$$L=1,\quad a_0=0,\quad a_1=1,\quad M=5,\quad q=0, \\ c^{(3)}_{002}=3,\quad c^{(3)}_{011}=9,\quad c^{(5)}_{00001}=3, \tag{2.14b}$$

and all other constants vanishing. Of course in these cases (with the indicated choices of M) the formula (2.7b) becomes exact; namely, the term $o(\epsilon^M)$ in its right-hand side need not be present. Note that the first four of these equations are S integrable; the first, (2.9), is the second nonlinear PDE of the KdV hierarchy; the second, (2.10), is the second nonlinear PDE of the mKdV hierarchy; the third, (2.11), is the sine–Gordon equation; the fourth, (2.12), is the Boussinesq equation (see, for instance, Ref. 7, and the literature quoted there). As for the last two, (2.13) and (2.14), they are C-integrable.[4]

In the solution (2.4) of the *linear* equation (2.3) the amplitude a is constant (t and x independent). We now consider solutions of the *nonlinear* equation (2.1) that are *small* (of order ϵ) and that are *close* (in the limit of small ϵ) to the solution (2.4), or rather to *a superposition of N dispersive waves*, i.e., N solutions of type (2.4) of the linear equation (2.3) characterized by different values of the parameter k. The main effect of the (weak) nonlinearity is then to induce a (slow) variation of the amplitudes of these dispersive waves; our task below is to obtain the nonlinear PDE's that describe, in appropriate slow and coarse-grained variables, such evolution. Let us emphasize that the derivation of these evolution equations from the original nonlinear evolution equation (2.1) is *exact* (in an asymptotic sense, as the parameter ϵ that controls the weakness of the nonlinearity, vanishes); and we shall find that large classes of nonlinear evolution equations of type (2.1) yield, in this asymptotic limit, the *same* equation. Hence the evolution equations obtained in this manner have a universal character, that justifies the expectation that they be both *widely applicable and integrable* (for the reasons already mentioned above). Let us emphasize here that the *universality* of these limit equations may be considered a natural consequence of the *slow* and *coarse-grained* character of their (independent) variables, which imply that many specific details characterizing the original equations (2.1) get smoothed away [for instance, as we shall see below, most of the differential operators present in the original equations (2.1) get replaced by multiplicative constants in the limit equations].

To obtain these results, it is expedient to introduce the asymptotic/Fourier expansion

$$u(x,t) = \epsilon \sum_{\underline{n}=-\infty}^{\infty} \exp\left(i \sum_{j=1}^{N} n_j z_j\right) \epsilon^{r_{\underline{n}}} \varphi_{\underline{n}}(\xi,\tau), \tag{2.15a}$$

where

$$z_j \equiv k_j x - \omega(k_j)t, \tag{2.15b}$$

$$\xi = \epsilon^p x, \quad \tau = \epsilon^p t, \quad p>0, \tag{2.15c}$$

and the index $\underline{n}$ stands for the set $\{n_j; j = 1,2,...,N\}$. We moreover assume that there hold the conditions

$$r_{\underline{n}} = r_{-\underline{n}}, \tag{2.15d}$$

$$\varphi_{\underline{n}}(\xi,\tau) = [\varphi_{-\underline{n}}(\xi,\tau)]^*, \tag{2.15e}$$

where of course the index $-\underline{n}$ stands for the set $\{-n_j; j = 1,2,...,N\}$. These conditions are clearly necessary and sufficient to guarantee the reality of $u(x,t)$.

Consistently with our approach we moreover generally set

$$r_{\underline{n}} = r \geqslant 0, \quad \text{if } n_j = 0, \quad j = 1,2,...,N, \tag{2.16a}$$

$$r_{\underline{n}} = -1 + \sum_{j=1}^{N} |n_j|, \quad \text{otherwise.} \tag{2.16b}$$

Note that this implies that $r_{\underline{n}}$ vanishes if one of the indices n_j has unit modulus and all the others vanish. For notational convenience we set, in this case,

$$\varphi_{\underline{n}}(\xi,\tau) \equiv \Psi_j(\xi,\tau), \quad \text{if } n_j = 1 \text{ and } n_{j'} = 0 \text{ for } j' \neq j, \tag{2.17a}$$

and we also set

$$\varphi_{\underline{n}}(\xi,\tau) = \Psi_0(\xi,\tau), \quad \text{if } n_j = 0, \quad j = 1,2,...,N. \tag{2.17b}$$

Hence the *ansatz* (2.13) implies

$$\begin{aligned} u(x,t) &= \epsilon \sum_{j=1}^{N} [\exp(iz_j)\Psi_j(\xi,\tau) + \text{c.c.}] + \epsilon^{r+1}\Psi_0(\xi,\tau) \\ &\quad + \epsilon^2 \sum_{j,j'=1}^{N} \{\exp[i(z_j + z_{j'})]\chi_{jj'}(\xi,\tau) + \text{c.c.}\} \\ &\quad + \epsilon^2 \sum_{j,j'=1,j'\neq j}^{N} \{\exp[i(z_j - z_{j'})] \\ &\quad \times \tilde{\chi}_{jj'}(\xi,\tau) + \text{c.c.}\} + O(\epsilon^3), \end{aligned} \tag{2.18}$$

where, for notational convenience, we have also set

$$\varphi_{\underline{n}}(\xi,\tau) = \chi_{jj'}(\xi,\tau), \quad \text{if } n_j = n_{j'} = 1 \text{ and } n_{j''} = 0 \text{ for } j'' \neq j,j', \tag{2.19a}$$

$$\varphi_{\underline{n}}(\xi,\tau) = \tilde{\chi}_{jj'}(\xi,\tau), \quad \text{if } n_j = 1, \ n_{j'} = -1, \text{ and } n_{j''} = 0 \text{ for } j'' \neq j,j', \ j \neq j'. \tag{2.19b}$$

It is easily seen that the ansatz (2.15) provides an asymptotic expansion, applicable for small ϵ, which is generally consistent with the nonlinear evolution equation (2.1) with (2.7) and (2.8). The significance of this expansion is best understood by looking at the more explicit formula (2.18). It is then evident that the leading terms are the N dispersive waves $\epsilon[\Psi_j(\xi,\tau)\exp(iz_j) + \text{c.c.}]$ [characterized by N real parameters k_j; see (2.15b)], whose (*complex*) amplitudes $\Psi_j(\xi,\tau)$ are functions of the variables ξ and τ, which are coarse grained and slow on account of (2.15c). The actual value of the *positive* exponent p in (2.15c), which sets the degree of coarse grainedness and slowness, will be determined in each case (see below), to take properly into account the effect of the weak nonlinearity; note, however, that, in contrast to the treatments that lead to equations of nonlinear Schrödinger type,[1–4] we are assuming here the *same* rescaling for the space and time variables [see (2.15c)]. Also note that we reserve for the moment to set the value of the exponent r [see (2.16a)].

Our strategy is to insert the ansatz (2.15) [or, equivalently, (2.18)] in the nonlinear evolution equation (2.1) [with (2.2) and (2.7)], and to obtain (after having made appropriate choices for the number N of dispersive waves, for their parameters, and for the exponents p and r) nonlinear evolution equations for the amplitudes $\Psi_j(\xi,\tau)$, that are *exactly valid* in the asymptotic limit of vanishing ϵ.

The first task to obtain such equations is to treat the linear part of (2.1). This can be done quite generally by noting that the ansatz (2.15) implies

$$Du(x,t) = \epsilon \sum_{\underline{n}=-\infty}^{+\infty} \exp\left(i\sum_{j=1}^{N} n_j z_j\right) \epsilon^{r_{\underline{n}}} D_{\underline{n}} \varphi_{\underline{n}}(\xi,\tau), \tag{2.20}$$

with

$$\begin{aligned} D_{\underline{n}} &= -i \sum_{j=1}^{N} n_j \omega(k_j) + \epsilon^p \frac{\partial}{\partial \tau} \\ &\quad + \sum_{l=0}^{L} (-)^l a_l \left(i \sum_{j=1}^{N} n_j k_j + \epsilon^p \frac{\partial}{\partial \xi}\right)^{2l+1} \\ &\quad (\textit{odd case}), \end{aligned} \tag{2.21a}$$

$$\begin{aligned} D_{\underline{n}} &= \left(-i \sum_{j=1}^{N} n_j \omega(k_j) + \epsilon^p \frac{\partial}{\partial \tau}\right)^2 \\ &\quad + \sum_{l=0}^{L} (-)^l b_l \left(i \sum_{j=1}^{N} n_j k_j + \epsilon^p \frac{\partial}{\partial \xi}\right)^{2l} \\ &\quad (\textit{even case}). \end{aligned} \tag{2.21b}$$

There thus obtain, in the *odd case*, the equations

$$\left(\frac{\partial}{\partial \tau} + v_j \frac{\partial}{\partial \xi}\right)\Psi_j(\xi,\tau) + O(\epsilon^p) = \epsilon^{1-p}\left[\sum_{m=2}^{M} \epsilon^{m-2} f_j^{(m)} + o(\epsilon^{M-2})\right], \tag{2.22}$$

$$\left(\frac{\partial}{\partial \tau} + a_0 \frac{\partial}{\partial \xi}\right)\Psi_0(\xi,\tau) + O(\epsilon^{2p}) = \epsilon^{1-r-p+qp}\left[\sum_{m=2}^{M} \epsilon^{m-2} f_0^{(m)} + o(\epsilon^{M-2})\right], \tag{2.23}$$

$$iA_{jj'}\chi_{jj'}(\xi,\tau) + O(\epsilon^p) = \sum_{m=2}^{M} \epsilon^{m-2} g_{jj'}^{(m)} + o(\epsilon^{M-2}), \tag{2.24a}$$

$$i\tilde{A}_{jj'}\tilde{\chi}_{jj'}(\xi,\tau) + O(\epsilon^p) = \sum_{m=2}^{M} \epsilon^{m-2} \tilde{g}_{jj'}^{(m)} + o(\epsilon^{M-2}), \quad j \neq j', \tag{2.24b}$$

where $v_j \equiv v(k_j)$ is the group velocity [see (2.6a)] and

$$\begin{aligned} A_{jj'} &= \sum_{l=1}^{L} a_l [(k_j + k_{j'})^{2l+1} - k_j^{2l+1} - k_{j'}^{2l+1}] \\ &= \omega(k_j + k_{j'}) - \omega(k_j) - \omega(k_{j'}), \end{aligned} \tag{2.25a}$$

$$\widetilde{A}_{jj'} = \sum_{l=1}^{L} a_l[(k_j - k_{j'})^{2l+1} - k_j^{2l+1} + k_{j'}^{2l+1}]$$
$$= \omega(k_j - k_{j'}) - \omega(k_j) + \omega(k_{j'}), \quad j \neq j'. \tag{2.25b}$$

As for the terms in the right-hand side, their significance is self-evident: $f_j^{(m)}$ is the coefficient of $\exp(iz_j)$ in $\epsilon^{-1}F^{(m)}[u]$, when the ansatz (2.18) is inserted in (2.7); $f_0^{(m)}$ is, likewise, the coefficient of the term without any exponential; $g_{jj'}^{(m)}$ and $\tilde{g}_{jj'}^{(m)}$ are, respectively, the coefficients of $\exp[i(z_j + z_{j'})]$ and $\exp[i(z_j - z_{j'})]$. The evaluation of these quantities is deferred to the following sections.

The analogous formulas for the *even case* read as follows:

$$-2i\omega_j\left(\frac{\partial}{\partial\tau} + v_j\frac{\partial}{\partial\xi}\right)\Psi_j(\xi,\tau) + O(\epsilon^p)$$
$$= \epsilon^{1-p}\left[\sum_{m=2}^{M} \epsilon^{m-2} f_j^{(m)} + o(\epsilon^{M-2})\right], \tag{2.26}$$

$$\left[b_0 + \epsilon^{2p}\left(\frac{\partial^2}{\partial\tau^2} - b_1\frac{\partial^2}{\partial\xi^2}\right)\right]\Psi_0(\xi,\tau) + O(\epsilon^{4p})$$
$$= \epsilon^{1-r+qp}\left[\sum_{m=2}^{M} \epsilon^{m-2} f_0^{(m)} + o(\epsilon^{M-2})\right], \tag{2.27}$$

$$B_{jj'}\chi_{jj'}(\xi,\tau) + O(\epsilon^p) = \sum_{m=2}^{M} \epsilon^{m-2} g_{jj'}^{(m)} + o(\epsilon^{M-2}), \tag{2.28a}$$

$$\widetilde{B}_{jj'}\widetilde{\chi}_{jj'}(\xi,\tau) + O(\epsilon^p)$$
$$= \sum_{m=2}^{M} \epsilon^{m-2}\tilde{g}_{jj'}^{(m)} + o(\epsilon^{M-2}), \quad j \neq j', \tag{2.28b}$$

where of course $\omega_j \equiv \omega(k_j)$ and $v_j \equiv v(k_j)$, see (2.5b) and (2.6b), and

$$B_{jj'} = \sum_{l=0}^{L} b_l[(k_j + k_{j'})^{2l} - k_j^{2l} - k_{j'}^{2l}] - 2\omega_j\omega_{j'}$$
$$= [\omega(k_j + k_{j'})]^2 - [\omega(k_j) + \omega(k_{j'})]^2, \tag{2.29a}$$

$$\widetilde{B}_{jj'} = \sum_{l=0}^{L} b_l[(k_j - k_{j'})^{2l} - k_j^{2l} - k_{j'}^{2l}] + 2\omega_j\omega_{j'}$$
$$= [\omega(k_j - k_{j'})]^2 - [\omega(k_j) - \omega(k_{j'})]^2, \quad j \neq j'; \tag{2.29b}$$

while the quantities in the right-hand sides are defined as in the previous case.

III. THE NONRESONANT CASE

Let us now compute the nonlinear contributions [see the right-hand sides of (2.22)–(2.24) and (2.26)–(2.28)] in the *nonresonant case*, namely under the assumption that there exist no set of integer values $\underline{n}$ (not all vanishing, of course!) such that *both* of the following equalities hold:

$$\sum_{j=1}^{N} n_j k_j = 0, \tag{3.1a}$$

$$\sum_{j=1}^{N} n_j \omega(k_j) = 0. \tag{3.1b}$$

In fact, it is sufficient for the validity of the following results that this (negative!) condition be satisfied for any choice of $\underline{n}$ such that none of the integers n_j exceed 2 in modulus (see below). Of course, this condition can always be satisfied, for any N, since we are free to choose the values of the parameters k_j.

It is clear, from (2.18) and (2.7), that the following relations then hold:

$$f_j^{(2)} = (ik_j)^q[\epsilon^r\gamma_{jj}(0;1)\Psi_0\Psi_j]$$
$$+ \epsilon\sum_{j'=1}^{N} \gamma_{jj'}(1,1;-1)\chi_{jj'}\Psi_{j'}^*$$
$$+ \epsilon\sum_{j'=1, j'\neq j}^{N} \gamma_{jj'}(1,-1;1)\widetilde{\chi}_{jj'}\Psi_{j'}$$
$$+ O(\epsilon^{r+p},\epsilon^{1+p},\epsilon^3), \tag{3.2a}$$

$$f_j^{(3)} = (ik_j)^q\Psi_j\left[\sum_{j'=1}^{N} \gamma_{jj'j'}(1;1;-1)|\Psi_{j'}|^2\right.$$
$$\left. + \epsilon^{2r}\gamma_{jjj}(0;0;1)\Psi_0^2\right] + O(\epsilon^p,\epsilon^{r+1},\epsilon^2), \tag{3.2b}$$

$$f_0^{(2)} = \left(\frac{\partial}{\partial\xi}\right)^q\left[\sum_{j=1}^{N} \gamma_{jj}(1;-1)|\Psi_j|^2\right.$$
$$\left. + \epsilon^{2r}c_{00}^{(2)}\Psi_0^2\right] + O(\epsilon^p,\epsilon^2), \tag{3.3a}$$

$$f_0^{(3)} = \epsilon^r\left(\frac{\partial}{\partial\xi}\right)^q\left[\Psi_0\sum_{j=1}^{N} \gamma_{jjj}(0;1;-1)|\Psi_j|^2\right.$$
$$\left. + \epsilon^{2r}c_{000}^{(3)}\Psi_0^3\right] + O(\epsilon^{r+p},\epsilon), \tag{3.3b}$$

$$g_{jj'}^{(2)} = [i(k_j + k_{j'})]^q\gamma_{jj'}(1;1)\Psi_j\Psi_{j'} + O(\epsilon^{1+r},\epsilon^2,\epsilon^p), \tag{3.4a}$$

$$g_{jj'}^{(3)} = O(\epsilon^r,\epsilon), \tag{3.4b}$$

$$\tilde{g}_{jj'}^{(2)} = [i(k_j - k_{j'})]^q\gamma_{jj'}(1;-1)\Psi_j\Psi_{j'}^*$$
$$+ O(\epsilon^{1+r},\epsilon^2,\epsilon^p), \quad j \neq j', \tag{3.5a}$$

$$\tilde{g}_{jj'}^{(3)} = O(\epsilon^r,\epsilon), \quad j \neq j'. \tag{3.5b}$$

Here and below

$$\gamma_{jj'}(n;n') = \sum_{l_1=0}\sum_{l_2=l_1} i^{l_1+l_2}c_{l_1l_2}^{(2)}$$
$$\times[(nk_j)^{l_1}(n'k_{j'})^{l_2} + (nk_j)^{l_2}(n'k_{j'})^{l_1}], \tag{3.6a}$$

$$\gamma_{jj'}(n_1,n_2;n') = \sum_{l_1=0}\sum_{l_2=l_1} i^{l_1+l_2}c_{l_1l_2}^{(2)}[(n_1k_j + n_2k_{j'})^{l_1}$$
$$\times(n'k_{j'})^{l_2} + (n_1k_j + n_2k_{j'})^{l_2}(n'k_{j'})^{l_1}], \tag{3.6b}$$

$$\gamma_{jj'j''}(n;n';n'')$$
$$= \sum_{l_1=0}\sum_{l_2=l_1}\sum_{l_3=l_2} i^{l_1+l_2+l_3}c_{l_1l_2l_3}^{(3)}[(nk_j)^{l_1}(n'k_{j'})^{l_2}$$
$$\times(n''k_{j''})^{l_3} + (nk_j)^{l_1}(n'k_{j'})^{l_3}(n''k_{j''})^{l_2}$$
$$+ (nk_j)^{l_2}(n'k_{j'})^{l_1}(n''k_{j''})^{l_3} + (nk_j)^{l_2}$$
$$\times(n'k_{j'})^{l_3}(n''k_{j''})^{l_1} + (nk_j)^{l_3}(n'k_{j'})^{l_1}(n''k_{j''})^{l_2}$$
$$+ (nk_j)^{l_3}(n'k_{j'})^{l_2}(n''k_{j''})^{l_1}]. \tag{3.7}$$

In these formulas, by convention, $(0)^0 = 1$; hence, for instance,

$$\gamma_{jj}(0;1) = c_{00}^{(2)} + \sum_{l=0} i^l c_{0l}^{(2)}(k_j)^l$$
$$= 2c_{00}^{(2)} + \sum_{l=1} i^l c_{0l}^{(2)}(k_j)^l. \quad (3.6c)$$

Note that these constants γ are generally *complex*. However, if the nonlinear part of (2.1) is *even* (*odd*), namely if the total number of derivatives in each term in the right-hand side of (2.1) is *even* (*odd*), implying that the (*real*) constants $c_{l_1 l_2 \cdots l_m}^{(m)}$ vanish unless the sum $\Sigma_{\mu=1}^m l_\mu$ has the *same* (*opposite*) parity as q [see (2.7a)–(2.7c)], then *all* (nonvanishing) constants $i^q\gamma$, appearing in the right-hand sides of (3.2)–(3.5) are *real* (*imaginary*). Also note that $\gamma_{jj}(n;-n)$ is necessarily *real*, since only terms with $l_1 + l_2 =$ even contribute in the rhs of (3.6a) if $j' = j$, $n' = -n$.

Let us now consider the case when the linear part is *even* [namely, (2.1) with (2.2b)], with moreover $b_0 \neq 0$ [see (2.2b)] and $q = 0$ [see (2.7a)]. It is then clear [see (2.26)–(2.28) and (3.2)–(3.6)] that the appropriate assignment for the exponents p [see (2.15c)] and r [see (2.16a)] is $p = 2$, $r = 1$, so that, in the limit of vanishing ϵ, (2.26)–(2.28) with (3.2)–(3.6) yield

$$-2i\omega_j\left(\frac{\partial}{\partial\tau} + v_j\frac{\partial}{\partial\xi}\right)\Psi_j$$
$$= \gamma_{jj}(0;1)\Psi_0\Psi_j + \sum_{j'=1}^N \gamma_{jj'}(1,1;-1)\chi_{jj'}\Psi_{j'}^*$$
$$+ \sum_{j'=1,j'\neq j'}^N (\gamma_{jj'}(1,-1;1)\tilde{\chi}_{jj'}\Psi_{j'}$$
$$+ \Psi_j \sum_{j'=1}^N \gamma_{jj'j'}(1;1;-1)|\Psi_{j'}|^2, \quad j = 1,2,\ldots,N, \quad (3.8a)$$

$$b_0\Psi_0 = \sum_{j=1}^N \gamma_{jj}(1;-1)|\Psi_j|^2, \quad (3.8b)$$

$$B_{jj'}\chi_{jj'} = \gamma_{jj'}(1;1)\Psi_j\Psi_{j'}, \quad j,j' = 1,2,\ldots,N, \quad (3.8c)$$

$$\tilde{B}_{jj'}\tilde{\chi}_{jj'} = \gamma_{jj'}(1;-1)\Psi_j\Psi_{j'}^*, \quad j,j' = 1,2,\ldots,N, \quad j\neq j'. \quad (3.8d)$$

Hence there obtain for the amplitudes Ψ_j the equations

$$\left(\frac{\partial}{\partial\tau} + v_j\frac{\partial}{\partial\xi}\right)\Psi_j = i\Psi_j\sum_{l=1}^N \alpha_{jl}|\Psi_l|^2, \quad j = 1,2,\ldots,N, \quad (3.9a)$$

with the constants α_{jl} defined as follows:

$$\alpha_{jl} = (2\omega_j)^{-1}[\gamma_{jj}(0;1)\gamma_{ll}(1;-1)/b_0$$
$$+ \gamma_{jl}(1,1;-1)\gamma_{jl}(1;1)/B_{jl}$$
$$+ (1-\delta_{jl})\gamma_{jl}(1,-1;1)\gamma_{jl}(1;-1)/\tilde{B}_{jl}$$
$$+ \gamma_{jll}(1;1;-1)]. \quad (3.9b)$$

In writing these equations, and always in the following, we use the synthetic notation $\omega_j \equiv \omega(k_j)$, $v_j \equiv v(k_j)$.

It is easily seen that these equations are explicitly solvable if the constants α_{jl} are all *real* (see Appendix A). This happens necessarily if the right-hand side of (2.1) is *even* (as defined above), since, as noted above, all the quantities γ in (3.9b) are then real. An instance belonging to this class is the (S-integrable) sine–Gordon equation (2.11) [in which case $\omega_j = s + k_j^2$, $v_j = k_j/\omega_j$, $\alpha_{jl} = s/(2\omega_j)$]. More generally, Eqs. (3.9a) are explicitly solvable if the imaginary part of the matrix α_{jl} is diagonal, or if it is proportional to the difference $v_j - v_l$ (see Appendix A).

Next let us consider the case when the linear part is *odd* [namely, (2.1) with (2.2a)], with moreover $q = 1$ [see (2.7a)]. It is then clear [see (2.22),(2.23) and (3.2),(3.3)] that the appropriate assignment for the exponents p [see (2.15c)] and r [see (2.16a)] is again $p = 2$, $r = 1$, so that, in the limiting of vanishing ϵ, (2.22), (2.23) with (3.2),(3.3) yield

$$\left(\frac{\partial}{\partial\tau} + v_j\frac{\partial}{\partial\xi}\right)\Psi_j$$
$$= ik_j\left\{\gamma_{jj}(0,1)\Psi_0\Psi_j + \sum_{l=1}^N \gamma_{jl}(1,1;-1)\chi_{jl}\Psi_l^*\right.$$
$$+ \sum_{l=1,l\neq j}^N \gamma_{jl}(1,-1;1)\tilde{\chi}_{jl}\Psi_l$$
$$\left.+ \Psi_j\sum_{l=1}^N \gamma_{jl}(1;1;-1)|\Psi_l|^2\right\}, \quad j = 1,2,\ldots,N, \quad (3.10a)$$

$$\left(\frac{\partial}{\partial\tau} + a_0\frac{\partial}{\partial\xi}\right)\Psi_0 = \left(\frac{\partial}{\partial\xi}\right)\sum_{j=1}^N \gamma_{jj}(1;-1)|\Psi_j|^2, \quad (3.10b)$$

$$A_{jl}\chi_{jl} = (k_j + k_l)\gamma_{jl}(1;1)\Psi_j\Psi_l, \quad j,l = 1,2,\ldots,N, \quad (3.10c)$$

$$\tilde{A}_{jl}\tilde{\chi}_{jl} = (k_j - k_l)\gamma_{jl}(1;-1)\Psi_j\Psi_l^*,$$
$$j,l = 1,2,\ldots,N, \quad l\neq j, \quad (3.10d)$$

implying

$$\left(\frac{\partial}{\partial\tau} + v_j\frac{\partial}{\partial\xi}\right)\Psi_j$$
$$= i\Psi_j\left[\beta_j\Psi_0 + \sum_{l=1}^N \alpha_{jl}|\Psi_l|^2\right], \quad j = 1,2,\ldots,N, \quad (3.11a)$$

$$\left(\frac{\partial}{\partial\tau} + a_0\frac{\partial}{\partial\xi}\right)\Psi_0 = \left(\frac{\partial}{\partial\xi}\right)\sum_{l=1}^N \eta_j|\Psi_j|^2, \quad (3.11b)$$

with

$$\beta_j = k_j\gamma_{jj}(0;1), \quad (3.11c)$$
$$\eta_j = \gamma_{jj}(1;-1), \quad (3.11d)$$
$$\alpha_{jl} = k_j\{(k_j + k_l)\gamma_{jl}(1,1;-1)\gamma_{jl}(1;1)/A_{jl}$$
$$+ (1-\delta_{jl})(k_j - k_l)\gamma_{jl}(1,-1;1)$$
$$\times\gamma_{jl}(1;-1)/\tilde{A}_{jl} + \gamma_{jll}(1;1;-1)\}. \quad (3.11e)$$

Here the function $\Psi_0(\xi,\tau)$, as well as the constants η_j, are of course *real*; while the constants β_j and α_{jl} may be *complex*.

If the constants β_j and α_{jl} are *real*, the general solution of this system of coupled nonlinear PDE's can be explicitly obtained (see Appendix B). Note that this happens necessarily if the right-hand side of (2.1) is *odd* (as defined above), since, as noted above, the constants γ are then all real (in this case with $q = 1$). An instance is the (S-integrable) equation (2.9) (in which case $a_0 = 0$, $\omega_j = a_1 k_j^3$

$+ a_2 k_j^5$, $v_j = 3a_1 k_j^2 + 5a_2 k_j^4$, $\beta_j = -2v_j/k_j$, $\eta_j = \beta_j/k_j$, $\alpha_{jl} = \delta_{jl} v_j/k_j^3 - 50 a_2 k_j)$.

Another instance belonging to this class is the *C*-integrable equation (2.13); in this case, however, both η_j and α_{jl} vanish, so that a different (higher) assignment of the exponents p and r becomes more appropriate. This case will be treated in a subsequent paper.

Next let us consider the case when the linear part is *odd* [i.e., (2.1) with (2.2a)], with moreover $q = 1$ [see (2.7a)] and the additional condition that the sum in the right-hand side of (2.7b) start from $m = 3$ (i.e., $c^{(2)}_{l_1 l_2} = 0$, implying $\widetilde{F}^{(2)}[u] = 0$; see (2.7c)). It is then easily seen that the proper assignment of the exponents p [see (2.15c)] and r [see (2.16a)] is $p = 2$, $r = 3$, so that, in the limit of vanishing ϵ, (2.22),(2.23) with (3.2),(3.3) yield

$$\left(\frac{\partial}{\partial\tau} + v_j \frac{\partial}{\partial\xi}\right)\Psi_j = i\,\Psi_j \sum_{l=1}^{N} \alpha_{jl}|\Psi_l|^2, \quad j = 1,2,...,N, \tag{3.12a}$$

with

$$\alpha_{jl} = k_j \gamma_{jll}(1;1;-1). \tag{3.12b}$$

As we have already noted, this system of coupled nonlinear PDE's is explicitly solvable if the constants α_{jl} are *real* (see Appendix A). An instance is the (*S*-integrable) equation (2.10) (in which case $\alpha_{jl} = 4s\, k_j[3a_1 + 5a_2(k_j^2 + k_l^2)]$).

Next let us consider the case when the linear part is *odd* [namely, (2.1) with (2.2a)], with $q = 0$. It is then clear [see (2.22),(2.23) and (3.2),(3.3)] that the appropriate assignment for the exponents p [see (2.15c)] and r [see (2.16a)] is $p = 1$, $r = 0$, so that, in the limit of vanishing ϵ, (2.22),(2.23) with (3.2),(3.3) yield

$$\left(\frac{\partial}{\partial\tau} + v_j \frac{\partial}{\partial\xi}\right)\Psi_j = \gamma_{jj}(0;1)\Psi_0\Psi_j, \quad j = 1,2,...,N, \tag{3.13a}$$

$$\left(\frac{\partial}{\partial\tau} + a_0 \frac{\partial}{\partial\xi}\right)\Psi_0 = c^{(2)}_{00}\Psi_0^2 + \sum_{j=1}^{N} \gamma_{jj}(1;-1)|\Psi_j|^2. \tag{3.13b}$$

Note, however, that if the *nonlinear* part of the original differential equation (2.1) is *odd* (as defined above), the constants $\gamma_{jj}(1;-1)$ vanish; then a different assignment for the exponents p and r becomes more appropriate. An instance of this kind is provided by the nonlinear PDE

$$\begin{aligned} u_t - u_{xxx} &= 3uu_{xxx} + 3u^2 u_{xxx} + 2cuu_x u_{xx} \\ &\quad + u^3 u_{xxx} + cu^2 u_x u_{xx}, \end{aligned} \tag{3.14}$$

that is *C*-integrable if $c = 3$ or $c = \frac{3}{2}$, and is *S*-integrable if $c = 0$ (Ref. 4); this case will be treated in a subsequent paper.

Next let us consider the case when the linear part is *odd* [namely, (2.1) with (2.2a)] and $q = 0$, but with the additional restriction that the right-hand side of (2.7b) contain only terms with $m \geqslant 3$ [namely, that all constants $c^{(2)}_{l_1 l_2}$ vanish; see (2.7c)]. It is then clear [see (2.22),(2.23) and (3.2),(3.3)] that the proper assignment for the exponents p [see (2.15c)] and r [see (2.16a)] is $p = 2$, $r = 0$, so that, in the limit of vanishing ϵ, (2.22),(2.23) with (3.2),(3.3) yield

$$\left(\frac{\partial}{\partial\tau} + v_j \frac{\partial}{\partial\xi}\right)\Psi_j = \Psi_j\left[\alpha_{j0}\Psi_0^2 + \sum_{l=1}^{N} \alpha_{jl}|\Psi_l|^2\right], \quad j = 1,2,...,N, \tag{3.15a}$$

$$\left(\frac{\partial}{\partial\tau} + a_0 \frac{\partial}{\partial\xi}\right)\Psi_0 = \Psi_0\left[\alpha_{00}\Psi_0^2 + \sum_{l=1}^{N} \alpha_{0l}|\Psi_l|^2\right], \tag{3.15b}$$

with

$$\begin{aligned} &\alpha_{jl} = \gamma_{jll}(1;1;-1), \quad \alpha_{j0} = \gamma_{jjj}(0;0;1), \\ &\alpha_{0l} = \gamma_{lll}(0;1;-1), \alpha_{00} = c^{(3)}_{000}, \quad j,l = 1,2,...,N. \end{aligned} \tag{3.15c}$$

Note that the constants a_{0l} and α_{00} are real; this is consistent with the reality of Ψ_0 [see (3.15b)].

The system (3.15a),(3.15b) can be written in the more compact form

$$\left(\frac{\partial}{\partial\tau} + v_j \frac{\partial}{\partial\xi}\right)\Psi_j = \Psi_j \sum_{l=0}^{N} \alpha_{jl}|\Psi_l|^2, \quad j = 0,1,2,...,N, \tag{3.15d}$$

after having set

$$v_0 = a_0 = v(0) \tag{3.15e}$$

[see (2.6a)]. The similarities and differences of the systems (3.15d) and (3.9a) should be noted; a particularly important difference is the presence of the factor i in the right-hand side of (3.9a), and its absence in (3.15d).

An example of a nonlinear PDE that belongs to the class we are now considering is the *C*-integrable equation (2.14a); it is easily seen that in this case $v_0 = 0$; $v_j = 3k_j^2$, $j = 1,2,...,N$; and $\alpha_{jl} = -2(v_j - v_l)$, $j,l = 0,1,2,...,N$. Hence in this case the general solution of the nonlinear system (3.15d) can be explicitly obtained (see Appendix A). This confirms the expectation that any limit equation obtained from a class of nonlinear equations of type (2.1) that contains a *C*-integrable equation, must itself be *C*-integrable; indeed we have discovered how to solve the system (3.15d) in the case with α_{jl} proportional to $v_j - v_l$ (see Appendix A), just by inserting the ansatz (2.18) in the technique appropriate to solve the *C*-integrable equation (2.14).[4,8]

Finally let us consider the case when the linear part is *even* [namely, (2.1) with (2.2b)], but now with $b_0 = 0$ [see (2.2b)] and $q = 2$ [see (2.7a)]. It is then clear [see (2.26),(2.27)] and (3.2),(3.3)] that the appropriate assignment for the exponents p [see (2.15c)] and r [see (2.16a)] is $p = 2$, $r = 1$, so that, in the limit of vanishing ϵ, (2.26)–(2.28) with (3.2)–(3.5) yield

$$\begin{aligned} -2i\omega_j\left(\frac{\partial}{\partial\tau} + v_j \frac{\partial}{\partial\xi}\right)\Psi_j &= -k_j^2\Big\{\gamma_{jj}(0;1)\Psi_0\Psi_j + \sum_{j'=1}^{N} \gamma_{jj'}(1,1;-1)\chi_{jj'}\Psi_{j'}{}^* \\ &\quad + \sum_{j'=1, j'\neq j}^{N} \gamma_{jj'}(1,-1;1)\widetilde{\chi}_{jj'}\Psi_{j'} \\ &\quad + \Psi_j \sum_{j'=1}^{N} \gamma_{jj'j'}(1;1;-1)|\Psi_{j'}|^2\Big\}, \quad j = 1,2,...,N, \end{aligned} \tag{3.16a}$$

$$\left(\frac{\partial^2}{\partial\tau^2}-b_1\frac{\partial^2}{\partial\xi^2}\right)\Psi_0=\left(\frac{\partial}{\partial\xi}\right)^2\sum_{j=1}^N\gamma_{jj}(1;-1)|\Psi_j|^2, \tag{3.16b}$$

$$B_{jj'}\chi_{jj'}=\gamma_{jj'}(1;1)\Psi_j\Psi_{j'},\quad j,j'=1,2,\dots,N, \tag{3.16c}$$

$$\widetilde{B}_{jj'}\widetilde{\chi}_{jj'}=\gamma_{jj'}(1;-1)\Psi_j\Psi_{j'}^*,\quad j,j'=1,2,\dots,N,\quad j'\neq j, \tag{3.16d}$$

implying

$$\left(\frac{\partial}{\partial\tau}+v_j\frac{\partial}{\partial\xi}\right)\Psi_j=i\Psi_j\left[\beta_j\Psi_0+\sum_{l=1}^N\alpha_{jl}|\Psi_l|^2\right],\quad j=1,2,\dots,N, \tag{3.17a}$$

$$\left(\frac{\partial^2}{\partial\tau^2}-b_1\frac{\partial^2}{\partial\xi^2}\right)\Psi_0=\left(\frac{\partial}{\partial\xi}\right)^2\sum_{j=1}^N\eta_j|\Psi_j|^2, \tag{3.17b}$$

with

$$\beta_j=-k_j^2(2\omega_j)^{-1}\gamma_{jj}(0;1), \tag{3.17c}$$

$$\eta_j=\gamma_{jj}(1;-1), \tag{3.17d}$$

$$\begin{aligned}\alpha_{jl}=&-k_j^2(2\omega_j)^{-1}\{\gamma_{jl}(1,1;-1)\gamma_{jl}(1;1)/B_{jl}\\&+(1-\delta_{jl})\gamma_{jl}(1,-1;1)\gamma_{jl}(1;-1)/\widetilde{B}_{jl}\\&+\gamma_{jll}(1,1;-1)\}.\end{aligned} \tag{3.17e}$$

Here the function $\Psi_0(\xi,\tau)$, as well as the constants η_j, are of course *real*, while the constants β_j and α_{jl} may be *complex*.

If the constants β_j and α_{jl} are real, the general solution of this system of coupled nonlinear PDE's can be explicitly obtained (see Appendix C). Note that this happens necessarily if the right-hand side of (2.1) is even (as defined above), since as noted above, the constants γ are then all real (in this case with $q=2$). An instance of this kind is provided by the (S-integrable) equation (2.12), in which case

$$\begin{aligned}&\omega_j=k_j(1-k_j^2)^{1/2},\quad v_j=k_j(1-2k_j^2)/\omega_j,\\&\beta_j=-k_j^2/\omega_j,\quad \eta_j=2,\\&\alpha_{jl}=-(k_j^2/\omega_j)[3+2/B_{jl}+2(1-\delta_{jl})/\widetilde{B}_{jl}],\\&B_{jl}=2k_jk_l\{1-2k_j^2-2k_l^2-3k_jk_l\\&\qquad-[(1-k_j^2)(1-k_l^2)]^{1/2}\},\\&\widetilde{B}_{jl}=-2k_jk_l\{1-2k_j^2-2k_l^2+3k_jk_l\\&\qquad-[(1-k_j^2)(1-k_l^2)]^{1/2}\}.\end{aligned}$$

IV. THREE-WAVE RESONANT INTERACTION

In this section we discuss the results that obtain when the number N of dispersive waves that constitute the basic approximation is three, $N=3$, and moreover the parameters k_j of these three waves satisfy the *resonant conditions*

$$\sum_{j=1}^3 k_j=0, \tag{4.1a}$$

$$\sum_{j=1}^3\omega_j=0. \tag{4.1b}$$

Note that these conditions imply [see (2.15b)]

$$z_j=-z_{j+1}-z_{j+2},\quad j=1,2,3. \tag{4.1c}$$

In this section the indices run from 1 to 3, with the usual cyclic convention setting $j+3\equiv j$. Let us recall that we always restrict attention to *real* values of the parameters k_j and of the corresponding $\omega_j=\omega(k_j)$, see (2.5a), (2.5b); and we moreover assume that none of the parameters k_j vanishes and that they differ in modulus,

$$k_j\neq 0,\quad j=1,2,3;\quad k_j\neq\pm k_l,\quad \text{if } j\neq l. \tag{4.1d}$$

In the *odd* case, these resonant conditions cannot be satisfied if $L=1$ [with $a_1\neq 0$; see (2.5a)], since in such a case (4.1a), (4.1b) and (2.5a) yield $k_1k_2k_3=0$. If $L=2$ these conditions can instead be satisfied provided $a_1/a_2<0$; for instance, a solution is

$$\begin{aligned}&k_1=k,\quad k_2=2k,\quad k_3=-3k,\quad \omega_1=4\omega,\quad \omega_2=23\omega,\\&\omega_3=-27\omega,\quad k^2=-3a_1/(35a_2),\quad \omega=8a_1k^3/35.\end{aligned} \tag{4.1e}$$

And they can of course be satisfied *a fortiori* for $L>2$.

In the *even* case, these resonant conditions can be already satisfied for $L=1$ [see (2.5b)], provided $b_0>0$ and $b_1<0$; for instance a solution is

$$\begin{aligned}&k_1=k,\quad k_2=2k,\quad k_3=-3k,\quad \omega_1=-5\omega,\quad \omega_2=4\omega,\\&\omega_3=\omega,\quad k^2=-3b_0/(28b_1),\quad \omega=(b_0/28)^{1/2}.\end{aligned} \tag{4.1f}$$

And they can of course be satisfied *a fortiori* for $L>1$.

It is then clear that the expressions of the quantities $f_j^{(m)}$ and $f_0^{(m)}$ that appear in the right-hand sides of (2.22), (2.23) and (2.26), (2.27), are now given, rather than by (3.2a), (3.2b) and (3.3a), (3.3b), by the following formulas:

$$\begin{aligned}f_j^{(2)}=&(ik_j)^q[\gamma_{j+1,j+2}(-1;-1)\Psi_{j+1}^*\Psi_{j+2}^*\\&+\epsilon^r\gamma_{jj}(0;1)\Psi_0\Psi_j]+O(\epsilon^p,\epsilon),\end{aligned} \tag{4.2a}$$

$$\begin{aligned}f_j^{(3)}=&(ik_j)^q\Big[\Psi_j\sum_{l=1}^3\gamma_{jll}(1;1;-1)|\Psi_l|^2\\&+\epsilon^r\gamma_{j,j+1,j+2}(0;-1;-1)\Psi_0\Psi_{j+1}^*\Psi_{j+2}^*\\&+\epsilon^{2r}\gamma_{jjj}(0;0;1)\Psi_0^2\Psi_j\Big]+O(\epsilon^p,\epsilon^2),\end{aligned} \tag{4.2b}$$

$$\begin{aligned}f_0^{(2)}=&\left(\frac{\partial}{\partial\xi}\right)^q\left[\sum_{j=1}^3\gamma_{jj}(1;-1)|\Psi_j|^2+\epsilon^{2r}c_{00}^{(2)}\Psi_0^2\right]\\&+O(\epsilon^p,\epsilon^2),\end{aligned} \tag{4.3a}$$

$$\begin{aligned}f_0^{(3)}=&\left(\frac{\partial}{\partial\xi}\right)^q\Big[\sum_{j=1}^3\gamma_{j,j+1,j+2}(1;-1;-1)\Psi_j\Psi_{j+1}^*\\&\times\Psi_{j+2}^*+\text{c.c.}+\epsilon^r\Psi_0\sum_{j=1}^3\gamma_{jjj}(0;1;-1)|\Psi_j|^2\\&+\epsilon^{3r}c_{000}^{(3)}\Psi_0^3\Big]+O(\epsilon^p,\epsilon^2).\end{aligned} \tag{4.3b}$$

Here, and always below, the quantities γ are defined as in the preceding section [see (3.6), (3.7)]. Note that (4.3a) coincides in fact with (3.3a).

Let us consider now the case when the linear part is *even* [namely, (2.1) with (2.2b)], with moreover $b_0\neq 0$ [see (2.2b)] and $q=0$ [see (2.7a)]. It is then clear [see (2.26), (2.27) and (4.2), (4.3)] that the appropriate assignment for the exponents p [see (2.15c)] and r [see (2.16a)] is $p=1$, $r=1$, so that, in the limit of vanishing ϵ, (2.26) with (4.2) yield

$$\left(\frac{\partial}{\partial\tau}+v_j\frac{\partial}{\partial\xi}\right)\Psi_j=\alpha_j\Psi_{j+1}{}^*\Psi_{j+2}{}^*,\quad j=1,2,3,\tag{4.4}$$

with

$$\alpha_j=(i/2)\gamma_{j+1,j+2}(-1;-1)/\omega_j.\tag{4.5}$$

Equation (4.4) is of course the well-known three-wave resonant interaction equation. Note that, if the three constants α_j appearing in the right-hand side of (4.4) have the same phase mod(π), $\alpha_j=s_j|\alpha_j|\exp(i\theta), j=1,2,3, s_j=\pm$, then by the trivial rescaling $\Psi_j\to|\alpha_{j+1}\alpha_{j+2}|^{-1/2}\exp(i\theta/3)\Psi_j$ one can rewrite (4.4) with all the constants α_j replaced by the signs s_j.

The S-integrable sine–Gordon equation (2.11) belongs to the class we are now considering [actually, not quite so, since b_1 is positive; see (2.11b) and (4.1f)]; one might therefore infer that this fact explains why the three-wave resonant interaction equation (4.4) is itself S-integrable. But such an inference would not be justified, since in the case of the sine–Gordon equation (whose nonlinear part contains only terms that are cubic or of higher order), the quantities α_j vanishing identically [see (4.5) and (3.6a)], so that (4.4) in this case becomes linear. Indeed it is easily seen that a more appropriate assignment, in the case of the sine–Gordon equation, is $p=2, r=2$, yielding again (3.9a) [with $N=3$ and $\alpha_{jl}=s/(2\omega_j)$].

Next, let us consider the case when the linear part is *odd* [namely, (2.1) with (2.2a)], with moreover $q=1$ [see (2.7a)]. It is then clear [see (2.22),(2.23) and (4.2),(4.3)] that the proper assignment for the exponents p [see (2.15c)] and r [see (2.16a)] is again $p=1, r=1$, so that, in the limit of vanishing ϵ, (2.26) with (4.2) yields again (4.4), but now with

$$\alpha_j=ik_j\gamma_{j+1,j+2}(-1;-1).\tag{4.6}$$

The S-integrable equation (2.9) belongs to the class we are now considering; one might therefore infer that this fact explains why the three-wave resonant interaction equation (4.4) is itself S-integrable. But such an inference would again be unjustified, since in the case of (2.9) one obtains

$$\gamma_{j+1,j+2}(-1,-1)=-6a_1-10a_2(k_{j+1}{}^2+k_{j+2}{}^2+k_{j+1}k_{j+2}),\tag{4.7}$$

and it is easily seen that (4.1a),(4.1b) with (2.5a) (with $L=2$) imply that the right-hand side of this equation vanishes. Hence also in this case the right-hand side of (4.4) vanishes. Indeed it is easily seen that in this case a more appropriate assignment for the exponents p and r is $p=2$, $r=1$, yielding back (3.11).

Next let us consider again the case when the linear part is *even* [namely, (2.1) with (2.2b)], but now with $b_0=0$ [see (2.2b)] and $q=2$ [see (2.7a)]. It is then clear [see (2.26),(2.27) and (4.2),(4.3)] that the appropriate assignment for the exponents p [see (2.15c)] and r [see (2.16a)] is $p=1, r=1$, so that, in the limit of vanishing ϵ, (2.26) with (4.2) yields again (4.4), but now with

$$\alpha_j=k_j{}^2\gamma_{j+1,j+2}(-1;-1)/(2i\omega_j).\tag{4.8}$$

An instance belonging to this class is the Boussinesq equation (2.12) [note incidentally that a choice of the parameters k_j is possible that is consistent with (4.1a)–(4.1d) and (2.5b), (2.12c); for instance, $k_1=k$, $k_2=2k$, $k_3=-3k, \omega_1=5\omega, \omega_2=-8\omega, \omega_3=3\omega$ with $k^2=\frac{3}{28}$ and $\omega^2=k^2/28$]. In this case one obtains

$$\alpha_j=-ik_j{}^2/\omega_j.\tag{4.9}$$

Hence the S-integrability of the three-wave resonant interaction (4.4) can be considered to follow from the S-integrability of the Boussinesq equation (2.12) (a more explicit analysis of this connection, including the derivation of the Lax pair for the three-wave resonant interaction from that of the Boussinesq equation, will be given elsewhere).

Finally let us consider once more the case with *odd* linear part [namely, (2.1) with (2.2a)], with $q=0$ and moreover with the assumption that the right-hand side of (2.7b) contain only terms with $m\geqslant3$ [namely, that all constants $c^{(2)}_{l_1l_2}$ vanish; see (2.7c)]. It is then clear [see (2.22),(2.23) and (4.2),(4.3)] that the appropriate assignment for the exponents p [see (2.15c)] and r [see (2.16a)] is $p=2, r=0$, so that, in the limit of vanishing ϵ, (2.22),(2.23) with (4.2),(4.3) yield

$$\begin{aligned}\left(\frac{\partial}{\partial\tau}+v_j\frac{\partial}{\partial\xi}\right)\Psi_j&=\Psi_j\sum_{l=1}^{3}\gamma_{jll}(1;1;-1)|\Psi_l|^2\\&\quad+\gamma_{j,j+1,j+2}(0;-1;-1)\Psi_0\Psi_{j+1}{}^*\Psi_{j+2}{}^*\\&\quad+\gamma_{jjj}(0;0;1)\Psi_0{}^2\Psi_j,\quad j=1,2,3,\end{aligned}\tag{4.10a}$$

$$\begin{aligned}\left(\frac{\partial}{\partial\tau}+a_0\frac{\partial}{\partial\xi}\right)\Psi_0&=\left[\sum_{j=1}^{3}\gamma_{j,j+1,j+2}(1;-1;-1)\Psi_j\Psi_{j+1}{}^*\Psi_{j+2}{}^*+\text{c.c.}\right]\\&\quad+\Psi_0\sum_{j=1}^{3}\gamma_{jjj}(0;1;-1)|\Psi_j|^2+c^{(3)}_{000}\Psi_0{}^3.\end{aligned}\tag{4.10b}$$

Note, however, that the C-integrable equation (2.12) does not belong to the class under consideration here [see the remark in the paragraph following Eq. (4.1d)].

V. FOUR-WAVE RESONANT INTERACTION

In this section we discuss the results that obtain when the number N of dispersive waves that constitute the baisc approximation is four, $N=4$, and moreover the parameters k_j of these four waves satisfy the resonant conditions

$$\sum_{j=1}^{4}k_j=0,\tag{5.1a}$$

$$\sum_{j=1}^{4}\omega_j=0,\tag{5.1b}$$

implying [see (2.25b)]

$$z_j=-z_{j+1}-z_{j+2}-z_{j+3}.\tag{5.1c}$$

In this section the indices run from 1 to 4, with the usual cyclic convention $j+4\equiv j$; and, as in the preceding section, we assume the real quantities k_j to be different in modulus among themselves and not to vanish:

$$k_j\neq0,\quad j=1,2,3,4;\quad k_j\neq\pm k_l,\quad\text{if } j\neq l.\tag{5.1d}$$

In the *odd* case, these resonant conditions cannot be satisfied if $L=1$ [with $a_1\neq 0$; see (2.5a)]. They can be satisfied for $L=2$, provided $a_1/a_2<0$; for instance, a solution is

$$k_1=k,\quad k_2=-2k,\quad k_3=-3k,\quad k_4=4k,\quad k^2=-a_1/(25a_2). \tag{5.1e}$$

And they can of course be satisfied *a fortiori* for $L>2$.

Also in the *even* case, these resonant conditions cannot be satisfied if $L=1$ [see (2.5b)].

It is now clear that the expressions of $f_j^{(3)}$ and $g_j^{(2)}$ that enter in the right-hand side of (2.22)–(2.24) and (2.26)–(2.28) must be modified as follows:

$$f_j^{(3)}=(ik_j)^q\Big\{\gamma_{j+1,j+2,j+3}(-1;-1;-1)\Psi_{j+1}{}^*\times\Psi_{j+2}{}^*\Psi_{j+3}{}^*+\Psi_j\Big[\sum_{j'=1}^{4}\gamma_{jj'j'}(1;1;-1)|\Psi_{j'}|^2+\epsilon^{2r}\gamma_{jjj}(0;0;1)\Psi_0{}^2\Big]\Big\}+O(\epsilon^p,\epsilon^{r+1},\epsilon^2), \tag{5.2}$$

$$g_{jj'}^{(2)}=[i(k_j+k_{j'})]^q\{\gamma_{jj'}(1;1)\Psi_j\Psi_{j'}+(1-\delta_{jj'})\times[\gamma_{j''j'''}(-1;-1)+\gamma_{j'''j''}(-1;-1)\Psi_{j''}{}^*\Psi_{j'''}{}^*\}. \tag{5.3}$$

In the right-hand side of the last equation, the indices j'' and j''' are different among themselves and different from j and j'; for instance, if $j=1$ and $j'=2$, then $j''=3$ and $j'''=4$ (or, equivalently, $j''=4$ and $j'''=3$). These formulas, (5.2) and (5.3), replace (3.2b) and (3.4a); note that the other formulas, namely, (3.2a), (3.3a), (3.3b), (3.4b) and (3.6a),(3.6b) remain applicable.

Let us now, as in the preceding sections, analyze the implications of these results in the context of some representative classes of nonlinear evolution equations of type (2.1).

Let us consider first the case when the linear part is *even* [namely, (2.1) with (2.2b)], with moreover $b_0\neq 0$ [see (2.2b)] and $q=0$ [see (2.7a)]. It is then clear that the appropriate assignment for the exponents p [see (2.15c)] and r [see (2.16a)] is $p=2$, $r=1$, so that, in the limit of vanishing ϵ, (2.26)–(2.28) yield

$$-2i\omega_j\Big(\frac{\partial}{\partial\tau}+v_j\frac{\partial}{\partial\xi}\Big)\Psi_j=\gamma_{jj}(0;1)\Psi_0\Psi_j+\sum_{l=1}^{4}\gamma_{jl}(1;1;-1)\chi_{jl}\Psi_l{}^*+\sum_{l=1,l\neq j}^{4}\gamma_{jl}(1;-1;1)\tilde{\chi}_{jl}\Psi_l+\gamma_{j+1,j+2,j+3}(-1;-1;-1)\times\Psi_{j+1}{}^*\Psi_{j+2}{}^*\Psi_{j+3}{}^*+\Psi_j\sum_{l=1}^{4}\gamma_{jll}(1;1;-1)|\Psi_l|^2,\quad j=1,2,3,4, \tag{5.4a}$$

$$b_0\Psi_0=\sum_{j=1}^{4}\gamma_{jj}(1;-1)|\Psi_j|^2, \tag{5.4b}$$

$$B_{jj'}\chi_{jj'}=[\gamma_{j''j'''}(-1;-1)+\gamma_{j'''j''}(-1;-1)]\Psi_{j''}{}^*\Psi_{j'''}{}^*+\gamma_{jj'}(1;1)\Psi_j\Psi_{j'},\quad j,j'=1,2,3,4, \tag{5.4c}$$

$$\tilde{B}_{jl}\tilde{\chi}_{jl}=\gamma_{jl}(1;-1)\Psi_j\Psi_l{}^*,\quad j,l=1,2,3,4,\quad l\neq j. \tag{5.4d}$$

Hence there obtain for the amplitudes Ψ_j the equations

$$\Big(\frac{\partial}{\partial\tau}+v_j\frac{\partial}{\partial\xi}\Big)\Psi_j=i\tilde{\alpha}_j\Psi_{j+1}{}^*\Psi_{j+2}{}^*\Psi_{j+3}{}^*+i\Psi_j\sum_{l=1}^{4}\alpha_{jl}|\Psi_l|^2,\quad j=1,2,3,4, \tag{5.5a}$$

with α_{jl} defined by (3.9b) and

$$\tilde{\alpha}_j=(2\omega_j)^{-1}\Big[\gamma_{j+1,j+2,j+3}(-1;-1;-1)+\sum{}'\gamma_{jj'}(1;1;-1)\gamma_{j''j'''}(-1;-1)/B_{jj'}\Big]. \tag{5.5b}$$

The symbol Σ' in the last equation indicates a sum over all values of the indices j', j'', and j''' different from j and among themselves:

$$\sum{}'\equiv\sum_{j'=1,j'\neq j}^{4}\ \sum_{j''=1,j''\neq j,j'}^{4}\ \sum_{j'''=1,j'''\neq jj'j''}^{4}. \tag{5.6}$$

Hence this sum contains generally six terms (actually, three different terms, each counted twice).

Next, let us consider the case when the linear part is *odd* [namely, (2.1) with (2.2a)], with $q=1$ [see (2.7a)]. It is then clear that the appropriate assignment for the exponents p [see (2.15c)] and r [see (2.16a)] is again $p=2$, $r=1$, so that, in the limit of vanishing ϵ, (2.22)–(2.24) yield

$$\Big(\frac{\partial}{\partial\tau}+v_j\frac{\partial}{\partial\xi}\Big)\Psi_j=ik_j\Big\{\gamma_{jj}(0,1)\Psi_0\Psi_j+\sum_{l=1}^{4}\gamma_{jl}(1,1;-1)\chi_{jl}\Psi_l{}^*+\sum_{l=1,l\neq j}^{4}\gamma_{jl}(1,-1;1)\tilde{\chi}_{jl}\Psi_l+\Psi_j\times\sum_{l=1}^{4}\gamma_{jll}(1;1;-1)|\Psi_l|^2+\gamma_{j+1,j+2,j+3}(-1;-1;-1)\times\Psi_{j+1}{}^*\Psi_{j+2}{}^*\Psi_{j+3}{}^*\Big\},\quad j=1,2,3,4, \tag{5.7a}$$

$$\Big(\frac{\partial}{\partial\tau}+a_0\frac{\partial}{\partial\xi}\Big)\Psi_0=\Big(\frac{\partial}{\partial\xi}\Big)\sum_{j=1}^{4}\gamma_{jj}(1;-1)|\Psi_j|^2, \tag{5.7b}$$

$$A_{jl}\chi_{jl}=(k_j+k_l)\{\gamma_{jl}(1;1)\Psi_j\Psi_l+(1-\delta_{jl})[\gamma_{j'l'}(-1;-1)+\gamma_{l'j'}(-1;-1)]\Psi_{j'}{}^*\Psi_{l'}{}^*\},\quad j,l=1,2,3,4, \tag{5.7c}$$

$$\tilde{A}_{jl}\tilde{\chi}_{jl}=(k_j-k_l)\gamma_{jl}(1;-1)\Psi_j\Psi_l{}^*,\quad j,l=1,2,3,4,\quad j\neq l. \tag{5.7d}$$

In (5.7c), the indices j' and l' are different from j and l (and since, in the relevant contribution, j and l are required to be different, this condition identifies them, up to an irrelevant permutation).

Hence there obtain for the amplitudes Ψ_j the equations

$$\Big(\frac{\partial}{\partial\tau}+v_j\frac{\partial}{\partial\xi}\Big)\Psi_j=i\tilde{\alpha}_j\Psi_{j+1}{}^*\Psi_{j+2}{}^*\Psi_{j+3}{}^*+i\Psi_j\Big[\beta_j\Psi_0+\sum_{l=1}^{4}\alpha_{jl}|\Psi_l|^2\Big],\quad j=1,2,3,4, \tag{5.8a}$$

$$\left(\frac{\partial}{\partial\tau}+a_0\frac{\partial}{\partial\xi}\right)\Psi_0=\left(\frac{\partial}{\partial\xi}\right)\sum_{l=1}^{4}\eta_j|\Psi_j|^2, \tag{5.8b}$$

with β_j, η_j, and α_{jl} defined by (3.11c)–(3.11e) and

$$\tilde{\alpha}_j=k_j\Big[\gamma_{j+1,j+2,j+3}(-1;-1;-1)+\sum{}'(k_j+k_{j'})\times\gamma_{jj'}(1,1;-1)\gamma_{j''j'''}(-1;-1)/A_{jj'}\Big]. \tag{5.8c}$$

Here Σ' is defined as above [see (5.6)]. Note that (5.8b) coincides essentially with (3.11b), and that these equations, (5.8b), constitute a generalization of (5.5a) (to which they reduce for $\beta_j=0$).

An equation belonging to this class is the (S-integrable) equation (2.9), in which case

$$a_0=0,\quad \omega_j=a_1k_j^3+a_2k_j^5,\quad v_j=3a_1k_j^2+5a_2k_j^4,$$

$$\beta_j=-\frac{2v_j}{k_j},\ \eta_j=\frac{\beta_j}{k_j},\ \alpha_{jl}=\frac{\delta_{jl}v_j}{k_j^3}-50a_2k_j,$$

$$\tilde{\alpha}_j=-60a_2k_j+24a_1\sum_{l=1,l\neq j}^{4}(k_l)^{-1}+20a_2\sum{}'\frac{(k_{j''}^2+k_{j'''}^2+k_{j''}k_{j'''})}{k_{j'}}.$$

Hence at least for these values of the constants $\tilde{\alpha}_j$, β_j, α_{jl}, and η_j [and values of the parameters a_1, a_2, and k_j consistent with (5.1a), (5.1b); see, for instance, (5.1e)], Eq. (5.8a), (5.8b) should be (at least) S-integrable. A further investigation of this question is postponed to a subsequent paper.

Another equation belonging to this class is the (S-integrable) equation (2.10). In this case, however, since the sum in the right-hand side of (2.7b) starts with $m=3$, it is easily seen that a more appropriate assignment for the exponent r [see (2.16a)] is $r=3$ (always with $p=2$); hence in this case the equations that obtain take the form (5.5a), with $\tilde{\alpha}_j$ and α_{jl} defined by the following simple formulas:

$$\tilde{\alpha}_j=k_j\gamma_{j+1,j+2,j+3}(-1;-1;-1), \tag{5.9a}$$

$$\alpha_{jl}=k_j\gamma_{jll}(1;1;-1) \tag{5.9b}$$

[which are merely special cases of (5.8c) and (3.11e)]. The values of these quantities that correspond to (2.10) are

$$\tilde{\alpha}_j=4s\,k_j[3a_1-5a_2(k_{j+1}^2+k_{j+2}^2+k_{j+3}^2+k_{j+1}k_{j+2}+k_{j+2}k_{j+3}+k_{j+3}k_{j+1})],$$

$$\alpha_{jl}=4s\,k_j[3a_1+5a_2(k_j^2+k_l^2)];$$

hence Eq. (5.5a), with these values of $\tilde{\alpha}_j$ and α_{jl}, and with $v_j=k_j^2(3a_1+5a_2k_j^2)$, should be (at least) S-integrable [provided the values of the parameters a_1, a_2, and k_j are consistent with (5.1a), (5.1b); see for instance (5.1e)]. This question will be further elaborated in a subsequent paper.

VI. OUTLOOK

As indicated in the Introduction, the main aim of this paper has been to show how the ideas and approach that lead to widely applicable and generally integrable universal equations of nonlinear Schrödinger type[1–4] can be extended to yield universal equations of N-wave interaction type; which of course turn out also to be widely applicable and generally integrable. We have outlined the context and approach that leads to such equations, and we have exhibited several examples. A more systematic treatment of this topic is postponed to a separate paper. Among the topics that will be covered there, in addition to a more complete survey of the various universal equations of N-wave interaction type that correspond to various classes of nonlinear evolution equations, the results will be reported that have been obtained from a more extended search for integrable equations of N-wave interaction type obtainable from known integrable equations. To this end it has also been convenient to push the approach beyond its "leading order" application by looking at special cases when some key parameters vanish. The extension to more than $1+1$ dimensions, which can be accommodated without any difficulty, has also proved fruitful. Also of interest is the exploration of situations characterized by the simultaneous presence of resonant and nonresonant waves; and the extensions that are obtained when one takes as a starting point of the treatment a system of coupled nonlinear evolution equations rather than a single equation.

ACKNOWLEDGMENTS

I wish to thank Attilio Maccari for spotting a mistake in an early draft of this paper.

APPENDIX A

In this Appendix we report the general solution of the *nonlinear* system of first-order PDE's

$$\left(\frac{\partial}{\partial\tau}+v_j\frac{\partial}{\partial\xi}\right)\Psi_j(\xi,\tau)=i\,\Psi_j(\xi,\tau)\sum_{l=1}^{N}\alpha_{jl}|\Psi_l(\xi,\tau)|^2,\quad j=1,2,...,N, \tag{A1}$$

when the (constant) matrix α_{jl},

$$\alpha_{jl}=\lambda_{jl}+i\mu_{jl}, \tag{A2}$$

has an imaginary part that is either diagonal,

$$\mu_{jl}=\mu_j\delta_{jl}, \tag{A3a}$$

or has the off-diagonal form

$$\mu_{jl}=(v_j-v_l)c_l^2. \tag{A3b}$$

Here the constants v_j, λ_{jl}, μ_{jl}, μ_j, and c_l are, of course, all *real.* Note that the case when α_{jl} is real, namely μ_{jl} vanishes, is included in both cases (A3a) and (A3b). Of course a change of the overall sign in the right-hand side of (A3b) can be easily accommodated by changing the signs of ξ and τ [see (A1)].

We moreover report a formula that linearizes (A1) in the general case (namely, when α_{jl} is an arbitrary complex matrix), but which we have not been able to invert, so that we cannot obtain from it explicit expressions for the solutions $\Psi_j(\xi,\tau)$.

Set

$$\Psi_j(\xi,\tau)=[\sigma_j(\xi,\tau)]^{1/2}\exp[i\theta_j(\xi,\tau)],\quad j=1,2,...,N, \tag{A4}$$

with σ_j *positive* and θ_j *real.* Then (A1) yields

$$\left(\frac{\partial}{\partial\tau}+v_j\frac{\partial}{\partial\xi}\right)\theta_j=\sum_{l=1}^{N}\lambda_{jl}\,\sigma_l, \tag{A5a}$$

$$\left(\frac{\partial}{\partial\tau}+v_j\frac{\partial}{\partial\xi}\right)\sigma_j=-2\sigma_j\sum_{l=1}^{N}\mu_{jl}\sigma_l. \tag{A5b}$$

The first of these two sets of equations, (A5a), can be solved for $\theta_j(\xi,\tau)$ if the $\sigma_l(\xi,\tau)$ are assumed known:

$$\theta_j(\xi,\tau)=\theta_j(\xi-v_j\tau,0)+\sum_{l=1}^{N}\lambda_{jl}\times\int_0^\tau d\tau'\,\sigma_l[\xi-v_j(\tau-\tau'),0]. \tag{A6}$$

Here the N *real* functions $\theta_j(\xi,0)$ can be chosen arbitrarily.

The general solution of the second set of equations, (A5b), is given by the following two formulas, (A7a) or (A7b), in the two cases (A3a) or (A3b) (as can be easily verified):

$$\sigma_j(\xi,\tau)=\sigma_j(\xi-v_j\tau,0)/[1+2\mu_j\tau\sigma_j(\xi-v_j\tau,0)], \tag{A7a}$$

$$\sigma_j(\xi,\tau)=\rho_j(\xi-v_j\tau)\times\left(\left[c_0^2+2\sum_{l=1}^{N}c_l^2\int_{\xi_0}^{\xi-v_l\tau}d\xi'\rho_l(\xi')\right]\right)^{-1}. \tag{A7b}$$

In (A7a), the N *positive* functions $\sigma_j(\xi,0)$ are arbitrary; note that this formula implies that $\sigma_j(\xi,\tau)$ is nonsingular for $\tau>0$ (for all real values of ξ), provided none of the constant μ_j is negative, $\mu_j\geqslant 0$ [assuming of course that the functions $\sigma_j(\xi,0)$ are themselves nonsingular for all real values of ξ]. In (A7b), the two *real* constants c_0 and ξ_0 are *real* but otherwise arbitrary, and the N *real* functions $\rho_j(\xi)$ are also arbitrary; they are related to the "initial" data $\sigma_j(\xi,0)$ by the formula

$$\sigma_j(\xi,0)=\rho_j(\xi)\left(\left[c_0^2+2\sum_{l=1}^{N}c_l^2\int_{\xi_0}^{\xi}d\xi'\rho_l(\xi')\right]\right)^{-1}, \tag{A8a}$$

implying

$$\rho_j(\xi)=c_0\sigma_j(\xi,0)\exp\left[-2\sum_{l=1}^{N}c_l^2\int_{\xi_0}^{\xi}d\xi'\,\sigma_l(\xi',0)\right]. \tag{A8b}$$

The C-integrability of (A5b) in the case (A3b) had been already pointed out in Ref. 9 [see Eq. (2.8b) of that paper].

In the general case, namely when μ_{jl} need not satisfy the constraints (A3a) or (A3b), it is still possible to linearize the nonlinear system (A5b). The formula that accomplishes this task reads

$$\rho_j(\xi,\tau)=\sigma_j(\xi,\tau)\exp\left\{2\sum_{l=1}^{N}\mu_{jl}\int_0^\tau dt'\times\sigma_l[\xi-v_j(\tau-\tau'),\tau']\right\}. \tag{A9}$$

Indeed it is easily seen that, via (A5b), it implies the trivial linear equations

$$\left(\frac{\partial}{\partial\tau}+v_j\frac{\partial}{\partial\xi}\right)\rho_j(\xi,\tau)=0,\quad j=1,2,...N, \tag{A10a}$$

whose general solution is of course [see also (A9)]

$$\rho_j(\xi,\tau)=\sigma_j(\xi-v_j\tau,0),\quad j=1,2,...,N, \tag{A10b}$$

where the functions $\sigma_j(\xi,0)$ can be assigned arbitrarily.

Let us finally note that these results can be easily generalized to the case with more than one spatial dimension.

APPENDIX B

In this Appendix we report the general solution of the *nonlinear* system of first-order PDE's

$$\left(\frac{\partial}{\partial\tau}+v_j\frac{\partial}{\partial\xi}\right)\Psi_j=i\Psi_j\left[\beta_j\Psi_0+\sum_{l=1}^{N}\alpha_{jl}|\Psi_l|^2\right],\quad j=1,2,...N, \tag{B1a}$$

$$\left(\frac{\partial}{\partial\tau}+a_0\frac{\partial}{\partial\xi}\right)\Psi_0=\left(\frac{\partial}{\partial\xi}\right)\sum_{l=1}^{N}\eta_j|\Psi_j|^2, \tag{B1b}$$

where the (given) constants v_j, β_j, α_{jl}, a_0, and η_j are *real*, and the function $\Psi_0(\xi,\tau)$ is also *real* [while the functions $\Psi_j(\xi,\tau)$ are *complex*].

Set

$$\Psi_j(\xi,\tau)=[\sigma_j(\xi,\tau)]^{1/2}\exp[i\theta_j(\xi,\tau)],\quad j=1,2,...,N, \tag{B2}$$

with σ_j *positive* and θ_j *real*. Then (B1a), (B1b) yield

$$\left(\frac{\partial}{\partial\tau}+v_j\frac{\partial}{\partial\xi}\right)\sigma_j=0,\quad j=1,2,...N, \tag{B3a}$$

$$\left(\frac{\partial}{\partial\tau}+v_j\frac{\partial}{\partial\xi}\right)\theta_j=\beta_j\Psi_0+\sum_{l=1}^{N}\alpha_{jl}\sigma_l,\quad j=1,2,...,N, \tag{B3b}$$

$$\left(\frac{\partial}{\partial\tau}+a_0\frac{\partial}{\partial\xi}\right)\Psi_0=\left(\frac{\partial}{\partial\xi}\right)\sum_{j=1}^{N}\eta_j\sigma_j. \tag{B3c}$$

The general solution of these equations reads as follows:

$$\sigma_j(\xi,\tau)=\sigma_j(\xi-v_j\tau,0),\quad j=1,2,...N, \tag{B4a}$$

$$\Psi_0(\xi,\tau)=\varphi(\xi-a_0\tau)+\sum_{j=1}^{N}\left[\frac{\eta_j}{a_0-v_j}\right]\sigma_j(\xi-v_j\tau,0), \tag{B4b}$$

$$\theta_j(\xi,\tau)=\theta_j(\xi-v_j\tau,0)+\int_0^\tau d\tau'\left\{\beta_j\Psi_0[\xi-v_j(\tau-\tau'),\tau']+\sum_{l=1}^{N}\alpha_{jl}\sigma_l[\xi-v_j(\tau-\tau'),\tau']\right\},\quad j=1,2,...,N. \tag{B4c}$$

Here the $2N+1$ real functions $\sigma_j(\xi,0)$, $\varphi(\xi)$, and $\theta_j(\xi,0)$ can be chosen arbitrarily. Note that we are implicitly assuming that none of the N constants v_j coincides with a_0.

Let us note, more generally, that the system (B1) can be solved even if the constants α_{jl} in (B1a) are not real,

$$\alpha_{jl}=\lambda_{jl}+i\mu_{jl}, \tag{B5}$$

provided the imaginary part of α_{jl} is either diagonal, $\mu_{jl}=\mu_j\delta_{jl}$, or proportional to v_j-v_l, $\alpha_{jl}=(v_j-v_l)c_l^2$; and it can be linearized if the complex matrix α_{jl} satisfies no restriction. The technique to deal with these cases [via (B2)] is sufficiently close to that described in Appendix A that it does not warrant an explicit treatment here.

APPENDIX C

In this Appendix we report the general solution of the *nonlinear* system of first-order PDE's

$$\left(\frac{\partial}{\partial\tau}+v_j\frac{\partial}{\partial\xi}\right)\Psi_j=i\Psi_j\left[\beta_j\Psi_0+\sum_{l=1}^{N}\alpha_{jl}|\Psi_l|^2\right],$$
$$j=1,2,\ldots,N, \qquad \text{(C1a)}$$
$$\left(\frac{\partial^2}{\partial\tau^2}-b_1\frac{\partial^2}{\partial\xi^2}\right)\Psi_0=\left(\frac{\partial}{\partial\xi}\right)^2\sum_{j=1}^{N}\eta_j|\Psi_j|^2, \qquad \text{(C1b)}$$
where the (given) constants v_j,β_j, α_{jl}, b_1, and η_j are *real*, and the function $\Psi_0(\xi,\tau)$ is also *real* [while the N functions $\Psi_j(\xi,\tau)$ are *complex*].

Set

$$\Psi_j(\xi,\tau)=[\sigma_j(\xi,\tau)]^{1/2}\exp[i\theta_j(\xi,\tau)],\quad j=1,2,\ldots,N, \qquad \text{(C2)}$$

with σ_j positive and θ_j real. Then (C1a), (C1b) yield

$$\left(\frac{\partial}{\partial\tau}+v_j\frac{\partial}{\partial\xi}\right)\sigma_j=0,\quad j=1,2,\ldots,N, \qquad \text{(C3a)}$$
$$\left(\frac{\partial}{\partial\tau}+v_j\frac{\partial}{\partial\xi}\right)\theta_j=\beta_j\Psi_0+\sum_{l=1}^{N}\alpha_{jl}\sigma_l,\quad j=1,2,\ldots,N, \qquad \text{(C3b)}$$
$$\left(\frac{\partial^2}{\partial\tau^2}-b_1\frac{\partial^2}{\partial\xi^2}\right)\Psi_0=\left(\frac{\partial}{\partial\xi}\right)^2\sum_{j=1}^{N}\eta_j|\Psi_j|^2. \qquad \text{(C3c)}$$

The general solution of these equations reads as follows:

$$\sigma_j(\xi,\tau)=\sigma_j(\xi-v_j\tau,0),\quad j=1,2,\ldots,N, \qquad \text{(C4a)}$$
$$\Psi_0(\xi,\tau)=\varphi(\xi-b_1^{1/2}\tau)+\chi(\xi+b_1^{1/2}\tau)-\sum_{j=1}^{N}\left[\frac{\eta_j}{(b_1+v_j^2)}\right]\sigma(\xi-v_j\tau,0), \qquad \text{(C4b)}$$
$$\theta_j(\xi,\tau)=\theta_j(\xi-v_j\tau,0)+\int_0^\tau d\tau'\left\{\beta_j\Psi_0[\xi-v_j(\tau-\tau'),\tau']+\sum_{l=1}^{N}\alpha_{jl}\sigma_l[\xi-v_j(\tau-\tau'),\tau']\right\},\quad j=1,2,\ldots,N. \qquad \text{(C4c)}$$

Here the $2N+2$ real functions $\sigma_j(\xi,0),\varphi(\xi),\chi(\xi)$, and $\theta_j(\xi,0)$ can be chosen arbitrarily. Note that we are implicitly assuming that b_1 is positive.

More generally, the system (C1) can be explicitly solved if α_{jl} is complex but its imaginary part is either diagonal or proportional to $(v_j-v_l)c_l^{\,2}$; and it can be linearized for an arbitrary α_{jl}. The treatment of these cases [via (C2)] is so close to that reported in Appendix A that it does not warrant further elaboration here.

[1] Galileo Galilei, *Il Saggiatore*, 1623 [see *Opere di Galileo Galilei*, Edizione nazionale (Barbera, Firenze, 1890-1909), Vol. VI, p. 232].

[2] F. Calogero and W. Eckhaus, "Nonlinear evolution equations, rescalings, model PDE's and their integrability, I.," Inverse Problems **3**, 229 (1987).

[3] F. Calogero and W. Eckhaus, "Nonlinear evolution equations, rescalings, model PDE's and their integrability. II.," Inverse Problems **4**, 11 (1988).

[4] F. Calogero, "Why are certain nonlinear PDE's both widely applicable and integrable?," *What is Integrability (For Nonlinear PDE's)?*, edited by V. E. Zakharov (Springer, Berlin, to be published).

[5] F. Calogero and W. Eckhaus, "Necessary conditions for integrability of nonlinear PDE's," Inverse Problems **3**, L27 (1987).

[6] F. Calogero and A. Maccari, "Equations of nonlinear Schroedinger type in 1 + 1 and 2 + 1 dimensions, obtained from integrable PDE's," *Inverse Problems: An interdisciplinary study* (Proceedings of the meeting on Inverse Problems, Montpellier, November 1986), Advances in Electronics and Electron Physics, Vol. 19, edited by P. C. Sabatier (Academic, New York, 1987), pp. 463–480.

[7] F. Calogero and A. Degasperis, *Spectral Transform and Solitons: Tools to Solve and Investigate Nonlinear Evolution Equations* (North-Holland, Amsterdam, 1982), Vol. 1.

[8] F. Calogero, "The evolution PDE $u_t=u_{xxx}+3(u_{xx}u^2+3u_xu^2)+3u_xu^4$," J. Math. Phys. **28**, 538 (1987).

[9] D. Levi, O. Ragnisco, and M. Bruschi, "Continuous and discrete matrix Burgers' hierarchies," Nuovo Cimento B **74**, 33 (1983).

A new generalization of the Hankel integral transform

A. A. Inayat-Hussain
Applied Mechanics and Electromagnetics Group, BHP Melbourne Research Laboratories, P. O. Box 264, Clayton, Victoria 3168, Australia

(Received 19 May 1988; accepted for publication 17 August 1988)

A study of the eigenfunction expansions associated with the Bessel equation defined on an unbounded composite region has yielded a new generalization of the Hankel integral transform. This generalization contains as a special case an integral transform that is the Neumann (in the sense of boundary value problem) counterpart of the (Dirichlet-type) Weber–Orr transform and which itself is new. It also contains as special cases certain curious integral representations of the Dirac δ function. These representations are in fact the orthogonality conditions for the quasiorthogonal (i.e., orthogonal with respect to a discontinuous weight function) kernels of the new integral transforms.

I. INTRODUCTION

Integral transforms provide a powerful means of obtaining the solutions of boundary value problems of mathematical physics.[1–4]

The Hankel transform, in particular, has been widely used for solving boundary value problems for homogeneous media with boundary conditions that possess cylindrical geometry.

The new generalization of the Hankel transform to be described here extends the class of exactly solved boundary value problems to include those in which the boundary conditions traverse an unbounded composite region; both the boundary conditions and the region are assumed to have cylindrical geometry. This generalization contains as a special case an integral transform that is the Neumann (in the sense of boundary value problem) counterpart of the (Dirichlet-type) Weber–Orr transform[5] and is itself new. It also contains as special cases certain curious integral representations of the Dirac δ function. These representations are in fact the orthogonality conditions for the quasiorthogonal[6] (i.e., orthogonal with respect to a discontinuous weight function) kernels of the new integral transforms.

Our new transform has emerged from a study of the eigenfunction expansions associated with the Bessel equation

$$\frac{1}{r}\frac{d}{dr}\left[r\frac{d}{dr}y(r)\right]+\left(\lambda^2-\frac{\nu^2}{r^2}\right)y(r)=0, \tag{1}$$

which is defined on the two disjoint regions, $0<r<a$ and $a<r<\infty$, where $a\geqslant 0$, $\lambda\geqslant 0$, and $\nu\geqslant 0$. The function $y(r)$ is bounded at $r=0$ and $r=\infty$, and at $r=a$ it satisfies

$$\lim_{r\to a^-} y(r) = \lim_{r\to a^+} y(r) \tag{2a}$$

and

$$\lim_{r\to a^-} \sigma\frac{dy(r)}{dr} = \lim_{r\to a^+} \frac{dy(r)}{dr}. \tag{2b}$$

The parameter σ ($\sigma\geqslant 0$) embodies the information on the physical properties of the two disjoint regions. For example, in the problem of steady electric current flow in a composite conductor comprised of a cylindrical conductor imbedded in an otherwise homogeneous infinite conductor, σ would correspond to the ratio of the electrical conductivities of the two conductors.

The eigenfunctions (corresponding to the continuous spectrum of eigenvalues λ^2) of (1) subject to the conditions in (2a) and (2b) are given by

$$\Psi_\nu(r,\lambda;a,\sigma)=\begin{cases}J_\nu(\lambda r), & 0<r<a,\\ (\pi\lambda a/2)\{[J_\nu(\lambda a)Y_\nu'(\lambda a)-\sigma J_\nu'(\lambda a)Y_\nu(\lambda a)]J_\nu(\lambda r)-(1-\sigma)J_\nu(\lambda a)J_\nu'(\lambda a)Y_\nu(\lambda r)\}, & r>a,\end{cases} \tag{3}$$

where the functions $J_\nu(z)$ and $Y_\nu(z)$ are, respectively, the Bessel (of the first kind) and Weber functions of order ν, and the prime denotes a derivative; for example, $J_\nu'(\lambda a)$ is $(d/dz)J_\nu(z)$ evaluated at $z=\lambda a$. These eigenfunctions can be shown, by a direct calculation[7] using Lommel's integral (12), to be orthogonal with respect to the discontinuous weight function $\omega(r)$, defined by

$$\omega(r)=\begin{cases}\sigma r, & 0<r<a,\\ r, & r>a.\end{cases} \tag{4}$$

Thus, the quasiorthogonal eigenfunctions in (3) can serve as a basis set for the expansion of an arbitrary function and, indeed, our generalization of the Hankel transform given in (5) and (6) below provides an example of this.

The main results of this paper, namely, the generalized Hankel transform and its special case (and their proofs) and the integral representations of the δ function are given in Sec. II. Section III contains a summary of this paper.

II. MAIN RESULTS

A. Generalized transform

The new generalization of the Hankel transform is given by the relations

 0022-2488/89/010041-04$02.50

$$F_\nu(\lambda) = \int_0^\infty dr\, \omega(r)\Psi_\nu(r,\lambda;a,\sigma)f(r) \tag{5}$$

and

$$f(r) = \int_0^\infty d\lambda \frac{\lambda}{\Omega_\nu(\lambda)} \Psi_\nu(r,\lambda;a,\sigma)F_\nu(\lambda)\,, \tag{6}$$

where the integrals are assumed to be convergent and $a \geqslant 0$, $\sigma \geqslant 0$, and $\nu \geqslant 0$; the kernel Ψ_ν is given in (3) and the positive definite[8] function $\Omega_\nu(\lambda)$ is defined by[9]

$$\begin{aligned}\Omega_\nu(\lambda) = (\pi\lambda a/2)^2\{&J_\nu^2(\lambda a)[J_\nu'^2(\lambda a) + Y_\nu'^2(\lambda a)] \\ &- 2\sigma J_\nu(\lambda a)J_\nu'(\lambda a)[J_\nu(\lambda a)J_\nu'(\lambda a) \\ &+ Y_\nu(\lambda a)Y_\nu'(\lambda a)] \\ &+ \sigma^2 J_\nu'^2(\lambda a)[J_\nu^2(\lambda a) + Y_\nu^2(\lambda a)]\}\,.\end{aligned} \tag{7}$$

Equation (5) gives the transform of order ν of the function $f(r)$, whereas (6) is the inversion formula for the transform. As can be seen from (5) and (6), the integral transform does not possess the self-reciprocal property of the Hankel transform.

The proof of the transform pair (5) and (6) is based on a contour integration technique similar to that used by MacRobert[10] in his elegant proof of the Hankel transform.

Consider then the double integral

$$\begin{aligned}I \equiv \int_0^\infty &dr\, \omega(r)\Psi_\nu(r,\mu;a,\sigma) \\ &\times \int_\alpha^\beta d\lambda \frac{\lambda}{\Omega_\nu(\lambda)} \Psi_\nu(r,\lambda;a,\sigma)F_\nu(\lambda)\,,\end{aligned} \tag{8}$$

where the function $F_\nu(\lambda)$ is assumed to be analytic in some portion of the complex λ plane containing the line between α and β $(0<\alpha<\beta)$. The Bessel functions in Ψ_ν can be written in terms of the Hankel functions $H_\nu^{(1)}$ and $H_\nu^{(2)}$ such that $\Psi_\nu = \Psi_\nu^{(1)} + \Psi_\nu^{(2)}$, where

$$\Psi_\nu^{(1)}(r,\lambda;a,\sigma) = \begin{cases} \frac{1}{2}H_\nu^{(1)}(\lambda r), \quad 0<r<a, \\ (\pi\lambda a/4)[J_\nu(\lambda a)Y_\nu'(\lambda a) - \sigma J_\nu'(\lambda a)Y_\nu(\lambda a) \\ \quad + i(1-\sigma)J_\nu(\lambda a)J_\nu'(\lambda a)]H_\nu^{(1)}(\lambda r), \quad r>a. \end{cases} \tag{9}$$

The corresponding expression for $\Psi_\nu^{(2)}$ is obtained from (9) be replacing all superscripts (1) with the superscript (2) and i with $-i$.

Following MacRobert,[10] the line contour between α and β can be deformed onto the two contours C_1 and C_2, as shown in Fig. 1, and the integral I becomes

$$\begin{aligned}I = \int_0^\infty &dr\, \omega(r)\Psi_\nu(r,\mu;a,\sigma)\int_{C_1} d\lambda \\ &\times \frac{\lambda}{\Omega_\nu(\lambda)} \Psi_\nu^{(1)}(r,\lambda;a,\sigma)F_\nu(\lambda) \\ &+ \int_0^\infty dr\, \omega(r)\Psi_\nu(r,\mu;a,\sigma)\int_{C_2} d\lambda \\ &\times \frac{\lambda}{\Omega_\nu(\lambda)} \Psi_\nu^{(2)}(r,\lambda;a,\sigma)F_\nu(\lambda)\,.\end{aligned} \tag{10}$$

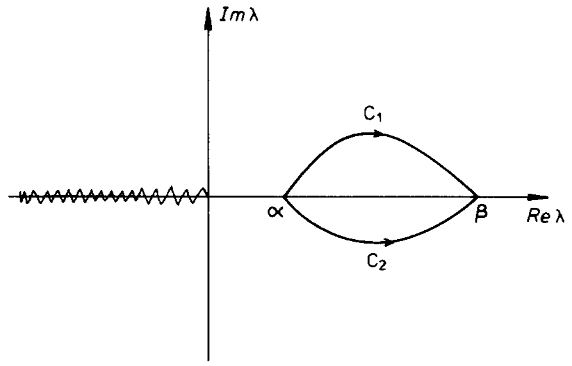

FIG. 1. Contours in the cut λ plane for the integral I.

The Hankel functions $H_\nu^{(1)}(\lambda r)$ and $H_\nu^{(2)}(\lambda r)$ (and therefore $\Psi_\nu^{(1)}$ and $\Psi_\nu^{(2)}$ decay exponentially with r on the contours C_1 and C_2, respectively, and we can interchange the order of integration in the two double integrals in (10). Thus

$$\begin{aligned}I = \int_{C_1} &d\lambda \frac{\lambda}{\Omega_\nu(\lambda)} F_\nu(\lambda) \int_0^\infty dr\, \omega(r) \\ &\times \Psi_\nu(r,\mu;a,\sigma)\Psi_\nu^{(1)}(r,\lambda;a,\sigma) \\ &+ \int_{C_2} d\lambda \frac{\lambda}{\Omega_\nu(\lambda)} F_\nu(\lambda) \int_0^\infty dr\, \omega(r) \\ &\times \Psi_\nu(r,\mu;a,\sigma)\Psi_\nu^{(2)}(r,\lambda;a,\sigma)\,.\end{aligned} \tag{11}$$

The evaluation of the integrals over r is straightforward but tedious and makes use of Lommel's integral[11]

$$\begin{aligned}(\lambda^2 - \mu^2)&\int_a^b dx\, xU_\nu(\lambda x)V_\nu(\mu x) \\ &= [\mu x U_\nu(\lambda x)V_\nu'(\mu x) - \lambda x U_\nu'(\lambda x)V_\nu(\mu x)]_a^b\,,\end{aligned} \tag{12}$$

where U_ν and V_ν are Bessel functions of order ν. This leads to

$$\begin{aligned}I = -i\int_C &d\lambda \frac{\lambda}{\Omega_\nu(\lambda)} \frac{F_\nu(\lambda)}{\lambda^2 - \mu^2} \\ &\times \left\{\xi_\nu(\lambda) + \frac{1}{2}\sigma\left[2\left(\frac{\mu}{\lambda}\right)^\nu (\pi)^{-1}\right.\right. \\ &\left.\left. + \mu a J_\nu'(\mu a)Y_\nu(\lambda a) - \lambda a J_\nu(\mu a)Y_\nu'(\lambda a)\right]\right\}\,,\end{aligned} \tag{13}$$

where

$$\begin{aligned}\xi_\nu(\lambda) &= (\pi\lambda a/4)\{\lambda a J_\nu(\lambda a)J_\nu(\mu a)[J_\nu'^2(\lambda a) + Y_\nu'^2(\lambda a)] \\ &\quad - \sigma[\lambda a J_\nu'(\lambda a)J_\nu(\mu a) + \mu a J_\nu(\lambda a)J_\nu'(\mu a)] \\ &\quad \times [J_\nu(\lambda a)J_\nu'(\lambda a) + Y_\nu(\lambda a)Y_\nu'(\lambda a)] \\ &\quad + \sigma^2 \mu a J_\nu'(\lambda a)J_\nu'(\mu a)[J_\nu^2(\lambda a) + Y_\nu^2(\lambda a)]\}\,,\end{aligned} \tag{14}$$

and C is the (anticlockwise) closed contour $C_2 - C_1$ between α and β. Noting the analyticity of $F_\nu(\lambda)$, the integral I can be evaluated by the residue theorem to yield

$$I=\begin{cases}F_\nu(\mu), & \alpha<\mu<\beta,\\ 0, & \mu<\alpha \text{ or } \mu>\beta,\end{cases} \tag{15}$$

thus completing the proof of the integral transform pair (5) and (6).

For the two separate choices of parameters $\sigma=1$ and $a=0$, (5) and (6) can be shown to reduce to the Hankel transform.

The generalized Hankel transform pair (5) and (6) does not satisfy any simple convolution-type relation. However, as in the case of the Hankel transform,[12] it is easy to derive the Parseval-type relation,

$$\int_0^\infty d\lambda\,\frac{\lambda}{\Omega_\nu(\lambda)}F_\nu(\lambda)G_\nu(\lambda)=\int_0^\infty dr\,\omega(r)f(r)g(r)\,, \tag{16}$$

where $F_\nu(\lambda)$ and $G_\nu(\lambda)$ are the generalized Hankel transforms of the functions $f(r)$ and $g(r)$, respectively.

B. A special case

The special case of the generalized Hankel transform (5) and (6) obtained by taking the limit $\sigma\to0$ (Ref. 13) yields the integral transform pair

$$F_\nu(\lambda)=\int_a^\infty dr\,r\phi_\nu(r,\lambda;a)f(r)\,, \tag{17a}$$

$$f(r)=\int_0^\infty d\lambda\frac{\lambda\phi_\nu(r,\lambda;a)F_\nu(\lambda)}{J_\nu'^2(\lambda a)+Y_\nu'^2(\lambda a)}, \tag{17b}$$

where

$$\phi_\nu(r,\lambda;a)\equiv Y_\nu'(\lambda a)J_\nu(\lambda r)-J_\nu'(\lambda a)Y_\nu(\lambda r)\,. \tag{17c}$$

This new result is the Neumann counterpart of the (Dirichlet-type) Weber–Orr transform[5] pair

$$F_\nu(\lambda)=\int_a^\infty dr\,r\zeta_\nu(r,\lambda;a)f(r)\,, \tag{18a}$$

$$f(r)=\int_0^\infty d\lambda\,\frac{\lambda\zeta_\nu(r,\lambda;a)F_\nu(\lambda)}{J_\nu^2(\lambda a)+Y_\nu^2(\lambda a)}, \tag{18b}$$

where

$$\zeta_\nu(r,\lambda;a)\equiv Y_\nu(\lambda a)J_\nu(\lambda r)-J_\nu(\lambda a)Y_\nu(\lambda r)\,; \tag{18c}$$

the kernel of the Weber–Orr transform satisfies the boundary condition $\zeta_\nu(r,\lambda;a)=0$ at $r=a$, while the kernel (17c) satisfies the boundary condition $(\partial/\partial r)\phi_\nu(r,\lambda;a)=0$ at $r=a$. Thus whereas the Weber–Orr transform is applicable to Dirichlet boundary value problems, the transform in (17a) and (17b) would be applicable to Neumann boundary value problems.

C. Dirac δ functions

The generalized Hankel transform given by (5) and (6) and its special case given by (17a) and (17b) can provide curious integral representations of the Dirac δ function. These representations, given here in (19a)–(19d), are in fact the orthogonality relations for the kernels of the integral transforms, and can easily be derived by taking the transforms of $f(r)=\delta(r-r_0)$ and of $F_\nu(\lambda)=\delta(\lambda-\lambda_0)$:

$$\delta(r-r_0)=\omega(r_0)\int_0^\infty d\lambda\,\lambda\,[\Omega_\nu(\lambda)]^{-1}\times\Psi_\nu(r,\lambda;a,\sigma)\Psi_\nu(r_0,\lambda;a,\sigma)\,, \tag{19a}$$

$$\delta(\lambda-\lambda_0)=\lambda\,[\Omega_\nu(\lambda)]^{-1}\int_0^\infty dr\,\omega(r)\times\Psi_\nu(r,\lambda;a,\sigma)\Psi_\nu(r,\lambda_0;a,\sigma), \tag{19b}$$

$$\delta(r-r_0)=r\int_0^\infty d\lambda\,\lambda\frac{\phi_\nu(r,\lambda;a)\phi_\nu(r_0,\lambda;a)}{J_\nu'^2(\lambda a)+Y_\nu'^2(\lambda a)}, \tag{19c}$$

$$\delta(\lambda-\lambda_0)=\lambda\,[J_\nu'^2(\lambda a)+Y_\nu'^2(\lambda a)]^{-1}\int_a^\infty dr\,r\times\phi_\nu(r,\lambda;a)\phi_\nu(r,\lambda_0;a)\,. \tag{19d}$$

The corresponding relations for the Hankel transform and the Weber–Orr transform are already established.[14]

III. CONCLUSIONS

We have presented a new generalization of the Hankel transform in terms of the quasiorthogonal eigenfunctions of a boundary value problem defined on a composite region.

This generalization has yielded as a special case an integral transform that is the Neumann counterpart of the Weber–Orr transform and which itself is new. It has also yielded for the Dirac δ function certain integral representations, which turn out to be the orthogonality conditions for the kernels of the new integral transforms.

The application of the generalized Hankel transform to a problem in steady current flow in a composite medium (arising in the context of the electrical resistivity method of geophysical exploration) is currently in preparation and will be published elsewhere.

ACKNOWLEDGMENT

I wish to thank Professor C. J. Thompson, Department of Mathematics, University of Melbourne for a critical reading of this paper.

[1]B. Davies, *Integral Transforms and their Applications* (Springer, New York, 1985), 2nd ed., and references therein.

[2]I. N. Sneddon, *The Use of Integral Transforms* (McGraw-Hill, New York, 1972).

[3]I. Stakgold, *Boundary Value Problems of Mathematical Physics* (MacMillan, New York, 1968), Vols. I and II.

[4]N. N. Lebedev, I. P. Skalskaya, and Y. S. Uflyand, *Problems of Mathematical Physics* (Prentice-Hall, Englewood Cliffs, NJ, 1965), revised, enlarged, and corrected English edition, translated and edited by R. A. Silverman.

[5]E. C. Titchmarsh, Proc. London Math. Soc. **22**, 15 (1923).

[6]Quasiorthogonal functions were first discussed in the context of finite transforms (as opposed to integral transforms considered here) by C. W. Tittle, J. Appl. Phys. **36**, 1486 (1965). Recent applications of finite transforms based on quasiorthogonal eigenfunctions are given in J. Baker-Jarvis and R. Inguva, J. Appl. Phys. **57**, 1569 (1985); M. D. Mikhailov, M. N. Özisik, and N. L. Vulchanov, Int. J. Heat Mass Transfer **26**, 1131 (1983).

[7]Alternatively, the differential equation (1) can be used to obtain, for $\lambda\neq\mu$,

$$\int_0^\infty dr\,\omega(r)\Psi_\nu(r,\lambda;a,\sigma)\Psi_\nu(r,\mu;a,\sigma)$$
$$= a(\lambda^2-\mu^2)^{-1}[\sigma\Psi_\nu(a^-,\lambda;a,\sigma)\Psi_\nu'(a^-,\mu;a,\sigma) - \sigma\Psi_\nu(a^-,\mu;a,\sigma)\Psi_\nu'(a^-,\lambda;a,\sigma) - \Psi_\nu(a^+,\lambda;a,\sigma)\Psi_\nu'(a^+,\mu;a,\sigma) + \Psi_\nu(a^+,\mu;a,\sigma)\Psi_\nu'(a^+,\lambda;a,\sigma)]\,,$$

the right-hand side of which is reduced to zero by the conditions (2a) and (2b).

[8]This can be demonstrated by use of the Cauchy–Schwartz inequality and the Wronskian relation $J_\nu(z)Y_\nu'(z) - J_\nu'(z)Y_\nu(z) = 2/(\pi z)$ that yields the bound

$$[J_\nu(\lambda)J_\nu'(\lambda) + Y_\nu(\lambda)Y_\nu'(\lambda)]^2 < [J_\nu^2(\lambda) + Y_\nu^2(\lambda)][J_\nu'^2(\lambda) + Y_\nu'^2(\lambda)],$$

which then implies

$$\Omega_\nu(\lambda) > (\pi\lambda a/2)^2\{J_\nu(\lambda a)[J_\nu'^2(\lambda a) + Y_\nu'^2(\lambda a)]^{1/2} - \sigma J_\nu'(\lambda a)[J_\nu^2(\lambda a) + Y_\nu^2(\lambda a)]^{1/2}\}^2 \geqslant 0.$$

[9]The notation $J_\nu^2(x)$ means $[J_\nu(x)]^2$, and $J_\nu'^2(x)$ means $[J_\nu'(x)]^2$.

[10]T. M. MacRobert, Proc. R. Soc. Edinburgh **51**, 116 (1931).

[11]G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U. P., Cambridge, England, 1958), 2nd ed., p. 134.

[12]See, for example, Ref. 1, p. 241.

[13]In this limit, Ψ_ν is defined only for $r \geqslant a$; see (1), (2a), and (2b).

[14]See, for example, Ref. 3, Vol. I, p. 316 and Vol. II, p. 243, respectively.

Green's functions of the edh operators

J. Ivancovich, C. Kozameh,[a] and E. T. Newman
Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

(Received 30 March 1988; accepted for publication 7 September 1988)

The spin-weight s Green's functions for the operator $\eth$ and powers of $\eth$ are obtained. The extension of these Green's functions to negative values of s and to the $\bar{\eth}^n$ operators, as well as a procedure for obtaining the Green's function for any combination of products of $\eth^n$ and $\bar{\eth}^n$, is also given.

I. INTRODUCTION

The differential operator edh, denoted by $\eth$, has played an extremely useful role in many areas of mathematical physics, as for example, in general relativity, Maxwell and Yang–Mills theory, integrable systems, etc. Since edh frequently appears in the form of differential equations, there has been interest in its Green's functions. Because edh acts on different classes of functions [the so-called spin-s functions, mapping them into the spin-$(s+1)$ functions] there will be different classes of Green's functions for edh that are labeled by s. Furthermore, there is interest in the Green's functions for powers of edh, i.e., for $\eth^n$. These Green's functions will be labeled by both s and n and written as $K_{s,-(s+n)}$. Though earlier work[1] on these Green's functions had been mainy confined to the cases $s = 0$ and $n = 1$ or 2, recent developments have led to the need to generalize this to arbitrary s and n. Furthermore, it has been possible to modify these Green's functions slightly and thereby improve them. It is the purpose of this paper to present these modified Green's functions (for arbitrary s, including s half-integer) for both edh and edh-bar and their powers and to provide a procedure for obtaining the Green's functions for any combination of products of the edh and edh-bar operators, e.g., $\eth^m\bar{\eth}^n$.

In Sec. II notation is described and a brief review of both spin-weighted functions and the action of the edh and edh-bar operators on spin-weighted functions is given. Section III contains a summary of some of the previous work done on Green's functions with a discussion of several problems associated with defining the Green's functions. Two alternative examples (from the same class of Green's functions) are presented and their differences contrasted. In Sec. IV, the new or generalized Green's functions for positive spin weight are described; these are seen as direct generalizations of the results of Sec. III. In Sec. V, extensions of these basic Green's functions to negative spin weights and to the Green's function for the edh-bar operator for both positive and negative spin weights are presented. In addition, an integration procedure is introduced that yields the Green's function for any combination of products of the edh and edh-bar operators. Appendix A outlines a proof that the expressions given in Sec. IV are the generalized Green's functions. In Appendix B it is shown that the generalized Green's functions possess the property that when integrated over the sphere against appropriate spin-weighted data, the resulting solution contains no elements of the kernel of the edh operator. This property is referred to as the kernel exclusion property of the Green's functions. Appendix C provides a convenient spinorial representation for the Green's functions defined in Secs. IV and V.

II. MATHEMATICAL PRELIMINARIES

In order to understand better the edh operator,[2] it is necessary to discuss spin-weighted functions on the sphere. Spin-weighted functions correspond to irreducible tensor representations of quantities on the sphere; the spin weight (which arises from the associated tensor type) refers to the behavior of the functions under rotations. If the complex vector $\mathbf{m}$ and its complex conjugate $\bar{\mathbf{m}}$ described by

$$\sqrt{2}\mathbf{m} = \mathbf{a} + i\,\mathbf{b}\,, \tag{1}$$

where $\mathbf{a}$ and $\mathbf{b}$ are orthonormal vector fields tangent to the sphere, are introduced, then a rotation of these vectors through an angle ψ is given by

$$\mathbf{m}' = e^{i\psi}\mathbf{m}\,. \tag{2}$$

Under this rotation, a spin-weight s function α_s, transforms according to

$$\alpha_s' = e^{is\psi}\alpha_s\,, \tag{3}$$

the spin weight of a function will be indicated by a subscript, i.e., α_s. (A simple example of a spin-weight 1 and -1 function is obtained by considering an arbitrary vector field $\mathbf{T}$, tangent to the sphere, and defining

$$\alpha_1 = \mathbf{T}\cdot\mathbf{m}\,, \quad \alpha_{-1} = \mathbf{T}\cdot\bar{\mathbf{m}}\,.)$$

(For a more mathematically complete definition of spin-weighted functions, see Refs. 3–5.) Any regular spin-weight s function can be expanded in a complete orthonormal basis set, the elements of this set are denoted ${}_sY_{lm}(\theta,\phi)$, $l \geqslant |s|$ and $-l \leqslant m \leqslant l$, and are referred to as the spin-s spherical harmonics; note that $s = 0$ corresponds to the ordinary spherical harmonics.

The edh (edh-bar) operator (which acts on these spin-s functions) is essentially the projection of the covariant derivative on the metric sphere along the complex direction $\mathbf{m}$ ($\bar{\mathbf{m}}$). If the sphere is coordinatized by the usual complex stereographic coordinates $(\zeta,\bar{\zeta})$, then the action of the edh operator on a spin-weight s function is given by

$$\eth_\zeta\alpha_s = (1+\zeta\bar{\zeta})^{1-s}[(1+\zeta\bar{\zeta})^s\alpha_s]_{,\zeta} \equiv P^{1-s}(P^s\alpha_s)_{,\zeta}\,, \tag{4}$$

where $\eth_\zeta$ denotes the edh operator taken with respect to the

[a] Current address: Facultad de Matematica, Astronomia y Fisica, Universidad Nacional de Cordoba, Argentina.

 0022-2488/89/010045-08$02.50

variables $(\zeta,\bar\zeta)$. Similarly, the action of the edh-bar operator can be given:

$$\bar\eth_\zeta \alpha_s = (1-\zeta\bar\zeta)^{1+s}[(1+\zeta\bar\zeta)^{-s}\alpha_s]_{,\zeta} \equiv P^{1+s}(P^{-s}\alpha_s)_{,\zeta}. \tag{5}$$

The action of the edh operator on a spin-weight s function $\eth_\zeta \alpha_s$ yields a spin-weight $s+1$ function; thus acting as a spin-weight raising operator. Similarly, the action of the edh-bar operator on a spin-weight s function $\bar\eth_\zeta \alpha_s$ yields a spin-weight $s-1$ function, thus acting as a spin-weight lowering operator. The actions of the edh and edh-bar operators on the spin-s spherical harmonics are given by

$$\eth_{\zeta s} Y_{lm}(\zeta,\bar\zeta) = [(l-s)(l+s+1)]^{1/2}{}_{s+1}Y_{lm}(\zeta,\bar\zeta)$$

and

$$\bar\eth_{\zeta s} Y_{lm}(\zeta,\bar\zeta) = -[(l+s)(l-s+1)]^{1/2}{}_{s-1}Y_{lm}(\zeta,\bar\zeta). \tag{6}$$

From the knowledge of the ordinary spherical harmonics and the actions of the edh and edh-bar operators, an iterative definition of the spin-s (s-integer) spherical harmonics is obtained. The spin-s spherical harmonics can be explicitly written as

$$
\begin{aligned}
{}_sY_{lm}(\zeta,\bar\zeta) = A_{slm} \sum_p (-1)^p \binom{l-s}{p} \\
\times \binom{l+s}{p+s-m} \frac{\zeta^p \bar\zeta^{p+s-m}}{(1+\zeta\bar\zeta)^l}
\end{aligned}
\tag{7}
$$

with

$$
\begin{aligned}
A_{slm} &= (-1)^{l+s}[[(l+m)!(l-m)!(2l+1)] \\
&\quad \times [4\pi(l-s)!(l+s)!]^{-1}]^{1/2}.
\end{aligned}
$$

Expression (7) can be applied to spinor harmonics for which the spin-weight s is half-integer. Note the important point that there exists certain spherical harmonics which are annihilated by edh (or edh-bar), namely those with $l=|s|$ parts, s positive in the case of edh and s negative in the case of edh-bar. These harmonics are elements of the kernel of the edh (or edh-bar) operator; the kernel being defined as the set of all functions annihilated by the edh (or edh-bar) operator, i.e., functions satisfying the equation $\bar\eth_\zeta f_s(\zeta,\bar\zeta)=0$ [or $\eth_\zeta f_s(\zeta,\bar\zeta)=0$].

III. SUMMARY OF PREVIOUS RESULTS

The previous work of Porter[1] concerns the integration of equations of the form

$$\eth_\zeta F_0(\zeta,\bar\zeta) = A_1(\zeta,\bar\zeta), \tag{8}$$

where $F_0(\zeta,\bar\zeta)$ is a regular function of spin-weight 0 and $A_1(\zeta,\bar\zeta)$ is a regular function of spin weight 1. The solution to (8) can be written as an integral over the sphere

$$F_0(\zeta,\bar\zeta) = \int_{S^2} K_{0,-1}(\zeta,\bar\zeta;\eta,\bar\eta)A_1(\eta,\bar\eta)\,d\mu_\eta, \tag{9}$$

where the Green's function of the edh operator $K_{0,-1}(\zeta,\bar\zeta;\eta,\bar\eta)$ is of spin-weight 0 in $(\zeta,\bar\zeta)$ and of spin-weight -1 in $(\eta,\bar\eta)$; note the notation of the Green's functions and its respective spin weights. The measure on the sphere is given by $d\mu_\eta = 2(d\eta\wedge d\bar\eta)/i(1+\eta\bar\eta)^2$. The solution to (8) is defined up to the freedom of the addition of spin-weight 0 functions that edh annihilates, the general solution to (8) would therefore be given by (9) plus the addition of an arbitrary element of the kernel of edh. An important point to note about (9) is that the data function $A_1(\eta,\bar\eta)$ is given as a regular function on the sphere and the sought for solution $F_0(\zeta,\bar\zeta)$ is also to be regular; this preservation of regularity is an important needed property of the Green's functions. The Porter expression for the Green's function is

$$K_{0,-1}(\zeta,\bar\zeta;\eta,\bar\eta) = (1/4\pi)[(1+\eta\bar\eta)/(\bar\zeta-\bar\eta)]. \tag{10}$$

A slight modification yields an alternate expression for the Green's function in (9), namely

$$K_{0,-1}(\zeta,\bar\zeta;\eta,\bar\eta) = (1/4\pi)[(1+\bar\zeta\eta)/(\bar\zeta-\bar\eta)]. \tag{11}$$

These expressions are two examples of $(0,-1)$ Green's functions of edh. The second expression (11) possesses two attractive properties not possessed by the Porter expression (10) that are now discussed.

(i) Using standard Cartesian coordinates x^a in Minkowski space, introduce the null tetrad, parametrized by $(\zeta,\bar\zeta)$ and defined by

$$
\begin{aligned}
l^a(\zeta,\bar\zeta) &= [(1+\zeta\bar\zeta,\zeta+\bar\zeta,i(\bar\zeta-\zeta),-1+\zeta\bar\zeta)] \\
&\qquad \times[\sqrt{2}(1+\zeta\bar\zeta)]^{-1}, \\
m^a(\zeta,\bar\zeta) &= \eth_\zeta l^a(\zeta,\bar\zeta), \\
\bar m^a(\zeta,\bar\zeta) &= \bar\eth_\zeta l^a(\zeta,\bar\zeta), \\
n^a(\zeta,\bar\zeta) &= l^a(\zeta,\bar\zeta) + \eth_\zeta\bar\eth_\zeta l^a(\zeta,\bar\zeta),
\end{aligned}
$$

where $l^a(\zeta,\bar\zeta)$ is assigned a spin weight of 0 and for each $(\zeta,\bar\zeta)$, $l^a(\zeta\bar\zeta)n_a(\zeta,\bar\zeta) = -m^a(\zeta,\bar\zeta)m_a(\zeta,\bar\zeta)=1$, all other products vanishing [note that a signature of $(+,-,-,-)$ is used throughout the discussion]. Given this null tetrad, (11) can be rewritten as

$$K_{0,-1}(\zeta,\bar\zeta;\eta,\bar\eta) = \frac{-1}{4\pi}\frac{l^a(\zeta,\bar\zeta)\bar m_a(\eta,\bar\eta)}{l^b(\zeta,\bar\zeta)l_b(\eta,\bar\eta)}. \tag{12}$$

This form of the Green's function is not only attractive but it is often useful when solving (8), since on occasion it is convenient to express the source term or data $A_1(\zeta,\bar\zeta)$ in terms of this null tetrad. (This happens frequently when dealing with the d'Adhemar form of the Maxwell equations or the self-dual Yang–Mills equations in the Sparling form.[6])

(ii) It can also be shown that the expression (11) gives the solution (9) in which $F_0(\zeta,\bar\zeta)$ has no $l=0$ part in its spherical harmonic decomposition, i.e., the solution contains no elements of the kernel of the edh (refer to Appendix A). This will be referred to as the kernel exclusion property of the Green's function. Therefore, the Green's function (11) yields the most basic form of the solution (9). (An $l=0$ part can always be added due to the freedom in the solution.) This property is particularly attractive in that it allows for the definition of a unique Green's function. When the initial data $A_1(\zeta,\bar\zeta)$ is purely an lth spherical harmonic, i.e., ${}_1Y_{lm}(\zeta,\bar\zeta)$, the solution (9) yields the lth spin-0 harmonic without any spurious elements of the kernel present and thus can be thought of as edh inverse.

IV. GREEN'S FUNCTIONS OF THE EDH OPERATORS

Now consider equations of the form

$$\eth_\zeta{}^n F_s(\zeta,\bar\zeta) = A_{s+n}(\zeta,\bar\zeta)\,, \tag{13}$$

where $F_s(\zeta,\bar\zeta)$ is a regular function of spin-weight s, s positive, and $A_{s+n}(\zeta,\bar\zeta)$ is a regular function of spin-weight $s+n$. The solution to (13) can be written in the form

$$F_s(\zeta,\bar\zeta) = \int_{S^2} K^+{}_{s,-(s+n)}(\zeta,\bar\zeta;\eta,\bar\eta) A_{s+n}(\eta,\bar\eta)\, d\mu_\eta\,, \tag{14}$$

where the Green's function $K^+{}_{s,-(s+n)}(\zeta,\bar\zeta;\eta,\bar\eta)$ is of spin-weight s in $(\zeta,\bar\zeta)$ and of spin-weight $-(s+n)$ in $(\eta,\bar\eta)$. The solution to (13) is defined up to the freedom of the addition of spin-weight s functions which the $\eth_\zeta{}^n$ operator annihilates and thus the general solution to (13) is given by (14) plus the addition of an arbitrary spin-weight s element of the kernel of $\eth_\zeta{}^n$ operator. Again, the initial data $A_{s+n}(\eta,\bar\eta)$ is a regular function on the sphere and a Green's function is sought that preserves this regularity in the solution $F_s(\zeta,\bar\zeta)$. A choice for the Green's function is given by

$$K^+{}_{s,-(s+n)}(\zeta,\bar\zeta;\eta,\bar\eta) = \frac{(-1)^{n-1}}{4\pi(n-1)!} \times \frac{(1+\zeta\bar\eta)^{2s+n}(\eta-\zeta)^{n-1}}{(1+\zeta\bar\zeta)^{s+n-1}(1+\eta\bar\eta)^{s+n-1}(\bar\eta-\bar\zeta)}\,. \tag{15}$$

A proof that this expression is indeed a Green's function is outlined in Appendix A. The Green's function, given by (15), can also be written in terms of the previously defined null tetrad, as

$$K^+{}_{s,-(s+n)}(\zeta,\bar\zeta;\eta,\bar\eta) = \frac{(-1)^{s+n-1}}{4\pi(n-1)!} \frac{[m^a(\zeta,\bar\zeta)\bar m_a(\eta,\bar\eta)]^s l^b(\zeta,\bar\zeta)\bar m_b(\eta,\bar\eta)[l^c(\zeta,\bar\zeta)\bar m_c(\eta,\bar\eta)]^{n-1}}{l^d(\zeta,\bar\zeta)l_d(\eta,\bar\eta)}\,. \tag{16}$$

This form is particularly attractive in that the expression is a product of three terms, the first term $[m^a(\zeta,\bar\zeta)\bar m_a(\eta,\bar\eta)]^s$ describes the generalization of the Green's function to spin-weight s, the second term $l^b(\zeta,\bar\zeta)l_b(\eta,\bar\eta)/l^d(\zeta,\bar\zeta)l_d(\eta,\bar\eta)$ represents a basic core Green's function, $K^+{}_{0,-1}(\zeta,\bar\zeta;\eta,\bar\eta)$, and the third term $[l^c(\zeta,\bar\zeta)\bar m_c(\eta,\bar\eta)]^{n-1}$ describes the generalization to higher orders of the edh operator. The disadvantage of the null tetrad version of the Green's function (16) is that it is only defined for integer spin weight, whereas the form (15) is defined for both integer and half-integer spin weight. The Green's function (15) also possesses the kernel exclusion property; the solution (14) gives an $F_s(\zeta,\bar\zeta)$ that has a spherical harmonic decomposition for which the $l\in\{s,\dots,s+n-1\}$ (i.e., the kernel) are absent. The proof that the Green's function (15) possesses this property is given in Appendix B.

V. EXTENSIONS OF THE GREEN'S FUNCTIONS

The Green's functions obtained in the previous section were associated with positive s. It is possible to extend their definition to include Green's functions for negative s. Care must be taken when seeking solutions to (13) for negative s because there exist choices of data $A_{s+n}(\zeta,\bar\zeta)$ for which solutions do not exist. When $2s+n<0$, data that contains in its spherical harmonic decomposition parts for which $l\in\{|s+n|,\dots,|s|-1\}$ will not yield a regular solution. If $2s+n\geqslant 0$ there is no restriction on the data and the extension of the Green's function to negative s merely involves letting s go to $-s$ in (15) or (16). The appropriate choice for the Green's function for negative s, $2s+n\geqslant 0$, is

$$K_{s,-(s+n)}(\zeta,\bar\zeta;\eta,\bar\eta) = \frac{(-1)^{n-1}}{4\pi(n-1)!} \frac{(1+\zeta\bar\eta)^{2|s|-n}(\eta-\zeta)^{n-1}}{(1+\zeta\bar\zeta)^{|s|-1}(1+\eta\bar\eta)^{|s|-1}(\bar\eta-\bar\zeta)}\,, \tag{17}$$

or, in terms of the null tetrad,

$$K^-{}_{s,-(s+n)}(\zeta,\bar\zeta;\eta,\bar\eta) = \frac{(-1)^{s+n-1}}{4\pi(n-1)!} \frac{[\bar m^a(\zeta,\bar\zeta)m_a(\eta,\bar\eta)]^{|s|-1}\bar m^b(\zeta,\bar\zeta)l_b(\eta,\bar\eta)[l^c(\zeta,\bar\zeta)\bar m_c(\eta,\bar\eta)]^{n-1}}{l^d(\zeta,\bar\zeta)l_d(\eta,\bar\eta)[l^e(\zeta,\bar\zeta)\eta_e(\eta,\bar\eta)]^{n-1}}\,. \tag{18}$$

The null tetrad version shows that the extension to negative s affects the term that describes the generalization of the Green's function to higher spin weights, the core Green's function, and the term that describes the generalization to higher orders of the edh operator. Again, the null tetrad version (18) is defined only for integer spin weight. The analog of the kernel exclusion property for these Green's functions is the property that if inappropriate initial data $A_{s+n}(\zeta,\bar\zeta)$ is chosen, that is, initial data that possesses in its spherical harmonic decomposition l values that do not yield regular solutions, the "solution," (14), yields zero. A brief treatment of this is given in Appendix B. Note that when $2s+n=0$, the expressions for the Green's functions for negative spin weights, (15) and (17), are equal.

The Green's functions for edh-bar can also be obtained, thus providing solutions for equations of the form

$$\bar\eth_\zeta{}^n F_s(\zeta,\bar\zeta) = A_{s-n}(\zeta,\bar\zeta)\,, \tag{19}$$

where $F_s(\zeta,\bar\zeta)$ is of spin-weight s, s negative, and $A_{s-n}(\eta,\bar\eta)$ is of spin-weight $s-n$. The solution to (19) can be written as

$$F_s(\zeta,\bar\zeta) = \int_{S^2} \bar K^-{}_{s,-(s-n)}(\zeta,\bar\zeta;\eta,\bar\eta) A_{s-n}(\eta,\bar\eta)\, d\mu_\eta\,, \tag{20}$$

where the Green's function of $\bar\eth_\zeta{}^n$, $\bar K^-{}_{s,-(s-n)}(\zeta,\bar\zeta;\eta,\bar\eta)$ is of spin-weight s in $(\zeta,\bar\zeta)$ and of spin-weight $-(s-n)$ in $(\eta,\bar\eta)$ and is obtained by taking the complex conjugate of expressions (15) and (16) and letting s go to $-s$,

$$\bar{K}^{-}{}_{s,-(s-n)}(\zeta,\bar{\zeta};\eta,\bar{\eta}) = \frac{(-1)^{n-1}}{4\pi(n-1)!}\,\frac{(1+\zeta\bar{\eta})^{-2s+n}(\bar{\eta}-\bar{\zeta})^{n-1}}{(1+\zeta\bar{\zeta})^{-s+n-1}(1+\eta\bar{\eta})^{-s+n-1}(\eta-\zeta)} \tag{21}$$

or

$$\bar{K}^{-}{}_{s,-(s-n)}(\zeta,\bar{\zeta};\eta,\bar{\eta}) = \frac{(-1)^{s+n-1}}{4\pi(n-1)!}\,\frac{[\bar{m}^{a}(\zeta,\bar{\zeta})m_{a}(\eta,\bar{\eta})]^{-s}\,l^{b}(\zeta,\bar{\zeta})m_{b}(\eta,\bar{\eta})[l^{c}(\zeta,\bar{\zeta})m_{c}(\eta,\bar{\eta})]^{n-1}}{l^{d}(\zeta,\bar{\zeta})l_{d}(\eta,\bar{\eta})}. \tag{22}$$

Expressions (21) and (22) are also valid for positive s when $2s-n\leqslant 0$. These Green's functions can also be extended to positive s, $2s-n\geqslant 0$, and are obtained by taking the complex conjugate of expressions (18) and (19) and letting $|s|$ go to s;

$$\bar{K}^{+}{}_{s,-(s-n)}(\zeta,\bar{\zeta};\eta,\bar{\eta}) = \frac{(-1)^{n-1}}{4\pi(n-1)!}\,\frac{(1+\bar{\zeta}\eta)^{2s-n}(\bar{\eta}-\bar{\zeta})^{n-1}}{(1+\zeta\bar{\zeta})^{s-1}(1+\eta\bar{\eta})^{s-1}(\eta-\zeta)} \tag{23}$$

or

$$\bar{K}^{+}{}_{s,-(s-n)}(\zeta,\bar{\zeta};\eta,\bar{\eta}) = \frac{(-1)^{s+n-1}}{4\pi(n-1)!}\,\frac{[m^{a}(\zeta,\bar{\zeta})\bar{m}_{a}(\eta,\bar{\eta})]^{s-1}m^{b}(\zeta,\bar{\zeta})l_{b}(\eta,\bar{\eta})[l^{c}(\zeta,\bar{\zeta})m_{c}(\eta,\bar{\eta})]^{n-1}}{l^{d}(\zeta,\bar{\zeta})l_{d}(\eta,\bar{\eta})[l^{e}(\zeta,\bar{\zeta})n_{e}(\eta,\bar{\eta})]^{n-1}}. \tag{24}$$

For $2s-n=0$, the expressions for the Green's functions for positive spin weight, (21) and (23), are equal.

Having obtained the Green's functions for both the edh and edh-bar operators for arbitrary positive or negative spin weights, the Green's functions for any combination of edh and edh-bar operators can be obtained from integrals of products of the individual Green's functions for the edh and edh-bar operators. For example, consider the following equation:

$$\eth_{\zeta}{}^{n}\bar{\eth}_{\zeta}{}^{m}F_{s}(\zeta,\bar{\zeta}) = A_{s+n-m}(\zeta,\bar{\zeta}), \tag{25}$$

where $F_{s}(\zeta,\bar{\zeta})$ is a regular function of spin-weight s and $A_{s+n-m}(\zeta,\bar{\zeta})$ is a regular function of spin-weight $s+n-m$ and for convenience s is positive and $s\geqslant m$. (Other cases could easily be considered.) The solution to (25) can be written as

$$F_{s}(\zeta,\bar{\zeta}) = \int_{S^2}\hat{K}^{+}{}_{s,-(s+n-m)}(\zeta,\bar{\zeta};\eta,\bar{\eta}) \times A_{s+n-m}(\eta,\bar{\eta})\,d\mu_{\eta}, \tag{26}$$

where $\hat{K}^{+}{}_{s,-(s+n-m)}(\zeta,\bar{\zeta};\eta,\bar{\eta})$ is the Green's function of the $\eth_{\zeta}{}^{n}\bar{\eth}^{m}$ operator and is spin-weight s in $(\zeta,\bar{\zeta})$ and of spin-weight $-(s+n-m)$ in $(\eta,\bar{\eta})$. Now consider the equation

$$\bar{\eth}_{\zeta}{}^{m}F_{s}(\zeta,\bar{\zeta}) = H_{s-m}(\zeta,\bar{\zeta}), \tag{27}$$

which is of the form (19) and whose solution (13) is given by

$$F_{s}(\zeta,\bar{\zeta}) = \int_{S^2}\bar{K}^{+}{}_{s,-(s-m)}(\zeta,\bar{\zeta};\eta,\bar{\eta})H_{s-m}(\eta,\bar{\eta})\,d\mu_{\eta}, \tag{28}$$

where the Green's function of the $\bar{\eth}_{\zeta}{}^{m}$ operator is given by (23) or (24). Now substituting (27) into (25) yields the equation

$$\eth_{\zeta}{}^{n}H_{s-m}(\zeta,\bar{\zeta}) = A_{s+n-m}(\zeta,\bar{\zeta}), \tag{29}$$

which is of the form (13) and whose solution (14) is given by

$$H_{s-m}(\zeta,\bar{\zeta}) = \int_{S^2}K^{+}{}_{s-m,-(s+n-m)}(\zeta,\bar{\zeta};\eta,\bar{\eta}) \times A_{s+n-m}(\eta,\bar{\eta})\,d\mu_{\eta}, \tag{30}$$

where the Green's function of the $\eth_{\zeta}{}^{n}$ operator is given by (15) or (16). Substituting (30) into (28) gives an expression for $F_{s}(\zeta,\bar{\zeta})$ in terms of the Green's functions of the $\bar{\eth}_{\zeta}{}^{m}$ and the $\eth_{\zeta}{}^{n}$ operators

$$F_{s}(\zeta,\bar{\zeta}) = \int_{S^2}\int_{S^2}\bar{K}^{+}{}_{s,-(s-m)}(\zeta,\bar{\zeta};\eta,\bar{\eta}) \times K^{+}{}_{s-m,-(s+n-m)}(\eta,\bar{\eta};\phi,\bar{\phi}) \times A_{s+n-m}(\phi,\bar{\phi})\,d\mu_{\eta}\,d\mu_{\phi}. \tag{31}$$

Comparing this expression to (26) produces a form for the Green's function of the $\eth_{\zeta}{}^{n}\bar{\eth}_{\zeta}{}^{m}$ operator,

$$\hat{K}^{+}{}_{s,-(s+n-m)}(\zeta,\bar{\zeta};\eta,\bar{\eta}) = \int_{S^2}\bar{K}^{+}{}_{s,-(s-m)}(\zeta,\bar{\zeta};\phi,\bar{\phi}) \times K^{+}{}_{s-m,-(s+n-m)}(\phi,\bar{\phi};\eta,\bar{\eta})\,d\mu_{\phi}. \tag{32}$$

This procedure can be applied to any combination of edh and edh-bar operators acting on arbitrary spin weight and so, in principle, the knowledge of a relatively few simple forms for the Green's functions associated with the edh and edh-bar operators allows one to calculate the Green's function for any combination of edh, edh-bar operators. In practice, this integration procedure can be both difficult and tedious.

The transformation properties of the Green's functions $K^{\pm}{}_{s,-(s+n)}(\zeta,\bar{\zeta};\eta,\bar{\eta})$ and $\bar{K}^{\pm}{}_{s,-(s-n)}(\zeta,\bar{\zeta};\eta,\bar{\eta})$ are now considered. Under ordinary coordinate transformations, the Green's functions transform as scalar functions and under tetrad rotations the Green's functions, having definite spin-weight s in $(\zeta,\bar{\zeta})$ and $-(s\pm n)$ in $(\eta,\bar{\eta})$, transform as proper spin-weighted functions. Due to these transformation properties of the Green's functions, the solutions to (13) and (19), $F_{s}(\zeta,\bar{\zeta})$, given by (14) and (20) transform accordingly; as scalar functions under ordinary coordinate transformations and as spin-weight s functions under tetrad rotations. While a particular coordinate representation was chosen in writing down the expressions for the Green's functions, choosing a different coordinatization is equivalent to a combined coordinate transformation and tetrad rotation yielding a suitably transformed Green's function and solution.

The primary motivation for considering these Green's functions comes from attempts to solve the matrix-valued Sparling equation with triangular initial data,[7] but the development of the machinery of the Green's functions has been of use in other investigations.[8,9] The edh operator and its corresponding Green's functions are intimately linked to the D-bar calculus[10]; methods and procedures developed for one area may prove useful in the other. The study of the structure of the edh operator is continuing[11] and some questions concerning these Green's functions are still outstanding.

ACKNOWLEDGMENT

The authors wish to thank Professor V. Ogievetsky, Professor A. Galperin, and Professor E. Ivanov.

APPENDIX A: EXISTENCE OF GREEN'S FUNCTIONS

In this section a proof that the expression (15) given in Sec. IV is the Green's function for the $\eth^n$ operator is outlined. The proof is accomplished by first showing that (15) for $n=1$, $K^+{}_{s,-(s+1)}(\zeta,\bar\zeta;\eta,\bar\eta)$ is the Green's function for the $\eth$ operator. It is then shown that for arbitrary n, $K^+{}_{s',-(s'+n-1)}(\zeta,\bar\zeta;\eta,\bar\eta)$ can be obtained by the application of the edh operator to $K^+{}_{s,-(s+n)}(\zeta,\bar\zeta;\eta,\bar\eta)$; therefore by $n-1$ applications of the edh operator, $K^+{}_{s'',-(s''+1)}(\zeta,\bar\zeta;\eta,\bar\eta)$ is obtained. This then completes the inductive proof that $K^+{}_{s,-(s+n)}(\zeta,\bar\zeta;\eta,\bar\eta)$ is the Green's function for the $\eth^n$ operator.

In order to show that the expressions given are indeed Green's functions of the edh operators, consider first the Green's function, expression (15) for $n=1$

$$K^+{}_{s,-(s+1)}(\zeta,\bar\zeta;\eta,\bar\eta) = \frac{1}{4\pi}\frac{(1+\bar\zeta\eta)^{2s+1}}{(1+\zeta\bar\zeta)^s(1+\eta\bar\eta)^s(\bar\eta-\bar\zeta)}. \tag{A1}$$

The solution to the equation

$$\eth_s F_s(\zeta,\bar\zeta) = A_{s+1}(\zeta,\bar\zeta) \tag{A2}$$

in terms of the integration over the sphere of this Green's function is written

$$\hat F_s(\zeta,\bar\zeta) = \int_{S^2} K^+{}_{s,-(s+1)}(\zeta,\bar\zeta;\eta,\bar\eta)A_{s+1}(\eta,\bar\eta)\,d\mu_\eta\,, \tag{A3}$$

the spin-weight s function $\hat F_s(\zeta,\bar\zeta)$ is to have no $l=s$ part in its spherical harmonic decomposition. Substituting (A2) into (A3) gives

$$\hat F_s(\zeta,\bar\zeta) = \int_{S^2} K^+{}_{s,-(s+1)}(\zeta,\bar\zeta;\eta,\bar\eta)\eth_\eta F_s(\eta,\bar\eta)\,d\mu_\eta\,, \tag{A4}$$

where the spin-weight s function $F_s(\zeta,\zeta)$ contained within the integral on the rhs of (A4) may contain an $l=s$ part. Explicitly writing the expression for $\eth_\eta F_s(\eta,\bar\eta)$ and substituting (A1) into (A4) yields

$$\hat F_s(\zeta,\bar\zeta) = \int_{S^2}\frac{(1+\bar\zeta\eta)^{2s+1}\partial_\eta[(1+\eta\bar\eta)^sF_s(\eta,\bar\eta)]}{4\pi(1+\zeta\bar\zeta)^s(1+\eta\bar\eta)^{2s+1}(\bar\eta-\bar\zeta)} \times d\eta\wedge d\bar\eta\,. \tag{A5}$$

The expression (A5) can be written (via an integration by parts) as the sum of two terms; the first term being the integral of

$$\eth_\eta[K^+{}_{s,-(s+1)}(\zeta,\bar\zeta;\eta,\bar\eta)F_s(\eta,\bar\eta)]$$

over the sphere, which can be evaluated using a complex version of Green's theorem, and the second term being the integral of

$$\eth_\eta[K^+{}_{s,-(s+1)}(\zeta,\bar\zeta;\eta,\bar\eta)]F_s(\eta,\bar\eta)$$

over the sphere. Note that the integral of

$$\eth_\eta[K^+{}_{s,-(s+1)}(\zeta,\bar\zeta;\eta,\bar\eta)F_s(\eta,\bar\eta)]$$

over the sphere would be zero if the quantity inside the square brackets was regular, but such is not the case. A more detailed calculation of this type can be found in Ref. 1. The expression (A5) becomes

$$\hat F_s(\zeta,\bar\zeta) = F_s(\zeta,\bar\zeta) - \int_{S^2}\frac{1}{4\pi}\frac{(2s+1)(1+\bar\zeta\eta)^{2s}}{[(1+\zeta\bar\zeta)^s(1+\eta\bar\eta)^s]}F_s(\eta,\bar\eta)\,d\mu_\eta$$

or

$$\hat F_s(\zeta,\bar\zeta) = F_s(\zeta,\bar\zeta) - \int_{S^2} P_s(\zeta,\bar\zeta;\eta,\bar\eta)\,F_s(\eta,\bar\eta)\,d\mu_\eta\,. \tag{A6}$$

Note that the first term on the rhs of (A6) may contain an $l=s$ part.

Contained within the integrand of the second term of (A6) is the function $P_s(\zeta,\bar\zeta;\eta,\bar\eta)$; this term is of spin-weight s in $(\zeta,\bar\zeta)$ and $-s$ in $(\eta,\bar\eta)$ and is annihilated by edh taken with respect to ζ and therefore contains only an $l=s$ part in its spherical harmonic decomposition in $(\zeta,\bar\zeta)$ and is annihilated by edh-bar taken with respect to η and therefore contains only an $l=-s$ part in its spherical harmonic decomposition in $(\eta,\bar\eta)$. This term can be expressed, using the definition of the spin-s spherical harmonics given in (7), as

$$P_s(\zeta,\bar\zeta;\eta,\bar\eta) = \sum_m {}_sY_{sm}(\zeta,\bar\zeta)\,{}_{-s}\bar Y_{sm}(\eta,\bar\eta)\,. \tag{A7}$$

The expression (A7) is a projection operator for the $l=s$ parts of the spherical harmonic decomposition of a spin-weight s function; when any spin-weight s function α_s is integrated against (A7) over the sphere, the result is the $l=s$ parts of the spherical harmonic decomposition of α_s.

The second term in (A6) projects out the $l=s$ parts of the spherical harmonic decomposition of $F_s(\zeta,\bar\zeta)$ and therefore the rhs of (A6) is $F_s(\zeta,\bar\zeta)$ minus its $l=s$ part or exactly what was defined as $\hat F_s(\zeta,\bar\zeta)$. Given that (A3) holds as shown above, it follows that $K^+{}_{s,-(s+1)}(\zeta,\bar\zeta;\eta,\bar\eta)$ must satisfy

$$\eth_\zeta K^+{}_{s,-(s+1)}(\zeta,\bar\zeta;\eta,\bar\eta) = \delta_{s+1,-(s+1)}(\zeta,\bar\zeta;\eta,\bar\eta)\,, \tag{A8}$$

where δ is a distribution of spin-weight $s+1$ in $(\zeta,\bar\zeta)$ and $-(s+1)$ in $(\eta,\bar\eta)$ with singular support at $(\zeta,\bar\zeta)=(\eta,\bar\eta)$; therefore (A1) is the Green's function for the $n=1$ edh operator.

Having established the validity of the expression (8) for the $n=1$ edh operator, the expression can be shown to be valid for arbitrary higher-order edh operators. Consider acting with the operator $\eth_\zeta$ on (15),

$$\eth_\zeta K^+_{s,-(s+n)}(\zeta,\bar\zeta;\eta,\bar\eta)=\eth\left(\frac{(-1)^{n-1}}{4\pi(n-1)!}\frac{(1+\bar\zeta\eta)^{2s+n}(\eta-\zeta)^{n-1}}{(1+\zeta\bar\zeta)^{s+n-1}(1+\eta\bar\eta)^{s+n-1}(\bar\eta-\bar\zeta)}\right),$$

and given the action of the edh operator on a spin-weight s function, (4), yields

$$\eth_\zeta K^+_{s,-(s+n)}(\zeta,\bar\zeta;\eta,\bar\eta)=K^+_{s+1,-(s+n)}(\zeta,\bar\zeta;\eta,\bar\eta)\,. \tag{A9}$$

Thus (A9) is expression (15) where $s\to s+1$ and $n\to n-1$; the action of edh on the Green's function is to decrease the order and increase the spin weight. Therefore the expression $\eth_\zeta{}^n K^+_{s,-(s+n)}(\zeta,\bar\zeta;\eta,\bar\eta)$ is equal to $\eth_\zeta K^+_{s+n-1,-(s+n)}(\zeta,\bar\zeta;\eta,\bar\eta)$; this is of the form of (A8) where $s\to s+n-1$ and $n\to 1$ and so

$$\eth_\zeta{}^n K^+_{s,-(s+n)}(\zeta,\bar\zeta;\eta,\bar\eta)=\delta_{s+n,-(s+n)}(\zeta,\bar\zeta;\eta,\bar\eta)\,, \tag{A10}$$

where δ is a distribution of spin-weight $s+n$ in $(\zeta,\bar\zeta)$ and $-(s+n)$ in $(\eta,\bar\eta)$ with singular support at $(\zeta,\bar\zeta)=(\eta,\bar\eta)$; therefore (15) is the Green's function for the order n edh operator. Similarly the expressions (17), (21), and (23) can be shown to be Green's functions; the technical details of these proofs follow closely those of the proof given above.

APPENDIX B: THE KERNEL EXCLUSION PROPERTY

This section provides a proof that the Green's function $K^+_{s,-(s+n)}(\zeta,\bar\zeta;\eta,\bar\eta)$ given in (15) possesses the kernel exclusion property; in order to show this it is necessary and sufficient to show that the Green's functions $K^+_{s,-(s+n)}(\zeta,\bar\zeta;\eta,\bar\eta)$ have no $l\in\{s,\ldots,s+n-1\}$ part in their spherical harmonic decomposition in $(\zeta,\bar\zeta)$. This can be accomplished by evaluating the integral

$$I=\int_{S^2}K^+_{s,-(s+n)}(\zeta,\bar\zeta;\eta,\bar\eta)\,{}_s\bar Y_{lm}(\zeta,\bar\zeta)\,d\mu_\zeta\,, \tag{B1}$$

where ${}_s\bar Y_{lm}(\zeta,\bar\zeta)$ is the spin-s spherical harmonic given by

$${}_s\bar Y_{lm}(\zeta,\bar\zeta)=A_{slm}\sum_p(-1)^p\binom{l-s}{p}\binom{l+s}{p+s-m}\frac{\bar\zeta^p\zeta^{p+s-m}}{(1+\zeta\bar\zeta)^l} \tag{B2}$$

with

$$A_{slm}=(-1)^{l+s}[[(l+m)!(l-m)!(2l+1)]\times[4\pi(l-s)!(l+s)!]^{-1}]^{1/2},$$

and showing that this integral yields zero for $l\in\{s,\ldots,s+n-1\}$. The remainder of this appendix provides the technical details for the evaluation of (B1).

Substituting (15) and (B2) into (B1) gives

$$I=\frac{(-1)^{n-1}A_{slm}}{(2\pi i)(n-1)!(1+\eta\bar\eta)^{s+n-1}}\times\sum_p(-1)^p\binom{l-s}{p}\binom{l+s}{p+s-m}I_p\,, \tag{B3}$$

where

$$I_p=\int_{S^2}\frac{\bar\zeta^p\zeta^{p+s-m}(1+\bar\zeta\eta)^{2s+n}(\eta-\zeta)^{n-1}}{(1+\zeta\bar\zeta)^{s+n+l+1}(\bar\eta-\bar\zeta)}\,d\zeta\wedge d\bar\zeta\,. \tag{B4}$$

The transformation $\eta\to-\eta$ and $\bar\eta\to-\bar\eta$ is now made for convenience. Consider the following rotation on the sphere:

$$\zeta'=(\zeta-\eta)/(1+\zeta\bar\eta)\quad\text{and}\quad\bar\zeta'=(\bar\zeta-\bar\eta)/(1+\bar\zeta\eta)\,.$$

This transformation is applied to (B4) and the numerator of the integrand is expanded in powers of ζ, then I_p becomes

$$I_p=(-1)^n(1+\eta\bar\eta)^{s+n-l-1}\times\sum_{k=n-1}^{s+l+n-1}\int_{S^2}\frac{C_k\zeta^k(\bar\zeta-\bar\eta)^p(1+\bar\zeta\eta)^{l-s-p}}{\bar\zeta(1+\zeta\bar\zeta)^{s+n+l+1}}\times d\zeta\wedge d\bar\zeta\,, \tag{B5}$$

where the coefficients of the expansion C_k need not be specified explicitly. This integral can be more easily evaluated by considering the following change of variables $\xi=1+\zeta\bar\zeta$, $d\xi=\bar\zeta\,d\zeta$, and $\zeta=(\xi-1)/\bar\zeta$. Under this change of variables, (B5) becomes

$$I_p=(-1)^n(1+\eta\bar\eta)^{s+n-l}\times\sum_{k=n-1}^{s+l+n-1}\int_{S^2}\frac{C_k(\bar\zeta-\bar\eta)^p(1+\bar\zeta\eta)^{l-s-p}(\xi-1)^k}{\bar\zeta^{k+2}\xi^{s+n+l+3}}\times d\xi\wedge d\bar\zeta\,. \tag{B6}$$

The term $(\xi-1)^k$ in the numerator of (B6) can be expanded in a binomial expansion yielding

$$I_p=(-1)^n(1+\eta\bar\eta)^{s+n-l-1}\times\sum_{k=n-1}^{s+l+n-1}\sum_{q=0}^{k}C_k\binom{k}{q}(-1)^{k-q}\times\int_{S^2}\frac{(\bar\zeta-\bar\eta)^p(1+\bar\zeta\eta)^{l-s-p}}{\bar\zeta^{k+2}\xi^{s+n+l+3-q}}\,d\xi\wedge d\bar\zeta\,. \tag{B7}$$

The integral over $d\xi$ can now be evaluated giving

$$I_p=(-1)^n(1+\eta\bar\eta)^{s+n-l-1}\times\sum_{k=n-1}^{s+l+n-1}\sum_{q=0}^{k}\frac{C_k\binom{k}{q}(-1)^{k-q}}{s+n+l-q}\times\oint\frac{(\bar\zeta-\bar\eta)^p(1+\bar\zeta\eta)^{l-s-p}}{\bar\zeta^{k+2}(1+\zeta\bar\zeta)^{s+n+l+3-q}}\,d\bar\zeta\,. \tag{B8}$$

Note that the substitution $\xi=1+\zeta\bar\zeta$ was made in (B8). The numerator of the integrand of (B8) can be expanded in powers of $\bar\zeta$, where the coefficients of the expansion again need not be specified explicitly,

$$I_p=(-1)^n(1+\eta\bar\eta)^{s+n-l-1}\times\sum_{k=n-1}^{s+l+n-1}\sum_{q=0}^{k}\sum_{r=0}^{l-s}\frac{C_kC_r\binom{k}{q}(-1)^{k-q}}{s+n+l-q}\times\oint\frac{d\bar\zeta}{\bar\zeta^{k+2-r}(1+\zeta\bar\zeta)^{s+n+l+3-q}}\,. \tag{B9}$$

Now as l ranges from s to $s+n-1$, $l\in\{s,\ldots,s+n-1\}$. The smallest value of the power of $\bar{\zeta}$ in the denominator of the contour integral is $n+s+1-l$; the smallest power of $\bar{\zeta}$ is obtained by considering the value $l=s+n-1$, $n+s+1-l=2$. The evaluation of the contour integral $\oint d\bar{\zeta}/[\bar{\zeta}^a(1+\zeta\bar{\zeta})^b]$ gives a nonzero contribution only for $a=1$. Since the smallest value of the power of $\bar{\zeta}$ in the denominator of the contour integral I_p is greater than 1 for all values of s, n, $l\in\{s,\ldots,s+n-1\}$, and m, $I_p=0$. Therefore the evaluation of the integral (B1) yields

$$I=\int_{S^2} K^+{}_{s,-(s+n)}(\zeta,\bar{\zeta};\eta,\bar{\eta})\,{}_s\bar{Y}_{lm}(\zeta,\bar{\zeta})\,d\mu_\zeta=0$$

for all $s,n\geqslant 1$, $l\in\{s,\ldots,s+n-1\}$, and m. Since this integral is zero, by orthogonality of the spin-s harmonics, the Green's functions $K^+{}_{s,-(s+n)}(\zeta,\bar{\zeta};\eta,\bar{\eta})$ have no $l\in\{s,\ldots,s+n-1\}$ part in its spherical harmonic decomposition in $(\zeta,\bar{\zeta})$.

A procedure comparable to that given above can be used to show that the Green's function extended to negative s, $K^-{}_{s,-(s+n)}(\zeta,\bar{\zeta};\eta,\bar{\eta})$, given by expression (17), will yield zero for inappropriate initial data $A_{s+n}(\zeta,\bar{\zeta})$; this is the analogous property to the kernel exclusion property for positive s. This can be accomplished by evaluating the integral

$$I=\int_{S^2} K^-{}_{s,-(s+n)}(\zeta,\bar{\zeta};\eta,\bar{\eta})\,{}_{s+n}\bar{Y}_{lm}(\eta,\bar{\eta})\,d\mu_\eta \tag{B10}$$

and showing that this integral yields zero for $l\in\{|s+n|,\ldots,|s|-1\}$. The technical details for the evaluation of (B10) are similar to those given for the evaluation of (B1) and are not formally presented.

APPENDIX C: SPINORIAL REPRESENTATION

In this section, for completeness, the Green's functions of Secs. III and IV are reexpressed in terms of homogeneous coordinates on the sphere. The spinor dyad $(\pi_{A'},\eta_{A'})$ is chosen and a particular representation for the spinors in terms of the complex stereographic coordinates $(\zeta,\bar{\zeta})$ is given. The spinorial representations of the Green's functions are then obtained. Note that some knowledge of Lorentzian spinors is assumed and the presentation that follows is not intended to be complete.

First, the spinor dyad $(\pi_{A'},\eta_{A'})$ is chosen subject to the condition that

$$\eta^{A'}\pi_{A'}=1\,. \tag{C1}$$

Raising of the spinor index is accomplished by the use of the nonzero skew two-index spinor $\epsilon^{A'B'}$ (the choice $\epsilon^{0'1'}=1$ being used throughout), so that

$$\pi^{A'}=\epsilon^{A'B'}\pi_{B'}\,. \tag{C2}$$

[A particular representation for this dyad can be given in terms of the coordinates $(\zeta,\bar{\zeta})$ and/or $(\eta,\bar{\eta})$,

$$\begin{aligned}&\pi_{A'}=P^{-1/2}(1,\zeta)\,,\quad \pi^*{}_{A'}=P^{*-1/2}(1,\eta)\,,\\&\eta_{A'}=P^{-1/2}(-\bar{\zeta},1)\,,\quad \eta^*{}_{A'}=P^{*-1/2}(-\bar{\eta},1)\,,\end{aligned} \tag{C3}$$

with $P=1+\zeta\bar{\zeta}$ and $P^*=1+\eta\bar{\eta}$.]

Note that the spinor $\eta_{A'}$ can be expressed in terms of $\bar{\pi}_{A'}$;

$$\eta_{A'}=-\bar{\epsilon}^{AB}t_{A'B}\bar{\pi}_A\,, \tag{C4}$$

where $\bar{\epsilon}^{AB}$ is the complex conjugate of $\epsilon^{A'B'}$ and $t_{A'B}$ is the unit matrix. Note that the choice of $t_{A'B}$ picks out a unique timelike direction in Minkowski space.

Having obtained the representation of the spinors, the following products can be formed:

$$\pi_{A'}\pi^{*A'}=\bar{\eta}_A\bar{\eta}^{*A}=(\eta-\zeta)P^{-1/2}P^{*-1/2}\,,$$
$$\bar{\pi}_A\bar{\pi}^{*A}=\eta_{A'}\eta^{*A'}=(\bar{\eta}-\bar{\zeta})P^{-1/2}P^{*-1/2}\,,$$
$$\pi_{A'}\eta^{*A'}=-\bar{\eta}_A\bar{\pi}^{*A'}=(1+\zeta\bar{\eta})P^{-1/2}P^{*-1/2}\,,$$

and

$$\bar{\pi}_A\bar{\eta}^{*A}=-\eta_{A'}\pi^{*A'}=(1+\bar{\zeta}\eta)P^{-1/2}P^{*-1/2}\,. \tag{C5}$$

The Green's function can be reexpressed in terms of these products.

For the edh operators, $2s+n\geqslant 0$, expression (15),

$$\begin{aligned}K^+{}_{s,-(s+n)}&(\pi_{A'},\bar{\pi}_A;\pi^*_{A'},\bar{\pi}^*_A)\\&=(-1)^{n-1}/4\pi(n-1)![\bar{\pi}_A\bar{\eta}^{*A}]^{2s+n}\\&\quad\times[\pi_{A'}\pi^{*A'}]^{n-1}[\bar{\pi}_A\bar{\pi}^{*A}]^{-1}\,,\end{aligned} \tag{C6}$$

which is of homogeneity $n-1$ in $\pi_{A'}$, $2s+n-1$ in $\bar{\pi}_A$, $2s+2n-1$ in $\pi^*{}_{A'}$, and -1 in $\bar{\pi}^*{}_{A'}$, and for $2s+n\leqslant 0$, expression (17)

$$\begin{aligned}K^-{}_{s,-(s+n)}&(\pi_{A'},\bar{\pi}_A;\pi^*_{A'},\bar{\pi}^*_A)\\&=(-1)^{n-1}/4\pi(n-1)![\pi_{A'}\eta^{*A'}]^{-2s-n}\\&\quad\times[\pi_{A'}\pi^{*A'}]^{n-1}[\bar{\pi}_A\bar{\pi}^{*A}]^{-1}\,,\end{aligned} \tag{C7}$$

which is of homogeneity $-2s-1$ in $\pi_{A'}$, -1 in $\bar{\pi}_A$, $n-1$ in $\pi^*{}_{A'}$, and $-2s-n-1$ in $\bar{\pi}^*{}_A$.

For the edh-bar operators, $2s-n\leqslant 0$, expression (21),

$$\begin{aligned}\bar{K}^-{}_{s,-(s-n)}&(\pi_{A'},\bar{\pi}_A;\pi^*{}_{A'},\bar{\pi}^*{}_A)\\&=(-1)^n/4\pi(n-1)![\pi_{A'}\eta^{*A'}]^{-2s+n}\\&\quad\times[\bar{\pi}_A\bar{\pi}^{*A}]^{n-1}[\pi_{A'}\pi^{*A'}]^{-1}\,,\end{aligned} \tag{C8}$$

which is of homogeneity $-2s+n-1$ in $\pi_{A'}$, $n-1$ in $\bar{\pi}_A$, -1 in $\pi^*{}_{A'}$, and $-2s+2n-1$ in $\bar{\pi}^*{}_A$, and for $2s-n\geqslant 0$, expression (23),

$$\begin{aligned}\bar{K}^+{}_{s,-(s+n)}&(\pi_{A'},\bar{\pi}_A;\pi^*_{A'},\bar{\pi}^*_A)\\&=(-1)^{n-1}/4\pi(n-1)![\bar{\pi}_A\bar{\eta}^{*A}]^{2s-n}\\&\quad\times[\bar{\pi}_A\bar{\pi}^{*A}]^{n-1}[\pi_{A'}\pi^{*A'}]^{-1}\,,\end{aligned} \tag{C9}$$

which is of homogeneity -1 in $\pi_{A'}$, $2s-1$ in $\bar{\pi}_A$, $2s-n-1$ in $\pi^*{}_{A'}$, and $n-1$ in $\bar{\pi}^*{}_A$. Note that in these expressions $\eta^{*A'}$ and $\bar{\eta}^{*A}$ are used explicitly rather than reexpressing them in terms of $\pi^{*A'}$ and $\bar{\pi}^*{}_A$.

[1] J. R. Porter, "Green's functions associated with the edth operators," Gen. Relativ. Gravit. 13, 531 (1981).
[2] E. T. Newman and R. Penrose, "Note on the Bondi–Metzner–Sachs group," J. Math. Phys. 7, 863 (1966).
[3] W. D. Curtis and D. E. Lerner, "Complex line bundles in relativity," J. Math. Phys. 19, 874 (1978).
[4] M. Eastwood and P. Tod, "Edth-a differential operator on the sphere," Math. Proc. Cambridge Philos. Soc. 92, 317 (1982).
[5] T. Dray, "A unified treatment of Wigner D functions, spin-weighted spherical harmonics, and monopole moments," J. Math. Phys. 27, 781 (1986).
[6] C. N. Kozameh, E. T. Newman, and J. R. Porter, "Maxwell's equations, linear gravity, and twistors," Found. Phys. 14, 1061 (1984).
[7] S. Chakravarty and E. T. Newman, "Quadratures for self-dual GL(2,C)

Yang–Mills fields," J. Math. Phys. **28**, 334 (1987).
[8]A. Galperin, E. Ivanov, V. Ogievetsky, and E. Sokatchev, "Harmonic supergraphs: Feynman rules and examples," Class. Quantum Gravit. **2**, 617 (1985).
[9]A. Galperin, E. Ivanov, V. Ogievetsky, and E. Sokatchev, "Harmonic supergraphs. Green functions," Class. Quantum Gravit. **2**, 601 (1985).
[10]R. Beals and R. R. Coifman, "The *D*-bar approach to inverse scattering and nonlinear evolutions," *Proceedings of the Conference on Solitons and Coherent Structures* (North-Holland, Amsterdam, 1986), p. 242.
[11]J. R. Porter, "Edth and its inverses," University of Pittsburgh preprint.

Time-dependent inverse scattering from gradient-type interfaces using an exact solution

K. L. Kreider
Applied Mathematical Sciences, Ames Laboratory—U.S. Department of Energy, Iowa State University, Ames, Iowa 50011

(Received 4 September 1987; accepted for publication 7 September 1988)

In the context of one-dimensional linear wave propagation, a gradient-type interface is a point at which the velocity profile of a scattering medium suffers a jump in first derivative. In this paper, a time domain approach to scattering from such an interface leads to an integrodifferential equation involving the kernel of the reflection operator (impulse response), which can be solved exactly if the velocity profile is piecewise parabolic. This special case provides the basis for an approximate numerical scheme to solve the inverse scattering problem (recover the velocity profile from reflection data) for a general velocity profile. The algorithm is fast and affords good results if the velocity profile is nearly parabolic or the interest is in short time only. A detailed error estimate for the approximation is provided.

I. INTRODUCTION

Several authors have dealt with one-dimensional scattering problems for the time-dependent wave equation involving discontinuous material parameters in a variety of contexts.[1–9] Much of this work developed from Weston's time domain approach,[10–12] which dealt with continuous parameter functions. However, little has been done with gradient-type interfaces, points at which the velocity profile suffers a jump in first derivative.

In this paper, the only material parameter considered is the propagation velocity, which in its most general form can have a jump discontinuity as well as a jump in first derivative at an interface. However, the effect of the jump in velocity can be eliminated by techniques described in Ref. 9, so the velocity profile will be considered continuous, retaining, of course, the jump in derivative. In addition, only one interface, located at $z = 0$, will be considered.

This paper describes a method of solving the inverse scattering problem for gradient-type media in such a way that the jump value at the interface is very easily obtained. The inverse scattering problem is this: a known incident wave penetrates the inhomogeneous scattering medium, producing reflected waves. From the reflection data, the velocity function is recovered. The solution method involves finding an exact solution of the r equation[13–15] (an integrodifferential equation relating the incident and reflected fields) for a special parabolic velocity profile that matches the jump value at the interface. The algorithm computes the difference between the true profile and this special profile. In the process, the jump value at the interface is easily obtained.

Section II briefly summarizes the derivation of the r equation. In Sec. III, an exact solution of the equation, corresponding to a piecewise parabolic velocity profile, is presented. This exact solution is then used to derive an approximate form of the r equation. Section IV discusses the numerical implementation of the approximate algorithm in solving the inverse problem. The approximation makes the computations extremely simple. Section V contains the error analysis for the approximation, providing a bound on the error in the reconstructed velocity profile in terms of the given data for the inverse problem. Numerical examples are provided in Sec. VI to illustrate the strengths and weaknesses of the approximation.

II. FORMULATION OF THE PROBLEM

The scattering problem in this paper consists of a semi-infinite scattering medium in the region $z > 0$, in which the velocity profile $c(z)$ is arbitrary, with $c(z)$ constant (equal to c_0) in $z < 0$. The gradient-type interface at $z = 0$ is characterized by a jump in c_z, denoted $[c_z]$. In addition, c itself is everywhere continuous, while c_z is continuous everywhere except at the interface. The form of the wave equation used here is

$$u_{zz} - c^{-2}(z)u_{tt} = 0\,.$$

Introducing the travel time coordinate x puts the wave equation in the form

$$v_{xx} - v_{tt} - b(x)v_x = 0\,,$$

where

$$\begin{aligned} v(x,t) &= u(z,t)\,, \\ b(x) &= c_z(z)\,, \\ x(z) &= \int_0^z c^{-1}(\zeta)d\zeta\,. \end{aligned} \tag{2.1}$$

The goal is to solve the inverse problem; that is, to reconstruct $c(z)$ a specified distance Z into the medium given some knowledge of the wave field (a more precise definition is given in Sec. IV). The strategy is to let a known right-moving incident wave propagate from $z = -\infty$ so that it impinges upon the inhomogeneous medium beginning at time $t = 0$. In the process, a left-moving reflected wave is produced. To model this physical situation, the wave field v is split into components $v^{\pm}$ by the decomposition[13,14]

$$v^{\pm} = \tfrac{1}{2}(u \mp \partial_t^{-1} u_x)\,,$$

where

$$\partial_t^{-1} u_x(x,t) = \int_0^t u_x(x,\tau)d\tau\,.$$

 0022-2488/89/010053-06$02.50

In the homogeneous region, v^+ may be interpreted as a right-moving wave, while v^- is left-moving. The relation between v^+ and v^- in the inhomogeneous region is defined by the reflection operator[15–17] R:

$$v^-(x,t) = Rv^+(x,t) = r*v^+(x,t) = \int_0^t r(x,s)v^+(x,t-s)ds\,. \quad (2.2)$$

The kernel r represents the impulse response of the slab of inhomogeneous medium $[x,\infty]$. The convolution form of R is dictated by Duhamel's principle[18]: a continuous incident field can be treated as a sequence of time-delayed delta pulses; the reflected field is then the time-delayed sum (convolution) of impulse responses of these "incident pulses."

The equation that r satisfies is

$$r_x - 2r_t = \tfrac{1}{2}b(x)r*r\,, \quad (2.3)$$

with initial condition

$$r(x,0) = \tfrac{1}{4}b(x)\,.$$

Finally, since the medium is penetrated to depth Z, it is convenient to define

$$L = \int_0^Z c^{-1}(\zeta)d\zeta\,.$$

III. AN EXACT SOLUTION

In this section, an exact solution of the r equation is sought for a particular velocity profile of a form that is as yet unknown. Once found, this particular velocity profile will form the basis of a scheme to solve the inverse problem for any gradient-type velocity profile.

While (2.3) may admit an exact solution of arbitrary form for a given function b, it is most expedient to assume that r has the form $\gamma(x)e^{\gamma(x)t}$ in order to simplify the convolution term. Substituting this expression into (2.3) yields

$$(\gamma_x + t\gamma_x\gamma - 2\gamma^2)e^{\gamma t} = \tfrac{1}{2}b\gamma^2te^{\gamma t}\,,$$

from which the following relations appear:

$$\gamma_x = 2\gamma^2, \quad \gamma_x = \tfrac{1}{2}b\gamma\,.$$

Equating these two expressions gives $\gamma = \frac{1}{4}b$, but, since an equation for c is desired, this value of γ is substituted back into either equation. Since $b_x = cc_{zz}$, the differential equation $c_{zz} = \frac{1}{2}c_z^2$ is obtained, the solution of which is $c(z) = c_0(1-\alpha z)^2$. This is the particular velocity profile desired. Denote it by

$$c_p(z) = c_0(1-\alpha z)^2\,,$$

and the corresponding reflection kernel by

$$r_p(x,t) = \gamma(x)e^{\gamma(x)t}\,,$$

where

$$\gamma(x) = \begin{cases} -c_0\alpha(1-\alpha z(x))/2, & \text{for } x\geqslant 0;\\ \gamma(0), & \text{for } x<0\,. \end{cases}$$

Note that $r_p(0,0) = \gamma(0) = [c_z]$, so the strength of the jump discontinuity at the interface is immediately available.

When the velocity profile is not parabolic, it and the reflection kernel may be written as follows:

$$c(z) = c_p(z) + \partial c(z)\,,$$
$$r(x,t) = r_p(x,t) + \partial r(x,t)\,. \quad (3.1)$$

It is important to set $\partial c(0) = 0$ and $\partial c_z(0) = 0$ so that the jump in c_z at the interface is contained completely within the parabolic portion of the profile.

Substituting (3.1) into the r equation yields the equation for ∂r,

$$\partial r_x - 2\partial r_t = \tfrac{1}{2}a(x)\gamma^2(x)e^{\gamma(x)t} + b(x)(r_p*\partial r) + \tfrac{1}{2}b(x)(\partial r*\partial r)\,, \quad (3.2)$$

with initial condition

$$\partial r(x,0) = \tfrac{1}{4}a(x)\,, \quad (3.3)$$

where

$$a(x) = \partial c_z(z(x))\,.$$

The equation holds for $0\leqslant x\leqslant L$ and $t\geqslant 0$.

At this point, the ∂r equation is ready for numerical solution. However, instead of solving the equation directly, the solution of an approximate equation will be considered. The equation is

$$\partial r_x^* - 2\,\partial r_t^* = 0, \quad \partial r^*(x,0) = \tfrac{1}{4}a(x)\,. \quad (3.4)$$

Here and in the remainder of the paper, all starred quantities are the approximations (computed using the approximate ∂r equation) of the respective unstarred quantities.

There are several reasons for using this approximate equation. First, the computation time is drastically reduced. Solving the full equation takes $O(N^3)$ operations, where N is the number of grid points; the approximate equation leads to an $O(N)$ algorithm. Second, solving the full equation is expected to yield nearly perfect results for the inverse problem, as is the case in previous work[1,8,17] on similar problems. Therefore, these two methods form a pair of extremes—fastest but least accurate versus slowest but most accurate—to judge other methods by. Any other approximate method [for example, retaining only the linear terms in (3.2)] will give results intermediate to those obtained by the extremes. Finally, making the approximation provides an opportunity to examine error estimates for the r equation, little of which has been done before.

Equation (3.4) is the one used in the numerical implementation and error analysis of the next two sections.

IV. SOLUTION OF THE INVERSE PROBLEM

The inverse problem is defined as follows: given the reflection kernel $\partial r(0,t)$, solve the ∂r equation for $\partial r(x,0)$, and use the initial condition on ∂r to calculate $\partial c(z)$ and $z(x)$ and hence $c(z)$. Assume that c_0 is known.

In applications, ideally one would measure the reflected wave and deconvolve (2.2) to retrieve the total reflection kernel $r(0,t)$, and then subtract $r_p(0,t)$ to obtain $\partial r(0,t)$. Since $\alpha = -2r(0,0)/c_0$, $r_p(x,0)$ could be computed as $z(x)$ is found, allowing the calculation of $\partial r(0,t)$ as needed. In addition, $[c_z]$ could be found directly, since $r(0,0) = r_p(0,0) = [c_z]$. In practice, this process is not always easily done. Here, $\partial r(0,t)$ was computed directly by discretizing the full ∂r equation (3.2) using the initial condition (3.3) with a specified ∂c.

The discretization employs a regular grid with $\Delta x = h$, $\Delta t = 2h$, and $h = 1/N$. The number of grid points along the x axis is $N+1$. Denote $\partial r_{i,j} = \partial r(x_i,t_j)$ with $x_i = ih$ and $t_j = 2jh$. For the inverse problem, then, (3.4) implies that

$$\partial r_{i,0} = \partial r_{0,i}\,.$$

Since the required ∂r values are obtained without any computation, the approximate inverse algorithm may be considered to be simply the reconstruction of $z(x)$ and $c(z)$. Approximating the travel-time integral (2.1) by the trapezoid rule gives

$$x_i - x_{i-1} = h = \tfrac{1}{2}(z_i - z_{i-1})(1/c_i + 1/c_{i-1})\,. \tag{4.1}$$

Also,

$$c_i = c_0(1-\alpha z_i)^2 + \partial c_i\,, \tag{4.2}$$

$$\partial c_i = \partial c_{i-1} + 4(z_i - z_{i-1})\partial r_{i,0}\,. \tag{4.3}$$

Equation (4.3) comes from a backward-difference approximation to the derivative ∂c_z in the initial condition (3.3).

Combining (4.1), (4.2), and (4.3) leads to the nonlinear equation for z_i,

$$2h = (z_i - z_{i-1})\left(\frac{1}{c_{i-1}} + \frac{1}{[c_0(1-\alpha z_i)^2 + \partial c_{i-1} + 4(z_i - z_{i-1})\partial r_{i,0}]}\right).$$

Starting with $z_0 = 0$, $\partial c_0 = 0$, and a given c_0, the equation is solved iteratively. Once z_i has been computed (by any standard nonlinear equation solver), it is a simple task to calculate c_i and ∂c_i.

V. ERROR ESTIMATE FOR THE APPROXIMATION

When the ∂r equation is approximated as in (3.4), it is useful to find a bound on $|c(z) - c^*(z^*)|$. Consider z and z^* to be fixed and correspond to the same value of x. Again, starred quantities indicate approximated values.

In the analysis that follows, any discretization error that might occur is ignored, most notably in the approximation of integrals by the trapezoid rule. Also, let $\alpha > 0$. The only modification needed for the case $\alpha < 0$ is noted directly below (5.1).

The first step is to split c into its component parts:

$$|c(z) - c^*(z^*)| \leqslant |c_p(z) - c_p(z^*)| + |\partial c(z) - \partial c^*(z^*)|\,.$$

The bound for the first term on the right is simple:

$$\begin{aligned}
&|c_p(z) - c_p(z^*)|\\
&\quad = c_0|(1-\alpha z)^2 - (1-\alpha z^*)^2|\\
&\quad = c_0|(1-\alpha z) - (1-\alpha z^*)|\,|(1-\alpha z) + (1-\alpha z^*)|\\
&\quad \leqslant c_0|(1-\alpha z) - (1-\alpha z^*)|\,|1+1|\\
&\quad = 2\alpha c_0|z - z^*|\,.
\end{aligned} \tag{5.1}$$

If $\alpha < 0$, this bound changes in that

$$|(1-\alpha z) + (1-\alpha z^*)| < (1+|\alpha|Z) + (1+|\alpha|z^*)\,,$$

so the coefficient 2 is replaced by $2 + |\alpha|(Z + z^*)$. All subsequent analysis follows as presented.

In order to bound $|z - z^*|$, note that

$$x = \int_0^z \frac{d\xi}{c(\xi)}\,, \tag{5.2}$$

$$x = \int_0^{z^*} \frac{d\xi}{c^*(\xi)}\,. \tag{5.3}$$

Ignoring error due to application of the trapezoid rule, these become

$$2x = z(1/c(z) + 1/c_0)\,, \tag{5.4}$$

$$2x = z^*(1/c^*(z^*) + 1/c_0)\,. \tag{5.5}$$

Subtract (5.5) from (5.4) and rearrange terms to get

$$0 = (z - z^*)\left(\frac{1}{c(z)} + \frac{1}{c_0}\right) - z^*\left(\frac{1}{c^*(z^*)} - \frac{1}{c(z)}\right),$$

which leads to the bound

$$\begin{aligned}
|z - z^*| &= \frac{z^* c_0|c(z) - c^*(z^*)|}{c^*(z^*)(c(z) + c_0)}\\
&\leqslant \frac{z^*|c(z) - c^*(z^*)|}{c^*(z^*)}\,.
\end{aligned}$$

Using this inequality, (5.1) may now be written as

$$|c_p(z) - c_p(z^*)| \leqslant \frac{2\alpha c_0 z^*|c(z) - c^*(z^*)|}{c^*(z^*)}\,. \tag{5.6}$$

Next, consider $|\partial c(z) - \partial c^*(z^*)|$. Finding a bound here is much more difficult than in the previous case. First, a bound is obtained for $|\partial r(x,0) - \partial r^*(x,0)|$, which is then related to the derivatives of ∂c and ∂c^* by the initial conditions (3.3) and (3.4). After some manipulation, the expression is integrated to give the desired bound.

To obtain the bound on $|\partial r(x,0) - \partial r^*(x,0)|$ begin by writing the full ∂r equation in integrated form

$$\begin{aligned}
&\partial r(x,t)\\
&\quad = \partial r(0,t+2x) + 2\int_0^x \partial r(y,0)\gamma^2(y)e^{\gamma(y)(t+2(x-y))}\,dy\\
&\qquad + 4\int_0^x (\gamma(y) + \partial r(y,0))\\
&\qquad \times (r_p * \partial r)(y,t+2(x-y))dy\\
&\qquad + 2\int_0^x (\gamma(y) + \partial r(y,0))\\
&\qquad \times (\partial r * \partial r)(y,t+2(x-y))dy\,,
\end{aligned} \tag{5.7}$$

where $a(x)$ and $b(x)$ have been replaced by $4\partial r(x,0)$ and $4(\gamma(x) + \partial r(x,0))$, respectively.

Write the approximate equation as

$$\partial r^*(x,t) = \partial r^*(0,t+2x)\,. \tag{5.8}$$

The boundary condition is the same for both equations:

$$\partial r(0,t) = \partial r^*(0,t)\,,$$

so subtracting (5.8) from (5.7) and letting $R(x,t) = \partial r(x,t) - \partial r^*(x,t)$ gives

$$R(x,t) = 2\int_0^x (R(y,0) + \partial r^*(y,0))\gamma^2(y)\, e^{\gamma(y)(t+2(x-y))}\, dy$$
$$+ \int_0^x (\gamma(y) + R(y,0) + \partial r^*(y,0))$$
$$\times [4(r_p * \partial r)(y,t + 2(x-y))$$
$$+ 2(\partial r * \partial r)(y,t + 2(x-y))]dy. \qquad (5.9)$$

With $T = \{(x,t): x \geqslant 0, t \geqslant 0, t + 2x \leqslant 2L\}$, define

$$P = \max_{t\in[0,2L]} |\partial r^*(0,t)|, \qquad (5.10)$$

$$m = \max_{(x,t)\in T} |R(x,t)|. \qquad (5.11)$$

P is known and M is desired. Now let

$$H = \max_{x'\in[0,x]} |\gamma(x')|,$$

$$K = \min_{x'\in[0,x]} |\gamma(x')|.$$

With these definitions, (5.9) implies, with $\beta = t + 2x - \tau$,

$$|R(x,t)| \leqslant 2(M+P)\int_0^x \gamma^2(y) e^{\gamma(y)(t+2(x-y))}\, dy$$
$$+ (M+P-K)\int_0^x \int_0^{t+2x} |R(y,\beta) + \partial r^*(y,\beta)|$$
$$\times [4|\gamma(y)|e^{\gamma(y)\tau} + 2|R(y,\tau) + \partial r^*(y,\tau)|]d\tau\, dy.$$

Now

$$\int_0^x \gamma^2(y) e^{\gamma(y)(t+2(x-y))}\, dy \leqslant H^2 \int_0^L e^{\gamma(y)(t+2(x-y))}\, dy$$
$$\leqslant H^2 L,$$

since the argument of the exponential is negative. In addition,

$$\int_0^x \int_0^{t+2x} |\gamma(y)|e^{\gamma(y)\tau}\, d\tau\, dy = \int_0^x 1 - e^{\gamma(y)(t+2x)}\, dy$$
$$\leqslant L(1 - e^{-2HL}).$$

With the definitions

$$F = H^2 L,$$
$$G = L(1 - e^{-2HL}),$$

the inequality becomes

$$|R(x,t)| \leqslant 2(M+P)F + 4(M+P$$
$$-K)((M+P)G + L^2(M+P^2))$$
$$= 4L^2M^3 + (4G + 12L^2P - 4KL^2)M^2$$
$$+ 2(F + 4PG + 6P^2L^2 - 2KG - 4L^2PK)M$$
$$+ 2PF + 4(P-K)PG + 4P^2L^2(P-K).$$

To force $|R(x,t)|$ to be less than M, write the inequality

$$4L^2M^3 + (4G + 12L^2P - 4KL^2)M^2$$
$$+ 2(F + 4PG + 6P^2L^2 - 2KG - 4L^2PK)M$$
$$+ 2PF + 4(P-K)PG + 4P^2L^2(P-K) \leqslant M. \qquad (5.12)$$

It is impossible to say anything about (5.12) in general, except that if the constant term is negative, then a bound for M is guaranteed. In this case

$$2P(2L^2P^2 + (2G - 2L^2K)P + F - 2GK) \leqslant 0,$$

so that

$$P \leqslant \frac{L^2K - G + \sqrt{(G + L^2K)^2 - 2L^2F}}{2L^2}.$$

If P is larger than this value, or if the value is negative, it might still be possible to obtain M; otherwise, the analysis is too crude to be meaningful.

Once M has been found, the examination of $|\partial c(z) - \partial c^*(z^*)|$ may continue. Equations (5.2) and (5.3) imply that

$$\partial r(x,0) = \tfrac{1}{4}\partial c_z(z) = \partial c_x(x)/4c(z), \qquad (5.13)$$

$$\partial r^*(x,0) = \tfrac{1}{4}\partial c^*_{z^*}(z^*) = \partial c^*_x(x)/4c^*(z^*). \qquad (5.14)$$

From (5.11), then,

$$|\partial r(x,0) - \partial r^*(x,0)| \leqslant M,$$

which is rewritten using (5.13) and (5.14) as

$$|\partial c_x/c(x) - \partial c^*_x(x)/c^*(x)| \leqslant 4M,$$

or

$$|c^*(x)\partial c_x(x) - c(x)\partial c^*_x(x)|$$
$$\leqslant 4Mc(x)c^*(x)$$
$$\leqslant 4MQ(Q + |c(x) - c^*(x)|), \qquad (5.15)$$

using $c(x) = c^*(x) + (c(x) - c^*(x))$ and setting

$$Q = \max_{x'\in[0,x]} c^*(x'),$$

$$S = \min_{x'\in[0,x]} c^*(x').$$

Now write

$$S|\partial c_x(x) - \partial c^*_x(x)|$$
$$\leqslant |c^*(x)\partial c_x(x) - c^*(x)\partial c^*_x(x)|$$
$$= |c^*(x)\partial c_x(x) - c^*(x)\partial c^*_x(x)$$
$$+ c(x)\partial c^*_x(x) - c(x)\partial c^*_x(x)|$$
$$\leqslant |c^*(x)\partial c_x(x) - c(x)\partial c^*_x(x)|$$
$$+ |\partial c^*_x(x)|\, |c(x) - c^*(x)|$$
$$\leqslant 4MQ(Q + |c(x) - c^*(x)|)$$
$$+ 4PQ|c(z) - c^*(z^*)|, \qquad (5.16)$$

where the final inequality comes from (5.15) for the first term, and (5.14) and (5.10) for the second term. Presumably, $|c(z) - c^*(z^*)|$ grows monotonically as x increases, so integrating (5.16) and recognizing that $x \leqslant L$ gives

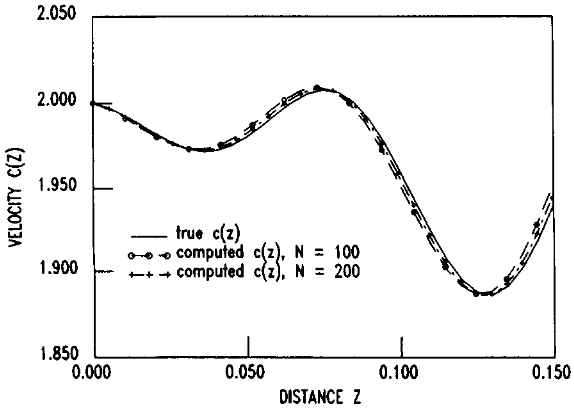

FIG. 1. This close-up of the reconstruction shows how increasing the number of grid points decreases discretization error. $c(z) = 2(1-0.1z)^2 - 0.5z\sin(20\pi z)$.

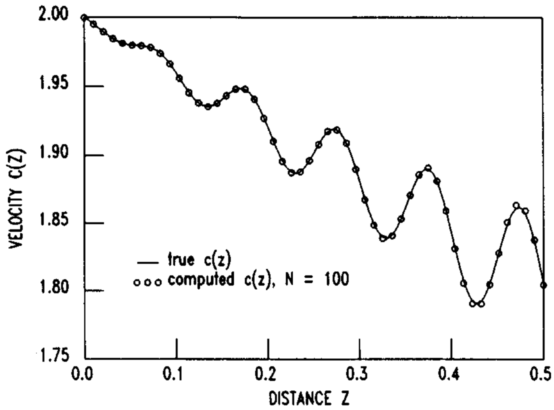

FIG. 3. When the magnitude of ∂c is reduced by a factor of 5, the algorithm is able to reconstruct the profile accurately. $c(z) = 2(1-0.1z)^2 - 0.1z\sin(20\pi z)$.

$$|\partial c(z) - \partial c^*(z^*)| \leqslant 4LMQ^2/S + 4(P+M)LQ\,|c(z) - c^*(z^*)|/S .$$

Should $|c(z) - c^*(z^*)|$ not be monotonic, the argument may be modified slightly to bound $[\max|c(x') - c^*(x')|, 0 \leqslant x' \leqslant x]$ instead. Combining this inequality with (5.6) gives the desired estimate

$$|c(z) - c^*(z^*)| \leqslant 2\alpha c_0 z^*|c(z) - c^*(z^*)| + 4LMQ^2/S + 4(P+M)LQ\,|c(z) - c^*(z^*)|/S ,$$

or

$$|c(z) - c^*(z^*)| \leqslant \frac{4LMQ^2}{S(1 - 2\alpha c_0 z^*/c^*(z^*) - 4(P+M)LQ/S)} . \tag{5.17}$$

Due to the nature of the bound, it is valid for z^* sufficiently small that the denominator is positive. It should be noted that although this estimate is relatively crude, it is significant because little has been done with error estimates for r equations in the past. A numerical example concerning the estimate is provided in Sec. VI.

VI. NUMERICAL EXAMPLES

Several specific profiles are examined in order to test the approximate algorithm. The sinusoidal forms of ∂c presented in these examples are the most interesting and illuminating of all the studied cases. Discussed are the effects of varying the number of grid points, altering the magnitude of ∂c, and changing the strength of the jump discontinuity at the interface. In all of the examples below, the medium is scaled so that $c_0 = 2$ and $Z = 0.5$.

Doubling the number of grid points halves the error due to discretization, as seen in the close-up profile of Fig. 1. However, the discretization error is negligible when compared to the error induced by the approximation, as is easily seen when Fig. 1 is expanded to the entire profile, shown in Fig. 2. This error is due to the information lost by dropping the convolution terms in the ∂r equation. Consequently, in the examples that follow, only the 100-point reconstruction is shown.

Reducing the magnitude of ∂c can dramatically improve the reconstruction process. As an example, compare Fig. 2, in which $\partial c(z) = -\frac{1}{2}z\sin(20\pi z)$, with Fig. 3, in which $\partial c(z) = -\frac{1}{10}z\sin(20\pi z)$. Naturally, the profiles are differ-

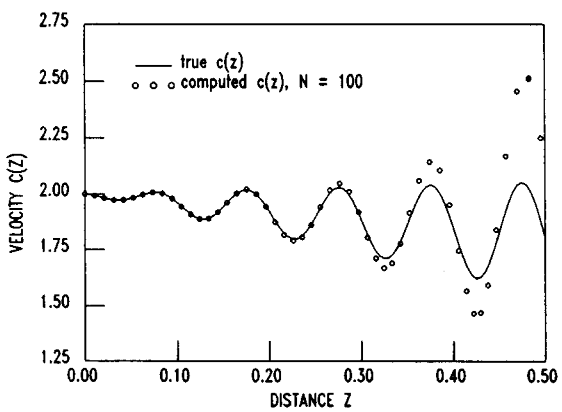

FIG. 2. Discretization error is insignificant compared to the error induced by the approximation. $c(z) = 2(1-0.1z)^2 - 0.5z\sin(20\pi z)$.

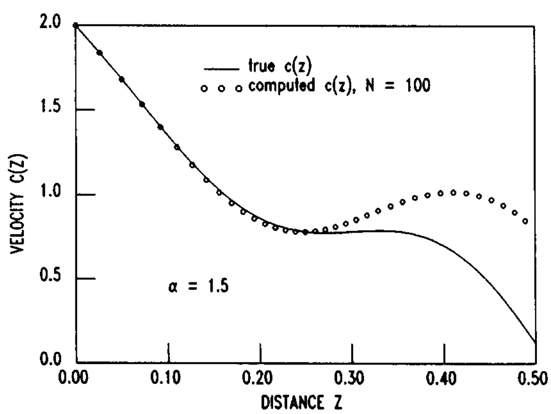

FIG. 4. With α small enough, the algorithm is able to reproduce, although poorly, the velocity profile's behavior around the inflection point. $c(z) = 2(1-1.5z)^2 - z\sin(4\pi z)$.

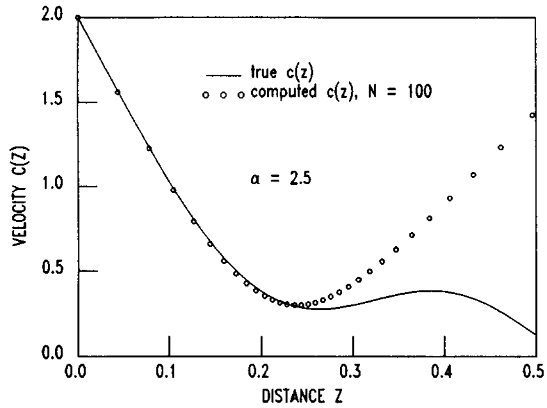

FIG. 5. When α is too large, the algorithm can no longer recover information about the reflection point. $c(z) = 2(1 - 2.5z)^2 - z\sin(4\pi z)$.

ent; the error is being compared. In Fig. 2 with the larger ∂c, the algorithm begins accurately, but the deeper into the medium it progresses, the worse the reconstruction is. When ∂c is reduced by a factor of 5, though, the algorithm can handle the oscillation of ∂c throughout the medium, resulting in a good reconstruction. Although it is not pictured, when $\partial c \equiv 0$ the reconstruction is exact, as expected.

It is also possible that changing the strength of the jump discontinuity at the interface can lead to profound changes in the reconstructed profile. Holding c_0 constant, the jump is characterized by α. Figures 4 and 5 illustrate this point, with $\partial c(z) = -z\sin(4\pi z)$. Although the true profiles are similar, the approximations are completely different. Perhaps the reason for this involves the fact that for certain values of α between the values given for the figures, 1.5 and 2.5, the profile dips negative and hence is not physically meaningful.

Finally, an example of the error estimate of Sec. V is presented. The velocity profile is $c(z) = 2(1 - 0.01z)^2 + z^2 - z^3$, for $z \in [0,0.5]$. Table I compares the estimated bound for the error induced by the approximation in reconstructing $c(z)$ with the actual error at various depths into the medium. The estimate is valid nearly halfway into the medium, at which time it breaks down because the denominator in (5.17) becomes negative.

VII. SUMMARY

Time-dependent scattering from a gradient-type interface is modeled using the r equation, which relates incident and reflected waves. An exact solution to the r equation corresponding to a parabolic velocity profile is used to provide an approximate equation. When so approximated, the algorithm for solving the inverse problem is very simple, so that when the velocity profile is nearly parabolic, the approximation gives good results at virtually no expense. Even when the velocity profile is not nearly parabolic, the approximation provides a good reconstruction for short time.

TABLE I. Comparison of estimated and actual error in $|c(z) - c^*(z^*)|$, for $c(z) = 2(1 - 0.01z)^2 + z^2 - z^3$.

x/L	$\lvert z - z^*\rvert$	Estimated error	Actual error
0.1	$3.15E-5$	0.618	$2.22E-4$
0.2	$4.02E-5$	1.235	$4.09E-4$
0.3	$6.57E-5$	2.776	$5.66E-4$
0.4	$6.91E-5$	7.445	$6.86E-4$
0.5	$8.36E-5$	···	$7.77E-4$

ACKNOWLEDGMENTS

The author thanks Anders Karlsson and Robert Krueger for helpful discusson and suggestions, as well as the referee for making several useful comments.

This work was begun at Purdue University and finished at Ames Laboratory, U.S. Department of Energy, Iowa State University. Ames Laboratory is operated for the U.S. Department of Energy by Iowa State University under Contract No. W-7405-ENG-82. This work was supported by the Applied Mathematical Sciences subprogram of the Office of Energy Research.

[1]E. Ammicht, J. P. Corones, and R. J. Krueger, "Direct and inverse scattering for viscoelastic media," J. Acoust. Soc. Am. **81**, 827 (1987).

[2]R. H. T. Bates and R. P. Millane, "Time domain approach to inverse scattering," IEEE Trans. Antennas Propag. **AP-29**, 359 (1981).

[3]R. A. Beezley and R. J. Krueger, "An electromagnetic inverse problem for dispersive media," J. Math. Phys. **26**, 317 (1985).

[4]R. J. Krueger, "An inverse problem for a dissipative hyperbolic equation with discontinuous coefficients," Quart. Appl. Math. **34**, 129 (1976).

[5]R. J. Krueger, "An inverse problem for an absorbing medium with multiple discontinuities," Quart. Appl. Math. **36**, 235 (1978).

[6]R. J. Krueger, "Inverse problems for nonabsorbing media with discontinuous material properties," J. Math. Phys. **23**, 396 (1982).

[7]G. Kristensson and R. J. Krueger, "Direct and inverse scattering in the time domain for a dissipative wave equation. I. Scattering operators," J. Math. Phys. **27**, 1667 (1986).

[8]G. Kristensson and R. J. Krueger, "Direct and inverse scattering in the time domain for a dissipative wave equation. II. Simultaneous reconstruction of dissipation and phase velocity profiles," J. Math. Phys. **27**, 1683 (1986).

[9]G. Kristensson and R. J. Krueger, "Direct and inverse scattering in the time domain for a dissipative wave equation. III. Scattering operators in the presence of a phase velocity mismatch," J. Math. Phys. **28**, 360 (1987).

[10]V. H. Weston, "On the inverse problem for a hyperbolic dispersive partial differential equation," J. Math. Phys. **13**, 1952 (1972).

[11]V. H. Weston and R. J. Krueger, "On the inverse problem for a hyperbolic dispersive partial differential equation. II," J. Math. Phys. **14**, 406 (1973).

[12]V. H. Weston, "On inverse scattering," J. Math. Phys. **15**, 209 (1974).

[13]J. P. Corones, R. J. Krueger, and V. H. Weston, "Some recent results in inverse scattering theory," in *Inverse Problems of Acoustic and Elastic Waves*, edited by F. Santosa, Y. Pao, W. Symes, and C. Holland (SIAM, Philadelphia, 1985), pp. 65–81.

[14]J. P. Corones, M. E. Davison, and R. J. Krueger, "Wave splittings, invariant imbedding and inverse scattering," in *Inverse Optics*, Proc. SPIE 413, edited by A. J. Devaney (SPIE, Bellingham, WA, 1983), pp. 102–106.

[15]J. Corones and R. J. Krueger, "Obtaining scattering kernels using invariant imbedding," J. Math. Anal. Appl. **95**, 393 (1983).

[16]R. Bellman and G. N. Wing, *An Introduction to Invariant Imbedding* (Wiley, New York, 1975).

[17]J. P. Corones, M. E. Davison, and R. J. Krueger, "Direct and inverse scattering in the time domain by invariant imbedding techniques," J. Acoust. Soc. Am. **74**, 1535 (1983).

[18]R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1962), Vol. II.

On a basic theorem of quaternion modules

Aharon Razon and L. P. Horwitz
School of Physics and Astronomy, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Ramat Aviv, Israel
L. C. Biedenharn
Department of Physics, Duke University, Durham, North Carolina 27707

(Received 26 July 1988; accepted for publication 31 August 1988)

A complete proof of the equivalence of the star operation in the operator algebra isomorphic to quaternions and the adjoint operation in a quaternion Hilbert module is given.

I. INTRODUCTION

The possibility of providing a natural framework for certain classes of non-Abelian gauge theories has motivated careful study and extension of the work of Finkelstein, Jauch, Schimonovich, and Speiser[1] on quaternion quantum mechanics. Adler[2] has found, for example, in his study of semiclassical forms for non-Abelian gauge quantum field theory, that a quaternion structure appears to be a good candidate for a prequark theory. In subsequent work, he has initiated the development of quaternion field theory.[3] Biedenharn and Horwitz[4] and Adler[5] have recently carried out basic investigations of the structure of quaternion quantum mechanics.

A basic theorem on the structure of quaternionic Hilbert modules was given in the first of Refs. 4 concerning the properties of the left-acting star algebra of operators isomorphic to the algebra of quaternions, i.e., that $Q^\dagger = Q^*$, where $Q^\dagger$ is the adjoint of the operator Q. The structure of the hierarchy of scalar products (real, complex, quaternionic) that exist in the quaternion Hilbert module depends on this theorem; for example, it follows from this result that the scalar products of "formally real" components of elements of this space are real valued.

A complete proof was not given in Ref. 4 for this theorem. In view of the current activity in this subject, it is of interest to provide a proof; we shall do this in what follows.

II. PROOF OF THE THEOREM

Consider a vector space $V_{\mathbf{H}}$ that is also a right module over the real quaternion algebra $\mathbf{H}$, generated by the elements e_1, e_2, with real coefficients, where

$$e_1^2 = e_2^2 = -1, \quad e_1e_2 = -e_2e_1. \tag{1}$$

The algebra has an involution

$$e_1^* = -e_1, \quad e_2^* = -e_2, \quad (e_1e_2)^* = e_2e_1^*, \tag{2}$$

and hence $e_3^* = -e_3$, where $e_3 \equiv e_1e_2$.

There exists a binary mapping (f,g) of $V_{\mathbf{H}} \times V_{\mathbf{H}}$ into $\mathbf{H}$ with the properties

$$(f,g) = (g,f),$$
$$(f,g+h) = (f,g) + (f,h),$$
$$(f,gq) = (f,g)q,$$

where $q \in \mathbf{H}$, and

$$(f,f) = \|f\|^2 \geqslant 0, \tag{3}$$

and is zero if and only if $f = 0$. If $V_{\mathbf{H}}$ is closed under the topology provided by the norm $\|f\|$, it is a right quaternion Hilbert module, which we shall call $H_{\mathbf{H}}$.

The theorem that we wish to discuss concerns the properties of a special set of mappings $H_{\mathbf{H}} \to H_{\mathbf{H}}$ represented by the left action of an associative algebra of bounded linear operators star isomorphic to $\mathbf{H}$. There is a set of operator-valued generators E_1, E_2, $E_3 \equiv E_1E_2$ that have the algebraic properties (2). The operators $(E_0 \equiv I, \lambda_i \in R)$

$$Q = \sum_{i=0}^{3} \lambda_i E_i \tag{4}$$

have the property

$$Q(Q^*f) = \|Q\|^2 f, \tag{5}$$

where $\|Q\|^2 = \Sigma \lambda_i^{\ 2}$, and $\|Q\|$ is the operator norm.

We now prove that $Q^\dagger = Q^*$. We may assume that $\|Q\|$ $(= \|Q^*\|) = 1$, and therefore, from (5), that $QQ^* = I$ and $Q^{*\dagger}Q^\dagger = I$. It then follows that

$$\|Q^*f\| = \|f\|$$

and

$$\|Q^\dagger f\| = \|f\|, \tag{6}$$

for all f. For if we assume [e.g., for the first of (6)] that there exists a g such that $\|Q^*g\| < \|g\|$, then $\|g\| = \|QQ^*g\| \leqslant \|Q\|\,\|Q^*g\| < \|g\|$, a contradiction. Therefore, since $(Q^\dagger f, Q^*f) = (Q^*f, Q^\dagger f) = \|f\|^2$,

$$\|Q^\dagger f - Q^*f\|^2 = \|Q^\dagger f\|^2 + \|Q^*f\|^2 - (Q^\dagger f, Q^*f) - (Q^*f, Q^\dagger f) = 0,$$

for all f. This completes the proof of the theorem.

[1] D. Finkelstein, J. M. Jauch, S. Schiminovich, and D. Speiser, J. Math. Phys. **3**, 207 (1962); **4**, 788 (1963).
[2] S. L. Adler, Phys. Rev. D **21**, 550 (1980).
[3] S. L. Adler, Phys. Rev. Lett. **55**, 783 (1985); Phys. Rev. D **34**, 1871 (1986).
[4] L. P. Horwitz and L. C. Biedenharn, Ann. Phys. (NY) **157**, 432 (1984). See also L. P. Horwitz and A. Soffer, J. Math. Phys. **24**, 2780 (1983); J. Rembielinski, Phys. Lett. B **88**, 279 (1979); J. Phys. A: Math. Gen. **13**, 15, 23 (1980); **14**, 2609 (1981); C. G. Nash and G. C. Joshi, J. Math. Phys. **28**, 463, 2883, 2886 (1987); C. S. Sharma and T. J. Coulson, *ibid.* **28**, 1941 (1987).
[5] S. L. Adler, Commun. Math. Phys. **104**, 611 (1986). See also S. L. Adler, "Scattering and decay theory for quaternionic quantum mechanics, and the structure of induced *T*-nonconservation," Institute for Advanced Study preprint IASSNS-HEP-87/67.

Quantum mechanics of the damped pulsating oscillator

M. Sebawe Abdalla
Department of Mathematics, College of Science, King Saud University, P. O. Box 2455, Riyadh 11451, Saudi Arabia
Ursula A. T. Ramjit
Department of Mathematics, King's College, Strand, London WC2R 2LS, England

(Received 13 January 1988; accepted for publication 24 August 1988)

A harmonic oscillator subject to the combined effects of damping and pulsating is represented by a Kanai–Caldirola Hamiltonian. The equations of motion are solved in the Heisenberg picture in the case of weak pulsation. The rotating-wave approximation (RWA) is used to obtain the motion in the neighborhood of the principal resonance. The RWA Schrödinger equation is solved exactly and pseudostationary and quasicoherent states are described. The transition probability between quasicoherent and coherent states is obtained and the gain in energy is discussed.

I. INTRODUCTION

The time-dependent harmonic oscillator has interesting applications in quantum optics; for example, a pulsating oscillator can be used to describe the Rabi modulation of a mode of the electromagnetic field owing to the emission and subsequent reabsorption of photons from resonant atoms.[1–5] The dynamics of such a system may be described by the well known Kanai–Caldirola Hamiltonian[6–11]

$$H(q,p,t) = \tfrac{1}{2} p^2/m(t) + \tfrac{1}{2} m(t)\omega_0^2 q^2 , \tag{1.1}$$

with an appropriate choice of $m(t)$ to reflect the periodic nature of the energy stored in the mode. Damping in a classical or quantum oscillator is another phenomenon that may be represented by a time-dependent Hamiltonian.[10–12] It is convenient to write $m(t)$ in the form[13]

$$m(t) = m_0 \exp[2\Gamma(t)] . \tag{1.2}$$

Then damping is described by $\Gamma(t) = \gamma t$ (γ constant) and a pulsation of strength μ corresponds to $\Gamma(t) = \mu \sin(\nu t)$.[2–4]

In the present paper we treat the combined effect of damping and pulsation by means of the fluctuation function[14]

$$\dot\Gamma(t) = \tfrac{1}{2}\dot m/m = \gamma + \mu\nu \cos \nu t , \tag{1.3}$$

or the mass parameter

$$m(t) = m_0 \exp[2(\gamma t + \mu \sin \nu t)] . \tag{1.4}$$

We shall concentrate on the quantum-mechanical aspects of the system represented by Eqs. (1.1) and (1.4), which describes the radiation in a Fabry–Pèrot cavity in the presence of a resonant atom when leakage through the walls is taken into account.

In the next section we shall consider the solution of the problem in the Heisenberg picture. The equations of motion are not amenable to exact solution and an analytical treatment must be confined to (a) the weakly pulsating case with $\mu \ll 1$, using perturbation theory, and (b) the near-resonance case, when the rotating-wave approximation (RWA) may be applied.

The remainder of the paper is concerned with the RWA solution. In Sec. III an alternative wave description is considered. The Schrödinger equation is solved and quasicoherent states are found. The uncertainty relations are examined in Sec. IV and energy expectation values are obtained. In Sec. V the transition probability is calculated between a coherent state of the simple harmonic oscillator and a quasicoherent state of the time-dependent oscillator. Finally, in Sec. VI, a brief discussion is given.

II. EQUATIONS OF MOTION

The Hamiltonian described by Eqs. (1.1) and (1.4) leads to the equation of motion

$$\ddot q + 2(\gamma + \mu\nu \cos \nu t)\dot q + \omega_0^2 q = 0 . \tag{2.1}$$

However, we shall find it convenient to make the canonical transformation

$$Q = q \exp(\gamma t + \mu \sin \nu t), \quad P = p \exp(-\gamma t - \mu \sin \nu t) , \tag{2.2}$$

which takes the Hamiltonian of Eq. (1.1) to the form

$$K(Q,P,t) = \frac{1}{2}\frac{P^2}{m_0} + \frac{1}{2} m_0\omega_0^2 Q^2 + \frac{\partial F_2}{\partial t} , \tag{2.3}$$

where F_2 is the generating function given by[12,15]

$$F_2(q,P,t) = \tfrac{1}{2}[m(t)/m_0]^{1/2}(qP + Pq) . \tag{2.4}$$

From Eqs. (2.3) and (2.4),

$$K(Q,P,t) = \tfrac{1}{2} P^2/m_0 + \tfrac{1}{2} m_0\omega_0^2 Q^2 + \tfrac{1}{2}\dot\Gamma(t)(QP + PQ) , \tag{2.5}$$

where $\dot\Gamma(t)$ is given by Eq. (1.3). The Heisenberg equations corresponding to the Hamiltonian (2.5) are

$$\dot Q = P/m_0 + \dot\Gamma Q, \quad \dot P = -m_0\omega_0^2 Q - \dot\Gamma P , \tag{2.6}$$

or, separately for Q and P,

$$\ddot Q + (\omega^2 + \mu\nu^2 \sin \nu t - 2\gamma\mu\nu \cos \nu t - \mu^2\nu^2 \cos^2 \nu t)Q = 0 , \tag{2.7a}$$

$$\ddot P + (\omega^2 - \mu\nu^2 \sin \nu t - 2\gamma\mu\nu \cos \nu t - \mu^2\nu^2 \cos^2 \nu t)P = 0 , \tag{2.7b}$$

where $\omega = (\omega_0^2 - \gamma^2)^{1/2}$ is the reduced frequency corresponding to undercritical damping.

 0022-2488/89/010060-06$02.50

A. Approximate solution using perturbation theory

If $\mu \ll 1$, Eq. (2.7a) is a damped Mathieu equation.[16] Taking μ as a perturbation parameter and working to first order, we write

$$Q(t) = Q_0(t) + \mu Q_1(t)\,. \tag{2.8}$$

(A second-order calculation is extremely tedious and has little value.) Substituting Eq. (2.8) into Eq. (2.7a) and equating μ^0 and μ terms gives

$$\ddot{Q}_0 + \omega^2 Q_0 = 0\,, \tag{2.9a}$$

$$\ddot{Q}_1 + \omega^2 Q_1 = \nu[2\gamma \cos \nu t - \nu \sin \nu t\,]Q_0\,. \tag{2.9b}$$

The arbitrary constants that arise can be fixed by setting $Q_0(0) = Q(0)$, $Q_1(0) = 0$, and

$$\dot{Q}(0) = \dot{Q}_0(0) + \mu \dot{Q}_1(0) = P(0)/m_0 + (\gamma + \mu\nu)Q(0)\,. \tag{2.10}$$

To first order in μ, we find

$$\begin{aligned}\frac{\omega}{\omega_0}Q(t) = Q(0)\Big\{&\cos(\omega t - \eta) + \frac{\mu}{4\omega^2 - \nu^2}\Big[2\nu\omega_0 \sin \omega t \\ &- \frac{1}{2}\bar{\nu}(2\omega - \nu)\cos((\nu + \omega)t + \delta - \eta) \\ &+ \frac{1}{2}\bar{\nu}(2\omega + \nu)\cos((\nu - \omega)t + \delta + \eta)\Big]\Big\} \\ &+ \frac{P(0)}{m_0\omega_0}\Big\{\sin \omega t + \frac{\mu}{4\omega^2 - \nu^2} \\ &\times\Big[2\nu\omega_0 \cos(\omega t - \eta) \\ &- \frac{1}{2}\bar{\nu}(2\omega - \nu)\sin((\nu + \omega)t + \delta) \\ &- \frac{1}{2}\bar{\nu}(2\omega + \nu)\sin((\nu - \omega)t + \delta)\Big]\Big\}\,,\end{aligned} \tag{2.11a}$$

$$\begin{aligned}\frac{\omega}{\omega_0}P(t) = P(0)\Big\{&\cos(\omega t + \eta) - \frac{\mu}{4\omega^2 - \nu^2}\Big[2\nu\omega_0 \sin \omega t \\ &+ \frac{\nu}{\omega_0}(4\omega^2 - \nu^2)\sin \omega t \cos \nu t \\ &+ \frac{\bar{\nu}\tilde{\nu}_+}{2\omega_0}(2\omega - \nu)\cos((\nu + \omega)t + \delta + \bar{\eta}_+) \\ &+ \frac{\bar{\nu}\tilde{\nu}_-}{2\omega_0}(2\omega + \nu)\cos((\nu - \omega)t + \delta + \bar{\eta}_-)\Big]\Big\} \\ &- m_0\omega_0 Q(0)\Big\{\sin \omega t \\ &- \frac{\mu}{4\omega^2 - \nu^2}\Big[2\nu\omega_0 \cos(\omega t + \eta) \\ &- \frac{\nu}{\omega_0}(4\omega^2 - \nu^2)\cos(\omega t - \eta)\cos \nu t \\ &+ \frac{\bar{\nu}\tilde{\nu}_+}{2\omega_0}(2\omega - \nu)\sin((\nu + \omega)t + \delta + \bar{\eta}_+ - \eta) \\ &- \frac{\bar{\nu}\tilde{\nu}_-}{2\omega_0}(2\omega + \nu) \\ &\times \sin((\nu - \omega)t + \delta + \bar{\eta}_- + \eta)\Big]\Big\}\,,\end{aligned} \tag{2.11b}$$

where

$$\begin{aligned}\eta &= \arctan(\gamma/\omega), \\ \delta &= \arctan(\nu/2\gamma), \\ \bar{\nu} &= (\nu^2 + 4\gamma^2)^{1/2}, \\ \bar{\eta}_\pm &= \arctan[\gamma/(\nu \pm \omega)], \\ \tilde{\nu}_\pm &= [(\nu \pm \omega)^2 + \gamma^2]^{1/2}.\end{aligned} \tag{2.12}$$

It is easy to see that the commutation relation

$$[Q(t),P(t)] = i\hbar \tag{2.13}$$

holds to first order in μ.

B. The RWA solution

In the near-resonance region the rotating-wave approximation provides a useful alternative to perturbation theory. For sufficiently large values of the time the RWA gives a solution in which all higher harmonics are neglected, but which otherwise satisfies the equation of motion to all orders in μ.[4]

We employ the Dirac boson operators

$$a(t) = (2m_0\omega_0\hbar)^{-1/2}(m_0\omega_0 Q + iP)\,, \tag{2.14a}$$

$$a^\dagger(t) = (2m_0\omega_0\hbar)^{-1/2}(m_0\omega_0 Q - iP)\,. \tag{2.14b}$$

Then the Hamiltonian (2.5) may be written

$$K(a,a^\dagger,t) = \hbar\omega_0(a^\dagger a + \tfrac{1}{2}) + \tfrac{1}{2}i\hbar(\gamma + \mu\nu \cos \nu t)(a^{\dagger 2} - a^2)\,, \tag{2.15}$$

and the equations of motion (2.6) become

$$\frac{da}{dt} = -i\omega_0 a + (\gamma + \mu\nu \cos \nu t)a^\dagger\,, \tag{2.16a}$$

$$\frac{da^\dagger}{dt} = i\omega_0 a^\dagger + (\gamma + \mu\nu \cos \nu t)a\,. \tag{2.16b}$$

These equations are simplified by making the canonical transformation[12]

$$\begin{bmatrix} a \\ a^\dagger \end{bmatrix} = (2\omega)^{-1/2}\begin{bmatrix} (\omega_0 + \omega)^{1/2} & -i(\omega_0 - \omega)^{1/2} \\ i(\omega_0 - \omega)^{1/2} & (\omega_0 + \omega)^{1/2} \end{bmatrix}\begin{bmatrix} b \\ b^\dagger \end{bmatrix} \tag{2.17}$$

followed by

$$c = b \exp\{i[\omega t - (\mu\gamma/\omega)\sin \nu t\,]\}\,. \tag{2.18}$$

Then Eq. (2.16a) reduces to

$$\frac{dc}{dt} = \left(\frac{\mu\nu\omega_0}{\omega}\right)\cos \nu t \exp\left\{2i\left[\omega t - \left(\frac{\mu\gamma}{\omega}\right)\sin \nu t\,\right]\right\}c^\dagger\,. \tag{2.19}$$

When all rapidly oscillating terms are ignored in accordance with the RWA, we are left with [cf. Ref. 4, Eq. (6.2)]

$$\frac{dc}{dt} = \frac{1}{2}\left(\frac{\mu\nu\omega_0}{\omega}\right)(M + N)\exp[-i(\nu - 2\omega)t\,]c^\dagger\,, \tag{2.20a}$$

where M and N are given by the series

$$M = \sum_{n=0}^{\infty}(-1)^n\left(\frac{\mu\gamma}{\omega}\right)^{2n}(n!)^{-2}\,, \tag{2.20b}$$

$$N = \sum_{n=0}^{\infty}(-1)^n\left(\frac{\mu\gamma}{\omega}\right)^{2n+2}(n!)^{-1}[(n+2)!]^{-1}\,, \tag{2.20c}$$

which converge for all values of $\mu\gamma/\omega$.

The solution of Eq. (2.20a) taken with its adjoint is

$$c(t) = e^{i(\omega - \nu/2)t}\{c(0)[\cosh(\rho t) - i(\omega - \nu/2)\sinh(\rho t)/\rho] + \tfrac{1}{2}(\mu\nu\omega_0/\omega)(M+N)c^\dagger(0)\sinh(\rho t)\}, \quad (2.21)$$

where the growth factor is

$$\rho = \tfrac{1}{2}[(\mu\nu\omega_0/\omega)^2(M+N)^2 - (2\omega - \nu)^2]^{1/2}. \quad (2.22)$$

Converting back to the coordinate Q and momentum P [related to the physical coordinate and momentum by Eq. (2.21)] via Eqs. (2.14), (2.17), (2.18), and (2.21), we obtain

$$Q_{\rm RWA}(t) = Q(0)F(t) + [P(0)/(m_0\omega_0)]G(t), \quad (2.23a)$$

$$P_{\rm RWA}(t) = P(0)K(t) - m_0\omega_0 Q(0)L(t), \quad (2.23b)$$

where

$$F(t) = (\omega_0/\omega)\{\cos(\Phi(t) - \eta)\cosh\rho t + R[\cos(\Phi(t))\cos\epsilon - \sin(\Phi(t) - \eta)\sin\epsilon]\sinh\rho t\}, \quad (2.24a)$$

$$G(t) = (\omega_0/\omega)\{\sin(\Phi(t))\cosh\rho t + R[\cos(\Phi(t))\sin\epsilon - \sin(\Phi(t) - \eta)\cos\epsilon]\sinh\rho t\}, \quad (2.24b)$$

$$K(t) = (\omega_0/\omega)\{\cos(\Phi(t) + \eta)\cosh\rho t - R[\cos(\Phi(t))\cos\epsilon + \sin(\Phi(t) + \eta)\sin\epsilon]\sinh\rho t\}, \quad (2.24c)$$

$$L(t) = (\omega_0/\omega)\{\sin(\Phi(t))\cosh\rho t + R[\cos\Phi(t)\sin\epsilon + \sin(\Phi(t) + \eta)\cos\epsilon]\sinh\rho t\}, \quad (2.24d)$$

$$\Phi(t) = \tfrac{1}{2}\nu t - (\mu\gamma/\omega)\sin\nu t, \quad (2.24e)$$

and in addition

$$R = (\Omega^2 + \Delta^2)^{1/2},$$
$$\Omega = \tfrac{1}{2}(\mu\nu\omega_0/\omega\rho)(M+N), \quad \Delta = (2\omega - \nu)/2\rho,$$
$$\epsilon = \arctan(\Delta/\Omega). \quad (2.25)$$

The term η has been defined in Eq. (2.12) and M,N has been defined in Eqs. (2.20b) and (2.20c). From Eq. (2.22) Ω and Δ are connected by the relation

$$\Omega^2 - \Delta^2 = 1. \quad (2.26)$$

At exact resonance ($\nu = 2\omega$) the solution simplifies considerably, since then

$$\rho \to \rho_0 = \mu\omega_0(M+N), \quad \Omega = 1,$$
$$\Delta = 0, \quad R = 1, \quad \epsilon = 0. \quad (2.27)$$

Inserting Eqs. (2.24) with the simplifications of Eqs. (2.27) into Eqs. (2.23) we find, for large values of the time and provided $\rho_0 > 0$,

$$Q_{\rm RWA}(t) \sim (m_0\omega)^{-1}(2\omega_0)e^{\rho_0 t}\cos[\Phi(t) - \eta/2] \times [m_0\omega_0(\omega_0 + \omega)^{1/2}Q(0) + (\omega_0 - \omega)^{1/2}P(0)], \quad (2.28a)$$

$$P_{\rm RWA}(t) \sim -\omega^{-1}(\omega_0/2)^{1/2}e^{\rho_0 t}\sin[\Phi(t) + \eta/2] \times [m_0\omega_0(\omega_0 + \omega)^{1/2}Q(0) + (\omega_0 - \omega)^{1/2}P(0)]. \quad (2.28b)$$

If $\rho_0 < 0$, the replacement $e^{\rho_0 t} \to \pm e^{|\rho_0| t}$ gives the correct asymptotic form (the $\pm$ being chosen according to whether $e^{-\rho t}$ comes from $\cosh\rho t$ or $\sinh\rho t$). It is interesting that the same combination of $Q(0)$ and $P(0)$ occurs in both $Q_{\rm RWA}$ and $P_{\rm RWA}$. The initial conditions could be arranged so that neither grows as $t \to \infty$.

We continue to focus our attention on this solvable RWA in the remainder of the paper. Since a solution has been found in the Heisenberg picture, it follows that a solution may be found in the Schrödinger picture. This requires a certain degree of ingenuity and is given in the next section.

III. WAVE FUNCTIONS FOR PSEUDOSTATIONARY AND QUASICOHERENT STATES

The main purpose of the present section is to give an exact solution of the Schrödinger equation,

$$K_{\rm RWA}\Psi(Q,t) = i\hbar\frac{\partial}{\partial t}\Psi(Q,t), \quad (3.1)$$

where it is to be understood that the pulsation frequency ν is near to the resonance value of 2ω. Following the method of Ref. 4, we evaluate the RWA version of the Hamiltonian (2.5) in the form

$$K_{\rm RWA}(Q,P,t) = \tfrac{1}{2}s(t)P^2/m_0 + \tfrac{1}{2}m_0\omega_0^2 r(t)Q^2 + \tfrac{1}{2}v(t)(QP + PQ), \quad (3.2)$$

where

$$s(t) = 1 - \bar{\mu}\cos\nu t - (\rho\Omega/\omega)\sin[2\Phi(t) - \eta], \quad (3.3a)$$

$$r(t) = 1 - \bar{\mu}\cos\nu t + (\rho\Omega/\omega)\sin[2\Phi(t) + \eta], \quad (3.3b)$$

$$v(t) = \gamma(1 - \bar{\mu}\cos\nu t) + (\rho\Omega\omega_0/\omega)\cos[2\Phi(t)], \quad (3.3c)$$

$$\bar{\mu} = \mu\nu\gamma/\omega^2, \quad (3.3d)$$

and η is given in Eqs. (2.12). From Eqs. (3.1) and (3.2) we find

$$s(t)\frac{\partial^2\Psi}{\partial Q^2} + i\frac{m_0}{\hbar}v(t)\left(2Q\frac{\partial\Psi}{\partial Q} + \Psi\right) - \left(\frac{m_0\omega_0}{\hbar}\right)^2 r(t)Q^2\Psi = -2i\frac{m_0}{\hbar}\frac{\partial\Psi}{\partial t}. \quad (3.4)$$

Now let us make the transformation

$$x = W(t)Q, \quad W(t) = [F^2 + G^2 - 2(\gamma/\omega_0)FG]^{-1/2}, \quad (3.5)$$

where $F(t)$ and $G(t)$ are given by Eqs. (2.24a) and (2.24b). The wave function $\Psi(Q,t)$ transforms to $\theta(x,t)$ and Eq. (3.4) becomes

$$\frac{\partial^2\theta}{\partial x^2} + 2i\frac{m_0}{\hbar}\delta(t)x\frac{\partial\theta}{\partial x} - \frac{r(t)}{s(t)}\left(\frac{m_0\omega_0 x}{\hbar W^2(t)}\right)^2\theta = -i\frac{m_0}{\hbar W^2(t)s(t)}\left(2\frac{\partial\theta}{\partial t} + v(t)\theta\right), \quad (3.6a)$$

where

$$\delta(t) = s^{-1}(t)W^{-3}(t)[\dot{W}(t) + v(t)W(t)]. \quad (3.6b)$$

We now seek a separation of the form

$$\theta(x,t) = X(x)T(t)\exp[-i(m_0/2\hbar)\delta(t)x^2] . \quad (3.7)$$

Substituting Eq. (3.7) into Eq. (3.6a) we obtain

$$\left(-\frac{\hbar^2}{2m_0}\right)\left(\frac{X''}{X}\right) + \frac{1}{2}m_0\omega^2x^2 = i\frac{\hbar}{sW^2}\left(\frac{\dot{T}}{T}\right) - i\frac{\hbar\dot{W}}{2sW^3} = \hbar\omega\left(n+\frac{1}{2}\right), \quad n = 0,1,2,..., \quad (3.8)$$

which gives a fundamental pseudostationary solution of Eq. (3.4) in the form

$$\Psi_n(Q,t) = \left(\frac{m_0\omega}{\pi\hbar}\right)^{1/4}(n!)^{-1/2}2^{-n/2} \times W^{1/2}(t)H_n[(m_0\omega/\hbar)^{1/2}W(t)Q] \times\exp[-(m_0/2\hbar)(\omega+i\delta(t))W^2(t)Q^2 - i(n+\tfrac{1}{2})\arctan J(t)] , \quad (3.9a)$$

where H_n denotes the Hermite polynomial of order n and

$$J(t) = (\omega/\omega_0)G(t)[F(t) - (\gamma/\omega_0)G(t)]^{-1} . \quad (3.9b)$$

The final phase factor in Eq. (3.9a) is important for the construction of quasicoherent states. In the absence of pulsation $\mu\to 0$ and Eq. (3.9a) reduces to

$$\Psi_n(Q,t) = \left(\frac{m_0\omega}{\pi\hbar}\right)^{1/4}(n!)^{-1/2}2^{-n/2}H_n[(m_0\omega/\hbar)^{1/2}Q] \times\exp[-(m_0/2\hbar)(\omega+i\gamma)Q^2 - i\omega(n+\tfrac{1}{2})t] . \quad (3.10)$$

This result may be compared with Eq. (3.7) of Ref. 12. Similarly, when $\gamma\to 0$, Eq. (3.9a) is exactly Eq. (11.16a) of Ref. 4 in the absence of the driving force.

A quasicoherent state Ψ_α may be expressed in terms of the pseudostationary states given by Eq. (3.9a) according to the relation[13]

$$\Psi_\alpha(Q,t) = \exp\left(-\frac{1}{2}|\alpha|^2\right)\sum_{n=0}^{\infty}(n!)^{-1/2}\alpha^n\Psi_n(Q,t) . \quad (3.11)$$

Substituting Eq. (3.9a) into Eq. (3.11) we obtain

$$\Psi_\alpha(Q,t) = (m_0\omega/\hbar\pi)^{1/4}W^{1/2}(t) \times\exp[-(m_0/2\hbar)(\omega+i\delta(t))W^2(t)Q^2 - \tfrac{1}{2}i\arctan J(t)] \times\exp[-\tfrac{1}{2}(\alpha^2(t)+|\alpha|^2) + (2m_0\omega/\hbar)^{1/2} \times\alpha(t)W(t)Q] , \quad (3.12a)$$

where

$$\alpha(t) = \alpha(0)\exp[-i\arctan J(t)] . \quad (3.12b)$$

IV. UNCERTAINTY RELATIONS AND ENERGY EXPECTATION VALUE

In the previous section we have obtained the wave function for a quasicoherent state $|\alpha(t)\rangle$ of the damped pulsating harmonic oscillator. Coherent or quasicoherent states are eigenstates of the operator given by Eq. (2.14a) or of any other suitably defined operator $A(t)$ which satisfies

$$[A(t),A^\dagger(t)] = 1 \quad (4.1)$$

at all times. Using Q and P in the RWA given by Eqs. (2.23a) and (2.23b) it is easily shown that, in the eigenstate $|\alpha\rangle$ corresponding to Eqs. (3.12), the expectation values are

$$\langle\alpha|Q(t)|\alpha\rangle = [\hbar/(2m_0\omega_0)]^{1/2}[(F(t)+iG(t))\alpha^* + (F(t)-iG(t))\alpha] , \quad (4.2a)$$

$$\langle\alpha|P(t)|\alpha\rangle = -[\hbar m_0\omega_0/2]^{1/2}[(L(t)-iK(t))\alpha^* + (L(t)+iK(t))\alpha] , \quad (4.2b)$$

where F, G, K, and L are defined by Eqs. (2.24).

We find that the uncertainties in Q and P do not depend on α and their product is given by

$$\Delta Q_\alpha\Delta P_\alpha = \tfrac{1}{2}\hbar\{[F^2(t)+G^2(t)][K^2(t)+L^2(t)]\}^{1/2} . \quad (4.3)$$

In the case of exact resonance ($\nu = 2\omega$) we can write Eq. (4.3) in the more explicit form

$$(\Delta Q)^2_\alpha(\Delta P)^2_\alpha = \tfrac{1}{4}\hbar^2\{[\omega_0^2\cosh 2pt + \gamma\omega_0\sin 2\Phi\sinh 2pt - \gamma^2\cos 2\Phi]^2/\omega^4 - [\omega_0\cos 2\Phi\sinh 2pt + \gamma\sin^2 2\Phi]^2/\omega^2\} . \quad (4.4)$$

In the absence of pulsation ($\mu\to 0$), Eq. (4.4) reduces to

$$\Delta Q_\alpha\Delta P_\alpha = \tfrac{1}{2}\hbar[1 + 4(\gamma\omega_0/\omega^2)^2\sin^4\omega t]^{1/2} , \quad (4.5)$$

which agrees with Eq. (25) of Ref. 13 and with Eq. (22) of Ref. 17. Also, in the absence of damping ($\gamma\to 0$), Eq. (4.4) is in agreement with Eq. (7.8) of Ref. 4.

The expressions for the potential energy and the kinetic energy are too lengthy to quote here. However, we can evaluate the expectation values of these quantities with respect to the number states, which satisfy

$$a|n\rangle = n^{1/2}|n-1\rangle, \quad a^\dagger|n\rangle = (n+1)^{1/2}|n+1\rangle , \quad a^\dagger a|n\rangle = n|n\rangle, \quad n = 0,1,2,..., \quad (4.6)$$

i.e., with respect to the quasistationary states $\langle Q|n\rangle = \Psi_n(Q,t)$ given by Eqs. (3.9). We find, with $\tilde{\Gamma} = \gamma/\omega_0$, $\Phi(t)$ and η given by Eqs. (2.24e) and (2.12), the following expressions for the expectation values of the kinetic energy $T = \frac{1}{2}p^2/m_0$, the potential energy $V = \frac{1}{2}m_0\omega_0^2q^2$, and the total energy $E = T + V$:

$$\langle n|T|n\rangle = \tfrac{1}{2}\hbar(n+\tfrac{1}{2})(\omega_0^3/\omega^2) \times\exp[-2(\gamma t+\mu\sin\nu t)]\{[1-\tilde{\Gamma}\sin(2\Phi+\eta)]\cosh^2 pt - (\Omega+\tilde{\Gamma}\Delta)\cos(2\Phi+\eta) \times\sinh 2pt + [(\Omega^2+\Delta^2)(1+\tilde{\Gamma}\sin(2\Phi+\eta)) + 4\Omega\Delta\cos\Phi\sin(\Phi+\eta)]\sinh^2 pt\} , \quad (4.7a)$$

$$\langle n|V|n\rangle = \tfrac{1}{2}\hbar(n+\tfrac{1}{2})(\omega_0^3/\omega^2) \times\exp[-2(\gamma t+\mu\sin\nu t)]\{[1+\tilde{\Gamma}\sin(2\Phi-\eta)]\cosh^2 pt + (\Omega+\tilde{\Gamma}\Delta)\cos(2\Phi-\eta)\sinh 2pt + [(\Omega^2+\Delta^2)(1-\tilde{\Gamma}\sin(2\Phi-\eta)) - 4\Omega\Delta\cos\Phi\sin(\Phi-\eta)]\sinh^2 pt\} , \quad (4.7b)$$

$$\langle n|E|n\rangle = \hbar(n+\tfrac{1}{2})(\omega_0^3/\omega^2)\times\exp[-2(\gamma t+\mu\sin\nu t)]\{(1-\tilde{\Gamma}^2\cos 2\Phi)\cosh^2\rho t+\tilde{\Gamma}(\Omega+\tilde{\Gamma}\Delta)\sin 2\Phi\sinh 2\rho t+[(\Omega^2+\Delta^2)(1+\tilde{\Gamma}^2\cos 2\Phi)+4\tilde{\Gamma}\Delta\Omega\cos^2\Phi]\sinh^2\rho t\}. \tag{4.7c}$$

At exact resonance ($\Omega = 1$, $\Delta = 0$), Eq. (4.7c) reduces to

$$\langle n|E|n\rangle = \hbar(n+\tfrac{1}{2})(\omega_0^3/\omega^2)\exp[-2(\gamma t+\mu\sin\nu t)]\times(\cosh 2\rho_0 t+\tilde{\Gamma}\sin 2\Phi\times\sinh 2\rho_0 t-\tilde{\Gamma}^2\cos 2\Phi). \tag{4.8}$$

The gain coefficient is $\rho_0 - \gamma$, where $\rho_0 = \mu\omega_0(M+N)$ and M and N are the series given by Eqs. (2.20b) and (2.20c). If the damping and pulsating are small enough to make $\mu\gamma/\omega$ small, then $M\approx 1$ and $N\approx 0$.

V. PROBABILITY FOR TRANSITION FROM COHERENT TO QUASICOHERENT STATE

We shall now calculate the transition probability $|\langle\alpha_0(t)|\alpha(t)\rangle|^2$ (as discussed in Ref. 18) between a quasicoherent state $|\alpha(t)\rangle$ of the damped pulsating oscillator, given by Eqs. (3.12a) and (3.12b), and a coherent state $|\alpha_0(t)\rangle$ of a simple harmonic oscillator. We may take $m_0 = 1$ without loss in generality; then for the state $|\alpha_0(t)\rangle$,

$$\Psi_{\alpha_0}(Q,t) = (\omega_0/\pi\hbar)^{1/2}\exp\{-(\omega_0/2\hbar)Q^2+(2\omega_0/\hbar)^{1/2}\alpha_0(t)Q-\tfrac{1}{2}[\alpha_0^2(t)+|\alpha_0|^2]\}, \tag{5.1a}$$

where

$$\alpha_0(t) = \alpha e^{-i\omega_0 t}. \tag{5.1b}$$

The transition amplitude is given by

$$\langle\alpha_0(t)|\alpha(t)\rangle = \int_{-\infty}^{\infty}\Psi_{\alpha_0}^*(Q,t)\Psi_\alpha(Q,t)dQ. \tag{5.2}$$

Substitution of Eqs. (3.12) and (5.1) into Eq. (5.2) yields the probability

$$|\langle\alpha_0(t)|\alpha(t)\rangle|^2 = [2W(t)/|\lambda(t)|](\omega\omega_0)^{1/2}\times\exp\{-2[(\mathrm{Re}\,\alpha_0(t))^2+(\mathrm{Re}\,\alpha(t))^2]+\mathrm{Re}[1/\lambda(t)][(2\omega_0)^{1/2}\alpha_0^*(t)+(2\omega)^{1/2}W(t)\alpha(t)]^2)\}, \tag{5.3a}$$

where

$$\lambda(t) = \omega_0 + W^2(t)[\omega + i\delta(t)]. \tag{5.3b}$$

VI. DISCUSSION

A time dependent replacement of the usual linear damping coefficient γ has been made in Eq. (1.3), corresponding to the equation of motion given by Eq. (2.1). This corresponds to a simple harmonic pulsation in the dissipation which we may identify with a slight Rabi pulsation in the strength of the radiation field in the presence of a resonant atom. The model can arise classically as an *LCR* electrical circuit in which the damped pulsation is the result of pumping the system by varying the capacitance.[14]

The Heisenberg equations have been solved to first order in the pulsation. Higher-order perturbative solutions[4] are not very useful, but the motion in the neighborhood of the primary resonance $\nu = 2\omega(\omega^2 = \omega_0^2 - \gamma^2)$ can be investigated using the rotation-wave approximation, which gives an accurate soluton for large values of the time.[4] The gain coefficient is $|\rho| - \gamma$ and ρ has been evaluated in Eq. (2.22). A maximum is reached at exact resonance, as expected. If $\mu\gamma/\omega < 1$ the series M and N given by Eqs. (2.20b) and (2.20c) converge rapidly and $\rho > 0$. If, however, $\mu\gamma/\omega > 1$ there is a possibility that M or N could become negative. For sufficiently large values of the time, the system gains energy at resonance provided $|\rho_0| > \gamma$, i.e., $\mu > \gamma/\omega_0$ in the case when $\mu\gamma/\omega \ll 1$. Further work is needed to evaluate the exact cycle-averaged rate of absorption of energy[14] and to investigate how the system operates away from resonance.

The Schrödinger picture enables us to discuss some purely quantum-mechanical aspects, such as expectation values, the construction of quasicoherent states from pseudostationary states as exhibited in Eqs. (3.12), the relaxation of minimum uncertainty as in Eq. (4.5), and the transition probability from a quasicoherent state $|\alpha\rangle$ to a coherent state $|\alpha_0\rangle$. If the system is in the state $|\alpha(t)\rangle$ (a state of maximum coherence), then at time t it can be observed to be in the strictly coherent state $|\alpha_0(t)\rangle$ with a probability $|\langle\alpha(t)|\alpha_0(t)\rangle|^2$ given by Eqs. (5.3).

The gain in energy of amount $|\rho| - \gamma$ shown in Eq. (4.7c) or (4.8) occurs near resonance in any quantum-mechanical state. It may be seen as a classical result from Eqs. (2.1) and (2.28). We have demonstrated that the energy may be treated in a completely satisfactory way using the Kanai–Caldirola Hamiltonian.[19]

ACKNOWLEDGMENTS

The authors are greatly indebted to Dr. R. K. Colegrave for comments and stimulating discussions.

One of us (M. S. A.) is also grateful to the Mathematics Department, King's College, London, for its invitation during the summer term where this work has been achieved.

[1] M. Sargent, M. O. Scully, and W. E. Lamb, Jr., *Laser Physics* (Addison-Wesley, Reading, MA, 1974).
[2] R. K. Colegrave and M. S. Abdalla, Opt. Acta **30**, 849 (1983).
[3] R. K. Colegrave and A. Vahabpous-Roudsari, Opt. Acta **32**, 485 (1985).
[4] M. S. Abdalla, R. K. Colegrave, and A. Khosravi, Nuovo Cimento B **93**, 195 (1986).
[5] M. S. Abdalla and R. K. Colegrave, Phys. Rev. A **32**, 1958 (1985).
[6] P. Caldirola, Nuovo Cimento **18**, 393 (1941).
[7] E. Kanai, Prog. Theor. Phys. **3**, 440 (1948).
[8] R. W. Hasse, J. Math. Phys. **16**, 2005 (1975).
[9] B. Remaud and E. S. Hernandez, J. Phys. A: Math. Gen. **13**, 201 (1980).
[10] H. Dekker, Phys. Rep. **80**, 1 (1981).
[11] P. Caldirola, Nuovo Cimento B **77**, 241 (1983).
[12] R. K. Colegrave and M. S. Abdalla, J. Phys. A: Math. Gen. **14**, 2269 (1981).

[13]V. V. Dodonov and V. I. Man'ko, Phys. Rev. A **20**, 550 (1979).
[14]R. K. Colegrave, P. Croxson, N. B. A. Maskell, and U. A. T. Ramjit, Phys. Lett. A **125**, 359 (1987).
[15]H. Goldstein, *Classical Mechanics* (Addison–Wesley, Reading, MA, 1980).
[16]A.H. Nayfeh and D. T. Mook, *Nonlinear Oscillations* (Wiley, New York, 1979).
[17]M. S. Abdalla and R. K. Colegrave, Lett. Nuovo Cimento **39**, 373 (1984).
[18]M. S. Abdalla, Phys. Rev. A **34**, 4598 (1986).
[19]J. R. Ray, Am. J. Phys. **47**, 627 (1979).

The quantum relativistic two-body bound state. I. The spectrum

R. Arshansky and L. P. Horwitz
School of Physics and Astronomy, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel

(Received 4 November 1987; accepted for publication 27 July 1988)

In the framework of a manifestly covariant quantum theory on space-time, it is shown that the ground state mass of a relativistic two-body system with O(3,1) symmetric potential is lower when represented by a wave function with support in an O(2,1) invariant subspace of the spacelike region. The wave functions for the relativistic bound states are obtained explicitly. Coulomb type binding, the harmonic oscillator, and the relativistic square well are treated as examples. The mass spectrum is determined by a differential equation in the invariant spacelike interval ρ, which can be put into correspondence with the radial part of a nonrelativistic Schrödinger equation with potential of the same form, where r is replaced by ρ. In the case that the binding is small compared to the particle masses, the mass spectrum (bounded below) is well-approximated by the results of the nonrelativistic theory. The eigenfunctions transform under the full Lorentz group as elements of an induced representation with O(2,1) little group. This representation is studied in a succeeding paper.

I. INTRODUCTION

In nonrelativistic quantum mechanics, the use of Schrödinger's time-independent equation with central potentials for the study of bound states has been very successful in the description of atomic spectra and in the construction of wave functions as a basis for perturbation theory for the treatment of non-spherically symmetric interactions and radiation. A corresponding relativistic theory, with O(3,1) symmetric direct action potentials, could be expected to offer analogous applications, with the advantage of maintaining covariance, essential for consistency in the determination of mass spectra and for its application to radiation theory.[1] Such a theory should include the nonrelativistic results when the binding is small compared to the particle masses, and provide bounds for the applicability of the nonrelativistic theory.

In this paper, we shall study the bound state problem in the framework of a manifestly covariant quantum theory[2,3] that treats *events* (the occurrence of physical phenomena locally at space-time points), rather than *particles* (the occurrence of physical phenomena with functional dependence along world lines), as the fundamental physical entities.[4]

The construction of a manifestly covariant mechanics, both classical and quantum, of the type that we shall use, was carried out by Stueckelberg in 1941,[2] for the case of a single particle in an external field. He considered the phenomena of pair annihilation and creation as a manifestation of the development, in each case, of a single world line that curves in such a way that in one half-space of time the line passes twice, and in the other, not at all. To describe such a curve, parametrization by the variable t is ineffective, since the trajectory is not single valued. He therefore introduced a parametric description, with parameter τ along the world line. Hence one branch of the curve is generated by motion in the positive sense of t as a function of increasing τ, and the other branch by motion in the negative sense of t. The second branch is identified with the antiparticle, a rule that also emerged in Feynman's quantum electrodynamics.[2]

The motion, in space-time, of the point generating the world line, which we shall call an *event* (and has properties of space-time position and energy momentum), is governed in the classical case by the Hamilton equations in space-time

$$\frac{dx^\mu}{d\tau}=\frac{\partial K}{\partial p_\mu},\quad \frac{dp^\mu}{d\tau}=-\frac{\partial K}{\partial x_\mu},\tag{1.1}$$

where $x^\mu=(t,\mathbf{x})$, $p^\mu=(E,\mathbf{p})$ [we take $c=1$ and $g_{\mu\nu}=(-1,1,1,1)$] and the evolution generator K is a function of the canonical variables x_μ,p_μ. For the special case of free motion,

$$K_0=p^\mu p_\mu/2M,\tag{1.2}$$

where M is an intrinsic parameter assigned to the generic event, and hence

$$\frac{dx^\mu}{d\tau}=\frac{p^\mu}{M}.\tag{1.3}$$

It then follows that

$$\frac{d\mathbf{x}}{dt}=\frac{\mathbf{p}}{E},\tag{1.4}$$

consistent with standard relativistic kinematics. We note, however, that the mass squared $m^2=-p^\mu p_\mu$ is a dynamical variable since $\mathbf{p}$ and E are considered to be kinematically independent, and therefore it is not taken to be equal to a given constant. The set of values taken by m^2 in a particular dynamical context is determined by initial conditions and the dynamical equations.

In the quantum theory, $\mathbf{x},t$ (and $\mathbf{p},E$) denote operators satisfying the commutation relations (we take $\hbar=1$)

$$[x^\mu,p^\nu]=ig^{\mu\nu}.\tag{1.5}$$

The state of a one-event system is described by a wave function $\psi_\tau(x)\in L^2(R^4)$, a complex Hilbert space with measure $d^4x=d^3x\,dt$ satisfying the equation[2]

$$i\frac{\partial\psi_\tau(x)}{\partial\tau}=K\psi_\tau(x).\tag{1.6}$$

This equation, designed to provide a manifestly covariant

 0022-2488/89/010066-15$02.50

description of relativistic phenomena, is similar in form to the nonrelativistic Schrödinger equation. Although free motion is determined by the operator form of K_0 of Eq. (1.2), i.e., the d'Alembertian, which is hyperbolic ($p_\mu p^\mu \equiv -\partial_\mu \partial^\mu$ instead of the elliptic operator $\mathbf{p}^2 \equiv -\nabla^2$), the same methods may be used for studying Eq. (1.6) as for the nonrelativistic Schrödinger equation.

The unperturbed evolution of a free event is described by a wave packet of the form

$$\psi_\tau(x) = \int d^4p f(p) \exp\left\{-i\left(\frac{p^2}{2M}\right)\tau\right\} e^{ip\cdot x}, \tag{1.7}$$

where $p^2 = p^\mu p_\mu$, $p\cdot x = p^\mu x_\mu$. The stationary phase contribution to $\psi_\tau(x)$ (Ehrenfest motion) is at the point

$$x^\mu{}_c \simeq (p^\mu{}_c/M)\tau, \tag{1.8}$$

where $p^\mu{}_c$ is the peak value in the distribution $f(p)$. In the case where $p^0{}_c = E_c < 0$, we see, as in Stueckelberg's classical example, that

$$\frac{dt_c}{d\tau} \simeq \frac{E_c}{M} < 0. \tag{1.9}$$

It has been shown[4] in the analysis of an evolution operator with minimal electromagnetic interaction, of the form

$$K = (p - eA(x))^2/2M, \tag{1.10}$$

that the CPT conjugate wave function is given by

$$\psi^{\mathrm{CPT}}{}_\tau(\mathbf{x},t) = \psi_\tau(-\mathbf{x}, -t), \tag{1.11}$$

with $e \to -e$. For the free wave packet, one has

$$\psi^{\mathrm{CPT}}{}_\tau(\mathbf{x},t) = \int d^4p f(p) \exp\left\{-i\left(\frac{p^2}{2M}\right)\tau\right\} e^{-ip\cdot x}. \tag{1.12}$$

The Ehrenfest motion in this case is

$$x^\mu{}_c \simeq -(p^\mu{}_c/M)\tau; \tag{1.13}$$

if $E_c < 0$, we see that the motion of the event in the CPT conjugate state is in the positive direction of time, i.e.,

$$\frac{dt_c}{d\tau} \simeq -\frac{E_c}{M} = +\frac{|E_c|}{M}, \tag{1.14}$$

and one obtains the representation of a positive energy generic event with the opposite sign of charge, i.e., the antiparticle.[5]

Equation (1.6), with K of the form (1.10), leads to the conservation law

$$\frac{\partial \rho}{\partial \tau} = -\partial_\mu j^\mu(x), \tag{1.15}$$

where

$$\rho(x) = |\psi_\tau(x)|^2 \tag{1.16}$$

and

$$j^\mu(x) = -(ie/2M)\{\psi^*{}_\tau(x)(\partial^\mu - ieA^\mu(x))\psi_\tau(x) - ((\partial^\mu + ieA^\mu(x))\psi^*{}_\tau(x))\psi_\tau(x)\}. \tag{1.17}$$

It is clear from (1.15) that $j^\mu(x)$ cannot be the source of a Maxwell field since

$$\partial_\nu F^{\mu\nu}(x) = J^\mu(x) \tag{1.18}$$

implies that

$$\partial_\mu J^\mu(x) = 0. \tag{1.19}$$

As observed by Stueckelberg, who gave a geometrical argument in his 1942 paper[2] (or by application of the Riemann–Lebesgue lemma[3]) $\rho_\tau(x) \to 0$ as $\tau \to \pm\infty$, and hence, for asymptotically free motion,[4,6]

$$J^\mu(x) = \int_{-\infty}^{\infty} d\tau\, j^\mu{}_\tau(x). \tag{1.20}$$

Since particles are observed in the laboratory, directly or indirectly, by means of electromagnetic interaction, we see that the notion of a *particle* is associated with the entire world line, i.e., the *set of events* generated by the motion over all τ. We have called this construction, of an object that has the properties of a particle, from a set of events constituting the world line, "concatenation."[4]

For the treatment of systems of more than one event (generating world lines of more than one particle), one assumes the unperturbed evolution generator to be of the form[3]

$$K_0 = \sum_{i=1}^{N} \frac{p_i{}^2}{2M_i}. \tag{1.21}$$

In the presence of electromagnetic interaction (for spinless particles) one uses the minimal coupling form, which is a generalization of (1.10),

$$K = \sum_{i=1}^{N} \frac{(p_i - e_i A(x_i))^2}{2M_i}. \tag{1.22}$$

As pointed out above, there is a class of model systems, for which solutions can be achieved using straightforward methods, which involve only effective action-at-a-distance (direct action) potentials, where the evolution generator is of the form

$$K = \sum_{i=1}^{N} \frac{p_i{}^2}{2M_i} + V(x_1, x_2, \ldots, x_N). \tag{1.23}$$

Note that in this case the potential function enters into the dynamical evolution equation as a term added to the generator of the free motion, and therefore corresponds to a space-time coordinate-dependent interaction mass.

Equations (1.1) become

$$\frac{dx_i{}^\mu}{d\tau} = \frac{\partial K}{\partial p_{i\mu}}, \quad \frac{dp_i{}^\mu}{d\tau} = -\frac{\partial K}{\partial x_{i\mu}}. \tag{1.24}$$

The program is to solve the dynamical equation (1.6) with the dynamical evolution operator (1.22) or (1.23) [or Eqs. (1.24) for the classical case] governing the motion of events in interaction with each other and with external fields; predictions of observable phenomena are then obtained *a posteriori* by concatenation of the historical sequence of events. We shall concentrate on the direct action form (1.23) in this paper in our treatment of two-body bound states. As we shall see, the *relative* motion of bound states is represented by τ-independent wave functions (up to a phase). The center of mass (since the evolution generator is quadratic in energy momentum, one may always carry out a separation of variables for the center of mass motion) evolves as a free event, however, and concatenation then provides a world history of the two-body bound state that consists of a straight world line for the (Ehrenfest motion of) the center of mass associated with a stationary distribution for the relative motion.

Nonrelativistic Schrödinger potential theory implicitly synchronizes points on the particle trajectories by assuming that interaction occurs between them at equal times, i.e., in the potential $V(|\mathbf{r}_1 - \mathbf{r}_2|^2)$, where $\mathbf{r}_1$, the position of the first particle, and $\mathbf{r}_2$, of the second, are to be taken as positions on the trajectories at the same time t. This synchronization cannot be maintained in a relativistic framework. The synchronization of space-time *events*, corresponding to points along the particle world lines, can, nevertheless, be consistently and covariantly maintained by means of the universal evolution parameter τ. The two-body potential function, which we choose for Poincaré invariance to be of the form $V(\rho^2)$, where

$$\rho = \sqrt{(x_2^\mu - x_2^\mu)(x_{1\mu} - x_{2\mu})} \equiv \sqrt{(x_1 - x_2)^2},$$

carries the implication that the events x_1^μ and x_2^μ interact at equal τ, and hence implies the existence of a synchronization of events.[7]

There have been many attempts to deal with the relativistic bound state problem. The Bethe–Salpeter method[8] and related techniques,[9] based on structures provided by quantum field theory, have been successful in describing spectra to high precision.[10] The quantum mechanical interpretation of the wave function in these approaches is, however, not completely clear.

Constraint Hamiltonian dynamics, introduced by Dirac,[11] for dealing with singular Lagrangians of the type arising in gauge theories, has been developed for relativistic mechanics in both the classical and quantum cases.[12] The form of the interaction potentials, however, which must be used in this approach, is highly restricted by the integrability conditions; possible forms for more than two particles are difficult to construct, and are not known in general.[13]

One of the advantages of the constraint formalism is that, in scattering processes, the asymptotic expectation value of p_i^2 for each of the particles is ensured to be the correct "on shell" value.[12] In the unconstrained form of mechanics that we shall use, there is no restriction on the structure of the potential function (other than the requirement that the resulting differential equations are mathematically well-defined) for any number of particles. The asymptotic behavior of the expectation value of p_i^2 for each particle in a scattering process (or in ionization from a bound state) is related to the asymptotic synchronization of events in the universal historical time τ.[14] Transitions, such as between μ and e masses, are admitted in this framework.

Some authors have discussed the relativistic two-body bound state in a framework similar to the one we use here.[15] In these works, it was assumed that the relative motion is free to penetrate the entire spacelike region. We shall show that, for the O(3,1) symmetric Coulomb-type potential, the ground state wave function with support in an O(2,1) invariant subregion of the full spacelike region has a lower mass eigenvalue than the ground state wave function with support in the full spacelike region. This phenomenon corresponds to a spontaneous breakdown of the O(3,1) symmetry of the differential equations.

The support of the wave function determines the range of synchronization of the two-event system, and our computation of excited states assumes that this synchronization is characteristic of the bound states and persists, i.e., their support also lies in the O(2,1) invariant subregion. The resulting mass spectrum, for the case in which the binding is small compared to the mass of the particles (as, for example, in atomic physics), essentially coincides with the nonrelativistic Schrödinger energy spectrum for the corresponding $V(r^2)$, for arbitrary $V(\rho^2)$. The method used here is applicable as well to the problem of the strong binding of light particles, such as light quarks in a hadron. If, however, the binding exceeds a critical strength (in case there *is* an ionization point), we find that the simple notion of a bound state as a composite of two systems with intrinsic properties determined asymptotically above the ionization point is untenable. Techniques will be presented elsewhere to take into account the effects of spin.[16]

Since the support of the bound state wave functions lies in a restricted O(2,1) invariant sector of the full spacelike region, the representations they provide for the full O(3,1) space-time symmetry must be of induced type [it is shown in the Appendix that an O(2,1) ladder cannot be constructed in the Hilbert space]. Under Lorentz transformations, the (unit) spacelike vector n_μ for which O(2,1) is the stabilizer subgroup transforms through all spacelike directions and covers the complete single sheeted unit hyperboloid. Under such transformations, the wave functions undergo an action of the O(2,1) little group, and are modified along orbits parametrized by this unit vector.

The induced representation is constructed as a family of Hilbert spaces with measure spaces restricted to a family of O(2,1) invariant sectors. The parameter n_μ appears, in this respect, to play the role of a continuous superselection rule.[17] In a sequel to this paper,[18] to be called II, the representations of O(3,1) obtained in this way are studied by classifying states according to the eigenvalues of the operators generating an O(3) subgroup of O(3,1). It is shown there that these constitute the canonical representations of Gel'fand of the principal series; they are unitary in the larger Hilbert space in which all of the Hilbert spaces labeled by n_μ are embedded with measure $d^4n\,\delta(n^2 - 1)$.

In Sec. II, we formulate the problem of reduced motion in an O(3,1) symmetric potential, and obtain the eigenvalue equation for the relative mass spectrum as a radial equation of Schrödinger type, with invariant ρ as the "radial" coordinate, and the O(3,1) Casimir operator $\frac{1}{2}M_{\mu\nu}M^{\mu\nu}$ as the coefficient of the "centrifugal" term. In Sec. III, the differential equations after separation of variables are obtained for a parametrization in terms of two angles θ, ϕ, and a hyperbolic angle β which, along with ρ, cover what we shall call the [O(2,1) invariant] restricted Minkowski space (RMS), a region obtained as the exterior of two hyperplanes tangent to the light cone and oriented along the z axis. This region may be visualized by folding the x, y coordinates together; in the resulting three-dimensional space, these hyperplanes become planes and intersect along the z axis (Fig. 1). Alterna-

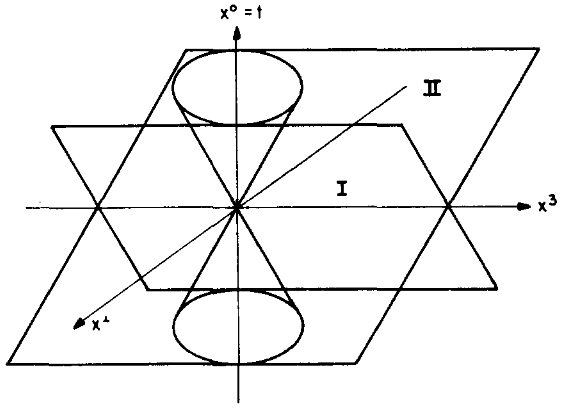

FIG. 1. The restricted Minkowski space (RMS) taken for the support of solutions of the eigenvalue equation in relative variables is designated as I, the region exterior to the two planes tangent to the light cone and intersecting along the x_3 axis ($\theta = 0,\pi$). The spatial coordinates x_1 and x_2 are folded into a single axis in this figure ($x_\perp$); in 3 + 1 dimensions the RMS is connected (but not simply connected, as seen from Fig. 2).

tively, we display this region in a projective space (Fig. 2).[19] The order of separation is first in ϕ, the azimuthal angle around the z axis, then in the O(2,1) boost parameter β to obtain the eigenvalue for the O(2,1) Casimir operator (the bound state levels are degenerate with respect to this quantum number). The separation equation for the remaining angle θ corresponds to the eigenvalue equation of the O(3,1) Casimir operator. The solutions and normalization conditions for these eigenvalue equations are given in Sec. IV. The separated equations for both θ and β variables have solutions that are associated Legendre functions, with "magnetic quantum number" determined by the O(2,1) Casimir. The separation function of θ has order determined by the O(3,1) Casimir. A geometrical interpretation is given in this section relating these quantum numbers to the usual nonrelativistic magnetic and orbital quantum numbers. In the nonrelativistic limit, these functions survive intact to play the usual role of the Legendre functions in the description of the nonrelativistic bound states.

In Sec. V the radial equation and invariant relative mass spectrum is discussed, and, in Sec. VI, we treat the examples of an O(3,1) invariant Coulomb-type potential (which reduces to the ordinary Coulomb potential in the nonrelativistic limit), the relativistic oscillator (where we find that no subsidiary conditions are required), and an O(3,1) invariant version of a square well. The lowest-order relativistic corrections to the corresponding nonrelativistic results are given in case the binding is small compared to the particle masses. For very large binding, exceeding a critical strength, we show that the simple idea of a bound state as a composite of two systems with intrinsic properties determined asymptotically above the ionization point (in case, as in the first and third examples, there is an ionization point) becomes untenable.

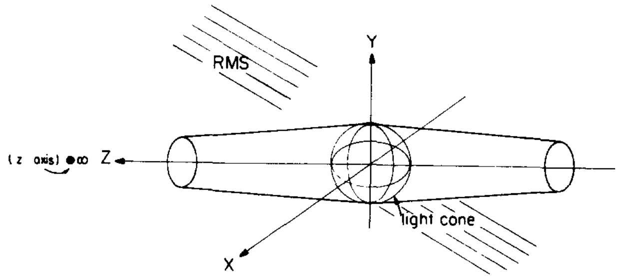

FIG. 2. The RMS in the projective space $\mathbf{R} = \mathbf{r}/t$; the unit sphere corresponds to the light cone. Each point corresponds to a line in Minkowski space. The point at ∞ along the Z axis is the z axis, and the point at $\mathbf{R} = 0$ is the t axis. The RMS is outside the cylinder $X^2 + Y^2 = 1$, i.e., $x^2 + y^2 \geqslant t^2$.

II. O(3,1) SYMMETRIC EQUATION OF MOTION AND THE EIGENVALUE EQUATION FOR REDUCED MOTION

We shall study in this section the evolution equation,[3]

$$i\frac{\partial}{\partial\tau}\Psi_\tau(x_1,x_2) = K\,\Psi_\tau(x_1,x_2), \tag{2.1}$$

where $(p_i^2 = p_i^\mu p_{i\mu} \equiv -\partial_i^\mu \partial_{i\mu})$,

$$K = p_1^2/2M_1 + p_2^2/2M_2 + V, \tag{2.2}$$

and $\Psi_\tau \in L^2(R^8)$.

We shall take the direct action potential V to have the O(3,1) symmetric form

$$V = V(\rho^2), \tag{2.3}$$

where

$$\rho^2 = (x_1 - x_2)^2 = (x_1 - x_2)^\mu (x_1 - x_2)_\mu. \tag{2.4}$$

We now separate the center of mass motion by defining the relative and center of mass variables with the natural choice[20]

$$P^\mu = p_1^\mu + p_2^\mu, \quad X^\mu = \frac{M_1 x_1^\mu + M_2 x_2^\mu}{M_1 + M_2}, \tag{2.5}$$

$$p^\mu = \frac{M_2 p_1^\mu - M_1 p_2^\mu}{M_1 + M_2}, \quad x^\mu = x_1^\mu - x_2^\mu, \tag{2.6}$$

where $m = M_1M_2/(M_1 + M_2)$ and $M = M_1 + M_2$. Equation (2.1) can be represented as a direct integral over Hilbert spaces $L^2(R^4)$, with measure d^4x, labeled by values of the absolutely conserved P^μ. One obtains the family of equations

$$i\frac{\partial}{\partial\tau}\Psi_{P'\tau}(x) = \left[\frac{P'^2}{2M} + K_{\rm rel}\right]\Psi_{P'\tau}(x). \tag{2.7}$$

In this way, we have separated out the center of mass motion. The operator K has, in general, continuous spectrum, but on the Hilbert spaces that are elements of the direct sum, i.e., for each value P'^μ, $K_{\rm rel}$ may have discrete or continuous spectrum. This spectrum corresponds to the contribution of the relative motion to the mass spectrum; we shall call it the "mass spectrum of the relative motion." We shall study the discrete spectrum of this operator, and the corresponding eigenstates.

For the discrete spectrum, we write

$$\Psi_{P'\tau}(x) = \exp(-i(P'^2/2M)\tau)e^{-iK_a\tau}\psi_{P'}^{(a)}(x); \tag{2.8}$$

Eq. (2.7) then becomes (we suppress reference to P' in the following)

$$K_a\psi^{(a)}(x) = (-(1/2m)\partial_\mu\partial^\mu + V(\rho^2))\psi^{(a)}(x). \tag{2.9}$$

The (invariant) relative radial coordinate can be separated from the angular and hyperbolic angular variables in the d'Alembertian with the help of the O(3,1) Casimir operator,

$$\Lambda = \tfrac{1}{2} M_{\mu\nu} M^{\mu\nu}, \tag{2.10}$$

where

$$M^{\mu\nu} = x^\mu p^\nu - x^\nu p^\mu. \tag{2.11}$$

With the help of the commutation relations (1.5) [valid also for the relative coordinates defined by (2.5)], one obtains

$$\Lambda = x^2 p^2 + 2i x\cdot p - (x\cdot p)^2, \tag{2.12}$$

where

$$x\cdot p \equiv x^\mu p_\mu = -i\rho \frac{\partial}{\partial \rho}, \tag{2.13}$$

valid for spacelike or timelike values of x^μ. We therefore see that ($\Box \equiv -\partial^\mu \partial_\mu, \rho^2 = x^\mu x_\mu$)

$$\Lambda = \rho^2 \Box + 3\rho \frac{\partial}{\partial \rho} + \rho^2 \frac{\partial^2}{\partial \rho^2}, \tag{2.14}$$

or

$$\Box = -\frac{\partial^2}{\partial \rho^2} - \frac{3}{\rho}\frac{\partial}{\partial \rho} + \frac{\Lambda}{\rho^2}. \tag{2.15}$$

Note that $\rho^2 \geqslant 0$ in the spacelike region [in the timelike region, ρ should be replaced by $i\rho'$, where $\rho'^2 = -x^\mu x_\mu$; in terms of the new variable ρ', it appears that the expression for $\Box$ in (2.15) has changed sign].

It then follows that Eq. (2.9) can be written as

$$K_a \psi^{(a)}(x) = \left\{ \frac{1}{2m} \left[-\frac{\partial^2}{\partial \rho^2} - \frac{3}{\rho}\frac{\partial}{\partial \rho} + \frac{\Lambda}{\rho^2} \right] + V(\rho^2) \right\} \psi^{(a)}(x). \tag{2.16}$$

III. SEPARATION OF VARIABLES

Further separation of variables depends on the choice of the sector of Minkowski space in which one studies the differential equation (2.9) and the corresponding parametrization of these sectors by hyperbolic angular (which we shall call hyperangular) and angular variables.[21] Each sector is associated with a spectrum determined by its structure and the boundary conditions applied to the solutions in that sector.

In addition to the more widely used decomposition of Minkowski space into the timelike and full spacelike regions, we shall use a further decomposition of the spacelike region into two subregions [invariant under an O(2,1) subgroup of O(3,1)]. One of these sectors (I) consists of the space-time points external (in spacelike directions) to two hyperplanes tangent to the light cone that are oriented along the z axis (the direction must be chosen to define this space). The second (II) consists of the space-time points in the sector interior (timelike direction) to these hyperplanes, but excluding the light cone. In Fig. 1 this decomposition is shown schematically by folding the two space axes x, y together (defining the coordinate $x_\perp$); in the resulting three-dimensional space, the two hyperplanes become planes and intersect along the z axis.

Alternatively, one may represent the light cone in a projective three-dimensional space[19] by dividing the equation $|\mathbf{r}|^2 - t^2 = 0$ by t^2 to obtain $|\mathbf{R}|^2 = 1$ ($\mathbf{R} = \mathbf{r}/t$), the equation for the unit sphere. The region I is characterized by $x^2 + y^2 - t^2 \geqslant 0$, translationally invariant in z. In the projective space, this region is mapped to $X^2 + Y^2 \geqslant 1$, the space exterior to the cylinder, parallel to the Z axis, which circumscribes the unit sphere. The space interior to the cylinder, excluding the unit sphere, corresponds to region II. We remark that the point at infinity on the Z axis ($z/t = \infty$) corresponds to the z axis, and the point at the center of the unit sphere ($\sqrt{x^2+y^2+z^2}/t = 0$) corresponds to the t axis. This representation is shown in Fig. 2.

The subgroup O(2,1) of O(3,1) leaving sectors I and II invariant has been used by Bargmann[22] as a little group for the construction of an induced representation of the Poincaré group with the direction of the z axis (momentum) providing the parameter along the orbit. In this construction, he used functions with support in the interior sector II. Zmuidzinas,[23] in his study of the unitary representations of the Lorentz group using differential equations, studied both the interior sector II and the exterior sector I. We shall see that solutions of Eq. (2.16) with support in the exterior sector I are associated with the physical bound states of the two-body problem with O(3,1) symmetric potential. We shall call this sector the restricted Minkowski space (RMS) oriented, as we have described it here, along the z axis.

The parametrization

$$\begin{aligned} x^0 &= \rho \sin\theta \sinh\beta, \quad x^1 = \rho \sin\theta \cos\phi \cosh\beta, \\ x^2 &= \rho \sin\theta \sin\phi \cosh\beta, \quad x^3 = \rho\cos\theta \end{aligned} \tag{3.1}$$

covers the RMS for $0 \leqslant \theta \leqslant \pi$, $0 \leqslant \phi < 2\pi$, $-\infty < \beta < \infty$, and $0 \leqslant \rho = \sqrt{|\mathbf{r}|^2 - t^2} < \infty$ (we shall use $\mathbf{x}$ and $\mathbf{r}$ interchangeably). We record, for comparison, the parametrization

$$\begin{aligned} x^0 &= \rho \sinh\beta, \quad x^1 = \rho\cosh\beta\cos\phi\sin\theta, \\ x^2 &= \rho\cosh\beta\sin\phi\sin\theta, \quad x^3 = \rho\cosh\beta\cos\theta, \end{aligned} \tag{3.2}$$

for the full spacelike region.

The properties of the wave functions and the spectrum of $K_{\rm rel}$ obtained in the full spacelike region[15] have important differences from those expected of physical bound states for spinless particles. In particular, separation of variables in the full spacelike region parametrized by (3.2) leads to degeneracy in $\mathbf{L}^2$ for *every* O(3,1) symmetric potential and the nonrelativistic limit of the spectrum obtained does not coincide with the nonrelativistic hydrogen spectrum in the case $V \propto 1/\rho$.

We therefore proceed to study Eq. (2.16) in the case that the wave functions have support in sector I, the RMS. Introducing the usual three-vector notation

$$L_i = \tfrac{1}{2}\epsilon_{ijk}(x^j p^k - x^k p^j), \tag{3.3}$$

$$A^i = x^0 p^i - x^i p^0, \tag{3.4}$$

for i,j,k running from 1 to 3, and ϵ_{ijk} the totally antisymmetric (unit) tensor in three dimensions, the nonvanishing O(3,1) Casimir operator (the second Casimir operator $\tfrac{1}{2}\epsilon^{\mu\nu\lambda\sigma} M_{\mu\nu} M_{\lambda\sigma} = A\cdot L$ is identically zero for the spinless case) is

$$\Lambda = \mathbf{L}^2 - \mathbf{A}^2. \tag{3.5}$$

In terms of the parameters of the RMS, the differential operators $\partial/\partial x^\mu$ are

$$\frac{\partial}{\partial x^0} = -\sin\theta\sinh\beta\frac{\partial}{\partial\rho} - \frac{1}{\rho}\cos\theta\sinh\beta\frac{\partial}{\partial\theta} + \frac{\cosh\beta}{\rho\sin\theta}\frac{\partial}{\partial\beta},$$

$$\frac{\partial}{\partial x^1} = \cos\phi\Big(\sin\theta\cosh\beta\frac{\partial}{\partial\rho} + \frac{1}{\rho}\cos\theta\cosh\beta\frac{\partial}{\partial\theta} - \frac{\sinh\beta}{\rho\sin\theta}\frac{\partial}{\partial\beta}\Big) - \sin\phi\frac{1}{\rho\sin\theta\cosh\beta}\frac{\partial}{\partial\phi}, \tag{3.6}$$

$$\frac{\partial}{\partial x^2} = \sin\phi\Big(\sin\theta\cosh\beta\frac{\partial}{\partial\rho} + \frac{1}{\rho}\cos\theta\cosh\beta\frac{\partial}{\partial\theta} - \frac{\sinh\beta}{\rho\sin\theta}\frac{\partial}{\partial\beta}\Big) + \cos\phi\frac{1}{\rho\sin\theta\cosh\beta}\frac{\partial}{\partial\phi},$$

$$\frac{\partial}{\partial x^3} = \cos\theta\frac{\partial}{\partial\rho} - \frac{1}{\rho}\sin\theta\frac{\partial}{\partial\theta}.$$

It then follows that

$$\Lambda = -\frac{\partial^2}{\partial\theta^2} - 2\cot\theta\frac{\partial}{\partial\theta} + \frac{1}{\sin^2\theta}N^2, \tag{3.7}$$

where

$$N^2 = L_3^2 - A_1^2 - A_2^2 \tag{3.8}$$

is the Casimir operator of the O(2,1) subgroup of O(3,1) leaving the z axis (and the RMS) invariant. In terms of the variables of sector I, this operator is given by

$$N^2 = \frac{\partial^2}{\partial\beta^2} + \tanh\beta\frac{\partial}{\partial\beta} - \frac{1}{\cosh^2\beta}\frac{\partial^2}{\partial\phi^2}. \tag{3.9}$$

We emphasize that these operators are not "restrictions," in the sense of projection, of the operators defined on functions with support on all of space-time, or on the full spacelike region. They are constructed as operators on functions with support in the RMS as their natural domain.

Since the operator Λ defined in (3.7) (and associated boundary conditions) is essentially different from the corresponding operator applicable to functions with support on the whole spacelike region, its spectrum is different as well.

The invariant measure in $L^2(R^4)$ on sector I of the Minkowski space is

$$d\mu = \rho^3\sin^2\theta\cosh\beta\, d\rho\, d\phi\, d\beta\, d\theta. \tag{3.10}$$

As a complete commuting set of dynamical variables, we use the subset of symmetric operators (we assume they are self-adjoint in the following and shall explicitly find their spectra),

$$\{K_{\rm rel}, L_3, N^2, \Lambda\}. \tag{3.11}$$

The generators of the O(2,1) subgroup, leaving the quadratic form $x_1^2 + x_2^2 - x_0^2$ invariant, are

$$H_\pm \equiv A_1 \pm iA_2 = e^{\pm i\phi}\Big(-i\frac{\partial}{\partial\beta} \pm \tanh\beta\frac{\partial}{\partial\phi}\Big) \tag{3.12}$$

and

$$L_3 = -i\frac{\partial}{\partial\phi}. \tag{3.13}$$

The remaining generators of O(3,1) are

$$A_3 = -i\Big(\cot\theta\cosh\beta\frac{\partial}{\partial\beta} - \sinh\beta\frac{\partial}{\partial\theta}\Big) \tag{3.14}$$

and

$$L_\pm \equiv L_1 \pm iL_2 = e^{\pm i\phi}\Big(\pm\Big(\cosh\beta\frac{\partial}{\partial\theta} - \sinh\beta\cot\theta\frac{\partial}{\partial\beta}\Big) + i\frac{\cot\theta}{\cosh\beta}\frac{\partial}{\partial\phi}\Big). \tag{3.15}$$

Let us take, for a solution of (2.16) in the RMS, the form

$$\psi(x) = R(\rho)\Theta(\theta)B(\beta)\Phi(\phi). \tag{3.16}$$

From (3.7) and (3.9) it follows that

$$\frac{\partial^2}{\partial\phi^2}\Phi_m(\phi) = -\Big(m+\frac{1}{2}\Big)^2\Phi_m(\phi), \tag{3.17}$$

i.e.,

$$\Phi_m(\phi) = (1/\sqrt{2\pi})e^{i(m+(1/2))\phi}, \quad 0\leqslant\phi<2\pi, \tag{3.18}$$

where we have indexed the solutions by the separation constant m.

For the case m integer, $\Phi_m(2\pi+\epsilon) = -\Phi_m(\epsilon)$ (it is a two-valued function); we shall see that this is the interesting case. One must, in fact, use (3.18) for $m\geqslant 0$ and $\Phi_m(\phi)^*$ for $m<0$.

The operator Λ in (2.16) contains N^2; with (3.17), the action of N^2 on (3.16) is determined, for $m>0$, by

$$N^2 B_{mn}(\beta) = \Big[\frac{\partial^2}{\partial\beta^2} + \tanh\beta\frac{\partial}{\partial\beta} + \frac{(m+\frac{1}{2})^2}{\cosh^2\beta}\Big]B_{mn}(\beta) = (n^2 - \tfrac{1}{4})B_{mn}(\beta), \tag{3.19}$$

where n^2 labels the separation constant for the variable β. The term $(m+\frac{1}{2})^2$ is to be replaced by $(m-\frac{1}{2})^2 = (|m|+\frac{1}{2})^2$ for $m<0$. We study explicitly only the case $m\geqslant 0$ in what follows.

As a final step in our treatment of the Casimir operator Λ in (2.16), it follows from (3.7) and (3.19) that

$$\Lambda\Theta(\theta) = \Big[-\Big(\frac{\partial^2}{\partial\theta^2} + \cot\theta\frac{\partial}{\partial\theta}\Big) + \frac{1}{\sin^2\theta}\Big(n^2 - \frac{1}{4}\Big)\Big]\Theta(\theta). \tag{3.20}$$

IV. SOLUTIONS OF THE ANGULAR AND HYPERANGULAR EIGENVALUE EQUATIONS IN THE RMS

In this section, we shall solve the eigenvalue equations obtained by separation of variables. For the treatment of Eq. (3.19), it is convenient to introduce the variable

$$\zeta = \tanh\beta, \tag{4.1}$$

where we note that $-1\leqslant\zeta\leqslant 1$, and define

$$B_{mn}(\beta) = (1-\zeta^2)^{1/4}\hat{B}_{mn}(\zeta). \tag{4.2}$$

Equation (3.19) then becomes

$$(1-\zeta^2)\frac{\partial^2\widehat{B}_{mn}(\zeta)}{\partial\zeta^2}-2\zeta\frac{\partial\widehat{B}_{mn}(\zeta)}{\partial\zeta}+\left[m(m+1)-\frac{n^2}{1-\zeta^2}\right]\widehat{B}_{mn}(\zeta)=0. \tag{4.3}$$

The solutions of this well-known equation are the associated Legendre functions of the first and second kind,[24] $P_m{}^n(\zeta)$ and $Q_m{}^n(\zeta)$.

The normalization condition for the wave functions (3.16) [with the measure (3.10)] is

$$1=\int\rho^3\,d\rho\,d\phi\,d\beta\,d\theta\sin^2\theta\times\cosh\beta\,|R(\rho)|^2|\Theta(\theta)|^2|B(\beta)|^2|\Phi(\phi)|^2 \tag{4.4}$$

and hence we must require that

$$\int\cosh\beta\,|B(\beta)|^2\,d\beta<\infty. \tag{4.5}$$

In terms of the variable ζ, this condition is

$$\int_{-1}^{1}(1-\zeta^2)^{-1}|\widehat{B}(\zeta)|^2\,d\zeta<\infty. \tag{4.6}$$

For $\nu>0$, and $\mu=0,1,2,\ldots$, one has[25]

$$\int_{-1}^{1}(1-\zeta^2)^{-1}|P_{\mu+\nu}{}^{-\nu}(\zeta)|^2\,d\zeta=\frac{1}{\nu}\frac{\Gamma(1+\mu)}{\Gamma(1+\mu+2\nu)}. \tag{4.7}$$

We shall show in an Appendix that the solutions for $\mu=m+n$ integer build the irreducible representations for the O(2,1) subgroup, which constitute the admissible physical states. The associated Legendre functions of the second kind do not satisfy the normalization condition (4.6).

We may choose for the normalized solutions (it is sufficient to consider only $n\geqslant 0$)

$$\widehat{B}_{mn}(\zeta)=\sqrt{n}\,\sqrt{[\Gamma(1+m+n)/\Gamma(1+m-n)]}\times P_m{}^{-n}(\zeta), \tag{4.8}$$

where $m\geqslant n$.

The case $n=0$ must be treated with some care. For $n=0$, the associated Legendre functions $P_m{}^{-n}(\zeta)$ become the Legendre polynomials $P_m(\zeta)$. The end points of integration in (4.6), $\zeta=\pm 1$, correspond to $\beta\to\pm\infty$. In terms of integration on β, e.g., in (4.5), the factor $\cosh\beta=(1-\zeta^2)^{-1/2}$ in the measure is canceled by the square of the factor $(1-\zeta^2)^{1/4}$ in (4.2), so the integration appears as

$$\int_{-\infty}^{\infty}|\widehat{B}(\zeta)|^2\,d\beta. \tag{4.9}$$

The Legendre polynomials do not vanish at $\zeta=\pm 1$, and hence if $\widehat{B}$ and P_m are related by a finite coefficient, this integral would diverge. When n goes to zero, the wave function spreads along the hyperbola labeled by ρ, going asymptotically to the light plane; the probability density with respect to intervals of β becomes constant for large $|\beta|$. Events associated with the two particles may therefore be found (for sufficiently large separation in space) with 2 + 1 lightlike separation out to remote regions of the tangent planes. To maintain the normalization, the Legendre functions must be multiplied by a vanishing factor, and the probability goes pointwise to zero (the wave function approaches a generalized eigenfunction). We shall therefore use, for this case, the function defined by

$$\widehat{B}_m(\zeta)=\sqrt{\epsilon}(1-\zeta^2)^{\epsilon/2}P_m(\zeta), \tag{4.10}$$

where it is understood that the limit $\epsilon\to 0$ is to be taken after the computation of scalar products; the factor $(1-\zeta^2)^{\epsilon/2}$ is a residue of the formula (AS 8.6.6; see also 8.1.4)

$$P_m{}^{-n}(\zeta)=(-1)^n(1-\zeta^2)^{n/2}\frac{d^n}{d\zeta^n}P_m(\zeta). \tag{4.11}$$

From (3.12) and (3.13), we see that the operators of O(2,1) leave θ invariant. We show in Appendix B that the functions $\{\widehat{B}(\zeta)\Phi(\phi)\}$ constitute the discrete series of irreducible projective representation of O(2,1), and that it is not possible to use these representations to construct a ladder representation of O(3,1).

We now turn to the solution of Eq. (3.20). Let us define the variable

$$\xi=\cos\theta \tag{4.12}$$

and the function

$$\widehat{\Theta}(\theta)=(1-\xi^2)^{1/4}\Theta(\theta). \tag{4.13}$$

Equation (3.20) then becomes

$$\frac{d}{d\xi}\left((1-\xi^2)\frac{d}{d\xi}\widehat{\Theta}(\xi)\right)+\left(l(l+1)-\frac{n^2}{1-\xi^2}\right)\widehat{\Theta}(\xi)=0, \tag{4.14}$$

where we have set

$$\Lambda=l(l+1)-\tfrac{3}{4}. \tag{4.15}$$

The solutions of Eq. (4.14) are proportional to the associated Legendre functions of the first and second kind, $P_l{}^n(\xi)$, $Q_l{}^n(\xi)$. For $n\neq 0$, the second kind of functions are not normalizable [the measure, according to (3.10) and (4.13) is the usual one for Legendre functions], and we therefore reject these. It follows from the requirement of unitarity for the representations of O(2,1) that we shall obtain, and normalizability, that l must be a non-negative integer (including 0) or positive half-integer.

To understand the geometrical and physical meaning of the quantum numbers l and n, consider the set of events parametrized by (3.1) with $\beta=0$ (these correspond to *equal time* correlations),

$$\begin{aligned}x^0&=0,\\x^1&=\rho\sin\theta\cos\phi,\\x^2&=\rho\sin\theta\sin\phi,\\x^3&=\rho\cos\theta.\end{aligned} \tag{4.16}$$

This set of events lies in a three-dimensional subspace parametrized by the usual spherical polar angles. The factor

$$Y_l{}^n(\theta,\phi)=(1/\sqrt{2\pi})e^{in\phi}\widehat{\Theta}_l{}^n(\cos\theta) \tag{4.17}$$

in the separated solution (3.16), where

$$\widehat{\Theta}_l{}^n(\theta)=\left(\frac{2l+1}{2}\frac{(l-n)!}{(l+n)!}\right)^{1/2}P_l{}^n(\cos\theta) \tag{4.18}$$

transforms under rotations according to

$$Y_l{}^{n}(\theta,\phi) = \sum_{n'} D^{l}{}_{nn'}(\eta_1,\eta_2,\eta_3)\, Y_l{}^{n'}(\theta',\phi'), \tag{4.19}$$

where the $D^{l}{}_{nn'}$ are the Wigner rotation functions of Euler angles η_1,η_2,η_3.[26] Note that the Legendre functions of the second kind do not admit this interpretation. We recognize that the Casimir operator of the Lorentz group labels the irreducible representation of the rotation group here, and the Casimir operator of the O(2,1) subgroup labels the magnetic quantum number corresponding to orientations of the three-dimensional space parametrized in (4.16). A general point in the RMS is obtained from such a representative point by performing a boost in the (x^1,x^2) plane. For

$$x_\perp \equiv \sqrt{(x^1)^2 + (x^2)^2} = \rho \sin\theta, \tag{4.20}$$

a boost with parameter β in the direction **x** results in

$$x'_\perp = x_\perp \cosh\beta,$$
$$x'^0 = x_\perp \sinh\beta, \tag{4.21}$$

corresponding to the general form (3.1) (for some ϕ). Each event in the three-space parametrized by (4.16) can be mapped in this way into a corresponding set of points in the RMS. Conversely, each point in the RMS is projected into this three-space by taking $\beta = 0$.

A reorientation of the three-dimensional space of (4.16) by the transformation (4.19) admits the same construction. A mapping from points represented in the reoriented space into general points in the RMS can be carried out by a set of active boosts in the *new* (x^1,x^2) plane.

The result of the reorientation of the three-dimensional equal time space is a reorientation of the entire RMS. After the transformation, the new RMS is constructed, with boundary planes tangent to the light cone, oriented along the new z axis (we shall show in II that all possible orientations must be considered in the specification of the two-body state).

V. THE RADIAL EQUATION AND INVARIANT SPECTRUM

The remaining "radial" equation obtained from (2.16) after separation of the angular and hyperangular variables, taking into account (4.15), is

$$\left[\frac{1}{2m}\left(-\frac{\partial^2}{\partial\rho^2} - \frac{3}{\rho}\frac{\partial}{\partial\rho} + \frac{l(l+1)-\frac{3}{4}}{\rho^2}\right) + V(\rho^2)\right] \times R^{(a)}(\rho) = K_a R^{(a)}(\rho). \tag{5.1}$$

Let us put

$$R^{(a)}(\rho) = (1/\sqrt{\rho})\hat{R}^{(a)}(\rho). \tag{5.2}$$

Equation (5.1) then becomes

$$\frac{d^2\hat{R}^{(a)}(\rho)}{d\rho^2} + \frac{2}{\rho}\frac{d\hat{R}^{(a)}(\rho)}{d\rho} - \frac{l(l+1)}{\rho^2}\hat{R}^{(a)}(\rho) + 2m(K_a - V(\rho^2))\hat{R}^{(a)}(\rho) = 0, \tag{5.3}$$

which is exactly of the form of the nonrelativistic spherically symmetric Schrödinger equation [the measure for the normalization of $\hat{R}$, according to (5.2) and (3.10), is just $\rho^2\, d\rho$]. The lowest mass eigenvalue for the case $V \propto 1/\rho$ occurs for the $l = 0$ state of the sequence $l = 0,1,2,3,\ldots$, and therefore the quantum number l plays a role analogous to that of orbital angular momentum in Eq. (5.3). In the interior region II, the spectrum of Λ is continuous.[27] In the full spacelike region, the last step of separation of variables associates the eigenvalues of Λ, which we have labeled with l, with a differential equation in the noncompact independent variable β [this can be seen from the structure of the parametrization (3.2) of the spacelike region, where β occurs in all four variables]. In this case,[28] $\Lambda = (l+\frac{1}{2})(l+\frac{3}{2}) - \frac{3}{4}$, for $l = 0,1,2,\ldots$, and hence the lowest achievable mass state is higher than the one we have obtained for wave functions with support in the RMS. This is the source of the spontaneous breaking of the O(3,1) space-time symmetry of the dynamical equations that selects the RMS subspace of the spacelike region.

For each nonrelativistic spherically symmetric potential problem, one obtains a corresponding direct action potential problem by the replacement of the relative radial coordinate r by ρ.

We shall argue below that the value of the full K operator (2.6) is usually determined (within a narrow interval) by intrinsic properties of the constituents. It then follows from the relation

$$K = P_a{}^2/2M + K_a \tag{5.4}$$

that the mass spectrum of the two-body system is determined by the spectrum K_a of the reduced motion. The two-body invariant mass squared (center of mass energy squared) is then given by

$$s_a \equiv -P_a{}^2 = 2M(K_a - K); \tag{5.5}$$

it is therefore quantized according to the spectrum of the relative motion, which coincides with the corresponding nonrelativistic energy spectrum.

Our argument that K is determined by intrinsic properties of the constituents is as follows. Transitions between bound state levels, involving changes in K_a, are induced by perturbation, such as coupling to electromagnetism. To treat such perturbations, we consider the addition of a τ-independent operator $\Delta V(x_1,x_2)$ that has non-negligible values in some limited space-time region (analogous to an adiabatic perturbation) near, for example, $x_i = 0$. Suppose, furthermore, that the wave function for the two-body system does not significantly overlap this perturbation for τ large and negative. It is in this range of τ values that we can consider the stationary bound state problem that we have studied here. At later τ, the wave function overlaps the perturbation, and transitions among the states of the stationary problem become possible. At large positive τ, the wave function no longer overlaps the ΔV and hence the system may again be found in a stationary state, perhaps different from the initial one (for example, radiation may have occurred). Since, however, ΔV is independent of τ, the value of K is conserved throughout the evolution. This situation is significantly different from the usual treatment of perturbations in nonrelativistic quantum theory, where the turning off and on of the perturbation in time causes transitions among values of the Hamiltonian operator. We therefore see that the relation between $P_a{}^2$ and K_a should be determined by (5.4), with a fixed value of K.

To determine this fixed value of K, we now suppose that the system is exposed to a τ-independent (but space-time-dependent) perturbation that brings the state of the system past the ionization point, if such a point exists. In this state, the constituent events may be separated by a large spacelike distance, where the potential is negligible (provided, as we shall see, a critical bound is not exceeded). Hence (see also Reuse[12])

$$K \simeq \frac{P^2}{2M} + \frac{p^2}{2m} = \frac{p_1{}^2}{2M_1} + \frac{p_2{}^2}{2M_2} \simeq -\frac{Mc^2}{2}, \tag{5.6}$$

where the last approximate equality follows from the assignment of each of the particles to a small interval in the neighborhood of its mass shell specified by its corresponding mass parameter M_i (if K varies over a small range, the two-body invariant mass squared varies over the same range; for each value of K, the quantization is determined by the discrete values of K_a). With (5.6), the mass squared spectrum (5.5) is

$$s_a \simeq M^2c^2 + 2MK_a. \tag{5.7}$$

If the nonrelativistic energy spectrum has values small compared to the particle rest masses, i.e., $|K_a| \ll Mc^2/2$, an *invariant* condition for nonrelativistic binding, the two-body center of mass energy spectrum is well approximated by

$$E_a \simeq Mc^2 + K_a - \tfrac{1}{2}K_a{}^2/Mc^2. \tag{5.8}$$

Up to the additive constant Mc^2, the center of mass energy thus coincides with the nonrelativistic energy spectrum to order $1/c^2$.

The families of functions Φ_m, B_{nm} for all values of m,n consistent with a given value of l form a degenerate set of solutions. The quantum numbers m,n of O(2,1) are a generalization of the magnetic quantum number that plays an analogous role in the corresponding nonrelativistic problem [the quantum number m changes under the action of the intrinsic O(2,1) subgroup].

It is interesting to note that the functions $\hat{R}$, $\hat{\Theta}$, and Φ have a correspondence interpretation. If the density $|\psi(x)|^2$ is used to study the expectation value of an observable that is a function on space-time that is independent of β [for example, a function of the O(2,1) invariant $x_1{}^2 + x_2{}^2 - x_0{}^2$], one may use the effective three-dimensional density given by [the probability of occurrence of an event in d^4x is $|\psi(x)|^2\, dx^0\, d^3x$]

$$\begin{aligned}\int |\psi(x)|^2 \frac{\partial x^0}{\partial \beta}\, d\beta &= \int |R(\rho)|^2 |\Theta(\theta)|^2 |\Phi(\phi)|^2 |B(\beta)|^2 \\ &\quad \times \rho \cosh\beta \sin\theta\, d\beta \\ &= (1/2\pi)|\hat{R}(\rho)|^2 |\hat{\Theta}(\theta)|^2,\end{aligned} \tag{5.9}$$

where $\hat{R}$, $\hat{\Theta}(\theta)$ coincide with the nonrelativistic wave functions (with the remaining measure $\rho^2\, d\rho \sin\theta\, d\theta\, d\phi$) for which ρ is the radial coordinate, and, as we found for the equal time correlation points at the end of Sec. IV, l is the orbital angular momentum, and n the magnetic quantum number [viz. (4.14) and (5.3)].

VI. SOME EXAMPLES

In this section, we give mass spectra for some exactly soluble problems, in particular, for the relativistic analog of the Coulomb potential, for which

$$V(\rho^2) = -Ze^2/\rho, \tag{6.1}$$

the four-dimensional space-time harmonic oscillator[29]

$$V(\rho^2) = \tfrac{1}{2}m\omega^2\rho^2, \tag{6.2}$$

and the relativistic analog of the three-dimensional square well potential, which has, in the relativistic case, a hyperboloidal boundary,[21] and for which

$$V(\rho^2) = \begin{cases} -U & \rho \leqslant a, \\ 0, & \rho > a. \end{cases} \tag{6.3}$$

In order to find the mass spectra and radial wave functions for these examples, it is not necessary to solve new differential equations. The radial equation (5.3) is exactly of the form of the corresponding nonrelativistic problem, and the solutions are known.

For the relativistic analog of the Coulomb potential, the relative mass spectrum is given by

$$K_a = -Z^2me^4/2\hbar^2(l+1+n_a)^2, \tag{6.4}$$

where $n_a = 0,1,2,\ldots$. The wave functions $\hat{R}\rho^{(a)}$ are the usual hydrogen functions[30]

$$\hat{R}_{n_a,l}(\rho) = C_{n_a,l}e^{-x/2}x^{l+1}L_{n_a}{}^{2l+1}(x), \tag{6.5}$$

where $L_{n_a}{}^{2l+1}$ are Laguerre polynomials. The variable x is defined by

$$x = (2Z\rho/a_0)/(n_a + l + 1), \tag{6.6}$$

where $a_0 = \hbar^2/me^2$, and

$$C^2{}_{n_a,l} = Z(n_a)!/(n_a+l+1)^2(n_a+2l+1). \tag{6.7}$$

The *size* of the bound state, which is related to the atomic form factor, is measured according to the invariant ρ. For the lowest level, $n_a = l = 0$,

$$\langle\rho\rangle_{n_a=l=0} = \tfrac{3}{2}a_0. \tag{6.8}$$

The total mass spectrum is then given by (5.7), i.e.,

$$s_{l,n_a} \simeq M^2c^2 - mMZ^2e^4/\hbar^2(n_a+l+1)^2. \tag{6.9}$$

For the case that the nonrelativistic energy spectrum has value small compared to the particle rest masses, we may use the approximate relation (5.8) to obtain

$$\begin{aligned}E_a \simeq Mc^2 - Z^2 \frac{me^4}{2\hbar^2(l+1+n_a)^2} \\ - \frac{1}{8}\frac{Z^4m^2e^8}{Mc^2\hbar^4(l+1+n_a)^4}.\end{aligned} \tag{6.10}$$

The lowest-order relativistic correction to the rest energy of the two-body system with Coulomb-like potential is therefore

$$\frac{\Delta(E_a - Mc^2)}{E_a - Mc^2} = \frac{Z^2\alpha^2}{4}\left(\frac{m}{M}\right)\frac{1}{(l+1+n_a)^2}. \tag{6.11}$$

For spinless atomic hydrogen ($Z = 1$), $\Delta(E - Mc^2) \simeq 9.7\times10^{-8}$ eV, and $E - Mc^2 \simeq 13.6$ eV for the ground state. The relativistic correction is therefore of the order of one part in 10^8. It is, however, about 10% of the hyperfine splitting $\hbar c/21$ cm $\simeq 9.4\times 10^{-7}$ eV. For positronium, $\Delta(E - Mc^2)\simeq 2\times 10^{-5}$ eV and $E - Mc^2 \simeq 6.8$ eV, so the relativistic correction is of the order of one part in 10^5. It is about 2% of the positronium hyperfine splitting $\frac{7}{6}\alpha^2$ Ry $\simeq 8.4\times 10^{-4}$ eV.[31]

For the four-dimensional harmonic oscillator, Eq. (5.3) has the form

$$\frac{d^2\hat{R}^{(a)}}{d\rho^2} + \frac{2}{\rho}\frac{d\hat{R}^{(a)}}{d\rho} + \left(\frac{2mK_a}{\hbar^2} - \frac{m^2\omega^2}{\hbar^2}\rho^2 - \frac{l(l+1)}{\rho^2}\right)\hat{R}^{(a)} = 0. \quad (6.12)$$

As for the nonrelativistic case, we make the transformation

$$\hat{R}^{(a)}(\rho) = x^{l/2}e^{-x/2}\, w^{(a)}(x), \quad (6.13)$$

where

$$x = (m\omega/\hbar)\rho^2, \quad (6.14)$$

to obtain

$$x\frac{d^2w^{(a)}}{dx^2} + \left(l + \frac{3}{2} - x\right)\frac{dw^{(a)}}{dx} + \frac{1}{2}\left(l + \frac{3}{2} - \frac{K_a}{\hbar\omega}\right) w^{(a)} = 0. \quad (6.15)$$

Normalizable solutions, the Laguerre polynomials $L_{n_a}^{l+1/2}(x)$, exist[30] when the coefficient of $w^{(a)}$ is a negative integer, i.e.,

$$K_a = \hbar\omega(l + 2n_a + \tfrac{3}{2}), \quad (6.16)$$

for $n_a = 0,1,2,\ldots$. The total mass spectrum is given by (5.7) (the choice of K is arbitrary here since there is no ionization point):

$$s_{l,n_a} = -2MK + 2M\hbar\omega(l + 2n_a + \tfrac{3}{2}). \quad (6.17)$$

For the case where the nonrelativistic energy spectrum has values small compared to K, which we surmise may be of the order of the particle rest masses,

$$E_a \simeq \sqrt{-2Mc^2K} + \hbar\omega\sqrt{(Mc^2/2|K|)}\,(l + 2n_a + \tfrac{3}{2}) - \tfrac{1}{2}(\hbar\omega)^2\sqrt{(Mc^2/8|K|^3)}\,(l + 2n_a + \tfrac{3}{2})^2. \quad (6.18)$$

Arbitrarily setting $K = -Mc^2/2$, one obtains

$$E_a \simeq Mc^2 + \hbar\omega\left(l + 2n_a + \frac{3}{2}\right) - \frac{1}{2}\frac{\hbar^2\omega^2(l + 2n_a + \frac{3}{2})^2}{Mc^2}. \quad (6.19)$$

Feynman, Kislinger, and Ravndal, Kim and Noz, and others[29] have studied the relativistic oscillator and obtained a positive spectrum [as in (6.17)] by imposing a subsidiary condition suppressing time excitations; although the mechanism is different, the restriction of the support of the wave functions to the O(2,1) invariant RMS plays an analogous role. No additional subsidiary condition is required; the set of solutions forms a complete orthogonal set in every Lorentz frame[32] (corresponding, in this case, to the induced representation to be described in II).

We now turn to the O(3,1) symmetric square well. In this case, the radial equation (5.3), with $V(\rho^2)$ given by (6.3), has solutions of the form (for $-U < K_a < 0$)[33]

$$\hat{R}^{(a)}(\rho) = \begin{cases} Aj_l(\sqrt{2m(K_a + U)/\hbar^2}\,\rho), & \rho \leqslant a, \\ Bh_l^{(1)}(i\sqrt{(-2mK_a)/\hbar^2}\,\rho), & \rho > a, \end{cases} \quad (6.20)$$

where j_l are spherical Bessel functions and $h_l^{(1)}$ are spherical Hankel functions of the first kind [the radial measure for $\hat{R}^{(a)}(\rho)$ is the same as for the nonrelativistic case]. Continuity of the wave function and its derivative with respect to ρ at the boundary $\rho = a$ provides the condition for the allowed values of K_a.

Let us call

$$\kappa_1 = \left(\frac{2m(K_a + U)}{\hbar^2}\right)^{1/2}, \quad \kappa_0 = \left(\frac{-2mK_a}{\hbar^2}\right)^{1/2}. \quad (6.21)$$

For $z_1 \equiv \kappa_1 a$, $z_0 \equiv \kappa_0 a \gg 1$, we may use the asymptotic forms

$$j^l(z) \sim (1/z)\cos(z - l\pi/2 - \pi/2), \qquad h_l^{(1)}(z) \sim (1/z)e^{i(z - l\pi/2 - \pi/2)}, \quad (6.22)$$

to obtain the eigenvalue conditions

$$-\cot\kappa_1 a \simeq \kappa_0/\kappa_1 \quad (l \text{ even}), \qquad \tan\kappa_1 a \simeq \kappa_0/\kappa_1 \quad (l \text{ odd}). \quad (6.23)$$

Since $\kappa_1^2 + \kappa_0^2 = 2mU/\hbar^2$, the large z_1, z_0 approximation requires that

$$\xi^2 \equiv (2mU/\hbar^2)a^2 \gg 1. \quad (6.24)$$

Defining

$$\epsilon = z_1 - \xi/\sqrt{2}, \quad (6.25)$$

the condition $\epsilon/\xi \ll 1$ then ensures, with (6.24), that z_0 and z_1 are both large. It then follows that

$$\frac{\kappa_0}{\kappa_1} = \left(\frac{\xi^2}{z_1^2} - 1\right)^{1/2} \simeq 1 - 2\sqrt{2}\,\frac{\epsilon}{\xi}. \quad (6.26)$$

For $\epsilon/\xi = 0$, solutions of (6.23) for l even are at $(4n - 1)/4\pi$ for integer $n \geqslant 1$, and for l odd, at $(4n + 1)/4\pi$ for integer $n \geqslant 0$. Expanding the trigonometric functions in the neighborhood of these values, and comparing with (6.25), we obtain

$$z_1(n) \simeq n\pi \mp \pi/4 - \sqrt{2}\epsilon/\xi$$

for l even or odd. Since, however, ϵ depends on z_1, we may substitute (6.25) and solve for $z_1(n)$, obtaining

$$z_1(n) \simeq (1 - \sqrt{2}/\xi)(n\pi + 1 \mp \pi/4) \simeq n\pi, \quad (6.27)$$

where $n\pi \gg 1$. Since $\epsilon/\xi = z_1/\xi - 1/\sqrt{2} \ll 1$, our solution is valid for values of n such that $n\pi/\xi \simeq 1/\sqrt{2}$.

For this set of high levels, the spectrum is given by

$$K_a \simeq -\{U - n_a^2\pi^2\hbar^2/2ma^2\}. \quad (6.28)$$

From (5.8), it follows that

$$E_a \simeq Mc^2 - \left(U - \frac{n_a^2\pi^2\hbar^2}{2ma^2}\right) - \frac{1}{2Mc^2}\left(U - \frac{n_a^2\pi^2\hbar^2}{2ma^2}\right)^2 \quad (6.29)$$

and the lowest-order relativistic correction to the relativistic spectrum is

$$\frac{\Delta(E_a - Mc^2)}{E_a - Mc^2} \simeq \frac{1}{2Mc^2}\left(U - \frac{n_a^2\pi^2\hbar^2}{2ma^2}\right). \quad (6.30)$$

The result (6.28) illustrates in a simple and explicit way a rather remarkable relativistic effect. Since an indefinite increase in the well depth U is in the framework of the approximation we have made in arriving at (6.28), which can be written alternatively as

$$K_a \simeq -\frac{\hbar^2}{2ma^2}\frac{\xi^2}{2}\left(1-2\sqrt{2}\,\frac{\epsilon(n)}{\xi}\right),$$

it is evident that the center of mass energy squared,

$$s_a = 2M(K_a - K), \tag{6.31}$$

can eventually become negative for any fixed value of K, for example, $-Mc^2/2$, as asserted in (5.6) (in this case, for $U \gtrsim Mc^2$). The argument leading to $K \simeq -Mc^2/2$ cannot, therefore, be justified in case the well depth U exceeds Mc^2 by a significant amount. This argument assumed that at, or above, the ionization point, the two particles may separate, and that the corresponding free motion can be consistent with $p_1{}^2 \sim -M_1{}^2c^2$ and $p_2{}^2 \sim -M_2{}^2c^2$. This would imply that the interpretation of the bound state as a composite system of the two particles with normal asymptotic behavior could be tenable. In this example, however, we see that if the potential well is sufficiently deep, this argument must fail, and ionization results in quasifree particle states for which the asymptotic values of $p_1{}^2$, $p_2{}^2$ must depend on the well depth (since the potential is bounded by a hyperboloid in space-time, only asymptotically approaching the light cone, it may be argued that unless there is compact support in t, there is always some small overlap of the wave function with the potential well no matter how large the spacelike separation). The drift of the particles out of the interaction region may be entirely suppressed, in fact, if there is a mechanism (such as self-energy) that induces a strong spectral enhancement of the asymptotic states of the two particles in the neighborhood of a definite value of the mass. In any case, the notion of a bound state as a composite of two particles with intrinsic properties determined in their free states becomes untenable when the binding potential is sufficiently strong. In this case, K must be treated as an unknown parameter, to be fixed to the observed spectrum. In the nonrelativistic limit, for which $c \to \infty$ (relative to all velocities), there is no U sufficiently large for this phenomenon to occur, and hence it must be understood as a relativistic effect.

The same phenomenon occurs for the Coulomb type potential, e.g., for Z sufficiently large, as can be seen from (6.9). The assignment of $K \simeq -Mc^2/2$ becomes untenable at

$$Z \gtrsim (M/\sqrt{M_1M_2})(1/\alpha). \tag{6.32}$$

If $M_1 \ll M_2$, the condition (6.32) becomes

$$Z \gtrsim \sqrt{(M_2/M_1)}\,(1/\alpha). \tag{6.33}$$

so that for one electron in the Coulomb field of a nucleus (for $M_2 \sim 2ZM_p$) the bound on Z for tenability of compositeness is very high ($\sim 5\times 10^5$).

For a system of two particles of equal mass parameter,

$$Z \gtrsim 2/\alpha, \tag{6.34}$$

which is of the order of magnitude of the value at which the spectrum of the Dirac equation becomes unstable. For a Coulomb-type strong interaction, where $\alpha \sim 1$, one sees that a simple picture of compositeness becomes questionable for any $Z \geqslant 1$.

VII. SUMMARY AND DISCUSSION

The eigenvalue equation for reduced motion (2.9), where $V(\rho^2)$ is an O(3,1) symmetric potential function, can be solved by separation of variables in the angular and hyperbolic angular coordinates (3.1) that range over the restricted Minkowski space (RMS) shown in Figs. 1 and 2 (the relativistic Coulomb-like problem can also be separated in hyperparaboloidal coordinates in this region; we shall discuss this procedure, along with the dynamical group of relativistic hydrogen, making use of a relativistic Runge–Lenz vector, elsewhere). The sequence of separation equations is in order ϕ,β,θ,ρ where β is a hyperbolic variable [in the full spacelike region, described by (3.2), the order of separation is ϕ,θ,β,ρ]. After the last stage of separation of variables, we are left with an equation in ρ that determines the spectrum. In the case of the full spacelike region, this radial equation depends on the separation constant for the β dependence; in the RMS, it is the separation constant for the θ dependence [which corresponds to the O(3,1) Casimir operator] that enters. In the nonrelativistic limit, O(3,1) is deformed to O(3) (the relative variables t, p^0 vanish in this limit), and the eigenvalues of the O(3,1) Casimir operator become eigenvalues of the O(3) Casimir operator, i.e., the angular momentum. Separation of variables in the RMS therefore has a clear correspondence to the nonrelativistic problem. The spectrum one finds in the full spacelike region and in the RMS are different. The lowest bound state in the RMS is lower than that found in the full spacelike region for $V \propto 1/\rho$, the relativistic generalization of the Coulomb potential.

Cook[15] has studied an equation similar to (2.9) with gauge invariant form for the electromagnetic interaction. In his approximations, the problem can be put into correspondence with the relativistic Coulomb potential problem we have studied. He obtains a mass spectrum proportional to a quantity of the form $-(n_a + l + \frac{1}{2})^{-2}$. This denominator is always half-integer squared and does not go to the Balmer form in the nonrelativistic limit. Its lowest value is higher than that of (6.4). As pointed out by Cook, the replacement of one of his quantum numbers (l) by a half-integer to compensate for this problem would lead to incorrect angular dependence.

Cook furthermore estimated the relativistic corrections both for Bohr–Sommerfeld quantization of his classical solutions (in the full spacelike region) and for a modified version of the treatment of the differential equations in the quantum case with extended sources admitting half-integer values for his analog of our n_a. He found that the $(\alpha/n)^4$ term [which we obtained in (6.10)] cannot be accounted for in his treatment.

The angular functions $P_l{}^n(\cos\theta)$ appearing in the solutions of the O(3,1) symmetric problem are in precise correspondence with those of the nonrelativistic case. The quantum number l specifies the O(3,1) Casimir operator, but it occurs in the relativistic radial equation in the same way that orbital angular momentum enters the nonrelativistic radial equation; in the nonrelativistic limit, O(3,1) is deformed to O(3), and l becomes the orbital angular momentum. The quantum number n specifies the O(2,1) Casimir operator; it becomes the magnetic quantum number in the nonrelativistic limit. The mass levels for the relativistic problem are degenerate in the O(2,1) quantum number, but not, in general, in l.

The restriction of the relative coordinates to the RMS corresponds to a restricted range of correlations available to the two events propagating in a bound state, i.e., to the range of $x_1{}^\mu - x_2{}^\mu$ available at each τ. We have assumed, in computing the full spectrum with functions whose support is restricted to the RMS, that this correlation is maintained for excited states as well.

The selection of wave functions defined on the O(2,1) invariant RMS corresponds to spontaneous symmetry breaking of the O(3,1) Lorentz invariance of the dynamical differential equation. The representations of O(3,1) generated by the solutions of the differential equation are, as we shall show in II, of induced type. Under the action of the full O(3,1), the solutions defined on the RMS specified by a spacelike unit vector (e.g., a unit vector along the z axis, as for the coordinate system used in this paper) undergo a Wigner transformation under the little group O(2,1), and are transported along an orbit parametrized by this spacelike vector whose range, under Lorentz transformation, is a single sheeted hyperboloid.

Due to the success of our choice of the RMS for the relativistic Coulomb problem, we have assumed that this region provides the correct correlations for two-body bound state O(3,1) symmetric potential problems in general, and a few examples are worked out.

Previous treatments of the relativistic harmonic oscillator problem[29] have imposed a subsidiary condition to ensure that timelike excitations are suppressed. Imbedding the bound state in the RMS instead of the full spacelike region eliminates the need for this condition. It replaces an explicit constraint by the introduction of coordinates whose free variation has sufficient structure to ensure that all excitations lie within a Hilbert space that has a consistent physical interpretation (positive norm); the spectrum corresponds to the excitations of just three harmonic degrees of freedom.

The relative mass eigenvalues of the relativistic square well potential problem were computed for a range of high levels for which the transcendental equations for the spectrum can be solved explicitly. It was found that, with the condition that the total K of the system takes on its asymptotic expected value for free particles approximately on mass shell above the ionization point, the well depth can be chosen sufficiently deep (in this case, $U \gtrsim Mc^2$) that the total invariant rest energy squared of the system can become negative. The assumption that the constituent particles behave asymptotically, above ionization, as free, therefore becomes untenable. A similar phenomenon occurs for Coulomb-type binding [at $Z \gtrsim (1/\alpha)M/\sqrt{M_1 M_2}$]. For particles of equal mass, this criterion is met at the order of magnitude at which the Dirac spectrum becomes unstable, but for an electron in the Coulomb-type field of a heavy nucleus, the bound is very high ($\sim 5\times 10^5$). For strong coupling, of the order $\alpha \sim 1$, the assumption that the constituents can be assigned on-shell values asymptotically becomes questionable for any $Z \geqslant 1$.

We emphasize that this critical value of the binding does not correspond to an instability in the spectrum of the dynamical evolution operator. It implies a limit to the depth of binding for which the simple notion of a bound state as a composite system of two particles with intrinsic properties determined as independent free particles above ionization becomes untenable. In the nonrelativistic limit, no bounded potential can produce this phenomenon, and hence it must be understood as a relativistic effect.

The solution of the problem of the relativistic bound state in an O(3,1) symmetric potential that we have given provides a *mass* spectrum that is the same as the corresponding nonrelativistic Schrödinger energy spectrum; this mass spectrum, up to the additive constant Mc^2, becomes the energy spectrum, and the wave functions acquire their usual nonrelativistic interpretation (for which l becomes the angular momentum, and n the magnetic quantum number), in the nonrelativistic limit. The structure of the theory therefore satisfies a correspondence principle.

ACKNOWLEDGMENTS

We wish to especially thank P. Winternitz for discussions at an early stage of this work, and V. Rittenberg for his encouragement. We are grateful to H. Bacry for very helpful discussions, and for his suggestion for the representation of the geometry of the RMS in projective coordinates. We also wish to thank our colleagues at Tel Aviv University, Syracuse University, and the University of Texas at Austin, particularly S. Nussinov, Y. Aharonov, I. Pietetski-Shapiro, M. Usher, B. Rosenstein, A. Kovner, A. Soffer, F. Rohrlich, A. Ashtekar, and W. C. Shieve, for valuable discussions. We are also pleased to have had the opportunity to discuss our results with C. Piron during his recent visit to Tel Aviv, and we wish to thank L. Lusanna and Y. S. Kim, respectively, for the opportunity to present them at the Workshop on Constraint's Theory and Relativistic Dynamics in Florence, 28–30 May 1986, and at the Workshop on Lorentz Transformations and Spacetime Geometry, University of Maryland, 27–30 May 1987.

This work was supported in part by the U.S.–Israel Binational Science Foundation, Jerusalem, Israel.

APPENDIX: DISCRETE SERIES OF IRREDUCIBLE REPRESENTATIONS OF O(2,1), THE QUANTUM NUMBERS, AND THE NONEXISTENCE OF A LADDER REPRESENTATION FROM THIS SERIES FOR O(3,1)

The representations of SO(2,1) and its double covering SU(1,1) have been studied by many authors.[22,34] Bargmann,[22] in particular, has discussed the basis functions with support in sector II, where $x_0{}^2 - x_\perp{}^2 \geqslant 0$. We are interested in the wave functions on a Hilbert space in the RMS, where $x_0{}^2 - x_\perp{}^2 \leqslant 0$.

We show explicitly in this Appendix that the solutions (3.18) and (4.8) that we have obtained for the β,ϕ parts of the differential equation (2.16) constitute the double-valued discrete series of irreducible projective representations of O(2,1).

The operators H_- and H_+ defined in Eq. (3.12) act as raising and lowering operators for the index m, since

$$[L_3, H_\pm] = \pm H_\pm . \tag{A1}$$

We now show that the $k=0$ element of the sequence (it is convenient to replace m by $n+k$)

$$\chi_{n+k}{}^{-n}(\zeta,\phi) \equiv B_{n+k,n}(\beta)\Phi_{n+k}(\phi) = (1-\zeta^2)^{1/4}\hat{B}_{n+k,n}(\zeta)\Phi_{n+k}(\phi) \tag{A2}$$

satisfies

$$H_-\chi_n{}^{-n}(\zeta,\phi) = 0. \tag{A3}$$

In terms of the variable $\zeta = \tanh\beta$,

$$H_\pm = e^{\pm i\phi}\left(-i(1-\zeta^2)\frac{\partial}{\partial\zeta} \pm \zeta\frac{\partial}{\partial\phi}\right), \tag{A4}$$

and Eq. (A3) becomes

$$\left[(1-\zeta^2)\frac{\partial}{\partial\zeta} + \left(n+\frac{1}{2}\right)\zeta\right]\chi_n{}^{-n}(\zeta,\phi) = 0. \tag{A5}$$

Using the relation (AS 8.6.17),

$$P_n{}^{-n}(\zeta) = \frac{1}{\Gamma(1+n)}\frac{(1-\zeta^2)^{n/2}}{2^n},$$

and (4.8), (A3) follows immediately.

We now study the action of H_+ on this lowest state:

$$\begin{aligned} H_+\chi_n{}^{-n}(\zeta,\phi) &= e^{i\phi}\left(-i(1-\zeta^2)\frac{\partial}{\partial\zeta} + i\zeta\left(n+\frac{1}{2}\right)\right)\chi_n{}^{-n}(\zeta,\phi) \\ &= i\sqrt{2n+1}\;\chi_{n+1}{}^{-n}(\zeta,\phi). \end{aligned} \tag{A6}$$

In general,

$$\begin{aligned} H_+\chi_{n+k}{}^{-n}(\zeta,\phi) &= i\sqrt{n}\left(\frac{\Gamma(1+k+2n)}{\Gamma(1+k)}\right)^{1/2}(1-\zeta^2)^{1/4} \\ &\quad\times\Big\{(n+k+1)\zeta P_{n+k}{}^{-n}(\zeta) \\ &\quad - (1-\zeta^2)\frac{\partial}{\partial\zeta}P_{n+k}{}^{-n}(\zeta)\Big\}\Phi_{n+k+1}(\phi). \end{aligned} \tag{A7}$$

It follows from (AS 8.5.3) and (AS 8.5.4) that

$$\begin{aligned} (1-\zeta^2)\frac{\partial}{\partial\zeta}P_{n+k}{}^{-n}(\zeta) &= (n+k+1)\zeta P_{n+k}{}^{-n}(\zeta) \\ &\quad - (2n+k+1)P_{n+k+1}{}^{-n}(\zeta), \end{aligned}$$

and hence (A7) becomes

$$H_+\chi_{n+k}{}^{-n}(\zeta,\phi) = i\sqrt{(k+1)(2n+k+1)}\,\chi_{n+k+1}{}^{-n}(\zeta,\phi). \tag{A8}$$

The Hermiticity of A_1, A_2 then implies that

$$\begin{aligned} (\chi_{n+k}{}^{-n}, H_-\chi_{n+k+1}{}^{-n}) &= (\chi_{n+k+1}{}^{-n}, H_+\chi_{n+k}{}^{-n})^* \\ &= -i\sqrt{(k+1)(2n+k+1)} \end{aligned}$$

and hence [since H_- can only lower the k value, according to (A1)]

$$H_-\chi_{n+k+1}{}^{-n} = -i\sqrt{(k+1)(2n+k+1)}\,\chi_{n+k}{}^{-n}. \tag{A9}$$

This result is, of course, consistent with the commutation relation

$$[H_+,H_-] = -2L_3, \tag{A10}$$

which follows from the formal commutation relations of the Lorentz group algebra

$$[M^{\mu\nu},M^{\alpha\beta}] = -i(g^{\nu\alpha}M^{\mu\beta} - g^{\beta\mu}M^{\alpha\nu} - g^{\nu\beta}M^{\mu\alpha} + g^{\alpha\mu}M^{\beta\nu}). \tag{A11}$$

For the O(2,1) subalgebra it follows from (A8) and (A9) that

$$(H_+H_- - H_-H_+)\chi_{n+k}{}^{-n} = -2(n+k+\tfrac{1}{2})\chi_{n+k}{}^{-n}. \tag{A12}$$

We now note that the complex conjugate of $\{\chi_{n\pm k}{}^{-n}\}$ transforms under $H_\pm$ in a similar way. We obtain in this way another, inequivalent, representation with the same value of the Casimir operator for O(2,1) [these elements correspond to the replacement of $m+\frac{1}{2}$ by $m-\frac{1}{2}$ for $m<0$ in (3.18) and (3.19); as we remarked after (3.19), we shall continue to consider m as positive]. Since the functions $B_{n+k,n}$ are real, we consider

$$\chi_{n+k}{}^{-n*}(\zeta,\phi) = (1-\zeta^2)^{1/4}\hat{B}_{n+k,n}(\zeta)\Phi_{n+k}{}^*(\phi). \tag{A13}$$

Since, according to (A4),

$$H_+{}^* = -H_-, \tag{A14}$$

it follows from (A2) that

$$H_+\chi_n{}^{-n*}(\zeta,\phi) = 0, \tag{A15}$$

and hence there is a sequence with a highest element. The Clebsch–Gordan coefficients are determined by (A8) and (A9). Using (A14), one obtains

$$\begin{aligned} H_-\chi_{n+m}{}^{-n*}(\zeta,\phi) &= i\sqrt{(k+1)(2n+k+1)}\,\chi_{n+k+1}{}^{-n*}(\zeta,\phi), \\ H_+\chi_{n+k+1}{}^{-n*}(\zeta,\phi) &= -i\sqrt{(k+1)(2n+k+1)}\,\chi_{n+k}{}^{-n*}(\zeta,\phi). \end{aligned} \tag{A16}$$

In fact, this complementary representation corresponds to charge conjugation. Since the operators A_1, A_2, L_3 are Hermitian, complex conjugation is equivalent to the transpose. Replacing the operators by their negative transpose, which corresponds to group theoretical charge conjugation (to be denoted by C), leaves the commutation relations invariant. Under this action,

$$\begin{aligned} H_-{}^C &= -H_+{}^* = H_-, \quad H_+{}^C = -H_-{}^* = H_+, \\ L_3{}^C &= -L_3{}^* = L_3, \end{aligned} \tag{A17}$$

where the last follows from (3.13). The two representations are therefore related by charge conjugation.

The O(2,1) Casimir operator defined in (3.12) is, in this set of representations, given by

$$\begin{aligned} N^2 = L_3{}^2 - A_1{}^2 - A_2{}^2 &= L_3{}^2 - \tfrac{1}{2}(H_+H_- + H_-H_+) \\ &= L_3(L_3-1) - H_+H_-. \end{aligned} \tag{A18}$$

With the help of (A8) and (A9) [or, correspondingly, (A16)], and the action of L_3, one obtains, as required by (3.19),

$$N^2 = n^2 - \tfrac{1}{4}. \tag{A19}$$

The unitary irreducible representations of O(2,1) are single or double valued, and hence m must be half-integer or integer, the latter corresponding to the double-valued representation. As we have seen, k is integer valued, and therefore n must be half-integer or integer, also. Normalizability conditions on the associated Legendre functions then require that l be, respectively, half-integer or integer. As we have remarked in Sec. V, the lowest mass state (for the soluble problems we have considered) corresponds to $l = 0$, and hence we shall only consider the integer values of l. This is consistent with our identification of the spectrum of K_a, of (5.3), with that of the corresponding nonrelativistic potential problem, and the correct behavior of the angular functions in that limit. We are therefore dealing with the double-valued representations of O(2,1).

In the following, we show that the operators A_3 and $L_\pm$ [which are not in the algebra of O(2,1)] move the set of eigenfunctions we have found out of the Hilbert space.

In terms of the variables ξ,ζ,ϕ, it follows from (3.14) and (3.15) that

$$A_3 = -i\left(\xi\left(\frac{1-\zeta^2}{1-\xi^2}\right)^{1/2}\frac{\partial}{\partial\zeta} + \zeta\left(\frac{1-\xi^2}{1-\zeta^2}\right)^{1/2}\frac{\partial}{\partial\xi}\right) \tag{A20}$$

and

$$L_\pm = -e^{\pm i\phi}\left(\pm\left(\frac{1-\xi^2}{1-\zeta^2}\right)^{1/2}\frac{\partial}{\partial\xi} \pm \left(\frac{1-\zeta^2}{1-\xi^2}\right)^{1/2}\zeta\xi \times\frac{\partial}{\partial\zeta} - i\left(\frac{1-\zeta^2}{1-\xi^2}\right)^{1/2}\xi\frac{\partial}{\partial\phi}\right). \tag{A21}$$

The action of these operators on the normalized eigenstates discussed above does not lead to a ladder representation for O(3,1) [unlike the case of the reduction O(3,1)⊂O(3) (Refs. 35 and 36)]. Let us study, for example, the action of A_3 on the normalized wave function $f_{l,m}{}^{-n}$ (taking again $m = n + k$),

$$f_{l,n+k}{}^{-n}(\theta,\beta,\phi) = \Theta_l{}^n(\theta)B_{n+k,n}(\beta)\Phi_{n+k}(\phi), \tag{A22}$$

where

$$\Theta_l{}^n(\theta) = \left[\left(\frac{2l+1}{2}\right)\left(\frac{(l-n)!}{(l+n)}\right)\right]^{1/2} P_l{}^n(\xi)(1-\xi^2)^{-1/4}. \tag{A23}$$

With the definitions (4.2) and (4.8), and (A20), we obtain

$$\begin{aligned} A_3 f_{l,n+k}{}^{-n}(\theta,\beta,\phi) &= -i\sqrt{n}\left(\frac{\Gamma(1+2n+k)}{\Gamma(1+k)}\right)^{1/2}\left[\left(\frac{2l+1}{2}\right)\left(\frac{(l-n)!}{(l+n)!}\right)\right]^{1/2}\Phi_{n+k}(\phi)\left\{\xi\left(\frac{1-\zeta^2}{1-\xi^2}\right)^{3/4}P_l{}^n(\xi)\frac{\partial}{\partial\zeta}P_{n+k}{}^{-n}(\zeta)\right.\\ &\quad\left.+\zeta\left(\frac{1-\xi^2}{1-\zeta^2}\right)^{1/4}P_{n+k}{}^{-n}(\zeta)\frac{\partial}{\partial\xi}P_l{}^n(\xi)\right\}. \end{aligned} \tag{A24}$$

Using identities for the associated Legendre functions,[36,37] we may write (A24) as

$$\begin{aligned} A_3 f_{l,n+k}{}^{-n}(\theta,\beta,\phi) &= \frac{i}{2}\left\{\left(\frac{k(2n+k+1)(l-n)(l+n+1)}{n(n+1)}\right)^{1/2}\right.\\ &\quad\times f_{l,n+k}{}^{-n-1}(\theta,\beta,\phi)\\ &\quad-\left(\frac{(1+k)(2n+k)(l+n)(l-n+1)}{n(n-1)}\right)^{1/2}\\ &\quad\left.\times f_{l,n+k}{}^{-n+1}(\theta,\beta,\phi,)\right\}. \end{aligned} \tag{A25}$$

This recursion relation is similar in form to that obtained from the ladder representation based on O(3) (Ref. 36) which, for the spinless case ($l_1' = 0$, so that $A_{l'} = 0$), is given by[36]

$$\begin{aligned} A_3\xi_{l',m'} &= C_{l'}\sqrt{l'^2 - m'^2}\,\xi_{l'-1,m'} - C_{l'+1}\\ &\quad\times\sqrt{(l'+1)^2 - m'^2}\,\xi_{l'+1,m'}, \end{aligned} \tag{A26}$$

where

$$C_{l'} = i\left(\frac{(l'^2 - l_0'^2)}{4l'^2 - 1}\right)^{1/2}, \quad A_{l'} = 0,$$

and $m' = -l', -l'+1,\ldots,l',\ l' = l_0', l_0'+1,\ldots$.

The correspondence can be easily seen by recalling that $k = m - n$, and that $n + \frac{1}{2}$ [which determines the value of the O(2,1) Casimir operator] should be put into correspondence with the angular momentum quantum number l' of O(3). Hence, in the sense of this correspondence,

$$k(2n+k+1) \sim m'^2 - l'^2, \tag{A27}$$

where $m' \sim m + \frac{1}{2}$. In the second coefficient, $k \to k - 1$ is equivalent to $l' \to l' + 1$. The second pair of factors in the radical of the first coefficient of (A25) corresponds to

$$(l-n)(l+n+1) \sim l_0'^2 - l'^2, \tag{A28}$$

where l_0', the lowest angular momentum of the corresponding tower of O(3) representations, is identified with $l + \frac{1}{2}$ [we are considering the $(l + \frac{1}{2}, 0)$ double-valued representation]. The corresponding factors of the second term are similarly obtained by the substitution $l' \to l' + 1$, inducing $n \to n - 1$.

The recursion relation (A25), however, cannot be used to generate a proper ladder representation based on O(2,1), since, for example, applying A_3 to $f_{l,n+k}{}^{-n}$ for $n = 1$ produces a term proportional to $f_{l,1+k}{}^{0}$. As we have pointed out in the discussion following Eq. (4.8), we can consider this function to be normalizable by the procedure of using the function $f_{l,1+k}{}^{-\epsilon}$ and taking the limit $\epsilon \to 0$ after integration.

The compensation for the singularity generated by the measure (3.10) for this function is obtained from the normalization factor in (4.8). The explicit appearance of the singularity $1/\sqrt{n-1}$ in (A25) for $n \to 1$ is precisely from this normalization. Since no such regularization procedure (by normalization) is available after operation with A_3, we see that this operator is not defined on $f_{l,1+k}{}^{-1}$; it shifts this function out of the Hilbert space.

[1]L. P. Horwitz, Found. Phys. **14**, 1027 (1984). See also L. C. Biedenharn and L. P. Horwitz, *ibid.* **14**, 953 (1984), for a discussion of the spin-$\frac{1}{2}$ case.

[2]E. C. G. Stueckelberg, Helv. Phys. Acta **14**, 372, 588 (1941); **15**, 23 (1942); R. P. Feynman, Phys. Rev. **80**, 440 (1950).

[3]L. P. Horwitz and C. Piron, Helv. Phys. Acta. **46**, 316 (1973); J. R. Fanchi, Phys. Rev. D **20**, 3108 (1979); L. P. Horwitz and Y. Lavie, *ibid.* **26**, 819 (1981), and references listed there.

[4]A discussion of the basic physical concepts, such as the notion of a particle, current, localization, and the discrete symmetries, have been given in R. Arshansky, L. P. Horwitz, and Y. Lavie, Found. Phys. **15**, 701 (1985). Applications to statistical mechanics are worked out in L. P. Horwitz, W. C. Schieve, and C. Piron, Ann. Phys. (NY) **137**, 306 (1981); L. P. Horwitz, S. Shashoua, and W. C. Schieve, Tel Aviv University preprint TAUP 1408-85, 1985. The existence of the wave operators in scattering theory is proven in L. P. Horwitz and A. Soffer, Helv. Phys. Acta **53**, 112 (1980); and completeness is proved in A. Soffer, Lett. Math. Phys. **8**, 517 (1984).

[5]This result can be obtained directly from examination of the classical equations.

[6]See also J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1975), 2nd ed., for the corresponding classical relation.

[7]The asymptotic synchronization of scattering states is associated, among other things, with inelasticity (e.g., conservation of individual particle masses); R. Arshansky and L. P. Horwitz, J. Math. Phys. **30**, 213 (1989).

[8]H. A. Bethe and E. Salpeter, Phys. Rev. **84**, 1232 (1951). See C. Itzykson and J.-B. Zuber, *Quantum Field Theory* (McGraw-Hill, New York, 1980), for a critical review.

[9]For example, M. Bander, D. Silverman, B. Klima, and U. Maor, Phys. Rev. D **29**, 2038 (1984); Phys. Lett. B **134**, 258 (1984).

[10]For a recent review, see G. T. Bodwin, D. R. Yennie, and M. A. Gregorio, Rev. Mod. Phys. **57**, 723 (1985).

[11]P. A. M. Dirac, Can. J. Math. **2**, 129 (1950); *Lectures on Quantum Mechanics*, Belfer Graduate School of Science, Monograph Series (Yeshiva University, New York, 1964). See, also, I. Todorov, JINR Report E2-10175, Dubna, 1976 and F. Rohrlich, Phys. Rev. D **23**, 1305 (1980).

[12]*Relativistic Action at a Distance: Classical and Quantum Aspects, Lecture Notes in Physics,* Vol. 162, edited by J. Llosa (Springer, Berlin, 1982), contains reviews and references. See, also, K. Sundermeyer, *Constraint Dynamics, Lecture Notes in Physics,* Vol. 169 (Springer, Berlin, 1982); L. P. Horwitz and F. Rohrlich, Phys. Rev. D **24**, 1528 (1981); **26**, 1452 (1982), and references listed there. H. Crater and P. Van Alstine, Ann. Phys. (NY) **148**, 57 (1983); Phys. Rev. D **30**, 2585 (1984), studied the bound state problem explicitly. See G. Longhi and L. Lusanna, preprint DFF 86n.30, Univ. di Firenze, 1986, for a useful review of Hamiltonian constraint mechanics in addition to its discussion of the bound state problem, scalar products, and conserved currents. For more recent work, see *Proceedings of the Workshop on Constraint's Theory and Relativistic Dynamics,* Arcetri, Firenze, Italy, 1986, edited by G. Longhi and L. Lusanna (World Scientific, Singapore, 1987).

[13]H. Sazdjian, Lett. Math. Phys. **5**, 319 (1981); F. Rohrlich, Ref. 11. See, also, Ph.Droz-Vincent, Lett. Nuovo Cimento **38**, 177 (1983); A. Komar, Phys. Rev. D **24**, 2330 (1981); L. Lusanna, Nuovo Cimento A **64**, 65 (1981); T. Biswas, F. Rohrlich, and M. King, *ibid.* **77**, 49 (1983); T. Biswas and F. Rohrlich, Lett. Math. Phys. **6**, 325 (1982), studied the three-body problem for a special choice of central three-body potential.

[14]R. Arshansky and L. P. Horwitz, Ref. 7; R. Arshansky, Tel Aviv University preprint TAUP 1479-86, 1986.

[15]S. Blaha, Phys. Rev. D **12**, 3921 (1975); J. L. Cook, Aust. J. Phys. **25**, 141 (1972). See, also, F. Reuse, Helv. Phys. Acta **53**, 416 (1980), where the interaction is taken to depend on $x^2 - (n\cdot x)^2 = (x - (n\cdot x)n)^2$ for n^μ timelike. After separating $n\cdot x$ ($\sim t$), one is left with an effectively three-dimensional problem.

[16]L. P. Horwitz, C. Piron, and F. Reuse, Helv. Phys. Acta **48**, 546 (1975); R. Arshansky and L. P. Horwitz, J. Phys. A: Math. Gen. **15**, L659 (1982).

[17]C. Piron, Helv. Phys. Acta **42**, 330 (1969).

[18]R. Arshansky and L. P. Horwitz, J. Math. Phys., in press (Paper II in series).

[19]We thank H. Bacry for suggesting this representation of the RMS to us.

[20]Other choices are possible. See, e.g., I. Todorov, Ref. 11 and in the first reference of Ref. 12.

[21]R. Arshansky and L. P. Horwitz, Phys. Rev. D **29**, 2860 (1984) used hyperbolic variables to achieve a separated solution of the O(1,1) symmetric potential problem in 1 + 1 dimensions.

[22]V. Bargmann, Ann. Math. **48**, 568 (1947).

[23]J. S. Zmuidzinas, J. Math. Phys. **7**, 764 (1966). We wish to thank P. Winternitz for bringing this important work to our attention.

[24]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1972), p. 332. We shall refer to specific formulas in this reference as (AS X.Y.Z) in the sequel.

[25]I. S. Gradstyn and I. M. Ryzhik, *Table of Integrals, Series and Products* (Academic, Orlando, 1980), 4th ed., p. 796.

[26]See, e.g., M. Hamermesh, *Group Theory and its Applications to Physical Problems* (Addison–Wesley, Reading, MA, 1962), p. 335.

[27]See, e.g., Zmuidzinas, Ref. 23.

[28]See Cook, Ref. 15.

[29]R. P. Feynman, M. Kislinger, and F. Ravndal, Phys. Rev. D **3**, 2706 (1971). See also, Y. S. Kim and M. E. Noz, *ibid.* **12**, 129 (1975); S. H. Oh, Y. S. Kim, and M. E. Noz, Found. Phys. **10**, 635 (1980); H. Leutwyler and J. Stern, Ann. Phys. (NY) **112**, 94 (1978); Nucl. Phys. B **133**, 115 (1978); **157**, 327 (1979).

[30]For example, L. Landau and E. Lifshitz, *Quantum Mechanics* (Pergamon, New York, 1958), Vol. III, p. 146ff.

[31]See, for example, C. Itzykson and J.-B. Zuber, Ref. 8.

[32]F. Rotbart, Phys. Rev. D **23**, 3078 (1981) has pointed out that the time excitations are necessary for maintaining completeness in a covariant framework.

[33]For example, see E. Merzbacher, *Quantum Mechanics* (Wiley, New York, 1970).

[34]L. C. Biedenharn, J. Nuyts, and N. Straumann, Ann. Inst. H. Poincaré **III**, 13 (1965); F. Boyce, J. Math. Phys. **8**, 6756 (1967); A. O. Barut and R. Raczka, *Theory of Group Representations and Applications* (Polish Scientific, Warsaw, 1977), pp. 365, 377.

[35]M. A. Naimark, *Linear Representations of the Lorentz Group* (Pergamon, New York, 1964).

[36]I. M. Gel'fand, R. A. Minlos, and P. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Application* (Pergamon, New York, 1963).

[37]N. Vilenkin, *Special Functions and the Theory of Group Representations* (Am. Math. Soc., Providence, RI, 1968), p. 144.

Extended Schrödinger equations leading to critical phenomena

B. Baseia
Departamento de Física-Centro de Ciencias Exatas e da Natureza, Universidade Federal da Paraîba, 58.000-João Pessoa, PB, Brazil

(Received 2 February 1988; accepted for publication 24 August 1988)

The role played by integrodifferential Schrödinger equations, as simulating critical phenomena, is investigated.

I. INTRODUCTION

There are many situations where the electrons in a condensed phase pass from a localized (bound) phase to an unlocalized (unbound) one. As examples, we mention the recently discovered polymers and oxides passing from an insulating phase to a conducting phase. Among the several attempts to explain a kind of phase transition one can cite Anderson[1] who employed a Schrödinger equation having a random potential in order to investigate critical phenomena. He showed that if the energies in a model solid were sufficiently random, some of the energy eigenfunctions, which for a regular solid would be Bloch states extending throughout the solid, become localized; an electron can no longer participate in electrical conduction.

In this paper we will employ a different scheme (and model) in order to investigate the possibility of simulating critical phenomena, through the use of extended integrodifferential Schrödinger equations (IDSE's).

II. MODEL

Take the following, one-dimensional, time-independent IDSE:

$$-\alpha\frac{d^2}{dx^2}\psi(x)+\int v(x,x')\psi(x')dx'=E\psi(x)\,, \qquad (1)$$

where $\alpha=h^2/2m$ and $v(x,x')$ is a kernel corresponding to a nonlocal potential.[2] According to the traditional interpretation, it includes correlations (due to interactions of the particle with its surrounding medium) between the particle placed at x and x'.

Next, set the nonlocal potential as

$$v(x,x')=g(x,x')G_\sigma(x-x') \qquad (2)$$

in such a way that

$$\lim_{\sigma\to 0}G_\sigma(\xi)=\delta(\xi)\,, \qquad (3)$$

which leads Eq. (1) to the local equation

$$-\alpha\frac{d^2}{dx^2}\psi+V(x)\psi(x)=E\psi(x)\,, \qquad (4)$$

where $V(x)=g(x,x)=v(x,x)/G_\sigma(0)$ is the local potential. So, to the extent of a constant $G_\sigma(0)=N(\sigma)$, the local potential coincides with the diagonal part of the kernel $v(x,x')$. (For integral equations the diagonal kernel is related to the Fredholm determinant.)

We call Eq. (1) an extended IDSE for the local Schrödinger equation given in Eq. (4).

There is an *a priori* large class of functions $G_\sigma(\xi)$ satisfying Eq. (3), yielding the same limit (for $\sigma\to 0$) of a given local problem. This would lead to ambiguities when defining an extended IDSE for a given local problem. However, it can be shown that this apparently large class of functions $G_\sigma(\xi)$ is strongly restricted by assuming, e.g., (i) the (Hermitian) property for the nonlocal potential $v(x,x')=\bar{v}(x',x)$; (ii) the asymptotic boundary condition: $v(x,x')\to 0$, if $|x-x'|\to\infty$; etc.

Next, make the (allowed) choice for $G_\sigma(\xi)$, as the Gaussian function

$$G_\sigma(\xi)=N(\sigma)e^{-\xi^2/2\sigma^2}\,, \qquad (5)$$

where $N(\sigma)=1/\sqrt{2\pi\sigma^2}$ and σ is the half-width for the Gaussian line shape. We take σ as the length of the nonlocal potential. In what follows, we will investigate the possible behavior of the parameter σ as a critical parameter.

Take the (well-known) local problem of electrons in a simple model of a one-dimensional periodic potential

$$-\alpha\frac{d^2}{dx^2}\psi+\lambda\sum_n\delta(x+na)\psi(x)=E\psi(x)\,, \qquad (6)$$

and make the IDSE extension

$$-\alpha\frac{d^2}{dx^2}\psi(x)\int v(x,x')\psi(x')dx'=E\psi(x)\,, \qquad (7)$$

with

$$v(x,x')=\left[\frac{\lambda}{N(\sigma)}\sum_n\delta(x+na)\delta(x'-x)+(x-x')^2\right]G(x'-x)\,. \qquad (8)$$

The substitution of Eq. (8) into Eq. (7) yields

$$-\alpha\frac{d^2}{dx^2}\psi(x)+\lambda\sum_n\delta(x+na)\psi(x)+\int(x-x')^2G_\sigma(x'-x)\psi(x')dx'=E\psi(x) \qquad (9)$$

which, in the limit for $\sigma\to 0$, coincides with Eq. (6). Hence Eq. (9) is an IDSE extension for the local Schrödinger equation (6).

The application of the Fourier transform (plus convolution theorem) to Eq. (9) reads

$$\alpha k^2\phi(k)+\lambda\sum_n e^{ikna}\psi(na)-\left[\frac{d^2}{dk^2}g_\sigma(k)\right]\phi(k)=E\phi(k)\,, \qquad (10)$$

where $\phi(k)$ and $g_\sigma(k)$ stand for the Fourier transforms of $\phi(x)$ and $G_\sigma(\xi)$, respectively; $g_\sigma(k)=n(\sigma)\exp(-\sigma^2k^2/2)$. Then

 0022-2488/89/010081-02$02.50

$$\phi(k) = -\frac{\lambda \sum_n e^{ikna}\psi(na)}{F(k)}, \tag{11}$$

where

$$F(k) = E - \alpha k^2 + (\sigma^2 k^2 - 1)\sigma^2 g_\sigma(k). \tag{12}$$

Hence

$$\psi(x) = \mathscr{F}^{1}[\phi(k)] = -\lambda \sum_n \psi(na) \int_{-\infty}^{\infty} \frac{e^{-ik(x-na)}}{F(k)} dk. \tag{13}$$

Since $kf(k) \to 0$ uniformly as $|k| \to \infty$, we can proceed using an analytical continuation in Eq. (13) and obtain

$$\psi(x) = -\lambda \sum_n \psi(na) \oint_c \frac{e^{-iz(x-na)}}{F(z)} dz, \tag{14}$$

where c is a semicircle with its center at the origin and radius $R \to \infty$. The semicircle is in the upper (lower) half-plane if $x - na < 0$ $(x - na > 0)$.

It is easy to note in Eq. (14) that if the zeros of $F(z)$ are real, which happens for $E < 0$ and σ obeying the inequality

$$\sigma^4 n(\sigma) > (3\alpha/4)e^{3/2}, \tag{15}$$

then the corresponding energy eigenfunctions are unlocalized $[\psi(x) \notin \mathscr{H}]$. On the other hand, for $E < 0$ and σ satisfying the reverse of inequality (15), the roots of $F(z) = 0$ are complex and the corresponding energy eigenfunctions are localized $[\psi(x) \in \mathscr{H}]$.

Hence there is a critical value for σ, satisfying the equation

$$\sigma_c^4 n(\sigma_c) = (3\alpha/4)e^{3/2}, \tag{16}$$

that separates a bound phase from an unbounded one.

A deeper investigation, into the realm of a phenomenological treatment, of details, alternative examples, and models, is a point that deserves attention in the present perspective and will be the subject of a future analysis.

ACKNOWLEDGMENT

Financial support by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) is acknowledged.

[1] P. W. Anderson, Phys. Rev. **109**, 1492 (1958).

[2] See, e.g., V. Volterra, *Theory of Functionals and of Integral and Integro-Differential Equations* (Dover, New York, 1959); G. Wataghin, Nuovo Cimento **11**, 635 (1934); Z. Phys. B **153**, 1377 (1967); M. Razavy, Nucl. Phys. **78**, 256 (1966).

Derivation of the real form of Schrödinger's equation for a nonconservative system and the unique relation between Re(ψ) and Im(ψ)

Robert L. W. Chen
Physics Department, Emory University, Atlanta, Georgia 30322

(Received 15 March 1988; accepted for publication 24 August 1988)

A unique relationship between the real part and the imaginary part of a wave function that obeys the time-dependent Schrödinger equation is derived. Thence the *real* form of the Schrödinger equation for the case of a nonconservative time-dependent potential $V = V(x,t)$ is obtained. Earlier work on this subject is found to be inadequate, applicable only to the conservative system. The results obtained here were first sought by Schrödinger but can be used in ways other than his original intention and purposes. Some unresolved issues relating to the nature of the dependency of $\mathrm{Im}(\psi)$ and $\mathrm{Re}(\psi)$ are also discussed.

I. INTRODUCTION

This paper will prove that the real part of a wave function carries full information on the state of a quantum system, as does the full complex wave function itself (assuming, of course, that the Hamiltonian is known). For every $\mathrm{Re}(\psi(x,t))$ there is one and only one $\mathrm{Im}\,(\psi(x,t))$. The latter can be obtained from the former by the use of an appropriate operator, to be shown below. Inseparable from this is the existence of a wave equation for $\mathrm{Re}(\psi)$ that Schrödinger called the "real wave equation."[1] Once the above relation is obtained, this equation automatically follows. It is only a matter of eliminating the $\mathrm{Im}(\psi)$ from the full Schrödinger equation. Schrödinger sought this relation and the real wave equation but was unable to obtain them in the case of a system with a time-dependent potential $V = V(x,t)$ (see Refs. 1–3). He obviously attached great importance to this matter, since he discussed this on at least three occasions (same references).

There is ample evidence that this relation is not generally known in the physics community. Textbooks on quantum mechanics often make statements that amount to a *denial* of its truth.[4–7] For example, one book[4] holds that one cannot obtain the probability density from $\mathrm{Re}(\psi)$ alone. This would imply that $\mathrm{Re}(\psi)$ contains *less* informational content than ψ. (See further remarks on these references.)

The issue came up while the author was working on computer graphics for wave mechanics.[8–10] Since it is possible to graph $\mathrm{Re}(\psi)$ but impossible (or at least extremely difficult) to graph the complex ψ, one wonders whether the former contains all the information on the quantum state. The author did prove this[11] and then found it to be in agreement with Schrödinger as found in Ref. 2. Most recently, the author found Ref. 1, which indicates that the nonconservative case has not been worked out. His earlier derivation needs to be reworked to achieve full generality.

Before proceeding to the derivations, we shall first discuss the general nature of the mathematical question at hand. The Schrödinger equation can be written as two coupled differential equations of two real variable [see Eq. (2) below]. There exist mathematical theorems on the conditions of equivalence of two first-order partial differential equations to one differential equation of second order. The best known example that meets the condition is the pair of Cauchy–Riemann (CR) equations

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y},$$
$$\frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.$$

They have the equivalent equation

$$\nabla^2 u = 0.$$

For comparison, we shall give a pair of equations that *do not* meet the requirement. From Courant and Hilbert,[12] we can find one by adding to the CR equations two terms on the right,

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} + a(x,y)v,$$
$$\frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} + b(x,y)v,$$

with the proviso that the two known functions a and b obey

$$\frac{\partial a}{\partial y} \neq \frac{\partial b}{\partial x}.$$

For these, no equivalent second-order equation exists.

According to Ref. 12, for equations of this kind, boundary conditions of the following type do not suffice:

$$u(0,y) = \Phi(y),$$
$$\left(\frac{\partial u}{\partial x}\right)(0,y) = \Psi(y),$$

where Φ and Ψ are two known functions. Rather than ensuring a unique solution, they yield a one-parameter family of solutions. The proper boundary condition will have to be on $u(0,y)$ and $v(0,y)$. The question is, then, to which category of equations does the Schrödinger equation belong? If [in the form of Eq. (2) below] it resembles the second pair of equations, then we could say that only u and v, not u and $\partial u/\partial t$, can specify the state of the system at a given time. In this case, one could justify a statement to the effect that the use of a real wave equation is not possible for wave mechanics.

Now the Schrödinger equation bears a formal resemblance to the CR equations [see (2) below]. There is nothing in it that resembles the av and bv terms that are present in the latter pair of equations. It is therefore plausible to assume it behaves like the CR equations, rather than the other equations. This, in fact, is what we are going to prove in the next section. An equivalent equation of second order in time, with $\partial^2/\partial t^2$, will be derived. It will satisfy Schrödinger's demand that "the state of the system is given by a real function and its time derivative" (Ref. 1, last paragraph).

II. DERIVATIONS

In Ref. 1, Schrödinger derived the real wave equation for the conservative case. In our notation it is

$$\frac{\partial^2 u}{\partial t^2} = -L^2 u, \tag{1}$$

where L is an operator,

$$L \equiv (1/\hbar)((\hbar/2m)\nabla^2 - V(x)),$$

$L^2 \equiv LL$, and u is Re (ψ). This is Schrödinger's Eq. (4) in Ref. 1. The complex wave equation is derived and is numbered (4″) (Ref. 1, p. 104) to distinguish it from his (4). Equation (4″) was destined to become the famous *Schrödinger equation*. It is ironic that Schrödinger himself *disliked* (4″) and preferred an equation of the type (4). At the end of Ref. 1 he remarked that "there is no doubt a certain crudeness in the use of a complex wave function . . . in (4″), we have before us only the substitute . . . for a real wave equation of probably the fourth order, which, however, I have not succeeded in forming for the non-conservative case."

This difficulty can be attributed to the lack of a proper, unique relationship that could link $\mathrm{Im}(\psi)$ to $\mathrm{Re}(\psi)$, a fact noted in Ref. 2. In Ref. 2, Schrödinger considered the possibility to "define the imaginary part unambiguously . . . referring only to the real quantity itself and its time and space derivative," and said it can be done except that "integration with respect to time would involve an undetermined wave function. *I do not know yet whether this can be fixed in a rational way.*" (See Ref. 2, pp. 56 and 57.)

Let us examine the relation between $\mathrm{Re}(\psi)$ and $\mathrm{Im}(\psi)$, to be denoted by $u(x,t)$ and $v(x,t)$. From the complex wave equation one obtains a pair of equations,

$$\frac{\partial u}{\partial t} = L_t v, \tag{2a}$$

$$\frac{\partial v}{\partial t} = -L_t u. \tag{2b}$$

(The complex equation is $i\,\partial\psi/\partial t = L_t\psi$.) The operator L_t is the same as L but with V dependent on time: $V(x,t)$. If one uses (2b), one could integrate with respect to time to find v, and the result would contain an arbitrary function, the same as what was found in Ref. 2. If we use (2a) instead, we may be able to find a unique inverse operator of L_t, which we shall call $(L_t)^{-1}$. Then we will get

$$v = (L_t)^{-1}\frac{\partial u}{\partial t}, \tag{3}$$

which is what we want. The nature of L_t and $(L_t)^{-1}$ can be revealed by putting them into matrix forms, and they will be diagonal if we adopt the *energy* representation, i.e., in eigenfunctions of L_t. We have

$$L_t f_i(x,t) = [E_i(t)/\hbar] f_i(x,t). \tag{4}$$

Equation (4) is the equation of the eigenfunction. Note that $f_i(x,t)$ is real. It is the same as the eigenfunction of L if $V(x)$ in L is equal to $V(x,t)$ at time t. The matrix is

$$L_{ij}(t) = [E_i(t)/\hbar]\delta_{ij}. \tag{5}$$

The inverse should be

$$(L_t)^{-1} = [\hbar/E_i(t)]\delta_{ij}. \tag{6}$$

There is no problem with any nonvanishing $E_i(t)$. For $E_i(t) = 0$ to be called $E_0(t)$, there can be an ambiguity. Any term $C_0(t) f_0(x,t)$ in an expansion $f = \Sigma_i C_i(t) f_i(x,t)$ will become zero on being operated on by L_t independent of $C_0(t)$. [We are assuming that $V(x,t) \to 0$ for $x \to \pm\infty$.] This problem can be solved if we restrict the domain of L_t to functions that are normalizable: $\int f^2\,dx =$ finite. Then if we expand f into $f_i(x,t)$, there will be a spectrum that is discrete for $E_i(t) < 0$ and continuous for $E_i(t) \geqslant 0$. The continuous part must be a function of finite value everywhere. The range of L_t must be a function that, if expanded into $f_i(x,t)$, should have a continuous part of the spectrum, which tends to zero as $E_i(t)$ goes to zero, i.e., reaching zero at the lower end of the continuous spectrum. Since it is forbidden for the spectrum to accumulate at $E_0(t)$, L_t and $(L_t)^{-1}$ are unique. Now we examine (3). The range of L_t has become the domain of $(L_t)^{-1}$. The question then arises, does $\partial u/\partial t$ belong to this domain? We can see that it does. In (2a), v must be normalizable and have a finite continuous spectrum, and $L_t v$ must have a spectrum that tends to zero at the lower end of its continuous part. Since $\partial u/\partial t$ is equal to $L_t v$, it must have this same property. Therefore, $\partial u/\partial t$ belongs to the admissible domain of $(L_t)^{-1}$. Thus the transformation of (3) from u to v is unique, not for any $u(x,t)$ function but for one that is the real part of a *normalized solution* ψ of a Schrödinger equation.

To obtain the real wave equation is now straightforward. We can simply take $\partial/\partial t$ on both sides of (2a) and get

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial}{\partial t}(L_t v).$$

Since the Laplacian in L_t commutes with $\partial/\partial t$, whereas the $V(x,t)$ part does not, exchanging the order of $\partial/\partial t$ and L_t on the right side would yield an extra term. We get

$$\frac{\partial^2 u}{\partial t^2} = L_t\frac{\partial v}{\partial t} + \frac{\partial V}{\partial t}v. \tag{7}$$

Using (2b) for $\partial v/\partial t$ and (3) for v, we get

$$\frac{\partial^2 u}{\partial t^2} = -L_t^2 u + \frac{\partial V}{\partial t}L_t^{-1}\frac{\partial u}{\partial t}. \tag{8}$$

This is the equation Schrödinger sought. It is indeed of the fourth order as he expected, since there is the L^2 operator. Note that the presence of $\partial V/\partial t$ in the last term was foreseen by him. (See Ref. 1, the middle of p. 104.) The boundary conditions for (8) should be of the Cauchy type for the spacelike surface at the initial time.

III. COMMENTS ON ITS SIGNIFICANCE

The main objective of this article is to prove the mathematical results shown in Sec. II. It seems desirable, however, to discuss its significance in relation to other issues of interest in quantum physics.

A. Schrödinger's motivation in seeking this result

From Schrödinger's writings[1–3] it is clear that it has to do with his epistemological view that a wave function represents some entity in reality, not just a potentiality as seen by others who later became known as the "Copenhagen school." As is well known, there are other important scientists who hold views similar to Schrödinger's; these will be referred to as "Group I" and the Copenhagen school as "Group II." From the standpoint of Group I, the relation Schrödinger sought is a very desirable one. If u and $\partial u/\partial t$ represent the state of a system containing all the information on the state, then $u(x,t)$ can be construed as representing the sole reality independently, without the use of $v(x,t)$. It is like the wave function of a classical wave, which represents a real local property in space and time. We wish to point out, however, that this result, though favorable to Group I, is *by no means inconsistent* with the views of Group II. Our proof is based on the nature of the continuous dynamics of the quantum-mechanical wave *uninterrupted* by the process of measurement. The collapse of the wave function remains a valid concept even if the wave is represented by a real function. By the same token, our relation should be valid *independently* of whether Group I or Group II is correct, because it is a purely mathematical result. If one disagrees with Schrödinger's epistemological view, one need not object to the $\mathrm{Re}(\psi)$–$\mathrm{Im}(\psi)$ relation that he believed should exist. It can, in fact, be used in other ways totally unrelated to epistemology—for instance, in making graphics of wave functions as we already alluded to.

B. Relation of the result to many statements in quantum mechanics literature

We select some typical examples from some textbooks.[4–7] It is often said that while the wave function of a classical wave is real, its counterpart in quantum mechanics must be complex. To be sure, the use of complex ψ brings simplicity and symmetry to the form of the wave equation. But this advantage does not manifest itself until one goes beyond the nonrelativistic regime with a conservative potential. For example, in Dirac's equation the operators $i\,\partial/\partial x_\mu$, $\mu = 1,2,3,4$, can each be paired with the electromagnetic potential A_μ. Now Refs. 4–7 seek an explanation of this from the mathematical nature of the nonrelativistic wave equation of a particle in conservative fields—in fact, usually a free particle with $V = 0$. Here the above reason as a possible justification for complex ψ is not applicable. Attempts to justify it lead to unwarranted statements. Thus Refs. 5 and 6 say it is impossible to construct a real wave equation, directly contradicting Eq. (1). Reference 7 does not make explicit statements on this as do Refs. 5 and 6, but the fact that its argument also starts from the wave motion of a free particle does *indirectly* imply that $\mathrm{Re}(\psi)$ is in some sense incomplete. Otherwise what could be the linkage of free particle wave motion and the complexity of ψ? In fact, the only conceivable reason one can give is that by adopting a complex ψ, one can avoid having to write wave equations of the *fourth* differential order, like Eq. (1). Yet the fourth-order wave equation had long been in use prior to the advent of wave mechanics, i.e., the equation for the vibrations of a solid plate.[1–3] It is, in fact, as Schrödinger pointed out, mathematically equivalent to the quantum wave equation of a free particle.[3]

IV. SOME ISSUES YET TO BE RESOLVED

The main question of interest is as follows: Can we formulate wave mechanics in terms of real wave function without inconsistency? There are some issues that must be resolved before one can give an affirmative answer.

Let us consider the nature of the dependence of v on u, namely, Eq. (3). First it can be shown to be *nonlocal*. Consider, for example, the special case of a free particle. The operator L_t now becomes $\nabla^2/2m$, and f_i the real part of $e^{i\mathbf{k}\cdot\mathbf{x}}$ with suitable constant factor. The eigenvalue E_i becomes $\hbar k^2/2m$. Equation (6) reduces to

$$(L_t)^{-1} = (1/k_i^2)\delta_{ij}\,. \tag{9}$$

To apply $(L_t)^{-1}$ to $\partial u/\partial t$ means first to decompose the latter into $e^{i\mathbf{k}\cdot\mathbf{x}}$ and then to multiply each term by $1/k_i^2$, i.e.,

$$\begin{aligned} v(\mathbf{x},t) &= \frac{1}{2\pi}\int \frac{e^{i\mathbf{k}\cdot\mathbf{x}}}{k^2}\int \frac{\partial u}{\partial t}\, e^{i\mathbf{k}\cdot\mathbf{x}'}\, d^3x'\, d^3k \\ &= K^2 \int \frac{1}{|\mathbf{x}-\mathbf{x}'|}\frac{\partial u}{\partial t}\, d^3x'\,, \end{aligned} \tag{10}$$

where we have used the identity[13]

$$\frac{1}{|\mathbf{x}-\mathbf{x}'|} = \frac{1}{2\pi^2}\int d^3k\, \frac{e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}}{k^2}\,,$$

and K^2 represents some appropriate constant.

Equation (10) can also be obtained directly by using the electrostatic analogy: $\partial u/\partial t$ to v is as the charge density ρ to the potential ϕ, and (10) is simply the Poisson solution of ϕ.[14] Equation (10) can be generalized to systems other than the free particle by replacing $1/|\mathbf{x}-\mathbf{x}'|$ by $G(\mathbf{x},\mathbf{x}',t)$, where G is the Green's function of the time-independent Schrödinger equation with its potential V set equal to $V(\mathbf{x},t)$ of our system:

$$v(\mathbf{x},t) = K^2 \int G(\mathbf{x},\mathbf{x}',t)\,\frac{\partial u}{\partial t}\, d^3x'\,. \tag{11}$$

Thus the dependence of v on $\partial u/\partial t$ is not only *nonlocal* but is also *dependent on the nature of the system*, since G varies with the system's Hamiltonian. One could then raise the following questions.

(1) The probability density P, which is $\psi^*\psi = u^2 + v^2$, is determined solely by the local values of u and v in the conventional formulation of wave mechanics. If we adopt a new formulation in terms of u alone, using the integral in (11) to replace v, it would cause P to depend on $\partial u/\partial t$ *over the whole space*. Do we thereby introduce absurdities, since P at $\mathbf{x}$ can now be changed by varying $\partial u/\partial t$ at other points in space?

(2) There is also the matter of gauge invariance. This refers to the invariance of the theory to the transformation

$$\psi = \psi' e^{i\theta}. \tag{12}$$

Let $\psi' = u' + iv'$. We then have from (12)

$$u + iv = (u' + iv')(\cos\theta + i\sin\theta)$$

$$\Rightarrow \begin{cases} u = u'\cos\theta - v'\sin\theta, \\ v = v'\cos\theta + u'\sin\theta. \end{cases}$$

Again, applying (11) to these, we get

$$\begin{aligned} u &= u'\cos\theta - K^2\sin\theta \int G(\mathbf{x},\mathbf{x}',t)\,\frac{\partial u'}{\partial t}\,d^3x', \\ v &= u'\sin\theta + K^2\cos\theta \int G(\mathbf{x},\mathbf{x}',t)\,\frac{\partial u'}{\partial t}\,d^3x', \end{aligned} \tag{13}$$

where the integral replaces v' on the right side of the equation. Since we could replace the system *arbitrarily* with another system, how can we expect (13) to remain valid? In other words, it appears that a *specific* factor has been injected into the relation that is supposed to be *general.*

A consistent formulation of wave mechanics in terms of the real wave function requires a rational explanation of (1) and (2). The author plans to fully address these questions soon in a sequel to this paper.

[1]The term first appeared in the abstract of the fourth of Schrödinger's basic papers on wave mechanics, in Ann. Phys. (Leipzig) **181**, 109 (1926). For an English translation see E. Schrödinger, *Collected Papers on Wave Mechanics* (Blackie, London, 1928), p. 102. There are also later uses of this term in the same paper.

[2]*Letters on Wave Mechanics*, edited by K. Przibram (Philosophical Library, New York, 1976), letter to Lorentz from Schrödinger, June 6, 1926, especially pp. 56–58.

[3]See Ref. 2, letter to Planck, June 11, 1926, p. 15.

[4]D. Bohm, *Quantum Theory* (Prentice-Hall, Englewood Cliffs, NJ, 1951), pp. 84–88.

[5]R. Eisburg and R. Resnick, *Quantum Physics of Atoms, Molecules, Solids, Nuclei, and Particles* (Wiley, New York, 1974), pp. 140–142.

[6]J. D. McGervey, *Introduction to Modern Physics* (Academic, New York, 1983), 2nd ed., pp. 138–141.

[7]L. I. Schiff, *Quantum Mechanics* (McGraw-Hill, New York, 1968), 3rd ed., pp. 20–22.

[8]R. L. W. Chen, Am. J. Phys. **50**, 902 (1982).

[9]R. L. W. Chen, Am. J. Phys. **51**, 570 (1983).

[10]R. L. W. Chen and M. B. Rhode, Am. J. Phys. **52**, 988 (1984).

[11]R. L. W. Chen, "Lagrangian formulation of Schrödinger equation," presented at a meeting of the American Physical Society, April 1985 (unpublished).

[12]R. Courant and D. Hilbert, *Methods of Mathematical Physics*, Vol. II (Interscience, New York, 1962), pp. 58–61.

[13]J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1975), 2nd ed., p. 120, Eq. 3.164.

[14]See Ref. 13, p. 17, Eq. 1.17.

De Sitter kinks

K. A. Dunn[a)]
Department of Mathematics, University of British Columbia, Vancouver, British Columbia V6T 1Y4, Canada

J. G. Williams
Department of Mathematics and Computer Science, Brandon University, Brandon, Manitoba R7A 6A9, Canada

(Received 24 June 1988; accepted for publication 24 August 1988)

In this paper a kink solution to the vacuum Einstein equations with cosmological constant $\Lambda > 0$ is discussed. The solution is well-defined on the whole space-time manifold. An appropriate coordinate transformation shows the solution to be locally the same as the de Sitter solution. Globally, the two are distinct, since the light-cone field for the kink solution is homotopically nontrivial.

I. INTRODUCTION

The general relativistic kinks of Finkelstein and Misner[1] arise when categorizing the cross sections of the Lorentz metric tensor bundle according to their homotopy class. For a topologically trivial space-time, $\mathbb{R}^4$ (or $\mathbb{R}^1 \times S^3$), the group of homotopy classes is just the group of integers, $\mathbb{Z}$. Even for a topologically more complex space-time, this same integer counting number $n \in \mathbb{Z}$ will still be present (as well, perhaps, as other topological indices). Metrics belonging to the nth homotopy class are said to have n kinks.

The simplest kind of kink metric can be written

$$g_{\mu\nu} = \delta_{\mu\nu} - 2\phi_\mu \phi_\nu, \quad \mu,\nu = 0,1,2,3,$$

where $\delta_{\mu\nu}$ denotes the Kronecker delta, $\|\delta_{\mu\nu}\| = \mathrm{diag}(1,1,1,1)$, and where the $\{\phi_\mu\}$ specify, at any instant of time, a mapping from $\mathbb{R}^3$ (or S^3 or possibly some more complicated three-manifold) into a three-sphere, S^3. This particular form of $g_{\mu\nu}$ is a special case of the well-known[2] representation of a Lorentz metric in terms of a Riemannian metric and a timelike vector field. The number of kinks present in $g_{\mu\nu}$ is equal to the degree of the mapping represented by $\{\phi_\mu\}$, and a suitable (and spherically symmetric) choice for $\{\phi_\mu\}$ is provided by the Skyrme hedgehog[3]:

$$\phi_0 = \cos\alpha(r),$$

$$\phi_i = (x^i/r)\sin\alpha(r), \quad i = 1,2,3,$$

with $r = (\Sigma x^i x^i)^{1/2}$. For a one-kink solution, α must change by π as r varies over its complete range.

Harriott and Williams[4] have recently presented the solution

$$\sin\alpha = r/A, \quad 0 \leqslant r \leqslant A,$$

as a perfect fluid solution for an object of mass A, the four-velocity being equated to ϕ_μ since $g^{\mu\nu}\phi_\mu\phi_\nu = -1$. (They connect this solution to an exterior vacuum solution in which $A \leqslant r < \infty$ so that α can vary from π to 0 as r varies from 0 to ∞, thereby allowing a complete kink to be present.) The equation of state is $\rho = -p$, with the constant positive energy density related to the scalar curvature by $\rho = R/32\pi$, which is suggestive of the de Sitter solution.[5] It is the purpose of this present paper to show that the $\sin\alpha = r/A$ kink metric is transformable into the de Sitter solution, although only by means of a singular transformation. The $\sin\alpha = r/A$ solution is made physically more plausible by introducing the cosmological constant and changing the topology of the background manifold so that the metric contains a complete kink without the need for attaching an external solution.

II. TRANSFORMATION TO DE SITTER FORM

Making the hedgehog substitution for the $\{\phi_\mu\}$, the expression for the kink metric becomes

$$ds^2 = -\cos 2\alpha\, dt^2 - 2\sin 2\alpha\, dt\, dr + \cos 2\alpha\, dr^2 + r^2\, d\Omega^2,$$

with $d\Omega^2 = d\theta^2 + \sin^2\theta\, d\phi^2$, where θ and ϕ are the usual spherical polar angles. For $\sin\alpha = r/A$,

$$ds^2 = -\left(1 - \frac{2r^2}{A^2}\right)dt^2 - 4\left(\frac{r}{A}\right)\left(1 - \frac{r^2}{A^2}\right)^{1/2} dt\, dr + \left(1 - \frac{2r^2}{A^2}\right)dr^2 + r^2\, d\Omega^2.$$

A first step in transforming this metric is to remove the $dt\, dr$ cross term by introducing a new time coordinate, $\bar{t} = t + f(r)$. Since $\bar{g}_{tr} = g_{tr} - g_{tt}\, df/dr$, the function $f(r)$ must be chosen so that

$$\frac{df}{dr} = \tan 2\alpha = 2\left(\frac{r}{A}\right)\left(1 - \frac{r^2}{A^2}\right)^{1/2}\left(1 - \frac{2r^2}{A^2}\right)^{-1} = 2^{1/2}\left(\frac{r}{a}\right)\left(1 - \frac{r^2}{2a^2}\right)^{1/2}\left(1 - \frac{r^2}{a^2}\right)^{-1},$$

where, for convenience, we have introduced the constant a such that $A = 2^{1/2}a$. The metric can now be written

$$ds^2 = -\left(1 - \frac{r^2}{a^2}\right)d\bar{t}^2 + \left(1 - \frac{r^2}{a^2}\right)^{-1} dr^2 + r^2\, d\Omega^2,$$

which is a well-known,[6] though not terribly convenient, form of the de Sitter metric. The apparent singularity in the metric at $r = a$ occurs because of a bad choice of coordinates and can be removed by the transformation

[a)] Permanent address: Department of Mathematics, Statistics and Computing Science, Dalhousie University, Halifax, Nova Scotia B3H 4H8, Canada.

 0022-2488/89/010087-03$02.50

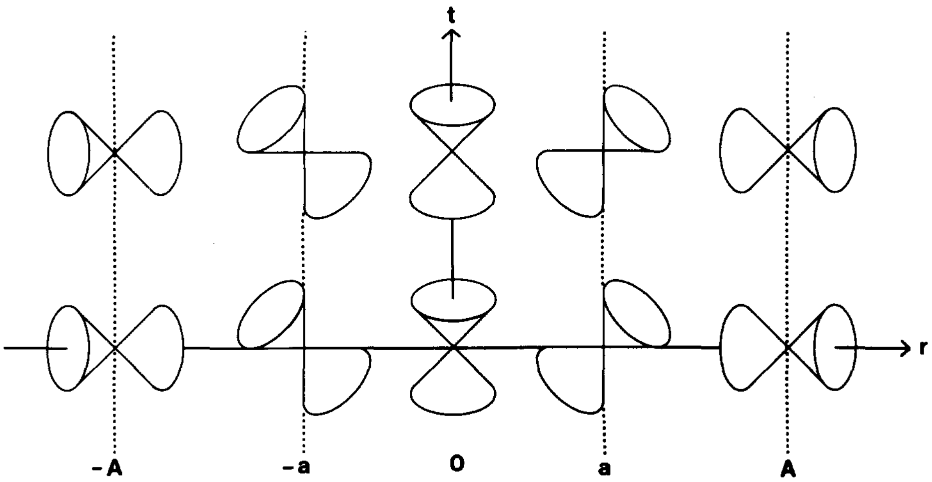

FIG. 1. Light cones for the kinked universe.

$r = a \sin\chi \cosh(a^{-1}T)$,

$\sinh(a^{-1}\bar{t}) = \pm \sinh(a^{-1}T)$
$\times\{1 - \sin^2\chi \cosh^2(a^{-1}T)\}^{-1/2}$.

The latter can also be written

$\cosh(a^{-1}\bar{t}) = \pm \cos\chi \cosh(a^{-1}T)$
$\times\{1 - \sin^2\chi \cosh^2(a^{-1}T)\}^{-1/2}$,

and it is easy to check that

$$dr = \sin\chi \sinh(a^{-1}T)dT + a\cos\chi \cosh(a^{-1}T)d\chi,$$

$$d\bar{t} = \{\cos\chi\, dT + a \sin\chi \sinh(a^{-1}T)\cosh(a^{-1}T)d\chi\} \times\{1 - \sin^2\chi \cosh^2(a^{-1}T)\}^{-1}.$$

In terms of the variables T,χ, the metric takes the usual de Sitter form[5]:

$$ds^2 = -dT^2 + a^2\cosh^2(a^{-1}T)\{d\chi^2 + \sin^2\chi\, d\Omega^2\}.$$

The coordinates T,χ,θ,ϕ cover the whole of space-time, $-\infty < t < \infty$, $0 \leqslant \chi \leqslant \pi$, $0 \leqslant \theta \leqslant \pi$, $0 \leqslant \phi \leqslant 2\pi$, without any singularities (except for the trivial ones associated with the spherical polar angles). The same is, of course, true for our original t,r,θ,ϕ coordinates.

III. DISCUSSION

Return now to the kink metric for $\sin\alpha = r/A$. Rather than assuming this to describe a fluid, it is perhaps preferable (as is usually done for the de Sitter universe) to regard it as a solution for empty space with a cosmological constant, $\Lambda = \frac{1}{4}R$, whence $a = (3/\Lambda)^{1/2}$. As r varies from 0 to A, $\alpha(r)$ varies from π to $\pi/2$. If one wishes a complete kink to be present, α must change by π, in which case one should no longer regard r as a radial coordinate but allow it to be negative as well as positive: $-A \leqslant r \leqslant A$. This requires a change in the topology of the space-time manifold, a point that has already been stressed by Finkelstein and McCollum.[7]

The light-cone behavior can be determined from the equation

$$g_{\mu\nu}\frac{dx^\mu}{ds}\frac{dx^\nu}{ds} = 0$$

by writing $x^t = t - t_0$, $x^r = r - r_0$. At the origin,

$$-(t - t_0)^2 + (r - r_0)^2 = 0,$$

so that $(r - r_0) = \pm (t - t_0)$ and the axis of the light-cone points upwards in the t direction ($\alpha = \pi$). At a distance a from the origin, $\cos 2\alpha = 0$ ($\alpha = 3\pi/4$), and the equation

$$(r - r_0)(t - t_0) = 0$$

shows that the directions of both the t axis and the r axis lie along the cone. Finally, at a distance A from the origin, one can check that the cone is tipped over with its axis pointing in the positive r direction ($\alpha = \pi/2$). The complete configuration is shown in Fig. 1. On the other hand, the light cones of the de Sitter metric are determined from

$$-(T - T_0)^2 + a^2\cosh^2(a^{-1}T)(\chi - \chi_0)^2 = 0,$$

so that the lines that specify the cone in the $T\chi$ plane have slope $\pm a\cosh(a^{-1}T)$. The cones do not tip as χ varies, but, as shown in Fig. 2, they narrow as T increases (or decreases) from $T = 0$.

Comparison of Fig. 1 and Fig. 2 shows the kink metric and the de Sitter metric to be topologically inequivalent. Although the Jacobian determinant is well defined and equals $a\cosh(a^{-1}T)$, the Jacobian *matrix* is clearly singular at $r = a$, or $\sin\chi\cosh(a^{-1}T) = 1$ (which, it is interesting to note, is the same point at which trouble occurred for the original version of the de Sitter metric). Since, by the very nature of a kink metric, α must vary to include some point(s) where $\cos 2\alpha = 0$, the singularity in the transfor-

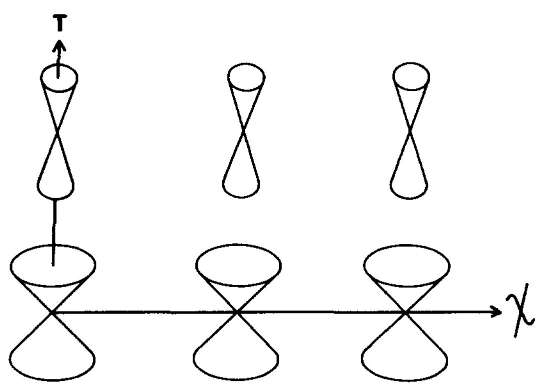

FIG. 2. Light cones for the de Sitter universe.

mation is unavoidable. If two metrics are transformable into each other only by a singular transformation, one should, according to Rosen,[8] regard them as describing different physical situations. It would be interesting to examine well-known metrics, other than the de Sitter one, to see if there exist similar singular changes in coordinates that will transform them into kink form.

ACKNOWLEDGMENTS

The authors would like to thank Dr. H. Goenner for his interesting comments on some earlier work.

The support of the Natural Sciences and Engineering Research Council of Canada through Grant No. A8732 (for KAD) and Grant No. OGP0009151 (for JGW) is gratefully acknowledged.

[1]D. Finkelstein and C. W. Misner, Ann. Phys. (NY) **6**, 230 (1959).
[2]C. T. J. Dodson, *Categories, Bundles and Spacetime Topology* (Shiva, Orpington, Kent, UK, 1980), pp. 162 and 190.
[3]T. H. R. Skyrme, Proc. R. Soc. London Ser. A **260**, 127 (1961).
[4]T. A. Harriott and J. G. Williams, J. Math. Phys. **29**, 179 (1988); "Homotopically nontrivial metric for a perfect fluid," in *Proceedings of the 2nd Canadian Conference on General Relativity and Relativistic Astrophysics*, University of Toronto, 14–16 May 1987, edited by A. Coley, C. Dyer, and T. Tupper (World Scientific, Singapore, 1988).
[5]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-time* (Cambridge U.P., London, 1973), pp. 124–131.
[6]E. Schrödinger, *Expanding Universes* (Cambridge U.P., London, 1957), Chap. 1.
[7]D. Finkelstein and G. McCollum, J. Math. Phys. **16**, 2250 (1975).
[8]N. Rosen, Found. Phys. **15**, 517 (1985).

The high-frequency limit in general relativity

Gregory A. Burnett
Enrico Fermi Institute and Department of Physics, University of Chicago, 5640 S. Ellis Avenue, Chicago, Illinois 60637

(Received 19 July 1988; accepted for publication 24 August 1988)

A precise characterization of the high-frequency limit in general relativity (vacuum space-times) is presented. The averaging schemes used in earlier works are avoided by the use of weak limits, giving the characterization the advantage that the high-frequency limit can be rigorously analyzed. Using this characterization, a theorem to the effect that a certain "effective stress-energy" tensor acts as a source of curvature of the smooth background space-time is proved. It is shown in the special case corresponding to the presence of a single wave that this tensor has the same form as the stress-energy tensor of a null fluid. Also, the extension of this characterization of the high-frequency limit to other fields, such as electromagnetic fields, and to nonvacuum space-times is briefly discussed.

I. INTRODUCTION

One of the basic features of general relativity is that space-time is dynamic: a star collapsing, two stars revolving about one another, or two black holes coalescing should all generate gravitational radiation. How would one describe quantitatively such radiating systems? In the case where the radiation is "weak," we may idealize the situation by taking the limit in which the amplitude of the radiation vanishes, thus turning the problem into one of solving the linearized Einstein equation on a fixed background space-time.[1] But, in the case in which the radiation is not weak, solving such a problem is generally very difficult. However, when the gravitational waves have small amplitudes and short wavelengths (high frequencies), a simplification similar to the weak case occurs. It is this limiting case that we shall characterize and study here.

As an example of the type of behavior with which we shall be concerned, consider the one-parameter family of metrics given, for parameter $\lambda > 0$, by

$$g_{ab}(\lambda) = B_\lambda^2(u)(e^{w_\lambda(u)}\nabla_a x\nabla_b x + e^{-w_\lambda(u)}\nabla_a y\nabla_b y) - \nabla_{(a}u\nabla_{b)}v\,, \qquad (1)$$

where $w_\lambda(u) = \lambda\alpha(u)\sin(u/\lambda)$ and $B_\lambda(u)$ satisfies $B''_\lambda(u) + (w'_\lambda(u))^2 B_\lambda(u) = 0$, with $B_\lambda(0) = 1$, $B'_\lambda(0) = 0$ (prime denoting differentiation with respect to u.) Here x, y, u, and v are smooth scalar fields and $\alpha(u)$ is a smooth function of its argument with compact support. These are all exact plane-wave solutions of the vacuum Einstein equation. We now consider the behavior of this family of metrics as $\lambda \to 0$. We see that $w_\lambda(u) = O(\lambda)$. It can also be shown that $B_\lambda(u) = B_0(u) + O(\lambda^2)$, where $B_0(u)$ is the solution of the differential equation

$$B_0''(u) + \tfrac{1}{2}\alpha^2(u)B_0(u) = 0\,, \qquad (2)$$

with $B_0(0) = 1$, $B_0'(0) = 0$. So, we have

$$g_{ab}(\lambda) = g_{ab}(0) + \lambda B_0^2(u)\alpha(u)\sin(u/\lambda) \times(\nabla_a x\nabla_b x - \nabla_a y\nabla_b y) + O(\lambda^2)\,, \qquad (3)$$

where

$$g_{ab}(0) = B_0^2(u)(\nabla_a x\nabla_b x + \nabla_a y\nabla_b y) - \nabla_{(a}u\nabla_{b)}v\,. \qquad (4)$$

We see that $g_{ab}(0)$ can be thought of as the smooth background part of $g_{ab}(\lambda)$ since $g_{ab}(\lambda) \to g_{ab}(0)$ uniformly as $\lambda \to 0$, while $g_{ab}(\lambda) - g_{ab}(0)$ can be thought of as the wave part since as $\lambda \to 0$ its amplitude and wavelength both approach zero. Notice that the derivative of the wave part does not vanish as $\lambda \to 0$. We might thus expect that these waves possess an "effective stress energy" in the high-frequency limit. Indeed, we have

$$G_{ab}[g(0)] = \alpha^2(u)\nabla_a u\nabla_b u\,, \qquad (5)$$

so an effective stress energy, $T_{ab} = \alpha^2(u)\nabla_a u\nabla_b u$, can be associated with the gravitational waves in the high-frequency limit which serves as the source for the background metric $g_{ab}(0)$.

One systematic treatment of the high-frequency limit in general relativity is that given by Isaacson.[2,3] Fix a manifold M, and consider a one-parameter family of metrics $g_{ab}(\lambda)$ thereon, of the form $g_{ab}(\lambda) = g_{ab}(0) + \lambda h_{ab}(\lambda)$. Let, for $\lambda > 0$, these metrics satisfy the vacuum Einstein equation. Here, $g_{ab}(0)$ is a smooth (background) metric, and $h_{ab}(\lambda)$ is a one-parameter family of symmetric tensor fields satisfying the conditions $h_{ab}(\lambda) = O(\lambda^0)$, $\nabla_m h_{ab}(\lambda) = O(\lambda^{-1})$, and $\nabla_m\nabla_n h_{ab}(\lambda) = O(\lambda^{-2})$, where ∇_m is the derivative operator associated with $g_{ab}(0)$. To examine the behavior of these metrics as $\lambda \to 0$, expand the Ricci tensor[4] of $g_{ab}(\lambda)$ in orders of λ, to find

$$R_{ab}[g(\lambda)] = R_{ab}{}^{(0)} + \lambda R_{ab}{}^{(1)}(\lambda) + \lambda^2 R_{ab}{}^{(2)}(\lambda) + O(\lambda)\,, \qquad (6)$$

where $R_{ab}{}^{(0)}$ is $R_{ab}[g(0)]$, $R_{ab}{}^{(1)}$ is a certain linear combination of contractions of the expression $g^{cd}(0)\nabla_m\nabla_n h_{ab}(\lambda)$, and $R_{ab}{}^{(2)}$ are certain linear combinations of contractions of the expressions $g^{ij}(0)g^{kl}(0)\nabla_m\nabla_n h_{ab}(\lambda)$ and $g^{ij}(0)g^{kl}(0)\nabla_m h_{ab}(\lambda)\nabla_n h_{cd}(1)$. Equating terms of each order in λ in Eq. (6) [noting that $R_{ab}{}^{(0)} = O(1)$, $R_{ab}{}^{(1)} = O(\lambda^{-2})$, and $R_{ab}{}^{(2)} = O(\lambda^{-2})$], we obtain

$$R_{ab}{}^{(1)}(\lambda) = 0\,, \qquad (7)$$

$$R_{ab}[g(0)] = -\lambda^2 R_{ab}{}^{(2)}(\lambda)\,. \qquad (8)$$

 0022-2488/89/010090-07$02.50

Finally, because we wish to ignore the high-frequency part of $R_{ab}{}^{(2)}$, we now average the right-hand side of Eq. (8), giving

$$R_{ab}[g(0)] = -\lambda^2 \langle R_{ab}{}^{(2)}(\lambda)\rangle, \tag{9}$$

where $\langle \cdots \rangle$ denotes a suitable average over some small region about each point.

This is Isaacson's analysis of the high-frequency limit. The resulting equations make good physical sense: The waves obey the linearized vacuum Einstein equation on the smooth background [Eq. (7)]; and the waves act as a source of curvature of the smooth background [Eq. (9)]. But, unfortunately the derivation of these equations is not completely convincing. For instance, in what sense is it proper to equate terms of each order in λ in Eq. (6)? Also, the averaging of the right-hand side of Eq. (8) is not clearly defined. What size and shape of a region are we to average over?

We here propose a simple and precise characterization of the high-frequency limit in general relativity. We shall use this characterization to prove a theorem that captures the essential content of Eq. (9). We briefly discuss the issue of gauge in this scheme, and show in the case where there is a single wave that the effective stress-energy tensor of the waves is that of a null fluid. We then remark on the extension of this characterization to other fields and to nonvacuum space-times.

II. CHARACTERIZATION OF THE HIGH-FREQUENCY LIMIT IN GENERAL RELATIVITY

Fix a manifold M, a smooth derivative operator ∇_m, and a one-parameter family of metrics $g_{ab}(\lambda)$ on M. We impose, on this system, four conditions which are intended to reflect the behavior of these metrics in the high-frequency limit in general relativity:

(i) $G_{ab}[g(\lambda)] = 0$, for all $\lambda > 0$,
(ii) $g_{ab}(\lambda)$ converges to $g_{ab}(0)$ uniformly as $\lambda \to 0$,
(iii) $\nabla_m(g_{ab}(\lambda) - g_{ab}(0))$ is uniformly bounded,
(iv) $\nabla_m(g_{ab}(\lambda) - g_{ab}(0))\nabla_n(g_{cd}(\lambda) - g_{cd}(0))$ converges weakly to some tensor field μ_{mnabcd}, as $\lambda \to 0$.

The first condition means that each metric $g_{ab}(\lambda)$, for $\lambda > 0$, is an exact solution of the vacuum Einstein equation. Notice we do not require that $G_{ab}[g(0)] = 0$. In fact, we will find that, in general, $G_{ab}[g(0)]$ is nonzero, as was the case in the example given above.

The second condition means that for every smooth tensor field t^{ab} and every smooth scalar field $\epsilon > 0$ there exists a smooth scalar field λ_0 such that

$$|(g_{ab}(\lambda) - g_{ab}(0))t^{ab}| < \epsilon, \tag{10}$$

for every smooth scalar field λ with $0 < \lambda < \lambda_0$. Physically, this condition requires that the amplitude of the wave part vanish in the high-frequency limit, so that $g_{ab}(0)$ is the smooth background metric.

The third condition means that for every smooth tensor field t^{mab}, there exist smooth scalar fields M and λ_0 such that

$$|\nabla_m(g_{ab}(\lambda) - g_{ab}(0))t^{mab}| < M, \tag{11}$$

for every smooth scalar field λ with $0 < \lambda < \lambda_0$. We interpret this condition physically as requiring that the amplitude of the waves approaches zero at least as fast as the wavelength. This condition not only captures our intuitive notion of the high-frequency limit, it will prove essential in the analysis of this limit.

Finally, the fourth condition means that there exists a smooth tensor field μ_{mnabcd}, such that for every test field t^{mnabcd} (smooth tensor density of weight $+1$ with compact support) we have

$$\lim_{\lambda \to 0} \int (\nabla_m(g_{ab}(\lambda) - g_{ab}(0))\nabla_n(g_{cd}(\lambda) - g_{cd}(0)) - \mu_{mnabcd})t^{mnabcd} = 0. \tag{12}$$

The sole role of this condition is to guarantee the existence of an effective stress-energy tensor of the gravitational waves.

Notice how these four conditions on the family of metrics considered compare to those used in Isaacson's analysis. Isaacson considers a family of metrics of the form $g_{ab}(\lambda) = g_{ab}(0) + \lambda h_{ab}(\lambda)$. Isaacson's requirement that $g_{ab}(\lambda)$ be a solution of Einstein's equation is exactly condition (i); that $\lambda h_{ab}(\lambda) = O(\lambda)$ is captured in condition (ii); and that $\lambda \nabla_m h_{ab}(\lambda) = O(\lambda^0)$ is captured in condition (iii). So, we see that the family of metrics considered by Isaacson is a special case of those allowed by conditions (i)–(iii). We then replace Isaacson's averaging, as in Eq. (9), by the use of weak limits, as in condition (iv).

Now consider again the one-parameter family of metrics given in Eq. (1). This family, we claim, satisfies our four conditions. Indeed, condition (i) is satisfied since each metric is an exact solution of the vacuum Einstein equation. Condition (ii) is satisfied, with $g_{ab}(0)$ given in Eq. (4). Condition (iii) is satisfied,[5] since $\cos(u/\lambda)$ is uniformly bounded. Finally, condition (iv) is satisfied,[5] with

$$\mu_{mnabcd} = \tfrac{1}{2}\alpha^2(u)B_0^2(u)(\nabla_m u)(\nabla_n u) \times (\nabla_a x \nabla_b x - \nabla_a y \nabla_b y)(\nabla_c x \nabla_d x - \nabla_c y \nabla_d y), \tag{13}$$

which follows from the fact that $\cos^2(u/\lambda) \to \tfrac{1}{2}$ weakly as $\lambda \to 0$.

To aid in the investigation of the consequences of the above conditions, we introduce a few facts about weak convergence, uniform convergence, and uniform boundedness (for proofs, see the Appendix.) Let $\alpha(\lambda)$ and $\beta(\lambda)$ be one-parameter families of smooth tensor fields (indices suppressed.) Then,

(a) $\alpha(\lambda) \to \alpha(0)$ weakly, if $\alpha(\lambda) \to \alpha(0)$ uniformly,
(b) $\nabla_m \alpha(\lambda) \to \nabla_m \alpha(\lambda)$ weakly, if $\alpha(\lambda) \to \alpha(0)$ weakly,
(c) $\alpha(\lambda)\beta(\lambda) \to 0$ uniformly, if $\alpha(\lambda) \to 0$ uniformly, and $\beta(\lambda)$ is uniformly bounded,
(d) $\alpha(\lambda)\beta(\lambda) \to \alpha(0)\beta(0)$ weakly, if $\alpha(\lambda) \to \alpha(0)$ uniformly, $\beta(\lambda) \to \beta(0)$ weakly, and $\beta(\lambda)$ is uniformly bounded.

With these facts, let us investigate some of the elementary consequences of conditions (ii)–(iv) that will be useful in what follows. Let $g_{ab}(\lambda)$ be a one-parameter family of metrics satisfying conditions (ii) and (iii), let ∇_a be the derivative operator compatible with $g_{ab}(0)$, and set

$$C^c_{ab}(\lambda) = g^{cm}(\lambda)[\nabla_{(a}g_{b)m}(\lambda) - \tfrac{1}{2}\nabla_m g_{ab}(\lambda)]. \tag{14}$$

Since $g_{ab}(\lambda) \to g_{ab}(0)$ uniformly as $\lambda \to 0$, we see, using (a), that $g_{ab}(\lambda)\to g_{ab}(0)$ weakly as $\lambda\to 0$. Hence, we see, using (b), that $\nabla_m g_{ab}(\lambda)\to 0$ weakly as $\lambda\to 0$. Next, using the fact that $\nabla_m g_{ab}(\lambda)$ is uniformly bounded we then see, using (d), that $g^{cd}(\lambda)\nabla_m g_{ab}(\lambda)\to 0$ weakly as $\lambda\to 0$. With this, we conclude that $C^c{}_{ab}(\lambda)\to 0$ weakly as $\lambda\to 0$ and that $C^c{}_{ab}(\lambda)$ is uniformly bounded. Using (b), we also see that $\nabla_m C^c{}_{ab}(\lambda)\to 0$ weakly as $\lambda\to 0$. Does $(g_{de}(\lambda)-g_{de}(0))\nabla_m C^c{}_{ab}(\lambda)\to 0$ weakly as $\lambda\to 0$? Not in general! Although $g_{de}(\lambda)-g_{de}(0)\to 0$ uniformly and $\nabla_m C^c{}_{ab}(\lambda)\to 0$ weakly, $\nabla_m C^c{}_{ab}(\lambda)$ need not be uniformly bounded. In fact, if $g_{ab}(\lambda)$ also satisfies condition (iv), then $(g_{de}(\lambda)-g_{de}(0))\nabla_m C^c{}_{ab}(\lambda)$ converges weakly to some expression in μ_{mnabcd}. We can see this by writing

$$\begin{aligned}(g_{de}(\lambda)-g_{de}(0))\nabla_m C^c{}_{ab}(\lambda)\\ &=\nabla_m[(g_{de}(\lambda)-g_{de}(0))C^c{}_{ab}(\lambda)]\\ &\quad - C^c{}_{ab}(\lambda)\nabla_m(g_{de}(\lambda)-g_{de}(0))\end{aligned}$$

and noting that the weak limit of the first term is zero and that the weak limit of the second term is some expression in μ_{mnabcd}.

What derivative operator are we to use in conditions (iii) and (iv)? It turns out that it does not matter. The validity of conditions (iii) and (iv) and the tensor field μ_{mnabcd} we obtain in the latter are independent of which smooth derivative operator we use. To show this, choose any two smooth derivative operators ∇_a and $\tilde{\nabla}_a$. Then these must be related by a smooth tensor field $K^c{}_{ab}$, so that in particular we have

$$\tilde{\nabla}_m(g_{ab}(\lambda)-g_{ab}(0))=\nabla_m(g_{ab}(\lambda)-g_{ab}(0)) - 2K^n{}_{m(a}(g_{b)n}(\lambda)-g_{b)n}(0)).$$

But $K^c{}_{de}(g_{ab}(\lambda)-g_{ab}(0))$ is uniformly bounded. Hence $\tilde{\nabla}_m(g_{ab}(\lambda)-g_{ab}(0))$ is uniformly bounded if and only if $\nabla_m(g_{ab}(\lambda)-g_{ab}(0))$ is uniformly bounded, i.e., condition (iii) is independent of derivative operator. Further, we see that

$$\begin{aligned}&\tilde{\nabla}_m(g_{ab}(\lambda)-g_{ab}(0))\tilde{\nabla}_n(g_{cd}(\lambda)-g_{cd}(0))\\ &\quad-\nabla_m(g_{ab}(\lambda)-g_{ab}(0))\nabla_n(g_{cd}(\lambda)-g_{cd}(0))\to 0\end{aligned}$$

weakly as $\lambda\to 0$ since $\nabla_m(g_{ab}(\lambda)-g_{ab}(0))$ is uniformly bounded and $K^c{}_{de}(g_{ab}(\lambda)-g_{ab}(0))\to 0$ uniformly. Thus the validity of condition (iv) and the tensor field μ_{mnabcd} we obtain are independent of our choice of derivative operator.

In condition (iv), the tensor field μ_{mnabcd} has, we claim, the symmetries $\mu_{mnabcd}=\mu_{(mn)(cd)(ab)}$. The symmetries $\mu_{mnabcd}=\mu_{nm(cd)(ab)}$ are manifest from the definition of μ_{mnabcd}. To show the remaining symmetries, $\mu_{mnabcd}=\mu_{(mn)abcd}$, consider

$$\begin{aligned}\mu_{[mn]abcd}&=\operatorname*{w-lim}_{\lambda\to 0}(\nabla_{[m}h_{|ab|}\nabla_{n]}h_{cd})\\ &=\operatorname*{w-lim}_{\lambda\to 0}(\nabla_{[m}(h_{|ab|}\nabla_{n]}h_{cd})-h_{ab}R_{mn(c}{}^p h_{d)p})\\ &=0,\end{aligned}\tag{15}$$

where $h_{ab}(\lambda)=g_{ab}(\lambda)-g_{ab}(0)$, $R_{abc}{}^d$ is the Riemann tensor associated with ∇_m, and w-$\lim_{\lambda\to 0}$ means the weak limit as $\lambda\to 0$. The first step follows from condition (iv). In the second step, we used the Leibnitz rule and replaced $\nabla_{[m}\nabla_{n]}h_{cd}$ by $R_{mn(c}{}^p h_{d)p}$. In the last step, the first term converges to zero weakly since $h_{ab}\to 0$ uniformly and $\nabla_n h_{cd}\to 0$ is uniformly bounded, and the second term converges to zero weakly since $h_{ab}\to 0$ uniformly. Hence $\mu_{mnabcd}=\mu_{(mn)(cd)(ab)}$.

So, our analysis of the high-frequency limit is based on a one-parameter family of metrics $g_{ab}(\lambda)$ satisfying conditions (i)–(iv). In the high-frequency limit, the information about the waves is contained in the fields $g_{ab}(0)$ and μ_{mnabcd}. In the next section, we show that there can be constructed from μ_{mnabcd} an effective stress-energy tensor that serves as a source for the background metric $g_{ab}(0)$.

III. THE EFFECT OF GRAVITATIONAL WAVES ON A BACKGROUND SPACE-TIME

We now introduce a theorem that shows the manner in which gravitational waves in the high-frequency limit act as a source of curvature for the background space-time.

Theorem I: Let $g_{ab}(\lambda)$ be a one-parameter family of metrics satisfying conditions (i)–(iv) above. Then, $G_{ab}[g(0)]=\mu_{[b|[am]|n]}{}^{mn}$ and $R[g(0)]=0$.

Proof: First, note that the Ricci tensors $R_{ab}[g(0)]$ and $R_{ab}[g(\lambda)]$ are related by

$$R_{ab}[g(0)]=R_{ab}[g(\lambda)]+2\nabla_{[a}C^m{}_{m]b}-2C^n{}_{b[a}C^m{}_{m]n},\tag{16}$$

where $C^c{}_{ab}$ is given by Eq. (14) and ∇_a is the derivative operator associated with $g_{ab}(0)$. Now, consider

$$\begin{aligned}R_{ab}[g(0)]\\ &=\operatorname*{w-lim}_{\lambda\to 0}(R_{ab}[g(\lambda)]+2\nabla_{[a}C^m{}_{m]b}-2C^n{}_{b[a}C^m{}_{m]n})\\ &=\tfrac{1}{4}(2\mu_{mn}{}^n{}_{ab}{}^m-2\mu_m{}^m{}_{nab}{}^n+\mu_{abmn}{}^{nm}\\ &\quad-2\mu_{m(ab)}{}^m{}_n{}^n+\mu_m{}^m{}_n{}^n{}_{ab}),\end{aligned}\tag{17}$$

where, in the last step, indices are raised and lowered with $g^{ab}(0)$ and $g_{ab}(0)$. The first step results by taking the weak limit of Eq. (16). The second step results as follows. The weak limit of the first term is zero by condition (i). The weak limit of the second term is also zero since $C^c{}_{ab}\to 0$ weakly, as we showed in the previous section. The weak limit of the third term is a certain linear combination of contractions of μ_{mnabcd}, namely, the expression given above, which follows from the fact that $C^n{}_{b[a}C^m{}_{m]n}$ is a certain linear combination of contractions of $g^{ij}(\lambda)g^{kl}(\lambda)\nabla_m g_{ab}(\lambda)\nabla_n g_{cd}(\lambda)$ and the use of conditions (ii), (iii), and (iv). This establishes Eq. (17).

This expression for $R_{ab}[g(0)]$ is now converted into the expression given in the theorem by means of the following. Consider

$$\begin{aligned}\mu_{[b|[am]|n]cd}g^{mn}(0)\\ &=\operatorname*{w-lim}_{\lambda\to 0}(\nabla_{[a}(g_{|cd|}(\lambda)-g_{|cd|}(0))C^m{}_{m]b})\\ &=\operatorname*{w-lim}_{\lambda\to 0}(\nabla_{[|a}[(g_{|cd|}(\lambda)-g_{|cd|}(0))C^m{}_{m]b}]\\ &\quad-(g_{cd}(\lambda)-g_{cd}(0))\nabla_{[a}C^m{}_{m]b})=0.\end{aligned}\tag{18}$$

In the first step, we used condition (iv). The second step is an identity. In the last step, the first term is zero since $g_{cd}(\lambda) - g_{cd}(0) \to 0$ uniformly and $C^c{}_{ab}$ is uniformly bounded, and the second term is also zero since, by virtue of Eq. (16), $\nabla_{[a} C^m{}_{m]b}$ is uniformly bounded. This establishes Eq. (18).

From Eqs. (17) and (18) the theorem now follows directly. First, note that μ_{mnabcd} is completely determined by the combination $\alpha_{abcdef} = \mu_{[c|[ab]|d]ef}$ and $\beta_{abcdef} = \mu_{(abcd)ef}$. Indeed, we have

$$\mu_{mnabcd} = -\tfrac{4}{3}(\alpha_{a(mn)bcd} + \alpha_{c(mn)dab} - \alpha_{c(ab)dmn}) + (\beta_{mnabcd} + \beta_{mncdab} - \beta_{abcdmn}) .$$

Next, note that from their definitions, $\alpha_{abcdef} = \alpha_{[ab][cd](ef)}$, $\alpha_{[abc]def} = 0$, $\alpha_{amc}{}^m{}_{ef} = 0$, and $\beta_{abcdef} = \beta_{(abcd)(ef)}$. Substitute the above expression for μ_{mnabcd} into the expression for $R_{ab}[g(0)]$ given in Eq. (17), to find

$$R_{ab}[g(0)] = \alpha_{ambn}{}^{mn} = \mu_{[b|[am]|n]}{}^{mn} ,$$

as claimed. Further, $R[g(0)] = \alpha_{am}{}^a{}_n{}^{mn} = 0$ since α_{abcdef} is trace-free on its first four indices. This completes the proof of Theorem 1. ◊

Theorem 1 admits a straightforward interpretation. Set $T_{ab} = \mu_{[b|[am]|n]}{}^{mn}$. Theorem 1 suggests that T_{ab} is to be thought of as the effective stress-energy tensor for the gravitational waves, for T_{ab} acts as the source of the curvature for the background space-time $(M, g_{ab}(0))$. Further, the theorem tells us that T_{ab} is trace-free, $T_a{}^a = 0$, and divergence-free, $\nabla_a T^{ab} = 0$.

IV. GAUGE

In condition (iv) of the characterization of the high-frequency limit presented above, we introduce the tensor field μ_{mnabcd}. This field is gauge dependent in the following sense. Fix a smooth manifold M. Regard two space-times (M, g_{ab}) and $(M, \hat{g}_{ab})$, which are related by a diffeomorphism ϕ, as being physically equivalent. So, in particular, any two one-parameter families of space-times $(M, g_{ab}(\lambda))$ and $(M, \hat{g}_{ab}(\lambda))$ that are related by a one-parameter family of diffeomorphisms ϕ_λ such that ϕ_0 is the identity map, are regarded as being physically equivalent. Yet, while $g_{ab}(\lambda)$ gives rise to the tensor field μ_{mnabcd} in condition (iv), $\hat{g}_{ab}(\lambda)$ gives rise to a tensor field $\hat{\mu}_{mnabcd}$ that is, in general, different from μ_{mnabcd}. In this section we find how μ_{mnabcd} and $\hat{\mu}_{mnabcd}$ are related.

Fix on M, a one-parameter family of metrics $g_{ab}(\lambda)$ satisfying conditions (i)–(iv). Consider one-parameter families of diffeomorphisms ϕ_λ: $M \to M$, and their associated maps $[\phi_\lambda^{-1}]^*$ (so, e.g., $([\phi_\lambda^{-1}]^*\xi)_a \in T_p^*$ if $\xi_a \in T^*_{\phi(p)}$) such that ϕ_0 is the identity map and the one-parameter family of metrics $\hat{g}_{ab}(\lambda) = ([\phi_\lambda^{-1}]^* g(\lambda))_{ab}$ also satisfies conditions (i)–(iv). Fix any smooth derivative operator ∇_m. Sufficient conditions that ensure that $\hat{g}_{ab}(\lambda)$ satisfy conditions (i)–(iv) are that for all smooth fields ξ and η_b,

(I) $([\phi_\lambda^{-1}]^*\xi)_a \to \xi_a$ uniformly as $\lambda \to 0$,

(II) $\nabla_m(([\phi_\lambda^{-1}]^*\xi)_a - \xi_a)$ is uniformly bounded,

(III) $\nabla_m(([\phi_\lambda^{-1}]^*\xi)_a - \xi_a)\nabla_n(([\phi_\lambda^{-1}]^*\eta)_c - \eta_c) \to \tau_{mna}{}^b{}_c{}^d \xi_b \eta_d$ weakly as $\lambda \to 0$, for some field $\tau_{mna}{}^b{}_c{}^d$,

(IV) $\nabla_m(([\phi_\lambda^{-1}]^*\xi)_a - \xi_a)\nabla_n(g_{cd}(\lambda) - g_{cd}(0)) \to \sigma_{mna}{}^b{}_{cd}\xi_b$ weakly as $\lambda \to 0$, for some field $\sigma_{mna}{}^b{}_{cd}$.

The reasoning behind this choice is that with the exception of condition (IV), the gauge is independent of which family of metrics $g_{ab}(\lambda)$ we are considering.

Just as conditions (i)–(iv) are independent of which derivative operator we use, so are conditions (I)–(IV). It follows, using the above conditions, that $\sigma_{mna}{}^b{}_{cd}$ and $\tau_{mna}{}^b{}_c{}^d$ have the symmetries $\sigma_{mna}{}^b{}_{cd} = \sigma_{(mna)}{}^b{}_{cd}$, and $\tau_{mna}{}^b{}_c{}^d = \tau_{(mna}{}^{(b}{}_{c)}{}^{d)}$. The symmetry $\sigma_{mna}{}^b{}_{[cd]} = 0$ is clear, while the symmetry $\sigma_{[mn]a}{}^b{}_{cd} = 0$ follows from an argument similar to that used to show $\mu_{[mn]abcd} = 0$. For the symmetry $\sigma_{n[ma]}{}^b{}_{cd} = 0$, on the other hand, first note that for any smooth dual vector field ξ_a we have

$$\begin{aligned} 2\nabla_{[m}(([\phi_\lambda^{-1}]^*\xi)_{a]} - \xi_{a]}) &= (d([\phi_\lambda^{-1}]^*\xi))_{ma} - (d\xi)_{ma} \\ &= ([\phi_\lambda^{-1}]^* d\xi)_{ma} - (d\xi)_{ma} , \end{aligned} \tag{19}$$

where d is the exterior derivative and we have used the fact that $d[\phi_\lambda^{-1}]^* = [\phi_\lambda^{-1}]^* d$. But, $(([\phi_\lambda^{-1}]^* d\xi)_{ma} - (d\xi)_{ma}) \to 0$ uniformly as $\lambda \to 0$ as follows from condition (I). Our symmetry $\sigma_{n[ma]}{}^b{}_{cd} = 0$ now follows. The symmetries of $\tau_{mna}{}^b{}_c{}^d$ follow similarly. In addition, we find that $\sigma_{[b|[a|ef|m]|n]} g^{mn}(0) = 0$ by the use of condition (i) and an argument similar to that used to derive Eq. (18).

The tensor fields $\hat{\mu}_{mnabcd}$ and μ_{mnabcd} are related through $\sigma_{mna}{}^b{}_{cd}$ and $\tau_{mna}{}^b{}_c{}^d$ by

$$\hat{\mu}_{mnabcd} = \mu_{mnabcd} + 2(\sigma_{mn(ab)cd} + \sigma_{mn(cd)ab}) + 4\tau_{mn(ab)(cd)} , \tag{20}$$

where indices are raised and lowered with $g^{ab}(0)$ and $g_{ab}(0)$. This is seen by writing

$$\hat{g}_{ab}(\lambda) - g_{ab}(0) = ([\phi_\lambda^{-1}]^*(g(\lambda) - g(0)))_{ab} + (([\phi_\lambda^{-1}]^* g(0))_{ab} - g_{ab}(0))$$

and using the above conditions. So, we see that the tensor field μ_{mnabcd} we obtain in condition (iv) is sensitive to a one-parameter family of diffeomorphisms, in the manner of Eq. (20). It is in this sense that μ_{mnabcd} is gauge dependent.

What $\sigma_{mna}{}^b{}_{cd}$ and $\tau_{mna}{}^b{}_c{}^d$ come from some one-parameter family of diffeomorphisms ϕ_λ, as in (III) and (IV) above? If we could answer this question we would know what part of μ_{mnabcd} is physical and what part is gauge. Clearly, it is necessary that $\sigma_{mna}{}^b{}_{cd}$ and $\tau_{mna}{}^b{}_c{}^d$ have the symmetries and traces given above. Further, by conditions (III) and (iv), it is necessary that

$$\tau_{mna}{}^b{}_c{}^d t^{ma}{}_b t^{nc}{}_d \geqslant 0 , \tag{21}$$

$$(\mu_{mnabcd} + 2(\sigma_{mn(ab)cd} + \sigma_{mn(cd)ab}) + 4\tau_{mn(ab)(cd)}) t^{mab} t^{ncd} \geqslant 0 , \tag{22}$$

for all smooth tensor fields $t^{ma}{}_b$. Are these conditions sufficient? Possibly for $\tau_{mna}{}^b{}_c{}^d$, but probably not for $\sigma_{mna}{}^b{}_{cd}$. It is plausible that we could obtain any $t_{mna}{}^b{}_c{}^d$ we desired (subject to the aforementioned constraints) by choosing ϕ_λ appropriately. But, in obtaining $\sigma_{mna}{}^b{}_{cd}$, we have no control over $g_{ab}(\lambda)$, so those $\sigma_{mna}{}^b{}_{cd}$ we can obtain are most likely restricted.

We now ask what part of μ_{mnabcd} is invariant under the addition of tensor fields $\sigma_{mna}{}^b{}_{cd}$ and $\tau_{mna}{}^b{}_c{}^d$ as in Eq. (20), where $\sigma_{mna}{}^b{}_{cd} = \sigma_{(mna)}{}^b{}_{(cd)}$, $\tau_{mna}{}^b{}_c{}^d = \tau_{(mna}{}^{(b}{}_{c)}{}^{d)}$, and $\sigma_{[b|[a|ef|m]|n]}g^{mn}(0) = 0$. That is, what part of μ_{mnabcd} is manifestly gauge invariant under the addition of tensors, in the proper manner, which have the symmetries and traces of gauge. To answer this, we note, as we did in the proof of Theorem 1, that μ_{mnabcd} can be decomposed uniquely into two parts: $\alpha_{abcdef} = \mu_{[c|[ab]|d]ef}$; and $\beta_{abcdef} = \mu_{(abcd)ef}$. The tensor field β_{abcdef} can be shown to be entirely gauge, in our algebraic sense, i.e., there is no algebraically gauge invariant part. So, we need only study α_{abcdef}.

It follows from the symmetries and traces of μ_{mnabcd} that α_{abcdef} has the symmetries and traces of the Weyl tensor on its first four indices and is symmetric on its last two. Under a gauge transformation α_{abcdef} changes as follows:

$$\hat{\alpha}_{abcdef} = \alpha_{abcdef} + 2\sigma_{[c|[a|(ef)|b]|d]}. \tag{23}$$

Define $T_{ab} = \alpha_{ambn}{}^{mn}$ and $S_{ab} = {}^*\alpha_{ambn}{}^{mn}$ (${}^*\alpha_{abcdef} = \frac{1}{2}\epsilon_{ab}{}^{mn}\alpha_{mncdef}$). It can be shown that if $\hat{\alpha}_{abcdef}$ and α_{abcdef} give rise to the same T_{ab} and S_{ab}, then $\hat{\alpha}_{abcdef}$ and α_{abcdef} are related as in Eq. (23) for some $\sigma_{mna}{}^b{}_{cd}$ satisfying the symmetries and traces given above. That is, T_{ab} and S_{ab} are the only gauge invariant parts of μ_{mnabcd}.

The tensor field T_{ab} is just the effective stress energy associated with the high-frequency gravitational waves—that this field is gauge invariant follows already from Theorem 1. But, what is the significance of S_{ab}? In the example given in the Introduction, $S_{ab} = 0$. Is this true in general? No argument has been found to show that S_{ab} always vanishes. Could S_{ab} be gauge invariant only for gauges satisfying conditions (I)–(IV)? If we consider the high-frequency limit of a spin-2 field in Minkowski space-time and construct the quantity in this case that is analogous to S_{ab}, one then finds that this quantity is gauge invariant in the broadest sense. So, it appears likely that S_{ab} is a gauge invariant quantity. But, its physical significance, assuming it does not always vanish, is not understood.

V. A SPECIAL CASE

On our fixed manifold M, consider a one-parameter family of metrics $g_{ab}(\lambda)$ satisfying conditions (i)–(iv). But now require that this family be such that in some region,

$$\mu_{mnabcd} = k_m k_n \gamma_{ab}\gamma_{cd} + (\text{gauge pieces}), \tag{24}$$

where k_a and $\gamma_{ab} = \gamma_{(ab)}$ are fields on M, and by (gauge pieces) we mean the addition of tensor fields $\sigma_{mna}{}^b{}_{cd}$ and $\tau_{mna}{}^b{}_c{}^d$ as in Eq. (20), where $\sigma_{mna}{}^b{}_{cd}$ and $\tau_{mna}{}^b{}_c{}^d$ have the symmetries and traces of gauge. The family of metrics given in the Introduction satisfies this property, for μ_{mnabcd} is given in this case by Eq. (13). We interpret this requirement physically as meaning that there is present a single gravitational wave. Equation (18) now demands

$$k^m k_m \gamma_{ab} = 2k^m\gamma_{m(a}k_{b)} - k_a k_b \gamma_m{}^m. \tag{25}$$

When k_a is spacelike or timelike, this condition is equivalent to $\gamma_{ab} = k_{(a}\xi_{b)}$ for some field ξ_b. In this case $T_{ab} = 0$, so that while μ_{mnabcd} is nonzero, the stress energy associated with these waves is zero. Is this μ_{mnabcd} all gauge? One might conjecture that this is so since $\mu_{mnabcd} = 4\tau_{mn(ab)(cd)}$, where $\tau_{mna}{}^b{}_c{}^d = \frac{1}{4}k_m k_n k_a \xi^b k_c \xi^d$ has all the required symmetries and positive definiteness properties of a gauge piece. When k_a is null the above condition is equivalent to $\gamma_{ab} = k_{(a}\xi_{b)} + p_{ab}$ for some fields ξ_b and p_{ab} such that $p_{ab} = p_{ab}$, $p_{ab}k^a = 0$, and $p^m{}_m = 0$ (this decomposition of γ_{ab} is not unique.) In this case, we find

$$T_{ab} = \tfrac{1}{4}p_{mn}p^{mn}k_a k_b. \tag{26}$$

Notice that T_{ab} is trace-free as required by Theorem 1. Since $p_{mn}p^{mn} \geqslant 0$, we see T_{ab} has the same form as the stress-energy tensor of a null fluid. Further, we can choose k^a such that, from the condition $\nabla_a T^{ab} = 0$, we have $k^m\nabla_m k^a = 0$ and $\nabla_a(a^2k^a) = 0$, where $a^2 = \frac{1}{4}p_{mn}p^{mn}$. That is, associated with the wave in the high-frequency limit is a null vector k^a that is geodetic, and a current $k^a a^2$ that is conserved.

This simple case raises a number of issues. First, Isaacson examined a case similar to this one and finds that Eq. (26) holds under the assumption that k_a is hypersurface orthogonal. Is hypersurface orthogonality a consequence of conditions (i)–(iv)? Second, does Eq. (26) have a simple analog in the general case? We suspect that it does.

Conjecture: For any tensor field $T_{ab} = \mu_{[b|[am]|n]}{}^{mn}$ obtained from a one-parameter family of metrics satisfying conditions (i)–(iv), there exists a scalar field $a^2(x,k)$ defined on the null cotangent bundle, such that

$$T_{ab}(x) = \int a^2(x,k)k_a k_b\, dV_k, \tag{27}$$

$$k^m\nabla_m a^2(x,k) = 0, \tag{28}$$

where $x \in M$, (x,k) is a point of the null cotangent bundle, and the integral is performed over the null cone.

Think of $a^2(x,k)$ as a "particle" distribution function on the null cotangent bundle. Then, if this conjecture is true, we see that these particles do not interact directly, but that they do affect one another by their effect on the background space-time. Further, we then have a complete system for describing the effect of high-frequency waves on the background space-time. That is, on a fixed smooth manifold M, we have a metric g_{ab}, and on the null cotangent bundle constructed from M, we have a scalar field $a^2(x,k)$. The fields g_{ab} and $a^2(x,k)$ then evolve together via $G_{ab}[g] = T_{ab}$, with T_{ab} given in Eq. (27), and via Eq. (28).

It would also be interesting to know if the converse of the above conjecture is true: Given any metric g_{ab} on a fixed manifold M, and a scalar field $a^2(x,k)$ defined on the null cotangent bundle constructed from M, such that $k^m\nabla_m a^2(x,k) = 0$ and

$$G_{ab}[g] = \int a^2(x,k)k_a k_b\, dV_k,$$

then there exists a one-parameter family of metrics $g_{ab}(\lambda)$ satisfying conditions (i)–(iv) with $g_{ab}(0) = g_{ab}$. If this is true, we do not need to impose any restrictions on what fields $a^2(x,k)$ and g_{ab} we use, other than (27) and (28) above.

If the conjecture above is true, then T_{ab} is positive semidefinite. Since T_{ab} is trace-free this would also imply that T_{ab} satisfies the dominant energy condition. Physically this property would mean that the effective stress-energy tensor contains no tensions (negative pressures). It seems plausible that this property of the effective stress-energy tensor could be proved directly from the properties of μ_{mnabcd}.

VI. DISCUSSION AND CONCLUSIONS

We know, from Theorem 1, that in the high-frequency limit, the gravitational waves act as a source of curvature for the background space-time $(M, g_{ab}(0))$. Do these waves have any other physical effects? For instance, suppose we fix a point $p \in M$ and a tangent vector ξ^a at p. Consider the one-parameter family of geodesics $\gamma(\lambda)$ that these initial conditions define in the space-times $(M, g_{ab}(\lambda))$. Does $\gamma(\lambda) \to \gamma(0)$, in some sense, as $\lambda \to 0$? Although the convergence of geodesics does hold in the case of the plane-wave example given in the Introduction (the curves converging uniformly), it is not clear what is true in general. Thus consider the geodesic equation

$$\xi^m \nabla_m \xi^a + C^a{}_{mn}(\lambda) \xi^m \xi^n = 0, \tag{29}$$

where ξ^a is the tangent vector to $\gamma(\lambda)$, ∇_m is the derivative operator compatible with $g_{ab}(0)$, and $C^a{}_{mn}(\lambda)$ is given by Eq. (14). We know that $C^a{}_{mn}(\lambda) \to 0$ weakly, but this is hardly enough to guarantee that $\gamma(\lambda) \to \gamma(0)$, e.g., pointwise, as $\lambda \to 0$. Does $\gamma(\lambda) \to \gamma(0)$ generically, in some sense? If not, can some additional condition be added to (i)–(iv) to guarantee a consistent limiting behavior of the geodesics as we approach the high-frequency limit?

One potential problem with the present characterization of the high-frequency limit is that it does not give any description of the polarization of the gravitational waves. Would such a polarization manifest any physical effects? If not, the description of the effective stress-energy tensor $\mu_{[b|[am]|n]}{}^{mn}$ would be all that is needed—our conjecture, if true, would then provide this description.

The characterization given here for high-frequency gravitational waves can be easily extended to other fields on a fixed background space-time (M, g_{ab}). For example, for the case of electromagnetism we would consider a one-parameter family of electromagnetic potentials $A_a(\lambda)$ such that

(i′) $\nabla^a \nabla_{[a} A_{b]}(\lambda) = 0$, for all $\lambda > 0$,
(ii′) $A_a(\lambda)$ converges to $A_a(0)$ uniformly as $\lambda \to 0$,
(iii′) $\nabla_m (A_a(\lambda) - A_a(0))$ is uniformly bounded,
(iv′) $\nabla_m (A_a(\lambda) - A_a(0)) \nabla_n (A_b(\lambda) - A_b(0))$ converges to some tensor field μ_{mnab} weakly as $\lambda \to 0$, where ∇_m is the derivative operator compatible with g_{ab}.

The analysis of the high-frequency limit for the electromagnetic case then proceeds in a manner similar to that which was used in the gravitational case. For instance, we find $\mu_{mnab} = \mu_{(mn)(ab)}$, which follows from an argument similar to that used to show $\mu_{mnabcd} = \mu_{(mn)(cd)(ab)}$ in the gravitational case. We find $\nabla^a \nabla_{[a} A_{b]}(0) = 0$, i.e., the "background" electromagnetic field satisfies Maxwell's equation, by taking the weak limit of the equation in condition (i′). From the definition of the stress-energy tensor,

$$T_{ab} = \nabla_{[a} A_{m]} \nabla_{[b} A_{n]} g^{mn} - \tfrac{1}{4} g_{ab} \nabla_{[m} A_{n]} \nabla^{[m} A^{n]}, \tag{30}$$

we find, using the above conditions, that $T_{ab}[A(\lambda)] \to T_{ab}[A(0)] + \mu_{[b|[am]|n]} g^{mn}$ weakly as $\lambda \to 0$. Further, we find $\mu_{[b|[am]|n]} g^{mn} g^{ab} = 0$, so that in the high-frequency limit, the electromagnetic waves have an effective stress energy $\mu_{[b|[am]|n]} g^{mn}$ that is trace-free and divergence-free.

One further extension is to consider the high-frequency limit of nonvacuum space-times in general relativity. For example, on a fixed manifold M, consider a one-parameter family of metrics $g_{ab}(\lambda)$ satisfying conditions (ii)–(iv), and electromagnetic potentials $A_a(\lambda)$ satisfying conditions (ii′)–(iv′), that further satisfy

$${}^{\lambda}\nabla^a \nabla_{[a} A_{b]}(\lambda) = 0, \quad \text{for } \lambda > 0, \tag{31}$$

and

$$G_{ab}[g(\lambda)] = T_{ab}[A(\lambda)], \quad \text{for } \lambda > 0. \tag{32}$$

Here, ${}^{\lambda}\nabla_a$ is the derivative operator compatible with $g_{ab}(\lambda)$, and $T_{ab}[A(\lambda)]$ is given by

$$\begin{aligned} T_{ab}[A(\lambda)] &= \nabla_{[a} A_{m]}(\lambda) \nabla_{[b} A_{n]}(\lambda) g^{mn}(\lambda) \\ &\quad - \tfrac{1}{4} g_{ab}(\lambda) \nabla_{[p} A_{m]}(\lambda) \nabla_{[q} A_{n]}(\lambda) g^{mn}(\lambda) g^{pq}(\lambda). \end{aligned} \tag{33}$$

Taking the weak limit of Eq. (31) we find $\nabla^a \nabla_{[a} A_{b]}(0) = 0$, where ∇_a is the derivative operator compatible with $g_{ab}(0)$. Taking the weak limit of Eq. (32) we find

$$\begin{aligned} G_{ab}[g(0)] &= T_{ab}[A(0)] + \mu_{[b|[am]|n]} g^{mn}(0) \\ &\quad + \mu_{[b|[am]|n]}{}^{mn}. \end{aligned} \tag{34}$$

So, in this case, there are three contributions to the curvature of the background space-time: the stress energy associated with the smooth background electromagnetic field; the effective stress energy associated with the high-frequency electromagnetic waves; and the effective stress energy associated with the high-frequency gravitational waves. Each term in the above expression is trace-free, and while in general the last two terms are not divergence-free, their sum necessarily is.

ACKNOWLEDGMENTS

I would like to thank Bob Geroch for all the helpful suggestions and encouragement he has given me. This paper was submitted in partial fulfillment of the degree requirements for a Ph.D. degree at the University of Chicago.

This research was supported by National Science Foundation Grant No. PHY84-16691 to the University of Chicago.

APPENDIX: PROOFS OF USEFUL FACTS

Here we prove some useful facts about weak convergence, uniform convergence, and uniform boundedness. Let $\alpha(\lambda)$ and $\beta(\lambda)$ be one-parameter families of smooth tensor fields (indices suppressed.) Then,

(a) $\alpha(\lambda)\rightarrow\alpha(0)$ weakly, if $\alpha(\lambda)\rightarrow\alpha(0)$ uniformly,

(b) $\nabla_m\alpha(\lambda)\rightarrow\nabla_m\alpha(0)$ weakly, if $\alpha(\lambda)\rightarrow\alpha(0)$ weakly,

(c) $\alpha(\lambda)\beta(\lambda)\rightarrow 0$ uniformly, if $\alpha(\lambda)\rightarrow 0$ uniformly, and $\beta(\lambda)$ is uniformly bounded,

(d) $\alpha(\lambda)\beta(\lambda)\rightarrow\alpha(0)\beta(0)$ weakly, if $\alpha(\lambda)\rightarrow\alpha(0)$ uniformly, $\beta(\lambda)\rightarrow\beta(0)$ weakly, and $\beta(\lambda)$ is uniformly bounded.

The proofs are straightforward. Without loss of generality, let $\alpha(\lambda)$ and $\beta(\lambda)$ be covariant vector fields.

Proof of (a): Let t^a be any test field. We have $t_a=\hat{t}^a e$, where $\hat{t}^a$ is a tensor field and e is a scalar test field each with common compact support C. Choose any number $\hat{\epsilon}>0$ and any smooth scalar field $\epsilon>0$ such that $\hat{\epsilon}=\int\epsilon|e|$. Since $\alpha_a(\lambda)\rightarrow\alpha_a(0)$ uniformly, there exists a scalar field λ_0 such that $|(\alpha_a(\lambda)-\alpha_a(0))\hat{t}^a|<\epsilon$ for all fields $\lambda<\lambda_0$. Let $\hat{\lambda}=\max_C(\lambda_0)$ (we are guaranteed that $\hat{\lambda}$ exists since λ_0 is continuous and C is compact). Consider

$$\begin{aligned}\left|\int(\alpha_a(\lambda)-\alpha_a(0))t^a\right| &\leqslant\int|(\alpha_a(\lambda)-\alpha_a(0))t^a|\\ &=\int|(\alpha_a(\lambda)-\alpha_a(0))\hat{t}^a||e|\\ &\leqslant\int\epsilon|e|=\hat{\epsilon},\quad\text{for }\lambda<\hat{\lambda}.\end{aligned}$$

In the first step we used the absolute value property of integrals, in the second step we used our decomposition of t^a, and in the third step we used the fact that $\alpha_a(\lambda)\rightarrow\alpha_a(0)$ uniformly. Notice that the third step holds only for $\lambda<\hat{\lambda}$. From this, we see that

$$\int(\alpha_a(\lambda)-\alpha_a(0))t^a\rightarrow 0\quad\text{as }\lambda\rightarrow 0,$$

thus establishing (a).

Proof of (b): Let t^{ma} be any test field, and consider

$$\begin{aligned}\lim_{\lambda\rightarrow 0}\int\nabla_m(\alpha_a(\lambda)-\alpha_a(0))t^{ma}\\ =-\lim_{\lambda\rightarrow 0}\int(\alpha_a(\lambda)-\alpha_a(0))\nabla_m t^{ma}=0.\end{aligned}$$

In the first step we integrated by parts and used the fact that t^{ma} has compact support. In the second step, we used the fact that $\alpha_a(\lambda)\rightarrow\alpha_a(0)$ weakly (test field $\nabla_m t^{ma}$). This establishes (b).

Proof of (c): Let t^{ab} be any smooth tensor field. We can write t^{ab} as a finite sum of products of smooth vector fields, i.e.,

$$t^{ab}=\sum_k \underset{k}{u^a}\underset{k}{v^b},$$

where $\underset{k}{u^a}$ and $\underset{k}{v^b}$ are smooth vector fields for each k. Since $\beta_b(\lambda)$ is uniformly bounded, there exist scalar fields $\underset{k}{M}$ and $\underset{k}{\lambda_0}$ such that $|\beta_b(\lambda)\underset{k}{v^b}|<\underset{k}{M}$ for all fields $\lambda<\underset{k}{\lambda_0}$. Then,

$$\begin{aligned}&\lim_{\lambda\rightarrow 0}|\alpha_a(\lambda)\beta_b(\lambda)t^{ab}|\\ &\quad\leqslant\lim_{\lambda\rightarrow 0}\sum_k|\alpha_a(\lambda)\underset{k}{u^a}|\,|\beta_b(\lambda)\underset{k}{v^b}|\\ &\quad\leqslant\sum_k\underset{k}{M}\lim_{\lambda\rightarrow 0}|\alpha_a(\lambda)\underset{k}{u^a}|=0.\end{aligned}$$

In the first step we used our decomposition of t^{ab} and the triangle inequality, in the second step we used the fact that $\alpha_a(\lambda)\rightarrow 0$ uniformly as $\lambda\rightarrow 0$. This establishes (c).

Proof of (d): Let t^{ab} be any test field, and consider

$$\begin{aligned}&\lim_{\lambda\rightarrow 0}\int(\alpha_a(\lambda)\beta_b(\lambda)-\alpha_a(0)\beta_b(0))t^{ab}\\ &\quad=\lim_{\lambda\rightarrow 0}\int(\alpha_a(\lambda)-\alpha_a(0))\beta_b(\lambda)t^{ab}\\ &\qquad+\lim_{\lambda\rightarrow 0}\int(\beta_b(\lambda)-\beta_b(0))\alpha_a(0)t^{ab}\\ &\quad=0.\end{aligned}$$

The first step is an identity. In the second step, the first term is zero since $\alpha_a(\lambda)-\alpha_a(0)\rightarrow 0$ uniformly and $\beta_b(\lambda)$ is uniformly bounded, so, from (c), $(\alpha_a(\lambda)-\alpha_a(0))\beta_b(\lambda)\rightarrow 0$ uniformly, and so, from (a), we conclude $(\alpha_a(\lambda)-\alpha_a(0))\beta_b(\lambda)\rightarrow 0$ weakly. The second term is also zero since $\beta_b(\lambda)\rightarrow\beta_b(0)$ weakly [test field $\alpha_a(0)t^{ab}$]. This establishes (d).

[1]See, for example, R. M. Wald, *General Relativity* (The University of Chicago Press, Chicago, 1984), p. 183.

[2]R. A. Isaacson, Phys. Rev. **166**, 1263, 1272 (1968).

[3]For other approaches, see M. A. H. MacCallum and A. H. Taub, Commun. Math. Phys. **30**, 153 (1973); Y. Choquet-Bruhat, *ibid.* **12**, 16 (1969).

[4]Our units are such that $8\pi G=c=1$. Our metrics have signature $+2$, and our convention for the Riemann and Ricci tensors are $\nabla_a\nabla_b w_c=\frac{1}{2}R_{abc}{}^d w_d$, and $R_{ab}=R_{amb}{}^m$.

[5]We will show later that it does not matter which smooth derivative operator we use in checking this condition.

Collisions of gravitational and electromagnetic waves that do not develop curvature singularities

Taxiarchis Papacostas
Department of Physics, University of Crete, Iraklion, Greece and Department of Mathematics, University of the Aegean, Karlovassi, Samos, Greece

Basilis C. Xanthopoulos
Department of Physics, University of Crete and Research Center of Crete, Iraklion, Greece

(Received 31 May 1988; accepted for publication 31 August 1988)

A five-parameter family of solutions is investigated, describing the collision of plane-fronted impulsive gravitational and shock electromagnetic waves. In the interaction region, to the future of the collision, it is a locally known solution of the Einstein–Maxwell electrovacuum equations of Petrov type D. The collision results in the formation of a Cauchy horizon. Extensions of the space-time are constructed beyond the Cauchy horizon and beyond certain two-dimensional surfaces that are mere coordinate singularities. It is found that in the extended space-time the following may occur: (i) no curvature singularities, (ii) two-dimensional spacelike curvature singularities, and (iii) two-dimensional timelike curvature singularities, according to the ranges of the parameters of the solution.

I. INTRODUCTION

In the last few years we have been studying the collision of plane-fronted gravitational waves, possibly coupled with electromagnetic or acoustic waves. Because gravity is always attractive it was expected that focusing of the waves would occur and one of the interesting questions is how much focusing does general relativity predict. Within the framework of general relativity, strong focusing would appear by the development of space-time curvature singularities.

In all studies of the problem until 1985 all obtained solutions, describing the collision of plane gravitational waves, shared the feature that they developed three-dimensional spacelike curvature singularities, formed within a finite time from the moment of the collision.[1–5] Note that three-dimensional spacelike curvature singularities are compulsory for any physical observer to encounter them. Our ideas about the singularities that colliding waves may form changed dramatically in early 1986 when Chandrasekhar and one of us (BCX) (Ref. 6) obtained a solution describing the collision of two mixtures of plane-fronted impulsive and shock gravitational waves which, instead of the curvature singularity, develop a Cauchy horizon. When the space-time is extended beyond the horizon one encounters two-dimensional timelike curvature singularities, i.e., singularities that almost all physical observers would miss. The interpretation of this result is that, at least in this solution, general relativity seems to cause much less focusing of the waves than it was previously thought. Since then we have also obtained solutions of the Einstein–Maxwell equations predicting the development of horizons and, subsequently, timelike singularities (actually both two and three dimensional).[7] More recently Ferrari, Ibañez, and Bruni[8] have also obtained solutions of the vacuum equations developing horizons.

Questions arise: Are these two the only kinds of singularity structures that may result in the collisions of plane-fronted gravitational waves? Are there other solutions describing similar physical situations and resulting in the formation of horizons (as opposed to three-dimensional spacelike singularities)? Which one of the two, horizons or singularities, would result more often? Or does one of the two situations appear generically in the collisions of plane-fronted gravitational waves? Since we lack the general theory that would rigorously answer these questions, we have to appeal to the investigation of more exact solutions.

In the present paper we shall answer the second of the previous questions affirmatively by presenting a five-parameter family of solutions of the Einstein–Maxwell equations. The solution describes the interaction of two plane-fronted gravitational and electromagnetic waves that result in the formation of horizons, while the extended space-time exhibits two-dimensional spacelike or timelike curvature singularities or no singularity at all.

The five-parameter family of solutions that describes the interaction region of the space-time is of Petrov type D. Of Petrov type D are also all the vacuum solutions developing horizons that have been obtained by Ferrari, Ibañez, and Bruni.[8] In fact, they have even obtained a family of solutions which, while it would generally develop a spacelike curvature singularity, does develop a horizon for a particular value of a free parameter for which it also becomes of Petrov type D. On the other hand, from the solutions developing horizons that have been studied by Chandrasekhar and Xanthopoulos, the only one that fails to be of Petrov type D is the infinite-parameter family of solutions of the Einstein–Maxwell equations with hypersurface orthogonal Killing fields described in §5 of Ref. 7. Is it conceivable that there exists an unexpected interrelationship between solutions (with two spacelike Killing fields) that develop horizons and their algebraic type, reminiscent of the situation in the black hole solutions (with one timelike and one spacelike Killing field) which are also of Petrov type D?

Recently Yurtsever[9] has argued that all the horizons that develop in the collisions of plane-fronted gravitational waves are unstable. He considers the evolution of plane sym-

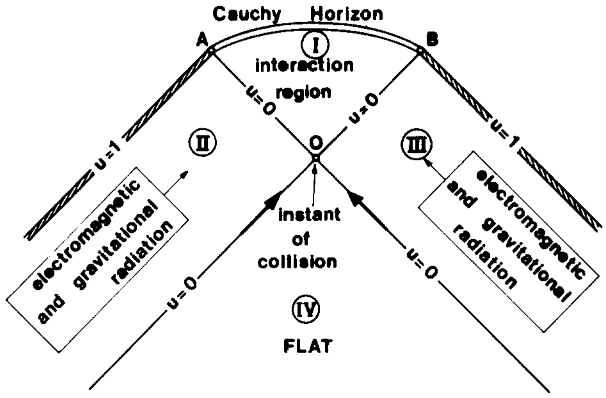

FIG. 1. The four regions of the space-time. Region IV is flat, where the waves propagate. The impulsive gravitational and shock electromagnetic waves propagate along the null boundaries $v=0$ and $u=0$, separating regions II and IV, and III and IV, respectively. In region II observers see the shower of pure gravitational and electromagnetic radiation following the wave front propagated along $v=0$; but they are not aware of the other wave. Symmetrical considerations apply in region III. The collision occurs at O and the interaction is described by region I. The Cauchy horizon develops within a finite time from the instance of the collision.

metric—but otherwise arbitrary—initial data in the null surfaces separating regions I and II and I and III (see Fig. 1). He argues that (as it has been presented so far, his analysis is completely rigorous only for the vacuum Einstein equations and for hypersurface orthogonal Killing fields) the evolution of these data would diverge logarithmically on the horizon. However, a recent detailed analysis of the perturbations of the Bell–Szekeres space-time[10] has shown that arbitrary initial data on the null boundaries does not seem to be compatible with the smoothness of the perturbations before the collision and that most probably these nonsmooth perturbations lead to the divergent behavior on the horizon. It is clear that more analysis, and in particular, more solutions developing horizons, would be needed to clarify these problems.

We shall follow the notation of Refs. 5–7. In fact, for the sake of brevity, some familiarity with these papers will be assumed. In Secs. II and III we describe the metric in the interaction region and in the regions before the collision, respectively. Section IV obtains the curvature scalars that are used in Sec. VI A in the study of the singularities in the extended space-time. Section V, the most technical, describes smooth extensions across the coordinate singularities. As far as we are aware, it is the first time that extensions of the space-time across timelike surfaces are obtained in problems of colliding waves.

II. REGION I

The space-time representing the collision of two gravitational waves consists of four regions (Fig. 1). Region IV is the part of space-time before the arrival of neither wave, regions II and III are the parts of space-time after the passage of only one of the waves, and region I is the interaction region, to the future of the collision of the two waves. The incoming waves propagate in that space-time (region IV) along the null boundaries separating regions II and IV, and III and IV.

For the space-time, we shall be considering that the metric in region I will be taken to be a special case of a family of solutions given by Debever,[11] Plebanski,[12] and Plebanski and Demianski,[13,14] (to be referred to, henceforth, as the DP solution), with two spacelike commuting Killing fields but without cosmological constant,

$$\begin{aligned} ds^2 = {} & (t^2+z^2)\left[\frac{(dt)^2}{E^2}-\frac{(dz)^2}{H^2}\right] \\ & -(t^2+z^2)^{-1}\,[\,E^2(dy-z^2\,dx)^2 \\ & +H^2(dy+t^2\,dx)^2], \end{aligned} \tag{2.1}$$

where a, b, c, f, and g are real constants, the charge of the metric is $e^2=-\frac{1}{2}(c+g)$, and we have introduced the notation

$$\alpha=(f^2+2ag)^{1/2},\quad \beta=(b^2+2ac)^{1/2}, \tag{2.2}$$

and

$$\begin{aligned} E^2=E^2(t) &= -\tfrac{1}{2}at^2+bt+c \\ &= -(1/2a)[(at-b)^2-\beta^2], \\ H^2=H^2(z) &= -\tfrac{1}{2}az^2+fz+g \\ &= -(1/2a)[(az-f)^2-\alpha^2]. \end{aligned} \tag{2.3}$$

To simplify our considerations we shall allow to the parameters of the DP solution the ranges $a>0$, $b^2+2ac>0$, and $f^2+2ag>0$. Note that $a>0$ is an assumption–restriction we are making while the other two inequalities are the conditions that the two Killing fields are spacelike. The case $a<0$ would split our considerations in a lot of subcases and it is not investigated here. The DP metric represents the most general type D solution of the Einstein–Maxwell electrovacuum equations (with nonsingular electromagnetic field tensor whose principal null directions are aligned with those of the Weyl tensor) for which the Hamilton–Jacobi equation is solvable by separation of variables.[15,16] In fact, this solution is a generalization of the metric $[A]$ of Carter[17] permitting the two-parameter invertible Abelian group to have spacelike orbits (two spacelike Killing fields), instead of timelike orbits of Carter's metric (one timelike and one spacelike Killing field). When the charge e^2 vanishes, the metric (2.1) represents a vacuum solution. [It should be noted that we have used the coordinates (t,z,y,x) instead of (u,v,w,x) of Debever and MacLenaghan[15] in order to indicate that the metric (2.1) admits the two spacelike Killing fields $(\partial/\partial x)$ and $(\partial/\partial y)$ spanning the wave fronts of the incoming gravitational and electromagnetic waves.]

The coordinates of the metric (2.1) take values in

$$at\in(b-\beta,b+\beta),\quad az\in(f-\alpha,f+\alpha),\quad x\in\mathbb{R},\quad y\in\mathbb{R}. \tag{2.4}$$

It will be necessary to express the metric (2.1) in suitable null coordinates. These coordinates can be most easily found by first expressing the metric (2.1) in the gauge and coordinates introduced by Chandrasekhar and Ferrari,[4] which were found very useful in the description of colliding waves. Thus setting

$$at-b=\beta\eta,\quad az-f=\alpha\mu,\quad \eta\in(-1,+1),\quad \mu\in(-1,+1), \tag{2.5}$$

we find that

$$E^2 = (\beta^2/2a)(1-\eta^2), \quad H^2 = (\alpha^2/2a)(1-\mu^2), \tag{2.6}$$

$$\frac{(dt)^2}{E^2} - \frac{(dz)^2}{H^2} = \frac{2}{a}\left[\frac{(d\eta)^2}{1-\eta^2} - \frac{(d\mu)^2}{1-\mu^2}\right], \tag{2.7}$$

and the metric (2.1) becomes

$$\begin{aligned} ds^2 = {} & 2a^{-1}(t^2+z^2)\left[\frac{(d\eta)^2}{1-\eta^2} - \frac{(d\mu)^2}{1-\mu^2}\right] \\ & - \frac{1}{2a(t^2+z^2)}\,[\beta^2(1-\eta^2)(dy - z^2\,dx)^2 \\ & + \alpha^2(1-\mu^2)(dy + t^2\,dx)^2]\,, \end{aligned} \tag{2.8}$$

where t and z should be expressed in terms of η and μ from Eqs. (2.5).

It is then straightforward to express the metric (2.8) in null coordinates. As in Ref. 6, Eq. (15), we set

$$\eta = u\sqrt{1-v^2} + v\sqrt{1-u^2}, \quad \mu = u\sqrt{1-v^2} - v\sqrt{1-u^2}, \tag{2.9}$$

for which the identity

$$\frac{(d\eta)^2}{1-\eta^2} - \frac{(d\mu)^2}{1-\mu^2} = \frac{4(du)(dv)}{\sqrt{1-u^2}\,\sqrt{1-v^2}} \tag{2.10}$$

establishes the nullness of the coordinates (u,v). Note that the determinant of the dx, dy part of the metric (2.8) is $\alpha^2\beta^2(1-\eta^2)(1-\mu^2)/(4a^2)$; therefore, except for the overall multiplicative constant factor, our gauge coincides with that used in Refs. 4–7.

III. EXTENSIONS TO THE REGIONS BEFORE THE COLLISION

The space-time metric in regions II, III, and IV is obtained from the metric in region I by expressing it in the null coordinates (u,v) and performing the Penrose extension

$$u \to u\,\mathscr{H}(u), \quad v \to v\,\mathscr{H}(v), \tag{3.1}$$

where $\mathscr{H}$ is the Heaviside unit step function [recall that $(u,v) = (0,0)$ represents the instant of the collision]. Upon the Penrose extension (3.1) the metric would be C^0 but not C^1 on the entire space-time. It was clarified in Refs. 5 and 6 that the Dirac δ-function singularities would appear only in the Weyl part of the curvature while the Ricci part would be actually continuous. Since $u = 0$ is the null boundary between regions I and II, the metric in region II is obtained by setting $u = 0$ in the metric of region I expressed in null coordinates.

We find that the obtained metric in region II is of the form studied in Sec. 8 of Ref. 5, i.e.,

$$ds^2 = e^{2\nu}(du)(dv) - e^{2\psi}(dx_1 - q_2\,dx_2)^2 - e^{2\mu_2}(dx_2)^2, \tag{3.2}$$

where

$$e^{2\nu} = \frac{8(t^2+z^2)}{a\sqrt{1-v^2}}, \quad e^{2\psi} = \frac{(1-v^2)(\alpha^2t^4+\beta^2z^4)}{2a(t^2+z^2)},$$

$$e^{2\mu_2} = \frac{\alpha^2\beta^2(1-v^2)(t^2+z^2)}{2a(\alpha^2t^4+\beta^2z^4)}, \quad q_2 = \frac{\beta^2z^2-\alpha^2t^2}{\alpha^2t^4+\beta^2z^4},$$

$$x_1 = x, \quad x_2 = y, \quad at = \beta v + b, \quad az = f - \alpha v. \tag{3.3}$$

For further comparisons with Sec. 8 of Ref. 5 it should be noted that

$$e^{\beta_{CX}} = e^{\psi}e^{\mu_2} = (\alpha\beta/2a)(1-v^2), \tag{3.4}$$

where β_{CX} represents what was denoted by β in Ref. 5.

It has been explained in Sec. 8 of Ref. 5 that any metric of the form (3.2) is of Petrov type N and its curvature is characterized by the two scalar quantities $L + M$ and $M - L + 2iN$, given by Eqs. (164) and (168), respectively, of Ref. 5. Note that $L + M$ describes the Ricci curvature and $M - L + 2iN$ the Weyl scalar Ψ_4 of the Weyl curvature (in a suitable null tetrad) of the metric (3.1). By using Eqs. (3.3) and (3.4) and Eqs. (164) and (168) of Ref. 5 we find that

$$\begin{aligned} L + M &= \frac{(\alpha^2+\beta^2-f^2-b^2)}{4a(t^2+z^2)^2\sqrt{1-v^2}} \\ &= -\frac{e^2}{(t^2+z^2)^2\sqrt{1-v^2}}, \end{aligned} \tag{3.5}$$

while $M - L + 2iN$ as a function of v is too complicated to be of any particular use. Note that the inequality $L + M < 0$, which is the necessary and sufficient condition for the metric in region II to be a solution of the Einstein–Maxwell equations, is always satisfied. When $e^2 = 0$, and the metric in region I is a solution of the vacuum Einstein equations, the metric in region II is also a solution of the vacuum equations.

Along the null boundary $v = 0$, $-\infty < u < 0$, separating regions II and IV, $L + M$ suffers a finite discontinuity while $M - L + 2iN$ suffers both a finite discontinuity and a δ-function singularity. These determine the amplitudes of the impulsive gravitational and the shock gravitational and electromagnetic waves that collide. By using the expressions (166), (168), (178), and (179) of Ref. 5 we find, after some considerable calculations, that the following occurs.

(a) The magnitude of the $\mathscr{H}$-function discontinuity (the shock gravitational wave) suffered by Ψ_4 is

$$\frac{3a^3(\alpha^2+\beta^2)}{4(b^2+f^2)^3M}\,(\mathbf{K} - i\Lambda)^2, \tag{3.6}$$

where

$$\begin{aligned} \mathbf{K} &= \beta f^3 + \alpha a^2 f - \alpha b^3 - a^2\beta b, \\ \Lambda &= (a^2 + bf)(\beta f + \alpha b), \\ \mathbf{M} &= (\alpha a^2 + \beta f^2)^2 + (\alpha b^2 + \beta a^2)^2. \end{aligned} \tag{3.7}$$

(b) The magnitude of the δ-function singularity suffered by Ψ_4 has real part

$$\begin{aligned} \frac{a^3}{4(f^2+b^2)^2M}\, & [\alpha\beta^2 f^3(f^2+2b^2) - \alpha^2\beta b^3(2f^2+b^2) \\ & + \alpha\beta^2a^2(a^2f - 2b^3) + \beta^3 b(f^4 - a^4) \\ & + \alpha^3 f(a^4 - b^4) + 2a^2\beta f^3(\alpha^2+\beta^2) \\ & - \alpha^2 a^2 b(2\alpha b^2 + \beta a^2)] \end{aligned} \tag{3.8}$$

and imaginary part

$$\frac{a^3(\alpha b+\beta f)}{4(f^2+b^2)^2M}[\alpha^2(b^4-a^4)+\beta^2(f^4-a^4)$$
$$-2\alpha\beta bf(b^2+f^2)-2a^2bf(\alpha^2+\beta^2)]. \tag{3.9}$$

(c) The magnitude of the $\mathscr{H}$-function discontinuity suffered by Φ_{22} is

$$e^2a^4/(b^2+f^2)^2. \tag{3.10}$$

The expressions (3.6)–(3.10) describe the characteristics of the two mixtures of plane gravitational and electromagnetic waves that collide. The considered waves have more degrees of freedom—determined by the five free parameters—than the solution for colliding waves considered so far. Note that the expression (3.10) points once more to the conclusion we have already mentioned: when $e^2=0$, there is no shock electromagnetic wave and the space-time is a solution of the vacuum Einstein equations.

IV. THE NEWMAN–PENROSE CURVATURE SCALARS IN REGION I

In our previous studies of collisions of plane waves the curvature scalars of the Newman–Penrose formalism were evaluated in a null tetrad suitably chosen to take advantage of the two Killing fields. In these tetrads $\Psi_1=\Psi_3=\phi_1=0$ while, generally, the remaining Weyl and Maxwell scalars are different from zero.

Since the family of solutions we are presently considering is of Petrov type D, a more suitable choice of the null tetrad will be the one in which the two real null vectors coincide with the repeated principal null directions of the Weyl tensor. This null tetrad can be found, for instance, in Debever and MacLenaghan,[16] Eqs. (3.24a)–(3.24c). The only nonvanishing Newman–Penrose curvature scalars are Ψ_2 and Φ_{11} given by

$$\Psi_2=-[2(t^2+z^2)^3]^{-1}[bt^3+3fzt+2(g+c)t^2-3bz^2t$$
$$-fz^3-2(g+c)z^2]-[i/3(t^2+z^2)^3]$$
$$\times[ft^3-3bzt^2-3fz^2t-4(g+c)zt+bz^3], \tag{4.1}$$
$$\Phi_{11}=e^2/(t^2+z^2)^2, \tag{4.2}$$

where e^2 is the charge of the metric. Obviously, $e=0$ is the necessary and sufficient condition for the metric (2.1) to be a vacuum solution.

V. EXTENSION OF THE SPACE-TIME BEYOND THE COORDINATE SINGULARITIES AT $E^2=0$ AND $H^2=0$

Obviously, the spacelike surface

$$t=(1/a)(b-\beta)\Leftrightarrow E^2(t)=0 \tag{5.1}$$

and the timelike surface

$$z=(1/a)(f-\alpha)\Leftrightarrow H^2(z)=0 \tag{5.2}$$

are, at least, coordinate singularities. Moreover, the results of Sec. IV show that all the Weyl and electromagnetic scalars Ψ_2 and Φ_{11} remain finite on these surfaces. Naturally, therefore, we wonder whether one could extend the space-time of region I beyond these surfaces and investigate the nature of these surfaces as well as the global structure (nature of singularities) of the extended space-time. Note, for instance, that in previous investigations of colliding waves, no one has considered extensions across timelike surfaces. Moreover, attention should be paid to the solution described in Ref. 18 in which, although all curvature scalars remain finite on the (at least) coordinate singularities, an extension of the space-time beyond these surfaces is lacking; and this unsatisfactory situation prevents us from understanding the nature of these surfaces.

Some preliminary considerations concerning the nature of the surfaces described by Eqs. (5.1) and (5.2) will be useful. The surface $E(t)=0$ corresponds to $\eta=1$ ($\eta=-1$ does not describe points within region I) or

$$u^2+v^2=1,\quad 0<u<1,\quad 0<v<1, \tag{5.3}$$

in the null coordinates. Hence extension beyond $E=0$ would correspond to an extension beyond the arc AB of Fig. 2, i.e., beyond a three-dimensional spacelike surface, to the future of the collision.

On the other hand, $H(z)=0$ corresponds to $\mu=\pm 1$. Within region I, $\mu=+1$ is the point $B=(v=0,u=1)$ and $\mu=-1$ is the point $A=(u=0,v=1)$ of Fig. 2. We expect, therefore, that the two surfaces $H(z)=0$ would be two-dimensional and that an extension across them would not open new space-time regions; it would merely show that they are regular surfaces and it would reveal their nature.

A. Extension across $E^2(t)=0$

We take the metric (2.1) in region I and we consider E and H, instead of t and z, as coordinates. The metric (2.1) is written in the alternative form

$$ds^2=4(t^2+z^2)\left[\frac{(dE)^2}{(b-at)^2}-\frac{(dH)^2}{(f-az)^2}\right]$$
$$-\frac{E^2H^2(t^2+z^2)}{E^2+H^2}(dx)^2-\frac{(E^2+H^2)}{t^2+z^2}$$
$$\times\left(dy+\frac{t^2H^2-z^2E^2}{E^2+H^2}dx\right)^2, \tag{5.4}$$

where t and z should now be considered as functions of E and H, respectively, given by

$$at=b+(\beta^2-2aE^2)^{1/2},\quad az=f+\epsilon(\alpha^2-2aH^2)^{1/2}. \tag{5.5}$$

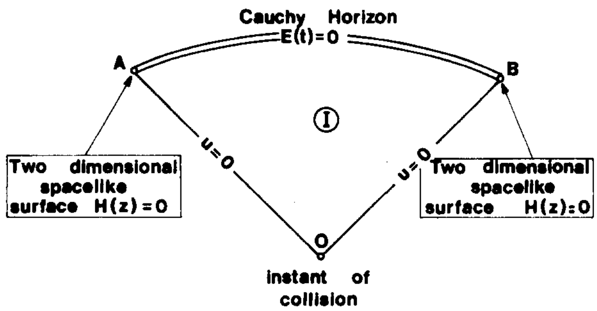

FIG. 2. The coordinate singularity at $E(t)=0$ corresponds to a Cauchy horizon. After the extension it is found that it consists of two null surfaces. The coordinate singularity at $H(z)=0$ corresponds to A and B, which are regular two-dimensional spacelike surfaces.

In Eqs. (5.5) ϵ stands for plus or minus one; $\epsilon = +1$ would correspond to $\mu = +1$ (point B) and $\epsilon = -1$ to $\mu = -1$ (point A) in Fig. 2. On the other hand, we do not have to consider two cases (plus or minus) in the first of Eqs. (5.5) because $E=0$ should correspond to $\eta = 1$ (and not $\eta = -1$ which is outside region I).

Near $E=0$, the metric (5.4) behaves like

$$ds^2 \simeq 4(t_0^2+z^2)\left[\frac{(dE)^2}{\beta^2} - \frac{E^2}{4}(dx)^2\right] - \frac{H^2}{(t_0^2+z^2)}(dy+t_0^2\,dx)^2 - \frac{4(t_0^2+z^2)}{(f-az)^2}(dH)^2, \tag{5.6}$$

where

$$t_0 = (b+\beta)/a. \tag{5.7}$$

The first two terms of the right-hand side of Eq. (5.6) suggest to perform the coordinate transformation

$$\begin{aligned}&(E,x,y,H)\to(\xi,\zeta,\tilde{y},H),\\ &\xi = E\exp(\beta x/2),\quad \zeta = E\exp(-\beta x/2),\\ &\tilde{y} = y + [(b+\beta)^2/a^2]x.\end{aligned} \tag{5.8}$$

Since $\xi\zeta = E^2$ and $\xi/\zeta = e^{\beta x}$ we obtain

$$dE = \frac{E}{2}\left(\frac{d\xi}{\xi}+\frac{d\zeta}{\zeta}\right),\quad dx = \frac{1}{\beta}\left(\frac{d\xi}{\xi}-\frac{d\zeta}{\zeta}\right). \tag{5.9}$$

By using the identity

$$\begin{aligned}&\frac{t^2H^2 - z^2E^2}{E^2+H^2} - \frac{(b+\beta)^2}{a^2}\\ &\quad = -\frac{E^2}{a^2(E^2+H^2)}\left\{a^2z^2 + 2aH^2 + (b+\beta)^2 + \frac{4abH^2}{\beta+\sqrt{\beta^2-2aE^2}}\right\},\end{aligned} \tag{5.10}$$

we find, after considerable calculations, that the metric (5.4) becomes

$$\begin{aligned}ds^2 = {}& \frac{(t^2+z^2)(\beta^2+2aH^2)}{\beta^2(E^2+H^2)(\beta^2-2aE^2)}(\zeta^2\,d\xi^2+\xi^2\,d\zeta^2)\\ &+\frac{2(t^2+z^2)[\beta^2(E^2+2H^2)-2aE^2H^2]}{\beta^2(E^2+H^2)(\beta^2-2aE^2)}\\ &\times(d\xi)(d\zeta) - \frac{4(t^2+z^2)}{(\alpha^2-2aH^2)}(dH)^2 - \frac{(E^2+H^2)}{(t^2+z^2)}\\ &\times\left\{d\tilde{y} - \frac{1}{a^2\beta(E^2+H^2)}\left[a^2z^2+2aH^2+(b+\beta)^2 + \frac{4abH^2}{\beta+\sqrt{\beta^2-2aE^2}}\right](\zeta\,d\xi - \xi\,d\zeta)\right\}^2.\end{aligned} \tag{5.11}$$

Obviously the metric (5.11) is smooth in a neighborhood of $E^2 = \xi\zeta = 0$. Moreover, its determinant is

$$-\frac{16H^2(t^2+z^2)^2}{\beta^2(\beta^2-2aE^2)(\alpha^2-2aH^2)}, \tag{5.12}$$

i.e., nonvanishing, and therefore the metric is also invertible in a neighborhood of $E=0$. We conclude, therefore, that the surface $E=0$ is a regular surface of the metric (2.1) and that the metric is extendible beyond this surface. The nature of the surface $E=0$ and of the extended space-time will be considered in Sec. VI.

B. Extension across $H^2(z)=0$

Near $H^2=0$ the metric (5.4) behaves like

$$\begin{aligned}ds^2 = {}& -4(t^2+z_0^2)\left[\frac{(dH)^2}{\alpha^2}+\frac{H^2}{4}(dx)^2\right]\\ &-\frac{E^2}{t^2+z_0^2}(dy - z_0^2\,dx)^2 + \frac{4(t^2+z_0^2)(dE)^2}{(b-at)^2},\end{aligned} \tag{5.13}$$

where

$$z_0 = (f+\epsilon\alpha)/a. \tag{5.14}$$

Note that the part of the metric in the squared brackets in the first term of Eq. (5.13) is now positive definite while the corresponding term in Eq. (5.6) was negative definite. We now perform the coordinate transformation

$$\begin{aligned}&(H,x,y,E)\to(\xi,\zeta,\tilde{y},E),\\ &H^2 = \xi^2+\zeta^2,\quad \tan(\alpha x/2) = \zeta/\xi,\quad y = z_0^2x+\tilde{y}.\end{aligned} \tag{5.15}$$

Using that

$$H(dH) = \xi(d\xi)+\zeta(d\zeta),\quad dx = \frac{2}{\alpha}\cdot\frac{\xi(d\zeta)-\zeta(d\xi)}{H^2}, \tag{5.16}$$

and the identities

$$\begin{aligned}&\frac{E^2}{\alpha^2(E^2+H^2)} - \frac{1}{\alpha^2-2aH^2}\\ &\quad = -\frac{H^2(\alpha^2+2aE^2)}{\alpha^2(E^2+H^2)(\alpha^2-2aH^2)},\\ &\frac{\xi^2(d\zeta)^2+\zeta^2(d\xi)^2}{(f-az)^2} + \frac{E^2}{\alpha^2(E^2+H^2)}\\ &\quad\times[\xi^2(d\zeta)^2+\zeta^2(d\xi)^2]\\ &\quad = \frac{H^2}{\alpha^2(\alpha^2-2aH^2)(E^2+H^2)}\\ &\quad\quad\times\{[\alpha^2(E^2+\xi^2)-2aE^2\zeta^2](d\xi)^2\\ &\quad\quad+[\alpha^2(E^2+\zeta^2)-2aE^2\xi^2](d\zeta)^2\},\end{aligned} \tag{5.17}$$

and

$$\begin{aligned}&z_0^2 + \frac{t^2H^2-z^2E^2}{E^2+H^2}\\ &\quad = \frac{H^2(t^2+z^2)}{E^2+H^2} + \frac{2H^2}{a}[1+2\epsilon f/(\alpha+\sqrt{\alpha^2-2aH^2})],\end{aligned}$$

we find that the metric (5.4) becomes

$$\begin{aligned}ds^2 = {}& \frac{4(t^2+z^2)(dE)^2}{(b-at)^2} - \frac{4(t^2+z^2)}{\alpha^2(E^2+H^2)(\alpha^2-2aH^2)}\\ &\times\{2(\alpha^2+2aE^2)\xi\zeta(d\xi)(d\zeta)\\ &+[\alpha^2(E^2+\xi^2)-2aE^2\zeta^2](d\xi)^2\\ &+[\alpha^2(E^2+\zeta^2)-2aE^2\xi^2](d\zeta)^2\}\end{aligned}$$

$$-\frac{E^2+H^2}{t^2+z^2}\left\{d\tilde{y}+\frac{2(\xi\, d\zeta-\zeta\, d\xi)}{\alpha}\left[\frac{t^2+z^2}{E^2+H^2}\right.\right.$$
$$\left.\left.+\frac{2}{a}\left(1+\frac{2\epsilon f}{\alpha+\sqrt{\alpha^2-2aH^2}}\right)\right]\right\}^2. \tag{5.18}$$

The determinant of the metric (5.18) is

$$-\frac{64E^2(t^2+z^2)^2}{\alpha^2(\alpha^2-2aH^2)(\beta^2-2aE^2)}. \tag{5.19}$$

Therefore the metric (5.18) is smooth and invertible on the two-dimensional surface

$$H^2=\xi^2+\zeta^2=0,\quad x\in\mathbb{R},\quad y\in\mathbb{R}. \tag{5.20}$$

VI. THE EXTENDED SPACE-TIME

First, we clarify the nature of the surfaces $E=0$ and $H=0$. For that purpose, of course, we should consider the space-time metric in the alternative coordinate forms (5.8), (5.11) and (5.15), (5.18), which cover, respectively, these surfaces.

A. The surface $E=0$

From the metric (5.11) we find the values of its contravariant components

$$g^{\xi\xi}=-\frac{(\beta^2+2aH^2)\xi^2}{4H^2(t^2+z^2)},\quad g^{\zeta\zeta}=-\frac{(\beta^2+2aH^2)\zeta^2}{4H^2(t^2+z^2)}. \tag{6.1}$$

The squared norms of the vector fields orthogonal to the surfaces $\xi=$ const and $\zeta=$ const are

$$g^{mn}(\nabla_m\xi)(\nabla_n\xi)=g^{\xi\xi}=-\frac{(\beta^2+2aH^2)\xi^2}{4H^2(t^2+z^2)},$$
$$g^{mn}(\nabla_m\zeta)(\nabla_n\zeta)=g^{\zeta\zeta}=-\frac{(\beta^2+2aH^2)\zeta^2}{4H^2(t^2+z^2)}. \tag{6.2}$$

Obviously, the vector fields $\nabla_m\xi$ and $\nabla_n\xi$ become null on the hypersurfaces $\xi=0$ and $\zeta=0$, respectively. We conclude, therefore, that the surface

$$E^2=\xi\zeta=0 \tag{6.3}$$

consists of two null surfaces. The situation resembles that described in Sec. B of Ref. 6 and depicted in Fig. 5 of that reference.

Region I, before the extension, corresponds to the $\xi>0$, $\zeta>0$ part of the extended space-time. By setting

$$\xi=-E\exp(\beta x/2),\quad \zeta=-E\exp(-\beta x/2), \tag{6.4}$$

in the expression (5.11) the metric takes the form (5.4). This observation shows that the regions I and I_0 are isometric.

The expressions (4.1) and (4.2) for the curvature scalars show that a curvature singularity occurs only when

$$t^2+z^2=0. \tag{6.5}$$

We now investigate when and where this singularity occurs, and what is its character, for different values of the parameters.

The condition (6.5) is equivalent to $t=z=0$ or, by Eqs. (5.5) and (5.8), to

$$(\beta^2-2a\xi\zeta)^{1/2}=-b\quad\text{and}\quad \alpha^2-2aH^2=f^2\Leftrightarrow H^2=g. \tag{6.6}$$

We immediately conclude that *when $b>0$ or when $g>0$, there is no curvature singularity.*

Next we consider the case

$$b<0\quad\text{and}\quad g>0. \tag{6.7}$$

The conditions (6.6) for the singularity then read

$$\xi\zeta=c\quad\text{and}\quad H^2=g. \tag{6.8}$$

If $c<0$, the singularity occurs for

$$\xi\zeta=c<0,\quad H=\sqrt{g},\quad \forall\,\tilde{y}\in\mathbb{R}: \tag{6.9}$$

this is a two-dimensional timelike surface, the singularity occurs in the two open regions that "glue" regions I and I_0 together, and the situation is similar to that described in Sec. B of Ref. 6.

An interesting possibility, which has not been encountered so far, arises when

$$b<0,\quad g>0,\quad c>0. \tag{6.10}$$

In this case a curvature singularity occurs for

$$\xi\zeta=c>0,\quad H=\sqrt{g},\quad \forall\,\tilde{y}\in\mathbb{R}, \tag{6.11}$$

i.e., in region I (and in region I_0 as well). It is more convenient to use the form (5.4) of the metric in which the singularity occurs for

$$E^2=c,\quad H^2=g,\quad \forall\,x\in\mathbb{R},\quad \forall\,y\in\mathbb{R}. \tag{6.12}$$

Obviously, we get a *two-dimensional spacelike curvature singularity* in region I, which, of course, will be missed practically by every observer.

B. The surface $H=0$

For the metric (5.18) we find that

$$g^{\xi\xi}=\frac{2aE^2\xi^2-\alpha^2(E^2+\xi^2)}{4E^2(t^2+z^2)},$$
$$g^{\zeta\zeta}=\frac{2aE^2\zeta^2-\alpha^2(E^2+\xi^2)}{4E^2(t^2+z^2)},$$
$$g^{\xi\zeta}=\frac{(\alpha^2+2aE^2)\xi\zeta}{4E^2(t^2+z^2)}.$$

The vector field orthogonal to the surfaces $H=\sqrt{\xi^2+\zeta^2}=$ const is

$$\nabla_a H=(1/H)(\xi\nabla_a\xi+\zeta\nabla_a\zeta)$$

with norm

$$(\nabla H)^2=[1/(\xi^2+\zeta^2)][\xi^2(\nabla\xi)^2+\zeta^2(\nabla\zeta)^2+2\xi\zeta(\nabla\xi)(\nabla\zeta)]$$
$$=[1/(\xi^2+\zeta^2)][\xi^2g^{\xi\xi}+\zeta^2g^{\zeta\zeta}+2\xi\zeta g^{\xi\zeta}]$$

or

$$(\nabla H)^2=-(\alpha^2-2aH^2)/4(t^2+z^2).$$

Obviously $(\nabla H)^2<0$ for $H=0$, and $\xi^2+\zeta^2=0$ is a two-dimensional spacelike regular surface of the space-time. As was explained in Sec. III, there is no need to describe the extension of the space-time beyond this surface. Our conclusion merely shows that $H=0$ is a regular surface of the space-time and that there is no need for worry.

VII. CONCLUDING REMARKS

Many solutions have been presented so far, describing the collision of plane-fronted gravitational and/or electromagnetic waves. And quite a few of them do develop Cauchy horizons. What is novel about the solution investigated in the present paper is that, for a particular range of the free parameters, *no curvature singularities occur in the extended space-time*!

Shall we draw some conclusion about the connection between solutions of Petrov type D and solutions that develop, as a result of the collision, Cauchy horizons? We think that the evidence accumulated so far could very well be considered circumstantial. For solutions of the Einstein–Maxwell equations with nonhypersurface orthogonal Killing fields, those of Petrov type D constitute one of the largest known families and this is why we decided to investigate them.

[1]K. Khan and R. Penrose, Nature **229**, 185 (1971).
[2]P. Szekeres, J. Math. Phys. **13**, 286 (1972).
[3]Y. Nutku and M. Halil, Phys. Rev. Lett. **39**, 1379 (1977).
[4]S. Chandrasekhar and V. Ferrari, Proc. R. Soc. London Ser. A **396**, 55 (1984).
[5]S. Chandrasekhar and B. C. Xanthopoulos, Proc. R. Soc. London Ser. A **398**, 223 (1985).
[6]S. Chandrasekhar and B. C. Xanthopoulos, Proc. R. Soc. London Ser. A **408**, 175 (1986).
[7]S. Chandrasekhar and B. C. Xanthopoulos, Proc. R. Soc. London Ser. A **410**, 311 (1987).
[8]V. Ferrari and J. Ibañez, Gen. Relativ. Gravit. **19**, 405 (1987); V. Ferrari, J. Ibañez, and M. Bruni, Phys. Rev. D **36**, 1053 (1987); V. Ferrari and J. Ibañez, Proc. R. Soc. London A **417**, 417 (1988).
[9]U. Yurtsever, Phys. Rev. D **36**, 1662 (1987); Phys. Rev. D **37**, 2803 (1988); **38**, 1706, 1731 (1988); CALTECH preprint GRP-184, 1988.
[10]S. Chadrasekhar and B. C. Xanthopoulos, Proc. R. Soc. London A **420**, 93 (1988).
[11]R. Debever, Bull. Soc. Math. Belg. **23**, 360 (1971).
[12]J. Plebanski, Ann. Phys. (NY) **90**, 196 (1975).
[13]J. Plebanski and M. Demianski, Ann. Phys. (NY) **98**, 9 (1976).
[14]D. Kramer, H. Stephani, E. Herlt, and M. Mac Callum, *Exact Solutions of Einstein's Field Equations* (Cambridge U.P., Cambridge, 1980), Sec. 19.1.
[15]R. Debever and R. G. MacLenaghan, J. Math. Phys. **22**, 1771 (1981).
[16]R. Debever, N. Kamran, and R. G. MacLenaghan, J. Math. Phys. **25**, 1955 (1984).
[17]B. Carter, Commun. Math. Phys. **10**, 280 (1968).
[18]S. Chandrasekhar and B. C. Xanthopoulos, Proc. R. Soc. London Ser. A **414**, 1 (1987).

Curvatures, Gauss–Bonnet and Chern–Simons multiplets in old-minimal $N=1$ supergravity

S. Ferrara
CERN, Geneva, Switzerland and Department of Physics, University of California, Los Angeles, California 90024

M. Villasante
Department of Physics, University of California, Los Angeles, California 90024

(Received 31 May 1988; accepted for publication 10 August 1988)

The complete form of the three curvature multiplets of old-minimal supergravity are given. The full super-Gauss–Bonnet theorem is derived and the Chern–Simons multiplets constructed. The latter construction allows a coupling to antisymmetric tensor fields that is the supersymmetric generalization of the Green–Schwarz mechanism occurring in superstring theories.

I. INTRODUCTION

Gravity theories with higher curvature terms[1] have recently received much attention in connection with the point-field limit of string theories. In particular, the Green–Schwarz[2] mechanism needed for anomaly cancellations in higher-dimensional theories implies specific couplings of Chern–Simons forms to antisymmetric (second-rank) tensor fields. In superstrings these properties must have a supersymmetric extension that leads to a generalization of supergravity theories including higher-order superinvariants.[3] The local supersymmetric completion of higher curvature theories requires a complete knowledge of the supercurvature multiplets[4] as well as the super-Chern–Simons ones.[5,6] The superspace form of these multiplets[7] has been found recently, but an explicit component determination has not yet been completed. Partial results including the full linearized expansion and the full nonlinear bosonic terms have been given in Refs. 3 and 4. Here we complete those results and give the complete proof of the super-Gauss–Bonnet theorem for old-minimal Poincaré supergravity. The full Chern–Simons multiplet is also obtained and its coupling to antisymmetric tensor fields via the Green–Schwartz mechanism described.

In dealing with supergravity theories with higher curvature terms, an important role is played by the so-called auxiliary fields of the gravity multiplets.[8] Different auxiliary field formulations correspond to different "compensators" of conformal supergravity. In canonical supergravity, when only terms linear in the curvatures are present, it is known that all these formulations are equivalent[9] in the sense that they can be transformed into each other via duality transformations or by adding extra multiplets.

In the minimal formulations, the auxiliary fields are dependent on a chiral and a linear multiplet, respectively.

For simple supergravity, without higher curvature terms, the two formulations are dual of each other. In fact, a local transformation can be performed that transforms a chiral into a linear multiplet and vice versa.

When higher curvature terms are present, as is the case in superstring theories, this is no longer true, due to the occurrence of derivatives on the compensators.[10] An explicit example of a Lagrangian quadratic in the curvatures was given that revealed this inequivalence.

In the present paper we confine our investigation to Poincaré supergravity in the so-called old minimal formulation. Results for the new minimal formulation are treated elsewhere.[11]

Our results are relevant in connection with a superspace formulation of background fields in superstrings with $N=1$ space-time supersymmetry in $D=4$ dimensions. Heterotic strings seem to have a natural formulation in the new-minimal formulation.[12] This does not mean that an old minimal formulation is impossible.[13] In fact, even in the presence of higher derivative terms, the two formulations may be transformed, in general, into one another provided a sufficient number of new degrees of freedom are introduced. The question is which formulation is more economical in terms of tensor fields. In view of the different possible formulations of superstring theories, we believe that a complete knowledge of the structure of both formulations is useful.

The plan of the paper is as follows. In Sec. II, we review some basics of the Poincaré calculus with old-minimal formulation, especially some properties of the supercovariant derivatives. In Sec. III, the curvature multiplets are explicitly displayed. In Sec. IV, the complete proof of the super-Gauss–Bonnet theorem is given. In Sec. V, we give the Lorentz Chern–Simons and Gauss–Bonnet multiplets and their couplings to an antisymmetric tensor.

II. ELEMENTS OF TENSOR CALCULUS

We will follow the conventions of Ref. 6. The transformation rules of the supergravity multiplet are given by[14]

$$\begin{aligned}
\delta e^a{}_\mu &= \tfrac{1}{2}\bar{\epsilon}\gamma^a \psi_\mu ,\\
\delta\psi_\mu &= (D_\mu + (i/2)\gamma_5 A_\mu)\epsilon - \tfrac{1}{2}\gamma_\mu \eta\epsilon ,\\
\delta S &= -\tfrac{1}{2}\bar{\epsilon}\gamma\cdot z,\\
\delta P &= (i/2)\bar{\epsilon}\gamma_5\gamma\cdot z,\\
\delta A_a &= (i/2)\bar{\epsilon}\gamma_5 z_a ,
\end{aligned} \tag{2.1}$$

where

$$\begin{aligned}
\eta &= -\tfrac{1}{3}(S - i\gamma_5 P - i\not{A}\gamma_5) ,\\
z_a &= \tfrac{3}{2}\left(R_a^{\,(\mathrm{cov})} - \tfrac{1}{3}\gamma_a\gamma\cdot R^{\,(\mathrm{cov})}\right) ,
\end{aligned} \tag{2.2}$$

and $R_a^{\,(\mathrm{cov})}$ is the supercovariant Rarita–Schwinger field strength

 0022-2488/89/010104-11$02.50

$$R^{(\mathrm{cov})a}=\tfrac{1}{2}\Gamma^{abc}\,\psi^{(\mathrm{cov})}_{bc}=\tfrac{1}{2}\epsilon^{abcd}\,\gamma_5\gamma_d\,\psi^{(\mathrm{cov})}_{bc}\,,$$
$$\psi^{(\mathrm{cov})}_{bc}=e_b{}^{\mu}\,e_c{}^{\nu}\,\psi^{(\mathrm{cov})}_{\mu\nu}\,, \tag{2.3}$$
$$\psi^{(\mathrm{cov})}_{\mu\nu}=2(D_{[\mu}+(i/2)A_{[\mu}\gamma_5-\tfrac{1}{2}\gamma_{[\mu}\eta)\psi_{\nu]}\,.$$

The tangent space covariant derivative is

$$D_\mu\psi_\nu=(\partial_\mu+\tfrac{1}{4}\omega_\mu{}^{ab}\,\Gamma_{ab})\psi_\nu \tag{2.4}$$

as usual. The spin connection $\omega_\mu{}^{ab}=\omega_\mu{}^{ab}(e,\psi)$ is given by the usual expression[14] and its transformation rule is[4,14,15]

$$\begin{aligned}\delta\omega_{\mu ab}&=-\tfrac{1}{4}(\bar\epsilon\gamma_\mu\,\psi^{(\mathrm{cov})}_{ab}-2\bar\epsilon\gamma_{[a}\,\psi^{(\mathrm{cov})}_{b]\mu})+\tfrac{1}{4}\bar\epsilon\{\Gamma_{ab},\eta\}\psi_\mu\\&=\tfrac{1}{2}\bar\epsilon(\Gamma_{[ab}\,z_{\mu]}-\gamma_\mu\,\psi^{(\mathrm{cov})}_{ab})+\tfrac{1}{4}\bar\epsilon\{\Gamma_{ab},\eta\}\psi_\mu\,.\end{aligned} \tag{2.5}$$

The Lagrangian for the supergravity multiplet that is invariant under the rules (2.1) is

$$\mathscr{L}_{\mathrm{sg}}=-\tfrac{1}{2}eR-\tfrac{1}{2}e\bar\psi_\mu\,R^{\mu}-\tfrac{1}{3}e(S^2+P^2-A_a\,A^a)\,, \tag{2.6}$$

where R is the scalar curvature and R^{μ} is the Rarita–Schwinger field strength:

$$R=g^{\mu\nu}R_{\mu\nu},\quad R_{\mu\nu}=R_{\mu\lambda}{}^{ab}\,e_a{}^{\lambda}\,e_{b\nu}\,,$$
$$R^{\mu}=\tfrac{1}{2}\,\Gamma^{\mu\nu\rho}\,\psi_{\nu\rho},\quad \psi_{\nu\rho}=2D_{[\nu}\,\psi_{\rho]}\,.$$

Now let us turn to the matter multiplets. A scalar multiplet $[A,B,\chi',F,G]$ consists of scalar (A) and pseudoscalar (B) physical fields a Majorana fermion χ', and scalar (F) and pseudoscalar (G) auxiliary fields with the transformation rules

$$\begin{aligned}&\delta A=\tfrac{1}{2}\bar\epsilon\chi',\\&\delta B=-(i/2)\bar\epsilon\gamma_5\chi',\\&\delta\chi'=\tfrac{1}{2}\hat{\not{D}}(A-i\gamma_5B)\epsilon+\tfrac{1}{2}(F+i\gamma_5G)\epsilon\,,\\&\delta F=\tfrac{1}{2}\bar\epsilon(\hat{\not{D}}-(i/2)\not{A}\gamma_5)\chi'+\tfrac{1}{2}\bar\epsilon\eta\chi'\,,\\&\delta G=(i/2)\bar\epsilon\gamma_5(\hat{\not{D}}-(i/2)\not{A}\gamma_5)\chi'-(i/2)\bar\epsilon\eta\gamma_5\chi'\,.\end{aligned} \tag{2.7}$$

The super-Poincaré covariant derivative $\hat{D}_\mu$ on any field with only tangent space indices is defined as

$$\hat{D}_\mu\phi\cdots=D_\mu\phi\cdots-\delta(\psi_\mu)\phi\cdots\,, \tag{2.8}$$

the first term being the tangent space covariant derivative and the second the subtraction of a supersymmetry transformation with parameter ψ_μ. The chiral multiplet $[z,\chi_L,h]$ is the complex combination,

$$z=i(A+iB),\quad \chi_L=(i/2)(I+\gamma_5)\chi',\quad h=i(F+iG)\,, \tag{2.9}$$

while the complex conjugate combination is an antichiral multiplet.

In the derivation of the components and transformation rules of the Ricci multiplet in Sec. III, we will need the supersymmetry transformations of the supercovariant derivative of these fields, and they are

$$\begin{aligned}\delta\hat{D}_aA&=\tfrac{1}{2}\bar\epsilon[\hat{D}_a-(i/2)A_a\gamma_5-\tfrac{1}{2}\eta\gamma_a]\chi'\,,\\\delta\hat{D}_aB&=-(i/2)\bar\epsilon\gamma_5(\hat{D}_a-(i/2)A_a\gamma_5)\chi'+(i/4)\bar\epsilon\eta\gamma_a\gamma_5\chi'\,,\\\delta\hat{D}_aF&=\tfrac{1}{2}\bar\epsilon(\hat{D}_a-(i/2)A_a\,\gamma_5-\tfrac{1}{2}\eta\gamma_a)\\&\quad\times(\hat{\not{D}}-(i/2)\not{A}\gamma_5+\eta)\chi'\,,\\\delta\hat{D}_aG&=(i/2)\bar\epsilon(\hat{D}_a-(i/2)A_a\,\gamma_5\\&\quad-\tfrac{1}{2}\eta\gamma_a)[\gamma_5(\hat{\not{D}}-(i/2)\not{A}\gamma_5)-\eta\gamma_5]\chi'\,,\end{aligned}$$

$$\begin{aligned}\delta\hat{D}_a\chi'&=\tfrac{1}{2}\{\hat{D}_a[\hat{\not{D}}(A-i\gamma_5B)+F+i\gamma_5G]\}\epsilon\\&\quad-\tfrac{1}{4}[\hat{\not{D}}(A-i\gamma_5B)+F+i\gamma_5G](iA_a\gamma_5-\gamma_a\eta)\epsilon\\&\quad+\tfrac{1}{8}\bar\epsilon(\Gamma_{[cd}\,z_{a]}-\gamma_a\,\psi^{(\mathrm{cov})}_{cd})\Gamma^{cd}\chi'\,.\end{aligned} \tag{2.10}$$

A locally supersymmetric Lagrangian for a chiral multiplet $[z,\chi_L,h]$ is given by

$$e^{-1}\mathscr{L}=h+\bar\psi_R\cdot\gamma\chi_L+\tfrac{1}{2}\bar\psi_{\mu R}\,\Gamma^{\mu\nu}\,\psi_{\nu R}z+uz \tag{2.11}$$

with

$$u=S-iP\,. \tag{2.12}$$

Density (2.11) is complex, thus usually one takes the real part, but we will use it as it stands since the imaginary part is important in providing the supersymmetric generalization of the Chern–Simons form. The latter in turn plays a crucial role in the Green–Schwartz mechanism, when formula (2.11) is applied to the square of the Weyl multiplet.[16] This will become transparent in Sec. V.

The second basic multiplet is the vector multiplet $[C,Z,H,K,B_a,\Lambda,D]$, where Z and Λ are Majorana fermions; C, H, K, and D are scalar or pseudoscalar fields and B_a is a vector one. Its transformation laws are[4,17]

$$\begin{aligned}&\delta C=(i/2)\bar\epsilon\gamma_5Z,\\&\delta Z=\tfrac{1}{2}(Hi\gamma_5-K-\not{B}+\hat{\not{D}}Ci\gamma_5)\epsilon\,,\\&\delta H=(i/2)\bar\epsilon\gamma_5(\hat{\not{D}}-(i/2)\not{A}\gamma_5)Z\\&\qquad+(i/2)\bar\epsilon\gamma_5\Lambda-(i/2)\bar\epsilon\eta\gamma_5Z\,,\\&\delta K=-\tfrac{1}{2}\bar\epsilon(\hat{\not{D}}-(i/2)\not{A}\gamma_5)Z-\tfrac{1}{2}\bar\epsilon\Lambda-\tfrac{1}{2}\bar\epsilon\eta Z\,,\\&\delta B_a=-\tfrac{1}{2}\bar\epsilon(\hat{D}_a-(i/2)A_a\gamma_5)Z-\tfrac{1}{2}\bar\epsilon\gamma_a\Lambda+\tfrac{1}{4}\bar\epsilon\eta\gamma_aZ\,,\\&\delta\Lambda=\tfrac{1}{4}[\Gamma^{ab}\,Y_{ab}+2i\gamma_5D]\epsilon\,,\\&\delta D=(i/2)\bar\epsilon\gamma_5(\hat{\not{D}}+(i/2)\not{A}\gamma_5)\Lambda\,,\end{aligned} \tag{2.13}$$

where

$$Y_{ab}=2\hat{D}_{[a}\,B_{b]}-\tfrac{1}{2}\bar{Z}\psi^{(\mathrm{cov})}_{ab}\,. \tag{2.14}$$

Let us also give the following transformation rules of covariant derivatives:

$$\begin{aligned}\delta\hat{D}_aC&=(i/2)\bar\epsilon\gamma_5(\hat{D}_a-(i/2)A_a\gamma_5)Z-(i/4)\bar\epsilon\eta\gamma_a\gamma_5Z\,,\\\delta\hat{D}_aZ&=\tfrac{1}{2}[\hat{D}_a(Hi\gamma_5-K-\not{B}+\hat{\not{D}}Ci\gamma_5)]\epsilon\\&\quad-\tfrac{1}{4}(Hi\gamma_5-K-\not{B}+\hat{\not{D}}Ci\gamma_5)(iA_a\gamma_5-\gamma_a\eta)\epsilon\\&\quad+\tfrac{1}{8}\bar\epsilon(\Gamma_{[cd}\,z_{a]}-\gamma_a\,\psi^{(\mathrm{cov})}_{cd})\Gamma^{cd}\,Z\,,\end{aligned} \tag{2.15}$$

where we left the last term in $\delta\hat{D}_aZ$ "unfierzed" for simplicity.

A locally supersymmetric Lagrangian for the vector multiplet is provided by the D density[4,17]

$$\begin{aligned}e^{-1}\mathscr{L}_D&=D-(i/2)\bar\psi\cdot\gamma\gamma_5\Lambda-\tfrac{2}{3}(HS-KP)+\tfrac{2}{3}A^a\,B_a\\&\quad-(i/3)\bar{Z}\gamma_5\gamma\cdot R+(i/4)e^{-1}\,\epsilon^{\mu\nu\rho\sigma}\,\bar\psi_\mu\,\gamma_\nu\,\psi_\rho\,B_\sigma\\&\quad-(i/8)e^{-1}\,\epsilon^{\mu\nu\rho\sigma}\,\bar\psi_\mu\,\gamma_\nu\,\psi_\rho\,\bar{Z}\,\psi_\sigma\\&\quad-\tfrac{2}{3}Ce^{-1}\,\mathscr{L}_{\mathrm{sg}}\,.\end{aligned} \tag{2.16}$$

A real function $\phi(\Sigma_i,\Sigma^{*j})$ of a set of chiral multiplets $\Sigma_i=[z_i,\chi_{iL},h_i]$ and their complex conjugates will be a vector multiplet with components

$$\phi(\Sigma_i,\Sigma^{*j})$$
$$= [\phi(z_i,z^{*j}), -2i(\phi^j \chi_{jL} + \phi_j \chi_R{}^j), -\phi^j h_j - \phi_j h^{*j} + \phi^{jk}\overline{\chi}_{jL}\chi_{kL} + \phi_{jk}\overline{\chi}_R{}^j\chi_R{}^k,$$
$$i\phi^j h_j - i\phi_j h^{*j} - i\phi^{jk}\overline{\chi}_{jL}\chi_{kL} + i\phi_{jk}\overline{\chi}_R{}^j\chi_R{}^k, i\phi^j \widehat{D}_\mu z_j - i\phi_j \widehat{D}_\mu z^{*j} + 2i\phi^j{}_k \overline{\chi}_{jL}\gamma_\mu \chi_R{}^k,$$
$$-2i\phi^j{}_k\{(h_j + \widehat{\not{D}}z_j)\chi_R{}^k - (h^{*k} + \widehat{\not{D}}z^{*k})\chi_{jL}\} + 2i\phi^{jk}{}_l\overline{\chi}_{jL}\chi_{kL}\chi_R{}^l - 2i\phi^l{}_{jk}\overline{\chi}_R{}^j\chi_R{}^k\chi_{lL},$$
$$2\phi^j{}_k\{h_j h^{*k} - \widehat{D}_\mu z_j \widehat{D}^\mu z^{*k} - (\overline{\chi}_{jL}\widehat{\not{D}}\chi_R{}^k + \overline{\chi}_R{}^k\widehat{\not{D}}\chi_{jL})\}$$
$$-2\phi^{jk}{}_l(h^{*l}\overline{\chi}_{jL}\chi_{kL} + \overline{\chi}_R{}^l(\widehat{\not{D}}z_j)\chi_{kL}) - 2\phi^l{}_{jk}(h_l\overline{\chi}_R{}^j\chi_R{}^k + \overline{\chi}_{lL}(\widehat{\not{D}}z^{*j})\chi_R{}^k) + 2\phi^{ij}{}_{kl}\overline{\chi}_{iL}\chi_{jL}\overline{\chi}_R{}^k\chi_R{}^l]\,. \quad (2.17)$$

One can obtain a chiral multiplet from a vector multiplet by chiral projection. This operation is used to define the so-called linear multiplet, which is the last basic type of multiplet in supergravity. We will postpone these matters until Sec. V, however, so that we will have a chance to introduce the scalar curvature multiplet that plays a significant role in all this.

III. THE CURVATURE MULTIPLETS

There are three basic curvature multiplets: the scalar curvature multiplet R,[18] the Weyl multiplet W_{ab},[16] and the Ricci multiplet W_a.[4] The first two are scalar multiplets, which we will present in their chiral form, and the last one is a vector multiplet.

Let us start with the scalar curvature multiplet R, whose components are

$$R = [S,P,-\gamma\cdot z,B,-\widehat{D}_a A^a]\,, \quad (3.1a)$$

or in chiral form

$$R = [u^*, -i\gamma\cdot z_R, B - i\widehat{D}_a A^a]\,, \quad (3.1b)$$

where u is given in (2.12) and

$$B = \eta^{ab}B_{ab} = \tfrac{1}{2}R^{(\mathrm{cov})} - \tfrac{2}{3}(|u|^2 + \tfrac{1}{4}A_a A^a)\,. \quad (3.2)$$

The symbol B_{ab} will be one of the components of the Ricci multiplet. The supercovariantized scalar curvature is

$$R^{(\mathrm{cov})} = R - \overline{\psi}^a\gamma^b\psi^{(\mathrm{cov})}_{ab} + \tfrac{1}{2}\overline{\psi}_a\Gamma^{ab}\eta\psi_b\,. \quad (3.3)$$

The product of R with its complex conjugate $\overline{R}$ is a vector multiplet whose components we can obtain by applying the formula (2.17) to (3.1),

$$R\overline{R} = [|u|^2, 2i\gamma_5(S - i\gamma_5 P)\gamma\cdot z, -2SB + 2P\widehat{D}_a A^a, 2PB$$
$$+ 2S\widehat{D}_a A^a, iu\overleftrightarrow{\widehat{D}}_b u^* - i\overline{z}\cdot\gamma\gamma_5\gamma_b\gamma\cdot z,$$
$$-2i\gamma_5\{B + i\gamma_5\widehat{D}_a A^a - \widehat{\not{D}}(S - i\gamma_5 P)\}\gamma\cdot z,$$
$$2\{B^2 + (\widehat{D}_a A^a)^2 - \widehat{D}_b u\widehat{D}^b u^*$$
$$+ \overline{z}\cdot\gamma(\widehat{\not{D}} + (i/2)\gamma_5\not{A})\gamma\cdot z\}]\,. \quad (3.4)$$

Next we turn to the Weyl multiplet $W_{ab}{}^\alpha$, which has two external vector indices a,b (it is antisymmetric in them) and one external spinor index. Its components are

$$W_{ab} = [R_{abL}(Q), \tfrac{1}{4}(1+\gamma_5)(\Gamma^{cd}W_{abcd}$$
$$- 3iR^{cd}(A)T_{cdab})C^{-1}, Z_{abL}]\,, \quad (3.5)$$

where C^{-1} is the inverse of the charge conjugation matrix and

$$Z_{ab} = 2T_{abcd}R^{(\mathrm{cov})cd}(S) - \tfrac{1}{2}(S - i\gamma_5 P)R_{ab}(Q)\,. \quad (3.6)$$

Here $R_{ab}(Q)$ and $R^{(\mathrm{cov})}_{ab}(S)$ are the curvatures of conformal supergravity [they correspond to what have been usually called $\overline{R}_{ab}(Q)$ and $\overline{R}^{(\mathrm{cov})}_{ab}(S)$]. The covariant S curvature $R^{(\mathrm{cov})}_{ab}(S)$ is

$$R^{(\mathrm{cov})}_{ab}(S) = R_{ab}(S) + (i/2)(\gamma_5 R_{c[a}(A) + \tilde{R}_{c[a}(A))\gamma^c\psi_{b]}\,, \quad (3.7)$$

while $R_{ab}(A)$ is the conformal A curvature. The Weyl tensor W_{abcd} is the traceless part of the covariant M curvature $R^{(\mathrm{cov})}_{ab}(M_{cd})$. Since that removes the dependence on $f^a{}_\mu$ and b_μ, W_{abcd} is actually the traceless part of

$$\widehat{R}^{(\mathrm{cov})ab}_{\mu\nu} = \widehat{R}_{\mu\nu}{}^{ab} - \overline{\psi}_{[\mu}\gamma_{\nu]}R^{ab}(Q) - \tfrac{2}{3}e^{[a}{}_{[\mu}\overline{\psi}_{\nu]}z^{b]}\,,$$
$$\widehat{R}_{\mu\nu}{}^{ab} = R_{\mu\nu}{}^{ab} - \tfrac{1}{3}\overline{\psi}_{[\mu}\Gamma^{ab}z_{\nu]} + \tfrac{1}{6}\overline{\psi}_\mu\Gamma^{ab}(S - i\gamma_5 P)\psi_\nu$$
$$+ (i/6)\epsilon^{abcd}A_c\overline{\psi}_\mu\gamma_d\psi_\nu\,, \quad (3.8)$$

and $R_{\mu\nu}{}^{ab}$ is the Riemann tensor for $\omega_\mu{}^{ab}(e,\psi)$. [There is a misprint in the last term of $\widehat{R}_{\mu\nu ab}$ in Eq. (4.10) of Ref. 4.]

Finally, the matrix projector T_{abcd} is given by[4,16]

$$T_{ab}{}^{cd} = \tfrac{1}{3}(\delta_{[a}{}^c\delta_{b]}{}^d - \tfrac{1}{2}\Gamma_{ab}{}^{cd} - \delta_{[a}{}^{[c}\Gamma_{b]}{}^{d]})$$
$$(\Gamma_{abcd} = \epsilon_{abcd}\gamma_5)\,, \quad (3.9)$$

which has the following properties:

$$T_{abcd}\gamma^d = \gamma^a T_{abcd} = 0\,,$$
$$T_{abcd}\gamma_5 = \gamma_5 T_{abcd} = -\tfrac{1}{2}\epsilon_{abef}T^{ef}{}_{cd}\,,$$
$$T_{abef}T^{ef}{}_{cd} = T_{abcd}, \quad T_{abef}T_{cd}{}^{ef} = \tfrac{1}{3}T_{abcd}\,, \quad (3.10)$$
$$CT_{abcd}C^{-1} = T_{abcd}{}^T\,,$$
$$(CT_{abcd})_{\alpha\beta} = -(CT_{cdab})_{\beta\alpha}\,,$$

where C is the charge conjugation matrix.

The Weyl multiplet satisfies $W_{ab} = T_{abcd}W^{cd}$ (which implies $\gamma^b W_{ab} = 0$ among other things) and is independent of the conformal dilation field b_μ. As a matter of fact, one has

$$R_{ab}(Q) = -\psi^{(\mathrm{cov})}_{ab} + \tfrac{2}{3}\gamma_{[a}z_{b]}\,. \quad (3.11)$$

Clearly (3.11) does not depend on b_μ, nor does $R_{ab}(A)$ or W_{abcd} [see Eq. (3.8)]. For more details, we refer the reader to Appendix A.

The square of the Weyl multiplet $\overline{W}_{ab}W^{ab}$ is also a chiral multiplet with no external indices and its components are[6,16]

$$\overline{W}_{ab}W^{ab}$$
$$= [\overline{R}_{ab}(Q)R_L{}^{ab}(Q), \tfrac{1}{2}(\Gamma^{cd}W_{abcd} - 3iR_{ab}(A))R_L{}^{ab}(Q),$$
$$4\overline{R}_L{}^{ab}(Q)R^{(\mathrm{cov})}_{ab}(S) + \tfrac{1}{2}W_{abcd}W^{abcd}$$

$$- \tfrac14 \epsilon^{abcd} W_{\mu\nu ab} W^{\mu\nu}{}_{cd} - \tfrac34 R_{ab}(A) R^{ab}(A) + \tfrac34 R_{ab}(A)\widetilde{R}^{ab}(A) - u\overline{R}_{ab}(Q) R_L{}^{ab}(Q) \,\big] \,. \quad (3.12)$$

If we put $u = 0$ in the last component, we get the conformal $\overline{W}_{ab} W^{ab}$ multiplet.

The third basic curvature multiplet in old-minimal Poincaré supergravity is the Ricci multiplet $W_a = [C_a, Z_a, H_a, K_a, B_{ab}, \lambda_a, d_a]$ given by

$$\begin{aligned}
&C_a = A_a, \quad Z_a = z_a, \\
&H_a = \widehat{D}_a P + A_a S, \quad K_a = \widehat{D}_a S - A_a P, \\
&B_{ab} = \tfrac12 (\widehat{R}_{ab} - \tfrac16 \eta_{ab} \widehat{R}\,) - \tfrac34 \bar\psi^c \gamma_a R_{bc}(Q) + (i/2)\widetilde{\widehat{F}}_{ab} \\
&\qquad - \tfrac16 (|u|^2 + A_c A^c)\eta_{ab} + \tfrac13 A_a A_b + \tfrac12 \bar\psi_a z_b \,, \\
&\lambda_a = 2\gamma^b (\widehat{D}_{[a} - (i/2)\gamma_5 A_{[a})z_{b]} + \tfrac12 (S + i\gamma_5 P) z_a \\
&\qquad + (i/2)\gamma_5 A\!\!\!/ z_a + (i/6)\gamma_5 A\!\!\!/ \Gamma_{ac} z^c \,,
\end{aligned}$$

$$\begin{aligned}
d_a = {}& \widehat{D}_a \widehat{D}_b A^b - \widehat{\Box} A_a - \tfrac23 B_{ac} A^c - \tfrac13 B A_a \\
& - \tfrac16 A_a |u|^2 - (i/12)\bar z^b \gamma_5 \Gamma_{abc} z^c + (i/6)\, \bar z_a \gamma_5 \gamma\cdot z \\
& - (i/6)\bar z^c \gamma_5 \gamma_a z_c + i\bar z^c \gamma_5 \psi_{ac}^{(\mathrm{cov})} \,,
\end{aligned} \quad (3.13)$$

where

$$\begin{aligned}
&\widehat{R}_{\mu\nu} = \widehat{R}_{\mu\lambda}{}^{ab} e_a{}^\lambda e_{b\nu} \,, \quad \widehat{R} = g^{\mu\nu}\widehat{R}_{\mu\nu} \,, \\
&\widehat{F}_{ab} = 2\widehat{D}_{[a} A_{b]} \,, \quad \widehat{\Box} A_a = \widehat{D}_b \widehat{D}^b A_a \,,
\end{aligned} \quad (3.14)$$

and $\widehat{R}_{\mu\lambda}{}^{ab}$ has been given in (3.8).

With the results of Appendix B it is easy to see that the antisymmetric part of B_{ab} is simply

$$B_{[ab]} = (i/2)\widetilde{\widehat{F}}_{ab} \quad (3.15)$$

and that $\widehat{F}_{ab}$ is related to the conformal A curvature,

$$\widehat{F}_{ab} = \tfrac32 R_{ab}(A). \quad (3.16)$$

The transformation rule of A_a has been given in (2.1) which indeed has the form dictated by (2.13). As already pointed out in Ref. 4, this is not true for the remaining components because of the external index. One gets

$$\begin{aligned}
\delta Z_a = {}& \tfrac12 (H_a i\gamma_5 - K_a - B_{ab}\gamma^b + i\widehat{D\!\!\!/} A_a \gamma_5)\epsilon - (i/6)\Gamma_{ab}\gamma_5 (S - i\gamma_5 P) A^b \epsilon \,, \\
\delta H_a = {}& (i/2)\bar\epsilon\gamma_5 (\widehat{D\!\!\!/} - (i/2)A\!\!\!/\gamma_5) z_a + (i/2)\bar\epsilon\gamma_5\lambda_a - (i/2)\bar\epsilon\eta\gamma_5 z_a - (i/12)\bar\epsilon(S - i\gamma_5 P)\gamma_5 \Gamma_{ac} z^c \,, \\
\delta K_a = {}& -\tfrac12 \bar\epsilon(\widehat{D\!\!\!/} - (i/2)A\!\!\!/\gamma_5) z_a - \tfrac12 \bar\epsilon\lambda_a - \tfrac12 \bar\epsilon\eta z_a - \tfrac1{12}\bar\epsilon(S - i\gamma_5 P)\Gamma_{ac} z^c \,, \\
\delta B_{ab} = {}& -\tfrac12 \bar\epsilon(\widehat{D}_b - (i/2)A_b\gamma_5) z_a - \tfrac12 \bar\epsilon\gamma_b\lambda_a + \tfrac14 \bar\epsilon\eta\gamma_b z_a + \tfrac1{12}\bar\epsilon(S - i\gamma_5 P)\Gamma_{ca}\gamma_b z^c \\
& + (i/6)\bar\epsilon\gamma_5 (A_b z_a - A\cdot z\eta_{ab}) - (i/2)\bar\epsilon\gamma_b\gamma_5 A^c \psi_{ac}^{(\mathrm{cov})} - (i/6)\bar\epsilon\gamma_5 A^c \Gamma_{ca} z_b \,, \\
\delta\lambda_a = {}& \tfrac12 \Gamma^{bc}\big[\widehat{D}_b B_{ac} - \tfrac14 \bar z_a \psi_{bc}^{(\mathrm{cov})} + \tfrac16 \bar z\cdot\gamma\gamma_5 z_c\gamma_5\eta_{ab} - \tfrac1{12}\bar z_b \gamma_a z_c \\
& - \tfrac16 (S - i\gamma_5 P)\widehat{D}_c (S + i\gamma_5 P) - (i/2)\gamma_5 (\widehat{R}\,{}^{(\mathrm{cov})}_{bcad} A^d + \tfrac13 |u|^2 \eta_{ab} A_c)\big]\epsilon + \tfrac16 A^c \widehat{F}_{ac}\epsilon + (i/2)\gamma_5 d_a \epsilon \\
& + \tfrac12 \{ (i/3)\big[\widetilde{\widehat{F}}_{ab} - \widehat{F}_{ab}\gamma_5 - \gamma_5 (A_b \delta_a{}^c - \eta_{ab} A^c)\widehat{D}_c - (2i/3)(A_a A_b - A^c A_c \eta_{ab}) \\
& + (i/3)(\delta_a{}^c \delta_b{}^d - \eta_{ab}\eta^{cd})(B_{cd} + i\gamma_5 \widehat{D}_d A_c)\big](S + i\gamma_5 P) + \tfrac12 \bar z^c \gamma_b \psi_{ac}^{(\mathrm{cov})} - \tfrac12 \bar z^c \gamma_5 \gamma_b \psi_{ac}^{(\mathrm{cov})}\gamma_5 \\
& - \tfrac1{12}\bar z^c \Gamma_{cb} z_a + \tfrac1{12}\bar z^c \Gamma_{cb}\gamma_5 z_a \gamma_5 + \tfrac1{12}\big[(\bar z_a z_b - \bar z_a \gamma_5 z_b \gamma_5) - \eta_{ab}(\bar z^c z_c - \bar z^c \gamma_5 z_c \gamma_5)\big]\}\gamma^b \epsilon \,, \\
\delta d_a = {}& (i/2)\bar\epsilon\gamma_5 (\widehat{D\!\!\!/} + (i/2)A\!\!\!/\gamma_5)\lambda_a - (i/6)\bar\epsilon(S - i\gamma_5 P)\gamma_5 \Gamma_{ab}\lambda^b - \tfrac16 \bar\epsilon A^c \Gamma_{cab}\lambda^b \\
& + (i/6)\bar\epsilon\gamma_5 (S - i\gamma_5 P)\gamma_{[b} (\widehat{D}^b - (i/2)A^b\gamma_5) z_{a]} + \tfrac13 \bar\epsilon A^b \Gamma_{c[a}(\widehat{D}_{b]} - (i/2)A_{b]}\gamma_5) z^c \\
& + \tfrac23 \bar\epsilon A^b (\widehat{D}_{[a} - (i/2)\gamma_5 A_{[a}) z_{b]} - (i/6)\bar\epsilon\gamma_5 (H_{[a} i\gamma_5 - K_{[a})\gamma_{b]} z^b \\
& + (i/6)\bar\epsilon(B_{(ac)} - B\eta_{ac})\gamma_5 z^c + (i/12)\bar\epsilon(B_{ac} - i\gamma_5 \widehat{D}_c A_a)\gamma_5 \Gamma^{cd} z_d \\
& - (i/12)\bar\epsilon(B_{bc} + i\gamma_5 \widehat{D}_b A_c)\gamma_5 \Gamma_a{}^c z^b - (i/12)\bar\epsilon(B - i\widehat{D}_c A^c \gamma_5)\Gamma_{ab}\gamma_5 z^b \\
& - \tfrac14 \bar\epsilon\gamma^b \gamma_a \widehat{F}_{bc} z^c - \tfrac13 \bar\epsilon\widehat{F}_{b[a}\Gamma^{bc} z_{c]} - (i/24)|u|^2 \bar\epsilon\gamma_5 \gamma_a \gamma\cdot z - \tfrac1{72}\bar\epsilon(S - i\gamma_5 P)\Gamma_{acd} A^c z^d + \tfrac1{24}\bar\epsilon(S - i\gamma_5 P)A\!\!\!/ z_a \\
& - \tfrac1{36}\bar\epsilon(S - i\gamma_5 P)A_{[a}\gamma_{b]} z^b + (i/6)\bar\epsilon\gamma_5 A^b A_{[a} z_{b]} + (i/9)\bar\epsilon\gamma_5 A_{[a}\Gamma_{b]c} A^b z^c \\
& - (i/36)\bar\epsilon\gamma_5 \Gamma_{ad} A^d A\cdot z + (i/6)\bar\epsilon A\!\!\!/\gamma_5 A^c \psi_{ac}^{(\mathrm{cov})} + (i/2)\bar\epsilon(B^{cd}\gamma_d - i\widehat{D\!\!\!/} A^c \gamma_5)\gamma_5 \psi_{ac}^{(\mathrm{cov})} \,,
\end{aligned} \quad (3.17)$$

where $B_{(ac)} = \frac12 (B_{ac} + B_{ca})$ is the symmetric part of B_{ac}.

The derivation of δz_a and δB_{ab} has been explained in Ref. 4. (The result for δB_{ab} is incorrectly reported, however.) The derivation of δH_a and δK_a is simple after using the rules (2.10) as applied to the multiplet in (3.1a). In order to obtain $\delta\lambda_a$ one needs the γ trace of the antisymmetric part of $\delta\widehat{D}_a z_b$ (the more general transformation is given in Appendix C),

$$\begin{aligned}
\gamma^b \delta\widehat{D}_{[a} z_{b]} = {}& \tfrac18 \big[\tfrac12 \widehat{D}_a |u|^2 + \tfrac13 \widehat{D}_a (A_c A^c) + \tfrac43 A_a \widehat{D}_c A^c - A^c \widehat{D}_c A^a - \tfrac56 \bar z_a \gamma\cdot z\big]\,\epsilon \\
& + (i/4)\gamma_5 \big[\widehat{D}_a \widehat{D}_b A^b - \widehat{\Box} A_a + \tfrac12 (P\widehat{D}_a S - S\widehat{D}_a P) - \tfrac14 A_a |u|^2 - \tfrac56 B_{ac} A^c + \tfrac13 B A_a \\
& - (i/8)\bar z^b \gamma_5 \Gamma_{abc} z^c + (7i/12)\bar z_a \gamma_5 \gamma\cdot z - (5i/24)\bar z^c \gamma_5 \gamma_a z_c + i\,\bar z^c \gamma_5 \psi_{ac}^{(\mathrm{cov})}\big]\epsilon \\
& + \tfrac1{12}\gamma^b \big[\{ i(\widehat{F}_{ab}\gamma_5 + \widetilde{\widehat{F}}_{ab}) + 2A_a A_b + B_{ab} - i\widehat{D}_b A_a
\end{aligned}$$

$$+\tfrac{1}{2}(B-i\gamma_5\hat{D}_cA^c)\eta_{ab}\}(S-i\gamma_5P)+A_b(H_a+i\gamma_5K_a)-2A_a(H_b+i\gamma_5K_b)$$
$$-\tfrac{1}{2}\eta_{ab}A^c(H_c+i\gamma_5K_c)-(i/2)\epsilon_{abcd}A^c\hat{D}^d(S-i\gamma_5P)+\tfrac{3}{2}\bar{z}^c\gamma_b\psi^{(\text{cov})}_{ac}$$
$$+\tfrac{3}{2}\bar{z}^c\gamma_5\gamma_b\psi^{(\text{cov})}_{ac}\gamma_5+\tfrac{1}{2}\bar{z}_b\Gamma_{ad}z^d-\tfrac{1}{2}\bar{z}^d\Gamma_{db}z_a-\tfrac{3}{8}\bar{z}_c\Gamma^{cd}z_d\eta_{ab}+\tfrac{5}{4}\bar{z}_bz_a$$
$$-\tfrac{1}{8}\bar{z}^cz_c\eta_{ab}+\tfrac{1}{4}\bar{z}_b\gamma_5\Gamma_{ad}z^d\gamma_5-\tfrac{1}{4}\bar{z}^d\Gamma_{db}\gamma_5z_a\gamma_5-\tfrac{1}{8}\bar{z}_c\gamma_5\Gamma^{cd}z_d\gamma_5\eta_{ab}+\tfrac{3}{4}\bar{z}_b\gamma_5z_a\gamma_5-\tfrac{3}{8}\bar{z}^c\gamma_5z_c\gamma_5\eta_{ab}]\epsilon$$
$$+\tfrac{1}{4}\Gamma^{bc}[\hat{D}_bB_{ac}-(i/2)\gamma_5\hat{R}^{\ (\text{cov})}_{bcad}A^d-\tfrac{1}{4}\bar{z}_a\psi^{(\text{cov})}_{bc}-\tfrac{1}{6}(S-i\gamma_5P)\eta_{ab}\hat{D}_c(S+i\gamma_5P)$$
$$+(i/6)\gamma_5\{(B_{ab}-i\gamma_5\hat{D}_bA_a)-(B-i\gamma_5\hat{D}_dA^d)\eta_{ab}\}A_c$$
$$-\tfrac{1}{6}\bar{z}^d\Gamma_{dca}z_b-\tfrac{1}{8}\bar{z}^d\Gamma_{dbc}z_a+\tfrac{1}{16}\bar{z}_d\Gamma_{abc}z^d-\tfrac{1}{12}\bar{z}\cdot\gamma z_c\eta_{ab}-\tfrac{1}{12}\bar{z}_b\gamma_cz_a-\tfrac{1}{24}\bar{z}_b\gamma_az_c]\epsilon, \tag{3.18}$$

where use has been made of the Bianchi identity,

$$\begin{aligned}&\hat{D}_aB-\hat{D}^bB_{ab}\\&\quad=-\tfrac{1}{2}\hat{D}_a(|u|^2+\tfrac{1}{3}A_cA^c)-\tfrac{1}{3}\hat{D}_b(A_aA^b)\\&\qquad+\bar{z}^c\psi^{(\text{cov})}_{ac}+\tfrac{1}{6}\bar{z}_a\gamma\cdot z\end{aligned}\tag{3.19}$$

and of

$$\tfrac{1}{4}\bar{z}^c\psi^{(\text{cov})}_{ac}+\tfrac{1}{8}\epsilon_{abcd}\bar{z}^d\gamma_5\psi^{(\text{cov})bc}=\tfrac{1}{12}\bar{z}_a\gamma\cdot z. \tag{3.20}$$

In (3.18) $\hat{R}^{\ (\text{cov})cd}_{ab}$ arises because of the commutator of two supercovariant derivatives, which is given by a Lorentz transformation with parameter $-R^{\ (\text{cov})cd}_{ab}$ plus a supersymmetry transformation with parameter $-\psi^{(\text{cov})}_{ab}$,

$$[\hat{D}_a,\hat{D}_b]=-\delta_L(\hat{R}^{\ (\text{cov})cd}_{ab})-\delta_{\text{sup}}(\psi^{(\text{cov})}_{ab}) \tag{3.21}$$

on any field with only tangent spaces indices. In particular,

$$\begin{aligned}&[\hat{D}_a,\hat{D}_b]A^c=\hat{R}^{\ (\text{cov})cd}_{ab}A_d-(i/2)\bar{z}^c\gamma_5\psi^{(\text{cov})}_{ab},\\&[\hat{D}_a,\hat{D}_b]z^c\\&\quad=\tfrac{1}{4}\hat{R}^{\ (\text{cov})de}_{ab}\Gamma_{de}z^c+\hat{R}^{\ (\text{cov})cd}_{ab}z_d\\&\qquad-\tfrac{1}{2}[H^ci\gamma_5-K^c-B^{cd}\gamma_d+i\hat{\not{D}}A^c\gamma_5\\&\qquad-(i/3)\Gamma^{cd}\gamma_5(S-i\gamma_5P)A_d]\,\psi^{(\text{cov})}_{ab}.\end{aligned}\tag{3.22}$$

The second commutator in (3.22) is used in the derivation of δd_a together with the Bianchi identities for $\hat{R}^{\ (\text{cov})}_{abcd}$ of Appendix B.

Finally, to obtain the transformation of d_a, one needs as additional ingredients $\delta\hat{\Box}A_a$ and $\delta\,\psi^{(\text{cov})}_{ab}$,

$$\begin{aligned}\delta\hat{\Box}A_a&=(i/2)\bar{\epsilon}(\hat{D}_c-(i/2)\gamma_5A_c-\tfrac{1}{2}\eta\gamma_c)\\&\quad\times[\gamma_5(\hat{D}^c-(i/2)A^c\gamma_5)-\tfrac{1}{2}\eta\gamma^c\gamma_5]z_a\\&\quad-\tfrac{1}{2}\bar{\epsilon}A^d(\hat{\not{D}}-(i/2)\gamma_5\not{A}-2\eta)\,\psi^{(\text{cov})}_{ad}\\&\quad+\tfrac{1}{2}\bar{\epsilon}\Gamma_{[ca}z_{d]}\hat{F}^{cd}-\tfrac{1}{4}\bar{\epsilon}\eta\gamma^c\Gamma_{[ca}z_{d]}A^d\\&\quad-\tfrac{1}{3}\bar{\epsilon}\gamma^d\gamma_{[d}z_{c]}\hat{D}^cA_a-\bar{\epsilon}\gamma_c\psi^{(\text{cov})}_{ad}\hat{D}^cA^d\\&\quad+\tfrac{1}{2}\bar{\epsilon}A^d\Gamma_{[ca}\hat{D}^cz_{d]},\end{aligned}\tag{3.23}$$

$$\begin{aligned}\delta\,\psi^{(\text{cov})}_{ab}&=[\tfrac{1}{4}\hat{R}^{\ (\text{cov})cd}_{ab}+(i/3)\,\delta^c_{\ [a}\hat{D}_{b]}A^d\gamma_5\\&\quad+\tfrac{1}{18}(|u|^2+A_eA^e)\delta_a^{\ c}\delta_b^{\ d}\\&\quad-\tfrac{1}{9}\delta^c_{\ [a}A_{b]}A^d]\Gamma_{cd}\epsilon+(i/3)\gamma_5\hat{F}_{ab}\epsilon\\&\quad+\tfrac{1}{3}[H_{[a}i\gamma_5+K_{[a}-(i/3)(S+i\gamma_5P)\gamma_5A_{[a}]\gamma_{b]}\epsilon\\&\quad+(i/9)\epsilon_{abcd}A^c\gamma^d(S-i\gamma_5P)\epsilon.\end{aligned}\tag{3.24}$$

In order to obtain the final form for δd_a, one must also use the Bianchi identities for $\hat{R}^{\ (\text{cov})}_{abcd}$ and $\psi^{(\text{cov})}_{ab}$ that we provide in Appendix B.

The square of the Ricci multiplet W_aW^a has to be derived from its first component $C^{\text{new}}=A_aA^a$, since the multiplication rule of vector multiplets without external indices does not apply here, as anticipated in Ref. 4. Using their nomenclature, we get

$$W_aW^a=[C^{\text{new}},Z^{\text{new}},H^{\text{new}},K^{\text{new}},B^{\ \text{new}}_b,\Lambda^{\text{new}},D^{\text{new}}], \tag{3.25}$$

with

$$\begin{aligned}&C^{\text{new}}=A_aA^a,\quad Z^{\text{new}}=2A_az^a,\\&H^{\text{new}}=2A^aH_a-\tfrac{1}{2}\bar{z}^az_a,\\&K^{\text{new}}=2A^aK_a+(i/2)\bar{z}^a\gamma_5z_a,\\&B^{\ \text{new}}_b=2A^aB_{ab}+(i/2)\bar{z}^a\gamma_5\gamma_bz_a,\\&\Lambda^{\text{new}}=2A^a\lambda_a+(H_a-i\gamma_5K_a-iB_{ab}\gamma^b\gamma_5-\hat{\not{D}}A_a)z^a,\\&D^{\text{new}}=H_aH^a+K_aK^a-B_{ab}B^{ab}-(\hat{D}_bA_a)(\hat{D}^bA^a)\\&\qquad+2A^a(\hat{D}_a\hat{D}_bA^b-\hat{\Box}A_a)-\tfrac{4}{3}B_{ab}A^aA^b\\&\qquad-\tfrac{2}{3}BA_aA^a-\tfrac{1}{3}|u|^2A_aA^a+\bar{z}^a(\hat{\not{D}}-(i/2)\not{A}\gamma_5)z_a\\&\qquad-2\bar{z}^a(\hat{D}_a+(i/2)A_a\gamma_5)\gamma\cdot z-\bar{z}^a(S+i\gamma_5P)z_a\\&\qquad-(i/3)A_a\bar{z}_b\gamma_5\Gamma^{abc}z_c-\tfrac{4}{3}i\bar{z}^a\gamma_5\not{A}z_a\\&\qquad+(i/3)\bar{z}\cdot A\gamma_5\gamma\cdot z+2i\bar{z}^b\gamma_5\psi^{(\text{cov})}_{ab}A^a.\end{aligned}\tag{3.26}$$

(The component D^{new} has not been correctly reported in Ref. 4.)

We stress that these components are derived directly. As it turns out, all the components except D^{new} can be expressed in terms of the components of W_a according to the naive multiplication rule. For D^{new} we have an extra term, though

$$\begin{aligned}D^{\text{new}}&=2A^ad_a-2\bar{z}^a\lambda_a+H_aH^a+K_aK^a-B_{ab}B^{ab}\\&\quad-(\hat{D}_bA_a)\hat{D}^bA^a-\bar{z}^a(\hat{\not{D}}-(i/2)\not{A}\gamma_5)z_a\\&\quad+(i/6)\bar{z}^a\gamma_5A^c\Gamma_{cab}z^b.\end{aligned}\tag{3.27}$$

IV. THE SUPER GAUSS–BONNET THEOREM

Applying the density formula (2.11) to the multiplet (3.12) we obtain after some manipulations

$$\begin{aligned}
e^{-1}\mathscr{L}(\overline{W}_{ab}\,W^{ab}) = {} & \tfrac{1}{2}\widehat{R}_{\mu\nu ab}\,\widehat{R}^{\,ab\mu\nu} - \widehat{R}_{ab}\widehat{R}^{\,ba} + \tfrac{1}{6}\widehat{R}^{\,2} - \tfrac{3}{4}R_{\mu\nu}(A)\,\widetilde{R}^{\,\mu\nu}(A) - \tfrac{1}{2}e^{-1}\,\epsilon^{\mu\nu\rho\sigma}R_{\nu\rho ab}\,\bar{\phi}_\mu\gamma_5\,\Gamma^{ab}\,\psi_\sigma \\
& - ie^{-1}\,\epsilon^{\mu\nu\rho\sigma}F_{\nu\rho}\,\bar{\phi}_\mu\,\psi_\sigma + \bar{\psi}_\mu\,\widehat{R}_{\nu\rho}\,\gamma^\rho R^{\,\mu\nu}(Q) + 4e^{-1}\,\epsilon^{\mu\nu\rho\sigma}\bar{\phi}_\sigma\,\gamma_\rho\gamma_5[D_\nu - (i/2)\gamma_5 A_\nu]\,\phi_\mu \\
& - \tfrac{1}{2}\overline{R}_{ab}(Q)\gamma_\mu\,\psi_\nu\overline{R}^{\,\mu\nu}(Q)\gamma^a\,\psi^b - \tfrac{3}{4}\,\overline{R}_{bd}(Q)\gamma_\nu\,\psi^d\,\overline{R}^{\,a\nu}(Q)\gamma^b\,\psi_a + \tfrac{1}{4}\bar{\psi}_\rho\,\Gamma^{\rho\sigma}\,\psi_\sigma\,\overline{R}_{\mu\nu}(Q)R^{\,\mu\nu}(Q) \\
& - \tfrac{1}{4}\bar{\psi}_\rho\,\Gamma^{\rho\sigma}\gamma_5\,\psi_\sigma\,\overline{R}_{\mu\nu}(Q)\gamma_5 R^{\,\mu\nu}(Q) + \tfrac{1}{2}e^{-1}\,\epsilon^{\mu\nu\rho\sigma}\,\partial_\mu\big[-\Omega^{\rm LOR}_{\nu\rho\sigma} + \tfrac{2}{3}\,\Omega_{\nu\rho\sigma}(A) \\
& - 8\bar{\phi}_\nu(1+\gamma_5)\mathscr{D}_\rho\,\psi_\sigma + 4\bar{\phi}_\nu\,\gamma_\rho\,\phi_\sigma\big]\,.
\end{aligned} \tag{4.1}$$

[The imaginary part of this Lagrangian as given here has already been given in Ref. 6. There is a misprint in the last term, though, in Eq. (7.6) there, which propagates to Eqs. (7.16), (7.21), and (7.24). Equation (7.6) also needs an overall factor 1/2.] The symbols $\widehat{R}_{\mu\nu ab}$, $\widehat{R}_{ab}$, and $\widehat{R}$ have been defined in (3.8) and (3.14) and

$$\begin{aligned}
&\Omega^{\rm LOR}_{\nu\rho\sigma} = \omega_{[\nu}{}^{ab}\,R_{\rho\sigma]ab} + \tfrac{2}{3}\,\omega_{\nu ab}\,\omega_\rho{}^{bc}\,\omega_{\sigma c}{}^{a}\,, \quad \Omega_{\nu\rho\sigma}(A) = A_{[\nu}\,F_{\rho\sigma]}\,, \quad F_{\rho\sigma} = 2\partial_{[\rho}\,A_{\sigma]}\,, \\
&\phi_\nu = \tfrac{1}{3}\gamma^\rho(S_{\nu\rho} + \tfrac{1}{2}\gamma_5\,\widetilde{S}_{\nu\rho})\,, \quad S_{\nu\rho} = -2\mathscr{D}_{[\nu}\,\psi_{\rho]}\,, \quad \mathscr{D}_\mu\,\psi_\nu = (\partial_\mu + \tfrac{1}{4}\omega_\mu{}^{ab}\,\Gamma_{ab} + (i/2)\gamma_5\,A_\mu)\psi_\nu\,.
\end{aligned} \tag{4.2}$$

The spin connection $\omega_\mu{}^{ab}$ contains ψ_μ torsion but no b_μ terms. The object ϕ_μ is obtained from the φ_μ field of conformal supergravity by putting the dilation field b_μ equal to zero and is related to z_a in (2.2),

$$\phi_\mu = \varphi_\mu(b_\nu = 0) = \tfrac{1}{3}\,z_\mu + \tfrac{1}{2}\eta\psi_\mu\,. \tag{4.3}$$

If we apply now the D-density formula (2.16) to the multiplet $R\overline{R}$ in (3.4), we get the following Lagrangian after the necessary algebra:

$$\begin{aligned}
&\tfrac{1}{2}e^{-1}\mathscr{L}(\overline{R}R) \\
&= \tfrac{1}{4}\widehat{R}^{\,2} + B\bar{\psi}^a\,z_a - \tfrac{1}{3}B(|u|^2 + 2A_aA^{\,a}) + \tfrac{1}{6}(3|u|^2 + 2A_c\,A^{\,c})\bar{\psi}^a\,z_a \\
&\quad - \tfrac{1}{9}[\,|u|^4 + 4|u|^2A_c\,A^{\,c} + (A_c\,A^{\,c})^2] + (\widehat{D}_a\,A^{\,a})^2 - \widehat{D}_\mu u\,\widehat{D}^\mu\,u^* + (i/3)A^{\,\mu}(u\overleftrightarrow{\widehat{D}}_\mu\,u^*) \\
&\quad + 2\bar{z}^a\,D_a\gamma\cdot z - \bar{z}^a\not{\widehat{D}}\,z_a + e^{-1}\,\partial_\mu\big[e\bar{z}\cdot\gamma z^\mu - \tfrac{3}{2}e\bar{\psi}_\nu\,\Gamma^{\mu\nu\rho}(S + i\gamma_5P)\phi_\rho \\
&\quad - (i/2)e\bar{z}_a\,\Gamma^{a\mu c}\not{A}\gamma_5\,\psi_c\big] + \tfrac{2}{3}\,\bar{z}_a(S + i\gamma_5P)R^{\,a} - \tfrac{1}{3}\,\bar{z}_a(S + i\gamma_5P)\Gamma^{ab}\,R_b \\
&\quad + 3ie^{-1}\,\varepsilon^{\mu\nu\rho\sigma}A_\mu\,\bar{\psi}_\nu\,D_\rho\,\phi_\sigma + 3i\bar{\psi}\cdot A\gamma_5\,\Gamma^{bc}\,D_b\,\phi_c - 6i\bar{\psi}_a\gamma_5A^{\,b}\,\Gamma^{ac}\,D_{[b}\,\phi_{c]} \\
&\quad + \bar{\psi}\cdot\gamma\widehat{D}_a(S + i\gamma_5P)z^a - \bar{\psi}_a\not{\widehat{D}}(S + i\gamma_5P)z^a + 9\bar{\phi}_\mu\,\Gamma^{\mu\nu\rho}\,D_\nu\,\phi_\rho - \tfrac{1}{2}\bar{\psi}\cdot A\gamma^b\,\psi^d\,D_b\,A_d \\
&\quad + \tfrac{1}{2}\bar{\psi}\cdot\gamma A^{\,b}\,\psi^d\,D_b\,A_d + \tfrac{1}{2}\bar{\psi}\cdot A\gamma\cdot\psi D_b\,A^{\,b} + \tfrac{1}{4}A_a\,A^{\,a}\,\bar{\psi}_b\,R^{\,b} - \tfrac{1}{2}\bar{\psi}\cdot A\,A_d\,\Gamma^{bcd}\,D_b\,\psi_c \\
&\quad + \bar{\psi}_aA_d\,\Gamma^{acd}\,A^{\,b}\,D_{[b}\,\psi_{c]} + (i/6)\,A_a\,\bar{z}_b\,\gamma_5\,\Gamma^{abc}\,z_c - (i/3)\bar{z}\cdot A\gamma_5\gamma\cdot z + (i/6)\bar{z}^a\gamma_5\not{A}z_a \\
&\quad + (i/24)|u|^2\,\bar{\psi}_a\,\Gamma^{abc}\gamma_5\,A_b\,\psi_c + \tfrac{1}{8}e^{-1}\varepsilon^{\mu\nu\rho\sigma}\,\bar{\psi}_\mu\gamma_\nu\,\psi_\rho\,\bar{z}^a\gamma_5\gamma_\sigma z_a + (i/8)\bar{\psi}_c\gamma_5\,\not{A}\Gamma^{cab}\,\psi_a\,\bar{\psi}_b\gamma\cdot z \\
&\quad - (i/8)\bar{\psi}_c\,\not{A}\Gamma^{cab}\psi_a\,\bar{\psi}_b\gamma_5\gamma\cdot z - \tfrac{1}{4}(\bar{\psi}^az_a)^2 + \tfrac{1}{4}\bar{\psi}_a\,\gamma_c\psi_b\bar{z}^a\,\gamma^cz^b - (3i/8)\bar{\psi}_b\gamma_dA\cdot\psi\bar{\psi}_c\gamma_5\Gamma^{bd}\,z^c \\
&\quad + (3i/16)\bar{\psi}_a\,\not{A}\psi_b\,\bar{\psi}_c\gamma_5\Gamma^{ab}\,z^c - (5i/8)\bar{\psi}\cdot\gamma A\cdot\psi\bar{\psi}_a\gamma_5z^a + (i/16)\epsilon^{acbe}A_e\,\bar{\psi}_a\,\gamma_c\,\psi_b\,\bar{\psi}_dz^d \\
&\quad - (i/8)\bar{\psi}\cdot\gamma\psi_b\,\bar{\psi}_c\gamma_5\,\Gamma^{bd}A_d\,z^c - (i/32)\,\bar{\psi}_a\,\Gamma_{\mu\nu}\,\psi_b\,\bar{\psi}_c\,\not{A}\Gamma^{\mu\nu}\,\Gamma^{ab}\gamma_5\,z^c \\
&\quad + (i/8)\bar{\psi}_a\,\Gamma^{\mu\nu}A\cdot\psi\bar{\psi}_d\,\gamma^a\,\Gamma_{\mu\nu}\,\gamma_5\,z^d + (i/4)\,\epsilon^{abce}\,\bar{\psi}_b\,\gamma_e\,\psi_c\bar{z}\cdot A\psi_a - (i/8)\epsilon^{abcd}\,\bar{\psi}_a\,\gamma_b\,\psi_c\bar{z}\cdot\gamma\not{A}\psi_d \\
&\quad + (i/8)\bar{\psi}_a\,\gamma_b\,\psi_c\,\bar{\psi}^b(S - i\gamma_5P)\Gamma^{ac}\gamma_5A\cdot\psi - (i/4)\bar{\psi}_a\,\gamma_bA\cdot\psi\bar{\psi}^b(S - i\gamma_5P)\Gamma^{ad}\,\gamma_5\,\psi_d\,.
\end{aligned} \tag{4.4}$$

The Lagrangian for the Ricci square multiplet (3.25) is obtained by using again (2.15) with the components (3.26) and the result can be cast in the form:

$$\begin{aligned}
&e^{-1}\mathscr{L}(W_a\,W^{a}) \\
&= \widehat{D}_\mu u\widehat{D}^\mu\,u^* - (i/3)A^{\,a}(u\overleftrightarrow{\widehat{D}}_a\,u^*) + \tfrac{1}{9}[\,|u|^4 + 4|u|^2A_cA^{\,c} + (A_c\,A^{\,c})^2] \\
&\quad + \tfrac{1}{3}B(|u|^2 + 2A_c\,A^{\,c}) - \tfrac{1}{6}(3|u|^2 + 2A_c\,A^{\,c})\bar{\psi}^a\,z_a - \tfrac{9}{4}(\widehat{R}_{ab}\,\widehat{R}^{\,ba} - \tfrac{2}{9}\widehat{R}^{\,2}) \\
&\quad + \tfrac{9}{4}\widehat{R}_{ab}\,\bar{\psi}\cdot\gamma R^{\,ab}(Q) + \tfrac{9}{4}\,\widehat{R}^{\,ab}\,\bar{\psi}^c\,\gamma_a\,R_{bc}(Q) - \tfrac{9}{64}\{\tfrac{1}{2}\bar{\psi}^c\,\psi_c\,\overline{R}^{\,bd}\,(Q)R_{bd}(Q) \\
&\quad - \tfrac{1}{2}\bar{\psi}^c\gamma_5\,\psi_c\,\overline{R}^{\,bd}(Q)\gamma_5R_{bd}(Q) - 2\bar{\psi}^c\,\gamma^a\,\gamma_5\,\psi_d\,\overline{R}^{\,bd}(Q)\gamma_a\,\gamma_5\,R_{bc}(Q) \\
&\quad - \tfrac{1}{2}\bar{\psi}_c\,\Gamma^{cd}\,\psi_d\,\overline{R}^{\,ab}(Q)R_{ab}(Q) + \tfrac{1}{2}\,\bar{\psi}_c\,\gamma_5\,\Gamma^{cd}\,\psi_d\,\overline{R}^{\,ab}(Q)\gamma_5\,R_{ab}(Q)\} \\
&\quad - (3i/2)\bar{\psi}_a\,\phi_b\,\widetilde{F}^{ab} + \tfrac{9}{16}(\bar{\psi}_a\,\Gamma^{ab}\,\psi_b\,\bar{\phi}_c\,\Gamma^{cd}\,\phi_d - 4\bar{\psi}_a\,\Gamma^{ac}\,\psi_b\,\bar{\phi}_c\,\Gamma^{bd}\,\phi_d \\
&\quad + \bar{\psi}_a\,\Gamma^{cd}\,\psi_b\,\bar{\phi}_c\,\Gamma^{ab}\,\phi_d) - \tfrac{1}{4}A^{\,a}\,A_a\,\bar{\psi}_b\,R^{\,b} - 3i\bar{\phi}^b\,\tilde{\psi}_{ab}\,A^{\,a} - (9i/2)A_a\,\bar{\phi}_b\,\gamma_5\,\Gamma^{abc}\,\phi_c
\end{aligned}$$

$$
\begin{aligned}
&+\tfrac{3}{4}\hat{F}_{ab}\hat{F}^{ab}-(\hat{D}_a A^a)^2+e^{-1}\partial_\mu[ee_a{}^\mu D_b(A^b A^a)]-ie^{-1}\partial_\mu(e\bar{\psi}\cdot A\gamma_5 z^\mu)\\
&-e^{-1}\partial_\mu[e\hat{D}^\mu(A^a A_a)]+\tfrac{1}{2}\bar{\psi}\cdot\gamma A\cdot\psi D_b A^b-\tfrac{1}{2}\bar{\psi}\cdot\gamma\psi_a A^b D_b A^a\\
&+\tfrac{1}{2}\bar{\psi}\cdot A\gamma^b\psi_c D_b A^c+\tfrac{1}{4}(\bar{\psi}^a z_a)^2-\tfrac{1}{4}\bar{\psi}_a\gamma_c\psi_b\bar{z}^a\gamma^c z^b-3i\bar{\psi}\cdot A\gamma_5\Gamma^{bc}D_b\phi_c\\
&+6i\bar{\psi}_b\gamma_5\Gamma^{bc}A^a D_{[a}\phi_{c]}+\tfrac{1}{2}\bar{\psi}\cdot A\Gamma^{bcd}A_d D_b\psi_c-\bar{\psi}_b\Gamma^{bcd}A^a A_d D_{[a}\psi_{c]}+\bar{z}^a\not{D}z_a\\
&-2\bar{z}^a D_a\gamma\cdot z-\bar{z}^a[\hat{\not{D}}(S-i\gamma_5P)+B]\psi_a-\bar{\psi}\cdot\gamma\hat{D}_a(S+i\gamma_5P)z^a-(i/6)\bar{z}^a\gamma_5\not{A}z_a\\
&+(i/3)\bar{z}\cdot A\gamma_5\gamma\cdot z-(i/6)A_a\bar{z}_b\gamma_5\Gamma^{abc}z_c-(i/24)|u|^2A_a\bar{\psi}_b\gamma_5\Gamma^{abc}\psi_c\\
&-\tfrac{2}{3}\bar{z}^a(S+i\gamma_5P)(R_a-\tfrac{1}{2}\Gamma_{ab}R^b)+(i/4)e^{-1}\epsilon^{\mu\nu\rho\sigma}\bar{\psi}_\mu\gamma_\nu\psi_\rho((i/2)\bar{z}^a\gamma_5\gamma_\sigma z_a-\bar{z}\cdot A\psi_\sigma)\\
&-(5i/8)\bar{\psi}\cdot A\gamma\cdot\psi\bar{\psi}_a\gamma_5 z^a-(3i/8)\bar{\psi}\cdot A\gamma_d\psi_b\bar{\psi}_a\gamma_5\Gamma^{bd}z^a+(i/8)\bar{\psi}\cdot A\Gamma_{cd}\psi_b\bar{\psi}_a\gamma^b\Gamma^{cd}\gamma_5 z^a\\
&-(i/8)\bar{\psi}_c\gamma_5\not{A}\Gamma^{cab}\psi_a\bar{\psi}_b\gamma\cdot z+(i/8)\bar{\psi}_c\not{A}\Gamma^{cab}\psi_a\bar{\psi}_b\gamma_5\gamma\cdot z+(i/8)\epsilon^{bdac}\bar{\psi}_b\gamma_d\psi_a\bar{z}\cdot\gamma\not{A}\psi_c\\
&-(3i/16)\bar{\psi}_c\not{A}\psi_a\bar{\psi}_b\gamma_5\Gamma^{ca}z^b-(i/16)\epsilon^{cdae}A_e\bar{\psi}_c\gamma_d\psi_a\bar{\psi}_b z^b+(i/32)\bar{\psi}_c\Gamma^{\mu\nu}\psi_a\bar{\psi}_b\not{A}\Gamma_{\mu\nu}\Gamma^{ca}\gamma_5 z^b\\
&+(i/8)\bar{\psi}_c\gamma\cdot\psi\bar{\psi}_a\gamma_5A_d\Gamma^{dc}z^a+(i/8)\bar{\psi}\cdot A\gamma_5\Gamma^{bc}(S-i\gamma_5P)\psi_d\bar{\psi}_b\gamma^d\psi_c-(i/4)\bar{\psi}_b\gamma_5\Gamma^{ba}(S-i\gamma_5P)\psi_c\bar{\psi}\cdot A\gamma^c\psi_a\,.
\end{aligned}
\tag{4.5}
$$

A considerable simplification occurs when one considers the combination $\mathscr{L}(\overline{W}_{ab}W^{ab})+\frac{1}{2}\mathscr{L}(R\overline{R})$, which gives the Lagrangian

$$
\begin{aligned}
&e^{-1}[\mathscr{L}(W_aW^a)+\tfrac{1}{2}\mathscr{L}(\overline{R}R)]\\
&\quad=-\tfrac{9}{4}(\hat{R}_{ab}\hat{R}^{ba}-\tfrac{1}{3}\hat{R}^2)+\tfrac{9}{4}\hat{R}_{ab}\bar{\psi}\cdot\gamma R^{ab}(Q)+\tfrac{9}{4}\hat{R}^{ab}\bar{\psi}^c\gamma_a R_{bc}(Q)-\tfrac{9}{64}\{\tfrac{1}{2}\bar{\psi}^c\psi_c\overline{R}^{bd}(Q)R_{bd}(Q)\\
&\qquad-\tfrac{1}{2}\bar{\psi}^c\gamma_5\psi_c\overline{R}^{bd}(Q)\gamma_5R_{bd}(Q)-2\bar{\psi}^c\gamma^e\gamma_5\psi_d\overline{R}^{bd}(Q)\gamma_e\gamma_5R_{bc}(Q)-\tfrac{1}{2}\bar{\psi}_c\Gamma^{cd}\psi_d\overline{R}^{ab}(Q)R_{ab}(Q)\\
&\qquad+\tfrac{1}{2}\bar{\psi}_c\gamma_5\Gamma^{cd}\psi_d\overline{R}^{ab}(Q)\gamma_5R_{ab}(Q)\}-(3i/2)\bar{\psi}_a\phi_b\tilde{F}^{ab}-3i\bar{\phi}^b\tilde{\psi}_{ab}A^a+3ie^{-1}\epsilon^{\mu\nu\rho\sigma}A_\mu\bar{\psi}_\nu D_\rho\phi_\sigma\\
&\qquad+9\bar{\phi}_\mu\Gamma^{\mu\nu\rho}(D_\nu-(i/2)A_\nu\gamma_5)\phi_\rho+\tfrac{3}{4}\hat{F}_{ab}\hat{F}^{ab}+e^{-1}\partial_\mu[ee_a{}^\mu D_b(A^bA^a)-e\hat{D}^\mu(A^aA_a)-ie\bar{\psi}\cdot A\gamma_5z^\mu]\\
&\qquad+\tfrac{9}{16}(\bar{\psi}_a\Gamma^{ab}\psi_b\bar{\phi}_c\Gamma^{cd}\phi_d-4\bar{\psi}_a\Gamma^{ac}\psi_b\bar{\phi}_c\Gamma^{bd}\phi_d+\bar{\psi}_a\Gamma^{cd}\psi_b\bar{\phi}_c\Gamma^{ab}\phi_d)\\
&\qquad+e^{-1}\partial_\mu[e\bar{z}\cdot\gamma z^\mu-\tfrac{3}{2}e\bar{\psi}_\nu\Gamma^{\mu\nu\rho}(S+i\gamma_5P)\phi_\rho-(i/2)e\bar{z}_a\Gamma^{a\mu c}\not{A}\gamma_5\psi_c]\,.
\end{aligned}
\tag{4.6}
$$

The super Gauss–Bonnet (GB) theorem is obtained when we take the combination

$$
\begin{aligned}
&\mathscr{L}(\overline{W}_{ab}\overline{W}^{ab})+\tfrac{4}{9}[\mathscr{L}(W_aW^a)+\tfrac{1}{2}\mathscr{L}(R\overline{R})]\\
&\quad=\partial_\mu K^\mu_{(r)}+i\,\partial_\mu K^\mu_{(i)}\,,
\end{aligned}
\tag{4.7}
$$

which is a total divergence. The real part is the four-divergence of

$$
\begin{aligned}
K^\mu_{(r)}=\epsilon^{\mu\nu\rho\sigma}(&\tfrac{1}{2}\Omega^{\rm GB}_{\nu\rho\sigma}+(4i/3)\bar{\psi}_\nu\phi_\rho A_\sigma-4\bar{\phi}_\nu\gamma_5\mathscr{D}_\rho\psi_\sigma)\\
&+\tfrac{4}{9}e[e_a{}^\mu D_b(A^bA^a)\\
&-\hat{D}^\mu(A_aA^a)-i\bar{\psi}\cdot A\gamma_5z^\mu\\
&+\bar{z}\cdot\gamma z^\mu-\tfrac{3}{2}\bar{\psi}_\nu\Gamma^{\mu\nu\rho}(S+i\gamma_5P)\phi_\rho\\
&-(i/2)\bar{z}_\nu\Gamma^{\nu\mu\rho}\not{A}\gamma_5\psi_\rho]\,,
\end{aligned}
\tag{4.8}
$$

and the imaginary part is the divergence of

$$
\begin{aligned}
K^\mu_{(i)}=-i\epsilon^{\mu\nu\rho\sigma}(&-\tfrac{1}{2}\Omega^{\rm LOR}_{\nu\rho\sigma}+\tfrac{1}{3}\Omega_{\nu\rho\sigma}(A)\\
&-4\bar{\phi}_\nu\mathscr{D}_\rho\psi_\sigma+2\bar{\phi}_\nu\gamma_\rho\phi_\sigma)\,.
\end{aligned}
\tag{4.9}
$$

The Gauss–Bonnet three-form $\Omega^{\rm GB}_{\nu\rho\sigma}$ is given by

$$
\Omega^{\rm GB}_{\nu\rho\sigma}=\tfrac{1}{2}\epsilon_{abcd}(\omega_{[\mu}{}^{ab}R_{\nu\rho]}{}^{cd}-\tfrac{2}{3}\omega_{[\mu}{}^{ab}\omega_\nu{}^{ce}\omega_{\rho]e}{}^d)
\tag{4.10}
$$

and satisfies

$$
\begin{aligned}
\epsilon^{\mu\nu\rho\sigma}\partial_\mu\Omega^{\rm GB}_{\nu\rho\sigma}&=\tfrac{1}{4}\epsilon^{\mu\nu\rho\sigma}\epsilon_{abcd}R_{\mu\nu}{}^{ab}R_{\rho\sigma}{}^{cd}\\
&=e(R_{abcd}R^{cdab}-4R_{ab}R^{ba}+R^2)\,.
\end{aligned}
\tag{4.11}
$$

In obtaining the super Gauss–Bonnet theorem (4.7), the following identity must be used:

$$
\begin{aligned}
&\hat{R}_{abcd}\hat{R}^{cdab}-4\hat{R}_{ab}\hat{R}^{ba}+\hat{R}^2\\
&\quad-e^{-1}\epsilon^{\mu\nu\rho\sigma}R_{\nu\rho ab}\bar{\phi}_\mu\gamma_5\Gamma^{ab}\psi_\sigma\\
&\quad=R_{abcd}R^{cdab}-4R_{ab}R^{ba}+R^2\\
&\qquad+2\bar{\psi}_a\Gamma^{db}\psi_b\bar{\phi}_d\Gamma^{ac}\phi_c-\tfrac{1}{2}\bar{\psi}_a\Gamma^{cd}\psi_b\bar{\phi}_c\Gamma^{ab}\phi_d\\
&\qquad-\tfrac{1}{2}\bar{\psi}_a\Gamma^{ab}\psi_b\bar{\phi}_c\Gamma^{cd}\phi_d\,.
\end{aligned}
\tag{4.12}
$$

V. GAUSS–BONNET AND CHERN–SIMONS MULTIPLETS

A vector multiplet $V=[C,Z,H,K,B_b,\Lambda,D]$ can be projected into a chiral one, which corresponds to the superspace operation $(\overline{\mathscr{D}}\overline{\mathscr{D}}-8R)V$, as shown in Ref. 19, whose formulas can be synthesized as

$$
\begin{aligned}
\Sigma(V)=[&-H+iK+\tfrac{2}{3}Cu^*,\\
&-i\Lambda_L-i(\hat{\not{D}}+(i/2)\not{A})Z_R+(i/3)u^*Z_L\\
&-\tfrac{1}{3}\not{A}Z_R-\tfrac{2}{3}C\gamma\cdot z_R\,,D+\hat{\Box}C+i\hat{D}_aB^a\\
&+\tfrac{2}{3}C(B-i\hat{D}_aA^a)+\tfrac{2}{3}A^a(B_a-i\hat{D}_aC)\\
&+\tfrac{2}{3}u(H-iK)-\tfrac{1}{3}u^*(H+iK)\\
&+(2i/3)\overline{Z}_L\gamma\cdot z+(i/3)\overline{Z}_R\gamma\cdot z]\,.
\end{aligned}
\tag{5.1}
$$

If we apply the density formula (2.11) to the chiral projection $\Sigma(V)$ (5.1), one obtains a complex Lagrangian whose real part is the D-density formula $\mathscr{L}_D$ in (2.16) plus a divergence[19] and whose imaginary part is a total divergence. Explicitly,

$$\begin{aligned}\mathscr{L}_h(\Sigma(V)) = \mathscr{L}_D(V) &+ \partial_\mu\left[e(\hat{D}^\mu C + (i/2)\bar{Z}\gamma_5\Gamma^{\mu\nu}\psi_\nu)\right] \\ &+ i\,\partial_\mu\left[e(B^\mu - \tfrac{2}{3}CA^\mu - \tfrac{1}{2}\bar{Z}\Gamma^{\mu\nu}\psi_\nu)\right. \\ &\left.+ (i/4)\epsilon^{\mu\nu\rho\sigma}C\bar{\psi}_\nu\gamma_\rho\psi_\sigma\right]. \qquad (5.2)\end{aligned}$$

The third basic type of multiplet in supergravity is the linear multiplet,[20] which is a vector multiplet whose chiral projection vanishes.[21] This means that only the C, Z, and B_b components are independent and the vector obeys the constraint

$$\hat{D}_a(B^a - \tfrac{2}{3}CA^a) + \tfrac{1}{2}\bar{Z}\gamma\cdot z = 0\,, \qquad (5.3)$$

which can be solved in terms of an antisymmetric tensor $b_{\mu\nu}$,[21]

$$\begin{aligned}B^\mu = \tfrac{2}{3}CA^\mu - (i/4)e^{-1}\epsilon^{\mu\nu\rho\sigma}(\partial_\nu b_{\rho\sigma} + C\bar{\psi}_\nu\gamma_\rho\psi_\sigma \\ - i\bar{Z}\gamma_5\Gamma_{\nu\rho}\psi_\sigma)\,. \qquad (5.4)\end{aligned}$$

The dependent components H, K, Λ, and D are easily read off from (5.1).

Equation (5.2) provides an alternative way to present the super Gauss–Bonnet theorem, if we just replace C, Z, and B_μ by the corresponding components of the multiplet $\tfrac{4}{3}(W_a W^a + \tfrac{1}{2}R\bar{R})$,

$$\begin{aligned}&\mathscr{L}_h\left[\bar{W}_{ab}W^{ab} + \tfrac{4}{3}\Sigma(W_a W^a + \tfrac{1}{2}R\bar{R})\right] \\ &\quad= \partial_\mu(eV^\mu_{(r)}) + i\partial_\mu(eV^\mu_{(i)})\,, \qquad (5.5)\end{aligned}$$

with

$$\begin{aligned}V^\mu_{(r)} &= e^{-1}K^\mu_{(r)} + \tfrac{4}{3}\left[\hat{D}^\mu(A_a A^a + \tfrac{1}{2}|u|^2)\right. \\ &\quad\left.+ i\bar{\psi}_\nu\gamma_5\Gamma^{\nu\mu}A\cdot z - \tfrac{1}{2}\bar{\psi}_\nu\Gamma^{\nu\mu}(S - i\gamma_5 P)\gamma\cdot z\right], \\ V^\mu_{(i)} &= e^{-1}K^\mu_{(i)} + \tfrac{4}{3}\left[B^{\text{new}\,\mu} + (i/2)u\overleftrightarrow{\hat{D}}^\mu u^*\right. \\ &\quad - (i/2)\bar{z}\cdot\gamma\gamma_5\gamma^\mu\gamma\cdot z - \tfrac{2}{3}(A_a A^a + \tfrac{1}{2}|u|^2)A^\mu \\ &\quad + (i/4)e^{-1}\varepsilon^{\mu\nu\rho\sigma}(A_a A^a + \tfrac{1}{2}|u|^2)\,\bar{\psi}_\nu\gamma_\rho\psi_\sigma \\ &\quad\left.- \bar{\psi}_\nu\Gamma^{\nu\mu}A\cdot z - (i/2)\,\bar{\psi}_\nu\gamma_5\Gamma^{\nu\mu}(S - i\gamma_5 P)\gamma\cdot z\right], \qquad (5.6)\end{aligned}$$

and $K^\mu_{(r)}$ and $K^\mu_{(i)}$ have been defined in (4.8) and (4.9) while $B_b{}^{\text{new}}$ is given in (3.26).

Now we turn to the Chern–Simons (CS) and Gauss–Bonnet multiplets. Following Ref. 6 we define the Lorentz Chern–Simons multiplet Ω^{CS} by its chiral projection

$$\Sigma(\Omega^{\text{CS}}) = \bar{W}_{ab}W^{ab}\,. \qquad (5.7)$$

The Gauss-Bonnet Chern–Simons multiplet $\Omega_{(i)}$ will be defined by the modification of (5.7) implied by (5.5),

$$\Sigma(\Omega_{(i)}) = \bar{W}_{ab}W^{ab} + \tfrac{4}{3}\Sigma(W_a W^a + \tfrac{1}{2}R\bar{R})\,, \qquad (5.8)$$

while the Gauss–Bonnet multiplet $\Omega_{(r)}$ will satisfy instead,

$$\Sigma(\Omega_{(r)}) = i\bar{W}_{ab}W^{ab} + \tfrac{4}{3}i\Sigma(W_a W^a + \tfrac{1}{2}R\bar{R})\,. \qquad (5.9)$$

All these multiplets are defined up to a linear multiplet of course. In order to find Ω^{CS}, $\Omega_{(i)}$, and $\Omega_{(r)}$, one must proceed to solve Eqs. (5.7), (5.8), and (5.9) following the method described in Ref. 6. Suppose we are looking for the solution $\Omega = [C^\Omega, Z^\Omega, H^\Omega, K^\Omega, B_b{}^\Omega, \Lambda^\Omega, D^\Omega]$ to the problem

$$\Sigma(\Omega) = [z, \chi_L, h] = Y\,, \qquad (5.10)$$

where the multiplet $Y = [z, \chi_L, h]$, $\chi_L = \tfrac{1}{2}(I + \gamma_5)\chi$, is given and furthermore we want Ω real. From (5.1) and (5.10) one gets

$$\begin{aligned}H^\Omega &= \tfrac{2}{3}C^\Omega S - \tfrac{1}{2}(z + z^*)\,, \\ K^\Omega &= -\tfrac{2}{3}C^\Omega P - (i/2)(z - z^*), \\ \Lambda^\Omega &= i\gamma_5\chi - (\hat{\not{D}} - (i/2)\not{A}\gamma_5)Z^\Omega \\ &\quad + \tfrac{1}{3}(S + i\gamma_5 P - i\not{A}\gamma_5)Z^\Omega + \tfrac{2}{3}iC^\Omega\gamma_5\gamma\cdot z\,, \qquad (5.11) \\ D^\Omega &= \tfrac{1}{2}(h + h^*) - \hat{\Box}C^\Omega - \tfrac{2}{3}C^\Omega B \\ &\quad - \tfrac{2}{3}A^a B_a{}^\Omega - \tfrac{2}{9}C^\Omega|u|^2 \\ &\quad + \tfrac{1}{6}(zu + z^*u^*) - (i/6)\bar{Z}^\Omega\gamma_5\gamma\cdot z\,.\end{aligned}$$

Also there is a constraint on $B_a{}^\Omega$:

$$\begin{aligned}&-(i/2)(h - h^*) \\ &\quad= \hat{D}^a B_a{}^\Omega - \tfrac{2}{3}C^\Omega\hat{D}_a A^a - \tfrac{2}{3}A^a\hat{D}_a C^\Omega \\ &\quad\quad + (i/2)(zu - z^*u^*) + \tfrac{1}{2}\bar{Z}^\Omega\gamma\cdot z\,. \qquad (5.12)\end{aligned}$$

In order to solve this constraint we exploit the fact that if Ω is a solution to (5.10), so is $\Omega - L$ where L is a linear multiplet,

$$L = [C^L, Z^L, H^L, K^L, B_b{}^L, \Lambda^L, D^L]\,. \qquad (5.13)$$

Thus one can choose

$$C^L = C^\Omega, \quad Z^L = Z^\Omega\,, \qquad (5.14)$$

but, since Ω and L are different multiplets, the supercovariant derivatives differ:

$$\begin{aligned}&\hat{D}_a C^L - \hat{D}_a C^\Omega = 0\,, \\ &\hat{D}_a Z^\Omega - \hat{D}_a Z^L = -(i/2)z\psi_{aR} + (i/2)z^*\psi_{aL} \\ &\qquad\qquad + \tfrac{1}{2}(\not{B}^\Omega - \not{B}^L)\psi_a\,, \qquad (5.15) \\ &\hat{\Box}C^\Omega - \hat{\Box}C^L = \tfrac{1}{4}z\bar{\psi}_R{}^a\psi_a + \tfrac{1}{4}z^*\bar{\psi}_L{}^a\psi_a\,.\end{aligned}$$

[This gives explicit ψ_μ terms in the last two components of the multiplet that are missing in Eq. (5.25) of Ref. 6.]

Then, if we call $V_a = B_a{}^\Omega - B_a{}^L$, the constraint (5.12) becomes

$$\partial_\mu(eV^\mu) = \text{Im}\,\mathscr{L}_h(\Sigma(\Omega)). \qquad (5.16)$$

Equation (5.16) implies Eq. (5.10) has a local solution if and only if the imaginary part of the density for the chiral multiplet in (5.10) is a total divergence which in turn determines V^μ. The multiplets in (5.7) and (5.8) satisfy this condition, so we can readily write particular solutions for Ω^{CS} and $\Omega_{(i)}$ in (5.7) and (5.8) from the particular solution to the general problem (5.10)

$$\begin{aligned}\Omega = [&0, 0, -\tfrac{1}{2}(z + z^*), -(i/2)(z - z^*), V_b, \\ &i\gamma_5\chi + (i/2)z\gamma\cdot\psi_R - (i/2)z^*\gamma\cdot\psi_L - \tfrac{1}{2}\gamma^a\not{V}\psi_a, \\ &\tfrac{1}{2}(h + h^*) + \tfrac{1}{6}(zu + z^*u^*) - \tfrac{2}{3}A^a V_a \\ &- \tfrac{1}{4}z\bar{\psi}_R{}^a\psi_a - \tfrac{1}{4}z^*\bar{\psi}_L{}^a\psi_a]\,, \qquad (5.17)\end{aligned}$$

and one only needs to replace z, χ, and h by the components of the corresponding multiplets on the right-hand side of

Eqs. (5.7) and (5.8), which can be extracted from (5.1) and the results of Sec. III, as well as

$$V^\mu = e^{-1} K^\mu_{(i)}, \quad \text{for } \Omega^{\text{CS}}, \qquad V^\mu = V^\mu_{(i)}, \quad \text{for } \Omega_{(i)} . \tag{5.18}$$

In order to obtain the Gauss–Bonnet multiplet $\Omega_{(r)}$ in (5.9), we need the solution to the modified problem,

$$\Sigma(\Omega') = iY. \tag{5.19}$$

Following exactly the same procedure used to solve (5.10), one finds the particular solution to (5.19),

$$\begin{aligned}\Omega' = [&0,0, -(i/2)(z-z^*), \tfrac{1}{2}(z+z^*), V_b,\\ &-\chi - \tfrac{1}{2}\gamma\cdot\psi_R z - \tfrac{1}{2}\gamma\cdot\psi_L z^* - \tfrac{1}{2}\gamma^a \not{V}\psi_a ,\\ &(i/2)(h-h^*) + (i/6)(zu - z^* u^*) - \tfrac{2}{3}A^a V_a\\ &-(i/4)z\bar{\psi}_R{}^a\psi_a + (i/4)z^*\bar{\psi}_L{}^a\psi_a] ,\end{aligned} \tag{5.20}$$

if and only if

$$\partial_\mu (eV^\mu) = \text{Re}\,\mathscr{L}_h(Y). \tag{5.21}$$

From Eqs. (5.21), (5.5), and (5.9), we get

$$V^\mu = V^\mu_{(r)}, \quad \text{for } \Omega_{(r)}, \tag{5.22}$$

and the components z, χ, h in (5.20) are the ones of the multiplet $\overline{W}_{ab} W^{ab} + \frac{4}{9}\Sigma(W_a W^a + \frac{1}{2}R\overline{R})$, which are easily derived from (5.1) and the formulas of Sec. III.

Finally, let us mention that

$$\Omega^{\text{CS}} + \tfrac{4}{9}(W_a W^a + \tfrac{1}{2}R\overline{R}) \tag{5.23}$$

is also a solution to Eq. (5.8) for $\Omega_{(i)}$ as it is obvious from (5.7) and (5.8). However, the solution from (5.17) looks simpler than (5.23), while the difference between the two is a linear multiplet.

When we add the multiplets Ω^{CS}, $\Omega_{(i)}$, or $\Omega_{(r)}$ to a linear multiplet L, we obtain generalizations of the field strength for the antisymmetric tensor $b_{\mu\nu}$ in (5.4), giving three different versions of the Green–Schwarz mechanism. The one corresponding to Ω^{CS} has been described in detail in Ref. 6.

VI. CONCLUSION

We have completed the derivation of the Ricci multiplet in old-minimal supergravity, as well as presented a complete and explicit proof of the super-Gauss–Bonnet theorem. Our main results rely on the full calculation of the fermionic terms of these multiplets which have never been reported before. In particular, we bring the attention of the reader to Eqs. (4.7), (4.8), and (4.9), which give the full expression of the supertopological densities. Also the full component expression of the Chern–Simons multiplet $\Omega_{(i)}$ [Eqs. (5.17) and (5.18)] and the Gauss–Bonnet multiplet $\Omega_{(r)}$ [Eqs. (5.20) and (5.22)] are new results. These multiplets as well as Ω^{CS} couple to linear multiplets through different supersymmetric generalizations of the Green–Schwarz mechanism.

APPENDIX A: RELATIONS BETWEEN CONFORMAL QUANTITIES AND POINCARÉ QUANTITIES

The Q curvature of conformal supergravity is given by[4,22,23]

$$\begin{aligned}R_{\mu\nu}(Q) = &-2[D_{[\mu}(\omega(e,\psi,b)) + (i/2)\gamma_5 A_{[\mu}]\psi_{\nu]}\\ &-2\gamma_{[\mu}\varphi_{\nu]} - b_{[\mu}\psi_{\nu]} ,\end{aligned} \tag{A1}$$

where b_μ is the dilatation gauge field which is independent, while the Lorentz connection $\omega_\mu{}^{ab}(e,\psi,b)$ and the S-supersymmetry gauge field φ_μ are dependent as a result of solving the constraints of conformal supergravity, which give

$$\begin{aligned}&\omega_\mu{}^{ab}(e,\psi,b) = \omega_\mu{}^{ab}(e,\psi) + 2e^{[a}{}_\mu b^{b]} ,\\ &\varphi_\mu = \tfrac{1}{2}(\mathscr{R}_\mu - \tfrac{1}{3}\gamma_\mu\gamma\cdot\mathscr{R}) ,\\ &\mathscr{R}^\mu = e^{-1}\epsilon^{\mu\nu\rho\sigma}\gamma_5\gamma_\nu\\ &\qquad\times[\partial_\rho + \tfrac{1}{4}\omega_\rho{}^{ab}(e,\psi,b)\Gamma_{ab} + (i/2)\gamma_5 A_\rho + \tfrac{1}{2}b_\rho]\psi_\sigma .\end{aligned} \tag{A2}$$

Actually φ_μ can be expressed in terms of z_μ,[4]

$$\varphi_\mu = \tfrac{1}{3}z_\mu + \tfrac{1}{2}(\eta + \not{b})\psi_\mu , \tag{A3}$$

and it is a simple matter to verify that $R_{\mu\nu}(Q)$ is in fact independent of b_μ. The relation between $R_{\mu\nu}(Q)$ and $\psi^{(\text{cov})}_{\mu\nu}$ has been given in the text:

$$R_{ab}(Q) = -\psi^{(\text{cov})}_{ab} + \tfrac{2}{3}\gamma_{[a}z_{b]} . \tag{A4}$$

The covariantized M curvature of conformal supergravity is

$$R^{(\text{cov})}_{\mu\nu}(M_{ab}) = R_{\mu\nu}(M_{ab}) - \bar{\psi}_{[\mu}\gamma_{\nu]}R_{ab}(Q) , \tag{A5}$$

where $R_{\mu\nu}(M_{ab})$ is given by gauging the superconformal algebra

$$\begin{aligned}R_{\mu\nu}(M^{ab}) = &R_{\mu\nu}{}^{ab}[\omega(e,\psi,b)] - 8e^{[a}{}_{[\mu}f^{b]}{}_{\nu]}\\ &-\bar{\psi}_{[\mu}\Gamma^{ab}\phi_{\nu]} ,\end{aligned} \tag{A6}$$

with $R_{\mu\nu}{}^{ab}[\omega(e,\psi,b)]$ being the Riemann tensor constructed with the connection in (A2). The conformal boost gauge field $f^b{}_\mu$ is also a dependent field and can be expressed in terms of Poincaré quantities[4]

$$\begin{aligned}f_{a\mu} = &-\tfrac{1}{6}e^c{}_\mu(B_{ca} - i\tilde{\tilde{F}}_{ca}) - \tfrac{1}{36}(|u|^2 + A_c A^c)e_{a\mu} + \tfrac{1}{18}A_\mu A_a\\ &+\tfrac{1}{12}\bar{\psi}_\mu z_a - \tfrac{1}{2}D_\mu b_a + \tfrac{1}{2}b_\mu b_a - \tfrac{1}{4}b_c b^c e_{a\mu} ,\end{aligned} \tag{A7}$$

where the connection in $D_\mu b_a$ is $\omega(e,\psi)$. When one considers the traceless combination to construct the Weyl tensor, the $f^a{}_\mu$ terms vanish identically. Furthermore, by replacing (A7) into (A6) one sees that the curvature $R_{\mu\nu}(M^{ab})$ is actually independent of b_μ (to our knowledge this has not been pointed out in the literature) and since, as already mentioned, $R_{\mu\nu}(Q)$ is also independent of b_μ, then $R^{(\text{cov})}_{\mu\nu}(M_{ab})$ does not depend of b_μ either. Hence

$$\begin{aligned}&R_{ab}(M^{cd})\\ &\quad= \widehat{R}_{ab}{}^{cd} + \tfrac{4}{3}\delta^{[c}{}_{[a}[B_{b]}{}^{d]} - i\hat{\tilde{\tilde{F}}}_{b]}{}^{d]}\\ &\qquad+\tfrac{1}{6}\delta^{d]}{}_{b]}(|u|^2 + A_e A^e) - \tfrac{1}{3}A_{b]}A^{d]}] ,\end{aligned} \tag{A8}$$

where

$$\widehat{R}_{ab}{}^{cd} = R_{ab}(M^{cd})|_{f^a{}_\mu = b_\mu = 0} = R_{ab}{}^{cd} - \bar{\psi}_{[a}\Gamma^{cd}\phi_{b]} \tag{A9}$$

and

$$\phi_\mu = \varphi_\mu(b_\nu = 0) . \tag{A10}$$

Combining (A9), (A10), and (A3) we get for $\widehat{R}_{ab}{}^{cd}$ the expression given in the text. For the covariantized curvature, we get simply

$$R_{ab}^{(\mathrm{cov})}(M^{cd}) = \hat{R}_{ab}^{(\mathrm{cov})cd} + \tfrac{4}{3}\delta^{[c}{}_{[a}[B_{b]}{}^{d]} - i\hat{\tilde{F}}_{b]}{}^{d]} + \tfrac{1}{6}\delta^{d]}{}_{b]}(|u|^2 + A_e A^e) - \tfrac{1}{3}A_{b]}A^{d]}] . \tag{A11}$$

From the cancellation of $f^a{}_\mu$ and b_μ terms we see that the Weyl tensor is in fact the traceless part of $\hat{R}_{abcd}^{(\mathrm{cov})}$.[16]

An interesting property of $\hat{R}_{\mu\nu}{}^{ab}$ is that the antisymmetric part of its contraction $\hat{R}_{\mu\rho} = e_a{}^\nu e_{b\rho}\hat{R}_{\mu\nu}{}^{ab}$ is

$$\hat{R}_{[\mu\nu]} = -\bar{\psi}_{[\mu}\phi_{\nu]} - \tfrac{1}{4}\bar{\psi}\cdot\gamma R_{\mu\nu}(Q) , \tag{A12}$$

from which is easily derived the identity given in the text:

$$B_{[ab]} = (i/2)\hat{\tilde{F}}_{ab} . \tag{A13}$$

The complete expression for $\hat{R}_{\mu\nu}$ is

$$\hat{R}_{\mu\nu} = R_{\mu\nu} + \tfrac{1}{2}\bar{\psi}_\mu\Gamma_{\nu c}\phi^c + \tfrac{1}{2}\bar{\psi}^c\Gamma_{c\nu}\phi_\mu , \tag{A14}$$

which has the trace

$$\hat{R} = g^{\mu\nu}\hat{R}_{\mu\nu} = R + \bar{\psi}_a\Gamma^{ab}\phi_b . \tag{A15}$$

Two more conformal curvatures that do not depend on the b_μ field are the A curvature

$$R_{\mu\nu}(A) = \tfrac{2}{3}(2\,\partial_{[\mu}A_{\nu]} - 3i\bar{\psi}_{[\mu}\gamma_5\varphi_{\nu]}) \tag{A16}$$

and the D curvature related to it,

$$R_{\mu\nu}(D) = -(i/2)\tilde{R}_{\mu\nu}(A) . \tag{A17}$$

As mentioned in the text, there is the relation[4]

$$R_{ab}(A) = \tfrac{2}{3}\hat{F}_{ab} . \tag{A18}$$

The covariantized S curvature is

$$R_{\mu\nu}^{(\mathrm{cov})}(S) = R_{\mu\nu}(S) + (i/2)(\gamma_5 R_{\lambda[\mu}(A) + \tilde{R}_{\lambda[\mu}(A))\gamma^\lambda\psi_{\nu]} , \tag{A19}$$

while $R_{\mu\nu}(S)$ is simply

$$R_{\mu\nu}(S) = 2D_{[\nu}(\omega(e,\psi,b))\varphi_{\mu]} - 2\gamma_a f^a{}_{[\mu}\psi_{\nu]} + b_{[\mu}\varphi_{\nu]} + i\gamma_5 A_{[\mu}\varphi_{\nu]} . \tag{A20}$$

The dependence of these curvatures on the dilatation field can be displayed by writing $R_{\mu\nu}^{(\mathrm{cov})}(S)$ in terms of Poincaré quantities,

$$R_{ab}^{(\mathrm{cov})}(S) = -\tfrac{2}{3}(\hat{D}_{[a} - (i/2)\gamma_5 A_{[a} + \tfrac{1}{2}\eta\gamma_{[a})z_{b]} + \tfrac{1}{2}(\eta + \not{b})R_{ab}(Q) . \tag{A21}$$

Then we can see that even though $R_{ab}^{(\mathrm{cov})}(S)$ is not independent of b_μ, the Weyl multiplet is, because the constraint on $R_{\mu\nu}(Q)$ (see Appendix B) implies

$$T_{abcd}\gamma^\mu R^{cd}(Q) = 0 , \tag{A22}$$

so that $T_{ab}{}^{cd}R_{cd}^{(\mathrm{cov})}(S)$ is independent of b_μ. With all these elements it is a trivial matter to see that the action of conformal supergravity is independent of the dilatation gauge field b_μ.[22]

APPENDIX B: BIANCHI IDENTITIES

The basic property of the Q curvature is the constraint[23]

$$\gamma^\nu R_{\mu\nu}(Q) = 0, \tag{B1}$$

from which we can derive several more identities

$$\begin{aligned} &\Gamma^{ab}R_{ab}(Q) = 0, \quad R_{ab}(Q) + \gamma_5\tilde{R}_{ab}(Q) = 0, \\ &T_{abcd}R^{cd}(Q) = R_{ab}(Q), \quad T_{abcd}\gamma^\mu R^{cd}(Q) = 0 , \\ &\gamma_{[a}R_{bc]}(Q) = 0, \quad \Gamma^{ab}\gamma_c R_{ab}(Q) = 0 , \\ &\Gamma^{ab}\gamma_\mu\gamma_\nu R_{ab}(Q) = -8R_{\mu\nu}(Q) . \end{aligned} \tag{B2}$$

These identities provide additional ones when we consider quadratic objects in $R_{ab}(Q)$,

$$\begin{aligned} &\bar{\psi}^c\gamma^a\psi_d\bar{R}^{bd}(Q)\gamma_a R_{bc}(Q) = 0 , \\ &\bar{\psi}^c\psi_d\bar{R}^{bd}(Q)R_{bc}(Q) = \tfrac{1}{4}\bar{\psi}^c\psi_c\bar{R}^{bd}(Q)R_{bd}(Q) , \\ &\bar{\psi}^c\gamma_5\psi_d\bar{R}^{bd}(Q)\gamma_5 R_{bc}(Q) = \tfrac{1}{4}\bar{\psi}^c\gamma_5\psi_c\bar{R}^{bd}(Q)\gamma_5 R_{bd}(Q) , \\ &\bar{\psi}_a\Gamma^{ac}\psi^d\bar{R}_{bd}(Q)R_c{}^b(Q) = -\tfrac{1}{4}\bar{\psi}_a\Gamma^{ac}\psi_c\bar{R}_{bd}(Q)R^{bd}(Q) . \end{aligned} \tag{B3}$$

For the covariantized M curvature, the basic constraint is

$$R_{ac}^{(\mathrm{cov})}(M^c{}_b) = (i/2)\tilde{R}_{ab}(A) , \tag{B4}$$

which trivially implies

$$R_{ab}^{(\mathrm{cov})}(M^{ab}) = 0 . \tag{B5}$$

These are complemented by two more,

$$\begin{aligned} &R_{ab}^{(\mathrm{cov})}(M^{cd}) - R^{(\mathrm{cov})cd}(M_{ab}) = -2i\delta_{[a}{}^{[c}\tilde{R}_{b]}{}^{d]}(A) , \\ &\epsilon^{ebcd}R_{ab}^{(\mathrm{cov})}(M_{cd}) = iR^e{}_a(A) , \end{aligned} \tag{B6}$$

and another one implied by the last one

$$\epsilon^{abcd}R_{ab}^{(\mathrm{cov})}(M_{cd}) = 0 . \tag{B7}$$

The basic Bianchi identity for $R_{ab}^{(\mathrm{cov})}(S)$ is

$$\Gamma_{[ab}R_{cd]}^{(\mathrm{cov})}(S) = 0 , \tag{B8}$$

from which we can obtain a set of additional ones

$$\begin{aligned} &\Gamma^{ab}R_{ab}^{(\mathrm{cov})}(S) = 0, \\ &\Gamma_a{}^{bc}R_{bc}^{(\mathrm{cov})}(S) + 2\gamma^b R_{ab}^{(\mathrm{cov})}(S) = 0 . \end{aligned} \tag{B9}$$

The identities for $R_{ab}(Q)$ together with the relation (A4) imply corresponding identities for $\psi_{ab}^{(\mathrm{cov})}$

$$\begin{aligned} &\gamma^b\psi_{ab}^{(\mathrm{cov})} = -z_a - \tfrac{1}{3}\Gamma_{ab}z^b , \\ &\Gamma^{ab}\psi_{ab}^{(\mathrm{cov})} = -2\gamma\cdot z, \\ &\psi_{ab}^{(\mathrm{cov})} + \gamma_5\tilde{\psi}_{ab}^{(\mathrm{cov})} = \tfrac{2}{3}(\gamma_{[a}z_{b]} + \tfrac{1}{2}\epsilon_{abcd}\gamma_5\gamma^c z^d) , \\ &\gamma_{[a}\psi_{bc]}^{(\mathrm{cov})} = \tfrac{2}{3}\Gamma_{[ab}z_{c]} , \\ &T_{abcd}\psi^{(\mathrm{cov})cd} = \psi_{ab}^{(\mathrm{cov})} - \tfrac{2}{3}\gamma_{[a}z_{b]} , \\ &T_{abcd}\gamma_e\psi^{(\mathrm{cov})cd} = \tfrac{4}{3}T_{abed}z^d , \\ &\Gamma^{ab}\gamma_c\psi_{ab}^{(\mathrm{cov})} = 2(z_c - \tfrac{1}{3}\Gamma_{cb}z^b) , \\ &\Gamma^{ab}\Gamma_{cd}\psi_{ab}^{(\mathrm{cov})} + 8\psi_{cd}^{(\mathrm{cov})} = \tfrac{2}{3}(8\gamma_{[c}z_{d]} + \Gamma_{cd}\gamma\cdot z) . \end{aligned} \tag{B10}$$

In addition, $\psi_{ab}^{(\mathrm{cov})}$ obeys the useful differential Bianchi identity,

$$(\hat{D}_{[c} + (i/2)\gamma_5 A_{[c} - \tfrac{1}{2}\gamma_{[c}\eta)\psi_{ab]}^{(\mathrm{cov})} = 0 . \tag{B11}$$

By using the relation (A11) between $R_{ab}^{(\mathrm{cov})}(M_{cd})$ and $\hat{R}_{abcd}^{(\mathrm{cov})}$ as well as (B4) through (B7), we can derive a set of Bianchi identities for $\hat{R}_{abcd}^{(\mathrm{cov})}$,

$$
\begin{aligned}
&\hat{R}^{(\text{cov})}_{abcd} - \hat{R}^{(\text{cov})}_{cdab} = 0, \quad \varepsilon^{ebcd}\hat{R}^{(\text{cov})}_{abcd} = 0\,,\\
&\hat{R}^{(\text{cov})}_{ab} = \hat{R}^{(\text{cov})c}_{ac}{}_{b} = \tfrac{1}{3}(2B_{ab} - i\tilde{F}_{ab}) + \tfrac{1}{3}(B + |u|^2)\eta_{ab}\\
&\qquad - \tfrac{2}{9}(A_a A_b - A_c A^c \eta_{ab})\,, \qquad \text{(B12)}
\end{aligned}
$$

and from the latter we deduce as well,

$$
\begin{aligned}
&\hat{R}^{(\text{cov})}_{[ab]} = 0\,,\\
&\hat{R}^{(\text{cov})} = \hat{R}^{(\text{cov})a}_{a} = 2B + \tfrac{4}{3}|u|^2 + \tfrac{2}{3}A_c A^c\,.
\end{aligned}
\qquad \text{(B13)}
$$

Aside from these we have the differential identity mentioned in the text,

$$
\begin{aligned}
\hat{D}_a B - \hat{D}^b B_{ab} = &-\tfrac{1}{2}\hat{D}_a(|u|^2 + \tfrac{1}{3}A_c A^c) - \tfrac{1}{3}\hat{D}_b(A_a A^b)\\
&+ \bar{z}^c \psi^{(\text{cov})}_{ac} + \tfrac{1}{6}\bar{z}_a \gamma\cdot z\,. \qquad \text{(B14)}
\end{aligned}
$$

Lastly, from the identities for $R^{(\text{cov})}_{ab}(S)$ and relation (A21) we can derive identities for $\hat{D}_a z_b$ like

$$
\begin{aligned}
&\Gamma^{ab}(\hat{D}_a - (i/2)\gamma_5 A_a + \tfrac{1}{2}\eta\gamma_a)z_b = 0\,,\\
&\gamma^b \epsilon_{abcd}(\hat{D}^c - (1/2)\gamma_5 A^c)z^d\\
&\quad = \gamma^b\{2\gamma_5(\hat{D}_{[a} - (1/2)\gamma_5 A_{[a})z_{b]}\\
&\qquad - (S - i\gamma_5 P)\gamma_5 \psi^{(\text{cov})}_{ab}\\
&\qquad - \eta(\gamma_5\gamma_{[a}z_{b]} + \tfrac{1}{2}\varepsilon_{abcd}\gamma^c z^d)\}.
\end{aligned}
\qquad \text{(B15)}
$$

APPENDIX C: ADDITIONAL TRANSFORMATION RULES

For reference, we provide the transformation rules of $\hat{D}_\mu z_a$, $\hat{D}_b A_a$, and $\hat{D}_a \hat{D}_b A_c$, from which one can derive some of the transformations used in Sec. III,

$$
\begin{aligned}
\delta\hat{D}_\mu z_a = &\tfrac{1}{2}[\hat{D}_\mu H_a i\gamma_5 - \hat{D}_\mu K_a - \gamma^b \hat{D}_\mu B_{ab} + i\gamma^b\gamma_5 \hat{D}_\mu \hat{D}_b A_a - (i/3)\Gamma_{ab}\gamma_5 \hat{D}_\mu\{(S - i\gamma_5 P)A^b\}]\epsilon\\
&- \tfrac{1}{4}[H_a i\gamma_5 - K_a - B_{ab}\gamma^b + i\gamma^b\gamma_5 \hat{D}_b A_a - (i/3)\Gamma_{ab}\gamma_5(S - i\gamma_5 P)A^b](iA_\mu\gamma_5 - \gamma_\mu\eta)\epsilon + \tfrac{1}{2}\{-\tfrac{5}{24}\bar{z}_a z_\mu - \tfrac{1}{8}\bar{z}_a\gamma_\mu\gamma\cdot z\\
&+ \tfrac{1}{12}\bar{z}_\mu\gamma_a\gamma\cdot z + [-\tfrac{5}{24}\bar{z}_a\gamma_5 z_\mu - \tfrac{1}{8}\bar{z}_a\gamma_5\gamma_\mu\gamma\cdot z + \tfrac{1}{12}\bar{z}_\mu\gamma_5\gamma_a\gamma\cdot z]\gamma_5 + \tfrac{1}{12}[\bar{z}^b\Gamma_{d\mu a}z_b + \tfrac{1}{2}\bar{z}_a\gamma_\mu\Gamma_{dc}z^c\\
&- \bar{z}_\mu\gamma_a\Gamma_{dc}z^c - \tfrac{5}{2}\bar{z}_a\gamma_\mu z_d + \tfrac{7}{2}\bar{z}_a\gamma_d z_\mu - \bar{z}_d\gamma_a z_\mu]\gamma^d + [\tfrac{1}{8}\bar{z}_a\gamma_5\gamma_\mu\Gamma_{dc}z^c - \tfrac{1}{12}\bar{z}_\mu\gamma_5\gamma_a\Gamma_{dc}z^c\\
&+ \tfrac{1}{24}\bar{z}_a\gamma_5\gamma_\mu z_d - \tfrac{1}{8}\bar{z}_a\gamma_5\gamma_d z_\mu + \tfrac{1}{12}\bar{z}_d\gamma_5\gamma_a z_\mu - \tfrac{1}{6}\bar{z}^b\gamma_5\eta_{d[c}\gamma_{a]}z_b e^c{}_\mu]\gamma^d\gamma_5 + \tfrac{1}{24}[\bar{z}_a\Gamma^{cd}z_\mu + \tfrac{1}{2}\bar{z}_a\Gamma_{\mu b}\Gamma^{cd}z^b\\
&- \bar{z}_\mu\Gamma_{ab}\Gamma^{cd}z^b - \bar{z}^b\Gamma^{cd}\Gamma_{\mu a}z_b]\Gamma_{cd}\}\epsilon + \tfrac{1}{8}\{\bar{z}^b\gamma_\mu\psi^{(\text{cov})}_{ab} + \bar{z}^b\gamma_5\gamma_\mu\psi^{(\text{cov})}_{ab}\gamma_5 + (2\bar{z}_a\psi^{(\text{cov})}_{\mu d} - \bar{z}^b\gamma_d\gamma_\mu\psi^{(\text{cov})}_{ab})\gamma^d\\
&- (2\bar{z}_a\gamma_5\psi^{(\text{cov})}_{\mu d} + \bar{z}^b\gamma_d\gamma_5\gamma_\mu\psi^{(\text{cov})}_{ab})\gamma^d\gamma_5 + \tfrac{1}{2}\bar{z}^b\Gamma^{cd}\gamma_\mu\psi^{(\text{cov})}_{ab}\Gamma_{cd}\}\epsilon + \tfrac{1}{2}\bar{\epsilon}\gamma^d\psi_\mu\hat{D}_d z_a\,. \qquad \text{(C1)}
\end{aligned}
$$

One can replace μ by a latin index everywhere provided one drops the last term. The remaining two are

$$
\delta\hat{D}_b A_a = (i/2)\bar{\epsilon}\gamma_5(\hat{D}_b - (i/2)A_b\gamma_5)z_a - (i/4)\bar{\epsilon}\eta\gamma_b\gamma_5 z_a + \tfrac{1}{2}\bar{\epsilon}(\Gamma_{[ac}z_{b]} - \gamma_b\psi^{(\text{cov})}_{ac})A^c\,, \qquad \text{(C2)}
$$

and

$$
\begin{aligned}
\delta\hat{D}_a\hat{D}_b A_c = &(i/2)\bar{\epsilon}(\hat{D}_a - (i/2)\gamma_5 A_a - \tfrac{1}{2}\eta\gamma_a)[\gamma_5(\hat{D}_b - (i/2)A_b\gamma_5) - \tfrac{1}{2}\eta\gamma_b\gamma_5]z_c + \tfrac{1}{2}\bar{\epsilon}(\hat{D}_a - (i/2)\gamma_5 A_a - \tfrac{1}{2}\eta\gamma_a)\\
&\times[(\Gamma_{[bc}z_{d]} - \gamma_b\psi^{(\text{cov})}_{cd})A^d] + \tfrac{1}{2}\bar{\epsilon}(\Gamma_{[ad}z_{e]} - \gamma_a\psi^{(\text{cov})}_{de})(\delta^d{}_b\hat{D}^e A_c + \delta_c{}^d\hat{D}_b A^e)\,. \qquad \text{(C3)}
\end{aligned}
$$

[1] B. Zwiebach, Phys. Lett. B **156**, 315 (1985); S. Cecotti, S. Ferrara, L. Girardello, and M. Porrati, *ibid.* **164**, 46 (1985); L. Romans and N. Warner, Nucl. Phys. B **273**, 320 (1986); S. Deser, in *Proceedings of the Nuffield Workshop on Supersymmetry and its Applications: Superstring Anomalies and Supergravity*, edited by G. W. Gibbons, S. W. Hawking, and P. K. Townsend (Cambridge U.P., Cambridge, 1986), p. 455; E. Bergshoeff, A. Salam, and E. Serzgin, Nucl. Phys. B **279**, 659 (1986); M. Daniel and N. E. Mavromatos, Phys. Lett. B **173**, 405 (1986); S. J. Gates, Jr. and H. Nishino, *ibid.* **173**, 52 (1986); H. Nishino and S. J. Gates, Jr., *ibid.* **173**, 417 (1986); R. E. Kallosh and B. E. W. Nilsson, *ibid.* **167**, 46 (1986); B. E. W. Nilsson, *ibid.* **175**, 319 (1986); B. E. W. Nilsson and A. K. Tollsten, *ibid.* **181**, 63 (1986); S. Ferrara, P. Fré, and M. Porrati, Ann. Phys. (NY) **175**, 112 (1987); S. Ferrara, Phys. Scr. **T15**, 132 (1987).

[2] M. B. Green and J. Schwarz, Phys. Lett. B **149**, 117 (1985).

[3] S. Cecotti, S. Ferrara, L. Girardello, A. Pasquinucci, and M. Porrati, Phys. Rev. D **33**, 2504 (1986); preprint CERN-TH-4718/87, 1987.

[4] P. K. Townsend and P. van Nieuwenhuizen, Phys. Rev. D **19**, 3592 (1979).

[5] S. Ferrara and M. Villasante, Phys. Lett. B **186**, 85 (1987).

[6] S. Cecotti, S. Ferrara, and M. Villasante, Int. J. Mod. Phys. A **2**, 1839 (1987).

[7] G. Girardi and R. Grimm, Nucl. Phys. B **292**, 181 (1987); P. Binetruy, G. Girardi, M. Müller, and R. Grimm, Phys. Lett. **195**, 389 (1987).

[8] S. Ferrara and P. van Nieuwenhuizen, Phys. Lett. B **74**, 333 (1978); K. Stelle and P. C. West, *ibid.* **74**, 330 (1978); M. F. Sohnius and P. C. West, *ibid.* **105**, 353 (1981).

[9] S. Ferrara, L. Girardello, T. Kugo, and A. van Proeyen, Nucl. Phys. B **222**, 191 (1983).

[10] S. Cecotti, S. Ferrara, M. Porrati, and S. Sabharwal, to appear in Nucl. Phys.

[11] S. Ferrara, S. Sabharwal, and M. Villasante, to appear in Phys. Lett. B; S. Ferrara and S. Sabharwal, preprint UCLA/88/TEP/11, CERN-TH-4995-88, 1988.

[12] S. Cecotti, S. Ferrara, and L. Girardello, Phys. Lett. B **198**, 336 (1987).

[13] W. Siegel, Maryland preprint, UMDEPP 88-180, 1988.

[14] For a review of supergravity, see P. van Nieuwenhuizen, Phys. Rep. C **68**, 189 (1981).

[15] M. Fischler, Phys. Rev. D **20**, 1842 (1979).

[16] S. Ferrara and P. van Nieuwenhuizen, Phys. Lett. B **78**, 573 (1978).

[17] K. S. Stelle and P. C. West, Phys. Lett. B. **77**, 376 (1978).

[18] S. Ferrara and P. van Nieuwenhuizen, Phys. Lett. B **76**, 404 (1978).

[19] K. S. Stelle and P. C. West, Nucl. Phys. B **145**, 175 (1978).

[20] S. Ferrara, J. Wess, and B. Zumino, Phys. Lett. B **51**, 239 (1974); W. Siegel, *ibid.* **85**, 333 (1979).

[21] M. F. Sohnius and P. C. West, Nucl. Phys. B **198**, 493 (1982).

[22] M. Kaku, P. K. Townsend, and P. van Nieuwenhuizen, Phys. Rev. D **17**, 3179 (1978).

[23] For a review of conformal tensor calculus, see T. Kugo and S. Uehara, Nucl. Phys. B **226**, 49 (1983). Useful properties of the superconformal curvatures can be found in Appendix B of T. Kugo and S. Uehara, Prog. Theor. Phys. **73**, 235 (1985).

Free energy density for continuous systems with and without superstability assumptions

Roman Gielerak
Research Center Bielefeld-Bochum-Stochastics, University of Bielefeld, Bielefeld, Federal Republic of Germany and Institute of Theoretical Physics, University of Wroclaw, Wroclaw, Poland

(Received 28 August 1987; accepted for publication 11 May 1988)

The thermodynamic limit of the free energy density for a large class of continuous, charged systems with stable and sufficiently regular many-body forces is studied. The main effort is placed on demonstrating rigorously that thermodynamic functions do not depend on the typical boundary conditions. Among the results the most important and new ones are the following. (1) For any *superstable* or *superstable in an extended sense* and sufficiently regular many-body interactions, the thermodynamic limit of the conditioned free energy density exists and is equal to that corresponding to the free boundary condition case. (2) A new criterion for the nondependence of the limiting free energy density on the typical boundary condition is formulated and proved. This criterion does not require any superstability type of assumptions. Among several applications of the results we list only two: (a) to the case of two-dimensional, neutral Yukawa gas in the monopole phase, which is not even stable according to the standard notion of stability, and (b) the uniqueness of a limiting Gibbs grand canonical state (modulo some technical conjecture) is proved for a class of continuous systems with two-body potentials of positive type.

I. INTRODUCTION

Let us consider a classical system formed by a finite number of particles enclosed in some bounded region $\Lambda \subset R^d$, which are interacting through a collection of many-body forces. The particles might have some internal degrees of freedom, such as charges or, more generally, some multipole moments, indexed by some set Σ. The distribution of these internal degrees of freedom is given by some *a priori* measure λ on Σ.

Statistical mechanics yields an expression for the corresponding free energy density at the volume Λ at fixed thermodynamic parameters that depend on the choice of the particular Gibbs ensemble. In this paper we will consider only the case of the grand canonical ensemble where the appropriate thermodynamical variables are temperature and chemical activity(ies). Then we have to control the corresponding expression for the free energy as Λ expands to R^d in a suitable way.

The question of the existence and the properties of the limiting free energy density have been studied first by Van Hove[1] (canonical ensemble with hard-core two-body forces) and also by Yang and Lee[2] (grand canonical ensemble, two-body forces with hard-core condition). Only after many years have these problems come to be studied in a more systematic way again. To get some idea how our knowledge about this problem has evolved we refer to Refs. 3–10 and references therein.

In a certain sense the progress culminates with the fundamental paper of Ruelle,[11] where for a general class of two-body potentials that are superstable and regular (in a suitable sense), the existence and shape independence (among the class of Van Hove sequences) of the limiting free energy density has been proved and some fundamental properties of the limiting free energy density has been established.

Ruelle's proofs are based on so-called probability estimates[11,12] (the tool we will explain in Sec. IV). Throughout Ruelle's fundamental study, the effects arising from imposing nontrivial boundary conditions have never been taken into account. This gap will be filled in Secs. IV and V of the present paper. We would like to stress the importance of establishing the independence of the limiting free energy density on the typical boundary condition, since this limit gives us all the thermodynamic functions of the system under consideration.

It is a general belief that thermodynamics should not depend on the typical boundary condition. This can be proved easily for lattice systems, both classical and quantum mechanical, with compact fiber space and short-ranged forces.[13,14] Some problems, however, might arise in the case of long range forces. Much less investigated is the situation concerning continuous systems. In this case, the configurational space is noncompact and this causes several complications in trying to extend the corresponding methods worked out for lattice systems with compact fiber space.

In the present paper we try to treat this fundamental problem of the nondependence of the limiting thermodynamics on the typical boundary condition for the classical, continuous, charged system. In Secs. III and IV we adopt the existing methods based on the fundamental probability estimates of Ruelle[11] to treat the above problem in the case of general multibody superstable and sufficiently well localized potentials. As the concept of superstability in the standard sense is no longer applicable to the charged systems, we introduce in Sec. IV the concept of superstability in an extended sense as more suitable for the case of charged particles. Presumably at this point it is hard to convince a potential reader of the significance of these results, but we find it very instructive to adopt the existing techniques to prove in detail such results, since we could not find a com-

 0022-2488/89/010115-19$02.50

plete proof in the existing literature.

The essentially new material is contained in Sec. V of the present paper. There we formulate a new criterion for the nondependence of the limiting thermodynamics on the typical boundary condition for a huge class of stable and sufficiently well localized interactions without any reference to the superstability properties. This criterion will be applied to several situations where the superstability assumption is not valid. Among our applications presented in Sec. V are the following.

(a) Proof of the nondependence of the limiting free energy density on the typical boundary condition in the case of neutral, many-component Yukawa plasma [in a monopole phase in two-dimensions (see Theorem 5.4)]. This seems to be a very interesting example as this interaction is even unstable from the classical point of view.

(b) Proof that for arbitrary stable, many-body potentials that are sufficiently well localized, the limiting thermodynamics does not depend on the typical boundary condition on the resolvent set of the corresponding Kirkwood–Salsburg operator (see Theorem 5.7).

Roughly speaking, our criterion states that a good control of the limiting Gibbs states and limiting free energy density obtained with the empty boundary condition is sufficient to prove the independence of the limiting thermodynamic of the typical boundary condition.

In the three Appendixes to this paper we have included (a) in Appendix A: a proof of probability estimates using the language of Poisson integration exclusively; (b) in Appendix B: proof of Theorem 5.5; and (c) in Appendix C: proof of Theorem 5.4.

Section II includes the basic definitions and notations used in this paper.

For the sake of completeness, let us mention that similar problems for the noncompact case have been treated in the literature for (a) quantum, continuous systems,[15–20] (b) lattice spin systems with noncompact spin space,[20–22] (c) quantum, Euclidean fields,[23–26] and (d) one-dimensional continuous systems.[27,28]

See also Ref. 29 where the very special case of Theorem 4.2 below has been treated and see also the recent papers discussing similar problems.[30,31]

II. PRELIMINARIES (REFS. 32 AND 33)

A. Configurational space

Let Σ be some Borel subset of some real, finite-dimensional space R^d. The set Σ will be called the space of charges. On the Borel σ algebra of sets of Σ there is given *a priori* some regular measure λ such that $\lambda(\Sigma) < \infty$.

In the sets $(R^d \otimes \Sigma)^{\otimes k}$, for $k = 1,2,\ldots$, let us introduce the following equivalence relation $\sim$. We will say that $\omega,\omega' \in (R^d \otimes \Sigma)^{\otimes k}$ are equivalent iff they differ only by the permutation of the elements composing them. For a given $\Lambda \subset R^d$, let $\hat{\omega}(\Lambda)$ be the restriction of ω to the set Λ. The subset $\hat{\Omega}_\infty$ of $2^{(R^d \times \Sigma)}/\sim$ having the property that card $\hat{\omega}(\Lambda) \equiv |\hat{\omega}(\Lambda)| < \infty$ (counting with the multiplicities) for every bounded $\Lambda \subset R^d$ is called the configurational space of the system. The subset $\hat{\Omega}(\Lambda)$ of $\hat{\Omega}_\infty$: $\hat{\omega} \in \hat{\Omega}(\Lambda) \Leftrightarrow \hat{\omega} = \hat{\omega}(\Lambda)$. Then, for every regular $\Lambda \subset R^d$ we have the natural decomposition $\hat{\Omega} = \hat{\Omega}(\Lambda) \otimes \hat{\Omega}(\Lambda^c)$, which corresponds to the notation $\omega = \hat{\omega}(\Lambda) \vee \hat{\omega}(\Lambda^c)$, where we denoted $\hat{\omega} \vee \hat{\omega}'$ as a union of the elements composing ω and ω'. We have also

$$\hat{\Omega}(\Lambda) = \sum_{n=0}^{\infty} \hat{\Omega}_n(\Lambda),$$

where $\hat{\Omega}_n(\Lambda) = \{\hat{\omega} \in \hat{\Omega}(\Lambda) \mid |\hat{\omega}| = n\}$. Moreover, the set $\hat{\Omega}_k(\Lambda)$ can be identified with $(\Lambda \otimes \Sigma)^{\otimes k}/\sim$. Therefore, it is possible to transform the measurable and topological structures of $(\Lambda \otimes \Sigma)^{\otimes k}/\sim$ into the set $\hat{\Omega}_k(\Lambda)$ and then also to $\hat{\Omega}(\Lambda)$. The corresponding σ algebras in $\hat{\Omega}(\Lambda)$ are denoted by $\hat{\mathscr{F}}(\Lambda)$. We notice that the measure and topological structures coincide here. By $\hat{\mathscr{F}}_\infty$ we denote the σ algebra $\hat{\mathscr{F}}(R^d)$.

Let a number $\delta > 0$ be given. Then we define the δ lattice $\mathbf{Z}_\delta = \{\underline{x} \in R^d \mid x_i = n_i\delta/2,\ i = 1,2,\ldots,d\}$, where n_i are integers, and we also define the δ cube,

$$\square_\delta(\mathbf{n}) = \{x \in R^d \mid n_i\delta - \delta/2 \leqslant x_i \leqslant n_i\delta + \delta/2\},$$

and δ covers:

$$C_\delta(R^d) = \bigcup_n \square_\delta(n), \tag{2.1}$$

$$C_\delta(\Lambda) = \bigcup_{n:\ \square_\delta(n) \cap \Lambda \neq \phi} \square_\delta(n). \tag{2.2}$$

A subset Λ of R^d will be called a regular δ polygon iff Λ is connected, one connected, and $\Lambda = C_\delta(\Lambda)$. For a given $\hat{\omega} \in \Omega$ we denote by $\hat{\omega}_\delta(n)$ the restriction of $\hat{\omega}$ to $\square_\delta(n)$ and by $n_\delta(\hat{\omega},n)$ the cardinality of $\hat{\omega}_\delta(n)$.

The corresponding notation without superscript caret will refer to projected $(R^d \times \Sigma) \to (R^d)$ objects involving $\hat{\omega}$. For example, if

$$\hat{\omega} = (x_1,\alpha_1,\ldots,x_n,\alpha_n)$$

then

$$\omega = (x_1,\ldots,x_n).$$

This convention will be used in Sec. IV of the present paper.

The subset of $\hat{\omega} \in \Omega_\infty$ such that $|\hat{\omega}| < \infty$ will be denoted as $\hat{\Omega}_f$.

B. Free, charged systems

Let us consider a system of cylindrical sets $C_\Lambda^n = \{\hat{\omega} \in \hat{\Omega} \mid |\omega(\Lambda)| = n\}$, where n runs over integers and Λ runs over bounded regular subsets of R^d. This system of sets with fixed Λ generates then some σ algebra of sets $\mathscr{F}'(\Lambda)$. On the generators C_Λ^n of this σ algebra we then define a function

$$\bar{\lambda}_{0,\Lambda}(C_\Lambda^n) \equiv \frac{1}{n!}\lambda(\Lambda)^{\otimes n}\left(\int_\Sigma \mu(d\alpha)\right)^{\otimes n}, \tag{2.3}$$

where λ is some Borel and regular measure on R^d. The system $\{\mathscr{F}'(\Lambda),\bar{\lambda}_{0,\Lambda}\}$ defines the projective family of measure spaces [we have identified $\bar{\lambda}_{0,\Lambda}$ with its extension to $\hat{\mathscr{F}}'(\Lambda)$ as a measure] whose projective limit can be defined on some $\{\hat{\Omega}'_\infty,\hat{\mathscr{F}}'_\infty,\bar{\lambda}_0\}$. It is easy to see that one can identify $\hat{\Omega}'_\infty = \hat{\Omega}_\infty$, $\hat{\mathscr{F}}'_\infty = \hat{\mathscr{F}}'_\infty$. The measure $\bar{\lambda}_{0,\Lambda}$ has the following remarkable property: if $\Delta \subset \Lambda$ then $\bar{\lambda}_{0,\Lambda} = \bar{\lambda}_{0,\Delta} \otimes \bar{\lambda}_{0,\Lambda-\Delta}$. The most popular choice for λ is the Lebesgue meaure multiplied by some positive constant z called chemical activity. This case we denote by $\bar{\pi}^z_{0,\Lambda}$. The different choices of λ leads to the description of non-self-interacting gas in some external field.

The system $\{\hat{\Omega}_\infty, \hat{\Omega}(\Lambda), \hat{\mathscr{F}}(\Lambda), \hat{\mathscr{F}}_\infty, \bar{\lambda}_{0,\Lambda}\}$ will be called the charged free system. To complete our expression let us note that

$$\bar{\lambda}_{0,\Lambda}(\Omega(\Lambda)) = \exp\left(\lambda(\Lambda)\left(\left(\int_\Sigma d\mu(\alpha)\right)\right)\right),$$

which follows easily via the above mentioned identification. In this paper we choose $\bar{\lambda}_{0,\Lambda}$ equal to $\bar{\pi}^z_{0,\Lambda}$. However, our results are valid for more general choices of λ as well.

Remark 2.1: For the many component systems as above, we should index every component by its own chemical activity $z = z(\alpha)$, $\alpha \in \Sigma$. But in order to simplify the notation we use $z = \text{const}$ on Σ.

C. Interactions

Any measurable function $\mathscr{E}: \Omega_f \to (-\infty, +\infty)$ will be called an interaction, and the value of $\mathscr{E}$ at the given point $\hat{\omega} \in \hat{\Omega}$ will be called the energy of the configuration $\hat{\omega}$. For statistical mechanics the most interesting interactions are those that are stable.

Definition 2.1: (a) An interaction $\mathscr{E}$ is stable $\Leftrightarrow$

$$\underset{B \in R_+}{\exists} : \underset{\hat{\omega} \in \hat{\Omega}_f}{\forall} E(\hat{\omega}) \geqslant -|\hat{\omega}| \cdot B. \tag{2.4}$$

(b) An interaction $\mathscr{E}$ is superstable iff

$$\underset{\delta > 0}{\forall} : \underset{\substack{A_\delta > 0 \\ B_\delta \in R_+}}{\exists} : \underset{\hat{\omega} \in \hat{\Omega}_f}{\forall} \mathscr{E}(\hat{\omega}) \geqslant \sum_{r \in \mathbf{Z}^d} A_\delta n_\delta^2(\hat{\omega}, r) + \sum_{r \in \mathbf{Z}^d} B n_\delta(\hat{\omega}, r). \tag{2.5}$$

Remark 2.2: The constant A_δ is, in general, dependent on δ. However, when $\mathscr{E}$ is superstable on some scale $\delta > 0$, then it is superstable on any scale $\delta' > 0$. For example, taking $\delta = 1$ and $\delta' = k^{-n}$ with $k,n \in N$, then if $\mathscr{E}$ is superstable on the scale $\delta = 1$ with the superstability constant A_1, then it is also superstable on the scale $\delta' = k^{-n}$ with the superstability constant $A_{\delta'} = k^{-nd} A_1$.

Definition 2.2: (1) A given interaction $\mathscr{E}$ is two-regular iff there exists a continuous, positive, and decreasing function ψ_2: $[0,\infty) \to [0,\infty)$ such that

$$\int_0^\infty r^{d-1} \psi_2(r) dr < \infty,$$

and moreover for any two configurations $\hat{\omega}_1, \hat{\omega}_2 \in \hat{\Omega}_f$, $\hat{\omega}_1 \cap \hat{\omega}_2 = \phi$ their interaction energy $\mathscr{E}(\hat{\omega}_1 | \hat{\omega}_2)$, defined by

$$\mathscr{E}(\hat{\omega}_1 | \hat{\omega}_2) = \mathscr{E}(\hat{\omega}_1 \vee \hat{\omega}_2) - \mathscr{E}(\hat{\omega}_1) - \mathscr{E}(\hat{\omega}_2), \tag{2.6}$$

can be estimated as follows:

$$|\mathscr{E}(\hat{\omega}_1 | \hat{\omega}_2)| \leqslant \frac{1}{2} \sum_{r \in \mathbf{Z}^d_\delta} \sum_{s \in \mathbf{Z}^d_\delta} \psi_2(|r-s|)(n_\delta^2(\hat{\omega}_1, r) + n_\delta^2(\hat{\omega}_2, s)). \tag{2.7}$$

(2) A given interaction $\mathscr{E}$ is called N regular iff there exists a sequence of continuous, positive, and decreasing functions $(\psi_2, \psi_3, \ldots, \psi_N)$, each defined on $\mathbb{R}_+$ and such that

$$\text{(i)} \quad \underset{a < k < N}{\forall} \sum_{\substack{A \subset \mathbf{Z}^d \\ |A| = k, A \ni 0}} \psi_k(\operatorname{diam} A) \equiv A_k < \infty, \tag{2.8}$$

and a sequence of integers $(k_2, \ldots, k_N)$ such that

(ii) for any two $\hat{\omega}_1, \hat{\omega}_2 \in \Omega_f$, $\hat{\omega}_1 \cap \hat{\omega}_2 = \phi$,

$$\begin{aligned} |\mathscr{E}(\hat{\omega}_1 | \hat{\omega}_2)| &\leqslant \sum_{p=2}^N \sum_{\substack{A \subset \mathbf{Z}^d, |A| = p \\ A = \{r_1, \ldots, r_p\}}} \psi_p(\operatorname{diam} A) \\ &\quad \times \{n^{k_p}(\omega_1, r_1) + \cdots + n^{k_p}(\omega_1, r_p) \\ &\quad + n^{k_p}(\omega_2, r_1) + \cdots + n^{k_p}(\omega_2, r_p)\}. \end{aligned} \tag{2.9}$$

Remark 2.3: The notion of two-regularity is a very restrictive assumption made on the energy function and it states that the effect of two-body interactions is predominating over all distance and that the effects arising from the many-body forces are negligibly small. On the other hand, there is no reason to expect this condition to hold in nature. It is clear that the proper notion of regularity for systems with many-body interactions is that of N regularity, or even better the following definition.

Definition 2.3: A given interaction $\mathscr{E}$ is ∞ regular iff there exists a sequence (infinite) $(\psi_2, \psi_3, \ldots)$ of continuous, positive, and decreasing functions ψ_k: $[0,\infty) \to (0,\infty)$ such that

$$\text{(i)} \quad \underset{k = 2, \ldots}{\forall} \sum_{\substack{A \subset \mathbf{Z}^d \\ |A| = k, A \ni 0}} \psi_k(\operatorname{diam} A) = A_k < \infty \tag{2.10}$$

and

$$\sum_{k=2}^\infty A_k < \infty.$$

(ii) For any $\hat{\omega}_1, \hat{\omega}_2 \in \hat{\Omega}_f$, $\hat{\omega}_1 \cap \hat{\omega}_2 = \phi$:

$$\begin{aligned} |\mathscr{E}(\hat{\omega}_1 | \hat{\omega}_2)| &\leqslant \sum_{p=2}^\infty \sum_{\substack{A \subset \mathbf{Z}^d \\ |A| = p \\ A = \{r_1, \ldots, r_p\}}} \psi_p(\operatorname{diam} A) \\ &\quad \times \left\{\sum_{j=1}^p n^{k_p}(\hat{\omega}_1, r_j) + \sum_{j=1}^p n^{k_p}(\hat{\omega}_2, r_p)\right\}, \end{aligned} \tag{2.11}$$

where $(k, k_2, \ldots)$ is a sequence of positive integers. According to this definition we should try to change the notion of superstability.

Definition 2.4: A given interaction $\mathscr{E}$ is N superstable iff there exists a positive constant $A_\delta > 0$ and a sequence of reals $B_\delta^1, \ldots, B_\delta^{2N-1}$ such that

$$\begin{aligned} \underset{\hat{\omega} \in \hat{\Omega}_f}{\forall} : \mathscr{E}(\omega) &\geqslant \sum_{r \in \mathbf{Z}^d} A_\delta n_\delta^{2N}(\hat{\omega}, r) \\ &\quad + \sum_{r \in \mathbf{Z}^d} \sum_{j=1}^{2N-1} B_\delta^j n^{N-j}(\hat{\omega}, r). \end{aligned} \tag{2.12}$$

Note that we have no constructive description of the N-regular ($N > 2$) and N-superstable interactions. The exceptions are well known; quite sharp criteria for the superstability and regularity are given[11,12] in the case of two-body chargeless systems. But if we introduce charges, then already in the case of two-body interactions some difficulties arise in trying to establish the superstabilty or extended superstability (this concept is discussed in Sec. III of the present paper) as our simple example in Sec. IV shows. See also the Appendix to our paper.[29]

Having in mind the decomposition $\widehat{\Omega}_\infty = \cup_{k=0}^{\infty} \widehat{\Omega}_k$ we can write,

$$\mathscr{E}(\hat{\omega}) = \sum_{k=1}^{\infty} \sum_{\substack{\hat{\omega}_k \subset \hat{\omega} \\ |\hat{\omega}_k| = k}} \mathscr{E}_k(\omega_k). \tag{2.13}$$

Then the functions $\mathscr{E}_k$ are called k-particle potentials. The k-particle potentials can be defined on the space $(R^d \otimes \Sigma)^{\otimes k}/\sim$ and then can be naturally extended to the space $(R^d \otimes \Sigma)^{\otimes k}$. These extensions are again called k-particle potentials and denoted by V_k.

Then the notion of N regularity can easily be expressed in terms of the corresponding decay properties of the corresponding k-particle potentials, but the constructive criteria for N superstability are much more difficult to obtain. We hope to discuss those questions elsewhere.

Let us note that for chargeless systems the concept of lower N regularity should be sufficient for the purposes of the present paper. However, it is because of our future applications that we use here the more restrictive notion of regularity.

D. Infinite volume grand canonical Gibbs measures

The finite volume equilibrium Gibbs measure $\mathscr{V}_\Lambda$ corresponding to the stable interaction $\mathscr{E}$ is defined on $\{\widehat{\Omega}(\Lambda), \widehat{\mathscr{F}}(\Lambda)\}$ by the following formula:

$$\mathscr{V}_\Lambda(d\omega) = (Z_\Lambda(z))^{-1} \exp[-\mathscr{E}(\hat{\omega})] \hat{\pi}^z_{0,\Lambda}(d\hat{\omega}), \tag{2.14}$$

where

$$Z_\Lambda(z) = \int_{\widehat{\Omega}(\Lambda)} \hat{\pi}^z_{0,\Lambda}(d\hat{\omega}) \exp[-\mathscr{E}(\hat{\omega})] = e^{p_\Lambda(z,\mathscr{E})|\Lambda|}, \tag{2.15}$$

where $p_\Lambda(z,\mathscr{E})$ is defined as the free energy density of the finite volume system corresponding to the empty boundary condition.

For any regular $\Delta \subset \Lambda$ we can define a finite volume measure $r_\Delta^{(\Lambda)}(d\omega)$ on $\Omega(\Delta)$ as a projection of $\mathscr{V}_\Lambda$ onto $\{\widehat{\Omega}(\Delta), \widehat{\mathscr{F}}(\Delta)\}$ by the formula

$$r_\Delta^{(\Lambda)}(d\omega(\Delta)) = (Z_\Lambda(z))^{-1} \int_{\Omega(\Lambda-\Delta)} \hat{\pi}^z_{0,\Lambda}(d\omega) e^{-\mathscr{E}(\omega(\Delta) \vee \omega(\Lambda-\Delta))}. \tag{2.16}$$

The measure $r_\Delta^{(\Lambda)}$ is then absolutely continuous with respect to the measure $\hat{\pi}^z_{0,\Lambda}$ with the corresponding Radon–Nikodym derivative

$$g_\Delta^{(\Lambda)}(\hat{\omega}) = \frac{dr_\Delta^{(\Lambda)}}{d\hat{\pi}^z_{0,\Lambda} \restriction \mathscr{F}(\Delta)}(\hat{\omega}) = Z_\Lambda^{-1} \int_{\Omega(\Lambda-\Delta)} \pi^z_{0,\Lambda-\Delta}(d\omega') e^{-\mathscr{E}(\hat{\omega} \vee \omega'(\Lambda-\Delta))}. \tag{2.17}$$

The system of densities $(g_\Delta^{(\Lambda)})_{(\Lambda)}$ forms a compatible system in the sense that starting from $r_{\Lambda'}$ $(\Lambda' \subset \Lambda)$ and successively projecting onto Λ and then onto Δ we always get the same $g_\Delta^{(\Lambda')}$. Of special interest are weak (sub)-limits $\lim_{\Lambda \uparrow R^d} \mathscr{V}_\Lambda = \mathscr{V}_\infty$ of the finite volume Gibbs measure $\mathscr{V}_\Lambda$. Several criteria for the existence of such limits are known with some special assumptions made about the interaction $\mathscr{E}$. (See Refs. 11, 32, 34, and 35.) Let $f(R^d)$ be a collection of all sequences (Λ_n) of bounded subsets of R^d and such that $\Lambda_n \to R^d$ monotonously and by inclusion. Let us denote by $\mathscr{G}^0(\mathscr{E})$ the set of all weak (sub)-limits $\lim_{n\to\infty} \mathscr{V}_{\Lambda_n} = r_\infty$ as (Λ_n) varies over $f(R^d)$. The elements of the set $\mathscr{G}^0(\mathscr{E})$ are called the infinite volume Gibbs measures corresponding to the empty boundary conditions.

The notion of a general infinite volume Gibbs measure corresponding to the stable interaction $\mathscr{E}$ is more subtle. Let $\widehat{\Xi} \subset \widehat{\Omega}_\infty$ be a subset for which the unique limit

$$\lim_{n\to\infty} \mathscr{E}(\hat{\omega}|\hat{\omega}(\tilde{\Lambda}_n)) = \mathscr{E}(\hat{\omega}|\hat{\omega}(\Lambda^c)) \tag{2.18}$$

exists uniformly in $\dot{\omega}$, for every $\hat{\omega} \in \widehat{\Omega}(\Lambda)$, every bounded Λ and every $(\Lambda_n) \in f(\Lambda^c)$ and the limit is $(\Lambda)_n$ independent. The question of measurability of the set Ξ that might arise from the above definition can be avoided here by introducing the corresponding constructions of Preston.[32]

Definition 2.5: Any probabilistic Borel measure μ on $(\widehat{\Omega}_\infty, \widehat{\mathscr{F}}_\infty)$ is called a Ξ-regular canonical Gibbs measure corresponding to the interaction $\mathscr{E}$ iff

(DLR 1) $\mu(\Xi) = 1$;

(DLR 2) $\forall_{|\Lambda|<\infty}\ \mu_{\restriction \mathscr{F}(\Lambda)} \alpha \hat{\pi}^z_{0,\Lambda}$

(α means absolute continuity);

(DLR 3) For any $\Lambda \subset R^d$: $|\Lambda| < \infty$, any $\hat{\omega} \in \Xi$: the conditional expectation value of μ with respect to the σ algebra $\widehat{\mathscr{F}}(\Lambda^c)$ is given by formula

$$E_\mu\{-|\widehat{\mathscr{F}}(\Lambda^c)\}(\hat{\omega}) = E_{\mathscr{V}_\Lambda}\{-|\widehat{\mathscr{F}}(\Lambda^c)\}(\hat{\omega}). \tag{2.19}$$

The set of canonical equilibrium Gibbs measures corresponding to $\{\Sigma, d\lambda\}, \mathscr{E}$ will be denoted by $\mathscr{G}_\Xi(\mathscr{E})$. The set of those $\mu \in \mathscr{G}^0(\mathscr{E})$ that have the property $\mu(\Xi) = 1$ is obviously included in $\mathscr{G}_\Xi(\mathscr{E})$ and denoted as $\mathscr{G}^0_\Xi(\mathscr{E})$.

By an easy calculation, we obtain

$$E_{\mathscr{V}_\Lambda}(-|\widehat{\mathscr{F}}(\Lambda^c))(\hat{\omega}) = (Z_\Lambda(\hat{\omega}))^{-1} \int_{\widehat{\Omega}(\Lambda)} \hat{\pi}^z_{0,\Lambda}(d\hat{\eta})(-) \times \exp[-\mathscr{E}(\hat{\eta})] \exp[-\mathscr{E}(\hat{\eta}|\hat{\omega}(\Lambda^c))], \tag{2.20}$$

$$Z_\Lambda^{\hat{\omega}}(\mathscr{E}) = \int_{\widehat{\Omega}(\Lambda)} \hat{\pi}^z_{0,\Lambda}(d\hat{\eta}) \times \exp[-\mathscr{E}(\hat{\eta}|\hat{\omega}(\Lambda^c))] \exp[-\mathscr{E}(\hat{\eta})] \tag{2.21}$$

Definition 2.6: Free energy density, conditioned by $\hat{\omega} \in \Xi$, in the finite volume $\Lambda \subset \mathbf{R}^d$, is given by

$$P_\Lambda^{\hat{\omega}}(z,\mathscr{E}) = (1/|\Lambda|) \ln \mathscr{Z}_\Lambda^{\hat{\omega}}(z). \tag{2.22}$$

III. GENERAL TWO-SUPERSTABLE, TWO-REGULAR (STRONGLY) INTERACTION CASE

A. Probability estimates

The fundamental paper of Ruelle[11] deals with the existence questions for infinite volume free energy density and the infinite volume Gibbs measures for the case of two-body superstable and regular interactions V_2.

Throughout this section we will assume that the interaction $\mathscr{E}$ is such that (i) $\mathscr{E}$ is superstable (two superstable); (ii) $\mathscr{E}$ is two-regular; or (iii) $\mathscr{E}$ is strongly two-regular, which means that the function ψ_2 has at least an asymptotic decay

$r^{-d-\varepsilon}$, for some $\varepsilon>0$ as $r\uparrow\infty$.

Before beginning we wish to explain the main ideas concerning the probability estimates. Let $\Lambda\subset R^d$ be a bounded region in R^d such that $\Lambda=\Lambda_1\cup\Lambda_2$, where $\Lambda_1\cap\Lambda_2=\phi$. Then we have

$$\mathscr{E}(\hat{\omega}(\Lambda))=\mathscr{E}(\hat{\omega}(\Lambda_1))+\mathscr{E}(\hat{\omega}(\Lambda_2))+\mathscr{E}(\hat{\omega}(\Lambda_1)|\hat{\omega}(\Lambda_2)).$$

Assume now that we are able to find some subset $\hat{\Omega}^{(\Lambda)}(\Lambda_1,\Lambda_2)$ of $\hat{\Omega}(\Lambda)$ such that

$$\mathscr{V}_\Lambda(\chi_{\hat{\Omega}(\Lambda)_{(\Lambda_1,\Lambda_2)}})>\tfrac{1}{2}, \tag{3.1}$$

and that we are able to find a bound of the following type:

$$\inf_{\hat{\omega}\in\hat{\Omega}(\Lambda)_{(\Lambda_1,\Lambda_2)}}\mathscr{E}(\hat{\omega}(\Lambda_1)|\hat{\omega}(\Lambda_2))\geqslant\Psi^{(\Lambda)}(\Lambda_1,\Lambda_2), \tag{3.2}$$

where $\Psi^{(\Lambda)}(\Lambda_1,\Lambda_2)$ fulfills requirements of the type to be clarified by the consideration below. Using (3.1) and (3.2) we easily conclude that

$$\begin{aligned}\int_{\hat{\Omega}(\Lambda)}&\tilde{\pi}^z_{0,\Lambda}(d\hat{\omega})e^{-\mathscr{E}(\hat{\omega})}\\ &\leqslant e^{\ln 2}e^{-\Psi^{(\Lambda)}(\Lambda_1,\Lambda_2)}\int_{\hat{\Omega}(\Lambda_1)}e^{-\mathscr{E}(\hat{\omega})}\tilde{\pi}^z_{0,\Lambda_1}(d\hat{\omega})\\ &\qquad\times\int_{\hat{\Omega}(\Lambda_2)}\tilde{\pi}^z_{0,\Lambda_2}(d\hat{\omega})e^{-\mathscr{E}(\hat{\omega})},\end{aligned} \tag{3.3}$$

which leads to the extended notion of subadditivity of $P_\Lambda(\mathscr{E})$,

$$\begin{aligned}P_\Lambda(\mathscr{E})\leqslant&\frac{2}{|\Lambda|}-\frac{\Psi^{(\Lambda)}(\Lambda_1,\Lambda_2)}{\Lambda}\\ &+\frac{|\Lambda_1|}{|\Lambda|}P_{\Lambda_1}(\mathscr{E})+\frac{|\Lambda_2|}{|\Lambda|}P_{\Lambda_2}(\mathscr{E}),\end{aligned} \tag{3.4}$$

from which (eventually after iterations) we are often able to conclude the existence of the (unique) thermodynamic limit for $P_\Lambda(\mathscr{E})$ whenever (in a suitable sense) $\Lambda\uparrow R^d$ and $\Psi^\Lambda(\Lambda_1,\Lambda_2)\to 0$.

Thus the core of the whole method consists in finding an appropriate decomposition of the configurational space $\hat{\Omega}(\Lambda)=\hat{\Omega}^{(\Lambda)}(\Lambda_1,\Lambda_2)\cup(\hat{\Omega}(\Lambda)-\hat{\Omega}^\Lambda(\Lambda_1,\Lambda_2))$ on which (3.1) and (3.2) are valid.

Lemma 3.1: Let the interaction $\mathscr{E}$ be superstable and regular. Then there exist constants $\gamma>0$, $\rho\in R$ such that for every $\Delta\subset\Lambda$ and $\hat{\omega}\in\hat{\Omega}(\Lambda)$ we have

$$\rho_\Delta^{(\Lambda)}(\hat{\omega})\leqslant\exp\left[-\left(\sum_{r\in c_\delta(\Delta)}\gamma n_\delta^2(\hat{\omega},r)+\rho\sum_{r\in c_\delta(\Delta)}n_\delta(\hat{\omega},r)\right)\right], \tag{3.5}$$

where $\rho_\Delta^{(\Lambda)}$ are given by (2.17).

The proof of this lemma is presented in Appendix A. Obvious corollaries of this theorem are the following ones.

Lemma 3.2: (1) Let the interaction $\mathscr{E}$ be superstable and regular. Define the set, $(\Delta\subset\Lambda)$,

$$\Omega_N(\Lambda,\Delta)=\left\{\hat{\omega}\in\hat{\Omega}(\Delta)\,\middle|\,\sum_{r\in c_\delta(\Delta)}n_\delta^2(\hat{\omega},r)\geqslant N^2|\Delta|\right\}$$

with N integer.

There exist constants $\gamma'>0$ and $\rho'\in R$ such that

$$r_\Lambda(\{\Omega_N(\Lambda,\Delta)\})\leqslant\exp(-(\gamma' N^2+\rho')|\Delta|).$$

(2) If additionally $\mathscr{E}$ is translationally invariant, then the constants γ, γ', ρ, and ρ' (γ,ρ are from Lemma 3.1) can be chosen independently of Λ.

Lemma 3.3: Let $\mathscr{E}$ be a translational invariant, superstable, and regular interaction. Then there exists $a>0$ such that the set

$$\hat{\Omega}(a)=\{\hat{\omega}\in\hat{\Omega}_\infty\mid\exists_{R:}\ \forall_{|r|>R}\ n^2(\omega,r)\leqslant a\log|r|\},$$

has the property

$$\mathscr{V}_\Lambda\left(\bigcup_{a>0}\hat{\Omega}(a)\equiv\hat{\Omega}^T_\infty\right)=1\ \text{uniformly in the volume }\Lambda.$$

Corollary 3.1: Let $\mathscr{G}^0(\mathscr{E})$ be a set of all Gibbs measures obtained from $\mathscr{V}_\Lambda$ as described in Sec. II. Assume that $\mathscr{E}$ is a translationally invariant, superstable, and regular interaction. Then for any $\mu_\infty\in\mathscr{G}^0(\mathscr{E})$ we have

(1) $\mu_\infty(\hat{\Omega}^T_\infty)=1$,

(2) μ_∞ is $(\hat{\Omega}^T_\infty)$ regular.

We will not elaborate on the proofs of these lemmas having established Lemma 3.1. They follow in more or less standard way from Lemma 3.1 (see Ref. 11, for Lemma 3.3, see especially Ref. 36).

B. General version of the van Hove theorem. The case of pure boundary condition

The main result of this subsection is the following one.

Theorem 3.1: Let $\mathscr{E}$ be a given superstable, strongly regular, and translationally invariant interaction.

Let $(\Lambda_n)_n$ be an arbitrary sequence of δ-polygonal bounded regions in R^d and such that $\Lambda_n\to R^d$ in the sense of van Hove. Then for any $\hat{\Omega}^T_\infty\ni\hat{\omega}$ the unique thermodynamic limit

$$\lim_{n\to\infty}P^{\hat{\omega}}_{\Lambda_n}(\mathscr{E})=P^{\hat{\omega}}_\infty(\mathscr{E}) \tag{3.6}$$

exists and is equal to

$$\lim_{n\to\infty}P^{\hat{\omega}=\phi}_{\Lambda_n}(\mathscr{E})=P_\infty(\mathscr{E}). \tag{3.7}$$

Remark 3.1: In Sec. V we present a much simpler proof of a similar theorem (see Theorem 5.7) than the one presented now. The proof below is an adaptation of the corresponding theorem of Presutti–Lebowitz from Ref. 20.

Proof: Let a regular δ-polygonal region Λ be given. Assume that $\Lambda=\Delta_1\cup\Delta_2$, $\Delta_1\cap\Delta_2=\phi$, where Δ_i are again some δ-polygonal regular sets. Then we define inductively the following sets: Let

$$V^{(\Lambda)}(\Delta_1,\Delta_2)\equiv\max\left\{\sum_{s\in c_\delta(\Delta_2)}\psi_2(|r-s|)\,\middle|\, r\in c_\delta(\Delta_1)\right\};\tag{3.8}$$

then a set $\overline{v_1}(\Delta_1,\Delta_2)$ is defined by

$$\overline{v_1}(\Delta_1,\Delta_2)=\left\{r\in\Delta_1\,\middle|\,\sum_{s\in c_\delta(\Delta_2)}\psi_2(|r-s|)=V_1^{(\Lambda)}(\Delta_1,\Delta_2)\right\}.\tag{3.9}$$

Having defined $V_n^{(\Lambda)}$ and $\overline{v_n}$ we define

$$V_{n+1}^{(\Lambda)}(\Delta_1,\Delta_2)\equiv\max\left\{\sum_{s\in c_\delta(\Delta_2)}\psi_2(|r-s|)\,\middle|\, r\in\Delta_1-\bar v_n\right\}\tag{3.10}$$

and then

$$\bar v_{n+1}(\Delta_1,\Delta_2)=\left\{r\in\Delta_1\,\middle|\,\sum_{s\in c_\delta(\Delta_2)}\psi_2(|r-s|)\geqslant V_{n+1}^{(\Lambda)}(\Delta_1,\Delta_2)\right\}.\tag{3.11}$$

From these definitions it follows that

(i) $V_1^{(\Lambda)}(\Delta_1,\Delta_2)\geqslant V_2^{(\Lambda)}(\Delta_1,\Delta_2)\geqslant\cdots$; (3.12)

(ii) $\bar v_1(\Delta_1,\Delta_2)\subseteq\cdots\subseteq\Delta_1$;

(iii) $|\bar v_n(\Delta_1,\Delta_2)|\geqslant n$ (if $\bar v_{n-1}\neq\Delta$); (3.13)

(iv) for bounded Δ_1 induction is finite and ends when $\bar v_n=\Delta_1$.

Similarly, we can define $V_n^{(\Lambda)}(\Delta_2,\Delta_1)$ and $\bar v_n(\Delta_2,\Delta_1)$. The sets $\bar v_n$ then induce definitions of some subsets of $\hat\Omega(\Lambda)$ of nonzero $\mathscr{V}_\Lambda$ measure

$$\hat\Omega_{N,n}^{(\Lambda)}(\Delta_1,\Delta_2)=\left\{\hat\omega\in\hat\Omega(\Lambda)\,\middle|\,\sum_{r\in\bar v_n(\Delta_1,\Delta_2)}n^2(\omega(\Delta,r))\leqslant N^2|\bar v_n(\Delta_1,\Delta_2)|\right\}\tag{3.14}$$

and similarly for $\Omega_{N,n}^{(\Lambda)}(\Delta_2,\Delta_1)$.

Finally, let us define

$$\Omega_N^{(\Lambda)}(\Delta_1;\Delta_2)=\Big(\bigcap_n\Omega_{N,n}^{(\Lambda)}\Big)\cap\Big(\bigcap_n\Omega_{N,n}^{(\Lambda)}(\Delta_2,\Delta_1)\Big).\tag{3.15}$$

From Lemma 2.1 it follows that

$$\begin{aligned}\mathscr{V}_\Lambda(\{\Omega^{(\Lambda)}(\Delta_1;\Delta_2)\})&\leqslant\sum_j\mathscr{V}_\Lambda(\{\Omega_{N,j}^{(\Lambda)}(\Delta_1;\Delta_2)\})\\&\leqslant\sum_j\exp(-(\gamma'N^2-\delta')j),\end{aligned}\tag{3.16}$$

therefore there exists an integer N_0, independent of Δ_1 and Δ_2 and such that, for any $N>N_0$,

$$\mathscr{V}_\Lambda(\{\Omega_N^{(\Lambda)}(\Delta_1;\Delta_2)\})>\tfrac34.\tag{3.17}$$

Let us note that from the very definition of the sets $\Omega_N^{(\Lambda)}$ it follows that

$$\forall_{\hat\omega\in\hat\Omega_N^{(\Lambda)}(\Delta_1;\Delta_2)}\ |\mathscr{E}(\hat\omega(\Delta_1)|\hat\omega(\Delta_2))|\leqslant N^2\sum_{r\in c_\delta(\Delta_1)}\sum_{s\in c_\delta(\Delta_2)}\psi_2(|r-s|).\tag{3.18}$$

From now on let us choose $N>N_0$. Then we have with $\Delta_1\cup\Delta_2=\Lambda_0$, $\Delta_1\cap\Delta_2=\varnothing$

$$\begin{aligned}Z_\Lambda(\mathscr{E})&=\int_{\Omega_N^{(\Lambda)}(\Delta_1;\Delta_2)}\tilde\pi_{0,\Lambda}^z(\hat\omega(\Delta))e^{-\mathscr{E}(\hat\omega(\Lambda))}\\&\quad+\int_{\Omega(\Lambda)-\Omega_N^{(\Lambda)}(\Delta_1;\Delta_2)}\tilde\pi_{0,\Lambda}^z(d\hat\omega)e^{-\mathscr{E}(\hat\omega)}\\&\leqslant\exp\Big(N^2\sum_{r\in\Delta_1}\sum_{s\in\Delta_2}\psi_2(|r-s|)\Big)\\&\quad\times\int_{\Omega_N^{(\Lambda)}(\Delta_1;\Delta_2)}\tilde\pi_{0,\Lambda}^z(d\hat\omega)\\&\quad\times e^{-\mathscr{E}(\hat\omega(\Delta_1))-\mathscr{E}(\hat\omega(\Delta_2))}+\tfrac12Z_\Lambda(\mathscr{E}),\end{aligned}\tag{3.19}$$

where we have used (3.17) and (3.18). This leads to an upper bound of the form

$$Z_\Lambda(\mathscr{E})\leqslant2Z_{\Delta_1}(\mathscr{E})Z_{\Delta_2}(\mathscr{E})\exp N^2\sum_{r\in\Delta_1}\sum_{s\in\Delta_2}\psi_2(|r-s|).\tag{3.20}$$

To obtain a lower bound on $Z_\Lambda(\mathscr{E})$ we proceed similarly:

$$\begin{aligned}Z_\Lambda(\mathscr{E})&\geqslant\int_{\Omega_N^{(\Lambda)}(\Delta_1;\Delta_2)}\tilde\pi_{0,\Lambda}^z(d\hat\omega)e^{-\mathscr{E}(\hat\omega)}\\&\geqslant\exp\Big(-N^2\sum_{r\in\Delta_1}\sum_{s\in\Delta_2}\psi_2(|r-s|)\Big)\\&\quad\times\int_{\Omega_N^{(\Lambda)}(\Delta_1;\Delta_2)}\tilde\pi_{0,\Lambda}^z(d\hat\omega)e^{-\mathscr{E}(\hat\omega(\Delta_1))}e^{-\mathscr{E}(\hat\omega(\Delta_2))}\\&\geqslant\exp\Big(-N^2\sum_{r\in\Delta_1}\sum_{s\in\Delta_2}\psi(|r-s|)\Big)Z_{\Delta_1}(\mathscr{E})Z_{\Delta_2}(\mathscr{E})\\&\quad-\int_{(\Omega_N^{(\Lambda)}(\Delta_1;\Delta_2))^c}\tilde\pi_{0,\Lambda}^z(d\hat\omega)e^{-\mathscr{E}(\hat\omega(\Delta_1))}e^{-\mathscr{E}(\hat\omega(\Delta_2))}\\&\quad\times\exp\Big(-N^2\sum_{r\in\Delta_1}\sum_{s\in\Delta_2}\psi_2(|r-s|)\Big)\\&\geqslant\frac12Z_{\Delta_1}(\mathscr{E})Z_{\Lambda_2}(\mathscr{E})\\&\quad\times\exp\Big(-N^2\sum_{r\in\Delta_1}\sum_{s\in\Delta_2}\psi_2(|r-s|)\Big).\end{aligned}\tag{3.21}$$

Now we iterate this procedure. Let now Λ be decomposed into n disjoint, regular δ-polygonal regions Δ_i such that $\cup_i\Delta_i=\Lambda$. Then by simple induction we get

$$\begin{aligned}&\prod_{i=1}^n2^{-1}Z_{\Delta_i}\exp\Big(-N^2\sum_{r\in\Delta_i}\sum_{s\notin\Delta_i}\psi_2(|r-s|)\Big)\\&\quad\leqslant Z_\Lambda(\mathscr{E})\leqslant\prod_{i=1}^n2Z_{\Delta_i}\exp\Big(N^2\sum_{r\in\Delta_i}\sum_{s\notin\Delta_i}\psi_2(|r-s|)\Big)\end{aligned}\tag{3.22}$$

and this yields the existence of the limit $\lim_n P_\Lambda(\mathscr{E})=P_\infty(\mathscr{E})$ as $n\to\infty$, where Λ_n is any van Hove type sequence, and gives also the independence of $P_\infty(\mathscr{E})$ of the van Hove sequence (Λ_n) chosen.

Let us now proceed to control the limit

$$\lim_{\Lambda_n\uparrow R^d}P_{\Lambda_n}^{\hat\omega}(\mathscr{E}),\quad\text{for }\hat\omega\in\hat\Omega^T(R^d).$$

For a given regular, δ-polygonal bounded region Λ we define its B boundary as

$$\partial_B(\Lambda) = \left\{ x \notin \Lambda \;\middle|\; \sum_{r \in c_\delta((\Lambda \cup \partial_B(\Lambda)))^c} \psi_2(|r-x|) < 2B \right\}, \quad (3.23)$$

with $B < A$.

This relation does not determine the B boundary uniquely. However, when we have sequence $\{\Lambda_n\}_n$ of regular, δ polygonal, bounded subsets, then we can define a sequence of B boundaries $\{\partial_B(\Lambda_n)\}$ in such a way that

$$\lim_{n\to\infty} \frac{|\partial_B(\Lambda_n)|}{\Lambda_n} = 0, \quad \text{if } \lim_{n\to\infty} \frac{|\partial\Lambda_n|}{|\Lambda_n|} = 0, \quad (3.24)$$

and $\partial_B(\Lambda_n)$ are still δ polygonals.

Then the proof goes essentially as in the empty boundary case. We are looking for appropriate lower and upper bounds for $Z_\Lambda^{\hat\omega}(\mathscr{E})$.

The upper bound: Choose $B < A$, where A is the superstability constant of $\mathscr{E}$ on the scale δ,

$$\begin{aligned} Z_\Lambda^{\hat\omega}(\mathscr{E}) &= \exp \mathscr{E}(\hat\omega(\partial_B(\Lambda))) \\ &\quad\times \exp(-\mathscr{E}(\hat\omega(\partial_B(\Lambda)))Z_\Lambda^{\omega}(\mathscr{E})) \\ &= \exp\mathscr{E}(\hat\omega(\partial_B(\Lambda))) \int_{\hat\Omega(\Lambda)} \hat\pi_{0,\Lambda}^z(d\hat\eta) \\ &\quad\times \exp - \mathscr{E}(\hat\eta \vee \hat\omega(\partial_B(\Lambda))) \\ &\quad\times \exp - \mathscr{E}(\hat\eta | \hat\omega(\partial_B(\Lambda)))\exp(-\mathscr{E}(\hat\eta|\hat\omega(\Lambda^c))). \end{aligned} \quad (3.25)$$

Let $\Lambda' = \Lambda \cup \partial_B(\Lambda)$. Then on $\{\hat\Omega(\Lambda'),\hat{\mathscr{F}}(\Lambda')\}$ we can define a new measure $r_\Lambda^{+,B}$ indexed also by $\hat\omega(\partial_B(\Lambda))$ in the following way:

$$\begin{aligned} r_\Lambda^{+,B}&(d\hat\eta(\Lambda')) \\ &= (Z_\Lambda^+(\hat\omega))^{-1}\hat\pi_{0,\Lambda} \otimes \partial_\omega(\hat\eta(\partial_B(\Lambda))) \\ &\quad\times \exp - \mathscr{E}_{\hat\omega}^+(\hat\eta(\Lambda')), \end{aligned} \quad (3.26)$$

where $\delta_{\hat\omega}$ is the δ measure concentrated at $\hat\eta(\partial_B(\Lambda)) = \hat\omega(\partial_B(\Lambda))$, i.e., for any $\hat{\mathscr{F}}(\partial_B(\Lambda))$ measurable and integrable function F we have

$$\int_{\hat\Omega(\partial_B(\Lambda))} \delta_{\hat\omega}(\eta(\partial_B(\Lambda)))F(\hat\eta) = F(\hat\omega(\partial_B(\Lambda))), \quad (3.27)$$

where $\mathscr{E}_{\hat\omega}^+$ is the new interaction on $\hat\Omega(\Lambda)$ (now Λ, B, and $\hat\omega$ dependent) defined by

$$\begin{aligned} \mathscr{E}_{\hat\omega}^+&(\hat\eta(\Lambda')) \\ &= \mathscr{E}(\hat\eta(\Lambda')) - \mathscr{E}(\hat\eta(\Lambda)|\hat\eta(\partial_B(\Lambda))) \\ &\quad + \mathscr{E}(\hat\eta(\Lambda)|\hat\omega(\Lambda^c)) \\ &= \mathscr{E}(\hat\eta(\Lambda')) + \mathscr{E}(\hat\eta(\Lambda)|\hat\omega(\Lambda'^c)) \end{aligned} \quad (3.28)$$

and

$$\begin{aligned} Z_\Lambda^+(\hat\omega) = \int_{\hat\Omega(\Lambda')} &\hat\pi_{0,\Lambda}^z(d\eta(\Lambda)) \otimes \delta_{\hat\omega}(\eta(\partial_B(\Lambda))) \\ &\times \exp - \mathscr{E}_{\hat\omega}^+(\eta(\Lambda')). \end{aligned} \quad (3.29)$$

It is easy to observe that the new interaction $\mathscr{E}_{\hat\omega}^+$ is superstable and regular on $\hat\Omega(\Lambda')$. For example, let us check the superstability of $\mathscr{E}_{\hat\omega}^+$:

$$\begin{aligned} \mathscr{E}(\hat\eta(\Lambda')) &\geqslant A \sum_{r \in c_\delta(\Lambda')} n_\delta^2(\hat\eta(\Lambda'),r) + C \sum_{r \in c_\delta(\Lambda')} n_\delta(\hat\eta(\Lambda'),r) \\ &\quad - \frac{1}{2} \sum_{r\in c_\delta(\Lambda)} \sum_{s \in c_\delta(\Lambda'^c)} \psi(|r-s|)n_\delta^2(\hat\eta(\Lambda),r) \\ &\quad - \frac{1}{2} \sum_{r\in c_\delta(\Lambda)} \sum_{s \in c_\delta(\Lambda')} \psi(|r-s|)n_\delta^2(s,\omega(\Lambda'^c)) \\ &\geqslant (A-B) \sum_{r \in c_\delta(\Lambda')} n_\delta^2(\eta(\Lambda'),r) \\ &\quad + \sum_{r\in c_\delta(\Lambda')} C(\omega(\Lambda^c))n_\delta(\hat\eta(\Lambda'),r), \end{aligned} \quad (3.30)$$

where C stands for the B of formula (2.5) to avoid the confusion of symbols. We have used, in Eq. (3.30), the assumed regularity of $\mathscr{E}$ and the definition of $\partial_B(\Lambda)$ and B.

We conclude that the superstability estimates of Lemma 2.1 applies to the measure $r_\Lambda^{+,B}$ with some constants δ^+ and ρ^+. For a given Λ' let $\Omega_N(\Lambda';\Lambda'^c)$ be as in formula (3.15). Then for sufficiently large N we have

$$\begin{aligned} \exp[&-\mathscr{E}(\hat\omega(\partial_B(\Lambda)))Z_{\Lambda'}(\hat\omega)] \\ &\leqslant \int_{\hat\Omega_N(\Lambda';\Lambda'^c)} + \int_{\Omega_N^c(\Lambda';\Lambda'^c)} \hat\pi_{0,\Lambda}^z(d\hat\eta) \otimes \delta_{\hat\omega}(\hat\eta(\Lambda'-\Lambda)) \\ &\quad\times \exp[-\mathscr{E}_{\hat\omega}^+(\eta(\Lambda'))] \\ &\leqslant \int_{\hat\Omega_N(\Lambda',\Lambda'^c)} \hat\pi_{0,\Lambda}^z(d\hat\eta) \otimes \delta_{\hat\omega}(\Lambda'-\Lambda)\exp[-\mathscr{E}(\hat\eta(\Lambda'))] \\ &\quad - \mathscr{E}(\hat\eta(\Lambda)|\omega(\Lambda'^c)) + \tfrac{1}{2}Z_\Lambda^+(\hat\omega). \end{aligned} \quad (3.31)$$

But on $\Omega_N(\Lambda';\Lambda'^c)$ we have

$$\begin{aligned} |\mathscr{E}&(\hat\eta(\Lambda)|\hat\omega(\Lambda'^c))| \\ &\leqslant \frac{N^2}{2} \sum_{r\in c_\delta(\Lambda)} \sum_{s\in c_\delta(\Lambda'^c)} (\psi(|r-s|)) \\ &\quad + \frac{1}{2} \sum_{r\in c_\delta(\Lambda)} \sum_{s\in c_\delta(\Lambda'^c)} \psi(|r-s|)n_\delta^2(\hat\omega(\Lambda'^c),s). \end{aligned} \quad (3.32)$$

Therefore, using the regularity of $\mathscr{E}$ we obtain

$$\begin{aligned} Z_\Lambda^{\hat\omega}&(\mathscr{E}) \\ &\leqslant \exp(\mathscr{E}(\hat\omega(\partial_B(\Lambda))) \\ &\quad\times \exp\frac{N^2}{2} \sum_{r\in c_\delta(\Lambda)} \sum_{s\in c_\delta(\Lambda'^c)} \psi(|r-s|) \\ &\quad\times \exp\left(\frac{1}{2} \sum_{r\in c_\delta(\Lambda)} \sum_{s\in c_\delta(\Lambda'^c)} \psi(|r-s|)n_\delta^2(\omega'(\Lambda'^c),s)\right) \\ &\quad\times Z_\Lambda Z_{\hat\omega}' + \frac{1}{2}Z_\Lambda^+(\omega), \end{aligned} \quad (3.33)$$

where

$$\begin{aligned} Z_{\hat\omega}' &= \exp[-\mathscr{E}(\hat\omega(\partial_B(\Lambda)))] \\ &\quad\times \exp\frac{N^2}{2} \sum_{r\in c_\delta(\Lambda)} \sum_{s\in c_\delta(\Lambda'^c)} \psi(|r-s|), \end{aligned} \quad (3.34)$$

from which it follows that

$$Z_{\Lambda}^{\hat{\omega}}(\mathscr{E})\leqslant 2\exp\mathscr{E}(\hat{\omega}(\partial_B(\Lambda))\exp\frac{N^2}{2}$$
$$\times\sum_{r\in c_\delta(\Lambda)}\sum_{s\in c_\delta(\Lambda'^c)}\psi(|r-s|)\exp\frac{1}{2}$$
$$\times\sum_{r\in c_\delta(\Lambda)}\sum_{s\in c_\delta(\Lambda'^c)}\psi(|r-s|)n_\delta^2(\omega(\Lambda'^c),s)Z_{\Lambda}^{(\mathscr{E}')}Z_{\hat{\omega}}'. \tag{3.35}$$

The lower bound: Now let $\Lambda=\Lambda'\cup\partial_B(\Lambda')$ and let χ be a characteristic function of some event in $\hat{\Omega}(\partial_B(\Lambda'))$ to be defined below. Then, we have

$$Z_{\Lambda}^{\hat{\omega}}(\mathscr{E})\geqslant\int_{\hat{\Omega}(\Lambda)}\hat{\pi}_{0,\Lambda'}^{z}(d\hat{\eta}(\Lambda'))\otimes\hat{\pi}_{0,\partial_B(\Lambda')}^{z}(d\eta(\partial_B(\Lambda')))$$
$$\times\exp[-\mathscr{E}(\hat{\eta}(\Lambda))]\exp[-\mathscr{E}(\hat{\eta}(\Lambda)|\hat{\omega}(\Lambda^c))]\cdot\chi$$
$$\geqslant\int_{\hat{\Omega}(\Lambda)}\tilde{\pi}_{0,\Lambda}^{z}(d\hat{\eta}(\Lambda'))\cdot\chi\cdot\exp[-\mathscr{E}(\hat{\eta}(\Lambda'))]$$
$$\times\exp[-\mathscr{E}(\hat{\eta}(\Lambda')|\hat{\omega}(\partial_B(\Lambda')))]$$
$$\times\exp[-\mathscr{E}(\hat{\eta}(\Lambda-\Lambda'))]$$
$$\times\exp[-\mathscr{E}(\hat{\eta}(\Lambda)|\hat{\omega}(\Lambda^c))]. \tag{3.36}$$

Choose now

$$\chi(\hat{\eta}(\Lambda-\Lambda'))=\left\{\hat{\eta}\in\hat{\Omega}(\Lambda-\Lambda')\,\middle|\,\underset{r\in c_\delta(\Lambda-\Lambda')}{\forall}n_\delta^2(\hat{\eta},r)\leqslant N_-\right\}\equiv\chi_{N_-}. \tag{3.37}$$

Then we have

$$Z_{\Lambda}^{\hat{\omega}}(\hat{\omega})\geqslant\int_{\hat{\Omega}(\Lambda)}\tilde{\pi}_{0,\Lambda}^{z*}(d\hat{\eta})\exp[-\mathscr{E}_{\omega}^{-}(\hat{\eta})]\equiv Z_{\Lambda}^{-}(\hat{\omega}), \tag{3.38}$$

where a new measure

$$\tilde{\pi}_{0,\Lambda}^{z*}(d\hat{\eta})=\hat{\pi}_{0,\Lambda}^{z}(d\hat{\eta})\chi_{N_-} \tag{3.39}$$

is defined on $\{\hat{\Omega}(\Lambda),\hat{\mathscr{F}}(\Lambda)\}$ and a new interaction is introduced as

$$\mathscr{E}_{\hat{\omega}}^{-}(\hat{\eta}(\Lambda))=(\mathscr{E}(\hat{\eta}(\Lambda)))+\mathscr{E}(\hat{\eta}(\Lambda)|\hat{\omega}(\Lambda^c))\cdot\chi_N(\eta(\Lambda-\Lambda')). \tag{3.40}$$

The new interaction $\mathscr{E}_{\hat{\omega}}^{-}$ is again superstable and regular, therefore, procedures similar to the above can be applied.

This leads to the following bound:

$$Z_{\Lambda'\cup\partial_B(\Lambda')}^{\hat{\omega}}(\mathscr{E})$$
$$\geqslant\exp\left(-\frac{1}{2}N^2\sum_{r\in c_\delta(\Lambda)}\sum_{s\in c_\delta(\Lambda^c)}\psi(|r-s|)\right)$$
$$\times\exp\left(-N^2\sum_{r\in c_\delta(\Lambda')}\sum_{s\in c_\delta(\Lambda^c)}\psi(|r-s|)\right)$$
$$\times\exp\left(-\frac{N^2}{2}\sum_{r\in c_\delta(\Lambda')}\sum_{s\in c_\delta(\Lambda-\Lambda')}\psi(|r-s|)\right.$$
$$\left.\cdot Z_{\partial_B(\Lambda')}^{-}\cdot Z_{\Lambda}\right), \tag{3.41}$$

taking $N_-=N$ and $N>N_0$, where

$$Z_{\partial_B(\Lambda')}^{-}(\hat{\omega})=\int_{\hat{\Omega}(\partial_B(\Lambda'))}\tilde{\pi}_{0,\partial_B(\Lambda')}^{z}(d\hat{\omega})\cdot\chi_{N_-}(\Lambda-\Lambda')$$
$$\times\exp[-\mathscr{E}\hat{\omega}((\partial_B(\Lambda)))]. \tag{3.42}$$

The proof of the theorem is then completed using the upper bound (3.35), the lower bound (3.41), assumption $\hat{\omega}\in\hat{\Omega}_{\infty}^{T}$, and the assumed decay of ψ (i.e., the strong regularity of $\mathscr{E}$). Q.E.D.

Remark 3.2: It seems to be possible to extend this result to the situation, where (Λ_n) is an arbitrary sequence of bounded subsets such that $\partial\Lambda_n$ are integrable and moreover $\Lambda_n\uparrow R^d$ in the van Hove sense. The idea is the following. For any n, let δ_n be a δ_n covering of Λ_n. Then let $\Lambda_{\delta,n}^{-}=\cup_n\square_{\delta,n}$, where $\square_{\delta,n}\subset\Lambda_n$ and let $\Lambda_{\delta,n}^{+}=\cup_n\square_\delta(n)$, where at least $\square_\delta(n)\cap\Lambda_n\neq\phi$. Then we have two sequences of regular, δ_n-polygonal regions $\{\Lambda_{\delta,n}^{-}\}$ and $\{\Lambda_{\delta,n}^{+}\}$ that both tend to R^d in van Hove sense. Then the "three sequence lemma" should give a proof for the sequence (Λ_n). But this requires us to control the N_0 as a function of δ_n, which seems to be difficult.

Note that we ignored the problem of whether the constant δ^+ used in the proof is bounded from below and uniformly in Λ by a constant greater than zero. This can be proved but requires some additional arguments.

In Sec. V (see Theorem 5.7) of this paper we present proof of a similar theorem that does not require as many technicalities as were necessary here.

IV. SUPERSTABLE INTERACTIONS IN THE EXTENDED SENSE

In this section we introduce a different notion of superstability that seems to be better suited to the case of charged systems than that given by Definition 3.1(b), which seems to be sufficient for the chargeless or extremely non-neutral systems of particles. As we remarked before, with any given interactions $\mathscr{E}$ we can associate a sequence $V=(V_1,V_2,...)$ of K-particle potentials.

Definition 4.1: We say that a given interaction $\mathscr{E}$ is superstable in the extended sense iff the effective new interaction $\mathscr{E}^{\text{eff}}$ defined by

$$e^{-\mathscr{E}^{\text{eff}}(\omega)}=\int_{\Sigma|\hat{\omega}|}\otimes d\lambda(\alpha)e^{-\mathscr{E}(\hat{\omega})} \tag{4.1}$$

is superstable in the standard sense.

As the discussion in our paper[29] shows (see Appendix there), there exist interactions that are stable, but nonsuperstable in the sense of Definition 3.1(b), but which are superstable in the sense of Definition 4.1.

Let us recall our convention from Sec. II that the chargeless configuration corresponding to a given $\hat{\omega}\in\Omega_f(R^d)$: $\hat{\omega}=((x_1,\alpha_1),...,(x_n,\alpha_n))$ will be denoted as $\omega=(x)_n=(x_1,...,x_n)$.

Theorem 4.1: Let the system $\mathscr{E}$, $(\Sigma,d\lambda)$ be two-superstable in the extended sense and strongly two regular. Then for any $\hat{\omega}\in\hat{\Omega}_{\infty}^{T}$, any sequence of bounded, regular δ-polygonal sets ($\delta>0$ arbitrary) $(\Lambda_n)_{n=1}$, tending to R^d in the van Hove sense there exists unique limit

$$\lim_{n\to\infty}P_{\Lambda_n}^{\hat{\omega}}(\mathscr{E})=P_{\infty}^{\hat{\omega}}(\mathscr{E})$$

which does not depend on $\hat{\omega}\in\hat{\Omega}_{\Lambda}^{T}$, the particular choice of the van Hove sequence as above, and is equal to

$$\lim_{n\to\infty} P_{\Lambda_n}(\mathscr{E}) = P_\infty(\mathscr{E}).$$

We start with some simple remarks. From Definition (4.1) it follows that

$$Z_\Lambda(\mathscr{E}) = \int_{\Omega(\Lambda)} \pi^z_{0,\Lambda}(d\omega) e^{-\mathscr{E}^{\text{eff}}(\omega)} = \int_{\hat{\Omega}(\Lambda)} \hat{\pi}^z_{0,\Lambda}(d\hat{\omega}) e^{-\mathscr{E}(\hat{\omega})} \tag{4.2}$$

(Fubini–Tonelli Theorem!).

Let $(\Lambda)_{n=1,\ldots}$ be a given sequence of bounded regular, δ-polygonal subsets of R^d tending to R^d in the van Hove sense. We choose a corresponding sequence of B boundaries as in Sec. III, where B is the superstability constant of $\mathscr{E}^{\text{eff}}$.

The projections corresponding to $\hat{\Omega}^{(\Lambda)}_{N,n}(\Delta_1,\Delta_2)$ on the chargeless part of $\hat{\Omega}_\infty$ we denote according to our convention in the same way with the omission of the superscript $\hat{}$. Using (4.2) and the assumed superstability of $\mathscr{E}^{\text{eff}}$ we get the bounds (3.20) and (3.21), which prove the existence of $P^{(\mathscr{E})}_\infty$ by applying the iteration argument.

The scheme of the proof is the same as that used for proving Theorem 3.1. We are looking for appropriate upper and lower bounds for $Z^{\hat{\omega}}_\Lambda(\mathscr{E})$.

The upper bound:

$$\begin{aligned} Z^{\hat{\omega}}_\Lambda(\mathscr{E}) &= \exp \mathscr{E}(\hat{\omega}(\partial_B(\Lambda))) \times \exp[-\mathscr{E}(\hat{\omega}(\partial_B(\Lambda)))\cdot Z^{\hat{\omega}}_\Lambda(\mathscr{E})] \\ &= \exp \mathscr{E}(\hat{\omega}(\partial_B(\Lambda))) \int_{\hat{\Omega}(\Lambda)} \hat{\pi}^z_0(d\hat{\eta}(\Lambda)) \\ &\quad \times \exp[-\mathscr{E}(\hat{\eta}(\Lambda))] \exp[-\mathscr{E}(\hat{\omega}(\partial_B(\Lambda)))] \\ &\quad \times \exp[-\mathscr{E}(\hat{\eta}(\Lambda)|\hat{\omega}(\partial_B(\Lambda)))] \\ &\quad \times \exp(\mathscr{E}(\hat{\eta}(\Lambda)|\hat{\omega}(\partial_B(\Lambda)))) \\ &\quad \times \exp[-\mathscr{E}(\hat{\eta}(\Lambda)|\hat{\omega}(\Lambda^c))]. \end{aligned} \tag{4.3}$$

For a given $\Lambda' = \Lambda \cup \partial_B(\Lambda)$, let us define a new measure $\hat{r}^{+,B}_\Lambda(d\hat{\eta})$ on the space $\{\hat{\Omega}(\Lambda'),\hat{\mathscr{F}}(\Lambda')\}$ in the following way:

$$\hat{r}^{+,B}_\Lambda(d\hat{\eta}(\Lambda')) = (Z^{\hat{\omega},+,B}_\Lambda)^{-1} \hat{\pi}^z_{0,\Lambda}(d\eta(\Lambda)) \otimes \delta_{\hat{\omega}}(\hat{\eta}(\partial_B(\Lambda))) \times \exp[-\mathscr{E}(\hat{\eta}(\Lambda'))] - \mathscr{E}(\hat{\eta}(\Lambda)|\hat{\omega}(\Lambda^c)),$$

$$Z^{\hat{\omega},+,B}_\Lambda = \int_{\hat{\Omega}(\Lambda')} \hat{\pi}^z_{0,\Lambda}(d\hat{\eta}(\Lambda)) \otimes \delta_{\hat{\omega}}(\hat{\eta}(\partial_B(\Lambda))) \times \exp[-\mathscr{E}(\hat{\eta}(\Lambda'))] - \mathscr{E}(\hat{\eta}(\Lambda)|\hat{\omega}(\Lambda^c)). \tag{4.4}$$

Integrating out over the charge degrees of freedom we get

$$r^{+,B}_\Lambda(d\hat{\eta}(\Lambda')) = (Z^{\hat{\omega},+,B}_{\Lambda'})^{-1} \pi^z_{0,\Lambda}(d\eta(\Lambda)) \otimes \delta_\omega(\eta(\partial_B(\Lambda))) \exp\Big[-\check{\mathscr{E}}^{\text{eff}}_\omega(\hat{\eta}(\Lambda'))\Big], \tag{4.5}$$

where the new partially chargeless energy $\check{\mathscr{E}}^{\text{eff}}_\omega$ is defined in the following way:

$$\check{\mathscr{E}}^{\text{eff}}_\omega(\eta(\Lambda')) = -\log \int_{\Sigma|\hat{\eta}|} \hat{\otimes}\, d\lambda \exp[-\mathscr{E}(\hat{\eta}(\Lambda'))] \times \exp[-\mathscr{E}(\hat{\eta}(\Lambda)|\hat{\omega}(\Lambda^c))], \tag{4.6}$$

where $\hat{\otimes}\, d\lambda$ means integration over the charges that are located inside Λ.

It is not difficult to check that this new partially chargeless energy $\check{\mathscr{E}}^{\text{eff}}_{\hat{\omega}}(\eta(\Lambda'))$ is superstable and regular on $\Omega(\Lambda) \otimes \{\partial_B(\hat{\omega})\}$. This means that it is possible to apply the probability estimates of Sec. III. In this way we have

$$\begin{aligned} &\exp[-\mathscr{E}(\hat{\omega}(\partial_B(\Lambda)))Z^{\hat{\omega}}_\Lambda(\mathscr{E})] \\ &\leqslant \Big(\int_{\Omega_N(\Lambda',\Lambda'^c)} + \int_{\Omega^c_N(\Lambda';\Lambda'^c)}\Big) \pi^z_{0,\Lambda}(d\eta(\Lambda)) \otimes \delta_\omega(\eta(\Lambda-\Lambda') \exp\Big[-\check{\mathscr{E}}^{\text{eff}}_{\hat{\omega}}(\eta(\Lambda))\Big] \\ &\leqslant \int_{\Omega_N(\Lambda',\Lambda'^c)} \pi^z_{0,\Lambda}(d\eta(\Lambda)) \otimes \delta_\omega(\eta(\Lambda-\Lambda')) \times \exp[-\check{\mathscr{E}}^{\text{eff}}_{\hat{\omega}}(\eta(\Lambda))] + \frac{1}{2}(Z^{\hat{\omega},B}_{\Lambda'}) \end{aligned} \tag{4.7}$$

$$\begin{aligned} &\leqslant \exp\Big(N^2 \sum_{r\in c_\delta(\Lambda')} \sum_{s\in c_\delta(\Lambda')} \psi(|r-s|)\Big) \\ &\quad \times \exp\Big(\frac{1}{2} \sum_{r\in c_\delta(\Lambda)} \sum_{s\in c_\delta(\Lambda'^c)} \psi(|r-s|) \times n^2(\omega(\Lambda^c),s) Z_\Lambda(\mathscr{E})\cdot Z'_\omega\Big) + \frac{1}{2} Z^{\hat{\omega},B}_\Lambda \end{aligned} \tag{4.8}$$

from which it follows that

$$\begin{aligned} Z^{\hat{\omega}}_\Lambda(\mathscr{E}) &\leqslant 2 \exp \mathscr{E}(\partial_B(\Lambda)) \times \exp\Big(N^2 \sum_{r\in c_\delta(\Lambda)} \sum_{s\in c_\delta(\Lambda'^c)} \psi(|r-s|)\Big) \\ &\quad \times \exp\Big(\frac{1}{2} \sum_{r\in c_\delta(\Lambda)} \sum_{s\in c_\delta(\Lambda'^c)} \psi(|r-s|) \times n^2(\omega(\Lambda^c),s) Z_\Lambda(\mathscr{E})\cdot Z^{\hat{\omega}}_\Lambda\Big), \end{aligned} \tag{4.9}$$

where

$$Z^{\hat{\omega}}_\Lambda = \exp[-\mathscr{E}(\hat{\omega}(\partial_B(\Lambda)))] \times \exp\Big(N^2 \sum_{r\in c_\delta(\Lambda)} \sum_{s\in c_\delta(\Lambda')} \psi(|r-s|)\Big). \tag{4.10}$$

The lower bound: Let us assume now that the set Λ has the form $\Lambda = \Lambda \cup \partial_B(\Lambda')$ and let χ be a characteristic function of some event in $\hat{\Omega}(\partial_B(\Lambda))$ that will be defined later on. Then we have

$$\begin{aligned} Z^{\hat{\omega}}_\Lambda(\mathscr{E}) &\geqslant \int_{\Omega(\Lambda')} \hat{\pi}^z_{0,\Lambda'}(d\hat{\eta}(\Lambda')) \otimes \hat{\pi}^z_{0,\Lambda-\Lambda'}(\hat{\eta}(\partial_B(\Lambda'))) \\ &\quad \times \exp[-\mathscr{E}(\hat{\eta}(\Lambda))] \exp[-\mathscr{E}(\hat{\eta}(\Lambda)|\hat{\omega}(\Lambda^c))]\cdot\chi \\ &= \int_{\hat{\Omega}(\Lambda')} \hat{\pi}^z_{0,\Lambda}(d\hat{\eta}(\Lambda')) \otimes \hat{\pi}^z_{0,\Lambda-\Lambda'}(d\hat{\eta}(\Lambda-\Lambda'))\cdot\chi \\ &\quad \times \exp[-\mathscr{E}(\hat{\eta}(\Lambda'))] \exp[-\mathscr{E}(\hat{\eta}(\Lambda')|\hat{\eta}(\partial_B(\Lambda')))] \\ &\quad \times \exp[-\mathscr{E}(\hat{\eta}(\partial_B(\Lambda')))] \exp[-\mathscr{E}(\hat{\eta}(\Lambda)|\hat{\omega}(\Lambda^c))]. \end{aligned} \tag{4.11}$$

Integrating over the charge degrees of freedom, using the assumed regularity of $\mathscr{E}$, and choosing

$$\chi(\eta(\partial_B(\Lambda-\Lambda')))$$
$$=\left\{\hat{\omega}\in\hat{\Omega}(\Lambda-\Lambda')\,\middle|\,\forall_{r\in c_\delta(\Lambda-\Lambda')} n^2(\hat{\omega},r)\leqslant M\right\}$$
$$\equiv\chi_M(\Lambda-\Lambda')$$

we obtain

$$Z^{\hat{\omega}}_\Lambda(\mathscr{E})\geqslant\exp\left(-\frac{1}{2}\sum_{r\in c_\delta(\Lambda)}\sum_{s\in c_\delta(\Lambda^c)}\psi(|r-s|)n^2(\hat{\omega}(\Lambda^c),s)\right)$$
$$\times\exp\left(-\frac{1}{2}\sum_{r\in c_\delta(\Lambda-\Lambda')}\sum_{s\in c_\delta(\Lambda)}\psi(|r-s|)N^2\right.$$
$$\left.\times\int\pi^z_{0,\Lambda}(d\eta(\Lambda))\chi_M(\Lambda-\Lambda')\right)$$
$$\times\exp[-\hat{\mathscr{E}}^{\rm eff}_{\hat{\omega}}(\eta(\Lambda))],\tag{4.12}$$

where the new, partially chargeless $\hat{\mathscr{E}}^{\rm eff}_{\hat{\omega}}$ is defined in the following way:

$$\hat{\mathscr{E}}^{\rm eff}_{\hat{\omega}}=-\ln\int_{\Sigma|\eta(\Lambda)|}\overset{\wedge}{\otimes}\,d\lambda\,e^{-\mathscr{E}(\hat{\eta}(\Lambda'))}e^{-\mathscr{E}(\hat{\eta}(\Lambda'-\Lambda))}$$
$$\times\exp\left(-\frac{1}{2}\sum_{r\in c_\delta(\Lambda')}\sum_{s\in c_\delta(\Lambda^c)}\psi(|r-s|)n^2(\hat{\eta},r)\right)$$
$$\times\chi_M(\Lambda-\Lambda')\,.\tag{4.13}$$

Assuming that $\mathscr{E}$ is superstable in the extended sense, it is not difficult to check that $\hat{\mathscr{E}}_{\hat{\omega}}$ is then superstable on $\Omega(\Lambda)$. Therefore, the probability estimates are still applicable. In this way we obtain

$$Z^{\hat{\omega}}_{\Lambda'\cup\partial_B(\Lambda')}(\mathscr{E})$$
$$\geqslant\frac{1}{2}\exp\left(-N^2\sum_{r\in c_\delta(\Lambda)}\sum_{s\in c_\delta(\Lambda^c)}\psi(|r-s|)\right)$$
$$\times\exp\left(-N^2\sum_{r\in c_\delta(\Lambda')}\sum_{s\in c_\delta(\Lambda-\Lambda')}\psi(|r-s|)\right)$$
$$\times Z^{-,\hat{\omega}}_{\partial_B(\Lambda)}\cdot Z_\Lambda(z)\,,\tag{4.14}$$

where

$$Z^{-,\hat{\omega}}_{\partial_B(\Lambda)}=\int_{\Omega(\partial_B(\Lambda))}\pi^z_{0,\Lambda-\Lambda'}(d\eta(\Lambda-\Lambda'))\chi_M(\Lambda-\Lambda')$$
$$\times\exp[-\hat{\mathscr{E}}^{\rm eff}_{\hat{\omega}}(\eta(\partial_B(\Lambda)))]$$
$$\times\exp\left(-\frac{N}{2}\sum_{r\in c_\delta(\Lambda)}\sum_{s\in\partial_B(\Lambda)}\psi(|r-s|)\right).\tag{4.15}$$

Now we can proceed in full analogy with the proof of Theorem 3.1. Q.E.D.

For the sake of the completeness of our exposition we introduce the concept of a general, tempered boundary condition. Let λ^{bc} by a cylindric measure on $\{\hat{\Omega}_\infty,\hat{\mathscr{F}}(R^d),\hat{\mathscr{F}}(\Lambda)\}$. We say that the measure λ is tempered iff $\lambda^{bc}(\Omega^T(R^d))=1$. Let us denote by λ^{bc}_Λ the densities of the projections of the measure λ^{bc} (whenever they exist) on the corresponding σ algebras $\hat{\mathscr{F}}(\Lambda)$.

Definition 4.2: We say that such a measure λ^{bc} realizes a tempered, regular boundary condition iff (i) λ^{bc} is tempered; (ii) for every bounded, regular $\Lambda\subset R^d$ the corresponding densities λ^{bc}_Λ are absolutely continuous with respect to $\hat{\pi}^z_{0,\Lambda}$ and obey the superstability estimates of Lemma 3.1.

Then we have the following generalization of Theorems 3.1 and 4.1.

Theorem 4.2: Let the system $((\Sigma,d\lambda),\mathscr{E})$ be two-superstable, or superstable in the extended case, strongly two-regular. Let λ^{bc} realize the tempered, regular boundary condition.

Let (Λ_n) be an arbitrary sequence of regular δ-polygonal regions that tends to R^d in the van Hove sense.

Define the following free energy density $P^{\lambda^{bc}}_\Lambda(\mathscr{E})$ conditioned by λ^{bc}:

$$P^{\lambda^{bc}}_\Lambda(\mathscr{E})=+\frac{1}{|\Lambda|}\ln Z^{\lambda^{bc}}_\Lambda(\mathscr{E});$$
$$Z^{\lambda^{bc}}_\Lambda(\mathscr{E})=\int_{\hat{\Omega}(\Lambda^c)}\lambda^{bc}(d\omega)Z^{\omega}_\Lambda(\mathscr{E}).$$

Then the unique thermodynamic limit

$$P^{\lambda^{bc}}_\infty(\mathscr{E})=\lim_{n\to\infty}P^{\lambda^{bc}}_{\Lambda_n}(\mathscr{E})$$

exists and is equal to $P_\infty(\mathscr{E})$.

We will not elaborate on the proof of this theorem because it follows rather easily from the definition (proofs) of λ^{bc} from Theorem 3.1 (4.1).

Before closing our discussion of superstable systems let us remark that there is the possibility of treating general (n) superstable, (n)-regular systems [or (n) superstable in the extended sense] by a technique similar to those described above. But this will be the topic of a forthcoming paper.

V. BEYOND SUPERSTABILITY ASSUMPTIONS

A. Basic criterion

In this subsection we formulate a criterion for the nondependence of the infinite volume free-energy density on the typical boundary conditions without using any superstable-type condition on the many-body potential $V=(V_1,V_2,\ldots)$. Instead of this we assume some local bounds on the infinite sequence of the corresponding correlation functions (those corresponding to the empty boundary condition).

We start with the following assumptions. Let $V=(V_1,V_2,\ldots)$ be a given stable and strongly regular many-body potential defined on $\hat{\Omega}$ as in the previous sections. For a finite volume $\Lambda\subset R^d$, let us define the corresponding grand canonical correlation functions $\rho^n_{\Lambda-\Delta}((\underline{x})_n)$ by the following formulas:

$$\rho^\phi_{\Lambda-\Delta}(\underline{x})_n=(Z^\phi_\Lambda(z))^{-1}\sum_{p=0}^\infty\frac{z^{n+p}}{p!}\int_{(\Lambda-\Delta)^{\times p}}d(\underline{y})_p$$
$$\times\exp(-\mathscr{E}_V((\underline{x})_n\vee(\underline{y})_p|(\underline{y})_p\vee(\underline{x})_n))$$
$$\text{for } n=1,2,\ldots,\tag{5.1}$$

where Δ is a unit cube and $\Delta\subset\Lambda$.

Then the probability of finding more than k particles in the given cube $\Delta \subset \Lambda$ that belongs to some configuration $\hat{\omega}$ is given by the following formula:

$$\operatorname{Prob}_{\Lambda}^{\phi}\{\hat{\omega}|n(\hat{\omega},\Delta)>k\} = \sum_{n=k+1}^{\infty}\sum_{p=0}^{\infty}\left(\frac{1}{Z_{\Lambda}^{\phi}(z)}\right)^{-1}\cdot\frac{z^{n+p}}{p!n!} \times\int_{\Delta^{*n}} d(\underline{x})_n \cdot \rho_{\Lambda}^{\phi}((\underline{x})_n). \tag{5.2}$$

From this formula we have

$$\begin{aligned}\operatorname{Prob}_{\Lambda}^{\phi}\{\hat{\omega}|n(\hat{\omega},\Delta)=k\} &= \operatorname{Prob}_{\Lambda}^{\phi}\{\hat{\omega}|n(\hat{\omega},\Delta)>k-1\} \\ &\quad - \operatorname{Prob}_{\Lambda}^{\phi}\{\hat{\omega}|n(\hat{\omega},\Delta)>k\} \\ &= \frac{1}{k!}\int_{\Delta^{*k}} d(\underline{x})_k \rho_{\Lambda-\Delta}^{\phi}(\underline{x})_k.\end{aligned} \tag{5.3}$$

It is clear that

$$0 \leqslant \rho_{\Lambda-\Delta}^{n}(\underline{x})_k \tag{5.4}$$

if $z \geqslant 0$.

Let us formulate our basic assumptions now.

Hypothesis 1: For any $\mu_\infty \in \mathscr{G}^0_{(\Xi)}(\mathscr{E})$, the limiting correlation functions $\{\rho_{\Lambda}^{(\mu_\infty)}((\underline{x})_k)_{k=1,2,\dots}\}$ exist and obey bounds of the form

$$\sup_{(\Lambda)_n}\int_{\Delta^{*k}} d(\underline{x})_k \rho_{\Lambda_n}((\hat{\underline{x}})_k) \leqslant f_k(\Delta) < \infty, \tag{5.5}$$

where $(\Lambda)_n$ is any sequence $(\Lambda_n)_n \in f(R^d)$ determining the Gibbs measure μ_∞ (see Sec. II).

We start with the following lemma.

Lemma 5.1: Let $\mathscr{E}_V$ be a given stable, (two)-regular many body potential. Let us suppose that Hypothesis 1 is valid and moreover,

$$\text{(i)}\quad \forall_{\Delta \subset R^d}\ f(\Delta) \equiv \sum_{k=0}^{\infty}\frac{k^2}{k!}f_k(\Delta) < \infty, \tag{5.6}$$

$$\text{(ii)}\quad \sup_{\substack{\Delta \\ \mu_\infty \in \mathscr{G}^0(\mathscr{E})}} f(\Delta) = f^* < \infty. \tag{5.7}$$

Let us take $(\Lambda_n) \in f(R^d)$ and let us assume that (Λ_n) determines some $\mu_\infty \in \mathscr{G}^0(\mathscr{E})$ (i.e., $\mu_\infty = \omega - \lim_{n\to\infty}\mu_{\Lambda_n}^{\phi}$).

Then there exists a subsequence $(n') \subset (n)$ such that

$$\forall_{\hat{\omega}\in\Xi}:\ \lim_{n'\to\infty}\frac{1}{|\Lambda_{n'}|}R_{\Lambda_{n'}}(\hat{\omega}) = 0, \tag{5.8}$$

where

$$R_{\Lambda_{n'}}(\hat{\omega}) \equiv \mu_{\Lambda_n}^{\hat{\omega}}(|\mathscr{E}_V(-|\hat{\omega}(\Lambda_n^c))|) \tag{5.9}$$

is well defined on Ξ.

Proof: For a fixed n, let Σ_n be a bounded subset of $R^d - \Lambda_n$; let us define then

$$R'_{\Lambda_n}(\hat{\omega}) = R_{\Lambda_n}(\hat{\omega}(\Sigma_n)) \tag{5.10}$$

and let us estimate the $L_1(\mu_\infty)$ norm of the quantity $(1/|\Lambda_n|)R'_{\Lambda_n}(\hat{\omega})$. Using the regularity of $\mathscr{E}$ we have

$$\begin{aligned}\left\|\frac{1}{|\Lambda_n|}R'_{\Lambda_n}(\hat{\omega})\right\|_{L^1(\mu_\infty)} &\leqslant \frac{1}{2|\Lambda_n|}\sum_{r\in c_1(\Lambda_n)}\sum_{s\in c_1(\Sigma_n)}\psi_2(|r-s|)\mu_\infty \circ \mu_{\Lambda_n}^{\hat{\omega}}(n^2(-,r)) \\ &\quad + \frac{1}{2|\Lambda_n|}\sum_{r\in c_1(\Lambda_n)}\sum_{s\in c_1(\Sigma_n)}\psi_2(|r-s|)\mu_\infty(n^2(\hat{\omega}(\Sigma_n),s)) \\ &= (\text{by the DLR equations}) \\ &\quad + \frac{1}{2|\Lambda_n|}\sum_{r\in c_1(\Lambda_n)}\sum_{s\in c_1(\Sigma_n)}\psi_2(|r-s|)\mu_\infty^{\phi}(n^2(-,r)) \\ &\quad + \frac{1}{2|\Lambda_n|}\sum_{r\in c_1(\Lambda_n)}\sum_{s\in c_1(\Sigma_n)}\psi_2(|r-s|)\mu_\infty^{\phi}(n^2(\hat{\omega}(\Sigma_n)),s).\end{aligned} \tag{5.11}$$

Using (5.3) and (5.6) we obtain

$$\begin{aligned}\mu_\infty(n^2(-,r)) &= \sum_{k=0}^{\infty}k^2\operatorname{Prob}_{\infty}^{\phi}\{\hat{\omega}|n(\hat{\omega},r)=k\} \\ &\leqslant \sum_{k=0}^{\infty}\frac{k^2}{k!}\int_{\Delta^{*k}} d(\underline{x})_k \rho_\infty^{(\mu_\infty)}((\underline{x})_k) \leqslant f(\Delta).\end{aligned} \tag{5.12}$$

Substituting the last estimate into (5.11) we obtain

$$\left\|\frac{1}{|\Lambda_n|}R'_{\Lambda_n}(\hat{\omega})\right\|_{L^1(\mu_\infty)} \leqslant O(1)\sum_{s\in c_1(\Sigma_n)}\psi(|s|). \tag{5.13}$$

Suppose now that Σ_n are chosen in such a way that $\|\mathscr{E}(-|\hat{\omega}(\Lambda_n^c)) - \mathscr{E}(-|\hat{\omega}(\Sigma_n))\|_{L^1(\mu_\infty)} < \epsilon$ for some fixed $\epsilon > 0$ and any $n > 0$. The possibility of such a choice follows from the definition of the set Ξ. Using the $2-\epsilon$ argument and the regularity of ψ_2 we conclude that

$$\lim_{n\to\infty}\left\|\frac{1}{|\Lambda_n|}R_{\Lambda_n}(\hat{\omega})\right\|_{L_1(\mu_\infty)} = 0. \tag{5.14}$$

Q.E.D.

In the case of N-regular interactions the following generalization of Lemma 5.1 is valid.

Lemma 5.2: Let $\mathscr{E}_V$ be a stable and (N)-regular interaction in the sense of Definition 2.2(2).

Let us assume that for any $\mu_\infty \in \mathscr{G}^0_{(\Xi)}(\mathscr{E})$ the limiting correlation functions $\{\rho_{\Lambda_n}^{(\mu_\infty)}((\underline{x})_k)\}_{k=1,\dots}$ exist and obey bounds of the forms:

$$\text{(i)}\ \sup_{\substack{(\Lambda_n)_n \\ \Delta}}\int_{\Delta^{*k}} d(\underline{x})_k \rho_{\Lambda_n}((\hat{x})_k) = f_k < \infty; \tag{5.15}$$

$$\text{(ii)}\ \forall_{i=2,\dots,N}\ \sum_{p=0}^{\infty}\frac{p^{k_i}}{p!}f_p = g_i < \infty. \tag{5.16}$$

[Here $(\Lambda_n)_n$ is an $f(R^d)$ sequence determining the measure $\mu_\infty \in \mathscr{G}^0_{\Xi}(\mathscr{E})$.] Then, there exists a subsequence $(n') \subset (n)$ such that

$$\forall_{\hat{\omega}\in\Xi}:\ \lim_{n'\to\infty}\frac{1}{|\Lambda_{n'}|}\mu_{\Lambda_{n'}}^{\hat{\omega}}(|\mathscr{E}_V(-|\hat{\omega}(\Lambda_{n'}^c))|) = 0. \tag{5.17}$$

Proof: Same as the proof of Lemma 5.1.

Remark: It is possible to prove a corresponding variation of Lemma 5.1 even for (∞)-regular interactions, but we do not write down this result here.

Now we are ready to prove the main result of this subsection.

Theorem 5.1: Let $\mathscr{E}_V$ be a given stable, (N)-regular many-body potential. Assume the validity of Hypothesis 1 and the assumptions (i) and (ii) of Lemma 5.2. Let $(\Lambda_n)\in f(R^d)$ be given such that

$$\lim_{n\to\infty} P_{\Lambda_n}(\mathscr{E}) \equiv P_\infty(\mathscr{E}) \tag{5.18}$$

exists and is finite.

Let us denote by $\mathscr{G}^0_{\Xi,(\Lambda_n)}(\mathscr{E})$ the subset of $\mathscr{G}^0_\Xi(\mathscr{E})$ determined by the sequence (Λ_n). Then for any $\hat{\omega}\in\Xi$ the unique limit

$$\lim_{n\to\infty} P^{\hat{\omega}}_{\Lambda_n}(\mathscr{E}) \equiv P^{\hat{\omega}}_\infty(\hat{\mathscr{E}}) \tag{5.19}$$

exists and moreover

$$\forall_{\mu_\infty\in\mathscr{G}^0_{\Xi,(\Lambda_n)}(\mathscr{E})} \quad \mu_\infty\{\hat{\omega}\in\Xi \mid P^{\hat{\omega}}_\infty(\mathscr{E})\neq P_\infty(\mathscr{E})\}=0. \tag{5.20}$$

Proof: We will present the proof only for the case of (two)-regular interactions. Generalization to the general (N)-regular interactions is then a straightforward application of the proof given below. Let us consider the partition functions $Z^{\hat{\omega}}_{\Lambda_n}(z)$ and $Z^\phi_{\Lambda_n}(z)$. Then we have

$$\begin{aligned} Z^{\hat{\omega}}_{\Lambda_n}(z) &= (Z^{\hat{\omega}}_{\Lambda_n}/Z^\phi_{\Lambda_n})\cdot Z^\phi_{\Lambda_n} \\ &= Z^\phi_{\Lambda_n}\cdot\mu^\phi_{\Lambda_n}(\exp[-\mathscr{E}_V(-|\hat{\omega}(\Lambda^c))]). \end{aligned} \tag{5.21}$$

Applying the Jensen inequality we get

$$P^{\hat{\omega}}_{\Lambda_n}(z) - P^\phi_{\Lambda_n}(z) \geqslant -(1/|\Lambda_n|)\mu^\phi_{\Lambda_n}(\mathscr{E}_V(-|\hat{\omega}(\Lambda^c))). \tag{5.22}$$

Similarly, we have

$$\begin{aligned} Z^\phi_\Lambda &= \frac{Z^\phi_\Lambda}{Z^{\hat{\omega}}_\Lambda}\cdot Z^{\hat{\omega}}_\Lambda \\ &= Z^\phi_\Lambda\cdot\mu^{\hat{\omega}}_\Lambda(\exp \mathscr{E}_V(-|\hat{\omega}(\Lambda^c))), \end{aligned} \tag{5.23}$$

from which it follows:

$$P^\phi_{\Lambda_n}(z) - P^{\hat{\omega}}_{\Lambda_n}(z) \geqslant -\frac{1}{|\Lambda_n|}\mu^{\hat{\omega}}_{\Lambda_n}(\mathscr{E}_V(-|\hat{\omega}(\Lambda^c))). \tag{5.24}$$

Now applying Lemma 5.1 we find that there exists a subsequence $(n')\subset(n)$ such that for $\mu^\phi_\infty(z)$ every $\hat{\omega}\in\hat{\Omega}(\mathbb{R}^d)$, where $\mu_\infty\in\mathscr{G}^0_{\Xi,(\Lambda_n)}(\mathscr{E})$:

$$\lim_{n'\to\infty} |\text{rhs of } (4.24)| = 0. \tag{5.25}$$

By similar arguments applied to the right-hand side of (4.22) we prove the existence of a subsequence $(n'')\subset(n)$ such that

$$\lim_{n''\to\infty} |\text{rhs of } (4.22)| = 0.$$

There exists, therefore, a subsequence $(n''')\subset(n)$ such that

$$\lim_{n'''\to\infty} (P^{\hat{\omega}}_{\Lambda_{n'''}}(z) - P^\phi_{\Lambda_{n'''}}(z)) = 0, \tag{5.26}$$

for μ^ϕ_∞ almost every $\hat{\omega}\in\Xi$ with respect to any $\mu^\phi_\infty\in\mathscr{G}^0_{\Xi,(\Lambda_n)}(\mathscr{E})$. Because we have assumed that $\lim_{n\to\infty} P^\phi_{\Lambda_n}$ exists and is finite the proof follows. Q.E.D.

Sometimes we know that the unique limit $\lim_{n\to\infty} P^\phi_{\Lambda_n}(\mathscr{E})$ exists and is independent of the sequence $(\Lambda_n)\in f(R^d)$ chosen. In this situation we have the following result.

Theorem 5.2: Let $\mathscr{E}$ be a stable and N-regular many-body interaction and assume that all the assumptions of Theorem 5.1 are fulfilled and moreover for any $(\Lambda_n)\in f(R^d)$ a unique limit $P^\phi_\infty(\mathscr{E}) = \lim_{n\to\infty} P^\phi_{\Lambda_n}(\mathscr{E})$ exists and is (Λ_n) independent. Then for any $\mu^\phi_\infty\in\mathscr{G}^0_\Xi(\mathscr{E})$:

$$\begin{aligned} \mu^\phi_\infty(\{\hat{\omega}\in\Xi \mid & \lim_{n\to\infty} P^{\hat{\omega}}_{\Lambda_n}(\mathscr{E}) \text{ either does not exist} \\ & \text{or is different than } P^\phi_\infty(\mathscr{E})\}) = 0. \end{aligned}$$

Remark: It is possible to prove a similar theorem, even in the case of ∞-regular interactions.

B. Applications

1. Two-body interactions with positive definite potential

Let us consider systems defined by $\{\mathscr{E}_V,\{\Sigma,\lambda\}\}$, where $V=(0,\widetilde{V}_2,0,\ldots)$, i.e., only two-body forces are present and are described by the two-body potential $\widetilde{V}_2$ that we assume to be a positive definite function on $(R^d\otimes\Sigma)^{\otimes 2}$. Namely, for an arbitrary choice of complex numbers $z_1,\ldots,z_n\in\mathbb{C}^1$, the following inequality is valid:

$$\forall_{n\geqslant 1}: \ \forall_{(\hat{x})_n}: \ \sum_{i,j=1}^{n} z_i\bar{z}_j\widetilde{V}_2(x_i,\alpha_i|x_j,\alpha_j)\geqslant 0. \tag{5.27}$$

For the sake of notational simplicity only, we assume that $\Sigma\subset R^1$ is compact and that the potential energy of interaction between particles located at (x,α) and those located at (x',α') is given by

$$\widetilde{V}_2(x,\alpha|x',\alpha') = \alpha\cdot\alpha' V(x-x'). \tag{5.28}$$

Note that in (5.28) we have assumed translational invariance of $\widetilde{V}_2$. However, all our results are valid for more general interactions as well. Condition (4.27) is then equivalent to the request that the Fourier transform $\hat{V}(k)$ be non-negative.

Throughout this subsection we will assume that $\hat{V}\in L_1(R^d)$ (see, however, Theorem 5.4), which means $V(0)<\infty$. Applying the Riemann–Lebesgue lemma it follows then that $V(x)\to 0$ as $|x|\to\infty$, but for the sake of regularity we have to assume that $V\in L_1(R^d)$.

From the paper of Ruelle[3] we know that, in general, for one-component systems the positive definiteness of V yields the superstability property of the corresponding interaction. But this is not true for charged systems as the following simple example indicates.

Example: Let $\Sigma=\{-1,1\}$ and let $d\lambda(\alpha) = \frac{1}{2}\{\delta(\alpha-1)+\delta(\alpha+1)\}$. For a given configuration $\hat{\omega}=(x_1,\alpha_1,\ldots,x_{2n},\alpha_{2n})$, which we assume to be located inside some unit cube $\Delta\subset R^d$, let us define its total charge as

$$Q(\hat{\omega}) = \sum_{i=1}^{2n}\alpha_i. \tag{5.29}$$

The $Q(\hat{\omega})$ varies in $[-2n,2n]\cap Z$ as α_i varies in Σ. For the configuration $\hat{\omega}_1$ with $\alpha_i=1$ for all $i=1,\ldots,2n$ we have

$$\mathscr{E}(\hat{\omega}_1) = \sum_{i\neq j}^{2n} V(x_i-x_j), \tag{5.30}$$

which obviously fulfills the superstability estimate with the superstability constant A equal to $\hat{V}(0)$.[37] There are several configurations that realize the case $Q(\hat{\omega}) = 0$. One of them is given by choosing $\alpha_i = +1$ for $i = 1,2,...,n$ and $\alpha_i = -1$ for $i = n+1,...,2n$. Let us denote it as $\hat{\omega}_2$. Then we have

$$\mathscr{E}(\hat{\omega}) = \sum_{i \neq j = 1}^{n} V(x_i - x_j) + \sum_{j=n+1}^{2n} V(x_i - x_j) - \sum_{i=1}^{n} \sum_{j=1}^{2n} V(x_i - x_j). \tag{5.31}$$

From now on let us assume that $V_2(x) \geqslant 0$. From the properties of positive definite functions it follows that $V(0) \geqslant V(x)$ for any $x \in R^d$. Thus we have

$$-\sum_{i=1}^{n} \sum_{j=n+1}^{2n} V(x_i - x_j) \geqslant -V(0)n^2, \tag{5.32}$$

which yields then

$$\hat{\mathscr{E}}(\hat{\omega}_2) \geqslant (2A - V(0))n^2. \tag{5.33}$$

Thus the energy of the configuration $\hat{\omega}$ with the total charge $Q(\hat{\omega}) = 0$ fulfills the superstability estimates iff $\hat{V}(0) > \frac{1}{2}V(0)$, and this indicates that in most cases the superstability properties can be violated.

Of course the argument given above is not a proof because we have ignored the contributions from different configurations. As the detailed discussion in our paper[29] shows, extended superstability also does not hold in general for the systems at hand (see Ref. 29, Appendix).

The system $(V_2, \{\Sigma, d\lambda\})$ will be called neutral iff $\Sigma = -\Sigma \equiv \{x \in R^d \mid -x \in \Sigma\}$ and moreover $d\lambda(-\alpha) = d\lambda(\alpha)$. For the neutral systems with the positive-type two-body interaction we know from the paper by Fröhlich and Park[38] the following.

Proposition 5.1 (Ref. 38): Let (Λ_n) be any monotonic sequence of bounded subsets in R^d that tends to R^d monotonously and by inclusion. Assume moreover that the neutral system $\{V, \{\Sigma, d\lambda\}\}$ with the positive-definite two-body potential V is such that $V(0) < \infty$, and $z \geqslant 0$. Then (1) for any $z \geqslant 0$ $\lim_{n \to \infty} p^{\phi}_{\Lambda_n}(z) = p^{\phi}_{\infty}(z)$ exists and is independent of the particularly chosen sequence (Λ_n); and (2) pointwise on $(R^d \times \Sigma)^{\circledast k}$ the thermodynamic limits

$$\lim_{n \to \infty} \rho^{\phi}_{\Lambda_n}((\underline{x})_k) = \rho^{\phi}_{\infty}((\underline{x})_k), \tag{5.34}$$

$$\underset{\substack{\Delta \subset R^d \\ \text{unit cube}}}{\forall} \lim_{n \to \infty} \rho^{\phi}_{\Lambda_n - \Delta}((\underline{x})_k) = \rho^{\phi}_{\infty, \Delta}((\underline{x})_k), \quad k = 1,2,..., \tag{5.35}$$

exist and obey the uniform bounds

$$\sup_{(\underline{x})_k} \rho^{\phi}_{\infty}((\underline{x})_k) \leqslant z^k \exp(k\alpha_*^2/2) V(0), \tag{5.36}$$

and the same for $\rho^{\phi}_{\infty,\Delta}$, where $\alpha_* = \sup\{|\alpha| : \alpha \in \Sigma\}$.

Thus we see that the class of systems considered by Fröhlich and Park fulfills all assumptions needed to prove the following result.

Theorem 5.3: Let $(\Lambda_n)_n$ be an arbitrary sequence of bounded and regular subsets of R^d. Assume that the system $\{V, \Sigma, d\lambda\}$ is neutral, that V is positive definite on $(R^d \times \Sigma)^{\circledast 2}$ and moreover $V(0) < \infty$ and

$$\int_0^{\infty} |x|^{d-1} V(|x|) d|x| < \infty.$$

Then for any $z \geqslant 0$ and for $\mu^{\phi}_{\infty}(z)$ almost every $\hat{\omega} \in \hat{\Omega}(R^d)$, the limit $p^{\hat{\omega}}_{\Lambda_n}(z) = p^{\hat{\omega}}_{\infty}(z)$ exists and is equal to $p_{\infty}(z)$.

Proof: Using Proposition 5.1 we see that all the assumptions of Theorem 5.2 are fulfilled in the case in hand.

In some cases it is possible to relax the assumption that $V(0) < \infty$. For this let us consider a many-component, two-dimensional, neutral Yukawa plasma that corresponds to the assumption that

$$V(x) = \int \frac{\exp - ikx}{k^2 + m^2} dx, \quad \text{with } m > 0. \tag{5.37}$$

Then $V(x)$ has a rapid, exponentially fast decay to zero as $|x| \to \infty$ and $V(x) \to +\infty$ logarithmically slowly as $|x| \downarrow 0$. Let us assume moreover that $\Sigma \subset (-2\sqrt{\pi}, 2\sqrt{\pi})$ and that $d\lambda$ is even and with bounded variation. Then we have the following theorem.

Theorem 5.4: Let $(\Lambda_n)_n$ be an arbitrary sequence of bounded, log-normal subsets of R^2. Let $\{V, \{\Sigma, d\lambda\}\}$ be the two-dimensional, neutral Yukawa plasma as described above. Then for any $z \geqslant 0$, for $\mu^{\phi}_{\infty}(z)$ almost every $\hat{\omega} \in \hat{\Omega}(R^2)$ the unique, thermodynamic limit

$$\lim_{n \to \infty} p^{\hat{\omega}}_{\Lambda_n}(z) = p^{\hat{\omega}}_{\infty}(z)$$

exists and is equal to $\lim_{n \to \infty} p^{\phi}_{\Lambda_n}(z) = p^{\phi}_{\infty}(z)$. Moreover, the limits do not depend on the sequence (Λ_n) chosen.

Because the proof uses some concepts from constructive field theory[39] not involved before, we give it in Appendix C of the present paper.

Let us also mention the following application of the results proved in Sec. V A to the problem considered in Refs. 26 and 29. Let $\{V, (\Sigma, d\lambda)\}$ define a *two*-component, neutral system in which the positive-definite two-body potential V fulfills the following assumptions: $V(0) < \infty$, $\int_0^{\infty} r^{d-1} V(r) dr < \infty$. The value of the chemical activity z_0 is called a regular value iff $p^{\phi}_{\infty}(z)$ is differentiable at $z = z_0$. Note that $p^{\phi}_{\infty}(z)$, being a concave function in z, is differentiable almost everywhere, except at most at a countable number of values.

Let the parameter ζ range over some open region $\mathscr{O}_{\hat{\omega}}$ of the complex plane $\mathbb{C}^1$ containing the points $\zeta = 0$ and $\zeta = i$ in its interior and such that

$$Z^{\hat{\omega}}_{\Lambda}(\zeta, z) \equiv \int_{\hat{\Omega}(\Lambda)} \hat{\pi}^z_0(d\eta) \times \exp + i\zeta \mathscr{E}_V(\hat{\eta}(\Lambda) | \hat{\eta}(\Lambda) \vee (\omega(\Lambda^c))) \tag{5.38}$$

is different from zero as ζ varies in $\mathscr{O}_{\omega}$. Then we define the following correlation functions $\tilde{\rho}^{\hat{\omega}}_{\Lambda}$:

$$\tilde{\rho}^{\hat{\omega}}_{\Lambda}(\zeta, z | (\underline{x})_k) = \left(\sum_{n=0}^{\infty} \frac{z^{n+V}}{n!} \int_{\Lambda} d(\underline{y})_n \exp[-\mathscr{E}_V(\underline{\hat{y}}_n | \underline{\hat{y}}_n)] \times \exp[i\zeta \mathscr{E}_V(\underline{\hat{y}}_n | \underline{\hat{x}}_k)] \exp[-\mathscr{E}_V(\underline{x}_k | \underline{x}_k)] \right) \times \exp[i\zeta \mathscr{E}_V((\hat{\underline{x}})_k \vee (\hat{\underline{y}})_n | \hat{\omega}(\Lambda^c)) \times (Z^{\hat{\omega}}_{\Lambda}(\zeta, z))^{-1}]. \tag{5.39}$$

Hypothesis: Let us assume that the random regions $\mathscr{O}_{\hat{\omega}}$, for μ_∞ almost every $\hat{\omega}$ can be chosen in such a way that

$$\sup_{\Lambda,\xi\in\mathscr{O}_{\hat{\omega}}} |\tilde{\rho}_\Lambda^{\hat{\omega}}(z,\xi\,|\,(\underline{x})_k)| \leqslant C(\hat{\omega}) f(\hat{\omega},(\underline{x})_k), \tag{5.40}$$

where $C(\hat{\omega})$ and $f(\hat{\omega},(\underline{x})_k)$ are some random functions that are finite μ_∞^ϕ almost everywhere on $\hat{\Omega}(R^d)$.

Then we have the following theorem.

Theorem 5.5: Let $\{V,\{\Sigma,d\lambda\}\}$ define a two-component (i.e., card $\Sigma = 2$) neutral system in which V is a positive definite, with $V(0) < \infty$, $\int_0^\infty V(r) r^{-d-1}\,dr < \infty$. Let $z_0 > 0$ be a regular value of the chemical activity z.

Then the set of the limiting grand canonical Gibbs equilibrium measures $\mathscr{G}^T(z_0)$ that have invariant first moments consists of exactly one element $\mu_\infty^\phi(z_0)$.

Because the proof introduces the sine–Gordon transformation and correlation inequalities, topics not involved before, we have to decide to present the details in Appendix B to this paper.

There are some arguments supporting the above hypothesis. We hope to prove this hypothesis by using analogs of the probability estimates (see the remark in Sec. VI and Ref. 29).

2. Stable, M-regular and R-regular many-body interactions

In the paper[40] we have investigated the connection between spectral properties of the infinite-volume Kirkwood–Salsburg operator and the uniqueness of the tempered, Gibbs states corresponding to the general stable and regular (in a suitable sense, see below) interactions $\mathscr{E}$.

Let us recall some definitions from Ref. 40. A sequence of potential $V = (V_1, V_2,\ldots)$ corresponding to a given stable interaction $\mathscr{E}_V$ is called M regular iff $\exists Q > 0$, $P > 0$,

$$\forall_{k\geqslant 1, n\geqslant 1} : \sup_{(\hat{x})_n} \int d(\hat{y})_k |\mathscr{K}(\hat{x}_1|(\hat{x})'_n|(\hat{y})_k)| \leqslant Q^k \quad (M1), \tag{5.41}$$

where the kernels $\mathscr{K}(\cdot|\cdot|\cdot)$ are given by the following formulas:

$$\mathscr{K}(\hat{x}_1|(\hat{x})'_n|\hat{y}_k) = \sum_{i=0}^{k} \binom{k}{i}(-1)^{k-i} \sigma(\hat{x}_1|(\underline{\hat{x}})'_n \vee (\underline{\hat{y}})_i),$$

$$\sigma(\underline{\hat{x}}|(\underline{\hat{y}})_k) = \exp - \beta \sum_{\substack{(\underline{\hat{y}})_q \subset (\underline{\hat{y}})_k \\ q \geqslant 1}} \mathscr{E}_V(\hat{x}|(\hat{y})_q), \tag{5.42}$$

$$(\underline{\hat{x}})'_n = (\hat{x}_2, \hat{x}_3,\ldots,\hat{x}_n)$$

and

$$\sup_{(\hat{x})_n,(\hat{y})_k} \exp - \beta \mathscr{E}^1((\hat{x})_n|(\hat{y})_k) \leqslant P \quad (M2), \tag{5.43}$$

where

$$\mathscr{E}^1 = \sum_{\substack{(\hat{x})_q \ni (\hat{x}_1), \\ (\hat{y})_i \subset (\hat{y})_k}} V_{i+q}((\hat{x})_q \vee (\hat{y})_i).$$

In Ref. 40 we have introduced also the condition corresponding in our notation to (two)-regularity of $\mathscr{E}_V$. In fact, this assumption has been used explicitly only in the proof of Lemma 3.1 and Proposition 3.4 there.

A moment of reflection shows that all we really need to have is some localization property of V that is also fulfilled by the weaker assumption of the N (or even ∞) regularity of $\mathscr{E}_V$. Thus the main result of Ref. 40 expressed there as Theorem under the weaker condition of strong N regularity of $\mathscr{E}_V$ also holds. So in fact we have proved in Ref. 40 the following theorem.

Theorem 5.6: Let $\mathscr{E}_V$ be stable, M regular and N strongly regular (or even ∞ strongly regular) interactions. Then for every value of z, such that $z^{-1} \notin \mathrm{sp}(K_\infty)$, $z > 0$ real, there exists at most one Ξ regular canonical Gibbs measure. Moreover this unique (if it exists!) Gibbs measure is analytic in z in the sense that all its correlation functions are analytic in $z^{-1} \notin \mathrm{sp}(K_\infty)$.

Remarks: Here K_∞ is the Kirkwood–Salsburg operator corresponding to the given system and is defined by formulas (2.18) and (2.20) in Ref. 40, and $\mathrm{sp}(K_\infty)$ means the spectrum of K_∞ in a suitable Banach space used in Ref. 40.

Note that in the case of stable interaction, we have no good control on the support properties of the infinite volume Gibbs measure μ_∞ and this is why we have to assume the existence of the Ξ-regular measure.

Detailed inspection of the proof of Theorem 5.6 [identical (almost) to the proof given in Ref. 40] shows that in fact assuming $z^{-1} \notin \mathrm{sp}\, K_\infty$, $z > 0$, real assumptions of Theorem 5.1 hold. Thus we have the following result.

Theorem 5.7: Let $\mathscr{E}_V$ be a stable, M regular and N strongly regular interaction. Assume that the Ξ-regular canonical Gibbs measures corresponding to $\mathscr{E}_V$ exist. Then for every $\hat{\omega}\in\Xi$, every value of z such that $z \geqslant 0$, $z^{-1} \notin \mathrm{sp}(K_\infty)$ the unique thermodynamic limit $\lim_{n\to\infty} p_{\Lambda_n}^{\hat{\omega}}(\mathscr{E}) = p_\infty(\mathscr{E})$ exists, where (Λ_n) is an arbitrary monotonic sequence of bounded and regular subsets of R^d, which tends to R^d in the sense that for every compact set $\Sigma \subset R^d$ there exists N_0 such that for any $n \geqslant N_0$, $\Lambda_n \supset \Sigma$. Moreover, this limit does not depend on the particular choice of the sequence (Λ_n) with the above properties. Moreover, $p_\infty(\mathscr{E})$ as a function of $\bar{z}^1$ is analytic on the set $C - \mathrm{sp}(K_\infty)$.

3. Superstable and regular interactions again

Here we apply our technique again to the class of (two)-superstable and regular (in the sense defined below) interactions $\mathscr{E}_V$. The conditions involved on $\mathscr{E}_V$ below are a little bit more restrictive than the ones used in Theorem 3.1.

Theorem 5.8: Let the system $\{\mathscr{E}_V, \{\Sigma, d\lambda\}\}$ be such that (i) $\mathscr{E}_V$ is $(M1)$ and $(M2)$ regular, and (ii) $\mathscr{E}_V$ is superstable and (two)-strongly regular. Let (Λ_n) be an arbitrary van Hove type sequence. Then for every $\mu_\infty \in \mathscr{G}_0(\mathscr{E})$,

$$\mu_\infty \left\{ \hat{\omega}\in\hat{\Omega} \,\middle|\, \lim_{n\to\infty} P_{\Lambda_n}^{\hat{\omega}}(\mathscr{E}) \neq \lim_{n\to\infty} P_{\Lambda_n}(\mathscr{E}) \right\} = 0. \tag{5.44}$$

Proof: With our assumptions on $(\mathscr{E}_V, \{\Sigma, d\lambda\})$ all the results of the fundamental paper of Ruelle[11] hold with only minor modifications of his original proofs. We refer also to our paper,[40,41] where some arguments involving Kirkwood–Salsburg, Mayer–Montroll identities for many body potentials have been discussed in detail. Using this observation we will freely quote the results of Ref. 11 for the class of systems under consideration.

By analogs of Proposition 5.2 and Theorem 5.5 of Ref. 11 we have that the set of Gibbs measures $\mathscr{G}_0(\mathscr{E})$ consists exactly of measures that are supported on the set

$$\hat{\Omega}_\infty^{T'} = \bigcup_{N>0} \left\{ \hat{\omega} \in \hat{\Omega} \,\middle|\, \forall_l : \sum_{|r|<l} n^2(\hat{\omega},r) \leqslant N^2(2l+1)^\nu \right\}.$$

But then, applying again the analog of Proposition 5.2a and the analog of Corollary 5.3 of Ref. 11 together with the arguments of Lanford,[36] we see that really all $\mu_\infty \in \mathscr{G}_0(\mathscr{E})$ are supported on the set $\hat{\Omega}_\infty^T$. Thus all grand canonical Gibbs measures $\mu_\infty \in \mathscr{G}_0(\mathscr{E})$ are tempered.

Let $(\Lambda_n)_n$ be an arbitrary van Hove type sequence. Applying then Theorem 4.2 (analog of) extended the van Hove sequences of Ref. 11, we conclude that the unique limit $\lim_{n\to\infty} P_{\Lambda_n}(\mathscr{E})$ exists. By the analog of Corollary 5.3c (Ref. 11) we have the following estimate for the system under consideration:

$$\sup_{(\hat{x})_n \in R^d} |\rho_\infty^{(\mu_\infty)}(\hat{x}_1,...,\hat{x}_n)| \leqslant \xi^n < \infty, \tag{5.45}$$

where ξ is some constant and the bound (5.45) is valid for every $\mu_\infty \in \mathscr{G}_0(\mathscr{E})$.

The proof is completed by applying Theorem 5.1. Q.E.D.

In the case when the interaction $\mathscr{E}_V$ is given in terms of the two-body potential V_2 only, i.e., $V = (0,V_2,0,...)$ and V_2 fulfills some regularity conditions, one can prove the following result.

Theorem 5.8′: Let the system $(\mathscr{E}_V,\{\Sigma,d\lambda\})$ be such that $V=(0,V_2,0,...)$, where V_2 is superstable, $(M1)$ regular, and strongly regular.

Then all the claims of Theorem 5.8 are valid.

Remarks: In the case of a one-component system we can assume that V_2 is lower strongly regular, superstable, and $(M1)$ regular, and the result of Theorem 5.8′ is still valid.

The proof of Theorem 5.8′ is analogous to the corresponding proof of Theorem 5.8, but instead of the Kirkwood–Salsburg K_∞ operator we have to use K_∞ composed with the index-juggling operator $\mathscr{M}$ of Ruelle (see Ref. 13).

In the case of two-body forces $(M1)$ means the standard regularity condition

$$\int_{R^d} |e^{-\beta V_2(x)} - 1| dx < \infty, \tag{5.46}$$

while in the case of the charged system $((\Sigma,\lambda),V_2)$ it means

$$\sup_{\alpha\in\Sigma} \int_{R^d} d\lambda(\beta) |e^{-\alpha\beta V_2(x)} - 1| < \infty. \tag{5.47}$$

VI. ADDITIONAL REMARKS

(1) We see, that, analyzing the question of the independence of the thermodynamic functions on the typical boundary conditions in the case of charged systems several difficulties arise. The most serious one is the question of whether an analog of the fundamental probability estimates of Ruelle is valid. As our example in Sec. V shows, already on the level of the simplest charged systems such natural concepts as superstability or extended superstability are no longer applicable. But there are situations where we have a good enough control of the elements from the set $\mathscr{G}_0(\mathscr{E})$ to analyze the question stated above.

(2) It is an interesting mathematical problem to find conditions on the local behavior of the correlation functions for the corresponding limiting Gibbs measure that give probability estimates on quantities like $n^k(\hat{\omega},r)$, for $k=1,2,...$. It seems to be very likely that certain Tauberian-type theorems[42] should be applicable to obtain such estimates without referring to the notion of superstability. *Assuming* the existence of such estimates one can easily eliminate the hypothesis stated for Theorem 5.5.

(3) It seems to be very interesting to try to extend our results to cover the case of a neutral, two-dimensional Coulomb plasma or even the two-dimensional jellium case[43,44] as there is a hope that the formation of the crystalline structure can be observed on the level of the free energy density.[45] However, our results do not apply to these cases (regularity does not hold). In the one-dimensional Coulomb systems the free energy depends in a nontrivial way on the boundary charges[46] and this leads to the conclusions about the crystalline order[47,48] and the existence of the θ states.[49]

(4) Finally, let us mention one more application of the methods developed in Sec III A to the problems studied in Ref. 26. Using the method of Sec. III A we can extend the validity of Theorem 1 of Ref. 26 to the whole set $\mathscr{G}_r(z,\alpha)$ (for definition, see p. 5 of Ref. 26).

ACKNOWLEDGMENTS

I am happy to express my deep gratitude to Professor S. Albeverio, Professor Ph. Blanchard, and Professor L. Streit for giving me the opportunity to join the Bielefeld-Bochum-Stochastics (BiBoS) project, where this work was done.

APPENDIX A: SUPERSTABILITY ESTIMATES USING POISSON INTEGRATION

The text here conforms with the Ruelle papers, especially Ref. 11, Sec. 2. The notation is that of Ref. 11, Sec. 2, the only difference being that we will work on the configurational space $\hat{\Omega}$, while in Ref. 11 the language of $\cup_m \Omega_m$ space is used.

Definition A1: For $\hat{\omega} \in \hat{\Omega}(\Lambda)$ we define the finite volume correlation function

$$\rho^{(\Lambda)}(\hat{\omega}) = (Z_\Lambda)^{-1} \int_{\hat{\Omega}(\Lambda)} \bar{\pi}_{0,\Lambda}(d\hat{\omega}) e^{-\mathscr{E}(\hat{\omega}\vee\hat{\omega})}. \tag{A1}$$

Lemma A1: There exists a number $\xi > 0$ such that

$$\rho^{(\Lambda)}(\mathring{\omega}) \leqslant \xi^{|\omega|}. \tag{A2}$$

Proof: Let $\omega_1 \in \Omega_1(\Lambda)$. Then we define $\mathring{\omega}' = \mathring{\omega} \vee \omega_1$, assume that for $\mathring{\omega}$ Lemma A1 is valid and let us define

$$\overset{1}{\Omega}_\Lambda(\mathring{\omega}) = \left\{ \hat{\omega} \in \Omega(\Lambda) \,\middle|\, \exists p \sum_{j>p} \sum_{r\in[j]} n^2(\hat{\omega} \vee \hat{\omega},r) \leqslant \psi_j V_j \right\} \tag{A3}$$

and

$$\overset{2}{\Omega}_{\Lambda,q}(\mathring{\omega} \vee \omega_1)$$
$$= \left\{ \omega \in \Omega(\Lambda) | q \text{ is the maximal integer such that } \sum_{r\in[q]} n^2(\omega' \vee \hat{\omega},r) \geqslant \psi_q V_q \right\}. \tag{A4}$$

Then we have

$$\left[\overset{1}{\Omega}_{\Lambda}(\mathring{\omega}')\right]^c \subseteq \bigcup_{q=p}^{\infty} \overset{2}{\Omega}_{\Lambda,q}(\mathring{\omega}\vee\omega_1), \tag{A5}$$

for any choice of $\omega_1 \in \Omega_1(\Lambda)$, and therefore

$$\rho^{(\Lambda)}(\mathring{\omega}') \leqslant \overset{1}{\rho}{}^{(\Lambda)}(\mathring{\omega}') + \sum_{q=p}^{\infty} \rho_q^{(\Lambda)}(\mathring{\omega}')$$

$$\equiv \overset{1}{\rho}{}^{(\Lambda)}(\mathring{\omega}') + \overset{2}{\rho}{}^{(\Lambda)}(\mathring{\omega}'), \tag{A6}$$

where

$$\overset{1}{\rho}{}^{(\Lambda)}(\mathring{\omega}') = (Z_\Lambda)^{-1} \int_{\overset{1}{\Omega}_\Lambda(\omega_0)} e^{-\mathscr{E}(\omega\vee\mathring{\omega}')} \bar{\pi}_{0,\Lambda}(d\omega) \tag{A7}$$

and

$$\overset{2}{\rho}{}_q^{(\Lambda)}(\mathring{\omega}') = (Z_\Lambda)^{-1} \int_{\overset{2}{\Omega}_{\Lambda,q}} e^{-\mathscr{E}(\hat{\omega}\vee\omega')} \bar{\pi}_{0,\Lambda}(d\omega). \tag{A8}$$

Using the lemma of Ref. 11 and choosing Δ in such a way that $\hat{\omega}_1 \in \Omega_1(\Delta)$, we have

$$\overset{1}{\rho}{}^{(\Lambda)}(\mathring{\omega}')$$

$$= (Z_\Lambda)^{-1} \int_{\overset{(1)}{\Omega}_\Lambda(\omega)} e^{-\mathscr{E}(\omega\vee\omega\vee\omega_1)} \bar{\pi}_{0,\Lambda}(d\omega_1)$$

$$= (Z_\Lambda)^{-1} \int_{\overset{1}{\Omega}_\Lambda(\omega)} e^{-\mathscr{E}(\omega_1)} e^{-\mathscr{E}(\mathring{\omega}\vee\omega)}$$

$$\times \exp\left[\frac{1}{2}\sum_{r\in Z^d} \Psi(|r|) + \frac{1}{2}\Psi(0) \sum_{r\in[p]} n^2(\mathring{\omega},\vee\omega,r)\right.$$

$$\left. + \frac{1}{2}\sum_{r\notin[p]} \Psi(|r|) n^2(\mathring{\omega},\vee\omega,r)\right] \bar{\pi}_{0,\Lambda}(d\omega)$$

$$\leqslant (Z_\Lambda^{-1}) e^{\text{const}} \int_{\overset{1}{\Omega}_\Lambda(\mathring{\omega})} e^{-\mathscr{E}(\omega_1)} e^{-\mathscr{E}(\mathring{\omega}\vee\omega)} \bar{\pi}_{0,\Lambda}(d\omega)$$

$$\leqslant (Z_\Lambda)^{-1} e^{\text{const}} \int_{\Omega(\Delta)} \bar{\pi}_{0,\Lambda}(d\omega_1) e^{-\mathscr{E}(\omega_1)}$$

$$\times \int_{\Omega(\Lambda)} e^{-\mathscr{E}(\mathring{\omega}\vee\omega)} \bar{\pi}_{0,\Lambda}(d\omega)$$

$$\leqslant (Z_\Lambda)^{-1} e^{\text{const}} \rho^{(\Lambda)}(\mathring{\omega}). \tag{A9}$$

Now we proceed to estimate $\overset{2}{\rho}{}_q^{(\Lambda)}$. Let us introduce the notation $\Lambda_q = \Lambda \cap [q+1]$, $\Lambda - \Lambda_q = \Lambda_q^c$ and therefore we decompose

$$\mathring{\omega}' \vee \omega = \mathring{\omega}'(\Lambda_q) \vee (\mathring{\omega}' \vee \omega)(\Lambda_q^c) \equiv \omega' \vee \omega''.$$

Let us also denote $N_q \equiv |\mathring{\omega}'(\Lambda_q)| \geqslant 1$ (if we assume $\omega_1 \in [q]$). Then we have

$$\overset{2}{\rho}{}_q^{(\Lambda)}(\mathring{\omega}') = (Z_\Lambda)^{-1} \int_{\overset{2}{\Omega}_q(\omega')} e^{-\mathscr{E}(\omega')} e^{-\mathscr{E}(\omega'')} e^{-\mathscr{E}(\omega'|\omega'')} \bar{\pi}_{0,\Lambda}$$

$$\leqslant \exp(-C\psi_{q+1} V_{q+1}) \cdot Z_\Lambda^{-1}$$

$$\times \int_{\overset{2}{\Omega}_q(\mathring{\omega}')} \mathrm{e}^{-\mathscr{E}(\omega'')} \bar{\pi}_{0,\Lambda_q}(d\omega(\Lambda_q)) \otimes \bar{\pi}_{0,\Lambda_q^c}(\omega(\Lambda_q^c))$$

$$\leqslant \exp(-C\psi_{q+1} V_{q+1})$$

$$\times \exp O(1) |V_{q+1}| \rho^{(\Lambda)}(\mathring{\omega}' - \mathring{\omega}'(\Lambda_q))$$

$$\leqslant \exp(-C\psi_{q+1} V_{q+1})$$

$$\times \exp O(1) |V_{q+1}| \xi^{|\mathring{\omega}' - \omega'(\Lambda_q)|}. \tag{A10}$$

Taking into account that $\psi_q \to \infty$ as $q \to \infty$ we finally obtain the estimate

$$\overset{2}{\rho}{}^{(\Lambda)}(\mathring{\omega} \vee \omega_1) \leqslant E \cdot \xi^{|\mathring{\omega}|} \tag{A11}$$

for suitable $\xi > 0$. Therefore

$$\rho^{(\Lambda)}(\mathring{\omega} \vee \omega_1) \leqslant (1+E)\xi^{|\mathring{\omega}|} \leqslant \xi^{|\mathring{\omega}+1|}, \tag{A12}$$

if we choose $1 + E \leqslant \xi$.

As the reader may have noticed, everything works as in the language of $\cup_n \Lambda^n$ space. In a similar manner we are able to prove the following lemma.

Lemma A2: There exist $\gamma > 0$ and $\delta \in R$ such that (uniformly in Λ if $\mathscr{E}$ is translationally invariant)

$$\rho_\Lambda(\mathring{\omega}) \leqslant \exp \sum_{r\in Z^d} (-\gamma n^2(\mathring{\omega},r) + \delta n(\mathring{\omega},r)). \tag{A13}$$

Lemma A2 immediately leads to the proof of Lemma 3.1.

APPENDIX B

(See also Ref. 50.)

Let us denote by $S(R^d)$ the Schwartz space of fast decreasing C^∞ functions and let us denote by $S'(R^d)$ its strong dual, i.e. the space of tempered distributions.

From the Minlos theorem it follows that the functional

$$S(R^d) \ni f + \Gamma_v^0(f) = \exp[-\tfrac{1}{2}V(f,f)] \tag{B1}$$

is the Fourier transform of probabilistic, Borel, cylindric set Gaussian measure μ_v^0 supported on $S'(R^d)$ and such that

$$\Gamma_v^0(f) = \mu_v^0(e^{i(\varphi(f))}). \tag{B2}$$

Using the quasi-invariance of the measure μ_v^0 under the translation $\varphi \to \varphi + if * V$ we easily derive the following expressions for the corresponding conditioned objects of the grand canonical ensemble in terms of the μ_v^0 integration. They are given by

$$Z_\Lambda^{\hat{\omega}}(z)$$

$$= \mu_v^0\left(\exp z \int_\Lambda dx\colon \cos(\alpha\varphi(x) + i\alpha\mathscr{E}(x|\hat{\omega}(\Lambda^c)))\colon_V\right), \tag{B3}$$

where

$$:\cos(\alpha\varphi(x) + i\alpha\mathscr{E}(x|\hat{\omega}(\Lambda^c))):_V$$

$$= \exp(\alpha^2/2) V(0) \cos(\alpha\varphi(x) + i\alpha\mathscr{E}(x|\hat{\omega}(\Lambda^c))), \tag{B4}$$

$$\underline{dx} = dx \otimes d\lambda,$$

$$\rho_\Lambda^{\hat{\omega}}(z|(\hat{x})_n) = z^n \exp[-i\mathscr{E}((\hat{x})_n|\hat{\omega}(\Lambda^c))]$$

$$\times \mu_\Lambda^{i,\hat{\omega}}\left(\prod_{i=1}^n : e^{i\alpha_i\varphi(x_i)} :_V\right), \tag{B5}$$

where $(\hat{x}_n) = (x_1,\alpha_1;x_2,\alpha_2;\ldots;x_n;\alpha_n)$ and we have denoted

$$\mu_\Lambda^{i,\hat{\omega}}(d\varphi) = (Z_\Lambda^\omega(z))^{-1} \exp\left(z \int_\Lambda \underline{dx}\colon \cos(\alpha\varphi(x)\right.$$

$$\left. + i\mathscr{E}(\hat{x}|\hat{\omega}(\Lambda^c):)\right)\mu_v^0(d\varphi). \tag{B6}$$

From the definition of the set Ξ it follows that

$$\sup_{x\in\Lambda} \mathscr{E}(x|\hat{\omega}(\Lambda^c)) < \infty \tag{B7}$$

for a typical $\hat{\omega}\in\Xi$. By the application of the Jensen inequality the existence of region $\mathscr{O}_{\hat{\omega}}$ containing the points $\zeta = i$ and $\zeta = 0$ follows and such that

$$\inf_{\zeta\in\mathscr{O}_{\hat{\omega}}} |Z^{\hat{\omega}}_{\Lambda}(z)| > 0 . \tag{B8}$$

The random region $\mathscr{O}_{\hat{\omega}}$ might be in general Λ dependent. Equation (B8) enables us to define a new complex measure by

$$\mu^{\zeta,\hat{\omega}}_{\Lambda}(d\varphi) = (Z^{\hat{\omega}}_{\Lambda}(\zeta,z))^{-1} \exp\Big(z \int_{\Lambda} :\cos(\alpha\varphi(x) + \zeta\mathscr{E}(\hat{x}|\hat{\omega}(\Lambda^c)))_V: d\underline{x}\Big)\mu^0_v(d\varphi) , \tag{B9}$$

$$Z^{\hat{\omega}}_{\Lambda}(\zeta,z) = \mu^0_v\Big(\exp\Big(z \int_{\Lambda} d\underline{x}: \cos(\alpha\varphi(x) + \zeta\mathscr{E}(\hat{x}|\hat{\omega}(\Lambda^c)))_V:\Big)\Big) \tag{B10}$$

for at least $\zeta\in\mathscr{O}_{\hat{\omega}}$.

Remark: From the assumed decay of V it follows that choosing arbitary compact set $\Xi\subset R^{nd}$ we can then conclude that

$$\limsup_{\Lambda\uparrow R^d} \sup_{(\hat{x})_n\in\Xi} |\mathscr{E}((\hat{x})_n|\hat{\omega}(\Lambda^c))| = 0 \tag{B11}$$

for a typical $\hat{\omega}\in\hat{\Omega}_{\infty}$.

So the first factor will play no role in controlling the thermodynamic limit of the form

$$\lim_{\Lambda\uparrow R^d} (\chi_{\Xi}(\hat{x})_n \cdot \rho^{\hat{\omega}}_{\Lambda}(z|(\hat{x})_n)).$$

Therefore we introduce the following moments of the measure $\mu^{\zeta,\hat{\omega}}_{\Lambda}(d\varphi)$:

$$C^{n,m}_{\Lambda}(z,\zeta,\hat{\omega}|(\hat{x})_n \vee (\hat{y})_m) \equiv \int_{S'(R^d)} \mu^{\zeta,\hat{\omega}}_{\Lambda}(d\phi) \prod_{i=1}^{n} :\cos\alpha_i\varphi:_V(x_i) \times \prod_{i=1}^{m} :\sin\alpha_i\varphi:_V(y_i) . \tag{B12}$$

Then we have the following lemma.

Lemma B1: Let us assume that $z>0$ and the remaining assumption of Theorem 5.3 holds. Then for a given $\hat{\omega}\in\Xi$, if

$$\forall_{t\in\mathscr{O}_{\hat{\omega}}\cap\{\mathrm{Im}\,\zeta=0\}} \lim_{\Lambda\uparrow R^d} C^{1,0}_{\Lambda}(z,t,\hat{\omega}|\hat{x}) = C^{1,0}_{\infty}(z,\phi|(\alpha,0)) = \lim_{\Lambda\uparrow R^d} C^{1,0}_{\Lambda}(z,\phi|(\hat{x})), \tag{B13}$$

then for any $n,m\geqslant 1$ we have

$$\forall_{t\in\mathscr{O}_{\hat{\omega}}\cap\{\mathrm{Im}\,\zeta=0\}} : \lim_{\Lambda\uparrow R^d} C^{n,m}_{\Lambda}(z,t,\hat{\omega}|(\hat{x})_n \vee (\hat{y})_m) = \lim_{\Lambda\uparrow R^d} C^{n,m}_{\Lambda}(z,\phi|(\hat{x})_n \vee (\hat{y})_m), \tag{B14}$$

where $\lim_{\Lambda\uparrow R^d}$ is taken over sequences $(\Lambda_n)\in f(R^d)$. Additionally the above convergences are uniform on compacts of $R^{d(n+m)}$.

Outline of the proof: The finite volume measure corresponding to the free boundary condition $\hat{\omega} = \phi$ we will denote by μ_{Λ} and by $\langle\ \rangle^{(t,\hat{\omega}),\phi}_{\Lambda}(z)$ we denote the expectations with respect to the tensor product measure $\mu^{\hat{\omega},\zeta=t}_{\Lambda}\otimes\mu_{\Lambda}$. Here we are assuming that t is real. Then the following correlation inequalities hold:

$$\mu_{\Lambda}\Big\langle \prod_{i=1}^{n} :\cos\alpha_i\varphi(x_i):_V; \prod_{j=1}^{m} :\cos\beta_j\varphi(y_j):_V\Big\rangle^T_{\Lambda}(z)\geqslant 0 , \tag{B15}$$

$$\forall_{\delta\in R^1} : \Big\langle\Big(\prod_{j=1}^{n} :\cos\alpha_i\varphi(x_i):_V - \prod_{j=1}^{n} :\cos\alpha_j\varphi(y_j):_V\Big) \times\exp\pm\delta\int_{\Lambda} :\cos\alpha\varphi:(x) :\cos\alpha\varphi'_{\Lambda}:(x)\,dx\Big\rangle^{(t,\hat{\omega}),\phi} \times(z)\geqslant 0 , \tag{B16}$$

$$\forall_{\theta_i\in[0,2\pi)} : \Big|\Big\langle\prod_{i=1}^{n}\cos(\alpha_i\varphi(x_i)+\theta_i)\Big\rangle^{t,\hat{\omega}}_{\Lambda}(z)\Big| \leqslant\Big\langle\prod_{i=1}^{n}\cos\alpha_i\varphi(x_i)\Big\rangle_{\Lambda}(z) , \tag{B17}$$

$$Z_{\Lambda_1\cup\Lambda_2}(z)\geqslant Z_{\Lambda_1}(z)Z_{\Lambda_2}(z) . \tag{B18}$$

From the simple bound $|C^{n,n}_{\Lambda}(z,\phi|(\underline{x})_n \vee (\underline{y})_m)| \leqslant\exp(\alpha^2/2)(n+m)_{V(0)}$ and the correlation inequality (B10) the (pointwise) existence of limits on the right-hand sides of (B8) and (B9) easily follows and moreover these limits are independent of the sequence $(\Lambda_n)_n$ chosen. By standard methods (like Mayer–Montroll identities, see for example, Ref. 11) we can extend this convergence to the locally uniform one. Moreover we also have noted that $C^{1,0}_{\infty}(z,\phi)\geqslant 1$. That the correlation inequalities (B11) and (B12) lead to (B9), assuming (B8) was proved in our paper.[50]

From the correlation inequality (B13) it follows that for every sequence of bounded sets (Λ_n) that tends to R^d monotonously and by inclusion, the unique, finite limit $\lim_{n\to\infty} p_{\Lambda_n}(z) = p_{\infty}(z)$ exists and moreover is (Λ_n) independent. Now we see that all assumptions of Theorem 5.3 are fulfilled. Therefore we see that for a typical $\hat{\omega}\in\Omega_{\infty}$ there exists a limit $\lim_{n\to\infty} p^{\hat{\omega}}_{\Lambda_n}(z) = p_{\infty}(z)$ and is $(\Lambda)_n$ independent. From this, and assuming moreover that $z=z_0>0$ is a regular value of $p_{\infty}(z)$, we have using concavicity of $p_{\infty}(z)$ as a function of z, assumed card $\Sigma = 2$ and translational invariance of the first moment, that for any typical $\hat{\omega}$,

$$C^{1,0}_{\infty}(z,i,\hat{\omega}|(\underline{x})) = C^{1,0}_{\infty}(z,\phi|o) . \tag{B19}$$

Moreover it is not too hard to obtain the same result for some linear neighborhoods of $\zeta = i$ lying in $\mathscr{O}_{\hat{\omega}}$. Using the hypothesis and the Vitali theorem we can extend this to some real neighborhood of the point $\zeta = 0$ lying entirely inside $\mathscr{O}_{\hat{\omega}}$. Summarizing we have

$$\forall_{t\in\mathscr{O}_{\hat{\omega}}\cap\{\zeta|\mathrm{Im}\,\zeta=0\}} (C^{1,0}_{\infty}(z,t,\hat{\omega})|0) = C^{1,0}_{\infty}(z,\phi|0) . \tag{B20}$$

This is a starting point for the induction scheme of Lemma B1. Applying Lemma B1, hypothesis, and Vitali theorem we obtain

$$\forall_{n,m\geq 1} \lim_{\Lambda_n \uparrow R^d} C^{n,m}_{\Lambda_n}(z,i,\hat{\omega}|(\underline{x})_n \vee (\underline{y})_m)$$
$$= \lim_{\Lambda_n \uparrow R^d} \mu_{\Lambda_n}\left(\prod_{i=1}^{n} :\cos\alpha_i\varphi(x_i):_V \times \prod_{j=1}^{m} :\sin\beta_j\varphi(y_i):_V\right) \tag{B21}$$

pointwise on $(R^d \otimes \{-1,1\})^{\otimes n+m}$ and moreover the limit in (B21) did not depend on the sequence (Λ_n) chosen.

This pointwise convergence proven in (B21) can easily be extended to the ω-* convergence in the space $\mathbb{B}_\xi$ (see Ruelle[13] for the definition of $\mathbb{B}_\xi$) and then to a locally uniform one using Mayer–Montroll identities (see Ref. 50). This and the Remark of this appendix completes the proof of Theorem 5.5. Q.E.D.

APPENDIX C: PROOF OF THEOREM 5.4

The basic reference for the construction of the thermodynamic limit with the empty boundary conditions are Refs. 38 and 51. The corresponding functional Dobrushin–Lanford–Ruelle equations have been discussed in our papers.[25,26]

Lemma C1: Let $V(x) = (-\Delta + 1)^{-1}(x), d = 2$, supp $d\lambda \subset (-2\sqrt{\pi}, 2\sqrt{\pi})$. Then for any two disjoint bounded open regions $\Delta_1, \Delta_2 \subset R^2$ the following correlation inequality holds:

$$\forall_{z>0} : Z^{\phi}_{\Delta_1 \cup \Delta_2}(z) \geq Z^{\phi}_{\Delta_1}(z) Z^{\phi}_{\Delta_2}(z) . \tag{C1}$$

Proof: Let us define the $U-V$ regularized covariance V_χ given by

$$V_\chi(x) = \int_{|k|\leq\chi} dk \frac{e^{ikx}}{k^2+m^2} . \tag{C2}$$

Then $V_\chi(0) < \infty$ for any $\chi < \infty$ and for any bounded $|\Lambda| < R^2$, the partition function

$$Z_{\Lambda,\chi}(z) = \mu^0_V\left[\exp\left(z\int_\Lambda d\lambda(\alpha)dx : \cos\alpha\varphi_\chi :_{V_\chi}\right)\right], \tag{C3}$$

where φ_χ is a Gaussian with covariance V_χ, is entire analytic in z. For any $\chi_1 < \chi_2$, we have (for z real)

$$Z_{\Lambda,\chi_1}(z) \leq Z_{\Lambda,\chi_2}(z) \tag{C4}$$

by the conditioning comparison theorems of Ref. 52. Fröhlich[51] proved uniform in χ bound,

$$\sup_\chi |Z_{\Lambda,\chi}(z)| \leq \exp O(1)|\Lambda| . \tag{C5}$$

Applying the previous remark together with the Vitali theorem we have that for any $z \in \mathbb{C}^1$ the unique limit

$$\lim_{\chi\to\infty} Z_{\Lambda,\chi}(z) = Z_\Lambda(z)$$

exists and obeys the bound of the form (C5).

For $\chi < \infty$ the inequality (for z real and non-negative)

$$Z_{\Delta_1\cup\Delta_2,\chi}(z) \geq Z_{\Delta_1,\chi}(z) Z_{\Delta_2,\chi}(z) \tag{C6}$$

has been proved in Ref. 38. Using (C4) and (C5) we obtain (C1). Q.E.D.

From this we have in the standard way the following corollary.

Corollary C2: Let $z \geq 0$ and $\{V,\Sigma,d\lambda\}$ be as in Lemma C1. Let $(\Lambda_n)_n$ be an arbitrary sequence of bounded subsets of R^2 which tends to R^2 monotonously and by inclusion.

Then the unique thermodynamic limit

$$\lim_{n\to\infty} p_{\Lambda_n}(z) = p_\infty(z) \tag{C7}$$

exists and is $(\Lambda_n)_n$ independent.

Let us define a measure on the space $S'(R^2)$,

$$\mu_\Lambda(d\varphi) = (Z_\Lambda(z))^{-1} \times \exp\left(z\int_\Lambda d(\underline{x}) : \cos\alpha\varphi :_V(x)\right)\mu^0_v(d\varphi) . \tag{C8}$$

The correlation inequalities (B11) still hold for the moments of $\Pi{:}\cos\alpha_i\varphi(x_i):$. This can be proved by a similar approximation procedure as in the proof of Lemma C1. From these correlation inequalities it follows that the moments (for $z \geq 0$)

$$C^n((\underline{x})_n) = \int \mu_\Lambda(d\varphi) \prod_{i=1}^{n} \cos\alpha_i\varphi(x_i) :_V \tag{C9}$$

are monotonically increasing in volume Λ, fixing $(\underline{x})_n$ such that $x_1 \neq \cdots \neq x_n$. Moreover, Fröhlich[53] using chessboard estimates has proved the following uniform estimate on the moment generating functional for moments like $C^{n,m}(z,\phi|(\underline{x})_n \vee (\underline{y})_m)$,

$$\forall_{\theta\in[0,2\pi)} : \sup_\Lambda \mu_\Lambda(|\exp\xi : \cos\alpha(\varphi+\theta) : (f)|) \leq \exp O(1)(\|\xi f\|_1 + \|\xi f\|_{p(\alpha)}) \tag{C10}$$

for

$$p(\alpha) > 1/(1-\alpha^2/4\pi) .$$

Taking into account this uniform bound and the correlation inequality (B11) we conclude that there exist unique thermodynamic limits,

$$\lim_{\Lambda\uparrow R^d} C^{n,m}_\Lambda(z,\phi|(\underline{x})_n \vee (\underline{y})_m) = C^{n,m}_\infty(z|(\underline{x})_n \vee (\underline{y})_m) \tag{C11}$$

in the sense of complex measures. This also gives the existence of the infinite volume limit $\mu_\infty = \lim_{\Lambda\uparrow R^d}\mu_\Lambda$.

Taking $g_j f_i = \chi_\Delta$, where Δ is a unit cube, located in R^2 and applying Cauchy's integral formula we find the following estimates from (C10):

$$\left|\mu_\infty\left(\prod_{i=1}^{n} :\cos\alpha_i\varphi :_V(f_i) \prod_{j=1}^{m} :\sin\alpha_j\varphi :_V(g_j)\right)\right| \leq O(1)[(n+m)!]^{1/2} \tag{C12}$$

and this gives also that the corresponding $f_k(\Delta)$ from (5.5) are bounded by

$$f_k(\Delta) \leq O(1)\cdot O(1)^k(k!)^{1/2} \tag{C13}$$

uniformly in Δ.

Thus we have checked that all the assumptions of the Theorem are valid also for the two-dimensional Yukawa, neutral plasma in the monopole region. Q.E.D.

[1]L. Van Hove, "Quelques proprietés générales de l'intégrale de configuration d'un systéme de particules avec interaction," Physica **15**, 951 (1949).
[2]C. N. Yang and T. D. Lee, "Statistical theory of equations of state and phase transitions, I. Theory of condensation." Phys. Rev. **87**, 404 (1952).
[3]D. Ruelle, "Classical statistical mechanics of a system of particles," Helv. Phys. Acta. **36**, 183 (1963).
[4]D. Ruelle, "Statistical mechanics of quantum systems of particles," Helv. Phys. Acta **36**, 789 (1963).
[5]M. E. Fisher, "The free energy of a macroscopic system," Arch. Rat. Mech. Anal. **17**, 377 (1964).
[6]J. L. Lebowitz and O. Penrose, "Rigorous treatment of the Van Der Waals–Maxwell theory of the liquid-vapor transition," J. Math. Phys. **7**, 98 (1966).
[7]E. M. Fisher and D. Ruelle, "The stability of many-particle systems," J. Math. Phys. 7, 260 (1966).
[8]J. Ginibre, "Rigorous lower bound on the compressibility of a classical system," Phys. Lett. A **24**, 223 (1967).
[9]R. L. Dobrushin and R. A. Minlos, "Existence and continuity of pressure in classic statistical physics," Teor. Veroyatn. Ee Primen. **12**, 595 (1967).
[10]M. E. Fisher and J. L. Lebowitz, "Asymptotic free energy of a system with periodic boundary conditions," Commun. Math. Phys. **19**, 251 (1970).
[11]D. Ruelle, "Superstable interactions in classical statistical mechanics," Commun. Math. Phys. **18**, 127 (1970).
[12]D. Ruelle, "Probability estimates for continuous spin systems," Commun. Math. Phys. **50**, 189 (1976).
[13]D. Ruelle, *Statistical Mechanics: Rigorous Results* (Benjamin, New York, 1969).
[14]R. B. Israel, *Convexity in the Theory of Lattice Gases* (Princeton U.P., Princeton, NJ, 1979).
[15]D. Ruelle, "Statistical mechanics of quantum systems of particles," Helv. Phys. Acta **36**, 789 (1963).
[16]D. W. Robinson, "Statistical mechanics of quantum mechanical particles with hard cores," Commun. Math. Phys. **16**, 290 (1970).
[17]D. W. Robinson, *The Thermodynamic Pressure in Quantum Statistical Mechanics, Lecture Notes in Physics*, Vol. 9 (Springer, Berlin, 1971).
[18]S. P. Novikov, "Independence of free energy with respect to boundary conditions," Funct. Anal. Appl. **3**, 58 (1969).
[19]J. Ginibre, *Some Applications of Functional Integrations in Statistical Mechanics*, edited by C. Dewitt and R. Stora (Gordon and Breach, New York, 1971).
[20]J. Fröhlich and Y. M. Park, "Correlation inequalities and the thermodynamic limit for classical and quantum continuous systems," to be published in J. Stat. Phys.
[21]J. L. Lebowitz and E. Presutti, "Statistical mechanics of systems of unbounded spins," Commun. Math. Phys. **50**, 195 (1976).
[22]J. Bellisard and R. Hoegh-Kröhn, "Compactness and the maximal Gibbs state for random Gibbs fields on a lattice," Commun. Math. Phys. **84**, 297 (1982).
[23]F. Guerra, L. Rosen, and B. Simon, "Boundary conditions for the $P(\varphi)_2$ Euclidean quantum field theory," Ann. Inst. H. Poincaré **15**, 233 (1975).
[24]R. Gielerak, "Vacuum energy density in the class of $P(\varphi)_2$ theories independent of (nonclassical) boundary conditions," Rep. Math. Phys. **24**, 145 (1986).
[25]R. Gielerak, "On the DLR equation for the two-dimensional sine–Gordon model," J. Math. Phys. **27**, 2892 (1986).
[26]R. Gielerak, "Trigonometric perturbation of the Gaussian generalized fields, Part II. Large coupling results," Comm. JINR E-8-87-12, Dubna, 1987, also BiBoS preprint 274/87, 1987.
[27]R. L. Dobrushin, "Analyticity of correlation functions in one-dimensional classical systems with slowly decreasing potentials," Commun. Math. Phys. **32**, 269 (1973); M. Campanino, D. Capocaccia, and E. Olivieri, "Analyticity for one-dimensional systems with long-range superstable interactions," J. Stat. Phys. **33**, 437 (1983).
[28]Yu. Suhow, "Random point processes and DLR equations," Commun. Math. Phys. **50**, 113 (1976).
[29]R. Gielerak, "Equilibrium equation for a class of continual systems with superstable, positive definite pair potential" to appear in Theor. Math. Phys. (in Russian).
[30]S. Pogosian, "Asymptotic expansion of the logarithm of the partition function," Commun. Math. Phys. **95**, 227 (1984), and references therein.
[31]E. Collet and F. Dunlop, "Geometric expansion of the boundary free energy of a dilute gas," Commun. Math. Phys. **108**, 1 (1987).
[32]J. Preston, *Gibbs Fields, Lecture Notes in Mathematics*, Vol. 534 (Springer, Berlin, 1976).
[33]H. O. Georgii, *Canonical Gibbs Measures, Lecture Notes in Mathematics*, Vol. 760 (Springer, Berlin, 1979).
[34]R. L. Dobrushin and E. A. Pecherski, "A criterion for uniqueness of Gibbsian fields in noncompact case," in *Probability Theory and Mathematical Statistics, Lecture Notes in Mathematics*, Vol. 1021 (Springer, New York, 1983).
[35]D. Klein, "Convergence of grand canonical Gibbs measures," Commun. Math. Phys. **92**, 295 (1984), and references therein.
[36]O. E. Lanford III, Battelle Conference, 1974.
[37]J. T. Lewis, J. V. Pule, and Ph. de Smedt, "The superstability of pair-potentials of positive type," J. Stat. Phys. **35**, 381 (1984).
[38]J. Fröhlich and Y. M. Park, "Correlation inequalities and the thermodynamic limit for classical and quantum continuous systems," Commun. Math. Phys. **59**, 235 (1978).
[39]J. Glimm and A. Jaffe, *Quantum Physics: A Functional Integral Point of View* (Springer, New York, 1981).
[40]R. Gielerak, "On the Lee–Mayer–Yang hypothesis for a class of continuous systems," BiBoS preprint 261, Bielefeld, 1987.
[41]R. Gielerak, "Existence of the transfer matrix formalism for a class of continuous gases," BiBoS preprint 260, Bielefeld, 1987.
[42]J. A. Shohat and J. D. Tamarkin, *The Problem of Moments* (Interscience, New York, 1943); S. Karlin, and W. Studden, *Tchebycheff Systems: With Applications in Analysis and Statistics* (Interscience, New York, 1966).
[43]S. Johannesen and D. Merlini, "On the thermodynamics of 20 jelium," J. Phys. A **1983**, 1449.
[44]B. Jancovici, "Exact results for the two-dimensional one-component plasma," Phys. Rev. Lett. **46**, 386 (1981).
[45]S. Albeverio, D. Dürr, and D. Merlini, "Remark on the independence of the free energy from crystalline boundary conditions in two-dimensional one-component plasma," J. Stat. Phys. **31**, 389 (1983).
[46]A. Lenard, "Exact statistical mechanics of one-dimensional system with Coulomb forces," J. Math. Phys. **4**, 553 (1963); **2**, 682 (1961).
[47]H. Kunz, "The one-dimensional classical electron gas," Ann. Phys. **85**, 303 (1974).
[48]R. R. Sari, D. Merlini, and D. Calinon, J. Phys. A **9**, 1539 (1976).
[49]M. Aizenmann and J. Fröhlich, "States of one-dimensional Coulomb systems as simple examples of θ-vacua and confinement," J. Stat. Phys. **36**, 163 (1984).
[50]R. Gielerak, "Equilibrium equations for the class of continuous systems with positive-definite two-body interactions," Theor. Math. Phys. **67**, 289 (1986).
[51]J. Fröhlich, "Classical and quantum statistical mechanics in one and two dimensions: Two-component Yukawa and Coulomb system," Commun. Math. Phys. **47**, 233 (1976).
[52]B. Simon, *The $P(\varphi)_2$ Euclidean (Quantum) Field Theory* (Princeton U.P., Princeton, NJ, 1974).
[53]J. Fröhlich, *Phase Transitions, Goldstone Bosons and Topological Superselection Rules* (Schladming, Austria, 1976).

Fractional diffusion and wave equations

W. R. Schneider
Asea Brown Boveri Corporate Research, CH-5405 Baden, Switzerland
W. Wyss
Department of Physics, University of Colorado, Boulder, Colorado 80309

(Received 29 June 1987; accepted for publication 7 September 1988)

Diffusion and wave equations together with appropriate initial condition(s) are rewritten as integrodifferential equations with time derivatives replaced by convolution with $t^{\alpha-1}/\Gamma(\alpha)$, $\alpha = 1,2$, respectively. Fractional diffusion and wave equations are obtained by letting α vary in $(0,1)$ and $(1,2)$, respectively. The corresponding Green's functions are obtained in closed form for arbitrary space dimensions in terms of Fox functions and their properties are exhibited. In particular, it is shown that the Green's function of fractional diffusion is a probability density.

I. INTRODUCTION

Vaguely formulated, our objective is to replace first- and second-order time derivatives in the diffusion and wave equations, respectively, by a fractional derivative of order α with $0<\alpha<2$. For $0<\alpha<1$ we shall speak of a fractional diffusion equation and for $1<\alpha<2$ of a fractional wave equation.

To give a precise meaning to these terms we recall the following elementary fact: Let y be a continuous function on R_+ with values in an unspecified topological (Hausdorff) vector space. Define

$$y_0(t) = y(t)\,, \tag{1.1a}$$

$$y_m(t) = \frac{1}{(m-1)!}\int_0^t d\tau (t-\tau)^{m-1} y(\tau), \quad m \in N. \tag{1.1b}$$

Then y_m is m-times continuously differentiable and its k th derivative is given by

$$y_m^{(k)}(t) = y_{m-k}(t), \quad 0 \leqslant k \leqslant m\,. \tag{1.2}$$

Hence the initial value problem

$$z^{(m)}(t) = y(t)\,, \tag{1.3a}$$

$$z^{(k)}(0) = c_k, \quad 0 \leqslant k \leqslant m-1, \quad m \in N, \tag{1.3b}$$

has the (unique) solution

$$z(t) = \sum_{k=0}^{m-1} \frac{c_k}{k!} t^k + \frac{1}{(m-1)!}\int_0^t d\tau (t-\tau)^{m-1} y(\tau)\,. \tag{1.4}$$

Note that (1.4) is equivalent to (1.3a) and (1.3b). However, in the formulation (1.4) the restriction $m \in N$ need not be maintained. One possible generalization is

$$z(t) = \sum_{k=0}^{m-1} \frac{c_k}{k!} t^k + \frac{1}{\Gamma(\alpha)}\int_0^t d\tau (t-\tau)^{\alpha-1} y(\tau)\,, \tag{1.5a}$$

with

$$m-1 < \alpha \leqslant m, \quad m \in N\,. \tag{1.5b}$$

It is easily verified that z has at least $m-1$ continuous derivatives satisfying (1.3b).

If the function y in (1.3a) is not given explicitly but instead is related to z via

$$y = \Phi(z)\,, \tag{1.6}$$

where Φ maps the set of continuous functions on R_+ into itself, (1.4) and (1.5a), (1.5b) become integral equations with (1.4) being equivalent to (1.3a) and (1.3b). If the map Φ is linear then the integral equations are linear. For two particular linear maps Φ the associated integral equations (1.5a) and (1.5b) will be solved explicitly.

The simple example

$$\Phi(z) = -\lambda z, \quad \lambda \in R_+\,, \tag{1.7}$$

serves mainly as preparatory training, but is also interesting in itself. It will be solved at the end of this section. In the second example we deal with maps

$$z(t)\colon D \to R, \quad t \in R_+\,, \tag{1.8}$$

where D is a domain in R^n. We shall use the notation

$$[z(t)](\mathbf{x}) = u(\mathbf{x},t)\,, \tag{1.9}$$

i.e., u is a map from $D \times R_+$ into R. With the choice

$$[\Phi(z)(t)](\mathbf{x}) = \Delta u(\mathbf{x},t) = \sum_{i=1}^{n} \frac{\partial^2 u(\mathbf{x},t)}{\partial x_i^2} \tag{1.10}$$

for the map Φ we obtain from (1.3a) and (1.3b) the initial value problem

$$\frac{\partial^m u}{\partial t^m} = \Delta u\,, \tag{1.11a}$$

$$\frac{\partial^k u}{\partial t^k}(\mathbf{x},0) = f_k(\mathbf{x}), \quad 0 \leqslant k \leqslant m-1\,, \tag{1.11b}$$

i.e., the diffusion ($m=1$) and the wave equation ($m=2$), respectively. Conveniently, the initial data c_k have renamed f_k, being maps from D into R.

Reformulating (1.11a) and (1.11b) via (1.4) as the integral equation with incorporated initial data and generalizing via (1.5a) and (1.5b), we obtain

$$u(\mathbf{x},t) = \sum_{k=0}^{m-1} \frac{1}{k!} f_k(\mathbf{x}) t^k + \frac{1}{\Gamma(\alpha)}\int_0^t d\tau (t-\tau)^{\alpha-1} \Delta u(\mathbf{x},\tau)\,, \tag{1.12a}$$

with

$$m-1 < \alpha \leqslant m, \quad m = 1,2\,. \tag{1.12b}$$

Note the restricted range of m [it makes the factorials in (1.12a) actually superfluous]. We call (1.12a) fractional diffusion ($m=1$) or fractional wave equation ($m=2$).

 0022-2488/89/010134-11$02.50

In Sec. II it is shown that the solutions of these fractional initial value problems for the domain $D = R^n$ are given by

$$u(\mathbf{x},t) = \sum_{k=0}^{m-1} \int d^n y\, G_k^\alpha(|\mathbf{x}-\mathbf{y}|,t) f_k(\mathbf{y}) . \qquad (1.13)$$

Explicit expressions for the Green's functions G_0^α and G_1^α in terms of Fox functions are derived. By determining their space-time Fourier transforms it is shown that they are tempered distributions. For the special values $\alpha = 1$ and $\alpha = 2$ the classical results are recovered.

In Sec. III it is shown that G_0^α is a probability density for $0<\alpha\leqslant 1$, i.e., one of the basic features of ordinary diffusion ($\alpha = 1$) carries over to fractional diffusion. All moments of these probability laws are finite and explicitly determined. In particular, we find for the mean square displacement

$$\int d^n x\, G_0^\alpha(|\mathbf{x}|,t)|\mathbf{x}|^2 = \frac{2n}{\Gamma(1+\alpha)} t^\alpha , \qquad (1.14)$$

i.e., subdiffusive behavior. The one-dimensional case is particular insofar as G_0^α remains a probability density for $1<\alpha\leqslant 2$, whereas it becomes indefinite in higher dimensions. Another peculiarity of $n = 1$ is the representability of G_0^α in terms of one-sided Lévy densities.

In Sec. IV we shortly discuss the case of a half-space $D = R^{n-1}\times R_+$ with various boundary conditions on the hyperplane $x_n = 0$. Two special one-dimensional problems are solved explicitly, thus recovering and simplifying earlier results.[1]

In Sec. V we summarize our results and indicate a physical application. In addition, we point out an extension of the present work.

We end this section by a short discussion of the solution of (1.5a) and (1.5b) when Φ takes the simple form (1.7), i.e., $y = -\lambda z$, $\lambda\in R_+$. The Laplace and Mellin transforms of a function ϕ on R_+ are defined by

$$\tilde\phi(p) = \int_0^\infty dt\, e^{-pt}\phi(t) \qquad (1.15)$$

and by

$$\hat\phi(s) = \int_0^\infty dt\, t^{s-1}\phi(t) , \qquad (1.16)$$

respectively. They are related to each other by

$$\hat\phi(s) = \frac{1}{\Gamma(1-s)} \int_0^\infty dp\, p^{-s}\tilde\phi(p) . \qquad (1.17)$$

Laplace transforming (1.5a) yields

$$\tilde z(p) = \sum_{k=0}^{m-1} c_k (p^\alpha + \lambda)^{-1} p^{\alpha-k-1} . \qquad (1.18)$$

With (1.17) we obtain from (1.18)

$$\hat z(s) = \alpha^{-1} \sum_{k=0}^{m-1} c_k \lambda^{-(k+s)/\alpha} \times \frac{\Gamma((k+s)/\alpha)\Gamma(1-(k+s)/\alpha)}{\Gamma(1-s)} . \qquad (1.19)$$

Recalling the definition of Fox functions[1-4] in terms of their Mellin transform (2.16) we obtain

$$z(t) = \alpha^{-1} \sum_{k=0}^{m-1} c_k t^k H_{12}^{11}\left(\lambda^{1/\alpha} t \,\middle|\, {}^{(0,1/\alpha)}_{(0,1/\alpha),(-k,1)}\right) . \qquad (1.20)$$

Inserting the series representation of H_{12}^{11} yields

$$z(t) = \sum_{k=0}^{m-1} \sum_{j=0}^{\infty} \frac{c_k}{\Gamma(1+k+j\alpha)} (-\lambda)^j t^{k+j\alpha} . \qquad (1.21)$$

It is easily verified that (1.21) satisfies (1.5a) and (1.5b) with $y = -\lambda z$.

II. FRACTIONAL DIFFUSION AND WAVE EQUATIONS

The aim is to solve the fractional diffusion ($m = 1$, $0<\alpha\leqslant 1$) or wave ($m=2$, $1<\alpha\leqslant 2$) equation (1.12a), dropping the factorials according to the remark after (1.12b),

$$u(\mathbf{x},t) = \sum_{k=0}^{m-1} f_k(\mathbf{x}) t^k + \frac{1}{\Gamma(\alpha)} \int_0^t d\tau (t-\tau)^{\alpha-1} \Delta u(\mathbf{x},t) . \qquad (2.1)$$

Here, f_k, $0\leqslant k\leqslant m-1$, are the initial data, i.e.,

$$\frac{\partial^k u}{\partial t^k}(\mathbf{x},0) = f_k(\mathbf{x}), \quad 0\leqslant k\leqslant m-1 . \qquad (2.2)$$

Applying the Laplace transform (1.15) with respect to time t yields

$$\tilde u(\mathbf{x},p) = \sum_{k=0}^{m-1} f_k(\mathbf{x}) p^{-k-1} + p^{-\alpha}\Delta\tilde u(\mathbf{x},p), \quad p>0 , \qquad (2.3)$$

or

$$\Delta\tilde u(\mathbf{x},p) - p^\alpha \tilde u(\mathbf{x},p) = -\sum_{k=0}^{m-1} f_k(\mathbf{x}) p^{\alpha-k-1} . \qquad (2.4)$$

This equation is of the form

$$\Delta v(\mathbf{x}) - \lambda^2 v(\mathbf{x}) = -g(\mathbf{x}), \quad \lambda^2>0 , \qquad (2.5)$$

where the source term g is given. The general solution of (2.5) is

$$v(\mathbf{x}) = w(\mathbf{x}) + \int d^n y\, k(|\mathbf{x}-\mathbf{y}|,\lambda) g(\mathbf{v}) . \qquad (2.6)$$

Here, w is an arbitrary solution of the homogeneous equation

$$\Delta w(\mathbf{x}) - \lambda^2 w(\mathbf{x}) = 0 . \qquad (2.7)$$

The kernel $k(r,\lambda)$ in (2.6) is given by (see Appendix A)

$$k(r,\lambda) = (2\pi)^{-n/2}(r/\lambda)^{1-n/2} K_{1-n/2}(\lambda r) , \qquad (2.8)$$

where K_σ denotes the modified Bessel function of the second kind.

Setting $w = 0$ (which will be justified in Appendix A) we obtain from (2.4)–(2.8)

$$\tilde u(\mathbf{x},p) = \sum_{k=0}^{m-1} \int d^n y\, \tilde G_k^\alpha(|\mathbf{x}-\mathbf{y}|,p) f_k(\mathbf{y}) , \qquad (2.9)$$

with

$$\tilde G_k^\alpha(r,p) = k(r,p^{\alpha/2}) p^{\alpha-k-1} . \qquad (2.10)$$

A direct transition to the time domain (i.e., inverting the Laplace transform) does not seem to be feasible. This difficulty is circumvented by passing through the intermediate step of the Mellin transform (1.16), connected with the Laplace transform by (1.17). Thus we obtain

$$\hat{\tilde G}{}_k^\alpha(r,s) = \frac{1}{\Gamma(1-s)} \int_0^\infty dp\, k(r,p^{\alpha/2}) p^{\alpha-k-s-1} . \qquad (2.11)$$

Inserting (2.8) and using[5]

$$\tilde{K}_\sigma(s) = 2^{s-2}\Gamma((s-\sigma)/2)\Gamma((s+\sigma)/2) \tag{2.12}$$

yield

$$\tilde{G}_k^\alpha(r,s) = \alpha^{-1}\pi^{-n/2}2^{-2(k+s)/\alpha}r^{-n+2(k+s)/\alpha} \cdot [1/\Gamma(1-s)]\Gamma(1-(k+s)/\alpha) \times \Gamma(n/2-(k+s)/\alpha). \tag{2.13}$$

Inverting the Mellin transform leads to

$$G_k^\alpha(r,t) = \pi^{-n/2}2^{-1-2k/\alpha}r^{-n+2k/\alpha} \cdot H_{12}^{20}\left(\tfrac{1}{2}rt^{-\alpha/2}\Big|{}^{(1,\alpha/2)}_{(n/2-k/\alpha,1/2),(1-k/\alpha,1/2)}\right). \tag{2.14}$$

Here,

$$H_{PQ}^{MN}(z) = H_{PQ}^{MN}\left(z\Big|{}^{(a_j,\alpha_j)_{j=1,\dots,P}}_{(b_j,\beta_j)_{j=1,\dots,Q}}\right) \tag{2.15}$$

denotes the general Fox or H function,[1-4] characterized by its Mellin transform

$$\hat{H}_{PQ}^{MN}(s) = A(s)B(s)/(C(s)D(s)) \tag{2.16}$$

with

$$\begin{aligned} A(s) &= \prod_{j=1}^{M}\Gamma(b_j+\beta_j s), \\ B(s) &= \prod_{j=1}^{N}\Gamma(1-a_j-\alpha_j s), \\ C(s) &= \prod_{j=M+1}^{Q}\Gamma(1-b_j-\beta_j s), \\ D(s) &= \prod_{j=N+1}^{P}\Gamma(a_j+\alpha_j s). \end{aligned} \tag{2.17}$$

The integers M, N, P, Q are supposed to satisfy

$$0 \leqslant N \leqslant P, \quad 1 \leqslant M \leqslant Q; \tag{2.18}$$

empty products in (2.17) are set equal to unity. The parameters a_j, $1 \leqslant j \leqslant P$, and b_j, $1 \leqslant j \leqslant Q$, are arbitrary complex numbers, whereas α_j, $1 \leqslant j \leqslant P$, and β_j, $1 \leqslant j \leqslant Q$, are positive. The sets of poles of A and B, respectively, are supposed to be disjoint. Further, it is assumed that

$$\delta = \sum_{j=1}^{Q}\beta_j - \sum_{j=1}^{P}\alpha_j > 0; \tag{2.19}$$

in Ref. 2 the case $\delta = 0$ is also treated.

For the particular Fox function in (2.14) the condition (2.19) is equivalent to

$$\alpha < 2. \tag{2.20}$$

Under the above conditions $H_{PQ}^{MN}(z)$ is an analytic function for $z \neq 0$, in general multiple valued (one valued on the Riemann surface of $\log z$). It is given by

$$H_{PQ}^{MN}(z) = -\sum_{s\in P(A)}\operatorname{res}\left(\frac{A(-s)B(-s)}{C(-s)D(-s)}z^s\right) \tag{2.21}$$

with res standing for residuum and where $P(A)$ is the set of poles of $A(-s)$. Replacing $P(A)$ by $P(B)$ and changing the sign in (2.21) yields an asymptotic expansion of H_{PQ}^{MN}, $N \neq 0$, for large $|z|$, uniformly on every closed subsector of

$$|\arg z| < \frac{\pi}{2}\left\{\sum_1^M \beta_j - \sum_{M+1}^{Q}\beta_j + \sum_1^N \alpha_j - \sum_{N+1}^{P}\alpha_j\right\}, \tag{2.22}$$

where the quantity in curly brackets is assumed to be positive. For $N = 0$ the asymptotic behavior becomes exponentially small. We have

$$H_{PQ}^{M0}(z) \sim F z^{\gamma/\delta}\exp(-E^{1/\delta}\delta z^{1/\delta}), \tag{2.23}$$

for large $|z|$, uniformly on every closed sector (vertex at the origin) contained in $|\arg z| < \delta\pi/2$. The constants in (2.23) are given by (2.19) and

$$\begin{aligned} \gamma &= \sum_1^Q b_j - \sum_1^P a_j + \frac{(P-Q+1)}{2}, \\ E &= \prod_1^P \alpha_j^{\alpha_j}\prod_1^Q \beta_j^{-\beta_j}, \\ F &= (2\pi)^{(Q-P-1)/2}E^{\gamma/\delta}\delta^{-1/2}\prod_1^P \alpha_j^{1/2-a_j}\prod_1^Q \beta_j^{b_j-1/2}. \end{aligned} \tag{2.24}$$

Applying (2.23) to (2.14) yields

$$G_k^\alpha(r,t) \sim \Gamma_k^\alpha r^{-[(1-\alpha)n+2k]/(2-\alpha)}t^{-[n\alpha/2-2k]/(2-\alpha)} \cdot \exp\{-(2-\alpha)\alpha^{\alpha/(2-\alpha)}2^{-2/(2-\alpha)} \times r^{2/(2-\alpha)}t^{-\alpha/(2-\alpha)}\} \tag{2.25}$$

with

$$\Gamma_k^\alpha = \pi^{-n/2}2^{(2k-n)/(2-\alpha)}(2-\alpha)^{-1/2} \times \alpha^{[\alpha(n+1)/2-2k-1]/(2-\alpha)}. \tag{2.26}$$

Specializing to $\alpha = 1$ and $k = 0$ leads to

$$G_0^1(r,t) = (4\pi t)^{-n/2}\exp(-r^2/4t), \tag{2.27}$$

which is not only asymptotic but exact as may be seen from (2.13).

Further properties of the Green's functions are obtained by investigating their spatial Fourier transforms

$$g_k^\alpha(|\mathbf{q}|,t) = \int d^n x\, G_k^\alpha(|\mathbf{x}|,t)e^{i\mathbf{q}\mathbf{x}}. \tag{2.28}$$

Evaluating (2.28) in spherical coordinates yields with $q = |\mathbf{q}|$

$$g_k^\alpha(q,t) = (2\pi)^{n/2}q^{1-n/2}\int_0^\infty dr\, r^{n/2}J_{n/2-1}(qr)G_k^\alpha(r,t), \tag{2.29}$$

where J_σ denotes the Bessel function of the first kind. Performing a Mellin transformation (1.16) with respect to q yields

$$\hat{g}_k^\alpha(s,t) = \frac{1}{2}t^{k-\alpha s/2}\frac{\Gamma(s/2)\Gamma(1-s/2)}{\Gamma(1+k-\alpha s/2)}. \tag{2.30}$$

Hence, we used[5]

$$\hat{J}_\sigma(s) = 2^{s-1}\Gamma((\sigma+s)/2)/\Gamma(1+(\sigma-s)/2)). \tag{2.31}$$

Hence in view of (2.16) and (2.17) we obtain

$$g_k^\alpha(q,t) = \tfrac{1}{2}t^k H_{12}^{11}\left(qt^{\alpha/2}\Big|{}^{(0,1/2)}_{(0,1/2),(-k,\alpha/2)}\right). \tag{2.32}$$

Note that here (2.19) does not yield a restriction on α, in contrast to (2.20)

The following series representation of (2.32) is deduced from (2.21):

$$g_k^\alpha(q,t) = \sum_{j=0}^{\infty}\frac{(-1)^j}{\Gamma(1+k+\alpha j)}q^{2j}t^{k+\alpha j}. \tag{2.33}$$

Similarly, according to the remark after (2.21), we obtain the asymptotic behavior

$$g_k^\alpha(q,t) \sim \sum_{j=1}^{\infty} \frac{(-1)^{j-1}}{\Gamma(1+k-\alpha j)} q^{-2j} t^{k-\alpha j}, \tag{2.34}$$

for large q or t and $\alpha \neq 1,2$. Recall that $k = 1$ occurs only in connection with $\alpha > 1$, i.e., in all cases of interest only negative powers of t occur.

The relation

$$g_0^\alpha(q,t) = \frac{\partial}{\partial t} g_1^\alpha(q,t) \tag{2.35}$$

between g_0^α and g_1^α and the limits

$$g_0^\alpha(q,0) = 1, \quad g_1^\alpha(q,0) = 0, \tag{2.36}$$

as $t \downarrow 0$, may be drived either from (2.32) or (2.33).

For $\alpha = 1$ and $\alpha = 2$ the series (2.33) becomes elementary yielding

$$\begin{aligned} g_0^1(q,t) &= \exp(-q^2 t), \\ g_1^1(q,t) &= (1/q^2)[1 - \exp(-q^2 t)] \end{aligned} \tag{2.37}$$

(note that g_1^1 is not "needed"), and

$$\begin{aligned} g_0^2(q,t) &= \cos qt, \\ g_1^2(q,t) &= (1/q)\sin qt. \end{aligned} \tag{2.38}$$

Obviously, g_0^1, g_0^2, g_1^2 are the spatial Fourier transforms of the Green's function (2.27) for diffusion and of the Green's functions

$$\begin{aligned} G_0^2(r,t) &= \frac{\partial}{\partial t} G_1^2(r,t), \\ G_1^2(r,t) &= \frac{1}{2\pi^m}\left(\frac{\partial}{\partial t^2}\right)^{m-1} \delta(t^2 - r^2), \quad n = 2m+1, \\ G_1^2(r,t) &= \begin{cases} 0, & r > t, \\ (1/2\pi^m)(\partial/\partial t^2)^{m-1}(t^2 - r^2)^{-1/2}, & r < t, \end{cases} \\ & \quad n = 2m, \end{aligned} \tag{2.39}$$

for wave propagation, respectively.

The results (2.32)–(2.38) on the Fourier transforms of the Green's functions G_k^α imply that the solution

$$u(\mathbf{x},t) = \sum_{k=0}^{m-1} \int d^n y \, G_k^\alpha(|\mathbf{x}-\mathbf{y}|,t) f_k(\mathbf{y}) \tag{2.40}$$

of the fractional propagation equation (2.1) belongs to the Schwartz space $\mathscr{S}(R^n)$ for all $t > 0$ if the initial data f_k belong to it.

A particularly simple expression in connection with the Green's functions is obtained by applying the Laplace transform (1.15) with respect to the time t to the spatial Fourier transforms (2.32):

$$\tilde{g}_k^\alpha(q,p) = p^{\alpha-k-1}/(p^\alpha + q^2). \tag{2.41}$$

To obtain this result one may use (2.33) leading to a geometric series with sum (2.41) but converging only for $|q^2 p^{-\alpha}| < 1$, whereas (2.41) holds without this restriction. Alternatively, one may apply the following general result: The Laplace transform $\tilde{F}$ of

$$F(x) = x^\rho H_{PQ}^{MN}\left(a x^\sigma \middle| {}^{(a_j,\alpha_j)_{1,\dots,P}}_{(b_j,\beta_j)_{1,\dots,Q}}\right), \quad \sigma > 0, \tag{2.42}$$

is given by

$$\tilde{F}(p) = p^{-\rho-1} H_{P+1,Q}^{M,N+1}\left(a p^{-\sigma} \middle| {}^{(-\rho,\sigma),(a_j,\alpha_j)_{1,\dots,P}}_{(b_j,\beta_j)_{1,\dots,Q}}\right) \tag{2.43}$$

provided that

$$\mathrm{Re}(\rho + 1) > -\sigma \min_{1 \leqslant j \leqslant M} \mathrm{Re}(b_j/\beta_j) \tag{2.44}$$

in addition to (2.19), (2.22) holds. This result [Eq. (2.4.2) in Ref. 3 ($Q + 1$ should read Q as is clear from the parameter list following the argument)] follows from the path integral representation[2] of the Fox function in (2.42); (2.44) reflects the condition[2] on the path in the integral representation of the Fox function in (2.43).

Application to (2.32) yields

$$\tilde{g}_k^\alpha(q,p) = \tfrac{1}{2} p^{-k-1} H_{22}^{12}\left(q p^{-\alpha/2} \middle| {}^{(-k,\alpha/2),(0,1/2)}_{(0,1/2),(-k,\alpha/2)}\right) \tag{2.45}$$

or

$$\tilde{g}_k^\alpha(q,p) = \tfrac{1}{2} p^{-k-1} H_{11}^{11}\left(q p^{-\alpha/2} \middle| {}^{(0,1/2)}_{(0,1/2)}\right) \tag{2.46}$$

by taking the reduction formula

$$H_{P+1,Q+1}^{M,N+1}\left(z \middle| {}^{(a,\alpha),(a_j,\alpha_j)_{1,\dots,P}}_{(b_j,\beta_j)_{1,\dots,Q},(a,\alpha)}\right) = H_{PQ}^{MN}\left(z \middle| {}^{(a_j,\alpha_j)_{1,\dots,P}}_{(b_j,\beta_j)_{1,\dots,Q}}\right), \tag{2.47}$$

into account.

By looking at its Mellin transform we find for the Fox function in (2.46)

$$H_{11}^{11}\left(z \middle| {}^{(0,1/2)}_{(0,1/2)}\right) = 2/(1+z^2). \tag{2.48}$$

Combining (2.46) and (2.48) yields (2.41).

The function $\tilde{g}_k^\alpha(q,p)$ in (2.41) has an analytic continuation into the half-plane $\mathrm{Re}\, p > 0$. This follows from its definition as Laplace transform of the spatial Fourier transform (2.32) of the Green's function G_k^α. Elementary estimates on the behavior of (2.41) as $\mathrm{Re}\, p$ tends to zero yield the following result: The spatiotemporal Fourier transform

$$FG_k^\alpha(|\mathbf{q}|,q_0) = \lim_{\epsilon \downarrow 0} \tilde{g}_k^\alpha(|\mathbf{q}|,\epsilon - iq_0) \tag{2.49}$$

is a tempered distribution on $R^n \times R$. The inverse Fourier transform of (2.49), i.e., $G_k^\alpha(|\mathbf{x}|,t)$, is a tempered distribution on $R^n \times R_+$.

III. SPECIAL PROPERTIES OF FRACTIONAL DIFFUSION

The solution of the fractional diffusion equation

$$u(\mathbf{x},t) = f_0(\mathbf{x}) + \frac{1}{\Gamma(\alpha)} \int_0^t d\tau (t-\tau)^{\alpha-1} \Delta u(\mathbf{x},\tau), \tag{3.1}$$

$0 < \alpha \leqslant 1$, with incorporated initial condition

$$u(\mathbf{x},0) = f_0(\mathbf{x}), \tag{3.2}$$

is given by

$$u(\mathbf{x},t) = \int d^n y \, G_0^\alpha(|\mathbf{x}-\mathbf{y}|,t) f_0(\mathbf{y}). \tag{3.3}$$

The Green's function G_0^α, given in (2.14), may be rewritten as

$$G_0^\alpha(r,t) = \alpha^{-1}\pi^{-n/2}r^{-n}H_{12}^{20}\left(2^{-2/\alpha}r^{2/\alpha}t^{-1}\Big|_{(n/2,1/\alpha),(1,1/\alpha)}^{(1,1)}\right) \tag{3.4}$$

using the relation

$$H_{PQ}^{MN}\left(z\Big|_{(b_j,\beta_j)_{1,\dots,Q}}^{(a_j,\alpha_j)_{1,\dots,P}}\right) = \gamma H_{PQ}^{MN}\left(z^\gamma\Big|_{(b_j,\gamma\beta_j)_{1,\dots,Q}}^{(a_j,\gamma\alpha_j)_{1,\dots,P}}\right) \tag{3.5}$$

holding for $\gamma > 0$, which is easily deduced from (2.15)–(2.17). For $\alpha = 1$ (3.4) reduces to

$$G_0^1(r,t) = (4\pi t)^{-n/2}\exp(-r^2/(4t)). \tag{3.6}$$

This is the Green's function of ordinary diffusion with the well-known property of being a probability density on R^n. It is remarkable that this property carries over to the Green's function G_0^α, $0<\alpha<1$, of fractional diffusion. It is a consequence of the following theorem.

Theorem 3.1: The functions

$$\psi_{\mu,\beta}(x) = [1/\Gamma(\mu)]x^{-1}H_{12}^{20}\left(2^{-1/\beta}x^{-1}\Big|_{(\mu,1/2\beta),(1,1/2\beta)}^{(1,1)}\right), \tag{3.7}$$

$x>0$, are probability densities on R_+ for

$$\mu > 0, \quad 0 < \beta \leqslant \beta_c(\mu). \tag{3.8}$$

Above the critical value $\beta_c(\mu)$ with

$$\beta_c(\mu) = \tfrac{1}{2}, \quad \mu \geqslant 1, \tag{3.9a}$$

$$\beta_c(\mu) = 1/(2\mu), \quad \tfrac{1}{2} \leqslant \mu \leqslant 1, \tag{3.9b}$$

$$\beta_c(\mu) = 1, \quad 0 < \mu \leqslant \tfrac{1}{2}, \tag{3.9c}$$

the functions $\psi_{\mu,\beta}$ do not have a definite sign but remain normalized.

The proof of the theorem will be presented in Appendix B.

Corollary 3.2: The Green's function G_0^α, $0<\alpha\leqslant 1$, is a probability density on R^n. Equivalently, due to rotational invariance,

$$\rho_\alpha(r,t) = [2\pi^{n/2}/\Gamma(n/2)]G_0^\alpha(r,t)r^{n-1} \tag{3.10}$$

is a probability density on R_+.

Proof: Setting $r = x^{-\alpha/2}$ we have to show that

$$f_\alpha(x,t) = (\alpha/2)x^{-\alpha/2-1}\rho_\alpha(x^{-\alpha/2},t), \quad x>0, \tag{3.11}$$

is a probability density on R_+. From (3.4), (3.10), and (3.11) we obtain

$$f_\alpha(x,t) = t\psi_{n/2,\alpha/2}(tx), \quad t>0. \tag{3.12}$$

Hence by Theorem 3.1 f_α is a probability density for all dimensions $n\in N$ and $0<\alpha\leqslant 1$. Actually, for $n=1$ we obtain the stronger result that it remains a probability density in the extended range $0<\alpha\leqslant 2$.

To the best of the authors' knowledge the probability densities (3.7) have not been introduced before apart from the special case $\mu = \frac{1}{2}$ for which the following result holds:

$$\psi_{1/2,\beta}(x) = w_\beta(x). \tag{3.13}$$

Here, w_β with $0<\beta<1$ denotes the one-sided stable (or Lévy) probability density characterized by its Laplace transform

$$\tilde{w}_\beta(p) = \exp(-p^\beta). \tag{3.14}$$

The equality (3.13) follows from the equality for the Mellin transforms

$$\hat{\psi}_{1/2,\beta}(s) = \beta^{-1}\frac{\Gamma(\beta^{-1}(1-s))}{\Gamma(1-s)} = \hat{w}_\beta(s) \tag{3.15}$$

from which also

$$w_\beta(x) = (1/\beta)x^{-2}H_{11}^{10}\left(x^{-1}\Big|_{(-1/\beta,1/\beta)}^{(-1,1)}\right) \tag{3.16}$$

may be inferred.[4]

As a consequence we obtain for $n=1$ the following representation of G_0^α, $0<\alpha<2$:

$$G_0^\alpha(r,t) = \alpha^{-1}\,tr^{-1-2/\alpha}w_{\alpha/2}(tr^{-2/\alpha}). \tag{3.17}$$

The limiting case $\alpha = 2$ is obtained by setting $\beta = 1$ in (3.14) or (3.15) leading to

$$w_1(x) = \delta(x-1). \tag{3.18}$$

Correspondingly, we obtain

$$G_0^2(r,t) = \tfrac{1}{2}\,tr^{-2}\,\delta(tr^{-1}-1), \tag{3.19}$$

or equivalently

$$G_0^2(r,t) = \tfrac{1}{2}\delta(t-r) = t\delta(t^2-r^2). \tag{3.20}$$

The latter corresponds to the special case $m=0$, $n=1$, in (2.39). We conclude this section by determining the moments of the probability distribution G_0^α on R^n. The moments are defined by

$$M(k_1,k_2,\dots,k_n) = \int d^nx\, G_0^\alpha(|\mathbf{x}|,t)x_1^{k_1}x_2^{k_2}\cdots x_n^{k_n}, \tag{3.21}$$

with non-negative integers $k_1,k_2,\dots,k_n$. Obviously

$$M(k_1,k_2,\dots,k_n) = 0, \quad \text{if } k_j = \text{odd}, \tag{3.22}$$

for some j, $1<j\leqslant n$. For the even moments we obtain in spherical coordinates

$$M(2m_1,2m_2,\dots,2m_n) = M_0(2m_1,2m_2,\dots,2m_n)\overline{G}_0^\alpha(n+2m,t), \tag{3.23}$$

where

$$M_0(k_1,\dots,k_n) = \int_{S^{n-1}} d\Omega(\mathbf{e})e_1^{k_1}\cdots e_n^{k_n} \tag{3.24}$$

and

$$m = \sum_{j=1}^n m_j. \tag{3.25}$$

The term $\overline{G}_0^\alpha$ is the Mellin transform of G_0^α with respect to r,

$$\overline{G}_0^\alpha(s,t) = \int_0^\infty dr\, r^{s-1}G_0^\alpha(r,t), \tag{3.26}$$

which may be obtained from (3.4), (2.15)–(2.17),

$$\overline{G}_0^\alpha(s,t) = \alpha^{-1}\pi^{-n/2}2^{s-n-1}t^{\alpha(s-n)/2} \times\frac{\Gamma(s/2)\Gamma((s-n)/2)}{\Gamma(\alpha(s-n)/2)}. \tag{3.27}$$

For the spherical moments M_0 in (3.24) one may deduce

$$M_0(2m_1,\dots,2m_n) = 2\,\frac{\Pi_{j=1}^n\,\Gamma(m_j+\frac{1}{2})}{\Gamma(m+n/2)} \tag{3.28}$$

by induction. Inserting (3.27) and (3.28) into (3.23) yields

$$M(2m_1,\dots,2m_n) = \pi^{-n/2}2^{2m}\prod_{j=1}^n\Gamma\left(m_j+\frac{1}{2}\right) \times\frac{\Gamma(1+m)}{\Gamma(1+\alpha m)}t^{\alpha m}. \tag{3.29}$$

In particular, we find for the mean square displacement

$$\int d^n x\, G_0^\alpha(|\mathbf{x}|,t)|\mathbf{x}|^2 = \frac{2n}{\Gamma(1+\alpha)}\, t^\alpha . \tag{3.30}$$

IV. FRACTIONAL DIFFUSION IN A HALF-SPACE

In the previous sections we considered fractional propagation in the spatial region D with $D = R^n$, i.e., the full space. Here, we shall treat the problem of fractional diffusion in the half-space $D = R^{n-1}\times R_+$. As D has a boundary ∂D, given by $\partial D = R^{n-1}\times\{0\}$, the fractional diffusion equation

$$u(\mathbf{x},t) = f(\mathbf{x}) + \frac{1}{\Gamma(\alpha)}\int_0^t d\tau (t-\tau)^{\alpha-1}\Delta u(\mathbf{x},\tau) \tag{4.1}$$

has to be supplemented by a boundary condition

$$\lambda u(\mathbf{x}^0,t) - \mu \frac{\partial u}{\partial x_n}(\mathbf{x}^0,t) = v(\mathbf{x}^T,t), \quad t>0, \tag{4.2}$$

with given v. The notation is as follows:

$$\mathbf{x} = (x_1,x_2,\ldots,x_n)\in D, \quad \mathbf{x}^T = (x_1,x_2,\ldots,x_{n-1})\in R^{n-1},$$
$$\mathbf{x}^0 = (x_1,x_2,\ldots,x_{n-1},0)\in\partial D, \quad S\mathbf{x} = (x_1,x_2,\ldots,-x_n)\in\bar{D}, \tag{4.3}$$

with $\bar{D} = R^{n-1}\times R_-$.

As the propagation equation (4.1) is linear, it is sufficient to consider separately the problems

$$\begin{aligned}&(1)\ f\neq 0, \quad v=0,\\ &(2)\ f=0, \quad v\neq 0,\end{aligned} \tag{4.4}$$

which will be called the first and second type, respectively. These problems are further subdivided by the choice of the parameters λ,u in (4.2):

$$\begin{aligned}&(1)\ \lambda=1, \quad \mu=0;\\ &(2)\ \lambda=0, \quad \mu=1;\\ &(3)\ \lambda\neq 0, \quad \mu=1.\end{aligned} \tag{4.5}$$

A pair of indices i,j with $i = 1,2$ and $j = 1,2,3$ will refer to (i) of (4.4) and (j) of (4.5). The Green's function G_0^α for fractional diffusion in R^n will be denoted by G^α, given, e.g., by (3.4).

The problems of first type are solved by

$$u(\mathbf{x},t) = \int_D d^n y\, G_{1j}^\alpha(\mathbf{x},\mathbf{y},t) f(\mathbf{y}), \tag{4.6}$$

where

$$G_{11}^\alpha(\mathbf{x},\mathbf{y},t) = G^\alpha(|\mathbf{x}-\mathbf{y}|,t) - G^\alpha(|\mathbf{x}-S\mathbf{y}|,t), \tag{4.7a}$$

$$G_{12}^\alpha(\mathbf{x},\mathbf{y},t) = G^\alpha(|\mathbf{x}-\mathbf{y}|,t) + G^\alpha(|\mathbf{x}-S\mathbf{y}|,t), \tag{4.7b}$$

$$G_{13}^\alpha(\mathbf{x},\mathbf{y},t) = G_{12}^\alpha(\mathbf{x},\mathbf{y},t) + G_3^\alpha(\mathbf{x},\mathbf{y},t), \tag{4.7c}$$

with

$$G_3^\alpha(\mathbf{x},\mathbf{y},t) = -2\lambda e^{\lambda y_n}\int_{-\infty}^{-y_n} G^\alpha(|\mathbf{x}-(\mathbf{y}^T,z)|,t)e^{\lambda z}\, dz . \tag{4.8}$$

The verification of these results is rather elementary and therefore omitted.

As application we treat the problem $i=1$, $j=1$ for $n=1$, i.e., $D = R_+$, with the particular initial condition

$$f(x) = 1, \quad x\in R_+ . \tag{4.9}$$

From (4.6), (4.7a), and (4.9) we obtain

$$u(x,t) = \int_0^\infty dy\{G^\alpha(|x-y|,t) - G^\alpha(x+y,t)\} \tag{4.10}$$

or

$$u(x,t) = \int_0^x dy\, G^\alpha(x-y,t) + \int_x^\infty dy\, G^\alpha(y-x,t) - \int_0^\infty dy\, G^\alpha(x+y,t) . \tag{4.11}$$

Substituting z for $x-y, y-x, x+y$, respectively, we obtain

$$u(x,t) = 2\int_0^x dz\, G^\alpha(z,t) . \tag{4.12}$$

Recalling the representation (3.17) of G^α in one dimension we obtain, substituting $y = tz^{-2/\alpha}$,

$$u(x,t) = \int_{tx^{-2/\alpha}}^\infty dy\, w_{\alpha/2}(y) = W_{\alpha/2}^c(tx^{-2/\alpha}) \tag{4.13}$$

with $W^c(\xi)$ the complementary probability distribution

$$W^c(\xi) = \int_\xi^\infty dy\, w(y) \tag{4.14}$$

associated with a probability density w. In particular, we have for $\alpha = 1$,

$$u(x,t) = W_{1/2}^c(tx^{-2}) = \operatorname{erf}(x/2\sqrt{t}) \tag{4.15}$$

with

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z dt\, e^{-t^2} \tag{4.16}$$

denoting the error function. The solution (4.13) may also be represented by a Fox function

$$u(x,t) = H_{22}^{11}\left(t^{-\alpha/2}x\,\middle|\,{}^{(1,1),(1,\alpha/2)}_{(1,1),(0,1)}\right) . \tag{4.17}$$

This may be verified by determining the Mellin transforms of (4.13) and (4.17) with respect to the variable $z = t^{-\alpha/2}x$, using (3.15) and $\widehat{W}^c(s) = \widehat{w}(s+1)/s$. Taking the relation

$$\Gamma(1+s) = \pi^{-1/2}2^s\Gamma((1+s)/2)\Gamma((2+s)/2) \tag{4.18}$$

into account, (4.17) may be rewritten as

$$u(x,t) = \pi^{-1/2}H_{23}^{21}\left(\tfrac{1}{2}t^{-\alpha/2}x\,\middle|\,{}^{(1,1),(1,\alpha/2)}_{(1/2,1/2),(1,1/2),(0,1)}\right), \tag{4.19}$$

which is the form of the solution presented in Ref. 1.

The second type of problem is solved by a Fourier transform with respect to $\mathbf{x}^T$ and a Laplace transform with respect to t:

$$U(\mathbf{k},x_n,p) = \int_{R^{n-1}} d^{n-1}x^T e^{i\mathbf{k}\mathbf{x}^T}\int_0^\infty dt\, e^{-pt}u(\mathbf{x},t), \tag{4.20}$$

with $\mathbf{k}\in R^{n-1}$. Applying (4.20) to (4.1) with $f=0$ yields, after rearrangement and with $k^2 = \mathbf{k}^2$,

$$\frac{\partial^2}{\partial x_n^2}U(\mathbf{k},x_n,p) = (p^\alpha + k^2)U(\mathbf{k},x_n,p) . \tag{4.21}$$

The general solution of (4.21) is

$$U(\mathbf{k},x_n,p) = A(\mathbf{k},p)e^{-\sqrt{p^\alpha+k^2}x_n} + B(\mathbf{k},p)e^{\sqrt{p^\alpha+k^2}x_n} . \tag{4.22}$$

Here U is a Laplace transform for arbitrary $x_n > 0$ only if B is zero. The coefficient A is determined by the boundary condition (4.2)

$$(\lambda + \mu\sqrt{p^\alpha + k^2})A(\mathbf{k},p) = V(\mathbf{k},p) , \tag{4.23}$$

where V is obtained from v by applying the Fourier–Laplace transform (4.20). Going back to the space-time domain we obtain the solution

$$u(\mathbf{x},t) = \int_{R^{n-1}} d^{n-1}y \int_0^t d\tau\, G_2^\alpha((\mathbf{x}^T - \mathbf{y},x_n),t-\tau)v(\mathbf{v},\tau) , \tag{4.24}$$

with the Green's function G characterized by its Fourier–Laplace transform (4.20)

$$\overline{G}_x^\alpha(\mathbf{k},x_n,p) = (\lambda + u\sqrt{p^\alpha + k^2})^{-1} e^{-\sqrt{p^\alpha + k^2}x_n}. \tag{4.25}$$

Inserting the three cases of (4.5) yields $\overline{G}_{2j}$, $j = 1,2,3$.

We did not succeed to find explicit representations for G_{2j} except in special cases that will be treated below. For $\alpha = 1$ and arbitrary n we have from (4.25)

$$\overline{G}_2^1(\mathbf{k},x_n,p) = \overline{G}_2^1(0,x_n,p + \mathbf{k}^2) . \tag{4.26}$$

Hence

$$G_2^1(\mathbf{x},t) = (4\pi t)^{-(n-1)/2} \exp[-(\mathbf{x}^T)^2/(4t)]g_2(x_n,t) , \tag{4.27}$$

with g_2 characterized by its Laplace transform

$$\tilde{g}_2(x_n,p) = (\lambda + \mu\sqrt{p})^{-1} e^{-\sqrt{p}x_n} . \tag{4.28}$$

From Ref. 5 we obtain

$$g_{21}(x_n,t) = 2^{-1}\pi^{-1/2}x_n t^{-3/2} \exp[-x_n^2/(4t)] , \tag{4.29a}$$

$$g_{22}(x_n,t) = \pi^{-1/2}t^{-1/2} \exp[-x_n^2/(4t)] , \tag{4.29b}$$

$$g_{23}(x_n,t) = g_{22}(x_n,t) - \lambda e^{\lambda x_n + \lambda^2 t} \operatorname{erfc}(\tfrac{1}{2}x_n t^{-1/2} + \lambda t^{1/2}) , \tag{4.29c}$$

with $\operatorname{erfc}(z) = 1 - \operatorname{erf}(z)$ the complementary error function.

We note that G_{2j}^1, $j = 1,2$, are non-negative. The question arises whether this remains true for $0 < \alpha < 1$. For $n > 1$ this question remains open.

For $n = 1$ and arbitrary α (4.25) reduces to

$$\tilde{G}_2^\alpha(x,p) = (\lambda + \mu p^{\alpha/2})^{-1} e^{-p^{\alpha/2}x}, \quad x = x_1 , \tag{4.30}$$

leading to

$$G_{21}^\alpha(x,t) = x^{-2/\alpha} w_{\alpha/2}(x^{-2/\alpha}t) , \tag{4.31}$$

$$G_{22}^\alpha(x,t) = \int_x^\infty dy\, G_{21}^\alpha(y,t) . \tag{4.32}$$

These explicit representations answer the above question affirmatively for $n = 1$. For $\alpha = 1$, (4.31) and (4.32) reduce to (4.29a) and (4.29b).

As application we treat the problem with $i = 2$, $j = 1$, and the particular choice

$$v(t) = 1, \quad t > 0 , \tag{4.33}$$

in (4.2). From (4.24) and (4.31) we obtain

$$u(x,t) = W_{\alpha/2}(tx^{-2/\alpha}) , \tag{4.34}$$

with

$$W(\xi) = \int_0^\xi d\eta\, w(\eta) , \tag{4.35}$$

the probability distribution associated with a probability density w on R_+. Using $\widetilde{W}(p) = \widetilde{w}(p)/p$, (1.17) and (3.15) yield

$$u(x,t) = H_{11}^{10}\left(t^{-\alpha/2}x\,\middle|\,{}^{(1,\alpha/2)}_{(0,1)}\right) , \tag{4.36}$$

or with (4.18)

$$u(x,t) = \pi^{-1/2} H_{23}^{30}\left(\tfrac{1}{2}t^{-\alpha/2}x\,\middle|\,{}^{(1,1),(1,\alpha/2)}_{(0,1),(1/2,1/2),(1,1/2)}\right) , \tag{4.37}$$

which is the form of the solution presented in Ref. 1 [except for a change of signs in (4.33) and (4.37)].

V. SUMMARY

Formally, fractional diffusion and wave equations are obtained from their ordinary counterparts by replacing the first- and second-order time derivatives, respectively, by derivatives of fractional order α with $0 < \alpha < 1$ for fractional diffusion and $1 < \alpha < 2$ for fractional wave propagation. To take appropriate initial conditions into account these equations are reformulated as integrodifferential equations. Exact solutions for the Green's functions of the latter have been found and discussed in detail for arbitrary space dimension. As a by-product a new class of probability densities has been disclosed.

A possible physical application of fractional diffusion is the description of diffusion in special types of porous media as pointed out by Nigmatullin[6]. As is well-known,[7] there is an intimate connection between white noise and ordinary diffusion. An analogous connection between "grey" noise, characterized by the "greyness" parameter α, $0 < \alpha < 1$, and fractional diffusion has been established.[8] For $\alpha = 1$ white noise and ordinary diffusion are recovered.

APPENDIX A: THE KERNEL OF $(\lambda^2 - \Delta)^{-1}$

The kernel

$$k(|\mathbf{x}|,\lambda) = (2\pi)^{-n/2}(r/\lambda)^{1-n/2}K_{n/2-1}(\lambda r), \quad r = |\mathbf{x}| , \tag{A1}$$

introduced in (2.8) satisfies

$$(\Delta - \lambda^2)k(|\mathbf{x}|,\lambda) = -\delta(\mathbf{x}) , \tag{A2}$$

which is verified by calculating its Fourier transform

$$K(|\mathbf{q}|,\lambda) = \int d^n x\, e^{i\mathbf{q}\mathbf{x}} k(|\mathbf{x}|,\lambda) . \tag{A3}$$

Inserting (A1) and introducing spherical coordinates leads to ($q = |\mathbf{q}|$)

$$K(|\mathbf{q}|,\lambda) = (q/\lambda)^{1-n/2} \int_0^\infty dr\, r J_{n/2-1}(qr) K_{n/2-1}(\lambda r) . \tag{A4}$$

Using[5]

$$\int_0^\infty dr\, r J_\sigma(qr) K_\sigma(\lambda r) = \frac{(q/\lambda)^\sigma}{q^2 + \lambda^2}, \quad \operatorname{Re}\sigma > -1 , \tag{A5}$$

yields

$$K(|\mathbf{q}|,\lambda) = (q^2 + \lambda^2)^{-1} . \tag{A6}$$

Hence

$$(-q^2 - \lambda^2)K(|\mathbf{q}|,\lambda) = -1, \tag{A7}$$

which is the Fourier transform of (A2).

Consider now the homogeneous equation

$$(\Delta - \lambda^2)w(\mathbf{x},\lambda) = 0\,. \tag{A8}$$

In one dimension ($n=1$) this is an ordinary differential equation with the general solution

$$w(x,\lambda) = A(\lambda)e^{\lambda x} + B(\lambda)e^{-\lambda x}\,. \tag{A9}$$

Obviously, $w(\cdot,p^{\alpha/2})$ is the Laplace transform of some function only in the trivial case $A=B=0$.

For higher dimensions ($n\geqslant 2$) (A8) is separable in spherical coordinates.[9] We set accordingly

$$w(\mathbf{x},\lambda) = f(r,\lambda)\,Y_l(\mathbf{e})\,, \tag{A10}$$

with

$$r = |\mathbf{x}| > 0, \quad \mathbf{e} = r^{-1}\mathbf{x}\,, \tag{A11}$$

and Y_l a spherical harmonics of degree l, i.e., the restriction

$$Y_l(\mathbf{e}) = P_l(\mathbf{e}), \quad \mathbf{e}\in S^{n-1}\,, \tag{A12}$$

of a homogeneous polynomial $P_l(\mathbf{x})$, $\mathbf{x}\in R^n$, of degree l satisfying

$$\Delta P_l(\mathbf{x}) = 0\,. \tag{A13}$$

This leads to the radial equation

$$f'' + \frac{n-1}{r}f' - \frac{l(l+n-2)}{r^2}f - \lambda^2 f = 0, \quad r>0\,, \tag{A14}$$

primes denoting derivatives with respect to r. Two independent solutions of (A14) are

$$f_l^1(r,\lambda) = (\lambda r)^{1-n/2} I_{l+n/2-1}(\lambda r) \tag{A15}$$

and

$$f_l^2(r,\lambda) = (\lambda r)^{1-n/2} K_{l+n/2-1}(\lambda r)\,, \tag{A16}$$

with I_σ and K_σ modified Bessel functions of the first and second kind, respectively.

In accordance with (A10) we set

$$w_l^i(\mathbf{x},\lambda) = f_l^i(r,\lambda)\,Y_l(\mathbf{e}), \quad i=1,2\,. \tag{A17}$$

By construction they satisfy (A8) in the restricted domain $R^n\setminus\{0\}$. For w_l^1 we obtain from (A15) and the series representation[9] of I_σ

$$w_l^1(\mathbf{x},\lambda) = 2^{1-l-n/2}\lambda^{\,l} P_l(\mathbf{x}) \sum_{j=0}^{\infty} \frac{(\lambda^2 r^2/4)^j}{j!\,\Gamma(l+j+n/2)}\,, \tag{A18}$$

where we used

$$r^l Y_l(\mathbf{e}) = P_l(\mathbf{x})\,. \tag{A19}$$

As (A18) is converging for all $\mathbf{x}\in R^n$, it represents an entire function of $\mathbf{x} = (x_1,x_2,\ldots,x_n)\in R^n$; "complexification" is trivially achieved by letting x vary in C^n. Obviously, $(\Delta-\lambda^2)w_l^1$ is entire. As this expression vanishes in $R^n\setminus\{0\}$ the same is true in R^n and in C^n. A direct verification is also easily performed using (A18) and

$$\begin{aligned}\Delta(r^{2j}P_l(\mathbf{x})) &= \Delta(r^{2j+l}Y_l(\mathbf{e}))\\ &= 4j(j+l+n/2-1)r^{2j-2}P_l(\mathbf{x})\,. \end{aligned} \tag{A20}$$

For w_l^2, given by

$$w_l^2(\mathbf{x},\lambda) = (\lambda r)^{1-n/2} K_{l+n/2-1}(\lambda r)\,Y_l(\mathbf{e})\,, \tag{A21}$$

from (A16) and (A17), we shall show that

$$w_l^2(\mathbf{x},\lambda) = \lambda^{-l} P_l\Big(-\frac{\partial}{\partial\mathbf{x}}\Big) w_0^2(\mathbf{x},\lambda) \tag{A22}$$

holds. Comparing (A21) for $l=0$ and (A1) yields

$$w_0^2(\mathbf{x},\lambda) = (2\pi)^{n/2}\lambda^{2-n} k(|\mathbf{x}|,\lambda)\,. \tag{A23}$$

Hence combining (A2), (A22), and (A23) yields

$$(\Delta-\lambda^2)w_l^2(\mathbf{x},\lambda) = -(2\pi)^{n/2}\lambda^{2-n-l} P_l\Big(-\frac{\partial}{\partial\mathbf{x}}\Big)\delta(\mathbf{x})\,. \tag{A24}$$

This result implies that w_l^2 is not a solution of the homogeneous equation (A8) in R^n. The most general solution of this equation is therefore a (finite or infinite) linear combination of the solutions (A18) that are unbounded in λ for $\lambda\to\infty$. In particular, they do not represent Laplace transforms in the variable $\lambda = p^{\alpha/2}$. Hence their exclusion in (2.9) is justified.

It remains to show the validity of the relation (A22). We denote by $V_l(n)$ the vector space of homogeneous polynomials $P_l(x)$, $x\in R^n$, satisfying (A13). Its dimension is given by[9]

$$d_l(n) = \dim V_l(n) = \frac{2l+n-2}{l+n-2}\binom{l+n-2}{l}\,. \tag{A25}$$

Obviously, it is sufficient to prove (A22) for a basis of $V_l(n)$. We shall do this separately for $n=2$ and $n\geqslant 3$.

A basis of the two-dimensional vector space $V_l(2)$, $l\geqslant 1$, is

$$P_{l,\sigma}(\mathbf{x}) = r^l e^{i\sigma l\phi}, \quad \sigma = \pm\,, \tag{A26}$$

with $\mathbf{x} = (x_1,x_2)$, $x_1 + ix_2 = re^{i\phi}$. Accordingly, we have from (A21)

$$w_{l,\sigma}^2(\mathbf{x},\lambda) = K_l(\lambda r)e^{i\sigma l\phi}\,. \tag{A27}$$

Evaluation of the Fourier transform

$$W_{l,\sigma}^2(\mathbf{q},\lambda) = \int d^2x\, e^{i\mathbf{q}\mathbf{x}} w_l^2(\mathbf{x},\lambda) \tag{A28}$$

in polar coordinates yields

$$W_{l,\sigma}^2(\mathbf{q},\lambda) = 2\pi\lambda^{-l} P_{l,\sigma}(i\mathbf{q})(q^2+\lambda^2)^{-1}\,, \tag{A29}$$

where (A5) has been used. Hence

$$W_{l,\sigma}^2(\mathbf{q},\lambda) = \lambda^{-l} P_{l,\sigma}(i\mathbf{q})\, W_0^2(\mathbf{q},\lambda)\,, \tag{A30}$$

which is the Fourier transform of (A22).

The proceeding in higher dimensions is similar. Let $P_{l,\sigma}(\mathbf{x})$, $\sigma = 1,2,\ldots,d_l(n)$ be a real basis of $V_l(n)$ such that the associated spherical harmonics $Y_{l,\sigma}(\mathbf{e}) = P_{l,\sigma}(\mathbf{e})$, $\mathbf{e}\in S^{n-1}$, are orthonormalized as follows[9]:

$$\int_{S^{n-1}} d\Omega(\mathbf{e})\,Y_{l,\sigma}(\mathbf{e})\,Y_{l',\sigma'}(\mathbf{e}) = \delta_{ll'}\delta_{\sigma\sigma'}\,. \tag{A31}$$

The following addition theorem[9] holds: Define

$$Z_l(\mathbf{e},\mathbf{e}') = C_l^{n/2-1}(\mathbf{e}\cdot\mathbf{e}')\,, \tag{A32}$$

with C_l^α denoting the Gegenbauer polynomials.[9] Then $Z_l(\mathbf{e},\cdot)$ and $Z_l(\cdot,\mathbf{e}')$ are in $V_l(n)$ and the relation

$$Z_l(\mathbf{e},\mathbf{e}') = C_l^{n/2-1}(1)\,\frac{\Omega(n)}{d_l(n)} \sum_{\sigma=1}^{d_l(n)} Y_{l,\sigma}(\mathbf{e})\,Y_{l,\sigma}(\mathbf{e}') \tag{A33}$$

holds. Here

$$\Omega(n) = \int_{S^{n-1}} d\Omega(\mathbf{e}) = \frac{2\pi^{n/2}}{\Gamma(n/2)} \tag{A34}$$

is the surface of S^{n-1} and[9]

$$C^{n/2-1}(1) = \frac{(l+n/2-2)!}{l!(n/2-2)!}. \tag{A35}$$

Calculating the Fourier transform

$$W^2_{l,\sigma}(\mathbf{q},\lambda) = \int d^n x\, e^{i\mathbf{qx}} w^2_{l,\sigma}(\mathbf{x},\lambda), \tag{A36}$$

where

$$w^2_{l,\sigma}(\mathbf{x},\lambda) = (\lambda r)^{1-n/2} K_{l+n/2-1}(\lambda r) Y_{l,\sigma}(\mathbf{e}) \tag{A37}$$

yields in spherical coordinates ($\mathbf{x} = r\mathbf{e}$, $\mathbf{q} = q\mathbf{e}'$)

$$W^2_{l,\sigma}(\mathbf{q},\lambda) = \int_0^\infty dr\, r^{n-1} (\lambda r)^{n/2-1} K_{l+n/2-1}(\lambda r) \cdot \int_{S^{n-1}} d\Omega(\mathbf{e}) e^{iqr\mathbf{e}\cdot\mathbf{e}'} Y_{l,\sigma}(\mathbf{e}). \tag{A38}$$

Inserting

$$e^{iqr\mathbf{e}\cdot\mathbf{e}'} = \sum_{l=0}^{\infty} c_l (qr)^{1-n/2} J_{l+n/2-1}(qr) C_l^{n/2-1}(\mathbf{e}\cdot\mathbf{e}') \tag{A39}$$

(see below for a derivation) with

$$c_l = 2^{n/2-1} i^l (l+n/2-1)\Gamma(n/2-1), \tag{A40}$$

using the addition theorem (A32) and (A33), the orthogonality relation (A31), and (A5) leads to

$$W^2_{l,\sigma}(\mathbf{q},\lambda) = (2\pi)^{n/2} \lambda^{2-l-n} P_{l,\sigma}(i\mathbf{q})(q^2+\lambda^2)^{-1} \tag{A41}$$

or

$$W^2_{l,\sigma}(\mathbf{q},\lambda) = \lambda^{-l} P_{l,\sigma}(i\mathbf{q}) W^2_0(\mathbf{q},\lambda), \tag{A42}$$

which is the Fourier transform of (A22).

To complete the argument we shall derive (A39). Starting from the trivial fact that

$$w(\mathbf{x}) = \exp(-\lambda x_1) \tag{A43}$$

is a solution of (A8), we may expand w as series in terms of the solutions (A17) with $i = 1$:

$$w(\mathbf{x}) = \sum_{l=0}^{\infty} d_l f_l^1(r,\lambda) Y_l(\mathbf{e}). \tag{A44}$$

As w depends only on x_1 the spherical harmonics $Y_l(\mathbf{e})$ depend only on e_1, i.e.,

$$Y_l(\mathbf{e}) = y_l(e_1). \tag{A45}$$

Inserting $P_l(\mathbf{x}) = r^l Y_l(\mathbf{e})$ into $\Delta P_l = 0$ expressed in spherical coordinates[9] yields

$$(1-z^2)\frac{d^2}{dz^2} y_l(z) - (n-1)\frac{d}{dz} y_l(z) + l(l+n-2) y_l(z) = 0, \tag{A46}$$

with $z = \cos\theta_1 = e_1$. The regular solution of (A46) is the Gegenbauer polynomial[9]

$$y_l(z) = C_l^{n/2-1}(z). \tag{A47}$$

From (A42)–(A44) and (A46) we have

$$e^{-\lambda r z} = \sum_{l=0}^{\infty} d_l (\lambda r)^{1-n/2} I_{l+n/2-1}(\lambda r) C_l^{n/2-1}(z). \tag{A48}$$

Using the orthogonality relation[9]

$$\int_{-1}^{+1} dz (1-z^2)^{(n-3)/2} C_l^{n/2-1}(z) C_{l'}^{n/2-1}(z) = A_l(n)\delta_{ll'} \tag{A49}$$

yields

$$A_l(n) d_l (\lambda r)^{1-n/2} I_{l+n/2-1}(\lambda r) = \int_{-1}^{+1} dz (1-z^2)^{(n-3)/2} e^{-\lambda r z} C_l^{n/2-1}(z). \tag{A50}$$

Expanding both sides into power series in (λr) and comparing the coefficients of $(\lambda r)^l$, the lowest power occurring, leads to

$$2^{1-l-n/2} A_l(n) d_l \frac{1}{\Gamma(l+n/2)} = \frac{1}{l!}\int_{-1}^{+1} dz (1-z^2)^{(n-3)/2} z^l P_l(z). \tag{A51}$$

As

$$C_l^{n/2-1}(z) = \beta_l(n) z^l + \text{lower-order terms}, \tag{A52}$$

with

$$\beta_l(n) = \frac{2^l}{l!}\frac{\Gamma(l+n/2-1)}{\Gamma(n/2-1)} \tag{A53}$$

from the recursion relation[9] of the Gegenbauer polynomials, we obtain from (A51) by inserting the inverted relation of (A52)

$$2^{1-l-n/2} A_l(n) d_l \frac{1}{\Gamma(l+n/2)} = \frac{1}{l!}\frac{1}{\beta_l(n)} A_l(n) \tag{A54}$$

or with (A53)

$$d_l = 2^{n/2-1}(-1)^l (l+n/2-1)\Gamma(n/2-1). \tag{A55}$$

As (A48) is analytic (even entire) in λ we may set $\lambda = -iq$. Using

$$I_\sigma(-iqz) = (-i)^\sigma J_\sigma(z) \tag{A56}$$

leads to (A38).

APPENDIX B: PROOF OF THEOREM 3.1

The basic idea of the proof of Theorem 3.1 is to determine the Laplace transform $\tilde{\psi}_{\mu,\beta}$ of the functions $\psi_{\mu,\beta}$ defined in (3.7) and to show that they are normalized, i.e.,

$$\tilde{\psi}_{\mu,\beta}(0) = 1, \tag{B1}$$

and completely monotone (c.m.), i.e.,

$$\left(-\frac{d}{dp}\right)^m \tilde{\psi}_{\mu,\beta}(p) \geqslant 0, \quad p > 0, \quad m \in Z_+, \tag{B2}$$

for $\mu > 0$ and $0 < \beta \leqslant \beta_c(\mu)$. Above the critical value β_c, given in (3.9a), (3.9b), and (3.9c), complete monotonicity breaks down. According to Bernstein's theorem[10] (B1) and (B2) are equivalent to Theorem 3.1.

The Laplace transform of (3.7) is given by

$$\tilde{\psi}_{\mu,\beta}(p) = [2^{1-\mu}/\Gamma(\mu)] p^{\mu\beta} K_\mu(p^\beta), \tag{B3}$$

where K_μ is the modified Bessel function of the second kind. To verify (B3) we determine the Mellin transform $\hat{\psi}_{\mu,\beta}$ twice: namely, directly from (3.7) and indirectly from (B3) using (1.17) and (2.12). In both cases the result is

$$\hat{\psi}_{\mu,\beta}(s) = \frac{1}{\Gamma(\mu)} 2^{(1-s)/\beta} \Gamma\Big(\mu + \frac{1-s}{2\beta}\Big) \times \Gamma\Big(1 + \frac{1-s}{2\beta}\Big) [\Gamma(2-s)]^{-1} . \tag{B4}$$

From the small z behavior[9] of $K_\mu(z)$ we obtain (B1) from (B3). The proof of (B2) will be split into several steps and consists in (i) showing the existence of a critical value $\beta_c(\mu)$ such that for $\beta \leqslant \beta_c(\mu)$ (B2) holds whereas for $\beta > \beta_c(\mu)$ it breaks down; and (ii) determination of upper and lower bounds for $\beta_c(\mu)$ for various subsets of $\mu > 0$.

Lemma B1: Let $f_{\mu,\beta}$, $\mu,\beta > 0$, be a two-parameter family of real-valued functions defined on the half-line $x > 0$ and having continuous derivatives of all orders. Assume that the relation

$$f_{\mu,\beta}(x) = f_{\mu,1}(x^\beta) \tag{B5}$$

holds.

(a) If $f_{\mu,\gamma}$ is c.m. for some $\gamma > 0$, then $f_{\mu,\beta}$ is c.m. for $\beta \leqslant \gamma$.

(b) If $f_{\mu,\gamma}$ is not c.m., then $f_{\mu,\beta}$ is not c.m. for $\beta \geqslant \gamma$.

Proof: (a) From (B5) we have

$$f_{\mu,\beta}(x) = f_{\mu,\gamma}(x^{\beta/\gamma}) . \tag{B6}$$

By the composition law[10] $f_{\mu,\beta}$ is c.m. for $\beta/\gamma \leqslant 1$ if $f_{\mu,\gamma}$ is c.m.

From now on we consider exclusively the family

$$f_{\mu,\beta}(x) = x^{\mu\beta} K_\mu(x^\beta) , \tag{B7}$$

which obviously satisfies (B5).

Lemma B3: Let $\mu < \frac{1}{2}$. Then $f_{\mu,1}$ is c.m.

Proof: This is immediate from the representation[5]

$$f_{\mu,1}(x) = x^\mu K_\mu(x) = \int_1^\infty dy\, e^{-xy} \frac{2^\mu \sqrt{\pi}}{\Gamma(\frac{1}{2} - \mu)} (y^2 - 1)^{-\mu - 1/2} . \tag{B8}$$

Corollary B4: From Lemma B3 and Corollary B2 we obtain the lower bound

$$1 \leqslant \beta_c(\mu), \quad \mu < \tfrac{1}{2} . \tag{B9}$$

Lemma B5: Set $f^0_{\mu,\beta} = f_{\mu,\beta}$ and

$$f^m_{\mu,\beta}(x) = \Big(-\frac{d}{dx}\Big)^m f_{\mu,\beta}(x), \quad m \in N , \tag{B10}$$

with $f_{\mu,\beta}$ given by (B7). Define further

$$A^0_0(\beta) = 1 \tag{B11}$$

and $A^m_k(\beta)$, $m \in N$, $k = 1,2,...,m$, recursively by

$$\begin{aligned} A^{m+1}_1(\beta) &= (m - 2\beta) A^m_1(\beta), \quad m \geqslant 1 , \\ A^{m+1}_k(\beta) &= (m - 2\beta k) A^m_k(\beta) + \beta A^m_{k-1}(\beta), \quad m \geqslant 2, \quad 2 \leqslant k \leqslant m , \\ A^{m+1}_{m+1}(\beta) &= \beta A^m_m(\beta), \quad m \geqslant 0. \end{aligned} \tag{B12}$$

Then

$$f^m_{\mu,\beta}(x) = \sum_{k=1}^m A^m_k(\beta) x^{\beta(\mu+k) - m} K_{\mu-k}(x^\beta), \quad m \in N . \tag{B13}$$

Proof: We note[9] that K_σ is even in σ and

$$-\frac{d}{dz} K_\sigma(z) = K_{\sigma-1}(z) + \frac{\sigma}{z} K_\sigma(z) . \tag{B14}$$

Thus we obtain

$$f^1_{\mu,\beta}(x) = -\frac{d}{dx}[x^{\mu\beta} K_\mu(x^\beta)] = \beta x^{\beta(\mu+1)-1} K_{\mu-1}(x^\beta) . \tag{B15}$$

This coincides with (B13) for $m = 1$ as $A^1_1(\beta) = \beta$ from (B14) and (B12). Assuming now (B13) to hold for some $m \geqslant 2$, we obtain by differentiation using (B14)

$$f^{m+1}_{\mu,\beta}(x) = \sum_{k=1}^m A^m_k(\beta) [(m - 2\beta k) x^{\beta(m+k)-(m+1)} K_{\mu-k}(x^\beta) + \beta x^{\beta(\mu+k+1)-(m+1)} K_{\mu-k-1}(x^\beta)] , \tag{B16}$$

which is (B13) with m replaced by $m+1$ in view of the recursion relation (B12).

Lemma B6: The functions $f_{\mu,\beta}$ defined in (B7) are c.m. for $\mu > 0$ and $\beta \leqslant \frac{1}{2}$.

Proof: As $K_\mu(z)$ is positive for z positive, it is sufficient to show that the coefficients $A^m_k(\beta)$ in (B13) are non-negative, which follows recursively from (B12) for $\beta \leqslant \frac{1}{2}$.

Corollary B7: From Lemma B6 we obtain the lower bound

$$\tfrac{1}{2} \leqslant \beta_c(\mu), \quad \mu > 0 . \tag{B17}$$

Lemma B8: For $\mu \geqslant 1$ and $\beta > \frac{1}{2}$ the functions $f_{\mu,\beta}$ from (B7) are not c.m.

Proof: From (B12) and (B13) we obtain

$$f^2_{\mu,\beta}(x) = \beta x^{\beta(\mu+1)-2} K_{\mu-1}(x^\beta) \{(1 - 2\beta) + \beta g_\mu(x^\beta)\} \tag{B18}$$

with

$$g_\mu(z) = z K_{\mu-2}(z) / K_{\mu-1}(z) . \tag{B19}$$

From the small z behavior[9] of $K_\sigma(z)$ we obtain

$$g_\mu(z) = \begin{cases} O(z^2), & \mu > 2, \\ O(z^{2\mu-2}), & 1 < \mu \leqslant 2, \end{cases} \tag{B20}$$

and

$$g_1(z) = O(|\ln z|^{-1}) . \tag{B21}$$

Hence the curly bracket in (B18) tends to $1 - 2\beta$ as x tends to zero. This quantity is negative by assumption. Thus $f^2_{\mu,\beta}(x)$ is negative for x sufficiently small. This implies that $f_{\mu,\beta}$ is not c.m.

Corollary B9: From Lemma B8 we obtain the upper bound

$$\beta_c(\mu) \leqslant \tfrac{1}{2}, \quad \mu \geqslant 1 . \tag{B22}$$

Lemma B10: Let $f^m_{-\mu,\beta}$ be given by (B7) and (B10). Then

$$f^m_{\mu,\beta}(x) = \beta x^{\beta(\mu+1)-m} \{P_m(x^\beta) K_{1-\mu}(x^\beta) + Q_m(x^\beta) K_\mu(x^\beta)\} , \tag{B23}$$

where $P_m(z)$, and $Q_m(z)$ are polynomials of degree less or equal to $m - 1$. The coefficients in

$$P_m(z) = \sum_{k=0}^{m-1} p^m_k z^k, \quad Q_m(z) = \sum_{k=0}^{m-1} q^m_k z^k \tag{B24}$$

are given by

$$p_0^1 = 1, \quad q_0^1 = 0, \tag{B25}$$

and by the recursion

$$\begin{aligned} p_m^{m+1} &= \beta q_{m-1}^m, \\ p_k^{m+1} &= [m - (2\mu + k)\beta] p_k^m + \beta q_{k-1}^m, \quad 0 \leqslant k \leqslant m-1, \\ q_m^{m+1} &= \beta p_{m-1}^m, \\ q_k^{m+1} &= [m - (k+1)\beta] q_k^m + \beta p_{k-1}^m, \quad 0 \leqslant k \leqslant m-1. \end{aligned} \tag{B26}$$

By convention, $p_{-1}^m = q_{-1}^m = 0$.

Proof: Using (B14) we obtain by differentiating (B7) the relation (B23) for the case $m = 1$ with $P_1 = 1$ and $Q_1 = 0$, i.e., (B24) with coefficients given by (B25). Differentiating (B23) yields $f_{\mu,\beta}^{m+1}$ expressed in terms of P_m, Q_m and their derivatives P'_m, Q'_m. Note that in view of (B14) the derivatives K'_μ and $K'_{1-\mu}$ may be eliminated in favor of $K_\mu, K_{1-\mu}$. Inserting (B24) into P_m, Q_m, P'_m, and Q'_m leads to (B26) by comparison with the relations (B23) and (B24) for $m+1$.

For $m = 2$, we obtain, from (B23)–(B26),

$$f_{\mu,\beta}^2(x) = \beta x^{\beta(\mu+1)-2}\{(1 - 2\mu\beta) K_{1-\mu}(x^\beta) + \beta x^\beta K_\mu(x^\beta)\}. \tag{B27}$$

As x tends to zero the first term in curly brackets is $O(x^{\beta(\mu-1)})$ for $\mu < 1$. Hence it dominates the second term, which is $O(x^{\beta(1-\mu)})$. Accordingly, $f_{\mu,\beta}^2$ becomes negative for x sufficiently small if $1 - 2\mu\beta$ is negative. This implies that $f_{\mu,\beta}$ is not c.m. for $\mu < 1$ and $2\mu\beta > 1$. Thus the following corollary holds.

Corollary B11: From Lemma B10 we obtain the upper bound

$$\beta_c(\mu) \leqslant (2\mu)^{-1}, \quad \mu < 1. \tag{B28}$$

For

$$\beta \leqslant \inf\{1, (2\mu)^{-1}\}, \tag{B29}$$

the square brackets in the recursion (B26) are non-negative for arbitrary m and $0 \leqslant k \leqslant m-1$. Hence all coefficients p_k^m, q_k^m are also non-negative. In view of (B23), (B24) this implies that $f_{\mu,\beta}$ is c.m. if β satisfies (B29). This result leads to the following corollary.

Corollary B12: From Lemma B10 we obtain the lower bound

$$\inf\{1, (2\mu)^{-1}\} \leqslant \beta_c(\mu), \tag{B30}$$

i.e.,

$$1 \leqslant \beta_c(\mu), \quad \mu \leqslant \tfrac{1}{2}, \tag{B31}$$

and

$$(2\mu)^{-1} \leqslant \beta_c(\mu), \quad \mu > \tfrac{1}{2}. \tag{B32}$$

Lemma B13: Let $\mu > 0$ and $\beta > 1$. Then $f_{\mu,\beta}$ defined in (B7) is not c.m.

Proof: In view of Lemma B1, statement (b), it is sufficient to prove that $f_{\mu,\beta}$ is not c.m. for β in the interval $(1,2]$. This is done by contradiction. Assume $f_{\mu,\beta}$ to be c.m. According to Bernstein's theorem[10] $f_{\mu,\beta}$ is the Laplace transform of a positive measure on R_+ [finite as $f_{\mu,\beta}(0+)$ is finite]. Hence $f_{\mu,\beta}$ has an analytic continuation into the half-plane $\mathrm{Re}(z) > 0$ and is bounded there. On the other hand, the asymptotic behavior of $f_{\mu,\beta}$ is given by[9]

$$f_{\mu,\beta}(z) \sim \sqrt{(\pi/2)} z^{\beta(\mu-1/2)} e^{-z^\beta}, \quad \beta|\arg(z)| < 3\pi/2, \tag{B33}$$

for $|z|$ large. Hence $f_{\mu,\beta}$ is growing exponentially in the two sectors

$$-\pi/2 < \arg(z) < -\pi/2\beta, \quad \pi/2\beta < \arg(z) < \pi/2 \tag{B34}$$

of the half-plane $\mathrm{Re}(z) > 0$ as $\cos[\beta \arg(z)] < 0$ there.

Combining the results of Corollaries B4, B7, B11, and B12 and Lemma B13 yields the results for $\beta_c(\mu)$ given in (3.9).

[1]W. Wyss, "The fractional diffusion equation," J. Math. Phys. **27**, 2782 (1986).
[2]B. L. J. Braaksma, "Asymptotic expansions and analytic continuations for a class of Barnes-integrals," Compositio Math. **15**, 239 (1964).
[3]H. M. Srivastava, K. C. Gupta, and S. P. Goyal, *The H-Functions of One and Two Variables with Applications* (South Asian, New Delhi, 1982).
[4]W. R. Schneider, "Stable distributions: Fox function representation and generalization," in *Stochastic Processes in Classical and Quantum Systems, Lecture Notes in Physics*, Vol. 262, edited by S. Albeverio, G. Casati, and D. Merlini (Springer, Berlin, 1986).
[5]A. Erdély, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Tables of Integral Transforms* (McGraw-Hill, New York, 1954).
[6]R. R. Nigmatullin, "The realization of the generalized transfer equation in a medium with fractal geometry," Phys. Status Solidi: B **133**, 425 (1986).
[7]T. Hida, *Brownian Motion* (Springer, New York, 1980).
[8]W. R. Schneider, "Grey noise and fractional diffusion," in *Stochastic Processes, Physics and Geometry*, Proceedings of the 2nd Ascona–Como International Conference, 1988, to be published.
[9]A. Erdély, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953).
[10]W. Feller, *An Introduction to Probability Theory and Its Applications* (Wiley, New York, 1971), Vol. II.

Low-density expansion for a two-state random walk in a random environment

Gregory F. Lawler[a)]
Courant Institute of Mathematical Sciences, 251 Mercer Street, New York, New York 10012

(Received 17 May 1988; accepted for publication 24 August 1988)

A nearest neighbor random walk on $\mathbb{Z}^2$ is considered where points of the lattice are labeled "good" or "bad." A particle takes a vertical step with probability a_G or a_B and a horizontal step with probability $1 - a_G$ or $1 - a_B$, depending on whether its present site is good or bad. Steps of size $+1$ are as probable as steps of size -1. If the good and bad sites are placed randomly, with density ρ of bad sites, it is known that there exists an $\bar{a}$ such that for almost every placement of sites, the random walk in the long run behaves like a homogeneous walk with vertical probability $\bar{a}$ and horizontal probability $1 - \bar{a}$. Here the problem of estimating $\bar{a}$ as a function of ρ is considered; in particular when $a_G = \frac{1}{2}$, a_B fixed, the first two terms of the expansion of $\bar{a}(\rho)$ at $\rho = 0$ are rigorously derived.

I. INTRODUCTION

We consider a two-state random walk on $\mathbb{Z}^2$. Let $0 < a_G$, $a_B < 1$, and π: $\mathbb{Z}^2 \to \{B,G\}$ be given. We call x a "bad" or "good" site depending on whether $\pi_x = B$ or $\pi_x = G$. In some papers, the term "scatterer" is used for a bad site. Let X_j be the nearest neighbor random walk with symmetric increments whose transition probabilities are given by

$$P\{X_{j+1} - X_j = \pm e_2\} = \tfrac{1}{2}(1 - 2P\{X_{j+1} - X_j = \pm e_1\})$$
$$= \begin{cases} \frac{1}{2}a_B, & \text{if } \pi_{X_j} = B, \\ \frac{1}{2}a_G, & \text{if } \pi_{X_j} = G. \end{cases}$$

Here e_1, e_2 are the standard unit vectors in $\mathbb{Z}^2$.

We now assume the assignment of B's and G's is random, i.e., let $\rho \in [0,1]$ be the density of bad points and we assume that $\{\pi_x\}$ are independent, identically distributed random variables with $P\{\pi_x = B\} = \rho$. It has been shown[1] that there exists an $\bar{a}$ such that for almost every environment π, the corresponding random walk has the same limiting behavior as a homogeneous walk with environment $\pi \equiv G$, $a_G = \bar{a}$. Clearly, $\bar{a}$ is a function of a_G, a_B, and ρ. Unfortunately, the proof of the existence of $\bar{a}$ does not give an easy way to calculate $\bar{a}(a_G, a_B, \rho)$—in general, it is not true that $\bar{a} = \rho a_B + (1 - \rho)a_G$.

In this paper we consider estimates of $\bar{a}$. We will fix a_G, a_B, and consider $\bar{a}(\rho)$. From the derivation of $\bar{a}$, it can be seen that for each ρ,

$$\bar{a}(\rho) = \lim_{m \to \infty} \bar{a}_m(\rho),$$

where $\bar{a}_m$ denotes the average using a random periodic environment with period m. (Place B's and G's randomly on an $m \times m$ square of sites and then extend periodically.) Roerdink and Shuler[2] considered the low-density limit for $\bar{a}_m(\rho)$; in particular, $\bar{a}_m(\rho)$ can be written as a polynomial in ρ:

$$\bar{a}_m(\rho) = a_G + b_{m,1}\rho + b_{m,2}\rho^2 + \cdots + b_{m,m^2}\rho^{m^2}.$$

They calculated $b_{m,1}$ and

$$b_1 \doteq \lim_{m \to \infty} b_{m,1},$$

which gives the first-order term for a low-density expansion for a fixed m. If $a_G = \frac{1}{2}$, then $b_{m,1} = b_1 = a_G - \frac{1}{2}$; however, if $a_G \neq \frac{1}{2}$, $b_1 \neq a_B - a_G$. The linear approximation, for fixed m, is good only for $\rho \lesssim 1/m^2$; hence their argument is not sufficient to determine the low-density limit for $\bar{a}$.

In this paper we consider the case $a_G = \frac{1}{2}$ and write the second-order expansion in ρ. We first compute

$$b_2 = \lim_{m \to \infty} b_{m,2}$$

and then derive uniform estimates to prove for some $\epsilon > 0$,

$$\bar{a}(\rho) = \tfrac{1}{2} + b_1\rho + b_2\rho^2 + O(\rho^{2+\epsilon}). \tag{1.1}$$

The expression we get for b_2 is in terms of some probabilities for simple random walk. Although we cannot write explicitly the exact value of b_2, we will easily see from our expression that $b_2 > 0$ for $a_B > \frac{1}{2}$ and $b_2 < 0$ for $a_B < \frac{1}{2}$. It immediately follows for ρ sufficiently small that

$$\begin{aligned} &\bar{a}(\rho) > \rho a_B + (1 - \rho)\tfrac{1}{2}, \quad a_B > \tfrac{1}{2}, \\ &\bar{a}(\rho) < \rho a_B + (1 - \rho)\tfrac{1}{2}, \quad a_B < \tfrac{1}{2}. \end{aligned} \tag{1.2}$$

We expect the inequalities (1.2) to hold for all $\rho \in (0,1)$, but have not proved it. While the estimates are done for a fixed a_B, it can be seen from the proof that the convergence in (1.1) can be made uniform for $a_B \in [c, 1, -c]$ for any $0 < c < \frac{1}{2}$.

In the case $a_G \neq \frac{1}{2}$, we have not considered the second derivative. We do note, however, that these methods could be used to prove

$$\bar{a}(\rho) = a_G + b_1\rho + O(\rho^{1+\epsilon}),$$

showing that b_1 does give the correct first-order term.

The outline of this paper is as follows: Section II reviews the derivation of $\bar{a}(\rho)$. Section III lists some lemmas about simple random walk and derives a few results about small perturbations of simple walks. Section IV does the periodic calculations and Sec. V outlines the expansion showing what uniform estimates are needed. The hard work of proving those estimates is saved for Sec. VI.

We wish to set some conventions about the use of constants throughout this paper. We use $0 < c_1 < c_2 < \infty$ to rep-

[a)] Permanent address: Department of Mathematics, Duke University, Durham, North Carolina 27706.

 0022-2488/89/010145-13$02.50

resent two universal constants—they may change from line to line. Similarly $0<c_3<c_4<\infty$ are universal constants that depend only on the (somewhat arbitrary) choice of exponents $(\alpha,\delta,\gamma,\theta,\kappa,\nu)$ made in the proof. The exponent $\epsilon>0$ also may change from line to line—it is our goal only to show that there exists an $\epsilon>0$ such that (1.1) holds. We also make use of the $O(\ )$ notation. Whenever this notation is used, it is assumed that the estimate is uniform, i.e., does not depend on the lattice point x or choice of environment π (it may depend on α, δ, γ, θ, κ, ν). All constants may depend on a_B; however, one can check through the proof that one could make the constants uniform for all $a_B\in[c,1-c]$ for any $c\in(0,\frac{1}{2})$.

II. DERIVATION OF THE EFFECTIVE DIFFUSION CONSTANT

Here we sketch the derivation of $\bar{a}(\rho)$ as in Ref. 1, including the facts we will need later. An environment π on $A\subset\mathbb{Z}^2$ is a function $\pi: A\to\{B,G\}$. We let $\mathscr{C}$ be the set of all environments on $\mathbb{Z}^2$ and use π_x to denote the value of π at $x\in\mathbb{Z}^2$.

We use T_m to denote the $m\times m$ torus in $\mathbb{Z}^2$, i.e., the equivalence classes under the relation $(x_1,x_2)\sim(y_1,y_2)$ if $x_1-y_1=k_1m$, $x_2-y_2=k_2m$, for some integers k_1,k_2. Any function π: $T_m\to\{B,G\}$ induces a periodic environment π: $\mathbb{Z}^2\to\{B,G\}$. We let $\mathscr{C}_m$ be the set of such environments.

Throughout this section, we fix a_G,a_B, $0<a_G,a_B<1$. If $\pi\in\mathscr{C}$ is fixed, there is a Markov chain with transition probabilities

$$\begin{aligned}
&p^\pi(x,x\pm e_2)=\tfrac{1}{2}a_G,\\
&p^\pi(x,x\pm e_1)=\tfrac{1}{2}(1-a_G),\quad \pi_x=G,\\
&p^\pi(x,x\pm e_2)=\tfrac{1}{2}a_B,\\
&p^\pi(x,x\pm e_1)=\tfrac{1}{2}(1-a_B),\quad \pi_x=B.
\end{aligned}$$

We will use X_j for the position of the chain at time j and E^π,P^π to denote expectations and probabilities with respect to this chain. If the π is deleted, then it will be assumed that the environment is all good, i.e., $\pi_x=G$, $\forall x\in\mathbb{Z}^2$. If $\pi\in\mathscr{C}_m$, then the Markov chain can be thought of as taking values in the finite set T_m.

Another way of viewing these Markov chains is to consider the Markov chain taking values in $\mathscr{C}$ with transition probabilities

$$\begin{aligned}
&p(\pi,T_{\pm e_2}\pi)=\tfrac{1}{2}a_G,\\
&p(\pi,T_{\pm e_1}\pi)=\tfrac{1}{2}(1-a_G),\quad \text{if } \pi_0=G,\\
&p(\pi,T_{\pm e_2}\pi)=\tfrac{1}{2}a_B,\\
&p(\pi,T_{\pm e_1}\pi)=\tfrac{1}{2}(1-a_B),\quad \text{if } \pi_0=B.
\end{aligned}$$

Here $T_x\pi$ denotes the translated environment $(T_x\pi)_y=\pi_{y+x}$.

For any ρ, $0\leqslant\rho\leqslant1$, let μ_ρ denote the Bernoulli measure on $\mathscr{C}$ with density ρ of bad points, i.e., the probability measure such that $\{\pi_x\}_{x\in\mathbb{Z}^2}$ are independent random variables with

$$\mu_\rho\{\pi_x=B\}=\rho.$$

Similarly, let $\mu_{\rho,m}$ be the Bernoulli measure on $\mathscr{C}_m$ that can be thought of as a measure on $\mathscr{C}$. It is routine that $\mu_{\rho,m}\to\mu_\rho$ weakly. It can also be shown that the Markov chain described above is ergodic with respect to the measure μ_ρ. Hence there exists at most one invariant probability measure λ_ρ for the chain that is mutually absolutely continuous with respect to μ_ρ. We sketch how such a λ_ρ is obtained.

For any environment $\pi\in\mathscr{C}_m$, the Markov chain induced on T_m is irreducible; hence there exists a unique invariant probabilty measure $\varphi^\pi(x)$. It is standard that if X_j takes value in T_m and

$$\tau_x=\inf\{j\geqslant1:\ X_j=x\},$$

then

$$\varphi^\pi(x)=[E_x^\pi(\tau_x)]^{-1}. \tag{2.1}$$

Let $h^\pi(x)$ be the density of this measure with respect to the uniform probability measure on T_m, i.e., $h^\pi(x)$ satisfies

$$\sum_{x\in T_m}h^\pi(x)=m^2,$$

$$\sum_{y\in T_m}h^\pi(y)p^\pi(y,x)=h^\pi(x),\quad x\in T_m.$$

Then one can easily verify that for every m,ρ the probability measure on $\mathscr{C}_m$,

$$\lambda_{m,\rho}(\pi)=h^\pi(0)\mu_{m,\rho}(\pi),$$

is invariant for the Markov chain. It follows from Ref. 1 that there exists a constant $c_2>0$ independent of π such that

$$\frac{1}{m^2}\sum_{x\in T_m}(h^\pi(x))^2\leqslant c_2^2,$$

which implies by translation invariance of $\mu_{m,\rho}$ that

$$\int_{\mathscr{C}}[h^\pi(0)]^2d\mu_{m,\rho}(\pi)\leqslant c_2^2. \tag{2.2}$$

Now standard arguments give that the sequence of measures $\lambda_{m,\rho}$ (for fixed ρ) is tight and hence converge to the (necessarily unique) probability measure λ_ρ. The amount of time spent on bad points is $\lambda_\rho(D_B)$, where $D_B=\{\pi\in\mathscr{C}:\pi_0=B\}$. It follows from the definition of weak convergence that

$$\lambda_\rho(D_B)=\lim_{m\to\infty}\lambda_{m,\rho}(D_B).$$

The martingale convergence theorem and the ergodic theorem can then be used to show the convergence (for almost every π) to the homogeneous process with $\bar{a}=\bar{a}(\rho)$ given by

$$\begin{aligned}
&\bar{a}_m(\rho)=a_B\lambda_{\rho,m}(D_B)+a_G(1-\lambda_{\rho,m}(D_B)),\\
&\bar{a}(\rho)=a_B\lambda_\rho(D_B)+a_G(1-\lambda_\rho(D_B)).
\end{aligned}$$

Note that $\bar{a}(\rho)=\lim_{m\to\infty}\bar{a}_m(\rho)$ follows.

If $A\subset\mathbb{Z}^2$ is a finite set and $\tilde{\pi}$ is an environment on A, let

$$D_{\tilde{\pi}}=\{\pi\in\mathscr{C}:\ \pi_x=\tilde{\pi}_x,\ x\in A\}.$$

Then by weak convergence

$$\lim_{m\to\infty}\lambda_{m,\rho}(D_{\tilde{\pi}})=\lambda_\rho(D_{\tilde{\pi}}).$$

Let g_ρ and $g_{m,\rho}$ denote the densities of λ_ρ and $\lambda_{m,\rho}$ with respect to the Bernoulli measure on environments on A, i.e.,

$g_\rho(\tilde\pi)\rho^{d(\tilde\pi)}(1-\rho)^{|A|-d(\tilde\pi)} = \lambda_\rho(D_{\tilde\pi})$

[and similarly for $g_{\rho,m}(\tilde\pi)$], where $d(\tilde\pi) = |\{x\in A:\ \tilde\pi_x = B\}|$. Then

$$\lim_{m\to\infty} g_{m,\rho}(\tilde\pi) = g_\rho(\tilde\pi).$$

If $A_1\subset A_2$, and $\tilde\pi_1$: $A_1\to\{B,G\}$, $\tilde\pi_2$: $A_2\to\{B,G\}$ we write $\tilde\pi_1\subset\tilde\pi_2$ if $(\tilde\pi_2)_x = (\tilde\pi_1)_x$, $x\in A_1$. Let P_ρ denote the conditional Bernoulli measure on environments on A_2,

$$P_\rho(\pi_2) = \rho^{e(\pi_2)}(1-\rho)^{|A_2|-|A_1|-e(\pi_2)},$$

where $e(\pi_2) = |\{x\in A_2\setminus A_1:\ (\pi_2)_x = B\}|$. Then it follows that

$$g(\tilde\pi_1) = \sum_{\tilde\pi_1\subset\tilde\pi_2} g(\tilde\pi_2)P_\rho(\tilde\pi_2). \tag{2.3}$$

For fixed n, let

$R_n = \{x\in\mathbb{Z}^2:\ |x_i|\leqslant n,\ i=1,2\}$,

$\operatorname{int} R_n = \{x\in\mathbb{Z}^2:\ |x_i|\leqslant n-1,\ i=1,2\}$,

$\partial R_n = R_n\setminus\operatorname{int} R_n$.

Let $\pi\in\mathscr{C}_m$ $(m>2n+1)$, X_j be the Markov chain on T_m, and

$$\xi_0 = \inf\{j\geqslant 1:\ X_j\in\{0\}\cup\partial R_n\}.$$

Then by the Markov property,

$$E_0^\pi(\tau_0) = E_0^\pi(\xi_0) + \sum_{y\in\partial R_n} P_0^\pi\{X_{\xi_0} = y\}E_y^\pi(\tau_0).$$

Note that in the above formula, $E_0^\pi(\xi_0)$ and $P_0^\pi\{X_{\xi_0}=y\}$ depend only on the values of π for $x\in R_n$. Also as $m\to\infty$, the first term becomes insignificant (since it remains constant while the other grows), i.e.,

$$\varphi^\pi(0) = [E_0^\pi(\tau_0)]^{-1}\sim\Big(\sum_{y\in\partial R_n} P_0^\pi\{X_{\xi_0}=y\}E_y^\pi(\tau_0)\Big)^{-1}.$$

Plugging into the definition g_ρ we can see that if $\tilde\pi$: $R_n\to\{B,G\}$,

$$g_\rho(\tilde\pi) = \lim_{m\to\infty}\sum_{\substack{\pi\in\mathscr{C}_m\\ \pi\subset\tilde\pi}} m^2\Big(\sum_{y\in\partial R_n} P_0^\pi\{X_{\xi_0}=y\}E_y^\pi(\tau_0)\Big)^{-1} \times\rho^{e(\pi)}(1-\rho)^{K-e(\pi)},$$

where again $e(\pi) = |\{x\in T_m\setminus R_n:\ \pi_x = B\}|$, $K = |T_m\setminus R_n|$. We will not use this formula except to derive one basic fact: suppose $\tilde\pi_G,\tilde\pi_B$: $R_n\to\{B,G\}$ agree at every point except the origin where $(\tilde\pi_G)_0 = G$, $(\tilde\pi_B)_0 = B$. Suppose also we have constants $k_1,k_2>0$ such that for all $y\in\partial R_n$,

$$k_1P_0^{\tilde\pi_G}\{X_{\xi_0}=y\}\leqslant P_0^{\tilde\pi_B}\{X_{\xi_0}=y\}\leqslant k_2P_0^{\tilde\pi_G}\{X_{\xi_0}=y\}. \tag{2.4}$$

Then since $E_x^{\tilde\pi_G}(\tau_0) = E_x^{\tilde\pi_B}(\tau_0)$ for $x\neq 0$,

$$(k_2)^{-1}g_\rho(\tilde\pi_G)\leqslant g_\rho(\tilde\pi_B)\leqslant(k_1)^{-1}g_\rho(\tilde\pi_G). \tag{2.5}$$

Our main technique in the following sections for deriving estimates on g_ρ will be estimating hitting probabilities as in (2.4), and using (2.5).

III. LEMMAS ON SIMPLE RANDOM WALK IN THE PLANE

Let $R = R_n = \{(z_1,z_2)\in\mathbb{Z}^2:\ |z_i|\leqslant n\}$ and for $x\in\mathbb{Z}^2$, $R(x) = R_n(x) = \{z+x:\ z\in R_n\}$. Let ∂R and $\operatorname{int} R$ be defined as in the previous section. Let X_j denote a simple random walk (i.e., homogeneous walk with $a_G=\frac{1}{2}$), let

$$\sigma = \inf\{j\geqslant 1:\ X_j\in\partial R\},$$

and if $x\in\operatorname{int} R_n$ let

$$\tau_x = \inf\{j\geqslant 1:\ X_j = x\},$$
$$\xi_x = \sigma\wedge\tau_x = \inf\{j\geqslant 1:\ X_j\in\{x\}\cup\partial R\}.$$

We start with two standard lemmas, stated without proof, of the hitting time σ. The first is essentially Harnack's inequality for the discrete Laplacian and the second is a standard estimate on hitting probabilities in $\mathbb{Z}^2$. As a reference for these estimates, see Ref. 3, especially Chap. III.

Lemma 3.1: There exist constants c_1,c_2 such that (a) for every $x\in R_{n/2}$, $y\in\partial R_n$

$$c_1P_0\{X_\sigma = y\}\leqslant P_x\{X_\sigma=y\}\leqslant c_2P_0\{X_\sigma=y\},$$

(b) for every $x\in R_n$, if $d = \operatorname{dist}(x,\partial R_n)$ and $z\in R_{d/2}(x)$, $y\in\partial R_n$,

$$c_1P_z\{X_\sigma=y\}\leqslant P_x\{X_\sigma=y\}\leqslant c_2P_z\{X_\sigma=y\}.$$

Lemma 3.2: Let $\alpha>0$. Then there exist constants c_3,c_4 (depending on α) such that if $x,z\in R_n$, $|x-z|\geqslant n^\alpha$, $\operatorname{dist}(x,\partial R)\geqslant n^\alpha$, then

(a) $c_3(\log n)^{-1}\leqslant P_x\{X_{\xi_x}\neq x\}\leqslant c_4(\log n)^{-1}$,

(b) $P_x\{X_{\xi_z}=z\}\leqslant c_4(\log n)^{-1}$.

We will need some estimates on the derivatives of the hitting probabilities. A function $f: R_n\to\mathbb{R}$ is called harmonic on $\operatorname{int} R_n$ if $f(x) = \frac{1}{4}(f(x+e_2)+f(x-e_2)+f(x+e_1)+f(x-e_1))$ for $x\in\operatorname{int} R_n$. The following lemma was proved by Brandt.[4] It is the discrete version of estimates for the first and second derivatives of the Poisson kernel for the usual Laplacian.

Lemma 3.3: There exist constants c_1,c_2 such that if $f: R_n\to\mathbb{R}$ is harmonic on $\operatorname{int} R_n$, then (a) if $x\in R_{c_1n}$, $|e|=1$,

$$|f(x)-f(x+e)|\leqslant(c_2/n)\sup_{y\in\partial R_n}|f(y)|,$$

(b) $|f(e_i)+f(-e_i)-2f(0)|\leqslant(c_2/n^2)\sup_{y\in\partial R_n}|f(y)|$, $\quad i=1,2.$

By combining this lemma with Lemmas 3.1 and 3.2 we get the estimates we will need.

Lemma 3.4: There exist constants c_1,c_2 such that if $H(x,y) = P_x\{X_\sigma=y\}$, then (a) if $x\in R_{c_1n}$, $|e|=1$, $y\in\partial R_n$,

$$|H(x,y)-H(x+e,y)|\leqslant(c_2/n)H(0,y),$$

(b) if $y\in\partial R_n$,

$$|H(e_i,y)+H(-e_i,y)-2H(0,y)|\leqslant(c_2/n^2)H(0,y),\quad i=1,2,$$

(c) if $0<\alpha<1$, there exists a $c_4>0$, such that if $x\in R_{n^\alpha}$, $y\in\partial R_n$,

$$|H(x,y)-H(0,y)|\leqslant c_4n^{\alpha-1}H(0,y).$$

Proof: To get (a) and (b) consider the function $f_y(x) = P_x\{X_\sigma = y\}$ for $x \in R_{n/2}$. By Lemma 3.1,

$$\sup_{z\in\partial R_{n/2}} P_z\{X_\sigma = y\} \leqslant c_2 P_0\{X_\sigma = y\}.$$

Hence by Lemma 3.3 we get the results. Part (c) is clearly obtained by repeated applications of part (a) since $R_{n^\alpha} \subset R_{c_1 n}$ for n large.

We now estimate how slight changes in the environment affect the hitting probability. Let π be the environment $\pi_x = G$, $x \neq 0$, $\pi_0 = B$. (Here we assume $a_G = \frac{1}{2}$, $0 < a_B < 1$.)

Lemma 3.5: There exists a constant $c_2 > 0$ such that $\forall y \in \partial R_n$,

(a) $|P_0^\pi\{X_{\xi_0} = y\} - P_0\{X_{\xi_0} = y\}|$
$\leqslant c_2[(\log n)/n^2] P_0\{X_{\xi_0} = y\}$,

(b) $|P_0^\pi\{X_\sigma = y\} - P_0\{X_\sigma = y\}|$
$\leqslant c_2[(\log n)/n^2] P_0\{X_\sigma = y\}$.

Proof: For any environment $\tilde{\pi}$ on R_n, $x \in \operatorname{int} R_n$, the Markov property gives

$$P_x^{\tilde\pi}\{X_\sigma = y\} = P_x^{\tilde\pi}\{X_{\xi_0} = y\} + P_x^{\tilde\pi}\{X_{\xi_0} = 0\} P_0^{\tilde\pi}\{X_\sigma = y\},$$

which for $x = 0$ gives

$$P_0^{\tilde\pi}\{X_\sigma = y\} P_0^{\tilde\pi}\{X_{\xi_0} \neq 0\} = P_0^{\tilde\pi}\{X_{\xi_0} = y\}. \tag{3.1}$$

By symmetry,

$$\begin{aligned} P_0\{X_{\xi_0} = 0\} &= P_0^\pi\{X_{\xi_0} = 0\} \\ &= P_{\pm e_i}\{X_{\xi_0} = 0\} \\ &= P_{\pm e_i}^\pi\{X_{\xi_0} = 0\}. \end{aligned} \tag{3.2}$$

Now, if $a = a_B$,

$$\begin{aligned} &P_0^\pi\{X_{\xi_0} = y\} \\ &= \tfrac{1}{2}a[P_{e_2}^\pi\{X_{\xi_0} = y\} + P_{-e_2}^\pi\{X_{\xi_0} = y\}] \\ &\quad + \tfrac{1}{2}(1-a)[P_{e_1}^\pi\{X_{\xi_0} = y\} + P_{-e_1}^\pi\{X_{\xi_0} = y\}] \\ &= \tfrac{1}{2}a[P_{e_2}\{X_{\xi_0} = y\} + P_{-e_2}\{X_{\xi_0} = y\}] \\ &\quad + \tfrac{1}{2}(1-a)[P_{e_1}\{X_{\xi_0} = y\} + P_{-e_1}\{X_{\xi_0} = y\}] \\ &= P_0\{X_{\xi_0} = y\} + \tfrac{1}{2}(a - \tfrac{1}{2})[P_{e_2}\{X_{\xi_0} = y\} \\ &\quad + P_{-e_2}\{X_{\xi_0} = y\} - 2P_0\{X_{\xi_0} = y\}] \\ &\quad + \tfrac{1}{2}(\tfrac{1}{2} - a)[P_{e_1}\{X_{\xi_0} = y\} \\ &\quad + P_{-e_1}\{X_{\xi_0} = y\} - 2P_0\{X_{\xi_0} = y\}]. \end{aligned}$$

However, by (3.2),

$$\begin{aligned} &P_{e_i}\{X_\sigma = y\} + P_{-e_i}\{X_\sigma = y\} - 2P_0\{X_\sigma = y\} \\ &= P_{e_i}\{X_{\xi_0} = y\} + P_{e_i}\{X_{\xi_0} = 0\}P_0\{X_\sigma = y\} \\ &\quad + P_{-e_i}\{X_{\xi_0} = y\} + P_{-e_i}\{X_{\xi_0} = 0\}P_0\{X_\sigma = y\} \\ &\quad - 2P_0\{X_{\xi_0} = y\} - 2P_0\{X_{\xi_0} = 0\}P_0\{X_\sigma = y\} \\ &= P_{e_i}\{X_{\xi_0} = y\} + P_{-e_i}\{X_{\xi_0} = y\} - 2P_0\{X_{\xi_0} = y\}. \end{aligned}$$

By Lemma 3.4(b) the above is bounded by $(c_2/n^2)P_0\{X_\sigma = y\}$ which by Lemma 3.1(a) and (3.1) gives a bound of

$$[(c_2 \log n)/n^2] P_0\{X_{\xi_0} = y\}.$$

This gives (a), and (3.1), (3.2), and Lemma 3.1(a) then give (b).

We will also need an estimate for environments with two bad points. Let $x,z \in \operatorname{int} R_n$ and let

$$\begin{aligned} \xi_{x,z} &= \xi_x \wedge \xi_z \\ &= \inf\{j \geqslant 1:\ X_j \in \{x\} \cup \{z\} \cup \partial R_n\}. \end{aligned}$$

We start with a standard two-dimensional estimate which we do not prove here.

Lemma 3.6: There exists a c_1 such that for all $x,w,z \in \operatorname{int} R_n$,

$$P_x\{X_{\xi_{w,z}} \notin \{w,z\}\} \geqslant c_1(\log n)^{-1}.$$

If $x \in \operatorname{int} R_n \setminus \{0\}$, let π_B^x denote the environment on R_n,

$$(\pi_B^x)_z = \begin{cases} B, & z \in \{0,x\}, \\ G, & z \notin \{0,x\}. \end{cases}$$

Lemma 3.7: For every $0 < \alpha < 1$, there exists a $c_4 < \infty$ such that if $x \in R_{n^\alpha}$, $y \in \partial R_n$,

$$|P_0^{\pi_B^x}\{X_\sigma = y\} - P_0\{X_\sigma = y\}| \leqslant c_4 n^{\alpha-1}(\log n) P_0\{X_\sigma = y\}.$$

Proof: We first consider the homogeneous environment. Let $\xi = \xi_{0,x}$. Then for any $z \in R_{n^\alpha}$, Lemma 3.4(c) gives

$$P_z\{X_\sigma = y\} = P_0\{X_\sigma = y\}(1 + O(n^{\alpha-1})). \tag{3.3}$$

But by the Markov property for any environment π

$$\begin{aligned} P_z^\pi\{X_\sigma = y\} &= P_z^\pi\{X_\xi = y\} + P_z^\pi\{X_\xi = 0\}P_0^\pi\{X_\sigma = y\} \\ &\quad + P_z^\pi\{X_\xi = x\}P_z^\pi\{X_\sigma = y\}. \end{aligned}$$

Hence by (3.3),

$$P_z\{X_\xi = y\} = P_z\{X_\sigma = y\}(P_z\{X_\xi \notin \{0,x\}\} + O(n^{\alpha-1}))$$

or, since $P_z\{X_\xi \notin \{0,x\}\} \geqslant c_1(\log n)^{-1}$,

$$P_z\{X_\xi = y | X_\xi \notin \{0,x\}\} = P_z\{X_\sigma = y\}(1 + O(n^{\alpha-1}\log n)). \tag{3.4}$$

Similarly

$$\begin{aligned} &P_z^{\pi_B^x}\{X_\xi = y | X_\xi \notin \{0,x\}\} \\ &\quad = P_z^{\pi_B^x}\{X_\sigma = y\}(1 + O(n^{\alpha-1}\log n)). \end{aligned}$$

But it is easy to see for any environment π

$$\inf_z (P_z^\pi\{X_\xi = y | X_\xi \notin \{0,x\}\})$$
$$\leqslant P_0^\pi\{X_\sigma = y\} \leqslant \sup_z P_z^\pi\{X_\xi = y | X_\xi \notin \{0,x\}\},$$

where in each case the inf or sup is taken over the set

$$\{\pm e_1, \pm e_2, x \pm e_1, x \pm e_2\} \setminus \{0,x\}.$$

But over this set the lhs and rhs do not change if we replace the homogeneous environment with π_B^x. Hence

$$P_0^{\pi_B^x}\{X_\sigma = y\} = P_0\{X_\sigma = y\}(1 + O(n^{\alpha-1}\log n)).$$

If Y_j is a one-dimensional simple random walk starting at x, $0 < x < n$, and $\sigma = \inf\{j \geqslant 1:\ Y_j \in \{0,n\}\}$, then it is well known that $P_x\{X_\sigma = 0\} = (n-x)/n$. For X_j a two-dimensional simple random walk let

$$\bar{\sigma} = \inf\{j\geqslant 1:\ X_j\notin\partial R_n \cup \partial R_{n/2}\}.$$

Then by considering only one component, one can check that there exists a c_2 such that if $x\in R_n$, $\mathrm{dist}(x,\partial R_n)\leqslant an$, then

$$P_x\{X_{\bar\sigma}\notin \partial R_{n/2}\}\leqslant ac_2.$$

Since by Lemma 3.2, for every $y\in\partial R_{n/2}$, $P_y\{X_{\xi_0}=0\}\leqslant c_2(\log n)^{-1}$, we conclude

$$P_x\{X_{\xi_0}=0\}\leqslant ac_2(\log n)^{-1}. \tag{3.5}$$

We use this to prove the following lemma.

Lemma 3.8: For every $\alpha>0$, there exists a $c_4>0$ such that if $x\in R_{n^\alpha}$ and $\eta=\inf\{j\geqslant 1:\ X_j\in\{0,x\}\cup\partial R_n\}$,

$$|P_x\{X_\eta\in\partial R_n\}-P_0\{X_\eta\in\partial R_n\}| \leqslant c_4 P_0\{X_\eta\in\partial R_n\}n^{\alpha-1}(\log n)^{-1}.$$

Proof: Let $R_0=R_{n-n^\alpha}(0)$, $R_x=R_{n-n^\alpha}(x)$. Let

$$\eta_0=\inf\{j\geqslant 1:\ X_j\in\{0,x\}\cup\partial R_0\},$$
$$\eta_x=\inf\{j\geqslant 1:\ X_j\in\{0,x\}\cup\partial R_x\}.$$

Then by symmetry

$$P_0\{X_{\eta_0}\notin\{0,x\}\}=P_x\{X_{\eta_x}\notin\{0,x\}\}.$$

By the work leading to (3.5) we get

$$P_0\{X_\eta\notin\{0,x\}\}=P_0\{X_{\eta_0}\notin\{0,x\}\}(1-O(n^{\alpha-1}(\log n)^{-1})),$$
$$P_x\{X_\eta\notin\{0,x\}\}=P_x\{X_{\eta_x}\notin\{0,x\}\}(1-O(n^{\alpha-1}(\log n)^{-1})).$$

We end this section by noting that the proofs of Lemmas 3.7 and 3.8 imply that $\forall y\in\partial R_n$, $x\in\partial R_{n^\alpha}$,

$$P_x\{X_\eta=y\}=P_0\{X_\eta=y\}(1+O(n^{\alpha-1}\log n)), \tag{3.6}$$

where again $\eta=\inf\{j\geqslant 1:\ X_j\in\{0,x\}\cup\partial R_n\}$.

IV. PERIODIC CASE

In this section we fix m and consider

$$\bar a_m(\rho)=a_B\lambda_{\rho,m}(D_B)+a_G(1-\lambda_{\rho,m}(D_B))$$

as $\rho\to 0$. In particular, we take the first two derivatives with respect to ρ at $\rho=0$ (we take the second derivative only when $a_G=\frac12$). The first derivative was previously computed by Roerdink and Shuler.[2]

As before, if $\pi\in\mathscr{C}_m$, let $d(\pi)=|\{x\in T_m:\ \pi_x=B\}|$. Then

$$\lambda_{\rho,m}(D_B)=\sum_{\pi\in\mathscr{C}_m\cap D_B} h^\pi(0)\rho^{d(\pi)}(1-\rho)^{m^2-d(\pi)}. \tag{4.1}$$

Since this is a polynomial in ρ for fixed m, in order to take the first two derivatives at $\rho=0$ we need only consider terms of order ρ and ρ^2. Let $\bar\pi$ denote the environment $(\bar\pi)_0=B$, $(\bar\pi)_y=G$, $y\neq 0$. If $x\neq 0$, let π^x be the environment $(\pi^x)_0=(\pi^x)_x=B$, $(\pi^x)_y=G$, $y\notin\{0,x\}$. Then by expanding (4.1) we get

$$\lambda_{\rho,m}(D_B)=\rho h^{\bar\pi(0)}+\sum_{x\neq 0}(h^{\pi^x}(0)-1)\rho^2+O(\rho^3). \tag{4.2}$$

We first consider

$$h^{\bar\pi}(0)=m^2\varphi^{\bar\pi}(0)=m^2[E_0^{\bar\pi}(\tau_0)]^{-1}.$$

If we use $E_0(\tau_0)$ to denote the expectation assuming all points are good, we know since the uniform probability measure is invariant that

$$E_0(\tau_0)=m^2.$$

Also, by considering one step, we get

$$E_0(\tau_0)=\tfrac12 a_G(E_{e_2}(\tau_0)+E_{-e_2}(\tau_0)) +\tfrac12(1-a_G)(E_{e_1}(\tau_0)+E_{-e_1}(\tau_0))+1.$$

By standard but relatively tedious computations using Markov chains (essentially doing the calculations in the appendix of Ref. 2), one can show that

$$\lim_{m\to\infty}\frac{1}{m^2}E_{\pm e_2}(\tau_0)=\frac{4}{\pi a_G}\tan^{-1}\left(\frac{a_G}{1-a_G}\right)^{1/2}-1,$$
$$\lim_{m\to\infty}\frac{1}{m^2}E_{\pm e_1}(\tau_0)=\frac{4}{\pi(1-a_G)}\tan^{-1}\left(\frac{1-a_G}{a_G}\right)^{1/2}-1. \tag{4.3}$$

Similarly,

$$\begin{aligned}E_0^{\bar\pi}(\tau_0)&=\tfrac12 a_B(E_{e_2}^{\bar\pi}(\tau_0)+E_{-e_2}^{\bar\pi}(\tau_0)) +\tfrac12(1-a_B)(E_{e_1}^{\bar\pi}(\tau_0)+E_{-e_1}^{\bar\pi}(\tau_0))+1\\ &=\tfrac12 a_B(E_{e_2}(\tau_0)+E_{-e_2}(\tau_0)) +\tfrac12(1-a_B)(E_{e_1}(\tau_0)+E_{-e_1}(\tau_0))+1\\ &=m^2+\tfrac12(a_B-a_G)(E_{e_2}(\tau_0)+E_{-e_2}(\tau_0)) +\tfrac12(a_G-a_B)(E_{e_1}(\tau_0)+E_{-e_1}(\tau_0)).\end{aligned}$$

Hence

$$\begin{aligned}\lim_{m\to\infty} m^2[E_0^{\bar\pi}(\tau_0)]^{-1} &=\left[1+(a_B-a_G)\left(\frac{4}{\pi a_G}\tan^{-1}\left(\frac{a_G}{1-a_G}\right)^{1/2} -\frac{4}{\pi(1-a_G)}\tan^{-1}\left(\frac{1-a_G}{a_G}\right)^{1/2}\right)\right]^{-1}\\ &\doteq r(a_G,a_B)\end{aligned}$$

and

$$b_1=\lim_{m\to\infty}\left[\frac{\partial}{\partial\rho}\bar a_{\rho,m}\right]\Bigg|_{\rho=0}=(a_B-a_G)r(a_G,a_B). \tag{4.4}$$

Note that if $a_G=\frac12$, $r(a_G,a_B)=1$ and $b_1=a_B-a_G$.

For the remainder of this section we will assume $a_G=\frac12$ and we will compute the second derivative of $\lambda_{\rho,m}(D_B)$ at $\rho=0$, which, by (4.2), is

$$2\sum_{y\neq 0}(h^{\pi^y}(0)-1).$$

We first state the result. If $y=(y_1,y_2)$, let $\bar y=(y_2,y_1)$. For $y\neq 0$, let X_j be a simple (homogeneous) random walk,

$$\tilde\tau=\inf\{j\geqslant 0:\ X_j\in\{0,y\}\},$$
$$\tau=\inf\{j\geqslant 1:\ X_j\in\{0,y\}\},$$

and

$$\Phi_m(y)=\frac{P_{e_2}\{X_{\tilde\tau}=y\}+P_{-e_2}\{X_{\tilde\tau}=y\}-P_{e_1}\{X_{\tilde\tau}=y\}-P_{-e_1}\{X_{\tilde\tau}=y\}}{2P_0\{X_\tau=y\}}. \tag{4.5}$$

Also for fixed y, set

$$\Phi(y)=\lim_{m\to\infty}\Phi_m(y).$$

[$\Phi(y)$ can be defined equivalently by using the expression on the rhs of (4.5) taking X_j to be a simple random walk on all of $\mathbb{Z}^2$.] Note that $\Phi_m(\bar y)=-\Phi_m(y)$. Then the result is for $y\neq 0$,

$$(h^{\pi^y}(0)-1)+(h^{\pi^{\bar y}}(0)-1)=\frac{2((a_B-\frac12)\Phi_m(y))^2}{1-((a_B-\frac12)\Phi_m(y))^2} \tag{4.6}$$

and hence

$$B_2\doteq\lim_{m\to\infty}\sum_{y\neq0}(h^{\pi^y}(0)-1)=\sum_{y\in\mathbb{Z}^2}\frac{((a_B-\frac12)\Phi(y))^2}{1-((a_B-\frac12)\Phi(y))^2}.$$

We will show that for some $c_2>0$, $|\Phi(y)|<c_2(\log|y|)|y|^{-2}$; hence the sum is finite [one can actually prove that $|\Phi(y)|\leqslant c_2|y|^{-2}$, but we will not need this]. Note also that the sum is positive.

Let $y\neq0$ be fixed and let us derive (4.6). We need a slight generalization of (2.1): Suppose X is an irreducible Markov chain on a finite state space S and $A\subset S$. Let φ be the invariant probability measure on S. Then if $\tau_A=\inf\{j\geqslant1: X_j\in A\}$,

$$\sum_{x\in A}\varphi(x)E_x(\tau_A)=1. \tag{4.7}$$

(We omit the proof, leaving it as an exercise for the reader who has not seen this.) We will use this with $A=\{0,y\}$ or $\{0,\bar y\}$. Let

$$\tau_y=\inf\{j\geqslant1: X_j\in\{0,y\}\},\quad \tilde\tau_y=\inf\{j\geqslant0: X_j\in\{0,y\}\},$$
$$\sigma=\inf\{j\geqslant1: X_j=0\},\quad \tilde\sigma=\inf\{j\geqslant0: X_j=0\}.$$

By (4.7) we have

$$h^{\pi^y}(0)E_0^{\pi^y}(\tau_y)+h^{\pi^y}(y)E_y^{\pi^y}(\tau_y)=m^2.$$

However by symmetry $h^{\pi^y}(0)=h^{\pi^y}(y)$ and $E_0^{\pi^y}(\tau_y)=E_y^{\pi^y}(\tau_y)$. Hence for any $y\neq0$,

$$h^{\pi^y}(0)=(m^2/2)[E_0^{\pi^y}(\tau_y)]^{-1}. \tag{4.8}$$

By considering one step, we get

$$E_0^{\pi^y}(\tau_y)=\tfrac12a_B(E_{e_2}(\tilde\tau_y)+E_{-e_2}(\tilde\tau_y))+\tfrac12(1-a_B)(E_{e_1}(\tilde\tau_y)+E_{-e_1}(\tilde\tau_y))+1,$$

$$E_0^{\pi^{\bar y}}(\tau_{\bar y})=\tfrac12a_B(E_{e_2}(\tilde\tau_{\bar y})+E_{-e_2}(\tilde\tau_{\bar y}))+\tfrac12(1-a_B)(E_{e_1}(\tilde\tau_{\bar y})+E_{-e_1}(\tilde\tau_{\bar y}))+1.$$

Note on the right-hand side of the above expressions the expectations are with respect to the homogeneous random walk. By symmetry,

$$E_{e_1}(\tilde\tau_y)+E_{-e_1}(\tilde\tau_y)=E_{e_2}(\tilde\tau_{\bar y})+E_{-e_2}(\tilde\tau_{\bar y}),$$
$$E_{e_1}(\tilde\tau_{\bar y})+E_{-e_1}(\tilde\tau_{\bar y})=E_{e_2}(\tilde\tau_y)+E_{-e_2}(\tilde\tau_y).$$

Hence

$$E_0^{\pi^y}(\tau_y)+E_0^{\pi^{\bar y}}(\tau_{\bar y})=\tfrac12(E_{e_2}(\tilde\tau_y)+E_{-e_2}(\tilde\tau_y))+\tfrac12(E_{e_1}(\tilde\tau_y)+E_{-e_1}(\tilde\tau_y))+2=2E_0(\tau_y).$$

By using (4.8) for the simple random walk (with uniform invariant probability) and $A=\{0,y\}$, we see that $E_0(\tau_y)=\frac12m^2$; therefore

$$E_0^{\pi^y}(\tau_y)+E_0^{\pi^{\bar y}}(\tau_{\bar y})=m^2. \tag{4.9}$$

We also get

$$E_0^{\pi^y}(\tau_y)-E_0^{\pi^{\bar y}}(\tau_{\bar y})=(\tfrac12-a_B)(E_{e_1}(\tilde\tau_y)+E_{-e_1}(\tilde\tau_y))+(a_B-\tfrac12)(E_{e_2}(\tilde\tau_y)+E_{-e_2}(\tilde\tau_y)). \tag{4.10}$$

For the remainder we fix y and write $\tau,\tilde\tau$ for $\tau_y,\tilde\tau_y$. For any x,

$$E_x(\sigma)=E_x(\tau)+P_x\{X_\tau=y\}E_y(\sigma).$$

Similarly for $x\neq0$,

$$E_x(\tilde\tau)=E_x(\sigma)-P_x\{X_{\tilde\tau}=y\}E_y(\sigma).$$

Plugging into (4.10), noting that $E_{\pm e_1}(\sigma)=E_{\pm e_2}(\sigma)$, gives

$$E_0^{\pi^y}(\tau_y)-E_0^{\pi^{\bar y}}(\tau_{\bar y})=E_y(\sigma)(a_B-\tfrac12)[P_{e_2}\{X_{\tilde\tau}=y\}+P_{-e_2}\{X_{\tilde\tau}=y\}-P_{e_1}\{X_{\tilde\tau}=y\}-P_{-e_1}\{X_{\tilde\tau}=y\}].$$

Now

$$E_y(\sigma)=E_y(\tau)+P_y\{X_\tau=y\}E_y(\sigma)$$

or

$$E_y(\sigma)P_y\{X_\tau=0\}=E_y(\tau).$$

Using (4.7) again we see that $E_y(\tau)=\frac12m^2$. Also symmetry gives that $P_y\{X_\tau=0\}=P_0\{X_\tau=y\}$. Hence

$$E_0^{\pi^y}(\tau_y)-E_0^{\pi^{\bar y}}(\tau_{\bar y})=(a_B-\tfrac12)\Phi_m(y)m^2. \tag{4.11}$$

Therefore by (4.9) and (4.11) we can solve to get

$$E_0^{\pi^y}(\tau_y)=m^2\left(\frac12+\frac{(a_B-\frac12)}{2}\Phi_m(y)\right). \tag{4.12}$$

From (4.8), then

$$h^{\pi^y}(0)+h^{\pi^{\bar y}}(0)=[(1+(a_B-\tfrac12)\Phi_m(y))]^{-1}+[(1-(a_B-\tfrac12)\Phi_m(y))]^{-1}=2(1-((a_B-\tfrac12)\Phi_m(y))^2)^{-1}.$$

Or

$$(h^{\pi^y}(0)-1)+(h^{\pi^{\bar y}}(0)-1)=\frac{2((a_B-\frac12)\Phi_m(y))^2}{1-((a_B-\frac12)\Phi_m(y))^2}$$

which is (4.6). We end by proving the estimate $\Phi(y) \leqslant c_2(\log|y|)|y|^{-2}$ using the kind of argument that will be used in Sec. VI. We note that if $|y|>1$, $\tau=\inf\{j\geqslant 1: X_j=y\}$,

$$\tfrac12\Phi(y)(a_B-\tfrac12)=\frac{P_0^{\bar\pi}\{X_\tau=y\}-P_0\{X_\tau=y\}}{P_0\{X_\tau=y\}},$$

where $\bar\pi$ denotes the environment with a single bad point at 0. For any environment π,

$$P_0^\pi\{X_\tau=y\}=\sum_{z\in\partial R_{|y|/2}} P_0^\pi\{X_{\xi_0}=z\}P_z^\pi\{X_\tau=y\},$$

where $\xi_0=\inf\{j\geqslant 1: X_j\in\{0\}\cup\partial R_{|y|/2}\}$. By Lemma 3.5(a), if $z\in\partial R_{|y|/2}$

$$|P_0^{\bar\pi}\{X_{\xi_0}=z\}-P_0\{X_{\xi_0}=z\}| \leqslant c_2[(\log|y|)/|y|^2]P_0\{X_{\xi_0}=z\}.$$

But for such z, $P_z^{\bar\pi}\{X_\tau=y\}=P_z\{X_\tau=y\}$, hence

$$P_0^{\bar\pi}\{X_\tau=y\}=P_0\{X_\tau=y\}(1+O((\log|y|)/|y|^2)),$$

which gives the estimate.

V. LOW-DENSITY EXPANSION

Let $\frac14<\alpha<\frac27$ and let κ satisfy $4\alpha-1<\kappa<1-3\alpha$. Let

$$R=R_{\rho^{-\alpha}}=\{z\in Z^2: |z_i|\leqslant\rho^{-\alpha}\}.$$

Let $\mathscr{D}=\mathscr{D}_\rho$ be the set of environments on $R\setminus\{0\}$. Let $\pi\in\mathscr{D}$ become an environment on R by choosing a value at 0. We let π_B,π_G denote the two possibilities. Let $n=n_\rho=|R\setminus\{0\}|\sim 4\rho^{-2\alpha}$ and $d(\pi)=|\{x\in R\setminus\{0\}: \pi_x=B\}|$. We then have

$$\lambda_\rho(D_B)=\rho\left(\sum_{k=0}^n \rho^{k(1-\rho)^{n-k}}\sum_{\substack{\pi\in\mathscr{D}\\ d(\pi)=k}} g(\pi_B)\right)$$

$$=\rho+\rho\left(\sum_{k=0}^n u_k\,\rho^k(1-\rho)^{n-k}\right), \qquad (5.1)$$

where

$$u_k=\sum_{d(\pi)=k}(g(\pi_B)-1). \qquad (5.2)$$

We let $\bar\pi\in\mathscr{D}$ denote the environment $\bar\pi_y=G, y\in R\setminus\{0\}$ and for $x\in R\setminus\{0\}$, we let π^x be the environment $(\pi^x)_y=G$, $y\neq x$, $(\pi^x)_x=B$. The estimates we need are included in the following lemma which will be proved in Sec. VI.

Lemma 5.1:

(a) There exists a c_2 such that for every $\pi\in\mathscr{D}$,

$$g(\pi_B)\leqslant c_2^{d(\pi)},\quad g(\pi_G)\leqslant c_2^{d(\pi)}.$$

(b) There exists a c_4 such that if $\pi\in\mathscr{D}$ with $d(\pi)=2$, $\pi_x=B$, $x_y=B$, and $\min\{|x|,|y|,|x-y|,\mathrm{dist}(x,\partial R),\mathrm{dist}(y,\partial R)\}\geqslant\rho^{-\kappa}$,

$$g(\pi_B)=g(\bar\pi_G)(1+O(\rho^{2\kappa}\log(1/\rho))),$$

(c) $g(\bar\pi_B)=g(\bar\pi_G)(1+O(\rho^{1+\epsilon}))$,

(d) $\sum_{x\in R}(g(\pi_G^x)-g(\bar\pi_G))=O(\rho^\epsilon)$,

(e) $\sum_{x\in R}(g(\pi_B^x)-g(\bar\pi_G)=(1+O(\rho^\epsilon))B_2$,

where B_2 is defined as in Sec. IV.

We now proceed to use Lemma 5.1 to derive the expansion. A simple combinatorial estimate gives

$$|\{\pi\in\mathscr{D}: d(\pi)=k\}|\leqslant O(\rho^{-2\alpha k}).$$

Hence by Lemma 5.1(a)

$$u_k\leqslant c_2^k O(\rho^{-2\alpha k}),$$
$$\sum_{k=3}^\infty u_k\rho^k(1-\rho)^{n-k}\leqslant O(\rho^{3(1-2\alpha)})=O(\rho^{1+\epsilon}). \qquad (5.3)$$

For the $k=2$ term we note that the cardinality of environments with two bad points is $O(\rho^{-4\alpha})$. However, the set of environments such that the two bad points x and y do not satisfy $\min\{|x|,|y|,|x-y|,\mathrm{dist}(x,\partial R),\mathrm{dist}(y,\partial R)\}\geqslant\rho^{-\kappa}$ has cardinality $O(\rho^{-3\alpha-\kappa})$. Hence by Lemma 5.1(a) and (b),

$$u_2=\sum_{d(\pi)=2}(g(\pi_B)-1)=O(\rho^{-4\alpha})O(\rho^{2\kappa}\log(1/\rho)) + O(\rho^{-3\alpha-\kappa})c_2^2 \leqslant O(\rho^{-(1-\epsilon)}). \qquad (5.4)$$

Hence (5.1) becomes

$$\lambda_\rho(D_B)=\rho+\rho\Big[(g(\bar\pi_B)-1)(1-\rho)^n + \sum_x(g(\pi_B^x)-1)\rho(1-\rho)^{n-1}+O(\rho^{1+\epsilon})\Big]. \qquad (5.5)$$

The next lemma shows that the all good environment has measure near one.

Lemma 5.2: $g(\bar\pi_G)=1+O(\rho^{1+\epsilon})$.

Proof: Since λ_ρ is a probabilty measure,

$$1=\sum_{k=0}^n\rho^k(1-\rho)^{n-k}\Big(\sum_{d(\pi)=k}g(\pi_B)\rho+g(\pi_G)(1-\rho)\Big)$$

$$=g(\bar\pi_G)+\sum_{k=0}^n\rho^k(1-\rho)^{n-k}(s_k\rho+t_k(1-\rho)), \qquad (5.6)$$

where

$$s_k=\sum_{d(\pi)=k}(g(\pi_B)-g(\bar\pi_G)),$$

$$t_k=\sum_{d(\pi)=k}(g(\pi_G)-g(\bar\pi_G)).$$

The estimates used to get (5.3) can be used to show for $k\geqslant 3$,

$$\rho^k(1-\rho)^{n-k}(s_k\rho+t_k(1-\rho))=O(\rho^{1+\epsilon}).$$

Similarly an estimate like that to get (5.4) shows

$$\rho^2(1-\rho)^{n-2}(s_2\rho+t_2(1-\rho))=O(\rho^{1+\epsilon}),$$
$$\rho(1-\rho)^{n-1}(s_1\rho)=O(\rho^{1+\epsilon}).$$

Finally, Lemma 5.1(c) and (d) give

$$\rho(1-\rho)^{n-1}(t_1\rho) = O(\rho^{1+\epsilon}),$$
$$(1-\rho)^n(s_0\rho) = O(\rho^{1+\epsilon}).$$

Hence

$$1 = g(\bar{\pi}_G) + O(\rho^{1+\epsilon}).$$

Returning to (5.5) using Lemma 5.1(c) we get

$$\lambda_\rho(D_B) = \rho + \rho\Big[O(\rho^{1+\epsilon})(1-\rho)^n + \sum_x (g(\pi^x_B) - g(\bar{\pi}_G))\rho(1-\rho)^{n-1} \times (1 + O(\rho^\epsilon)) + O(\rho^{1+\epsilon})\Big],$$

and then using Lemma 5.1(e) we get

$$\lambda_\rho(D_B) = \rho + B_2\rho^2 + O(\rho^{2+\epsilon})$$

and hence

$$\bar{a}(\rho) = \tfrac{1}{2} + \rho + b_2\rho^2 + O(\rho^{2+\epsilon}),$$

where $b_2 = (a_B - \frac{1}{2})B_2$. (Note since $B_2 > 0$, $b_2 > 0$ if $\bar{a}_B > \frac{1}{2}$ and $b_2 < 0$ if $a_B < \frac{1}{2}$.)

VI. PROOF OF ESTIMATES

In this section we prove Lemma 5.1. We start with a crude estimate on the effect of changing one value of an environment.

Lemma 6.1: Let $a = \min(a_B, 1-a_B)$. Then if $\pi_1, \pi_2 \in \mathscr{C}_m$ and $(\pi_1)_x = (\pi_2)_x$ for $x \neq y$; $(\pi_1)_y = G$, $(\pi_2)_y = B$; then

$$h(\pi_2) \leqslant ((1-a)/a)h(\pi_1).$$

Proof: Suppose first that $y = 0$. Then

$$\begin{aligned} E_0^{\pi_2}(\tau_0) &= \tfrac{1}{2}a_B[E_{e_2}^{\pi_2}(\tau_0) + E_{-e_2}^{\pi_2}(\tau_0)] \\ &\quad + \tfrac{1}{2}(1-a_B)[E_{e_1}^{\pi_2}(\tau_0) + E_{-e_1}^{\pi_2}(\tau_0)] \\ &= \tfrac{1}{2}a_B[E_{e_2}^{\pi_1}(\tau_0) + E_{-e_2}^{\pi_1}(\tau_0)] \\ &\quad + \tfrac{1}{2}(1-a_B)[E_{e_1}^{\pi_1}(\tau_0) + E_{-e_1}^{\pi_1}(\tau_0)] + 1 \\ &\geqslant 2aE_0^{\pi_1}(\tau_0). \end{aligned}$$

Hence $h(\tau_2) \leqslant (2a)^{-1}h(\pi_1) \leqslant ((1-a)/a)h(\pi_1)$. Similarly, suppose $y \neq 0$ and let $\xi = \inf\{j > 0: X_j \in \{0, y\}\}$. Then

$$\begin{aligned} E_0^{\pi_2}(\tau_0) &= E_0^{\pi_2}(\xi) + P_0^{\pi_2}\{X_\xi = y\}E_y^{\pi_2}(\tau_0) \\ &= E_0^{\pi_1}(\xi) + P_0^{\pi_1}\{X_\xi = y\}E_y^{\pi_2}(\tau_0), \end{aligned} \tag{6.1}$$

$$E_y^{\pi_2}(\tau_0) = E_y^{\pi_2}(\xi) + P_y^{\pi_2}\{X_\xi = y\}E_y^{\pi_2}(\tau_0).$$

The second equation implies

$$E_y^{\pi_2}(\tau_0) = E_y^{\pi_2}(\xi)[P_y^{\pi_2}\{X_\xi \neq 0\}]^{-1}. \tag{6.2}$$

Using an argument as above one can show

$$E_y^{\pi_2}(\xi) \geqslant 2aE_0^{\pi_1}(\xi),$$
$$P_y^{\pi_2}\{X_\xi = 0\} \leqslant 2(1-a)P_0^{\pi_1}\{X_\xi = 0\}.$$

Therefore

$$E_0^{\pi_2}(\tau_0) \geqslant E_0^{\pi_1}(\xi) + [a/(1-a)]P_0^{\pi_1}\{X_\xi = y\}E_y^{\pi_1}(\tau_0) \geqslant [a/(1-a)]E_0^{\pi_1}(\xi)$$

and hence $h(\pi_2) \leqslant [(1-a)/a]h(\pi_1)$.

Corollary 6.2: Let π be an environment on $R \setminus \{0\}$ with $d(\pi) = k$. Then

$$g(\pi_G) \leqslant ((1-a)/a)^k g(\bar{\pi}_G),$$
$$g(\pi_B) \leqslant ((1-a)/a)^{k+1} g(\bar{\pi}_G).$$

It is not difficult to give an estimate like $g(\bar{\pi}_G) \leqslant 2$ as $\rho \to 0$. Combining this with Corollary 6.2 gives Lemma 5.1(a).

If a point being changed from G to B is surrounded by good points we can get a better estimate. Again let $R = R_{\rho^{-\alpha}}$ with α and κ as in Sec. V. We restate (2.5) as we will use it: Let

$$\xi_0 = \inf\{j \geqslant 1: X_j \in \{0\} \cup \partial R\},$$

then if there exists a v such that for all $y \in \partial R$,

$$P_0^{\pi_B}\{X_{\xi_0} = y\} = vP_0^{\pi_G}\{X_{\xi_0} = y\}(1 + O(f(\rho))),$$

then

$$g(\pi_B) = (1/v)g(\pi_G)(1 + O(f(\rho))). \tag{6.3}$$

Lemma 6.3: If $\bar{\pi}$ is the all good environment on R, then

$$g(\bar{\pi}_B) = g(\bar{\pi}_G)(1 + O(\rho^{2\alpha}\log(1/\rho))).$$

Proof: By Lemma 3.5

$$P_0^{\bar{\pi}_B}\{X_{\xi_0} = y\} = P_0^{\bar{\pi}_G}\{X_{\xi_0} = y\}(1 + O(\rho^{2\alpha}\log(1/\rho))).$$

Note that the lemma is not sharp enough to give us Lemma 5.1(c).

Lemma 6.4: If π is an environment in $\mathscr{D}$ such that $\pi_x = G$ for $0 \leqslant |x| \leqslant c$, then

$$|g(\pi_B) - g(\pi_G)| \leqslant c_4 g(\pi_G)O(c^{-2}\log c).$$

Proof: Let $\tilde{\xi}_0 = \inf\{j: X_j = \{0\}$ or $|X_j| > c\}$. Then for $c < |y| \leqslant c+1$,

$$\begin{aligned} P_0^{\bar{\pi}_B}\{X_{\xi_0} = y\} &= \sum_{z \neq 0} P_0^{\bar{\pi}_B}\{X_{\tilde{\xi}_0} = z\}P_z^{\bar{\pi}_B}\{X_{\xi_0} = y\} \\ &= \sum_{z \neq 0} P_0^{\bar{\pi}_G}\{X_{\tilde{\xi}_0} = z\} \\ &\quad \times (1 + O(c^{-2}\log c))P_z^{\bar{\pi}_G}\{X_{\xi_0} = y\} \\ &= P_0^{\bar{\pi}_G}\{X_{\xi_0} = y\}(1 + O(c^{-2}\log c)). \end{aligned}$$

Lemma 6.5: Let $x \in R$ satisfy $|x| \geqslant \rho^{-\kappa}$, $\mathrm{dist}(x, \partial R) \geqslant \rho^{-\kappa}$. Then if $\bar{\pi}$ denotes the all good environment and π^x the environment that is bad only at x, then

$$\begin{aligned} g(\pi^x_G) &= g(\bar{\pi}_G)(1 + O(\rho^{2\kappa}\log(1/\rho))) \\ &= g(\bar{\pi}_B)(1 + O(\rho^{2\kappa}\log(1/\rho))). \end{aligned}$$

Proof: The second inequality follows from Lemma 6.4. To prove the first we will prove the following stronger statement: If π_1 is any environment on $\mathscr{C}_m$ $(m > 2n+1)$ with $(\pi_1)_y = G$ for $y \in \bar{R} = R_{(1/2)\rho^{-\kappa}}(x)$ and π_2 is the environment that agrees with π_1 everywhere except $(\pi_2)_x = B$, then

$$E_0^{\pi_1}(\tau_0) = E_0^{\pi_2}(\tau_0)(1 + O(\rho^{2\kappa}\log(1/\rho))).$$

Let $\eta = \tau_0 \wedge \tau_x$. Then it is easy to see that

$$E_0^{\pi_1}(\tau_0) - E_0^{\pi_2}(\tau_0) = P_0^{\pi_1}\{X_\eta = x\}(E_x^{\pi_1}(\tau_0) - E_x^{\pi_2}(\tau_0)).$$

Let

$$\bar\sigma = \inf\{j \geqslant 1:\ X_j \in \partial \bar R\},$$
$$\bar\eta = \inf\{j \geqslant \bar\sigma:\ X_j \in \{0,x\}\}.$$

Then since $0 \notin \bar R$, it is easy to see that for any π

$$E_x^\pi(\tau_0) = E_x^\pi(\bar\eta) + P_x^\pi\{X_{\bar\eta} = x\} E_x^\pi(\tau_0),$$

i.e.,

$$E_x^\pi(\tau_0) = E_x^\pi(\bar\eta)[P_x^\pi\{X_{\bar\eta} = x\}]^{-1}.$$

For any $y \in \partial \bar R$, by Lemma 3.5,

$$P_x^{\pi_1}\{X_{\bar\sigma} = y\} = P_x^{\pi_2}\{X_{\bar\sigma} = y\}(1 + O(\rho^{2\kappa}\log(1/\rho))),$$

which, as before, gives

$$E_x^{\pi_1}(\bar\eta - \bar\sigma) = E_x^{\pi_2}(\bar\eta - \bar\sigma)(1 + O(\rho^{2\kappa}\log(1/\rho))),$$
$$P_x^{\pi_1}\{X_{\bar\eta} = x\} = P_x^{\pi_2}\{X_{\bar\eta} = x\}(1 + O(\rho^{2\kappa}\log(1/\rho))).$$

Symmetry gives $E_x^{\pi_1}(\bar\sigma) = E_x^{\pi_2}(\bar\sigma)$. Hence

$$E_x^{\pi_1}(\tau_0) = E_x^{\pi_2}(\tau_0)(1 + O(\rho^{2\kappa}\log(1/\rho)))$$

and since $E_0^{\pi_1}(\tau_0) \geqslant P_0^{\pi_1}\{X_\eta = x\} E_x^{\pi_1}(\tau_0)$,

$$E_0^{\pi_1}(\tau_0) = E_0^{\pi_2}(\tau_0)(1 + O(\rho^{2k}\log(1/\tau))).$$

To get Lemma 5.1(b) we need only do the above argument twice.

After this relatively easy warmup, we will proceed with the more delicate estimate, Lemma 5.1(c). For this purpose we will have to consider a rectangle larger than $R = R_{\rho^{-\alpha}}$. Let $\beta,\gamma,\theta > 0$ be chosen satisfying

$$\alpha + \beta < \gamma < \tfrac{1}{2} - \beta, \quad 2\gamma - \beta > 0,$$
$$2\beta < \theta < 4\alpha - 1, \quad \alpha + \beta + \tfrac{1}{2} < \tfrac{5}{6},$$

and let $\delta = \frac{1}{2} + \beta$. We write R_α for $R_{\rho^{-\alpha}}$, R_γ for $R_{\rho^{-\gamma}}$, and R_δ for $R_{\rho^{-\delta}}$.

Definition: An environment π on R_δ is sparse if

(1) $\pi_x = G$, $x \in R_\alpha$,

(2) For every $z \in R_\delta$, if $\pi_z = B$, then $\pi_y = G$ for $y \in R_\alpha(z)\setminus\{z\}$,

(3) $|\{z: \pi_z = B\}| \leqslant \rho^{-\theta}$.

The basic idea of the proof is similar to Lemma 6.3. Note if $\bar\pi$ is the "all good" environment on R_δ, then by Lemma 6.3,

$$g(\bar\pi_B) = g(\bar\pi_G)(1 + O(\rho^{2\delta}\log(1/\rho)))$$
$$= g(\bar\pi_G)(1 + O(\rho^{1+\epsilon})).$$

What we wish to show is (1) sparse environments do not differ too much from the homogeneous environment, and (2) most environments, under the Bernoulli measure, are sparse. Let $\mathscr{E}$ denote the set of environments on $R_\delta\setminus\{0\}$ that extend the all good environment $\bar\pi$ on $R_\alpha\setminus\{0\}$. Then by (2.3) we see

$$g(\bar\pi_B) - g(\bar\pi_G) = \sum_{\pi\in\mathscr{E}} (g(\pi_B) - g(\pi_G))P(\pi), \qquad (6.4)$$

where $P = P_\rho$ denotes the conditional Bernoulli probability on environemnts on $R_\delta \setminus R_\alpha$. We start by making estimates on P.

We call an environment sparse on R_γ if it satisfies the conditions for being sparse, except that condition (2) need only hold for $z \in R_\gamma$. Let

$$\mathscr{E}_0 = \{\pi\in\mathscr{E}: \text{ is sparse on } R_\delta\},$$
$$\mathscr{E}_1 = \{\pi\in\mathscr{E}\setminus\mathscr{E}_0: \pi \text{ is sparse on } R_\gamma\},$$
$$\mathscr{E}_k = \{\pi\in\mathscr{E}\setminus\mathscr{E}_0\cup\cdots\cup\mathscr{E}_{k-1}): \text{by changing } k-1 \text{ values of } \pi \text{ from } B \text{ to } G \text{ one can make } \pi \text{ sparse on } R_\gamma\},$$
$$\mathscr{F} = \{\pi\in\mathscr{E}: |\{z: \pi_z = B\}| \leqslant \rho^{-\theta}\}.$$

Finally if $\frac{1}{2} \leqslant r \leqslant 1$, let

$$R_\delta^r = \{z\in R_\delta: |z_i| \leqslant r\rho^{-\delta}\}$$

and similarly for R_γ^r. Let

$$\mathscr{G} = \{\pi\in\mathscr{E}: \exists r\in[\tfrac{1}{2},1] \text{ such that } \forall x \text{ with } \operatorname{dist}(x,\partial R_\delta^r) \leqslant 2\rho^{-\alpha} \text{ or } \operatorname{dist}(x,\partial R_\gamma^r) \leqslant 2\rho^{-\alpha},\ \pi_x = G\}.$$

We now estimate the Bernoulli probabilities of these sets,

$$P(\mathscr{E}_0^c) \leqslant \sum_{\substack{x,y\in R_\delta \\ |x-y|\leqslant 2\rho^{-\alpha}}} P\{\pi_x = B, \pi_y = B\}$$
$$= \rho^2 O(\rho^{-2(\alpha+\delta)}) = O(\rho^{1-2(\alpha+\beta)}).$$

Similarly,

$$P((\mathscr{E}_0\cup\mathscr{E}_1)^{cv}) \leqslant O(\rho^{2(1-\alpha-\gamma)}).$$

An environment in $\mathscr{E}_{k+1}$ is obtained by taking an environment in $\mathscr{E}_k$ by changing one value in R_γ from G to B. For any $\pi\in\mathscr{E}_k$ there are at most $O(\rho^{-2\gamma})$ ways of choosing this value. Hence

$$P(\mathscr{E}_{k+1}) \leqslant O(\rho^{1-2\gamma})P(\mathscr{E}_k).$$

We can estimate $P(\mathscr{F})$ using a standard estimate for large deviations of binomial probabilities that we just state: since $\theta > 2\beta$, there exists an $\epsilon > 0$, such that

$$P(\mathscr{F}^c) \leqslant O(e^{-\rho^{-\epsilon}})$$

and hence is $o(\rho^j)$ for any j. For any r, let

$$\mathscr{H}_r = \{\pi\in\mathscr{E}: \exists x \text{ with } \operatorname{dist}(x,\partial R_\gamma^4) \leqslant 2\rho^{-2\alpha} \text{ or } \operatorname{dist}(x,R_\gamma^r) \leqslant 2\rho^{-2\alpha} \text{ with } \pi_x = B\}.$$

Then a simple counting argument gives $P(\mathscr{H}_r) \leqslant O(\rho^{1-\alpha-\delta}) \leqslant O(\rho^{1/6})$. For ρ sufficiently small the sets $\mathscr{H}_r$, $r\in\{\frac{1}{2} + k/15: k = 0,\ldots,15\}$ are disjoint. Hence for these r,

$$P\Big(\bigcap_r \mathscr{H}_r\Big) \leqslant O(\rho^{8/3}).$$

Hence $P(\mathscr{G}^c) \leqslant O(\rho^{8/3})$.

It follows from (2.2) and Hölder's inequality that for any $\tilde{\mathscr{E}} \subset \mathscr{E}$,

$$\sum_{\pi\in\tilde{\mathscr{E}}} g(\pi)P(\pi) \leqslant (c_2 P(\tilde{\mathscr{E}}))^{1/2}.$$

In particular,

$$\sum_{\pi\in\mathscr{F}^c\cup\mathscr{G}^c} g(\pi)P(\pi)\leqslant O(\rho^{4/3}). \tag{6.5}$$

Below we will prove the following estimates for $\pi\in(\mathscr{E}_0\cup\mathscr{E}_1)\cap\mathscr{F}\cap\mathscr{G}$.

Lemma 6.6: (a) If $\pi\in\mathscr{E}_0\cap\mathscr{F}\cap\mathscr{G}$,

$$g(\pi_B)=g(\pi_G)(1+O(\rho^{1+\epsilon})).$$

(b) If $\pi\in\mathscr{E}_1\cap\mathscr{F}\cap\mathscr{G}$,

$$g(\pi_B)=g(\pi_G)(1+O(\rho^{2(\alpha+\beta)\epsilon})).$$

From Lemma 6.6(b) and Lemma 6.1, we get for $\pi\in\mathscr{E}_k\cap\mathscr{F}\cap\mathscr{G}$,

$$g(\pi_B)-g(\pi_G)\leqslant c_2^{k-1}.$$

From (6.4), (6.5), and Lemma 6.6 we get

$$\begin{aligned}
&|g(\bar\pi_B)-g(\bar\pi_G)|\\
&\leqslant\Bigg[\sum_{k=0}^{\infty}\sup_{\pi\in\mathscr{E}_k\cap\mathscr{F}\cap\mathscr{G}}|g(\pi_B)\\
&\quad-g(\pi_G)|P(\mathscr{E}_k)\Bigg]+O(\rho^{4/3})\\
&\leqslant O(\rho^{1+\epsilon})+O(\rho^{2(\alpha+\beta)+\epsilon})O(\rho^{1-2(\alpha+\beta)})\\
&\quad+\sum_{k=2}^{\infty}c_2^kO(\rho^{1-2(\alpha+\gamma)+(k-1)(1-2\gamma)})+O(\rho^{4/3})\\
&=O(\rho^{1+\epsilon}).
\end{aligned}$$

In order to prove Lemma 6.6(a), by (2.5) it is sufficient to show that for every $\pi\in\mathscr{E}_0\cap\mathscr{F}\cap\mathscr{G}$, there exists an $r\in[\frac{1}{2},1]$ such that if

$$\xi_0=\inf\{j\geqslant 1:\ X_j\in\{0\}\cup\partial R_\delta^r\},$$

then for $y\in\partial R_\delta^r$,

$$P_0^{\pi_B}\{X_{\xi_0}=y\}=P_0^{\pi_G}\{X_{\xi_0}=y\}(1+O(\rho^{1+\epsilon})). \tag{6.6}$$

The r we will choose is any r such that $\pi_x=G$ for all x satisfying $\operatorname{dist}(x,R_\delta^r)\leqslant 2\rho^{-\alpha}$. In order to simplify the notation we assume that $r=1$ satisfies this condition (the argument is the same for the other r and the estimates are clearly uniform in r). We will prove (6.6) by induction on the number of bad points in π—essentially we will add one bad point at a time starting with the homogeneous environment $\bar\pi$. We know that every $\pi\in\mathscr{F}\cap\mathscr{E}_0$ can be obtained from $\bar\pi$ by addition at most $\rho^{-\theta}$ bad points and then if we add them one at a time at each stage the environment obtained is sparse. In the remainder of this section we use π sparse to mean π sparse *and* $\pi_x=G$ for $\operatorname{dist}(x,\partial R_\delta)\leqslant 2\rho^{-\alpha}$.

Lemma 6.7: Let π be a sparse environment; $z\in R_\delta$ with $\pi_w=G$ for $w\in R_{\rho^{-\alpha}}(z)$; and $y\in\partial R_\delta$. Then

(a) $P_z^{\pi_G}\{X_{\xi_0}=y\}=P_z^{\bar\pi_G}\{X_{\xi_0}=y\}(1+O(\rho^\epsilon))$,

(b) $P_0^{\pi_G}\{X_{\xi_z}=z\}=P_0^{\bar\pi_G}\{X_{\xi_z}=z\}(1+O(\rho^\epsilon))$.

Proof: We prove (a); the proof for (b) is exactly the same (replace z with 0 and y and z). Assume $\pi_x=B$ and let $\tilde\pi$ be the environment that agrees with π everywhere except $\tilde\pi_x=G$. We first note that Harnack's inequality (Lemma 3.1) implies that

$$\sup_{w\in R_{1/2\rho^{-\alpha}}(x)}P_w^{\tilde\pi_G}\{X_{\xi_0}=y\}\leqslant c_2P_x^{\tilde\pi_G}\{X_{\xi_0}=y\}.$$

Now let $\eta=\inf\{j\geqslant 1:X_j\in\{0,x\}\cup\partial R_\delta\}$. Then

$$\begin{aligned}
&|P_z^{\pi_G}\{X_{\xi_0}=y\}-P_z^{\tilde\pi_G}\{X_{\xi_0}=y\}|\\
&\quad=P_z^{\tilde\pi_G}\{X_\eta=x\}|P_x^{\pi_G}\{X_{\xi_0}=y\}-P_x^{\tilde\pi_G}\{X_{\xi_0}=y\}|.
\end{aligned}$$

By an argument similar to Lemma 6.5, using Lemma 3.1 as above,

$$\begin{aligned}
&|P_x^{\pi_G}\{X_{\xi_0}=y\}-P_x^{\tilde\pi_G}\{X_{\xi_0}=y\}|\\
&\quad\leqslant O(\rho^{2\alpha}\log(1/\rho))P_x^{\tilde\pi_G}\{X_{\xi_0}=y\}.
\end{aligned}$$

Since

$$P_z^{\tilde\pi_G}\{X_{\xi_0}=y\}\geqslant P_z^{\tilde\pi_G}\{X_\eta=x\}P_x^{\tilde\pi_G}\{X_{\xi_0}=y\},$$

this implies

$$P_z^{\pi_G}\{X_{\xi_0}=y\}=P_z^{\tilde\pi_G}\{X_{\xi_0}=y\}(1+O(\rho^{2\alpha}\log(1/\rho))).$$

Since $d(\pi)\leqslant\rho^{-\theta}$, this process can be done at most $\rho^{-\theta}$ times to get

$$\begin{aligned}
P_z^{\pi_G}\{X_{\xi_0}=y\}&=P_z^{\bar\pi_G}\{X_{\xi_0}=y\}(1+O(\rho^{2\alpha}\log(1/\rho)))\rho^{-\theta}\\
&=P_z^{\bar\pi_G}\{X_{\xi_0}=y\}(1+O(\rho^\epsilon)).
\end{aligned}$$

Proof of Lemma 6.6: We prove (a); (b) is similar. Let π and $\tilde\pi$ be as in the proof of the above lemma and let $\pi_B,\pi_G,\tilde\pi_B,\tilde\pi_G$ be the four possible environments, obtained by varying values at 0 and x. Let $\eta=\inf\{j\geqslant 1:\ X_j\in\partial R_\delta\cup\{0,x\}\}$. Then for any environment $\hat\pi$ on R_δ

$$P_0^{\hat\pi}\{X_{\xi_0}=y\}=P_0^{\hat\pi}\{X_\eta=y\}+P_0^{\hat\pi}\{X_\eta=x\}P_x^{\hat\pi}\{X_{\xi_0}=y\}.$$

From this we get

$$\begin{aligned}
&[P_0^{\pi_B}\{X_{\xi_0}=y\}-P_0^{\pi_G}\{X_{\xi_0}=y\}]\\
&\quad-[P_0^{\tilde\pi_B}\{X_{\xi_0}=y\}-P_0^{\tilde\pi_G}\{X_{\xi_0}=y\}]\\
&\quad=[P_0^{\tilde\pi_B}\{X_\eta=x\}-P_0^{\tilde\pi_G}\{X_\eta=x\}]\\
&\qquad\times[P_x^{\pi_G}\{X_{\xi_0}=y\}-P_x^{\tilde\pi_G}\{X_{\xi_0}=y\}].
\end{aligned}$$

By Lemmas 6.3 and 6.7

$$\begin{aligned}
&|P_0^{\tilde\pi_B}\{X_\eta=x\}-P_0^{\tilde\pi_G}\{X_\eta=x\}|\\
&\quad\leqslant P_0^{\tilde\pi_G}\{X_\eta=x\}O(\rho^{2\alpha}\log(1/\rho))\\
&\quad\leqslant P_0^{\bar\pi_G}\{X_{\xi_x}=x\}O(\rho^{2\alpha}\log(1/\rho)).
\end{aligned}$$

Similarly,

$$\begin{aligned}
&|P_x^{\pi_G}\{X_{\xi_0}=y\}-P_x^{\tilde\pi_G}\{X_{\xi_0}=y\}|\\
&\quad\leqslant P_x^{\bar\pi_G}\{X_{\xi_0}=y\}O(\rho^{2\alpha}\log(1/\rho)).
\end{aligned}$$

But

$$\begin{aligned}
&P_0^{\bar\pi_G}\{X_{\xi_x}=x\}P_x^{\bar\pi_G}\{X_{\xi_0}=y\}\\
&\quad\leqslant P_0^{\bar\pi_G}\{X_\sigma=y\}\leqslant c_4(\log(1/\rho))P_0^{\bar\pi_G}\{X_{\xi_0}=y\}
\end{aligned}$$

(using Lemma 3.2). We therefore get

$$\begin{aligned}
&|[P_0^{\pi_B}\{X_{\xi_0}=y\}-P_0^{\pi_G}\{X_{\xi_0}=y\}]\\
&\quad-P_0^{\tilde\pi_B}\{X_{\xi_0}=y\}-P_0^{\tilde\pi_G}\{X_{\xi_0}=y\}|\\
&\quad\leqslant P_0^{\bar\pi_G}\{X_{\xi_0}=y\}O(\rho^{4\alpha}(\log(1/\rho))^3).
\end{aligned}$$

By performing this operation one point at a time at most $\rho^{-\theta}$ times we get

$$|(P_0^{\pi_B}\{X_{\xi_0}=y\}-P_0^{\pi_G}\{X_{\xi_0}=y\}) - (P_0^{\bar\pi_B}\{X_{\xi_0}=y\}-P_0^{\bar\pi_G}\{X_{\xi_0}=y\})| \leqslant P_0^{\bar\pi_G}\{X_{\xi_0}=y\}O(\rho^{1+\epsilon}).$$

By Lemma 3.5,

$$|P_0^{\bar\pi_B}\{X_{\xi_0}=y\}-P_0^{\bar\pi_G}\{X_{\xi_0}=y\}| = P_0^{\bar\pi_G}\{X_{\xi_0}=y\}O(\rho^{1+\epsilon}).$$

Hence by Lemma 6.7

$$|P_0^{\pi_B}\{X_{\xi_0}=y\}-P_0^{\pi_G}\{X_{\xi_0}=y\}|\leqslant P_0^{\bar\pi_G}\{X_{\xi_0}=y\}O(\rho^{1+\epsilon}) = P_0^{\pi_G}\{X_{\xi_0}=y\}O(\rho^{1+\epsilon})$$

and

$$P_0^{\pi_B}\{X_{\xi_0}=y\}=P_0^{\pi_G}\{X_{\xi_0}=y\}(1+O(\rho^{1+\epsilon})),$$

which gives (6.6) and hence proves Lemma 6.6(a).

The remainder of this section will be used to prove Lemma 5.1(d),(e). We will need a lemma on the effect of extending an environment.

Lemma 6.8: Let $A\subset R_{2\rho^{-\alpha}}$ and $\pi_1: A\to\{B,G\}$, $\pi_2: R_{2\rho^{-\alpha}}\to\{B,G\}$ with $\pi_1\subset\pi_2$ and $(\pi_2)_x=G$ for $x\notin A$. Then if $|\{x: (\pi_1)_x=B\}|\leqslant 2$,

$$g(\pi_1)=g(\pi_2)(1+O(\rho^{2\alpha+\epsilon})).$$

Proof: Let κ as before satisfy $4\alpha-1<\kappa<1-3\alpha$. By (2.3),

$$g(\pi_1)=\sum_{\bar\pi\supset\pi_1} g(\bar\pi)P_\rho(\bar\pi).$$

Let $d(\bar\pi)=|\{x\in R_{2\rho^{-\alpha}}\setminus A: \bar\pi_x=B\}|$. Then routine estimates as before give

$$P_\rho\{\bar\pi: d(\bar\pi)=k\}\leqslant O(\rho^{k(1-2\alpha)}).$$

Let $\mathscr{S}$ be the set of $\bar\pi$ such that $d(\bar\pi)=1$, and if $x\in R_{2\rho^{-\alpha}}\setminus A$ is the point with $\pi_x=B$ then $\bar\pi_y=G$ for all $y\in R_{\rho^{-\kappa}}(x)\setminus\{x\}$. Then another estimate gives

$$P_\rho\{\mathscr{S}\}\leqslant O(\rho^{1-2\alpha}),$$

$$P_\rho(\{\bar\pi: d(\bar\pi)=1\}\cap\mathscr{S}^c)\leqslant O(\rho^{1-\alpha-\kappa})=O(\rho^{2\alpha}+\epsilon).$$

For $\bar\pi\in\mathscr{S}$, estimates as in Lemma 5.1(b) give

$$g(\bar\pi)=g(\pi_2)(1+O(\rho^{2\kappa}\log(1/\rho))).$$

For other $\bar\pi$, Lemma 5.1(a) gives $g(\bar\pi)\leqslant c_2^{d(\bar\pi)}$. Hence

$$g(\pi_1)=g(\pi_2)+\sum_{\bar\pi\supset\pi_1}(g(\bar\pi)-g(\pi_2))P_\rho(\bar\pi) = g(\pi_2)+g(\pi_2)O(\rho^{2\alpha+\epsilon}).$$

Let X_j be an irreducible Markov chain on a finite space S with invariant probability φ. Let $A\subset S$, $x\in S$,

$$\tau_x=\inf\{j\geqslant 1,\ X_j=x\},$$
$$V_x=|\{1\leqslant j\leqslant\tau_x: X_j\in A\}|.$$

Then it is standard that

$$\varphi(x)E_x(V_x)=\sum_{y\in A}\varphi(y).$$

Let $\bar\pi: R_\alpha\setminus\{0\}\to\{B,G\}$ be the all good environment and let $A=R_\alpha\setminus R_\nu$, where $0\leqslant\nu\leqslant\alpha$. Let $\hat\pi$ be an environment on $\mathscr{C}_m$ $(m\geqslant 2\rho^{-\alpha}+1)$ that extends $\bar\pi$, let $\xi_0=\inf\{j\geqslant 1: X_j\in\{0\}\cup\partial R_\alpha\}$ and

$$V_1=|\{1\leqslant j\leqslant\xi_0: X_j\in A\}|,$$
$$V_2=|\{\xi_0<j\leqslant\tau_x: X_j\in A\}|.$$

Then $V_0=V_1+V_2$. By symmetry

$$E_0^{\hat\pi_G}(V_1)=E_0^{\hat\pi_B}(V_1).$$

By an argument as in Lemma 6.3,

$$E_0^{\hat\pi_B}(V_2)=E_0^{\hat\pi_G}(V_2)(1+O(\rho^{2\alpha}\log(1/\rho))),$$
$$\varphi^{\hat\pi_B}(0)=\varphi_0^{\hat\pi_G}(0)(1+O(\rho^{2\alpha}\log(1/\rho))).$$

Hence

$$\sum_{y\in A}\varphi^{\hat\pi_B}(y)=\left(\sum_{y\in A}\varphi^{\hat\pi_G}(y)\right)\left(1+O\left(\rho^{2\alpha}\log\frac{1}{\rho}\right)\right).$$

Similarly, if $\theta>0$ is given, $A=R_{(\alpha+\theta)}\setminus R_\nu$, and $\bar\pi_x=G$ for all $x\in R_{(\alpha+\theta)}\setminus\{0\}$, then

$$\sum_{y\in A}\varphi^{\hat\pi_B}(y)=\left(1+O\left(\rho^{2(\alpha+\theta)}\log\frac{1}{\rho}\right)\right)\sum_{y\in A}\varphi^{\hat\pi_G}(y) = (1+O(\rho^{2\alpha+\epsilon}))\sum_{y\in A}\varphi^{\hat\pi_G}(y). \tag{6.7}$$

For any $x\in R_\alpha$, let $R_\alpha^x=\{y+x: y\in R_\alpha\}$ and let $\bar\pi^x: R_\alpha^x\to\{B,G\}$ be the environment $(\bar\pi^x)_x=B$, $(\bar\pi^x)_y=G$ for $y\neq x$. Then by Lemma 6.8 (applied twice),

$$g(\bar\pi^x)=g(\pi_G^x)(1+O(\rho^{2\alpha+\epsilon}))$$

and hence

$$\sum_{x\in A}g(\pi_G^x)=(1+O(\rho^{2\alpha+\epsilon}))\sum_{x\in A}g(\bar\pi^x). \tag{6.8}$$

By the definition of g,

$$\sum_{x\in A}g(\bar\pi^x)=\lim_{m\to\infty}\left(\sum_{x\in A}\sum_{\substack{\bar\pi^x\subset\hat\pi\\ \hat\pi\in\mathscr{C}_m}}P_\rho(\hat\pi)m^2\varphi^{\hat\pi}(0)\right) = \lim_{m\to\infty}\left(\sum_{\substack{\bar\pi_B\subset\hat\pi\\ \hat\pi\in\mathscr{C}_m}}P_\rho(\hat\pi)\sum_{x\in A}m^2\varphi^{\hat\pi}(x)\right).$$

We now assume $0<\theta<\frac{1}{2}-\alpha$ and let $\mathscr{U}=\{\hat\pi\supset\pi_B: (\hat\pi)_x=G, x\in R_{(\alpha+\theta)}\setminus\{0\}\}$ and $\bar{\mathscr{U}}=\{\hat\pi\supset\pi_B: \hat\pi\notin\mathscr{U}\}$. By standard estimates

$$P_\rho(\bar{\mathscr{U}})\leqslant O(\rho^{1-2\alpha-2\theta})=O(\rho^\epsilon).$$

Then, using (6.7),

$$\sum_{\hat\pi\subset\pi_B} P_\rho(\hat\pi)\sum_{x\in A}\varphi^{\hat\pi}(x)$$
$$=\sum_{\hat\pi\in\mathscr{U}} P_\rho(\hat\pi)\sum_{x\in A}\varphi^{\hat\pi}(x)+\sum_{\hat\pi\in\mathscr{U}} P_\rho(\hat\pi)\sum_{x\in A}\varphi^{\hat\pi}(x)$$
$$=\sum_{\hat\pi\in\mathscr{U}} P_\rho(\hat\pi)\sum_{x\in A}\varphi^{\hat\pi_G}(x)(1+O(\rho^{2\alpha+\epsilon}))$$
$$+\sum_{\hat\pi\in\mathscr{U}} P_\rho(\hat\pi)\sum_{x\in A}\varphi^{\hat\pi_G}(x)(1+O(\rho^{2\alpha}\log(1/\rho)))$$
$$\times(1+O(\rho^{2\alpha+\epsilon}))\sum_{\hat\pi\in\mathscr{U}} P_\rho(\hat\pi)\sum_{x\in A}\varphi^{\hat\pi_G}(x).$$

(Here we have written $\hat\pi_G$ for the environment obtained by changing $\hat\pi$ at the origin from B to G.) We see then that

$$\sum_{x\in A} g(\bar\pi^x)=(1+O(\rho^{2\alpha+\epsilon}))\sum_{x\in A} g(\bar\pi_x),$$

where $\bar\pi_x$ denotes the all good environment on R_α^x. By Lemma 6.8, $g(\bar\pi_x)=g(\bar\pi_G)(1+O(\rho^{2\alpha+\epsilon}))$ and therefore (6.8) gives

$$\sum_{x\in A} g(\pi_G^x)=(1+O(\rho^{2\alpha+\epsilon}))\sum_{x\in A} g(\bar\pi_G).$$

But

$$\sum_{x\in A} g(\bar\pi_G)=O(\rho^{-2\alpha}).$$

Hence

$$\sum_{x\in A}(g(\pi_G^x)-g(\bar\pi_G))=O(\rho^\epsilon). \tag{6.9}$$

If we set $A=R_\alpha$, then (6.9) gives Lemma 5.1(d).

To obtain Lemma 5.1(e), we let $\frac{1}{2}\alpha<\nu<\alpha/(2\alpha+1)$ and set $A=R_\alpha\setminus R_\nu$. Then again by (6.9),

$$\sum_{x\in A}(g(\pi_G^x)-g(\bar\pi_G))=O(\rho^\epsilon).$$

For $x\in R_\alpha\setminus R_\nu$ an argument like that used in the proof of Lemma 6.6 can be used to show

$$g(\pi_B^x)=g(\pi_G^x)(1+O(\rho^{4\nu}(\log(1/\rho))^3))$$
$$=g(\pi_G^x)(1+O(\rho^{2\alpha+\epsilon})).$$

Hence

$$\sum_{x\in A}(g(\pi_B^x)-g(\bar\pi_G))=O(\rho^\epsilon). \tag{6.10}$$

For $x\in R_\nu$ we will prove the following:

$$g(\pi_B^x)=g(\bar\pi_G)(1+\psi_x)(1+O(\rho^{1-\nu/\alpha}\log(1/\rho))), \tag{6.11}$$

where ψ_x is obtained from (4.8) and (4.12) by

$$\psi_x=\lim_{m\to\infty}(h^\pi(0)-1)=[1+(a_B-\tfrac{1}{2})\Phi(y)]^{-1}.$$

From this it easily follows [using Lemma 5.2 and the fact that $\nu<\alpha/(2\alpha+1)$] that

$$\sum_{x\in R_\nu}(g(\pi_B^x)-g(\bar\pi_G))=\left(\sum_{x\in R_\nu}\psi_x\right)(1+O(\rho^\epsilon)). \tag{6.12}$$

The inequality $\psi_x+\psi_{\bar x}\leqslant O(|x|^{-4}(\log|x|)^2)$ implies that

$$\sum_{x\in R_\nu}\psi_x=B_2(1+O(\rho^\epsilon)),$$

which will give Lemma 5.1(e). Hence all we have left is to prove (6.11).

We now compare $g(\bar\pi_G)$ and $g(\pi_B^x)$ for $x\in R_\nu$. First, let $\hat\pi$ be any environment on $R=R_\alpha$ and $\pi\in\mathscr{C}_m$, m very large, an environment with $\pi\supset\hat\pi$. Let $\eta=\inf\{j\geqslant 1: X_j\in\{0,x\}\}$. Then two applications of the Markov property, as done many times in this paper, give

$$E_0^\pi(\tau_0)=E_0^\pi(\eta)+P_0^\pi\{X_\eta=x\}$$
$$\times[P_x^\pi\{X_\eta=0\}]^{-1}E_x^\pi(\eta).$$

Reasoning as at the end of Sec. II, we get

$$E_0^\pi(\eta)\sim\sum_{y\in\partial R}P_0^{\hat\pi}\{X_\xi=y\}E_y^\pi(\eta),$$
$$E_x^\pi(\eta)\sim\sum_{y\in\partial R}P_x^{\hat\pi}\{X_\xi=y\}E_y^\pi(\eta),$$

where $\xi=\inf\{j\geqslant 1: X_j\in\{0,x\}\cup\partial R\}$. Therefore for all $\pi\supset\hat\pi$,

$$E_0^\pi(\tau_0)\sim\sum_{y\in\partial R}P_0^{\hat\pi}\{X_\xi=y\}E_y^\pi(\eta)$$
$$+P_0^\pi\{X_\eta=x\}[P_x^\pi\{X_\eta=0\}]^{-1}$$
$$\times\sum_{y\in\partial R}P_x^{\hat\pi}\{X_\xi=y\}E_y^\pi(\eta). \tag{6.13}$$

By (3.4) we have for $y\in\partial R$, $x\in R_\nu$,

$$P_x^{\pi_B^x}\{X_\xi=y|X_\xi\notin\{0,x\}\}$$
$$=P_x^{\bar\pi_G}\{X_\xi=y|X_\xi\notin\{0,x\}\}(1+O(\rho^{1-\nu/\alpha}\log(1/\rho)))$$
$$=P_x^{\pi_B^x}\{X_\xi=y|X_\xi\notin\{0,x\}\}(1+O(\rho^{1-\nu/\alpha}\log(1/\rho)))$$
$$=P_0^{\bar\pi_B}\{X_\xi=y|X_\xi\notin\{0,x\}\}(1+O(\rho^{1-\nu/\alpha}\log(1/\rho)))$$
$$=H(0,y)(1+O(\rho^{1-\nu/\alpha}\log(1/\rho))).$$

By symmetry conditions and (3.6) we get

$$P_0^{\pi_G}\{X_\xi=x\}=P_x^{\pi_G}\{X_\xi=0\}(1+O(\rho^{1-\nu/\alpha})),$$
$$P_0^{\pi_B^x}\{X_\xi=x\}=P_x^{\pi_B^x}\{X_\xi=0\}(1+O(\rho^{1-\nu/\alpha})).$$

For any $\pi\in\mathscr{C}_m$,

$$P_0^\pi\{X_\eta=x\}=P_0^\pi\{X_\xi=x\}$$
$$+\sum_{y\in\partial R}P_0^\pi\{X_\xi=y\}P_y^\pi\{X_\eta=x\},$$
$$P_x^\pi\{X_\eta=0\}=P_x^\pi\{X_\xi=0\}$$
$$+\sum_{y\in\partial R}P_x^\pi\{X_\xi=y\}P_y^\pi\{X_\eta=0\}.$$

For any $\pi\in\mathscr{C}_m$ which extends $\bar\pi_G$ or π_B^x and any $y\in\partial R$, $P_y^\pi\{X_\eta=0\}$ is bounded above (below) by the supremum (infimum) of $P_z^\pi\{X_\xi=0\}$ over $z\in\partial R$. But by reversibility of simple random walk and (3.6),

$$P_z^{\pi}\{X_{\xi}=0\}=P_0^{\bar{\pi}_G}\{X_{\xi}=z\}$$
$$=P_x^{\bar{\pi}_G}\{X_{\xi}=z\}(1+O(\rho^{1-\nu/\alpha}\log(1/\rho))),$$
$$P_z^{\pi}\{X_{\xi}=x\}(1+O(\rho^{1-\nu/\alpha}\log(1/\rho))).$$

Hence

$$P_y^{\pi}\{X_\eta=0\}=P_y^{\pi}\{X_\eta=x\}(1+O(\rho^{1-\nu/\alpha}\log(1/\rho)))$$

and since their sum is 1,

$$P_y^{\pi}\{X_\eta=0\}=\tfrac{1}{2}+O(\rho^{1-\nu/\alpha}\log(1/\rho)).$$

Hence by Lemma 3.8 we can conclude for all such π

$$P_0^{\pi}\{X_\eta=x\}=P_x^{\pi}\{X_\eta=0\}(1+O(\rho^{1-\nu/\alpha}\log(1/\rho))).$$

We now let $\pi\supset\pi_G$ and let π^1 equal π everywhere except that $(\pi^1)_0=(\pi^1)_x=B$ (i.e., $\pi^1\supset\pi_B^x$). Then by (6.13),

$$E_0^{\pi}(\tau_0)\left[P_0^{\pi_B^x}\{X_{\xi}\in\partial R\}+P_x^{\pi_B^x}\{X_{\xi}\in\partial R\}\right]$$
$$\sim E_0^{\pi'}(\tau_0)\left[P_0^{\bar{\pi}_G}\{X_{\xi}\in\partial R\}+P_x^{\bar{\pi}_G}\{X_{\xi}\in\partial R\}\right]$$
$$\times(1+O(\rho^{1-\nu/\alpha}\log(1/\rho))).$$

(Here the error in the asymptotics can be made as small as we want for fixed ρ so we can change the $\sim$ to $=$.) Note also

$$P_0^{\pi_B^x}\{X_{\xi}\in\partial R\}=P_x^{\pi_B^x}\{X_{\xi}\in\partial R\}(1+O(\rho^{1-\nu/\alpha}\log(1/\rho))),$$
$$P_0^{\pi_G^x}\{X_{\xi}\in\partial R\}=P_x^{\pi_G^x}\{X_{\xi}\in\partial R\}(1+O(\rho^{1-\nu/\alpha}\log(1/\rho))).$$

Therefore

$$g(\pi_B^x)=g(\bar{\pi}_G)\frac{P_0^{\pi_B^x}\{X_{\xi}\in\partial R\}}{P_0^{\bar{\pi}_G}\{X_{\xi}\in\partial R\}}$$
$$\times(1+O(\rho^{1-\nu/\alpha}\log(1/\rho))).$$

We only need to compute the constant. Actually, we have already done it: if we do the same analysis as above with $\pi\supset\bar{\pi}_G$ being the all good environment on $\mathscr{C}_m$ (instead of averaging over all $\pi\supset\bar{\pi}_G$), we repeat the periodic calculations of Sec. IV. Hence we conclude

$$\frac{P_0^{\pi_B^x}\{X_{\xi}\in\partial R\}}{P_0^{\bar{\pi}_G}\{X_{\xi}\in\partial R\}}=(1+\psi_x)(1+O(\rho^{1-\nu/\alpha}\log(1/\rho)))$$

and

$$g(\pi_B^x)=g(\bar{\pi}_G)(1+\psi_x)(1+O(\rho^{1-\nu/\alpha}\log(1/\rho))),$$

which is (6.11).

VII. REMARKS

What we have shown is that the first term in a low-density expansion comes from "one-point" interactions and the second term from "two-point" interactions. For a random periodic environment, the k th-order term comes from considering environments with k or fewer bad points. We expect that the k th-order term for the random case will also come only from k (or fewer) point interactions. There are many places in this paper where estimates are far weaker than what is expected. More detailed analysis using the same basic ideas should give much finer results.

It is quite possible, although it is unclear how to prove, that $\bar{a}(a_G,a_B,\rho)$ is an analytic function in the region $c\leqslant a_G$, $a_B\leqslant 1-c$, $0\leqslant\rho\leqslant 1$, for any choice of $c\in(0,\frac{1}{2})$.

ACKNOWLEDGMENT

This research was supported by NSF Grant No. DMS-8502293 and an Alfred P. Sloan Research Fellowship.

[1]G. Lawler, Commun. Math. Phys. **87**, 81 (1982).

[2]J. B. T. M. Roerdink and K. E. Shuler, J. Stat. Phys. **41**, 581 (1985).

[3]F. Spitzer, *Principles of Random Walk* (Springer, New York, 1976), 2nd ed.

[4]A. Brandt, Math. Comp. **20**, 473 (1966).

Well-posedness of stationary and time-dependent Spencer–Lewis equations modeling electron slowing down

C. V. M. van der Mee[a)]
Dipartimento di Matematica, Università di Pavia, Strada Nuova 65, I-27100 Pavia, Italy

(Received 9 September 1987; accepted for publication 7 September 1988)

The unique solvability of the time-dependent and stationary Spencer–Lewis equations is established under natural assumptions on the solution and the data of the problem. The strategy used is the method of characteristics followed by perturbation and monotone approximation arguments. The evolution operator in the time-dependent Spencer–Lewis equation is proved to generate a strongly continuous contraction semigroup.

I. INTRODUCTION

In this article we prove that the Spencer–Lewis equation, originally derived in the 1950's by Spencer[1] and Lewis[2] to describe the continuous slowing down of electrons of intermediate energy in a semiconductor or metallic slab medium when the distribution function of the electrons at the upper end of the considered energy range is known, is uniquely solvable. In its present form the equation was formulated by Bartine *et al.*[3] (see also Arkuszewski *et al.*[4]), who replaced the original term $\beta\, \partial u/\partial E$ by the mathematically more convenient term $\partial(\beta u)/\partial E$. Here $\beta = \beta(x,E)$ represents the stopping power. Both the time-dependent and the stationary problem will be considered under natural initial and boundary conditions (see below).

The solution $u = u(x,\mu,E,t)$ of the Spencer–Lewis equation describes the electron distribution as a function of position $x\in[0,a]$, direction cosine of propagation $\mu\in[-1,1]$, energy $E\in[E_m,E_M] \subseteq (0,\infty)$ and, when the problem is time dependent, time $t\in[0,\infty)$. The equation takes account of the fact that incoming electrons may undergo elastic scattering of electrons by atomic nuclei, inelastic scattering by atomic electrons, and bremsstrahlung producing collisions with atomic nuclei and atomic electrons. Inelastic scattering between an incident electron and an atomic electron may cause ionization and thus add to the free electron population. However, the relatively small contribution to the electron distribution by the electrons stemming from the interaction of photons with matter, through the photoelectric effect, Compton scattering, and pair production, is neglected when deriving the Spencer–Lewis equation. The contribution of the so-called "soft" electron–electron and electron–atomic nuclei collisions leading to an energy transfer of the order of or less than the binding energy of the target electrons is described as a continuous slowing down so that the energy loss per unit distance due to such collisions rather than their cross sections appears in the equation.

In the time-dependent case the equation has the form

$$\frac{\partial u}{\partial t}(x,\mu,E,t) + \mu\frac{\partial u}{\partial x}(x,\mu,E,t) - \frac{\partial(\beta u)}{\partial E}(x,\mu,E,t) + \sigma(x,\mu,E)u(x,\mu,E,t)$$
$$= \int_{-1}^{1}\sigma_s(x,\mu,\mu',E)u(x,\mu',E,t)\,d\mu' + f(x,\mu,E,t), \tag{1}$$

where $\sigma = \sigma(x,\mu,E)$ is the total scattering cross section that usually does not depend on μ, $\sigma_s = \sigma_s(x,\mu,\mu',E)$ the (azimuthally integrated) scattering cross section, and $f(x,\mu,E,t)$ the distribution function for the internal electron sources of intermediate energy. Equation (1) is endowed with the boundary conditions

$$u(x=0,\mu,E,t) = g_0(\mu,E,t), \quad \mu>0, \quad E\in[E_m,E_M], \tag{2a}$$

$$u(x=a,\mu,E,t) = g_a(\mu,E,t), \quad \mu<0, \quad E\in[E_m,E_M], \tag{2b}$$

specifying the distribution of the incident electrons of intermediate energy, the boundary condition

$$u(x,\mu,E=E_M,t) = g_i(x,\mu,t), \quad x\in[0,a], \quad \mu\in[-1,1], \tag{2c}$$

specifying the distribution of the electrons incident at the higher end of the energy range, and the initial condition

$$u(x,\mu,E,t=0) = h_0(x,\mu,E), \quad x\in[0,a], \quad \mu\in[-1,1], \quad E\in[E_m,E_M]. \tag{3}$$

In the stationary case we have the boundary-value problem

$$\mu\frac{\partial u}{\partial x}(x,\mu,E) - \frac{\partial(\beta u)}{\partial E}(x,\mu,E) + \sigma(x,\mu,E)u(x,\mu,E) = \int_{-1}^{1}\sigma_s(x,\mu,\mu',E)u(x,\mu',E)\,d\mu' + f(x,\mu,E), \tag{4}$$

$$u(0,\mu,E) = g_0(\mu,E), \quad \mu>0, \quad E\in[E_m,E_M], \tag{5a}$$

$$u(a,\mu,E) = g_a(\mu,E), \quad \mu<0, \quad E\in[E_m,E_M], \tag{5b}$$

$$u(x,\mu,E_M) = g_i(x,\mu), \quad x\in[0,a] \quad \mu\in[-1,1]. \tag{5c}$$

Contrary to the situation of neutron transport theory, the integral term describing the gain of electrons due to collisions with the host medium does not involve an integration over energy but only over the direction cosine of propagation. Another difference with neutron transport theory is the presence of a term in the equation involving partial differentiation with respect to energy. Natural assumptions on the model are to require u, f, g_0, g_a, g_i, and h_0 as well as β, σ, and σ_s to be non-negative Borel functions and to adopt the hypothesis

[a)] Permanent address: Department of Mathematical Sciences, University of Delaware, Newark, Delaware 19716.

 0022-2488/89/010158-08$02.50

$$\sigma(x,\mu',E) \geqslant \int_{-1}^{1} \sigma_s(x,\mu,\mu',E)\,d\mu,$$
$$(x,\mu',E)\in[0,a]\times[-1,1]\times[E_m,E_M], \tag{6}$$

where the equality sign holds true if (and only if) there is no electron absorption at intermediate energies.

The natural functional setting for the above two problems is suggested by the fact that $u = u(x,\mu,E,t)$ is the electron distribution function for given incident electron fluxes $|\mu| g_0(\mu,E,t)$ and $|\mu| g_a(\mu,E,t)$. This means in particular that one should analyze the above problems in the Banach space $\mathscr{N} = L_1(\Lambda)$, where $\Lambda = (0,a)\times(-1,1)\times(E_m,E_M)$, consisting of the functions $u = u(x,\mu,E)$, which are finite with respect to the norm

$$\|u\|_1 = \int_0^a \int_{-1}^{1} \int_{E_m}^{E_M} |u(x,\mu,E)|\,dE\,d\mu\,dx,$$

while f, $h_0 \in \mathscr{N}$, $g_0 \in \mathscr{N}_+$, $g_a \in \mathscr{N}_-$, and $g_i \in \mathscr{N}_i$ are given functions. Here $\mathscr{N}_\pm$ is the Banach space of all functions $g = g(\mu,E)$ finite with respect to the norm

$$\|g\|_1 = \pm \int_0^{\pm 1} \int_{E_m}^{E_M} |\mu|\,|g(\mu,E)|\,dE\,d\mu,$$

while $\mathscr{N}_i$ is the Banach space of all functions $g = g(x,\mu)$ finite with respect to the norm

$$\mathscr{N}_i = \int_0^a \int_{-1}^{1} \beta(x,E_M)\,|g(x,\mu)|\,d\mu\,dx.$$

Note that the stopping power at the higher end of the energy range appears as a weight in the L_1 norm of $\mathscr{N}_i$.

In recent years there has been renewed activity on the Spencer–Lewis equation, in part because of the necessity of proving the convergence of the existing finite difference methods for solving Eqs. (4) and (5). Nelson and Seth[5] proved the convergence of certain finite difference schemes under the assumption that Eqs. (4) and (5) have a unique solution. For a simple rod model the well-posedness of the original problem was proved by Nelson.[6] After the emergence of the abstract time-dependent kinetic theory of Beals and Protopopescu,[7] which can also be found in the monograph of Greenberg *et al.*,[8] these results have been extended for the time-dependent and the stationary problem to the case where (i) $\beta = \beta(x,E)$ is piecewise constant in energy and Lipschitz continuous in position, (ii) $\sigma = \sigma(x,E)$ is bounded and independent of μ, and (iii) when treating the stationary problem, condition (6) is replaced by

$$\sigma(x,\mu',E) \geqslant \delta \int_{-1}^{1} \sigma_s(x,\mu,\mu',E)\,d\mu,$$
$$\text{for all } (x,\mu',E)\in[0,a]\times[-1,1]\times[E_m,E_M], \tag{7}$$

where $\delta\in(0,1)$. Condition (i) was imposed to make the vector fields appearing in Eqs. (1) and (4) divergence-free so that the theory of Refs. 7 and 8 goes through. Condition (ii) implies the boundedness of the integral operator at the right-hand side of Eqs. (1) and (4) on $\mathscr{N}$, which is another prerequisite of the theory of Refs. 7 and 8. Condition (iii) implies that the evolution semigroup of Eq. (1) is exponentially decreasing in time, which makes the corresponding stationary problem uniquely solvable. For these results we refer to Sec. XIII.3 of Ref. 8.

Recently a number of new developments in abstract kinetic theory have taken place that will enable us to drop the above rather artificial conditions (i)–(iii) from the theory of Spencer–Lewis equations and to prove the unique solvability of both the time-dependent and the stationary problem under more natural assumptions. It has become clear[9,10] how to treat non-divergence-free force fields and thus how to drop condition (i). Furthermore, abstract kinetic theory has been extended to the case where the integral term of the collision operator is a (positive) contraction from $\mathscr{N}_\sigma = L_1(\Lambda;\sigma(x,\mu,E)dx\,d\mu dE)$ into $\mathscr{N} = L_1(\Lambda;dx\,d\mu\,dE)$ (see Refs. 9 and 10 for treatments of similar situations). These novel developments will guide us in the construction of an existence and uniqueness theory for the solution of the Spencer–Lewis equation under the following assumptions.

(A) There exists a partition $E_m = E_0 < E_1 < \cdots < E_r = E_M$, possibly with $r = 1$, of the intermediate energy range such that β is non-negative and Lipschitz continuous on the closure of each set

$$\Lambda_i = (0,a)\times(-1,1)\times(E_{i-1},E_i), \tag{8}$$

where $i = 1,\ldots,r$.

(B) The stopping power is Lipschitz continuous on the *disjoint* union $\cup_{i=1}^{r}\overline{\Lambda}_i$ and has only finitely many zeros, all of them in the interior of $(0,a)\times(E_m,E_M)$. Thus in defining the stopping power one should distinguish between E_i^- and E_i^+, for $i = 1,\ldots,r-1$ if $r\geqslant 2$.

(C) σ and σ_s satisfy condition (6) and

$$\int_0^a \int_{-1}^{1} \int_{E_m}^{E_M} \sigma(x,\mu,E)\,dE\,d\mu\,dx < \infty.$$

Apart from this integrability condition, σ may be unbounded. When $r\geqslant 2$, we will also require the solutions u of Eqs. (1)–(3) and Eqs. (4) and (5) to be continuous at the energy jumps, i.e., to satisfy $u(E_i^-) = u(E_i^+)$, for $i = 1,\ldots,r-1$. The physical meaning of this requirement is that discontinuous jumps in the stopping power do not bring about (positive or negative) electron sources. We will apply the method of characteristics in such a way that this continuity requirement is incorporated in the mathematical formulation in such a way that it does not show up as a boundary condition any more. In this fashion we will accomplish a major simplification of the method of characteristics used in Sec. XIII 3 of Ref. 8.

In Sec. II we will solve the time-dependent problem using the method of characteristics. The stationary problem will be the topic of Sec. III. However, we will solve this problem by reformulating it as an *initial*-value problem and applying the method of characteristics in the usual way. Section IV is devoted to semigroup properties and Sec. V to a discussion of the results.

Remark: Recently E. Ringeisen (Centre de Mathématiques Appliquées, École Normale Supérieure, Paris) proved the unique solvability of the stationary Spencer–Lewis equation under the assumptions that (i) the stopping power $\beta = \beta(x,E)$ is continuously differentiable on

$[0,a]\times[E_m,E_M]$, (ii) the cross sections σ and σ_s are L_∞ functions, and

(iii) $$\int_{-1}^{1}\sigma_s(x,E,\mu,\mu')\times\exp\left[-\frac{\delta(x,\mu',E)(\|\sigma\|_\infty+\lambda_1)}{|\mu|}\right]d\mu'$$

is bounded away from zero for some positive constant λ_1. Here $\delta(x,\mu,E)$ is the length of the maximal integral curve of the vector field X (defined below) passing through (x,μ,E).

II. THE TIME-DEPENDENT PROBLEM

On the set $\Lambda_t=\cup_{i=1}^{r}\Lambda_i$ with the union thought of as disjoint and endowed with the Lebesgue measure we introduce the vector field

$$X=\mu\frac{\partial}{\partial x}-\beta(x,E)\frac{\partial}{\partial E},$$

which is clearly Lipschitz continuous on the closure Λ_t (when distinguishing between E_i^- and E_i^+, for $i=1,...,r$). Using time t as a parameter there is a unique integral curve of X through each point of Λ_t satisfying the characteristic equations

$$\frac{dx}{dt}=\mu,\quad\frac{d\mu}{dt}=0,\quad\frac{dE}{dt}=-\beta(x,E).$$

In contrast to the practice of Refs. 7–9, we will identify all points of the type (x,μ,E_i^-) with the corresponding points (x,μ,E_i^+), thus obtaining the original manifold Λ, and continue the integral curves of X across the energy interfaces $E=E_i$ $(i=1,2,...,r-1)$. The sets $D_\pm$ of left and right end points of the integral curves of X passing through an interior point of Λ_t are then given by

$$\begin{aligned}D_-&=[\{0\}\times(0,1)\times(E_m,E_M)]\\&\quad\cup[\{a\}\times(-1,0)\times(E_m,E_M)]\\&\quad\cup[(0,a)\times(-1,1)\times\{E_M\}],\\D_+&=[\{0\}\times(-1,0)\times(E_m,E_M)]\\&\quad\cup[\{a\}\times(0,1)\times(E_m,E_M)]\\&\quad\cup[(0,a)\times(-1,1)\times\{E_m\}].\end{aligned}$$

Along the integral curves the energy E is steadily decreasing. We now parametrize $\overline{\Lambda}$ as

$$\overline{\Lambda}=\{(z,t):\ z\in D_-,\ 0\leqslant t<l(z)\},$$

where $l(z)$ is the travel time along the entire trajectory of X starting from $z\in D_-$. From nonzero μ there is a maximal travel time along a trajectory having μ as a constant of motion which is bounded above by $a/|\mu|$. However, on the trajectory with $\mu=0$ and $x\in(0,a)$ as constants of motion the total travel time

$$l(z)=\int_{E_m}^{E_M}\frac{1}{\beta(x,E)}\,dE=+\infty,$$

because $\beta(x,E_M)\neq0$, β has only finitely many zeros on Λ, and β is Lipschitz continuous. [Note that $l(z)=+\infty$ if $\beta(x,E_0)=0$ for some $E_0\in(E_m,E_M)$, because the Lipschitz condition on β implies $|\beta(x,E)|\leqslant L|E-E_0|$ with $L>0$.] Similarly, one may parametrize $\overline{\Lambda}\times[0,T]$ as

$$\overline{\Lambda}\times[0,T]=\{(z,s):\ z\in[\Lambda\times\{0\}]\cup[D_-\times(0,T)],\ 0\leqslant s\leqslant\max(l(z),T)\},$$

where s is the travel time parameter along the trajectory of $Y=\partial/\partial t+X$ from its left end point on either $\Lambda\times\{0\}$ or $D_-\times(0,T)$. To avoid confusion between t as a variable appearing in the vector field Y and the parameter in the characteristic equations of Y, we will use t as the time variable and s as the travel time parameter.

For every $u\in L_1(\Sigma)$ with $\Sigma=\Lambda\times(0,T)$ one may define

$$Yu=\frac{\partial u}{\partial t}+Xu,\quad Xu=\mu\frac{\partial u}{\partial x}-\beta\frac{\partial u}{\partial E},$$

as distributional directional derivates by

$$\int_\Sigma\left\{(Yu)v+u(Yv)+\frac{\partial\beta}{\partial E}uv\right\}dx\,d\mu\,dE\,dt=0$$

and

$$\int_\Lambda\left\{(Xu)v+u(Xv)+\frac{\partial\beta}{\partial E}uv\right\}dx\,d\mu\,dE=0,$$

where v belongs to the test function space Φ_0 of all real Borel functions on Σ (resp. Λ) that are bounded, are continuously differentiable along the trajectories of Y (resp. X) with bounded directional derivative Yv (resp. Xv), vanish at the end points of each trajectory and have the property that the lengths of the trajectories meeting the support of v are bounded away from zero. The latter means in particular that $|\mu|$ is bounded away from zero on the support of each $v\in\Phi_0$. Note that $\partial\beta/\partial E$ exists almost everywhere as a result of the absolute continuity of β.

Below we will employ the spaces $\mathscr{M}$, $\mathscr{M}_\pm$, and $\mathscr{M}_i$. These spaces are defined in the same way as the $\mathscr{N}$ spaces, i.e., again as L_1 spaces but with F replaced by $F\times(0,T)$ and the underlying measure replaced by its product with the Lebesgue measure on $(0,T)$. These spaces may also be represented as the spaces of all Bochner integrable functions from $(0,T)$ into the corresponding $\mathscr{N}$ space endowed with the L_1 norm (cf. Ref. 11).

Lemma 2.1: Suppose $h_0\in\mathscr{N}$, $g_0\in\mathscr{M}_+$, $g_a\in\mathscr{M}_-$, $g_i\in\mathscr{M}_i$, and $f\in\mathscr{M}$. Then there exists a unique solution u of the initial-boundary-value problem

$$Yu+\left\{\sigma(x,\mu,E)-\frac{\partial\beta}{\partial E}\right\}u=f,\tag{9}$$

$$u(x,\mu,E,t=0)=h_0(x,\mu,E),\tag{10}$$

$$u(x=0,\mu,E,t)=g_0(\mu,E,t),\quad\mu>0,\tag{11}$$

$$u(x=a,\mu,E,t)=g_a(\mu,E,t)\quad\mu<0,\tag{12}$$

$$u(x,\mu,E=E_M,t)=g_i(x,\mu,t).\tag{13}$$

The solution u and the left-hand sides of Eqs. (10)–(13) have the following properties.

(i) $u\in\mathscr{M}$ while the left-hand sides of Eqs. (10)–(13) belong to $\mathscr{M}$, $\mathscr{M}_+$, $\mathscr{M}_-$, and $\mathscr{M}_i$, respectively.

(ii) Together with $u(x,\mu,E=E_m,t)$, which belongs to the space $L_1(M;\beta(x,E_m)dx\,d\mu\,dt)$ with $M=(0,a)$

$\times(-1,1)\times(0,T)$, and $u(x,\mu,E,t=T)$ in $\mathcal{N}$, these functions are related by the Green identity

$$\int_\Sigma \left(\frac{\partial u}{\partial t}+\mu\frac{\partial u}{\partial x}-\frac{\partial}{\partial E}(\beta u)\right)dx\,d\mu\,dE\,dt$$
$$=\int_M (\beta(x,E_m)u(x,\mu,E_m,t)$$
$$-\beta(x,E_M)u(x,\mu,E_M,t))dx\,d\mu\,dt$$
$$+\int_0^T\int_{E_m}^{E_M}\int_{-1}^{1}\mu\{u(a,\mu,E,t)$$
$$-u(0,\mu,E,t)\}d\mu\,dE\,dt$$
$$+\int_\Lambda \{u(x,\mu,E,T)-u(x,\mu,E,0)\}dx\,d\mu\,dE. \quad (14)$$

Here the functions $u(x=0,\mu,E,t)$ and $u(x=a,\mu,E,t)$ belong to the L_1 space $L_1((-1,1)\times(E_m,E_M)\times(0,T);|\mu|d\mu\,dE\,dt)$.

(iii) $\sigma u\in\mathcal{M}$ and

$$\|\sigma u\|_{\mathcal{M}}\leqslant\|f\|_{\mathcal{M}}+\|h_0\|_{\mathcal{N}}+\|g_0\|_{\mathcal{M}}+\|g_a\|_{\mathcal{M}_+}+\|g_i\|_{\mathcal{M}_i}. \quad (15)$$

Proof: Writing

$$h(x,\mu,E)=\sigma(x,\mu,E)-\frac{\partial\beta}{\partial E}$$

and using the above parametrization of Y, we reduce Eqs. (9)–(13) to the initial-value problem

$$\frac{du}{ds}=h(z,s)u(z,s)=f(z,s), \quad (16)$$

$$u(z,s=0)=g(z), \quad (17)$$

where, modulo the parametrization, $g(z)$ coincides with $h_0(x,\mu,E)$ on $\Lambda\times\{0\}$, $g_0(\mu,E,t)$ on $\{0\}\times(0,1)\times(E_m,E_M)\times(0,T)$, $g_a(\mu,E,t)$ on $\{a\}\times(-1,0)\times(E_m,E_M)\times(0,T)$, and $g_i(x,\mu,t)$ on $(0,a)\times(-1,1)\times(0,T)$. Since β is (piecewise) absolutely continuous on $\overline{\Lambda}$, it has an almost everywhere defined derivative $\partial\beta/\partial E$, which belongs to $L_1(\Lambda)$; hence, by assumption (C), $h\in L_1(\Lambda)$. As the unique solution we find

$$u(z,s)=\exp\left\{-\int_0^s h(z,\sigma)d\sigma\right\}g(z)+\int_0^s\exp\left\{-\int_\tau^s h(z,\sigma)d\sigma\right\}f(z,\tau)d\tau.$$

Here the uniqueness, with the derivatives in Eq. (16) taken in distributional sense, follows from the Green's identity applied to Eqs. (16) and (17) with $g=0$ and $f=0$. Hence $u(x,\mu,E,t=0)$, $u(x=0\,\mu,E,t)$, $u(x=a,\mu,E,t)$, and $u(x,\mu,E=E_M,t)$ have the appropriate properties and satisfy Eqs. (10)–(13). Further, $u(x,\mu,E=E_m,t)$, $u(x,\mu,E,t=T)$, $u(x=0,\mu,E,t)$, and $u(x=a,\mu,E,t)$ also have the appropriate properties and Eq. (14) is satisfied. In fact, Eq. (14) can be written as

$$\int_\Sigma\left(Y-\frac{\partial B}{\partial E}\right)u\,dx\,d\mu\,dE\,dt=\int_{D^+}u^+\,dv^+(z)-\int_{D^-}u^-\,dv^-(z), \quad (18)$$

where $D^-=[\Lambda\times\{0\}]\cup[D_-\times(0,T)]$ and $D^+=[\Lambda\times\{T\}]\cup[D_+\times(0,T)]$ for appropriate positive Borel measures $dv^\pm(z)$ which are weighted Lebesgue measures with $|\mu|$ as the weight on $\{x=0\}$ and $\{x=a\}$, 1 as the weight on $\{t=0\}$ and $\{t=T\}$, $\beta(x,E_M)$ as the weight on $\{E=E_M\}$, and $\beta(x,E_m)$ as the weight on $\{x,E_m\}$. Finally, to prove (15) it suffices to restrict onself to non-negative g_0, g_0, g_a, g_i, and f. For non-negative data we have

$$\|\sigma u\|=\left\|\left(h+\frac{\partial\beta}{\partial E}\right)u\right\|$$
$$=\|f\|-\left\|\left(Y-\frac{\partial\beta}{\partial E}\right)u\right\|$$
$$=\|f\|+\|u^-\|-\|u^+\|\leqslant\|f\|+\|u^-\|$$
$$=\|f\|+\|h_0\|+\|g_0\|+\|g_a\|+\|g_i\|,$$

which proves the lemma. □

We now define the positive operator

$$(Ju)(x,\mu,E)=\int_\Lambda\sigma_s(x,\mu,\mu',E)u(x,\mu',E)d\mu', \quad (19)$$

which satisfies

$$\|Ju\|_{\mathcal{N}}\leqslant\|\sigma u\|_{\mathcal{N}},\quad u\in\mathcal{N}_\sigma=L_1(\Lambda;\sigma\,dx\,d\mu\,dE). \quad (20)$$

We will denote the norm of J as a contraction from $\mathcal{N}_\sigma$ into $\mathcal{N}$ by $\|J\|_\#$.

Lemma 2.2: Suppose $\|J\|_\#<1$, i.e., suppose

$$\sigma(x,\mu',E)\geqslant\delta\int_{-1}^{1}\sigma_{s(x,\mu,\mu',E)}\,d\mu,$$

for some $\delta\in(0,1)$. Then there exists a unique solution u of the initial-boundary-value problem

$$Yu+\left\{\sigma(x,\mu,E)-\frac{\partial\beta}{\partial E}\right\}u=Ju+f, \quad (21)$$

$$u(x,\mu,E,t=0)=h_0(x,\mu,E), \quad (22)$$

$$u(x=0,\mu,E,t)=g_0(\mu,E,t),\quad \mu>0, \quad (23)$$

$$u(x=a,\mu,E,t)=g_a(\mu,E,t),\quad \mu<0, \quad (24)$$

$$u(x,\mu,E=E_M,t)=g_i(x,\mu,t). \quad (25)$$

The solution u and the left-hand sides of Eqs. (22)–(25) have the properties (i) and (ii) in the statement of Lemma 2.1, while (iii) is replaced by (iii') σu belongs to $\mathcal{M}$ and

$$\|\sigma u\|\leqslant(1-\|J\|_\#)^{-1}(\|f\|_{\mathcal{M}}+\|h_0\|_{\mathcal{N}}+\|g_0\|_{\mathcal{M}}+\|g_a\|_{\mathcal{M}_+}+\|g_i\|_{\mathcal{M}_i}). \quad (26)$$

Proof: Let us write Eqs. (21)–(25) as

$$(Y+h)u=Ju+f, \quad (27)$$

$$u^-=g, \quad (28)$$

where g is defined as in the proof of Lemma 2.1. Denoting the solution of Eqs. (16) and (17) as $u=S(f,g)$, we represent the solution of Eqs. (27) and (28) as $u=S(f^*,g)$. Then $f^*\in\mathcal{M}$ satisfies the equation

$$(\mathbb{1}+L)f^*=f+JS(0,g), \quad (29)$$

where

$$Lf^*=-JS(f^*,0).$$

Then

$$\|Lf^*\|_{\mathscr{N}} \leqslant \|J\|_\# \|\sigma S(f^*,0)\|_{\mathscr{N}} \leqslant \|J\|_\# \|f^*\|_{\mathscr{N}},$$

so that Eq. (29) has a unique solution f^* satisfying

$$\|f^*\| \leqslant \sum_{n=0}^{\infty} \|J\|_\#^n \|f + JS(0,g)\|$$

$$\leqslant (1 - \|J\|_\#)^{-1}(\|f\| + \|J\|_\# \|g\|) .$$

Hence

$$\|\sigma u\| = \|\sigma S(f^*,g)\| \leqslant \|f^*\| + \|g\|$$

$$\leqslant (1 - \|J\|_\#)^{-1}(\|f\| + \|g\|),$$

which proves the lemma. □

For non-negative data we directly obtain from Eq. (27)

$$\|(Y+b)u\| + \|(h-b)u\| + \|g\| = \|Ju\| + \|f\| + \|g\|,$$

where $b = -(\partial\beta/\partial E)$ and all norms are L_1 norms. Using the Green's identity (18) we have

$$\|u^+\| + \|\sigma u\| = \|Ju\| + \|f\| + \|g\|,$$

whence

$$\|u^+\| + (1 - \|J\|_\#)\|\sigma u\| \leqslant \|f\| + \|g\|. \tag{30}$$

Here

$$u^+ = (u(t=T), u(E=E_m),$$
$$u(x=0, \mu<0), u(x=a, \mu \geqslant 0))$$

on a direct sum of L_1 spaces with certain weights.

When $\|J\|_\# = 1$, we cannot apply the same perturbation arguments as in the proof of Lemma 2.2. Instead we approximate u monotonically by the unique solutions u_n of the initial-boundary-value problems

$$(Y+h)u_n = \beta_n J u_n + f, \tag{31}$$

$$(u_n)^- = g, \tag{32}$$

where $\{\beta_n\}_{n=1}^{\infty} \uparrow 1$. These solutions are non-negative, are nondecreasing with n, and satisfy

$$\|u_n^+\| + (1-\beta_n)\|\sigma u_n\| \leqslant \|f\| + \|g\|.$$

Hence there exists u^+ such that

$$\lim_{n\to\infty} \|u^+ - u_n^+\| = 0 \tag{33}$$

in the norm of $\mathscr{N}_+$ and Eq. (28) is satisfied. On the other hand,

$$f - \left(Y - \frac{\partial\beta}{\partial E}\right) u_n = \sigma u_n - \beta_n J u_n,$$

while

$$\int_\Lambda \left(Y - \frac{\partial\beta}{\partial E}\right)[u_k - u_n]\,dx\,d\mu\,dE\,dt = \|u_k^+ - u_n^+\| \to 0,$$

as $k,n \to \infty$. Hence there exists $w \in \mathscr{M}$ such that

$$\lim_{n\to\infty} \|\sigma u_n - \beta_n J u_n - w\| = 0. \tag{34}$$

Then we also have

$$\lim_{n\to\infty} \left\| \left(Y - \frac{\partial\beta}{\partial E}\right) u_n - [f - w] \right\| = 0. \tag{35}$$

We now solve the initial-boundary-value problem

$$\left(Y - \frac{\partial\beta}{\partial E}\right) u = f - w, \quad u^- = g,$$

and find a solution $u \in \mathscr{M}$ having the properties (i) and (ii) in the statement of Lemma 2.1. We will show that u is a solution of Eqs. (22)–(25), but in a rather weak sense.

Indeed, from (30) and $\|\beta_n J\|_\# = \beta_n$ it is clear that

$$(1-\beta_n)\|Ju_n\| \leqslant (1-\beta_n)\|\sigma u_n\| \leqslant \|f\| + \|g\|,$$

so that $\{(\sigma - J)u_n\}_{n=1}^{\infty}$ is a bounded sequence in $\mathscr{M}$. Since

$$\lim_{n\to\infty} \|[f - (\sigma - \beta_n J)u_n] - [f - w]\| = 0,$$

we have

$$\lim_{n\to\infty} \|u - u_n\| = 0.$$

To establish the uniqueness of the solution, we assume that u is a (real) solution of the homogeneous time-dependent problem

$$(Y+b)u + (\sigma - J)u = 0, \quad u^- = 0,$$

where $b = -\partial\beta/\partial E$. Then $u = \operatorname{sgn}(u)|u|$ and hence

$$(Y+b)u + (\sigma - J)|u| + \{J|u| - \operatorname{sgn}(u)Ju\} = 0,$$
$$|u|^- = 0.$$

Integrating over position–velocity–energy–time phase space and using $u^- = 0$ we get

$$\|u^+\| + \iiiint (\sigma - J)|u|\,dx\,d\mu\,dE\,dt$$
$$+ \|\{J|u| - \operatorname{sgn}(u)Ju\}\| = 0.$$

The second term on the left-hand side is non-negative [cf. Eq. (6)], so that $u^+ = 0$, $(\sigma - J)|u|$ has a zero integral over phase space, and $\{J|u| - \operatorname{sgn}(u)Ju\} = 0$, so that $|u|$ is a solution of Eqs. (27) with $f = 0$. Thus without loss of generality we may assume $u \geqslant 0$. We then find

$$u(z,s) = \int_0^s \exp\left\{-\int_\tau^s (\sigma + b)(z,\tau')d\tau'\right\}(Ju)(z,\tau)d\tau.$$

For $s = l(z)$ we get u^+, which vanishes. Moreover, $Ju \geqslant 0$. Hence $Ju \equiv 0$ and $u \equiv 0$, which settles the uniqueness issue.

We have therefore established the following.

Theorem 2.3: There exists a unique solution u of the initial-boundary-value problem (21)–(25). The solution u and the left-hand sides of Eqs. (22)–(25) have the properties (i) and (ii) in the statement of Lemma 2.1, $(\sigma u - Ju) \in \mathscr{M}$, but σu and Ju themselves need not belong to $\mathscr{M}$.

Corollary 2.4: If $g_0 \equiv 0$, $g_a \equiv 0$, $g_i \equiv 0$, and $f \equiv 0$, the unique solution of Eqs. (22)–(25) can be represented as $u(t) = S(t)h_0$, $t \geqslant 0$, where $\{S(t)\}_{t \geqslant 0}$ is a positive contraction semigroup on $\mathscr{N}$. This semigroup satisfies $\|S(t)h_0\| = \|h_0\|$, for all non-negative $h_0 \in \mathscr{N}$ and all $t \geqslant 0$, if and only if $\|Ju\| = \|\sigma u\|$ for all non-negative $u \in \mathscr{N}_\sigma$.

Proof: The first part is clear from the estimates

$$\|u(t=T)\| \leqslant \|u^+\| \leqslant \|f\| + \|g\| = \|h_0\|.$$

The second part is a simple consequence of the Green's identity for non-negative data. □

III. THE STATIONARY PROBLEM

The stationary problem is given by Eqs. (4) and (5) and can thus be written as

$$Xu + \left(\sigma - \frac{\partial\beta}{\partial E}\right) u = Ju + f, \tag{36}$$

$$u_{+} = g, \tag{37}$$

where $g = (g_0, g_a, g_i)$. We have to find a function $u \in \mathcal{N}$ such that $u_{-} \in \mathcal{N}_{-} \oplus \mathcal{N}_{+} \oplus \mathcal{N}_i$ and Eqs. (36) and (37) are satisfied. Here we assume that $f \in \mathcal{N}$ and $g \in \mathcal{N}_{-} \oplus \mathcal{N}_{+} \oplus \mathcal{N}_i$. As in the previous section we distinguish between the cases $\|J\|_{\#} < 1$ and $\|J\|_{\#} = 1$.

Lemma 3.1: Suppose $\|J\|_{\#} < 1$. Then Eqs. (36) and (37) have a unique solution u such that $\sigma u \in \mathcal{N}$ and

$$\|\sigma u\| \leqslant (1 - \|J\|_{\#})^{-1}\{\|f\| + \|g\|\}. \tag{38}$$

The solution is non-negative for non-negative data f and g.

Proof: Writing $h = \sigma - (\partial\beta/\partial E)$, we solve Eqs. (36) and (37) for $J = 0$ and obtain

$$u(z,s) = \exp\left\{-\int_0^s h(z,\sigma)d\sigma\right\} g(z) + \int_0^s \exp\left\{-\int_\tau^s h(z,\sigma)d\sigma\right\} f(z,\tau)d\tau,$$

which has the desired properties. The Green's identity for X gives as before

$$\|\sigma u\| \leqslant \|f\| + \|g\|.$$

We write $u = S(f,g)$.

As in the previous section we represent the solution of Eqs. (36) and (37) as $u = S(f^*,g)$, where $(\mathbb{1} + L) f^* = f + JS(0,g)$ and $\|L\| \leqslant \|J\|_{\#} < 1$. We then find a unique $f^* \in \mathcal{N}$, which is non-negative for non-negative f and g because $(-L) \geqslant 0$. A simple estimation then gives (38). □

To pass to the case $\|J\|_{\#} = 1$, we use monotone approximation by the solutions u_n of the stationary problem

$$Xu_n + \left(\sigma - \frac{\partial\beta}{\partial E}\right) u_n = \beta_n J u_n + f, \tag{39}$$

$$u_{n,-} = g, \tag{40}$$

where $(\beta_n)_{n=1}^{\infty} \uparrow 1$. Using the Green's identity for X we find

$$\|u_{n,-}\| + (1 - \beta_n)\|\sigma u_n\| \leqslant \|f\| + \|g\|,$$

so that $\{u_{n,-}\}_{n=1}^{\infty}$ converges monotonically to some $u_{-} \in \mathcal{N}_{-} \oplus \mathcal{N}_{+} \oplus \mathcal{N}_i$ in the strong sense. Then

$$\left\|\left(X - \frac{\partial\beta}{\partial E}\right)[u_n - u_k]\right\| \leqslant \|u_{n,-} - u_{k,-}\|$$

implies that $\{(X - \partial\beta/\partial E)u_n\}_{n=1}^{\infty}$ converges in $\mathcal{N}$ to some limit w. We then have

$$\lim_{n\to\infty} \|(\sigma u_n - \beta_n J u_n) - (f - w)\| = 0. \tag{41}$$

Letting u be the unique solution in $\mathcal{N}$ of the trivial stationary problem

$$\left(X - \frac{\partial\beta}{\partial E}\right) u = f - w,$$

$$u_{-} = g,$$

which depends continuously on $(f - w)$ and g, we find from (41) that

$$\lim_{n\to\infty} \|u_n - u\| = 0. \tag{42}$$

The uniqueness issue is settled in the same way as for the time-dependent problem with $\|J\|_{\#} = 1$.

We have the following theorem.

Theorem 3.2: Suppose $\|J\|_{\#} \leqslant 1$. Then Eqs. (36) and (37) have a unique solution u such that $(X - \partial\beta/\partial E)u \in \mathcal{N}$ and $(\sigma u - Ju) \in \mathcal{N}$. This solution is non-negative for non-negative data f and g.

IV. SEMIGROUP FORMULATION OF THE TIME-DEPENDENT SOLUTION

We have proved the unique solvability of Eqs. (1)–(3) in an L_1 space of functions u on $\Lambda \times [0,T]$. We have proved these solutions to have L_1 traces on each hyperplane $t = t_0$ with $t_0 \in [0,T]$. This follows from the inclusion $(t = T) \subseteq D^+$, the finiteness of $\|u^+\|$, and the arbitrariness of T (so that we may replace T by t_0). Nevertheless, we have not studied their continuity properties as a function of t. In this section we intend to do so. In order to apply the Hille–Yosida theorem (cf. Ref. 12) we will first study the stationary equation

$$Xu_\lambda + \left(\sigma + \lambda - \frac{\partial\beta}{\partial E}\right) u_\lambda = Ju_\lambda + f, \tag{43}$$

with boundary conditions

$$u_{\lambda_-} = g, \tag{44}$$

i.e., Eqs. (36) and (37) with σ replaced by $\sigma + \lambda$, where $\lambda > 0$. According to Lemma 3.1, Eqs. (43) and (44) have a unique solution $u_\lambda \in \mathcal{N}$ such that $(\sigma + \lambda)u_\lambda \in \mathcal{N}$ and

$$\|(\sigma + \lambda)u_\lambda\| \leqslant (1 - \|J\|_{\#})^{-1}\{\|f\| + \|g\|\}$$

whenever $\|J\|_{\#} < 1$, and this solution u_λ is non-negative for non-negative data f and g. Thus if f and g are non-negative, then u_λ satisfies Eq. (38) as well as the estimate

$$\|u_\lambda\| \leqslant \lambda^{-1}(1 - \|J\|_{\#})^{-1}\{\|f\| + \|g\|\}.$$

Hence if we define the operator $\mathcal{J}$ by

$$\mathcal{J} = -X - \left(\sigma(x,\mu,E) - \frac{\partial\beta}{\partial E}\right),$$

$$D(\mathcal{J}) = \left\{u \in \mathcal{N}: \left(X - \frac{\partial\beta}{\partial E}\right) u \in \mathcal{N},\ \sigma u \in \mathcal{N},\ u|_{D_-} \equiv 0\right\},$$

then for $g \equiv 0$ there exists a unique solution $u_\lambda \in D(\mathcal{J})$ such that

$$(\lambda - \mathcal{J} - J)u_\lambda = f,$$

which satisfies

$$\|u_\lambda\| \leqslant \lambda^{-1}(1 - \|J\|_{\#})^{-1}\|f\|.$$

Thus for $\|J\|_{\#} < 1$ the operator $\mathcal{J} + J$ with domain

$$D(\mathcal{J} + J) = D(\mathcal{J})$$

generates a bounded strongly continuous semigroup on $\mathcal{N}$, which we will denote as $\{S(t)\}_{t>0}$.

Let us apply Lemma 2.1 to Eqs. (10)–(13) for nonnegative data. We obtain immediately

$$\|f\| = \int_\Sigma \left(\frac{\partial u}{\partial t} + Xu + \left\{\sigma - \frac{\partial \beta}{\partial E} - J\right\}u\right)dx\, d\mu\, dE\, dt$$
$$= \|u^+\| - \|u^-\| + \|\sigma u\| - \|Ju\| \geqslant \|u^+\| - \|u^-\|,$$

so that

$$\|u(t=T)\| \leqslant \|u(t=0)\|$$

whenever $f\equiv 0$, $g_0 \equiv 0$, $g_a \equiv 0$, and $g_i \equiv 0$. Hence the semigroup $\{S(t)\}_{t>0}$ is a contraction semigroup.

We have the following theorem.

Theorem 4.1: Suppose $\|J\|_\# < 1$. Then the operator $-X - (\sigma - \partial\beta/\partial E + J)$ generates a strongly continuous contraction semigroup $\{S(t)\}_{t>0}$ on $\mathcal{N}$.

Suppose $\|J\|_\# = 1$ and let us approximate J from below by $\beta_n J$, where $\{\beta_n\}_{n=1}^\infty \uparrow 1$. Denoting the corresponding contraction semigroups on $\mathcal{N}$ by $\{S_n(t)\}_{t>0}$ we use that $S_n(t) \leqslant S_m(t)$, for $n \leqslant m$, as well as the upper bound $\|S_n(t)\| \leqslant 1$. We thus obtain the family of contraction operators $\{S_n(t)\}_{t>0}$ on $\mathcal{N}$ satisfying

$$\lim_{n\to\infty} \|[S(t) - S_n(t)]g\| = 0, \quad t \geqslant 0,$$

as well as the semigroup property. If we now define

$$R(\lambda)g = \int_0^\infty e^{-\lambda t} S(t)g\, dt, \tag{45}$$

for Re $\lambda \geqslant 0$, we obtain

$$\|R(\lambda)g\| \leqslant [1/(\operatorname{Re}\lambda)]\|g\|, \quad \operatorname{Re}\lambda > 0.$$

On the other hand,

$$(\lambda - [\mathcal{J} + \beta_n J])^{-1} g = \int_0^\infty e^{-\lambda t} S_n(t) g\, dt.$$

Thus by dominated convergence we obtain

$$\lim_{n\to\infty} \|[R(\lambda) - (\lambda - [\mathcal{J} + \beta_n J])^{-1}]g\| = 0, \quad \operatorname{Re}\lambda > 0.$$

We then find the resolvent identity

$$R(\lambda) - R(\mu) = -(\lambda - \mu) R(\lambda) R(\mu). \tag{46}$$

Using (46) we find that Ker $R(\lambda)$ and Ran $R(\lambda)$, the kernel and range of $R(\lambda)$, do not depend on λ. Since every $g \in \operatorname{Ker} R(\lambda)$ satisfies $S(t)g \equiv 0$ and $u \equiv S(t)g$ is a solution of Eqs. (1)–(3) in $\mathcal{M}$ for $f\equiv 0, g_0 \equiv 0, g_a \equiv 0, g_i \equiv 0$, and, given $u(t=0) = g$, we obtain $g \equiv 0$ by the unique solvability of the time-dependent problem so that Ker $R(\lambda) = \{0\}$. By a similar argument on the adjoint semigroup we get the density of Ran $R(\lambda)$ in $\mathcal{N}$. Hence $R(\lambda) = (\lambda - \mathcal{G})^{-1}$ for some closed and densely defined operator $\mathcal{G}$. Thus, by the Hille–Yosida theorem and the uniqueness of the Laplace transform, $\mathcal{G}$ is the generator of a strongly continuous semigroup of $\mathcal{N}$ that must necessarily coincide with $\{S(t)\}_{t>0}$. Thus $\{S(t)\}_{t>0}$ is a strongly continuous contraction semigroup on $\mathcal{N}$.

We have the following theorem.

Theorem 4.2: Suppose $\|J\|_\# = 1$. Then the closure of operator $-X - (\sigma - \partial\beta/\partial E + J)$ generates a strongly continuous contraction semigroup $\{S(t)\}_{t>0}$ on $\mathcal{N}$.

Proof: Clearly the generator $\mathcal{G}$ of $\{S(t)\}_{t>0}$ is a closed extension of $-X - (\sigma - \partial\beta/\partial E + J)$. It remains to prove its minimality. Indeed, observe that

$$D(\mathcal{G}) = \{R(\lambda)g : g \in \mathcal{N}\},$$

where Re $\lambda > 0$. Then for every $h \in D(\mathcal{G})$ we have

$$\lim_{n\to\infty} \|k_n - h\| = 0,$$

where $k_n = (\lambda - \mathcal{J} - \beta_n J)^{-1} g$ and g is the unique vector in $\mathcal{N}$ such that $R(\lambda)g = h$. Note that $k_n \in D(\mathcal{J} + J)$, which is true because $D(\mathcal{J}) \subseteq \mathcal{N}_\sigma$ so that J is well-defined on $D(\mathcal{J})$. Moreover,

$$(\lambda - \mathcal{J} - J)k_n = g + (1 - \beta_n) J k_n,$$

where $\{\sigma k_n\}$ is bounded; hence $(\lambda - \mathcal{J} - J)k_n \to g$. Thus every h belongs to the domain of the closure of $\mathcal{J} + J$ while $(\overline{\mathcal{J} + J})h = g$. But then we must have $\mathcal{G} = \overline{\mathcal{J} + J}$. □

V. DISCUSSION

We have established the unique solvability of the time-dependent and stationary Spencer–Lewis equations under natural assumptions on the stopping power and the cross section and in natural function spaces. These results are far more general than the existence and uniqueness results given by Nelson[6] and Greenberg *et al.*[8] On the other hand, Nelson and Seth[5] have established the convergence of a number of finite difference schemes for solving the stationary Spencer–Lewis equation numerically under the assumption that the corresponding stationary Spencer–Lewis equation is uniquely solvable. If we combine their conditional convergence proof with our well-posedness results, we obtain a convergence proof for the numerical schemes used by Nelson and Seth.[5]

The section on the stationary problem was very concise, because it appeared possible to treat both the time-dependent and the stationary problem by the method of characteristics as introduced in transport theory by Beals and Protopopescu.[7] Certain peculiarities of the Spencer–Lewis equation, however, forced us to go off the path followed by Ref. 7. The one rather artificial assumption left, assumption (B) on the number and position of the zeros of $\beta(x,E)$, may be dropped in the time-dependent case, provided one does not seek a restatement of the time-dependent result within the framework of semigroup theory. When adopting the semigroup framework or sticking to the stationary problem, assumption (B) is a necessary tool to avoid the intricacies of a singular vector field.

ACKNOWLEDGMENTS

The author wishes to express his gratitude to P. Nelson for stimulating discussions on the Spencer–Lewis equation. Also discussions with L. Arlotti and G. Frosali (University of Ancona) and S. L. Paveri-Fontana (University of Rome) who are dealing with similar mathematical questions in other transport problems turned out to be most useful.

This research was conducted while the author was a visiting professor supported by C.N.R. (Consiglio Nazionale per le Ricerche), Gruppo Nazionale per la Fisica Matematica.

[1]L. V. Spencer, Phys. Rev. **98**, 1597 (1955).
[2]H. W. Lewis, Phys. Rev. **78**, 526 (1950).
[3]D. E. Bartine, R. G. Alsmiller, Jr., F. R. Mynatt, W. W. Engle, Jr., and J. Barish, Nucl. Sci. Eng. **48**, 159 (1972).
[4]J. Arkuszewski, T. Kulikowska, and J. Mika, Nucl. Sci. Eng. **49**, 20 (1972).
[5]P. Nelson, Jr. and D. L. Seth, Transport Theory Stat. Phys. **15**, 211 (1986).
[6]P. Nelson, Jr., J. Math. Anal. Appl. **118**, 115 (1986).
[7]R. Beals and V. Protopopescu, J. Math. Anal. Appl. **121**, 370 (1987).
[8]W. Greenberg, C. V. M. Van der Mee, and V. Protopopescu, *Boundary Value Problems in Abstract Kinetic Theory* (Birkhäuser, Basel, 1987).
[9]C. V. M. Van der Mee, "Trace theorems and kinetic equations for non divergence free external forces," submitted to Appl. Anal.
[10]L. Arlotti, J. Math. Anal. Appl. **123**, 528 (1987).
[11]J. Diestel and J. J. Uhl, Jr., *Vector Measures* (Am. Math. Soc., Providence RI, 1977).
[12]T. Kato, *Perturbation Theory of Linear Operators* (Springer, Berlin, 1966).

Kac–Moody algebra from infinitesimal Riemann–Hilbert transform

Ling-Lie Chau and Mo-Lin Ge[a)]
Department of Physics, University of California, Davis, California 95616

(Received 11 August 1987; accepted for publication 18 May 1988)

A general formulation is given for deriving a Kac–Moody algebra in the spectrum space from the infinitesimal regular Riemann–Hilbert transform. Such a Kac–Moody algebra can be obtained for many nonlinear systems, e.g., the (super) principal chiral model with or without a Wess–Zumino–Witten term, the sine-Gordon, Liouville, and KdV equations; reduced two-dimensional gravity (Belinski–Zakharov gravity), and self-dual supersymmetric Yang–Mills theories.

I. INTRODUCTION

One of the common properties for many classical integrability systems is the existence of the loop algebras.[1,2,3] Many nonlinear equations in physics can be formulated as the integrability conditions of linearization systems.[4] The important feature is that there is an arbitrary spectral parameter λ in the linearization equations. In general, the λ dependence of a linearization equation is very simple. The λ dependence can be λ powerlike or simple singularity at a special value of λ. Observing the known linearization equations, for instance, we find a pole at $\lambda = 0$ for the Liouville equation and the sine-Gordon equation, poles at $\lambda = \pm 1$ for the chiral model, and no pole in the Landau–Lifshitz equation, etc. As for the gravity with two Killing vectors, the special form of the Lax pair can be introduced to avoid the cut along the real axis of λ.

This circumstance allows us to use the regular Riemann–Hilbert tranform (RRHT) as a powerful tool to treat the nonlinear phenomena. At present, the infinitesimal RRHT has been very useful in generating the loop algebras.[3]

In this paper we would like to develop a systematic approach to derive loop algebras for any linearization systems based on the RRHT. The point is to generate the loop algebras from the given algebraic structure for the group G.

The advantages of the approach will be seen explicitly in discussing the supersymmetric theories with a larger group and the two-dimensional gravity for Belinski–Zakharov (BZ) gravity (the Ernst formulation with a different matric sign).

II. INFINITESIMAL RRHT AND LOOP ALGEBRAS

The known linearization systems we discuss here can be summarized in the following generic form:

$$\bar{D}(\xi,\lambda)\Phi(\xi,\lambda) + \mathscr{A}(\xi,\lambda)\Phi(\xi,\lambda) = 0, \tag{2.1}$$

where λ is the spectral parameter and ξ represents the coordinate variables such as space-time coordinates, superspinor variables, and so on. In Eq. (2.1),

$$\bar{D}(\xi,\lambda) = D(\xi,\lambda) + d(\xi,\lambda), \tag{2.2}$$

where $\bar{D}(\xi,\lambda)$ stands for a derivative operator linear for the set of coordinates ξ and may be λ dependent. The term $d(\xi,\lambda) \sim \partial_\lambda$ is a linear derivative operator to λ with a ξ-dependent coefficient.

For instance, $\bar{D}(\xi,\lambda)$ represents $\lambda\,\partial_y - \partial_{\bar{z}}$ or $\lambda\,\partial_z + \partial_{\bar{y}}$ in the self-dual Yang–Mills (SDYM) case[5] and ∂_t or ∂_x in the chiral model (or ∂_ξ and ∂_η in light-cone coordinates). In Belinski–Zakharov gravity,

$$\bar{D}(\xi,\lambda) = \partial_\xi - \frac{2\lambda\alpha_\xi}{\lambda - \alpha}\partial_\lambda \quad \text{or} \quad \partial_\eta + \frac{2\lambda\alpha_\eta}{\lambda + \alpha}\partial_\lambda, \tag{2.3}$$

where α is the determinant of the matrix $\|g_{ab}\|$,[6] $\alpha_\xi \equiv \partial_\xi \alpha$, and $\alpha_\eta \equiv \partial_\eta \alpha$.

Under the Darboux-type transformation Eq. (2.1) submits to the transformations

$$\Phi'(\lambda) = X(\lambda)\Phi(\lambda), \tag{2.4}$$

$$\mathscr{A}'(\lambda) = X(\lambda)\mathscr{A}(\lambda)X^{-1}(\lambda) - (\bar{D}(\lambda)X(\lambda))X^{-1}(\lambda); \tag{2.5}$$

then Eq. (2.1) becomes

$$\bar{D}(\xi,\lambda)\Phi'(\xi,\lambda) + \mathscr{A}'(\xi,\lambda)\Phi'(\xi,\lambda) = 0, \tag{2.6}$$

for fixed ξ and λ. It is noted that in Eqs. (2.4), (2.5), and subsequent equations only λ dependence is stressed for simplicity. In general, $\mathscr{A}$, Φ, X, and $\bar{D}$ are all coordinate dependent.

For any linearization equation we can construct the RRHT[3] with the fixed point

$$X(\lambda = \infty) = 1, \tag{2.7}$$

$$X(\lambda) = 1 + W(\lambda), \quad \Phi' = X\Phi, \tag{2.8}$$

$$W(\lambda) = \frac{1}{2\pi i}\int_c \frac{dt}{t-\lambda}\,\tilde{G}(t), \quad \tilde{G}(t) = \Phi'(t)V(t)\Phi^{-1}(t), \tag{2.9}$$

[a)] On leave of absence from Theoretical Physics Division, Nankai Institute of Mathematics, Tianjin, People's Republic of China.

 0022-2488/89/010166-06$02.50

with

$$\bar{D}(\xi,\lambda)V(\xi,\lambda)=0. \tag{2.10}$$

Substituting Eqs. (2.7)–(2.10) into Eq. (2.5) we have

$$\mathscr{A}(\lambda)-\mathscr{A}'(\lambda)=\mathscr{A}'(\lambda)W(\lambda)-W(\lambda)\mathscr{A}(\lambda)+\bar{D}(\lambda)W(\lambda), \tag{2.11}$$

namely,

$$\mathscr{A}(\lambda)-\mathscr{A}'(\lambda)=\bar{D}(\lambda)\frac{1}{2\pi i}\int_c\frac{dt}{t-\lambda}\tilde{G}(t)+\frac{1}{2\pi i}\int_c\frac{dt}{t-\lambda}\times(\mathscr{A}'(t)\tilde{G}(t)-\tilde{G}(t)\mathscr{A}(t)). \tag{2.12}$$

For different fixed points other than $\lambda\to\infty$, say $\lambda=0$, a subtraction is needed. In general, $V(\xi,\lambda)$ must satisfy the requirements of the gruop and Eq. (2.10). In the infinitesimal case the infinitesimal parameter α^A in the transformation can be included in $V(\xi,\lambda)$. Sometimes α^A may be a Grassmann number where the supersymmetric theories are concerned.

The contour C in the complex λ plane can be chosen arbitrarily to avoid the possible singularities on the plane.

Now let us discuss the infinitesimal RRHT, i.e., $V(\lambda)$ belongs to the algebra of group G. In the case

$$\mathscr{A}'(\lambda)-\mathscr{A}(\lambda)=\alpha^a\delta_a\mathscr{A}(\lambda),\quad V_\alpha=\alpha^aV_a, \tag{2.13}$$

where α^a is an infinitesimal constant with the group index a. For the infinitesimal RRHT Eq. (2.12) can be reduced to

$$\delta_a\mathscr{A}(\lambda)=\frac{-1}{2\pi i}\bar{D}(\lambda)\int_c\frac{dt}{t-\lambda}G_a(t)-\frac{1}{2\pi i}\int_c\frac{dt}{t-\lambda}[\mathscr{A}(\lambda),G_a(t)], \tag{2.14}$$

where

$$G_a(t)=\Phi(t)V_a(t)\Phi^{-1}(t),\quad G_\alpha=\alpha^aG_a. \tag{2.15}$$

From Eqs. (2.8), (2.9), and (2.15) one obtains

$$\delta_bG_a(t)=\frac{1}{2\pi i}\int_c\frac{dt'}{t'-t}[G_b(t'),G_a(t)]. \tag{2.16}$$

Based on Eqs. (2.14) and (2.16) one can calculate the variation commutators $[\delta_\alpha,\delta_\beta]\mathscr{A}(\lambda)$. The methods used here are the extension of those presented in Ref. 7. Substituting Eqs. (2.14) and (2.16) into

$$-\delta_\beta\delta_\alpha\mathscr{A}(\lambda)=\frac{1}{2\pi i}\bar{D}(\xi,\lambda)\int_c\frac{dt}{t-\lambda}\delta_\beta G_\alpha(t)+\frac{1}{2\pi i}\int_c\frac{dt}{t-\lambda}\times\{[\delta_b\mathscr{A}(\lambda),G_\alpha(t)]+[\mathscr{A}(\lambda),\delta_\beta G_\alpha(t)]\}, \tag{2.17}$$

and considering Eq. (2.2), we obtain

$$\begin{aligned}[\delta_\alpha,\delta_\beta]\mathscr{A}(\lambda)=&(D+d(\lambda))\frac{1}{(2\pi i)^2}\int_c\frac{dt}{t-\lambda}\int_{c'}\frac{dt'}{t'-t}[G_\beta(t'),G_\alpha(t)]\\&-\frac{1}{(2\pi i)^2}\int_c\frac{dt}{t-\lambda}\left[(D+d(\lambda))\int_{c'}\frac{dt'}{t'-\lambda}G_\beta(t'),G_\alpha(t)\right]\\&-\frac{1}{(2\pi i)^2}\int_c\frac{dt}{t-\lambda}\int_{c'}\frac{dt'}{t'-\lambda}[[\mathscr{A}(\lambda),G_\beta(t')],G_\alpha(t)]\\&+\frac{1}{(2\pi i)^2}\int_c\frac{dt}{t-\lambda}\int_{c'}\frac{dt'}{t'-t}[\mathscr{A}(\lambda),[G_\beta(t'),G_\alpha(t)]]-(\alpha\leftrightarrow\beta).\end{aligned}$$

Explicitly,

$$\begin{aligned}[\delta_\alpha,\delta_\beta]\mathscr{A}(\lambda)=&\frac{1}{(2\pi i)^2}\int_c dt\left(d(\lambda)\frac{1}{t-\lambda}\right)\int_{c'}\frac{dt'}{t'-t}\{[G_\beta(t'),G_\alpha(t)]-[G_\alpha(t'),G_\beta(t)]\}\\&-\frac{1}{(2\pi i)^2}\int_c\frac{dt}{t-\lambda}\int_{c'}dt'\left(d(\lambda)\frac{1}{t'-\lambda}\right)\{[G_\beta(t'),G_\alpha(t)]-[G_\alpha(t'),G_\beta(t)]\}\\&+\frac{1}{(2\pi i)^2}D\int_c\frac{dt}{t-\lambda}\int_{c'}\frac{dt'}{t'-t}\{[G_\beta(t'),G_\alpha(t)]-[G_\alpha(t'),G_\beta(t)]\}\\&-\frac{1}{(2\pi i)^2}\int_c\frac{dt}{t-\lambda}\int_{c'}\frac{dt'}{t'-\lambda}\{[DG_\beta(t'),G_\alpha(t)]-[DG_\alpha(t'),G_\beta(t)]\}\\&-\frac{1}{(2\pi i)^2}\int_c\frac{dt}{t-\lambda}\int_{c'}\frac{dt'}{t'-\lambda}\{[[\mathscr{A}(\lambda),G_\beta(t')],G_\alpha(t)]-[[\mathscr{A}(\lambda),G_\alpha(t')],G_\beta(t)]\}\\&+\frac{1}{(2\pi i)^2}\int_c\frac{dt}{t-\lambda}\int_{c'}\frac{dt'}{t'-t}\{[\mathscr{A}(\lambda),[G_\beta(t'),G_\alpha(t)]]-[\mathscr{A}(\lambda),[G_\alpha(t'),G_\beta(t)]]\}.\end{aligned} \tag{2.18}$$

Because λ is outside the contours c and c', the second line of Eq. (2.18) is equal to

$$\left(\frac{-1}{2\pi i}\right) d(\lambda) \int_c \frac{dt}{t-\lambda} \int_{c'} \frac{dt'}{t'-\lambda} [G_\beta(t'), G_\alpha(t)],$$

which when combined with the first term of the first line in Eq. (2.18) gives

$$\frac{1}{(2\pi i)^2} d(\lambda) \int_c \frac{dt}{t-\lambda} \int_{c'} \frac{dt'}{t'-t} \{[G_\beta(t'), G_\alpha(t)] - [G_\alpha(t'), G_\beta(t)]\}.$$

Interchanging t and t', the first two lines of Eq. (2.18) contribute

$$\frac{1}{(2\pi i)^2} d(\lambda) \int_c \frac{dt}{t-\lambda} \int_{c_1} \frac{dt'}{t'-t} [G_\beta(t'), G_\alpha(t)],$$

where c_1 is a small contour around the singularity $t = t'$.[3,7] Thus the total contribution received from the first two lines in Eq. (2.18) consists in

$$\frac{1}{2\pi i} d(\lambda) \int_c \frac{dt}{t-\lambda} [G_\beta(t), G_\alpha(t)]. \tag{2.19}$$

Using a similar consideration we can calculate the contributions of singularities appearing at $t' \to t$ as interchanging the contour c with c' for other terms in Eq. (2.18). With the help of the Jacobi identity we finally derive

$$[\delta_\alpha, \delta_\beta] \mathscr{A}(\lambda) = \left[D(\lambda) + d(\lambda) + \mathscr{A}(\lambda), \frac{-1}{2\pi i} \int_c \frac{dt}{t-\lambda} \times [G_\alpha(t), G_\beta(t)] \right]. \tag{2.20}$$

Introducing the covariant derivative

$$\mathscr{D}(\lambda) = \bar{D} + \mathscr{A} = D + d + \mathscr{A}, \tag{2.21}$$

and noting

$$\delta_\alpha \mathscr{A}(\lambda) = \left[\mathscr{D}(\lambda), \left(\frac{-1}{2\pi i}\right) \int_c \frac{dt}{t-\lambda} G_\alpha(t) \right], \tag{2.22}$$

which is obtained from Eq. (2.14), we thus derive, for classical fields,

$$[\delta_\alpha, \delta_\beta] \mathscr{A}(\lambda) = \left[D(\lambda), \left(\frac{-1}{2\pi i}\right) \int_c \frac{dt}{t-\lambda} \times \Phi(t) [V_\alpha(t), V_\beta(t)] \Phi^{-1}(t) \right], \tag{2.23}$$

with

$$\bar{D}(\xi,\lambda) V_\alpha(\xi,\lambda) = 0. \tag{2.24}$$

Equation (2.23) tells us that if $V_\alpha = \alpha^A V_A$ satisfies certain algebraic structure of the group G, this algebraic structure will be reformulated to the algebraic structure of $G \otimes C(\lambda)$ generated by the integrability systems, through Eqs. (2.22) and (2.23).

In other words, there exist loop algebras for any linearization systems based on the infinitesimal RRHT, which is closely related to the analytic properties on the complex λ plane.

It should be noted that Eqs. (2.22) and (2.23) are model independent. The discussions tell us how to generate the algebras of $G \otimes C(\lambda)$ for given algebraic structures determined by $[V_\alpha(t), V_\beta(t)]$ for group G. The term $V_\alpha(t)$ is independent of particular forms of Lax pairs. It has to be chosen to satisfy the requirements of the group and Eq. (2.24). The role that the Lax pair plays here is in "dressing up" G to be $G \otimes C(\lambda)$. For $V_\alpha(\xi,\lambda)$ one can choose

$$V_\alpha(\xi,\lambda) = \alpha^A V_A(\xi,\lambda),$$

where α^A can be a Grassmann number, in such a way that

$$V_A(\xi,\lambda) = (f(\xi,\lambda))^{-m} I_A, \tag{2.25}$$

with

$$\bar{D}(\xi,\lambda) f(\xi,\lambda) = 0 \tag{2.26}$$

and I_A being the matrix basis of Lie algebras or differentiating operators. The variable m is an integer. Let us look at some examples.

A. The (1+1)-dimensional chiral model (including a Wess–Zumino–Witten term)

The equation of motion reads

$$(\lambda - k)\partial_\xi A_\eta + (\lambda + k)\partial_\eta A_\xi = 0. \tag{2.27}$$

In comparison with the notations used in Ref. 8 we have

$$\lambda = 1/2\beta^2, \quad k = n/8\pi$$

for the Lagrangian[8]

$$L = \frac{1}{4\beta^2} \int_S d^2x\, t_r(\partial_\mu g^{-1} \partial_\mu g) - \frac{n}{24\pi} \epsilon^{ijk} \int_B d^3x (g^{-1}\partial_i g g^{-1} \partial_j g g^{-1} \partial_k g).$$

Though the spectral parameter λ appears in the equation of motion, we can introduce the Lax pair

$$\partial_\xi \psi = 1/(\lambda + a) A_\xi \psi, \quad \partial_\eta \psi = 1/(\lambda - a) A_\eta \psi, \tag{2.28}$$

where $a = 1 - k$. It is easy to check by using $A_\xi = g^{-1}\partial_\xi g$ and $A_\eta = g^{-1}\partial_\eta g$ that the integrability condition of Eq. (2.28) gives Eq. (2.27), for any λ and k. As $k = 0$ we have the usual principal chiral model. In this case $\bar{D}(\xi,\lambda) = \partial_\xi$ or ∂_η, and $\mathscr{A}(\xi,\lambda) = 1/(\lambda + a)A_\xi$ or $-1/(\lambda - a)A_\eta$. Taking $f(\lambda) = \lambda$ and I_A as the generators of SU(2) that satisfy Eqs. (2.25) and (2.26), Eq. (2.23) yields

$$[\delta_A^{(m)}, \delta_B^{(n)}] A_u = C_{AB}^C \delta_C^{(m+n)} A_u, \quad u = \xi \text{ or } \eta. \tag{2.29}$$

B. The super chiral model (with a Wess–Zumino–Witten term)

The equation of motion in the case is[9]

$$(\lambda - k) D_1 A_2 - (\lambda + k) D_2 A_1 = 0, \tag{2.30}$$

where $D_\alpha = \partial/\partial\bar\theta^\alpha + i(\gamma^\mu\theta)_\alpha \partial_\mu$, $k/\lambda = a$, using the notation in Ref. 9.

The Lax pair can be taken as

$$\begin{aligned} D_1\Phi &= -1/[\lambda + (1-k)] A_1 \Phi, \\ D_2\Phi &= 1/[\lambda - (1-k)] A_2 \Phi. \end{aligned} \tag{2.31}$$

After manipulation the same algebra as shown by Eq. (2.29) is derived for the model. The simplicity is in choosing non-Grassmann variations for the model.

For $k = 0$ it returns to the super principal chiral model.

C. The SDYM field[5]

The Lax pair

$$
\begin{aligned}
&(\lambda\,\partial_z + \partial_{\bar y} + \lambda J^{-1}J_z)\Phi = 0,\\
&(\lambda\,\partial_y - \partial_{\bar z} + \lambda J^{-1}J_y)\Phi = 0
\end{aligned}
\tag{2.32}
$$

gives the SDYM equation

$$(J^{-1}J_y)_{\bar y} + (J^{-1}J_z)_{\bar z} = 0. \tag{2.33}$$

By choosing the I_A as the generators of SL(N,C) or some other appropriate group and

$$f(\xi,\lambda) = f(\bar y - (1/\lambda)z, \bar z + (1/\lambda)y), \tag{2.34}$$

we derive the loop algebra

$$[\delta_A^{(m)},\delta_B^{(n)}]B_u = C_{AB}^{C}\delta^{(m+n)}B_u, \tag{2.35}$$

where $u = z$ or y, and C_{AB}^{C} is the structure constant for the group.

D. BZ gravity[6]

For the element

$$-ds^2 = f(z,t)(-dt^2 + dz^2) + g_{ab}(z,t)dx^a\,dx^b, \tag{2.36}$$

where $a,b = 1,2$, the determinant of the 2×2 matrix $\|g_{ab}\|$ is

$$\det g = \alpha^2. \tag{2.37}$$

The Einstein vacuum equations can be decoupled into two groups. One of them reads

$$(\alpha g_\xi g^{-1})_\eta + (\alpha g_\eta g^{-1})_\xi = 0. \tag{2.38}$$

Here the light-cone coordinates $\zeta = z + t$ and $\eta = z - t$ have been used. The subindices ζ and η denote the derivatives. It is easy to see that

$$\alpha_{\xi\eta} = 0, \tag{2.39}$$

which has the solutions

$$\alpha = a(\zeta) + b(\eta), \quad \beta = a(\zeta) - b(\eta). \tag{2.40}$$

The Lax pair for Eq. (2.38) has the form

$$
\begin{aligned}
&\left[\left(\partial_\zeta - \frac{2\alpha_\zeta\lambda}{\lambda-\alpha}\partial_\lambda\right) + \frac{A}{\lambda-\alpha}\right]\Phi = 0,\\
&\left[\left(\partial_\eta + \frac{2\alpha_\eta\lambda}{\lambda+\alpha}\partial_\lambda\right) + \frac{B}{\lambda+\alpha}\right]\Phi = 0,
\end{aligned}
\tag{2.41}
$$

with

$$A = \alpha g_\zeta g^{-1}, \quad B = -\alpha g_\eta g^{-1}, \tag{2.42}$$

where α,g are dependent on ζ and η.

In accordance with Eqs. (2.25) and (2.26) one takes[6]

$$f(\xi,\lambda) = (\alpha^2/\lambda + 2\beta + \lambda) \tag{2.43}$$

and the I_A as the generators of SL$(2,R)$. The proof of the existence of a loop algebra for BZ gravity by other methods is not easy.[10]

E. The constraint equations for extended self-dual supersymmetric Yang–Mills (ESYM)

In four dimensions the linearization system under a special gauge can be written in the form[11]

$$\mathscr{D}^s\Phi \equiv (D_1^s + \lambda D_2^s + \lambda B D_2^s B^{-1})\Phi = 0, \tag{2.44a}$$

$$\bar{\mathscr{D}}_t\Phi \equiv (\bar D_{\dot 2 t} + \lambda'\bar D_{\dot 1 t} + B\bar D_{\dot 2 t}B^{-1})\Phi = 0, \tag{2.44b}$$

$$
\begin{aligned}
(1/2\dot{\mathrm{i}})\mathscr{P}\Phi \equiv [&\partial_{1\dot 2} + g\nabla_{1\dot 2}g^{-1} + \lambda(\partial_{2\dot 1} + B\,\partial_{2\dot 2}B^{-1})\\
&+ \lambda'\,\partial_{1\dot 1} + \lambda'\lambda(\partial_{2\dot 1} + g\nabla_{2\dot 1}g^{-1})]\Phi = 0,
\end{aligned}
\tag{2.44c}
$$

where

$$
\begin{aligned}
&D_\alpha^s \equiv \frac{\partial}{\partial\theta_s^\alpha} + i(\sigma^\mu)_{\alpha\dot\alpha}\bar\theta^{\dot\alpha s}\,\partial_\mu \quad (\mu = 0,1,2,3),\\
&\bar D_{\dot\alpha s} \equiv -\frac{\partial}{\partial\bar\theta^{\dot\alpha s}} - i\theta_s^\alpha(\sigma^\mu)_{\alpha\dot\alpha}\,\partial_\mu,\\
&\nabla_{\alpha\dot\beta} \equiv (\sigma^\mu)_{\alpha\dot\beta}(\partial_\mu + A_\mu),
\end{aligned}
\tag{2.45}
$$

and λ' and λ are independent spectral parameters. The notation used in Eqs. (2.44) and (2.45) is as given in Ref. 11. To consider the loop algebra for the full ESYM we must consider Riemann–Hilbert transforms with two complex variables,[12] which we are not ready to discuss now. Here we shall only consider the self-dual (or the anti-self-dual) ESYM, i.e., only Eq. (2.44a) alone [or Eq. (2.44b) alone] will be considered, and Φ is a function of coordinate and one complex spectral parameter. [The other equations of Eq. (2.44) are trivially satisfied in this special self-dual or anti-self-dual situation.] To generate the loop algebras the scalar function $f(\xi,\lambda)$ must satisfy Eq. (2.26), i.e.,

$$\mathscr{D}^s f(\xi,\lambda) = \bar{\mathscr{D}}_t f(\xi,\lambda) = \mathscr{P} f(\xi,\lambda) = 0. \tag{2.46}$$

The ESYM is especially interesting because of the large symmetry involved. As discussed above, for a given algebra formed by $[V_A,V_B]$ the linearization system can generate the corresponding loop algebras through the Lax pair. Now for four-dimensional supersymmetric theories we can choose the graded algebra of extended supersymmetry as the "starting" one. The generators are Q_α, $\bar Q_\beta$, D, S_α, K, P_μ, $M_{\mu\nu}$, and Π as shown in Ref. 13 for the case without internal symmetry. Such a choice implies that a subalgebra of extended supersymmetric theory has been chosen to be the starting algebra. Based on the commutation and anticommutation relations as shown in Ref. 13,

$$
\begin{aligned}
&[Q_\alpha,D] = \tfrac{1}{2}iQ_\alpha, \quad [S_\alpha,D] = -\tfrac{1}{2}iS_\alpha,\\
&[Q_\alpha,M_{\mu\nu}] = i(\sigma_{\mu\nu}Q)_\alpha, \quad [S_\alpha,M_{\mu\nu}] = i(\sigma_{\mu\nu}S)_\alpha,\\
&[Q_\alpha,P_\mu] = 0, \quad [S_\alpha,K_\mu] = 0,\\
&[Q_\alpha,K_\mu] = -i(\gamma_\mu S)_\alpha, \quad [S_\alpha,P_\mu] = i(\gamma_\mu Q)_\alpha,\\
&[Q_\alpha,\Pi] = -\tfrac{3}{2}i(\gamma_5 Q)_\alpha, \quad [S_\alpha,\Pi] = \tfrac{3}{2}i(\gamma_5 S)_\alpha,
\end{aligned}
$$

we have

$$
\begin{aligned}
&[P_\mu,M_{\rho\sigma}] = \delta_{\mu\rho}P_\sigma - \delta_{\mu\sigma}P_\rho,\\
&[M_{\mu\nu},M_{\rho\sigma}] = \delta_{\mu\sigma}M_{\nu\rho} + \delta_{\nu\rho}M_{\mu\sigma} - \delta_{\mu\rho}M_{\nu\sigma} - \delta_{\nu\sigma}M_{\mu\rho},
\end{aligned}
\tag{2.47}
$$

and

$$
\begin{aligned}
&\{Q_\alpha,\bar Q_\beta\} = -2\gamma_{\alpha\beta}^{\mu}P_\mu,\\
&\{S_\alpha,\bar S_\beta\} = 2\gamma_{\alpha\beta}^{\mu}K_\mu,\\
&\{Q_\alpha,\bar S_\beta\} = 2(\sigma^{\mu\nu}M_{\mu\nu} - D + 2\gamma_5\Pi)_{\alpha\beta},
\end{aligned}
$$

we define the following variations:

$$\delta_\alpha^{(m)}\mathscr{A}(\lambda)=\alpha^A\delta_A^{(m)}\mathscr{A}(\lambda)$$
$$=\Big[\mathscr{D}(\lambda),\frac{-1}{2\pi i}\int_c\frac{dt}{t-\lambda}\times f^{-m}(t,...)\alpha^A\Phi(t)I_A\Phi^{-1}(t)\Big], \quad (2.48)$$

where Φ satisfies the linearization system and the α^A's are infinitesimal constants (perhaps Grassmann numbers).

The α^A's are the usual infinitesimal constants when the I_A operators are D, P_μ, $M_{\mu\nu}$, Π, and K_μ, or Grassmann infinitesimal constants when the I_A's are Q_α, $\overline{Q}_\beta$, S_α, and $\overline{S}_\beta$.

We define

$$\begin{aligned}&I(D)=D,\quad I_\mu(p)=P_\mu,\quad I_{\mu\nu}^{(m)}=M_{\mu\nu},\\&I_\mu(K)=K_\mu,\quad I(\Pi)=\Pi,\end{aligned}\quad (2.49)$$

and $I_\alpha(Q)=Q_\alpha$, $I_\beta(\overline{Q})=\overline{Q}_\beta$, $I_\alpha(s)=s_\alpha$, $I_\beta(\overline{s})=\overline{s}_\beta$. By Eq. (2.48) we can obtain the corresponding variations related to Eq. (2.49). The general conclusion of Eq. (2.23a) gives the loop algebras corresponding to Eq. (2.49) for $A=BD^s_{\frac{1}{2}}B^{-1}$ or $B\overline{D}_{\frac{1}{2}t}B^{-1}$, as follows:

$$\begin{aligned}
&[\delta_\alpha^{(m)}(Q),\delta^{(n)}(D)]=\tfrac{1}{2}i\delta_\alpha^{(m+n)}(Q),\\
&[\delta_\alpha^{(m)}(Q),\delta_{\mu\nu}^{(n)}(M)]=i(\sigma_{\mu\nu}\delta^{(m+n)}(Q))_\alpha,\\
&[\delta_\alpha^{(m)}(S),\delta^{(n)}(D)]=-\tfrac{1}{2}i\delta_\alpha^{(m+n)}(S),\\
&[\delta_\alpha^{(m)}(S),\delta_{\mu\nu}^{(n)}(M)]=i(\sigma_{\mu\nu}\delta^{(m+n)}(S))_\alpha,\\
&[\delta_\alpha^{(m)}(Q),\delta_\mu^{(n)}(K)]=-i(\gamma_\mu\delta^{(m\,1\,n)}(S))_\alpha,\\
&[\delta_\alpha^{(m)}(Q),\delta^{(n)}(\Pi)]=-\tfrac{3}{2}i(\gamma_5\delta^{(m+n)}(Q))_\alpha,\\
&[\delta_\alpha^{(m)}(S),\delta^{(n)}(\Pi)]=\tfrac{3}{2}i(\gamma_5\delta^{(m+n)}(S))^\alpha,\\
&[\delta_\alpha^{(m)}(S),\delta_\mu^{(n)}(P)]=i(\gamma_\mu\delta^{(m+n)}(Q))_\alpha,\\
&[\delta_\alpha^{(m)}(Q),\delta_\mu^{(n)}(P)]=[\delta_\alpha^{(m)}(S),\delta_\mu^{(n)}(K)]=0,
\end{aligned}\quad (2.50)$$

and

$$\begin{aligned}
&\{\delta_\alpha^{(m)}(Q),\delta_\beta^{(n)}(\overline{Q})\}=-2\gamma^\mu_{\alpha\beta}\delta_\mu^{(m+n)}(P),\\
&\{\delta_\alpha^{(m)}(S),\delta_\beta^{(n)}(\overline{S})\}=2\gamma^\mu_{\alpha\beta}\delta_\mu^{(m+n)}(K),\\
&\{\delta_\alpha^{(m)}(Q),\delta_\beta^{(n)}(\overline{S})\}=2\{\sigma^{\mu\nu}\delta_{\mu\nu}^{(m+n)}(M)-\delta^{(m+n)}(D)\\
&\qquad+2\gamma_5\delta^{(m+n)}(\Pi)\}_{\alpha\beta},
\end{aligned}$$

where the variations acting on A, were not written out.

For the case of Eq. (2.44b) alone, we simply replace α by $\dot{\alpha}$ in Eqs. (2.49) and (2.50).

It remains a challenge to develop Riemann–Hilbert transforms in two complex variables for the full ESYM.

F. The Liouville equation

The Lax pair reads

$$\begin{aligned}&\partial_S\Phi=\frac{1}{\lambda}e^{2\phi}\begin{pmatrix}0&1\\0&0\end{pmatrix}\Phi,\\&\partial_\eta\Phi=\Big[\lambda\begin{pmatrix}0&0\\1&0\end{pmatrix}+\partial_\eta\phi\begin{pmatrix}1&0\\0&-1\end{pmatrix}\Big]\Phi.\end{aligned}\quad (2.51)$$

Correspondingly, $\mathscr{A}_\xi$ is taken to be

$$\frac{-1}{\lambda}e^{2\phi}\begin{pmatrix}0&1\\0&0\end{pmatrix},$$

and

$$A_\eta=-\Big[\lambda\begin{pmatrix}0&0\\1&0\end{pmatrix}+\partial_\eta\phi\begin{pmatrix}1&0\\0&-1\end{pmatrix}\Big].$$

Noting that

$$\sigma_3\Phi(\lambda)\sigma_3=\Phi(-\lambda), \quad (2.52)$$

Eqs. (2.22) and (2.23) give the result of Ref. 13. The same result holds for the sine–Gordon equation. As was pointed out in Ref. 13, the "broken loop algebra" will appear due to the reduction symmetry (2.52) for the Liouville equation and the sine–Gordon equation.

III. QUANTUM EFFECTS

Finally, we discuss the nonclassical aspect of the RRHT in our case. Because

$$\begin{aligned}&[\delta_\alpha,\delta_\beta]\mathscr{A}(x)\\&=-\overline{D}(\lambda)\frac{1}{2\pi i}\int_c\frac{dt}{t-\lambda}[G_\alpha(t),G_\beta(t)]\\&\quad-\frac{1}{2\pi i}\int_c\frac{dt}{t-\lambda}[\mathscr{A}(\lambda),[G_\alpha(t),G_\beta(t)]],\end{aligned}\quad (3.1)$$

when we change $G_\alpha(t)=\Phi v_\alpha\Phi^{-1}$ to

$$:G_\alpha(t):\,=\,:\Phi v_\alpha\Phi^{-1}:, \quad (3.2)$$

the commutator $[G_\alpha(t),G_\beta(t)]$ is no longer equal to $\Phi[V_\alpha,V_\beta]\Phi^{-1}$. The additional term appears due to the normal ordered product. As discussed in Ref. 14, this will give rise to an additional term in the algebras, which is the central extension of the algebra.

IV. CONCLUSION

We have developed a general formulation based on the infinitesimal RRHT to find the Kac–Moody algebra, for the linearization system with spectral parameter. This approach is general and model independent. In deriving the algebra only the linearization equation and the analyticity in the complex λ plane of the system are needed. From the known algebraic structures related to the group G, the infinitesimal RRHT generates the corresponding Kac–Moody algebra.

ACKNOWLEDGMENTS

We would like to thank Professor Bo-Yu Hou for enlightening discussions.

This work is partially supported by the U.S. Department of Energy.

[1]For the reviews see L.-L. Chau, *Nonlinear Phenomena, Lecture Notes in Physics*, Vol. 189, edited by K. B. Wolf (Springer, Berlin, 1983); L. Dolan, Phys. Rep. **109** (1984); J. Lepowsky, *Vertex Operators in Mathematics and Physics*, edited by J. Lepowsky, S. Mandelstam, and I. M. Singer (Springer, Berlin, 1984).

[2]For loop algebra in chiral models see L. Dolan, Phys. Rev. Lett. **47**, 1371 (1981); C. Davchand and D. B. Fairle, Nucl. Phys. B **194**, 232 (1982); M.-L. Ge and Y.-S. Wu, Phys. Lett. B **108**, 411 (1982). For loop algebra in self-dual Yang–Mills equations, see L.-L. Chau and Y.-S. Wu, Phys. Rev. D **26**, 3581 (1983); L.-L. Chau, M.-L. Ge, A. Sinha, and Y.-S. Wu, Phys. Lett. B **121**, 391 (1983). For the connection with Yang–Baxter algebras see H. J. de Vega, plenary talk given at The 14th International Conference on Group Theoretical Method in Physics, Seoul, 1985 and the references therein. For the conserved currents for the chiral model see M. C. Davies, P. J. Houston, J. M. Leinaas, and A. J. Macfarlane, Phys. Lett. B **119**, 187 (1982); M. Gomes and Y. K. Ha, Phys. Rev. D. **28**, 2683 (1983); H. J. de Vega, H. Eichenherr, and J. M. Maillet, Phys. Lett. B **132**, 337 (1983).

[3]Y. Nakamura and K. Ueno, Phys. Lett. B **107**, 273 (1982); Y.-S. Wu, Commun. Math. Phys. **90**, 461 (1983); L.-L. Chau, M.-L. Ge, and Z. Dopowicz, Phys. Rev. Lett. **52**, 1940 (1984).

[4]See L.-L. Chau, "Geometrical integrability and equations of motion in physics: A unifying view," talk at the Nankai Workshop on Geometrical Integrability, 18–22 August 1987, Nankai Institute of Mathematics, Nankai University, Tianjin, China; to appear in proceedings.

[5]L.-L. Chau, talks at the 8th Winter School of Theoretical Physics, Karpacz, Poland, 1981.

[6]For discussions on RRHT for the Ernst equation and BT in BZ gravity see C. M. Cosgrov, J. Math. Phys. **23**, 615 (1982); V. A. Belinski and V. E. Zakharov, Sov. Phys. JETP **48**, 985 (1978); **50**, 1 (1979).

[7]H. Eichenherr, Phys. Lett. B **115**, 375 (1982).

[8]E. Witten, Commun. Math. Phys. **92**, 455 (1984).

[9]L.-L. Chau and H. C. Yen, Phys. Lett. B **177**, 368 (1986).

[10]Y.-S. Wu and M.-L. Ge, J. Math. Phys. **24**, 1187 (1983).

[11]E. Witten, Phys. Lett. B **77**, 394 (1978); M. Sohnius, Nucl. Phys. B **136**, 461 (1978); I. V. Volovich, Lett. Math. Phys. **7**, 517 (1983); C. Devchand, Nucl. Phys. B **238**, 333 (1984); L.-L. Chau, M.-L. Ge, and C. S. Lim, Phys. Rev. D **33**, 1056 (1986).

[12]L.-L. Chau and H.-C. Yen (unpublished); J. Tafel, J. Math. Phys. **28**, 240 (1986).

[13]P. Fayet and S. Ferrara, Phys. Rep. **32**, 249 (1977).

[14]H. Bohr, M.-L. Ge, and J. V. Volovich, Nucl. Phys. B **260**, 701 (1985).

[15]G. Bbattacharya, L.-L. Chau, *Lecture Notes in Physics*, Vol. 201, edited by G. Denardo (Springer, Berlin, 1984), p. 153.

Curvature form and solutions of nonlinear models

K. Tanaka
Department of Physics, The Ohio State University, Columbus, Ohio 43210

(Received 17 May 1988; accepted for publication 31 August 1988)

Several equations of motions of bosonic two-dimensional nonlinear models are noted that emerge from the requirement that the curvature two-form $\Omega = 0$. This information is used to obtain solutions of these equations.

I. INTRODUCTION

It was noted that there is a connection between the nonlinear two-dimensional models and the scalar curvature r. In particular, when the r is set equal to $r = -2$, the equations of motion of the Liouville and the sine–Gordon models were obtained.[1] This method of generating equations of motion is limited because the metric g_{AB} in two dimensions has three independent components. On the other hand, the curvature two-form Ω obtained from the affine connection Γ has in general six independent parameters, so when one takes $\Omega = 0$, additional equations of motion such as the Korteweg–de Vries (KdV) equation and modified Korteweg–de Vries (MKdV) equation were obtained.[2] We are interested in obtaining equations of motion from $\Omega = 0$, because the information we gain in the process may be useful in obtaining their solutions.

The popular method of solution[3] for certain nonlinear differential equations is to set up the linear scattering problem in the x variable, choose the time dependence of eigenfunctions, solve at $t = 0$, and then determine the solution at later times from the scattering data. The two first-order equations are

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix}_x = \begin{pmatrix} \eta & q \\ r & -\eta \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \equiv -\Gamma^R_{S1} V, \tag{1}$$

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix}_t = \begin{pmatrix} A & B \\ C & -A \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \equiv -\Gamma^R_{S0} V, \tag{2}$$

where η is the eigenvalue, all the quantities are functions of x and t, and the subscripts are partial derivatives.

From $V_{xt} = V_{tx}$ and the requirement η is time independent, $\eta_t = 0$, we obtain

$$\begin{aligned} &-A_x + qC - rB = 0, \\ &q_t - B_x + 2\eta B - 2Aq = 0, \\ &r_t - C_x - 2\eta C + 2Ar = 0. \end{aligned} \tag{3}$$

One expands A, B, and C in terms of η and solves the equations.

Equations (1) and (2) can be written as

$$\frac{\partial V^R}{\partial x^m} + \Gamma^R_{Sm} V^s = 0, \quad m = 0,1, \qquad R,S = 1,2, \quad x^0 = t, \quad x^1 = x, \tag{4}$$

that is, vanishing of the covariant derivative of V, where Γ^R_{Sm} are components of the affine connection. The $\Gamma^R_S = \Gamma^R_{Sm}\, dx^m$ is the one-form with values in SL$(2,R)$.

The curvature two-form is

$$\Omega^R_S = d\Gamma^R_S + \Gamma^R_T \wedge \Gamma^T_S, \quad \Gamma = \theta_\alpha X_\alpha, \quad \alpha = 1,2,3, \tag{5}$$

where

$$X_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad X_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

It was shown from (1), (2), and (5) (Ref. 2),

$$\begin{aligned} \Omega = dx \wedge dt\{ &(-A_x + qC - rB) X_1 \\ &+ (q_t - B_x + 2\eta B - 2qA) X_2 \\ &+ (r_t - C_x - 2\eta C + 2rA) X_3 \}. \end{aligned} \tag{6}$$

Therefore, the condition $\Omega = 0$ is equivalent to the requirement that the eigenvalue of η is time independent as given in Eq. (3). The following choices of the components of the affine connections yield the equations of motion of various models considered here. There does not seem to be a systematic way of formulating a given nonlinear equation into the condition $\Omega = 0$. The structure of the coefficients of X_1, X_2, and X_3 are quite restrictive.

(a) SG (sine–Gordon): $u_{xt} = \sin u$,

$$\begin{aligned} -\Gamma^R_{S1} &= \begin{pmatrix} \eta & -\frac{1}{2}u_x \\ \frac{1}{2}u_x & -\eta \end{pmatrix}, \\ -\Gamma^R_{S0} &= \frac{1}{4\eta} \begin{pmatrix} \cos u & \sin u \\ \sin u & -\cos u \end{pmatrix}. \end{aligned} \tag{7}$$

The substitution of components of (7) in (6) gives

$$\Omega = dx \wedge dt\, \tfrac{1}{2}(u_{xt} - \sin u) \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{8}$$

(b) MKdV: $u_t + 6u^2 u_x + u_{xxx} = 0$,

$$\begin{aligned} -\Gamma^R_{S1} &= \begin{pmatrix} \eta & u \\ -u & -\eta \end{pmatrix}, \\ -\Gamma^R_{S0} &= \begin{pmatrix} -4\eta^3 - 2\eta u^2 & -u_{xx} - 2\eta u_x - 4\eta^2 u - 2u^3 \\ u_{xx} - 2\eta u_x + 4\eta^2 u + 2u^3 & 4\eta^3 + 2\eta u^2 \end{pmatrix}. \end{aligned} \tag{9}$$

(c) NLS [nonlinear Schrödinger equation (Ref. 4)]: $iu_t + u_{xx} + 2|u|^2 u = 0$,

$$-\Gamma_{S1}^R = \begin{pmatrix} \eta & u \\ -u^* & -\eta \end{pmatrix}, \qquad -\Gamma_{S0}^R = \begin{pmatrix} 2i\eta^2 + i|u|^2 & iu_x + 2i\eta u \\ iu_x^* - 2i\eta u^* & -2i\eta - i|u|^2 \end{pmatrix}. \tag{10}$$

(d) ϕ^4 model: $4u_{xt} + u^3 - u = 0$,

$$-\Gamma_{S1}^R = \begin{pmatrix} 0 & -u_x \\ u_x & 0 \end{pmatrix},$$

$$-\Gamma_{S0}^R = \begin{pmatrix} 0 & \frac{1}{4}\int_0^x (u^3 - u)dx' \\ -\frac{1}{4}\int_0^x (u^3 - u)dx' & 0 \end{pmatrix}. \tag{11}$$

Change variables to

$$t = T + X, \qquad x = T - X, \qquad 4\,\partial_{tx} u = (\partial_{TT} - \partial_{XX})u, \tag{12}$$

and obtain the equation of motion[5]

$$u_{TT} - u_{XX} + u^3 - u = 0.$$

(e) KdV: $u_t + 6uu_x + u_{xxx} = 0$,

$$-\Gamma_{S1}^R = \begin{pmatrix} \eta & u \\ -1 & -\eta \end{pmatrix}, \qquad -\Gamma_{S0}^R = \begin{pmatrix} -4\eta^3 - 2\eta u - u_x & -u_{xx} - 2\eta u_x - 4\eta^2 u - 2u^2 \\ 4\eta^2 + 2u & 4\eta^3 + 2\eta u + u_x \end{pmatrix}. \tag{13}$$

(f) Liouville equation: $\rho_{xt} - \frac{1}{2}e^\rho = 0$,

$$-\Gamma_{S1}^R = \begin{pmatrix} 0 & \rho_x \\ -\rho_x & 0 \end{pmatrix}, \qquad -\Gamma_{S0}^R = \begin{pmatrix} 0 & \frac{1}{2}\int_0^x e^\rho\, dx' \\ -\frac{1}{2}\int_0^x e^\rho\, dx' & 0 \end{pmatrix}. \tag{14}$$

II. SOLUTIONS OF THE NONLINEAR MODELS

In order to solve the equations for the models (a)–(d), one notes that all the terms are of odd powers in the amplitude u, or q of Γ_{S1}^R because $q = u$ or $q \sim u_x$. This means that if we write the solution q in terms of sech (hyperbolic secant) the equations are expressible in terms of odd powers of sech with the aid of $\tanh^2\theta = 1 - \mathrm{sech}^2\theta$. One then needs to match the coefficients of the odd powers of sech to satisfy the equations of motions. The sign of the argument θ is arbitrary. Equations (1) and (2) suggest $v_i = e^{\pm\eta x}$ and $v_i = e^{\pm a_i t}$ $(i = 1,2)$, respectively, where a_1 is the real part of the constant term A_c of A. So a reasonable choice for the argument is $\theta = -\eta x - a_1 t$. The imaginary part of A_c will be put in an exponential factor that multiplies the sech. The argument is 2θ because of the definition of the constant part of A,

$$A_c = a_1 + ia_2. \tag{15}$$

We thus start with a form of solution[6]

$$q = N\,\mathrm{sech}\, 2\theta e^{2ia_2 t}, \tag{16}$$

$$\theta = -\eta x - a_1 t. \tag{17}$$

(a) SG: Expand $\cos u$ in A of (2) and (7) and obtain from (1), (7), (15), (16), and (17),

$$a_1 = \frac{1}{4\eta}, \quad a_2 = 0, \quad u = -2\int q\, dx,$$

$$q = N\,\mathrm{sech}(-2\eta x - (1/2\eta)t),$$

$$u = -2N\int \mathrm{sech}\left(-2\eta x - \frac{1}{2\eta}t\right)dx = \frac{4N}{2\eta}\tan^{-1} e^{-2\eta x - (1/2\eta)t}.$$

The sine–Gordon equation is satisfied when $N = 2\eta$, so the solution is

$$u = 4\tan^{-1} e^{2\eta x - (1/2\eta)t}. \tag{18}$$

(b) MKdV: From (9), (15), (16), and (17), we get

$$a_1 = -4\eta^3, \quad a_2 = 0, \qquad \theta = -\eta x + 4\eta^3 t, \qquad u = q = N\,\mathrm{sech}(-2\eta x + 8\eta^3 t), \tag{19}$$

and upon substitution in the MKdV equation, obtain the solution

$$u = 2\eta\,\mathrm{sech}(-2\eta x + 8\eta^3 t). \tag{20}$$

(c) NLS: From (10), (15), (16), and (17), we get

$$a_1 = 0, \quad a_2 = 2\eta^2,$$

and

$$u = q = 2\eta\,\mathrm{sech}(-2\eta x)e^{4i\eta^2 t} \tag{21}$$

as a solution.

(d) ϕ^4 model: There is no information on A_c so we put

$$u = N\,\mathrm{sech}(-2\eta x - 2at)$$

and substitute in (12) which is written as

$$u_{tt} - u_{tt} + u^3 - u = 0 \tag{22}$$

and obtain

$$N(4a^2 - 4\eta^2) - N = 0, \qquad -2N(4a^2 - 4\eta^2) + N^3 = 0. \tag{23}$$

The solution of (23) is

$$N = 1, \quad a = \tfrac{1}{2}\sqrt{4\eta^2 + 1},$$

which gives the solution of (22)

$$u = \operatorname{sech}(-2\eta x - \sqrt{4\eta^2 + 1}\, t). \tag{24}$$

A related equation that has important physical applications is[5]

$$u_{tt} - u_{xx} + u - \tfrac{1}{6}u^3 = 0. \tag{25}$$

The solution of (25) is found to be

$$u = 2\sqrt{3}\, \operatorname{sech}(-2\eta x - \sqrt{4\eta^2 - 1}\, t). \tag{26}$$

(e) KdV: This equation does not consist of odd powers of u but a solution can be obtained from a solution of MKdV.

From (13), (16), and (17)

$$a_1 = -4\eta^3, \quad a_2 = 0,$$

$$\theta = -\eta x + 4\eta^3 t,$$

and we obtain the MKdV solution

$$q = 2\eta\, \operatorname{sech}(-2\eta x + 8\eta^3 t).$$

The Miura transformation[7]

$$u = q^2 + iq_x \tag{27}$$

gives a solution u of the KdV equation

$$u = 4\eta^2(\operatorname{sech}^2 2\theta + i \operatorname{sech} 2\theta \tanh 2\theta). \tag{28}$$

(f) Liouville equation: $\rho_{xt} - \frac{1}{2}e^{\rho} = 0$.

This example does not have a solution of the form (16) but is included for completeness. Change the variables

$$t = T + iX,$$

$$x - T - iX,$$

and obtain

$$\tfrac{1}{2}(\rho_{TT} + \rho_{XX}) = e^{\rho}$$

or

$$\rho_{TT} + \rho_{XX} = e^{2\rho}. \tag{29}$$

A solution of (29) is

$$\rho = \ln[2/(1 - T^2 - X^2)]. \tag{30}$$

The only model that has a breatherlike solution (a real solution confined in x, periodic in t) is the NLS model. The solutions (18), (20), (24), (26), and (28) are soliton solutions. They vanish at $x = \pm\infty$, $t = \pm\infty$ as $\operatorname{sech} 2\theta = (\cosh 2\theta)^{-1}$. As x and t increases the amplitude decreases, which means that energy is radiated away, or converted into other modes.

We found it interesting to obtain the equations of motion via $\Omega = 0$, rather than solve Eq. (3) directly, because we expect there are a variety of ways of solving the nonlinear equations once it is formulated in terms of $\Omega = 0$.

ACKNOWLEDGMENTS

The author thanks V. K. Mishra for discussions.

This work was supported in part by the U.S. Department of Energy under Contract No. EY-76-C-02-1415*00.

[1] V. K. Mishra and K. Tanaka, Phys. Rev. D **36**, 3722 (1987); L. Dolan, *ibid.* **15**, 2337 (1977).

[2] M. Crampin, F. A. E. Pirani, and D. C. Robinson, Lett. Math. Phys. **2**, 15 (1977).

[3] M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Phys. Rev. Lett. **31**, 125 (1973).

[4] V. E. Zakharov and A. B. Shabat, Sov. Phys. JETP **34**, 62 (1972); K. Konno and M. Wadati, Prog. Theor. Phys. **53**, 1652 (1975).

[5] Y. S. Kirshar and B. A. Malomed, Phys. Rev. Lett. **60**, 164 (1988).

[6] D. J. Kaup and A. C. Newell, Proc. R. Soc. London Ser. A **36**, 413 (1978).

[7] R. Sasaki and J. Yamanaka, in Proceedings of the "Confirmed field theory and solvable lattice models," Nagoya University, 1987; R. M. Miura, J. Math. Phys. **9**, 1202 (1968).

The Φ^4 equations of motion. II. The zero-, one-, and two-dimensional solutions

Marietta Manolessou
Institut des Hautes Etudes Scientifiques, 35, route de Chartres, 91440 Bures-sur-Yvette, France

(Received 15 December 1987; accepted for publication 25 May 1988)

This is the second paper of the series concerning the solution of the system of Φ^4 equations of motion for the Schwinger functions by a fixed-point method. These works constitute a part of a general program towards the construction of a Φ_4^4 Wightman quantum field theory (QFT). In the previous paper [J. Math. Phys. **29**, 2092 (1988)] (Paper I), a general outline of the program has been presented. Moreover, the "nice" properties of *signs*, *splitting*, and *norms* revealed "experimentally" by the Φ *iteration* have been analyzed. We have shown how this iterative procedure converges to the solution (if the coupling constant is fixed positive and smaller than a finite value) thanks to the conservation of these properties, which constitute in fact a complete system of self-consistent conditions. Taking into account this information, in the present paper, the answer to the zero-, one-, and two-dimensional problems is given. To be precise, the fixed-point method is constructed by formulating the properties of signs and splitting of the Φ iteration in terms of particular subsets $\Phi_{0\Lambda} \subset \mathscr{B}_0$ (zero dimensions) and $\Phi_\Lambda \subset \mathscr{B}$ (one and two dimensions) of the appropriated Banach spaces $\mathscr{B}_0$ and $\mathscr{B}$, defined exactly with the norms provided by the Φ iteration. The basic ingredients, both introduced already in I, are, on the one hand, the bounded positive sequences of the *splitting constants* and, on the other hand, the *sweeping factors*, which carry all the combinatorial information for the global terms. Their absolute and relative bounds yield the stability of the corresponding subsets and the conservation of the norms. *For the zero-dimensional problem* a simpler equivalent system (nonlinear map) of equations in the space of the splitting sequences is solved by ensuring the existence of the solution (resp. contractivity) inside the corresponding subset when the coupling constant Λ satisfies $0 < \Lambda \lesssim 0.1$ (resp. $0 < \Lambda \lesssim 0.01$). *For the two- (or one-) dimensional problem* it is shown that when $0 < \Lambda \lesssim 0.006$, the subset Φ_Λ is stable under the nonlinear mapping (represented by the equations of motion), which in turn is contractive inside Φ_Λ under weaker conditions on Λ. By this last result the convergence of the Φ iteration to the unique fixed point (presented in I) is reobtained in a direct way.

I. INTRODUCTION

This is the second paper (called Paper II) of the work concerning the solution of the equations of motion for a Φ_r^4 model ($0 \leqslant r \leqslant 2$) satisfied by the completely amputated connected Green's functions in the Euclidean momentum space (Schwinger functions).

The ultimate goal of our program, begun some years ago (and already with published results[1–3]), is to introduce a new method for the construction of a nontrivial Wightman field theory Φ_4^4. In Paper I[3] we have presented an extensive outline of the method and analyzed the following most important steps for its realization: (1) the definition of the renormalized normal product, the finite equations of motion for a Φ_4^4 model, and conservation by them of the linear axiomatic field theory properties; (2a) the zero-dimensional problem; (2b) the two-dimensional problem (which trivially includes the one-dimensional problem); (3) the four-dimensional problem (which also includes the three-dimensional problem); and (4) the conservation of Osterwalder–Schrader axioms[4] by the four-dimensional solution.

Step (1) has already been published.[1,2] In the present article we give the proofs of the fundamental statements in order to solve (2a) and (2b). In other words, we find the solutions of the zero- and two-dimensional systems of equations of motion, as we already announced in Paper I [cf. Eqs. (3.1) and (2.4) of Ref. 3].

In I we explained in detail how we collected crucial "experimental" results at zero external momenta by the Φ iteration, i.e., starting from the free solution and iterating the Φ_2^4 equations of motion. These results consisted mainly of signs and factorization properties (named "splitting" properties) for the Schwinger functions, properties which furthermore revealed some bounds and norms in the course of the Φ iteration. All these properties have shown the way to define, in the space of the Green's functions sequences, appropriate subsets [cf. Definitions 2(b) and 2(e) in Sec. II] in zero and two dimensions. So by a fixed-point method we can construct the solution of each of the problems within the corresponding particular subset.

Before giving the plan of the paper we recall briefly the fundamental definitions introduced in I that we shall use extensively throughout the sections to follow. Sometimes we shall only mention their names and refer the reader to the corresponding paragraph in I. We start with the fundamental notation of the whole work: a Schwinger function will be denoted by $H^{n+1}(q,\Lambda)$, with $n = 2k+1$, $k \in \mathbb{N}$, $\Lambda \in \mathbb{R}$, and $q \in \mathscr{E}^{rn}_{(q)}$. Here

$q \equiv \{q_i = (\mathbf{q}_i \in \mathbb{R}^{r-1},\ q_i^0 \in \mathbb{R}),\ \|q_i\| = |\mathbf{q}_i^2 + q_i^{02}|^{1/2},\ 1 \leqslant i \leqslant n\}$

 0022-2488/89/010175-22$02.50

is the set of independent r momenta, and $\mathscr{E}^{rn}_{(q)}$ means the corresponding Euclidean momentum space.

Definitions 1(a) (associated with the zero-dimensional problem): (i) The space $\mathscr{B}_0$. We consider the space $\mathscr{B}_0$ of the sequences $H_0 \equiv \{H_0^{n+1}(\Lambda)\}_n$ of C^∞ functions on $\mathbb{R}$ satisfying absolute bounds with respect to n, described in the following way: $\forall H_0 \in \mathscr{B}_0$, there exists a finite positive constant c_H such that, $\forall n = 1,3,5,\ldots$,

$$|H_0^{n+1}(\Lambda)| \leqslant n!\, c_H^{(n+1)/2}. \tag{1.1}$$

(ii) The zero-dimensional system of equations [cf. Definition 3(b) of I]. We define the nonlinear mapping $\mathscr{M}_0$: $\mathscr{B}_0 \to \mathscr{B}_0$ by the system of equations

$$H_0^2(\Lambda) = -\Lambda H_0^4 + 1, \tag{1.2a}$$

$$H_0^{n+1}(\Lambda) = A_0^{n+1} + B_0^{n+1}(\Lambda) + C_0^{n+1}(\Lambda), \quad \forall n \geqslant 3. \tag{1.2b}$$

Here

$$A_0^{n+1} = -\Lambda H_0^{n+3},$$

$$B_0^{n+1}(\Lambda) = -3\Lambda \sum_{w(J)} \theta^n_{j_1 j_2} H_0^{j_2+2} H_0^{j_1+1}, \tag{1.3}$$

$$C_0^{n+1} = -6\Lambda \sum_{w(I)} \theta^n_{i_1 i_2 i_3} \prod_{l=1}^{3} H_0^{i_l+1},$$

with

$$\theta^n_{j_1 j_2} = n!/j_1!\, j_2!, \quad \theta^n_{i_1 i_2 i_3} = n!/i_1!\, i_2!\, i_3!\, \sigma_{\mathrm{sym}}(i_1 i_2 i_3). \tag{1.3a}$$

The partitions of n $w(J)$ and $w(I)$ together with the symmetry factors $\sigma_{\mathrm{sym}}(i_1 i_2 i_3)$ are explicitly given by Definition 3(b) of I.

(iii) The sweeping procedure [cf. Definition 3(c) of I] in $\mathscr{B}_0$. In Sec. III of I we have introduced the "sweeping procedure" in terms of sequences of the so-called *sweeping factors* β_n by the formula

$$|C_0^{n+1}| = 3\Lambda n(n-1)\beta_n\, |H_0^{n-1}|\, |H_0^2|^2 \tag{1.4}$$

(and analogous definitions for every ordered partial sum of C_0^{n+1}). We analyzed there the double role of these quantities β_n. On one hand they give a simple form on the sum C_0^{n+1} (and the form of the splitting procedure recalled below), which is easily applied through the technique developed in Secs. II and III. On the other hand they carry precise combinatorial information for the terms of the ordered sum C_0^{n+1}.

(iv) The splitting procedure in $\mathscr{B}_0$: The space $\mathscr{B}_\delta$ [cf. Definition 3(d) of I]. Using the above definition of the sweeping factors, the splitting properties revealed by the Φ iteration have been reformulated in Sec. III of I as follows. We consider the space of infinite sequences of C^∞ functions of $\Lambda \in \mathbb{R}^+$, denoted by $\delta \equiv \{\delta_n(\Lambda)\}_{n=1,3,\ldots}$ and bounded as follows: $\forall \delta\ \exists 0 < c_\delta < \infty$ such that $|\delta_n(\Lambda)| \leqslant c_\delta n^4$, $\forall n$. We call this space $\mathscr{B}_\delta$. With every sequence $H_0 \in \mathscr{B}_0$ we associate a sequence $\delta \in \mathscr{B}_\delta$ such that the following recurrent definition (called the splitting procedure) of every $H_0^{n+1}(\Lambda)$ in terms of the corresponding H_0^{n+1} and H_0^2 holds:

$$\begin{aligned} &|H_0^2| = 1 + \delta_1(\Lambda)\Lambda, \quad |H_0^4| = \delta_3(\Lambda)\, |H_0^2|^2, \\ &|H_0^6| = 4\delta_5(\Lambda)\, |H_0^4(\Lambda)|\, |H_0^2(\Lambda)|, \end{aligned} \tag{1.5a}$$

and $\forall n \geqslant 7$,

$$|H_0^{n+1}| = \delta_n(\Lambda)\beta_n(\Lambda)\, |H_0^{n-1}|\, |H_0^2|. \tag{1.5b}$$

We called $\delta_n(\Lambda)$, $n = 1,3,\ldots$, the splitting constants.

The reader must realize that in what follows the one-dimensional problem is treated simultaneously with the two-dimensional one.

Definitions 1(b) (associated with the two- or one-dimensional problem): (i) The space $\mathscr{B}$ and the two- (or one-)dimensional system of equations. We consider the space $\mathscr{B}$ of the sequences $H \equiv \{H^{n+1}(q,\Lambda)\}_{n=1,3,\ldots}$ with $q \equiv \mathscr{E}^{rn}_{(q)}$, $r = 1,2$, $\Lambda \in \mathbb{R}$, satisfying the following bounds: $\forall H \in \mathscr{B}\ \exists$ a finite positive constant c_H such that

$$|H^2(q,\Lambda)(1+q^2)^{-1}| \leqslant c_H, \tag{1.6a}$$

$$|H^{n+1}(q,\Lambda)| \leqslant [n!]^2 c_H^n, \quad \forall n = 3,5,\ldots. \tag{1.6b}$$

The corresponding nonlinear mapping $\mathscr{M}: \to \mathscr{B}$ is defined by the following system of equations introduced in Sec. II of I [cf. Eqs. (2.4)]:

$$H^2(q,\Lambda) = -\Lambda[N_3 H^4](q,\Lambda) + q^2 + 1, \tag{1.7a}$$

$$\begin{aligned} H^{n+1}(q,\Lambda) = -\Lambda\Big\{&[N_3 H^{n+3}] \\ &+ 3\sum_{w_n(J)} [N_2 H^{j_2+2}][N_1 H^{j_1+1}] \\ &+ 6\sum_{w_n(I)} \prod_{l=1}^{3} [N_1 H^{i_l+1}]\Big\}. \end{aligned} \tag{1.7b}$$

The reader is referred to Eqs. (2.5)–(2.7) and Figs. 3–5 of I for the precise definitions and corresponding graphical representations of the so-called Φ^4-type operations,

$$[N_3 H^{n+3}](q,\Lambda), \quad [N_2 H^{i_2+2}][N_1 H^{j_1+1}](q,\Lambda),$$

$$\prod_{l=1}^{3} [N_1 H^{i_l+1}](q,\Lambda),$$

appearing on the rhs's of Eqs. (1.7).

(ii) The Φ convolution products (ΦC's)—the coherent sequences of ΦC's. In Sec. II of I we have also introduced the notion of Φ convolution products (ΦC's) associated with an arbitrary sequence $H \in \mathscr{B}$. More precisely, by definition 2(c) of I we have explicitly presented the recursive construction of an arbitrary ΦC by repeated application of an arbitrary number of the above-mentioned Φ^4-type operations. The corresponding graphical representation has been also precisely given in terms of the graphs G_Φ. Taking into account the particular structure of the Φ^4-type operations as they appear in Eqs. (1.7) of the mapping $\mathscr{M}$, it has been necessary for the consistency of the forthcoming definitions and proofs (cf. Sec. IV) to introduce the notion of coherent sequences of ΦC's. We do not give here their detailed description but refer to Definition 2(d) of I. We only recall that they are defined by reference to a fixed $\bar{n}$, and that the notation for them is $\{\Phi^{(\bar{n},n)}(H)H^{n+1}\}_n$, or Φ in abbreviated form. We shall also use the notation $\mathscr{F}\{\Phi^{(\bar{n},n)}(H)\}$ for the infinite family of coherent sequences of ΦC's associated with every $H \in \mathscr{B}$, or simply $\mathscr{F}$ in abbreviated form.

(iii) Reductions, the sweeping factors ${}^\Phi\beta_n$, and the splitting procedure in $\mathscr{B}$. When some or all of the H^{n+1} functions defining a ΦC are replaced by the constant 1, then we obtain the *partially* or *completely reduced* ΦC denoted by

$\widetilde{\Phi}^{(\bar{n},n)}$, or $[\Phi^{(\bar{n},n)}N_3]\widetilde{}$, etc. [cf. Definition 2(e) of I]. If $H_0 \in \mathscr{B}_0$, then we verify that every coherent sequence of ΦC's acting on H_0 factorizes out in the form of the corresponding completely reduced coherent sequence of ΦC's. This property allows to introduce a sweeping procedure analogous to (1.4) of the zero-dimensional case [cf. Sec. III of I, Eqs. (3.9) and (3.10)], namely,

$$|{}^{\Phi}C_0^{n+1}| = 3\Lambda^{\Phi}\beta_n n(n-1)\,|H_0^{n-1}|\,|H_0^2|^2 \times\left[\left|\Phi^{(\bar{n},n)}\prod_{l=1}^{3}N_1^{(\hat{i}_l)}\right|\right]_{q_{\bar{l}}=0}$$
$$(\text{with } \hat{i}_1 = n-2,\ \hat{i}_2 = \hat{i}_3 = 1). \tag{1.8}$$

Moreover, a formula analogous to Eqs. (3.5) of I holds for every partition of $n, (\bar{i}_1\bar{i}_2\bar{i}_3)$:

$$\left|\sum_{w(I)}\theta^n_{i_1i_2i_3}\left[\Phi^{(\bar{n},n)}(H_0)\prod_{l=1}^{3}N_1^{(i_l)}\prod_{l=1}^{3}H_0^{i_l+1}\right|_{q_{\bar{l}}=0}$$
$$= {}^{\Phi}\beta^n_{\bar{i}_1\bar{i}_2\bar{i}_3}\theta^n_{\bar{i}_1\bar{i}_2\bar{i}_3}\left|\prod_{l=1}^{3}H_0^{\bar{i}_l+1}\right|\left|\Phi^{(\bar{n},n)}(H_0)\prod_{l=1}^{3}N_1^{(\bar{i}_l)}\right|_{q_{\bar{l}}=0}$$
$$+ {}^{\Phi}\beta^n_{\bar{i}_1-2,\hat{i}_2,\hat{i}_3}\theta^n_{\bar{i}_1-2,\hat{i}_2,\hat{i}_3}\,|H_0^{\bar{i}_1-1}H_0^{\hat{i}_2+1}H_0^{\hat{i}_3+1}|$$
$$\times|[\Phi^{(\bar{n},n)}(H_0)N_1^{(\bar{i}_1-2)}N_1^{(\hat{i}_2)}N_1^{(\hat{i}_3)}]_{q_{\bar{l}}=0}|, \tag{1.8a}$$

and the analog of (3.5b) of I if $\bar{i}_3 \leqslant i_3 \min(\bar{i}_1 - 2) \equiv \hat{i}_3$. Finally, a splitting procedure analogous to Eqs. (1.5) can be defined [cf. Sec. III of I, Eqs. (3.11a)–(3.11c)] in terms of the sequences ${}^{\Phi}\delta$.

We now present the general plan of the paper. The purpose of Sec. II which follows is to reformulate, using the basic definitions of Sec. II of I, all the experimental information collected in I through the Φ iteration in more mathematically elegant language. More precisely, the signs, the splitting properties, and the absolute upper bounds of the Schwinger functions are now the characteristic features that describe the structure of the subsets $\Phi_{0\Lambda} \subset \mathscr{B}_0$ (cf. Sec. II A for zero dimensions) and $\Phi_\Lambda \subset \mathscr{B}$ (cf. Sec. II B for one or two dimensions). Let us repeat here our fundamental assertion, presented in a slightly different way in I: The structure and stability of these subsets under the mappings, $\mathscr{M}_0$ and $\mathscr{M}$, respectively, has been the sufficient condition for our method to work in both cases. In other words, the conservation of signs and precise bounds in terms of the splitting properties yields the conservation of the norms in the corresponding spaces $\mathscr{B}_0$, $\mathscr{B}$, which in turn implies the existence of fixed points (cf. Secs. III and IV) of $\mathscr{M}_0$, $\mathscr{M}$. In Sec. II A (devoted to the zero-dimensional case) we give the definition of the subsets $\Phi_{0\Lambda} \subset \mathscr{B}_0$, characterized by the signs of the sequences H_0, and the precise increase properties of the splitting sequences $\delta \in \mathscr{B}_\delta$. For every sequence $H_0 \in \Phi_{0\Lambda}$ we then show crucial bounds of the sweeping factors β_n and the global terms of the mapping $\mathscr{M}_0$ [defined in (1.3)]. In Sec. II B, using the above definitions 1(b), we proceed in an analogous way for the two-dimensional problem. The corresponding subset $\Phi_\Lambda \subset \mathscr{B}$ is introduced, and the properties for the sweeping factors and corresponding global terms of Eqs. (1.7) when $H \in \Phi_\Lambda$ are obtained.

Section III is devoted to the solution of the zero-dimensional system (1.2) inside $\Phi_{0\Lambda}$, through the solution of an equivalent system of δ sequences inside an appropriate subset of $\mathscr{B}_\delta$ space.

In Sec. IV, after the introduction of a precise norm in $\mathscr{B}$, we give the proof of the stability of Φ_Λ under $\mathscr{M}$ and its closedness. Then we find the solution of (1.7) through the application of the contractive mapping principle inside Φ_Λ. Moreover, a simpler proof of the convergence of the Φ iteration to the solution is presented. This result provides us immediately with an iterative construction of the above solution.

Finally, in Sec. V we introduce the new spaces $\hat{\mathscr{B}}_0 \subset \hat{\mathscr{B}}$ of double sequences, and by the definition of an appropriate norm we prove in a unified way the contractivity of both $\hat{\mathscr{M}}_0$ and $\hat{\mathscr{M}}$ mappings inside the corresponding subsets.

We conclude this introduction with the following remarks: All the statements and conclusions presented in this article have already appeared in the form of two Bielefeld University preprints.[5] Apart from some slight modifications in the arguments of certain proofs and the general presentation, the results are exactly the same. We essentially simplified and abbreviated the form of the paper. All the technical and combinatorial informations (concerning, in particular, the sweeping factors) are collected and written in the form of a separate paper that we call Paper III.[6]

II. THE SUBSETS $\mathscr{C}_\Lambda \subset \mathscr{B}_\delta$, $\Phi_{0\Lambda} \subset \mathscr{B}_0$, AND $\Phi_\Lambda \subset \mathscr{B}$

This section is essentially devoted to the definition of the subsets $\Phi_{0\Lambda}$ (resp. Φ_Λ) through which we shall demonstrate the existence of the solutions of the zero-dimensional (1.2) and the two-dimensional (1.7) systems in Secs. III and IV, respectively.

We begin by presenting the so-called subsets $\mathscr{C}_\Lambda$ described by all the detailed bounds of the splitting constant sequences δ in the space $\mathscr{B}_\delta$, because, as will become clear below, the definition of the subsets $\Phi_{0\Lambda}$ (resp. Φ_Λ) includes precisely the fine structure of $\mathscr{C}_\Lambda$. We shall then distinguish two parts of the section. In the first part (i.e., Sec. II A) we shall proceed to the description of $\Phi_{0\Lambda}$ and establish the nonemptiness of it. Then we prove some recursive formulas and bounds for the sweeping factors β_n when $H_0 \in \Phi_{0\Lambda}$. Moreover, we introduce the sweeping factors α_n's associated with the global terms B_0^{n+1} [cf. the definition (1.3)]. These sequences α_n play exactly the same role for the sums B_0^{n+1} as the β_n's play for C_0^{n+1}. Under the assumption $H_0 \in \Phi_{0\Lambda}$ we also show the corresponding recursive formulas and bounds for the sweeping factors α_n. We close Sec. II A with a statement concerning signs and bounds for the global terms $A_0^{n+1}, B_0^{n+1}, C_1^{n+1}$ of the definition (1.3) when $H_0 \in \Phi_{0\Lambda}$. All of these established properties concerning the sweeping factors and the global terms of (1.2) will play a crucial role in the proof of the main theorems of Sec. III.

The second part, i.e., Sec. II B, deals with the two- (or one-)dimensional case and starts with the definition and properties of the triplets $\{{}^{\Phi}a_n, {}^{\Phi}b_n, {}^{\Phi}c_n\}$ of parameters [cf. Definition 2(d) below] associated with every coherent sequence of ΦC's. These positive finite quantities express the supplementary information that we obtain only in the dimensional cases, and that comes from the presence of nontrivial Φ convolutions. They constitute the basic difference between the corresponding "zero external momenta" systems

(1.2) and (1.7) because in the zero-dimensional case they are trivially reduced to the constant 1.

The remaining part of Sec. II B proceeds in a way exactly analogous to the one explained above for Sec. II A. The subset Φ_Λ is introduced, characterized by the factorization of reduced ΦC's, by their signs, and, for every coherent sequence of ΦC's (denoted briefly by Φ), by precise bounds and increase properties for the splitting sequences ${}^\Phi\delta\in\mathscr{B}_\delta$.

Then we show the nonemptiness of Φ_Λ and (when $H\in\Phi_\Lambda$) the properties implied for ${}^\Phi\beta_n$'s,${}^\Phi\alpha_n$'s, and the global terms ${}^\Phi A_0^{n+1}$,${}^\Phi B_0^{n+1}$,${}^\Phi C_0^{n+1}$. Here, for a given $H_0\in\mathscr{B}_0$,

$$ {}^\Phi A_0^{n+1} \overset{\text{def}}{\equiv} -\Lambda H_0^{n+3}\{\Phi^{(n,\bar n)}(H_0)N_3^{(n+2)}\}_0, \quad \text{(a)} $$

$$ {}^\Phi B_0^{n+1} \overset{\text{def}}{\equiv} -3\Lambda\sum_{w(J)}\theta^n_{j_1j_2}H_0^{j_2+2}H_0^{j_1+1} \times\{\Phi^{(n,\bar n)}(H_0)N_2^{(j_2)}N_1^{(j_1)}\}_0, \quad \text{(b)} $$

$$ {}^\Phi C_0^{n+1} \overset{\text{def}}{\equiv} -6\Lambda\sum_{w(I)}\theta^n_{i_1i_2i_3}\prod_{l=1}^3 H_0^{i_l+1}\left\{\Phi^{(n,\bar n)}(H_0)\prod_{l=1}^3 N_1^{(i_l)}\right\}_0. \quad \text{(c)} $$

These properties will be extensively used in the proofs of Sec. IV.

Let us consider the space $\mathscr{B}_\delta$ of the sequences $\delta(\Lambda)\equiv\{\delta_n(\Lambda)\}_n$ presented in the Introduction. We define the following subsets $\mathscr{C}_\Lambda$ characterized by precise limit values, absolute bounds, and slow increase properties (i.e., relative bounds) of the splitting sequences δ.

Definition 2(a) [The subsets $\mathscr{C}_\Lambda(\delta^\Lambda_\infty,I_n,S_n)\subset\mathscr{B}_\delta$]: Let Λ be fixed positive. A sequence $\delta\in\mathscr{B}_\delta$ is said to belong to a subset $\mathscr{C}_\Lambda$ if the following properties are satisfied $\forall n=1,3,5\ldots$.

$\mathscr{C}_\Lambda$ (i) (limit values): (a) There exist finite real positive numbers c_n such that

$$ \lim_{\Lambda\to 0}\delta_n(\Lambda)/\Lambda = c_n. \quad \text{(2.1a)} $$

(b) There exists a finite positive constant δ^Λ_∞, $2\leqslant\delta^\Lambda_\infty<+\infty$, independent on the particular δ and such that

$$ \lim_{n\to\infty}\delta_n(\Lambda)=\delta^\Lambda_\infty. \quad \text{(2.1b)} $$

$\mathscr{C}_\Lambda$ (ii) (absolute bounds): There exist rational real positive functions of Λ, $I_n(\Lambda)$, $S_n(\Lambda)<1$, $\forall n\geqslant 3$, and such that

(a) $6\Lambda I_3\leqslant\delta_1(\Lambda)\leqslant 6\Lambda S_1$, $\quad 6\Lambda I_3\leqslant\delta_3(\Lambda)\leqslant 6\Lambda S_3$,

$$ 15\Lambda I_5\leqslant\delta_5(\Lambda)\leqslant 15\Lambda S_5; \quad \text{(2.2a)} $$

(b) $$ 3\Lambda n(n-1)I_n\leqslant\delta_n(\Lambda)\leqslant 3\Lambda n(n-1)S_n, \quad \forall n\geqslant 7. \quad \text{(2.2b)} $$

$\mathscr{C}_\Lambda$ (iii) (relative bounds—slow increasing behavior): $\forall n\geqslant 7$ and for $N_\Lambda\equiv 4/3\Lambda+1$,

(a) $\frac{42}{5}(1-\bar\gamma_7(\Lambda))\leqslant\delta_7/\delta_5\leqslant\frac{42}{5}(1-\hat\gamma_7(\Lambda))$,

$$ \frac{n(n-1)}{(n-2)(n-3)}(1-\bar\gamma_n(\Lambda)) \leqslant\frac{\delta_n}{\delta_{n-2}}\leqslant\frac{n(n-1)}{(n-2)(n-3)}(1-\hat\gamma_n(\Lambda)), \quad \text{if } 9\leqslant n\leqslant N_\Lambda, \quad \text{(2.3a)} $$

with $0\leqslant\hat\gamma_n(\Lambda)\leqslant\bar\gamma_n(\Lambda)\leqslant 4/(n-1)$;

(b) $$ 1+\frac{\bar\mu_n(\Lambda)}{n^2}\leqslant\frac{\delta_n}{\delta_{n-2}}\leqslant 1+\frac{\hat\mu_n(\Lambda)}{n^2}, \quad \text{if } n\geqslant N_\Lambda, \quad \text{(2.3b)} $$

with $0\leqslant\bar\mu_n\leqslant\hat\mu_n<\infty$.

We notice that, following the above definition and once δ^Λ_∞ is given, we can find an infinite number of pair sequences $\{I_n(\Lambda),S_n(\Lambda)\}$ of bounded real positive rational functions of Λ, and so also an infinite number of subsets $\mathscr{C}_\Lambda$ in $\mathscr{B}_\delta$ (for the same fixed Λ and δ^Λ_∞). Nevertheless, below (and resp. in Sec. II B) we shall specify a particular pair $\bar I_n(\Lambda),\bar S_n(\Lambda)$ so that the subset $\overline{\mathscr{C}}_\Lambda$ (resp. $\tilde{\mathscr{C}}_\Lambda$) and the corresponding $\Phi_{0\Lambda}$ (resp. Φ_Λ) will be uniquely defined by the set $(\delta^\Lambda_\infty,\bar S_n(\Lambda),\bar I_n(\Lambda))$ [resp. $(\tilde\delta^\Lambda_\infty,\tilde S_n(\Lambda),\tilde I_n(\Lambda))$].

A. Zero-dimensional problem

Definition 2(a1): Let δ^Λ_∞ be given. We consider the following pair of sequences $\{\bar I_n(\Lambda),\bar S_n(\Lambda)\}$:

$$ \bar I_3(\Lambda)\equiv[1+9\Lambda(1+6\Lambda^2)]^{-1}, $$
$$ \bar S_3(\Lambda)\equiv(1+6\Lambda^2)[1+9\Lambda-60\Lambda^2]^{-1}, $$
$$ \bar S_1(\Lambda)\equiv 4[1-24\Lambda^2S_3]^{-1/2}\times[1+(1-24\Lambda^2S_3)^{1/2}]^{-1}, \quad \text{(2.4a)} $$

$$ \bar I_5(\Lambda)\equiv(1-3\Lambda S_3)/1+15\Lambda(1+6\Lambda^2), $$
$$ \bar S_5(\Lambda)\equiv\frac{(1+6\Lambda^2\bar S_1)[3\Lambda I_3(14\Lambda-1)+1]}{1+15\Lambda(1+6\Lambda^2\bar S_1)(1-42\Lambda/5)}, \quad \text{(2.4b)} $$

and

$$ \bar I_n(\Lambda)\equiv 2[2+3\Lambda n(n-1)(1+6\Lambda^2)]^{-1}, $$
$$ \bar S_n(\Lambda)\equiv\frac{\delta^\Lambda_\infty(1+6\Lambda^2)}{\delta^\Lambda_\infty+3\Lambda n(n-1)(1+6\Lambda^2)}, \quad \forall n\geqslant 7. \quad \text{(2.4c)} $$

We denote by $\overline{\mathscr{C}}_\Lambda(\delta^\Lambda_\infty,\bar I_n(\Lambda),\bar S_n(\Lambda))$ the subset defined by Definition 2(a) and specified by the above sequences (2.4). We show the nontriviality of $\overline{\mathscr{C}}_\Lambda$ by the following lemma.

Lemma 2.1: The subset $\overline{\mathscr{C}}_\Lambda\subset\mathscr{B}_\delta$ of Definition 2(a1) is nonempty for every fixed $\Lambda\in(0;0.1]$.

Proof: We consider the following sequence $\delta^*\in\mathscr{B}_\delta$:

$$ \delta^*\equiv\{\delta_1^*(\Lambda)=6\Lambda\bar I_3,\ \delta_3^*(\Lambda)=6\Lambda\bar I_3,\ \delta_5^*(\Lambda)=15\Lambda\bar I_5, \text{ and } \forall n\geqslant 7,\ \delta_n^*(\Lambda)=3\Lambda n(n-1)\bar I_n(\Lambda)\}. \quad \text{(2.5)} $$

We first verify $\mathscr{C}_\Lambda$ (i):

(a) $\lim_{\Lambda\to 0}\delta_n^*(\Lambda)/\Lambda=u_n$, with $u_1=u_3=6$

and

$$ u_5=15,\quad u_n=3n(n-1),\quad\forall n\geqslant 7; \quad \text{(2.5a)} $$

(b) $\lim_{n\to\infty}\delta_n^*(\Lambda)\leqslant 2$.

Q.E.D.

Moreover, $\mathscr{C}_\Lambda$ (ii) is trivially proved in view of (2.4) and (2.5). Concerning $\mathscr{C}_\Lambda$ (iii) we find (a) for $9\leqslant n\leqslant 4/3\Lambda+1$,

$$\frac{\delta_n^*}{\delta_{n-2}^*}=\frac{n(n-1)}{(n-2)(n-3)}(1-\gamma_n^*(\Lambda)),$$

with

$$\gamma_n^*(\Lambda)\equiv 1-\frac{\bar I_n}{I_{n-2}}$$

$$=\frac{6\Lambda(1+6\Lambda^2)(2n-3)}{2+3\Lambda n(n-1)(1+6\Lambda^2)},\qquad (2.5b)$$

and analogous relations for $n=7$: (b) for $n\geqslant 4/3\Lambda+1$,

$$1\leqslant\delta_n^*/\delta_{n-2}^*\leqslant 1+2/(n-2)(n-3),\qquad (2.5c)$$

where the last inequality is obtained iff $n\geqslant 4/3\Lambda+1$.

Q.E.D.

From these results it follows that $\delta^*\in\overline{\mathscr{C}}_\Lambda$, and this ensures that $\mathscr{C}_\Lambda\neq\varnothing$. Q.E.D.

Definition 2 (a2) (The order $\mathscr{O}$): Two sequences $\delta_{(1)},\delta_{(2)}\in\mathscr{B}_\delta$ are said to be ordered following $\mathscr{O}$, and we write (resp. $\delta_{(1)}>\delta_{(2)}$), iff $\forall n=1,3,\ldots,\ \delta_{n(1)}\geqslant\delta_{n(2)}$ (resp. $\delta_{n(1)}\leqslant\delta_{n(2)}$).

We also introduce the ordered subsets $\mathscr{C}_\Lambda'\subset\mathscr{C}_\Lambda$. As is shown in Sec. III, these subsets are appropriate for the validity of the contractivity criterion (under sufficient conditions on Λ) by the nonlinear map of the splitting sequences.

Definition 2(a3) (The ordered subsets $\mathscr{C}_\Lambda'\subset\mathscr{C}_\Lambda$): The sequences of $\mathscr{C}_\Lambda'$ are ordered following $\mathscr{O}$, i.e., $\forall\delta_{(1)},\delta_{(2)}\in\mathscr{C}_\Lambda'$, either $\delta_{(1)}>\delta_{(2)}$ or $\delta_{(1)}<\delta_{(2)}$. Moreover, if $\delta_{(1)}>\delta_{(2)}$ and $\delta_{n(1)}\neq\delta_{n(2)}$ $\forall n\geqslant 5$, then

$$\frac{\delta_{n(2)}}{\delta_{n-2(2)}}\leqslant\frac{\delta_{n(1)}-\delta_{n(2)}}{\delta_{n-2(1)}-\delta_{n-2(2)}}\leqslant\begin{cases}\left[\dfrac{n(n-1)}{(n-2)(n-3)}\right]^2, & \text{if } n\geqslant 9,\qquad (\mathscr{C}_\Lambda'\ \text{a})\\ (\tfrac{42}{5})^2, & \text{if } n=7,\qquad (\mathscr{C}_\Lambda'\ \text{b})\end{cases}$$

and

$$\delta_{5(1)}/\delta_{3(1)}\geqslant\delta_{5(2)}/\delta_{3(2)}.$$

Using the above notion of the $\mathscr{C}_\Lambda$ subsets in $\mathscr{B}_\delta$ space, we now introduce the subset $\Phi_{0\Lambda}$ in the space $\mathscr{B}_0$ of the sequences H.

Definition 2(b) (The subset $\Phi_{0\Lambda}\subset\mathscr{B}_0$): A sequence $H\in\mathscr{B}_0$ belongs to the subset $\Phi_{0\Lambda}$ if it satisfies the following *signs* and *splitting properties*:

$\Phi_{0\Lambda}$ (i): $(-1)^{(n-1)/2}H_0^{n+1}(\Lambda)>0$;

$\Phi_{0\Lambda}$ (ii): The associated sequence of splitting constants $\delta\equiv\{\delta_n(\Lambda)\}_{n=1,3,\ldots}$ presented by (1.5) of the splitting procedure belongs to the subset $\overline{\mathscr{C}}_\Lambda$ of Definition 2(a1).

When a sequence $H_0\in\mathscr{B}_0$ belongs to the subset $\Phi_{0\Lambda}$, then it is possible to define the sweeping factors β_n (or $\beta^n_{i_1i_2i_3}$) by a recurrence in terms of only the ratios of the splitting constants $\delta_{\bar n}$ with $\bar n\leqslant n-2$, and all $\beta_{\bar n}$'s with $\bar n\leqslant n-2$ (or $\beta^{\bar n}_{\bar i_1\bar i_2\bar i_3}$ with $\bar i_1\leqslant i_1-2$, etc.). We establish this recursive sweeping procedure by Lemma 2.2 below. For the proof of this statement [where the sign properties $\Phi_{0\Lambda}$ (i) are essentially used through (1.4) and (1.5) of the Introduction] we refer the reader to III,[6] where we present all the detailed proofs of the lemmas concerning the sweeping factors.

In the example of Fig. 1 (for the case $n=21$) one can visualize the mechanism of this sweeping procedure and the proof of Lemma 2.2, inside the ordered $C_0^{22}(i_1\geqslant i_2\leqslant i_3)$.

Lemma 2.2 (Sweeping procedure in C_0^{n+1}): Let $H_0\in\Phi_{0\Lambda}$; $\forall n\geqslant 3$ the sweeping factors β_n [resp. $\beta^n_{i_1i_2i_3}$ for every partition $(i_1i_2i_3)$] are given recurrently as follows:

$$\beta_3=\beta_5=1,$$

and $\forall n\geqslant 7$,

$$\beta_n=\beta^n_{n-2,1,1},$$

where $\forall(i_1i_2i_3)$ with $i_1\geqslant i_2\geqslant i_3$,

$$\begin{aligned}\beta^n_{i_1i_2i_3}&=1+\zeta(i_1i_2)\frac{\theta^n_{i_1-2,i_2+2,i_3}}{\theta^n_{i_1i_2i_3}}\frac{\delta_{i_2+2}}{\delta_{i_1}}\frac{z(\beta_{i_2}+2)}{z(\beta_{i_1})}\beta^n_{i_1-2,i_2+2,i_3}\\&\quad+\zeta(i_2i_3)\frac{\theta^n_{i_1,i_2-2,i_3+2}}{\theta^n_{i_1i_2i_3}}\frac{\delta_{i_3+2}}{\delta_{i_2}}\frac{z(\beta_{i_3+2})}{z(\beta_{i_2})}\beta^n_{i_1,i_2-2,i_3+2}\\&\quad+\zeta(i_1i_3)\frac{\theta^n_{i_1-2,i_2,i_3+2}}{\theta^n_{i_1,i_2,i_3}}\frac{\delta_{i_3+2}}{\delta_{i_1}}\frac{z(\beta_{i_3+2})}{z(\beta_{i_1})}.\end{aligned}\qquad (2.6)$$

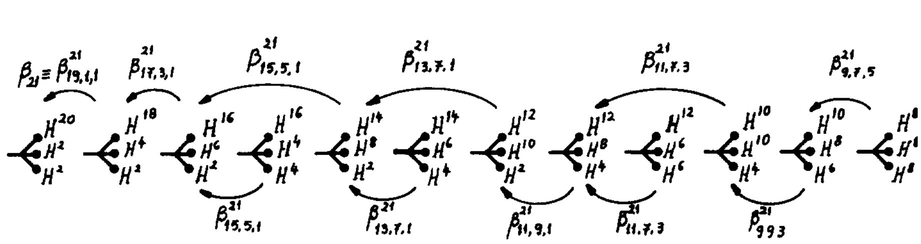

FIG. 1. Sweeping factors of C_0^{22}.

Here

$$z(\beta_i) = \begin{cases} \beta_i, & \text{for } i \neq 5, \\ 4, & \text{for } i = 5, \end{cases} \tag{2.6a}$$

$$\zeta(i_1 i_2) = \begin{cases} 1, & \text{if } \begin{cases} i_1 > (n+1)/2 \text{ and } i_2 = n - i_1 - 1, \\ i_1 \leqslant (n+1)/2 \text{ and } i_2 = i_1 - 4, \end{cases} \\ 0, & \text{otherwise,} \end{cases} \tag{2.6b}$$

$$\zeta(i_2 i_3) = \begin{cases} 1, & \text{if } i_3 \leqslant i_2 - 4, \\ 0, & \text{otherwise,} \end{cases} \quad \zeta(i_1 i_3) = \begin{cases} 1, & \text{if } i_1 - 2 = i_2 = n/3, \\ 0, & \text{otherwise.} \end{cases} \tag{2.6c}$$

Notice that $\beta^n_{i_1 - 2, i_2, i_3 + 2} = 1$ if $\zeta(i_1 i_3) = 1$.

Using extensively the above explicit expressions of β_n's ($\beta^n_{i_1 i_2 i_3}$'s), we have obtained through a double recursive procedure the absolute bounds for the sweeping factors presented by Lemma 2.3 below. These results translate an important combinatorial property: The β_n's (resp. $\beta^n_{i_1 i_2 i_3}$'s) are proportional, up to finite constants smaller than 1, to the number of terms swept by them inside C_0^{n+1} (resp. inside a partial sum of C_0^{n+1}). As we already mentioned in I, for every n this number, which equals the number of different possible configurations $(i_1 i_2 i_3)$ of n, is calculated in Appendix B of III and is denoted by $\mathscr{T}_n$ [resp. $\mathscr{T}^n_{i_1 i_2 i_3}$ for every given partition $(i_1 i_2 i_3)$]. Roughly speaking, following Appendix B of III we have, for n sufficiently large ($n > 25$),

$$\mathscr{T}_n \approx (n-3)^2/48 + (n-3)/4. \tag{2.7}$$

Combining (1.5) and $\mathscr{C}_\Lambda$ (ii) b of $\overline{\mathscr{C}}_\Lambda$ with this number, the structure of the space $\mathscr{B}_0$ (1.1) is justified, as we show in Lemma 2.4 below (i.e., consistency with the assumption $\Phi_{0\Lambda} \subset \mathscr{B}_0$). Moreover, this combinatorial bound later becomes a crucial tool for the theorems in Sec. III, where, in particular, we show the absolute upper bound δ^Λ_∞ of the splitting sequence $\delta_n(\Lambda)$ when $n \to \infty$ (cf. also the proof of Proposition 3.2 in III).

The proof of Lemma 2.3 is rather complicated; for this reason it is also left for Paper III. Nevertheless we can note at this moment that this technical demonstration, which presents the form of a double recurrence, is realized thanks to the fundamental relative bounds $\mathscr{C}_\Lambda$ (iii) of the splitting sequences $\delta \in \overline{\mathscr{C}}_\Lambda$.

Lemma 2.3: (i) Let $\delta \in \overline{\mathscr{C}}_\Lambda$; $\forall n \geqslant 7$ the following bounds are satisfied by the sweeping factors defined by Lemma 2.2 when Λ is fixed in $(0;0,1]$:

$$\overline{Y}_n(\Lambda) \mathscr{T}_n \leqslant \beta_n \leqslant \mathscr{T}_n Y_n(\Lambda). \tag{2.8}$$

The quantities $Y_n(\Lambda)$ [resp. $\overline{Y}_n(\Lambda)$] are defined recurrently by explicit expressions given in III, they decrease slowly with n (cf. III), and they satisfy

(a)

$$0 < \overline{Y}_n(\Lambda) \leqslant Y_n(\Lambda) \leqslant 1, \tag{2.9a}$$

(b)

$$\lim_{\Lambda \to 0} Y_n(\Lambda) = \lim_{\Lambda \to 0} \overline{Y}_n(\Lambda) = Y_n^{(0)} > 0, \tag{2.9b}$$

(c) there exist $Y_\infty(\Lambda) > 0$ such that

$$\lim_{n \to \infty} \overline{Y}_n(\Lambda) = \lim_{n \to \infty} Y_n(\Lambda) = Y_\infty(\Lambda). \tag{2.9c}$$

(ii) Let $\delta_{(1)} > \delta_{(2)} \in \overline{\mathscr{C}}'_\Lambda$; then $\forall n \geqslant 7$,

$$\beta_{n(2)} - \beta_{n(1)} \geqslant 0, \qquad \beta^n_{n-4,3,1}/\beta_{n-2(1)} \geqslant \beta^n_{n-4,3,1}/\beta_{n-2(2)}, \tag{2.9d}$$

when $0 < \Lambda \lesssim 0.01$.

As we mentioned before, the upper bounds (2.8) and the corresponding upper bounds $\mathscr{C}_\Lambda$ (ii) of slow increasing sequences $\delta \in \mathscr{C}_\Lambda$ yield the means to reproduce the bounds defining $\mathscr{B}_0$ [cf. (1.1)] and obtain the proof of Lemma 2.4 below. This statement establishes the nontriviality of $\Phi_{0\Lambda} \subset \mathscr{B}_0$ through a simple method of construction of a sequence $H_0^* \in \Phi_{0\Lambda}$ once a sequence $\delta^* \in \overline{\mathscr{C}}_\Lambda$ is given.

Lemma 2.4: (i) Let $^*\delta \in \overline{\mathscr{C}}_\Lambda$ and $\Lambda \in (0;0.1]$; every sequence $^*H_0$ defined by the following recursion belongs to the subset $\Phi_{0\Lambda}$ of Definition 2(b):

$$\begin{aligned} ^*H \equiv \{ ^*H^2 &= 1 + {}^*\delta_1 \Lambda, \ {}^*H^4 = -{}^*\delta_3 [{}^*H^2]^2, \\ ^*H^6 &= -4 {}^*\delta_5 {}^*H^4 {}^*H^2; \\ \forall n \geqslant 7, \ {}^*H^{n+1} &= -{}^*\delta_n {}^*\beta_n {}^*H^{n-1} {}^*H^2 \}. \end{aligned} \tag{2.10}$$

Here the sequences $\{^*\beta_n\}$ are given recurrently by Lemma 2.2 as functionals of all $^*\beta_{\bar n}$'s (resp. $^*\beta^{\bar n}_{\bar i_1 \bar i_2 \bar i_3}$ with $\bar i_1 \leqslant \bar n - 2$) and $^*\delta_{\bar n}$'s with $\bar n \leqslant n - 2$.

(ii) The subset $\Phi_{0\Lambda} \subset \mathscr{B}_0$ is nonempty.

Proof: (i) We first ensure that $^*H \in \mathscr{B}_0$. Following the hypothesis $^*\delta \in \overline{\mathscr{C}}_\Lambda$ we obtain by Definition 2(a)

$$|{}^*H^2| < 2, \quad |{}^*H^4| < 3!\, \delta^\Lambda_\infty, \quad |{}^*H^6| < 5! (\delta^\Lambda_\infty)^2. \tag{2.10a}$$

These bounds allow us to state the following recursion: $\forall 5 \leqslant \bar n \leqslant n - 2$ we suppose that

$$|H^{\bar n + 1}| \leqslant [\delta^\Lambda_\infty]^{(\bar n - 1)/2} \bar n!\,. \tag{2.10b}$$

To prove (2.10b) for $\bar n = n$ we first use the assumption in part (i) of the lemma that the $^*\beta_n$'s are defined by Lemma 2.2. So by application of Lemma 2.3 we obtain (cf. Appendix B of III for $\mathscr{T}_n$)

$$^*\beta_n < n(n-1)/10. \tag{2.10c}$$

From the latter and the recurrence hypothesis (2.10b), the general definition (2.10) yields

$$|{}^*H^{n+1}| \leqslant [\delta^\Lambda_\infty]^{(n-1)/2} n!\,.$$

By comparison of the last bound with the definition of $\mathscr{B}_0$ (1.1) we obtain that $^*H \in \mathscr{B}_0$ with $c_H = \delta^\Lambda_\infty$. Q.E.D.

On the other hand we note that by the definition of $^*H$ (2.10), properties $\Phi_{0\Lambda}$ (i) (signs) and $\Phi_{0\Lambda}$ (ii) are automatically satisfied. These conclusions allow us to state that $^*H \in \Phi_{0\Lambda}$. Q.E.D.

(ii) For the construction (2.10) we consider precisely the sequence $^*\delta \in \overline{\mathscr{C}}_\Lambda$ defined by (2.5) and ensuring the nonemptiness of $\overline{\mathscr{C}}_\Lambda$. In view of (i) we obtain that $\Phi_{0\Lambda} \neq \varnothing$. Q.E.D.

Definition 2(c) (The sweeping factors α_n for B_0^{n+1} [cf. (1.3)]*:* With (1.4) (the definitions of the sweeping factors β_n) and Lemma 2.2, we succeeded in replacing, in an explicit way, the sum C_0^{n+1} of (1.2b) for the system (1.2) by only one term that is proportional to the "dominant" contribution $H_0^{n-1}[H_0^2]^2$. Using an analogous combinatorial technique we introduce below the corresponding constants α_n [or resp. $\alpha^n_{j_1 j_2}$ for every partition (j_1,j_2)] which, inside the ordered sum B_0^{n+1}, play the analogous role as the β_n's (resp. $\beta^n_{i_1 i_2 i_3}$'s) do inside the ordered sum C_0^{n+1}: they "sweep," or carry, all the combinatorial information coming from the preceding contributions of the "ordered"—following the increasing values of j_2—sum B_0^{n+1} (resp. of the partial sum of B_0^{n+1}), and they replace it by one term proportional to the dominant contribution. So let us define the sweeping factors α_n by the following equations, for any $\Lambda > 0$ and $n \geqslant 3$:

$$|B_0^{n+1}| = 3\Lambda\alpha_n\theta^n_{1,n-1}\,|H_0^{n+1}|\,|H_0^2|. \tag{2.11}$$

Trivially, we obtain $\alpha_3 = 1$. Moreover, for every fixed partition $w(\bar{J}) = (\bar{j}_1,\bar{j}_2)$ with $[(n+1)/2] \leqslant \bar{j}_2 \leqslant n-1$ (here $[N] = N$ if N is even, $[N] = N-1$ otherwise), we define the corresponding sweeping factor $\alpha^n_{\bar{j}_1\bar{j}_2}$ by the equation

$$\Big|\sum_{w(J)} \theta^n_{j_1 j_2} H_0^{j_2+2} H_0^{j_1+1}\Big| = \alpha^n_{\bar{j}_1\bar{j}_2}\theta^n_{\bar{j}_1\bar{j}_2}\,|H_0^{\bar{j}_2+2}|\,|H_0^{\bar{j}_1+1}|,$$
$$\bar{j}_1 - 1 \leqslant j_2 \leqslant \bar{j}_2, \quad [(n+1)/2] \leqslant \bar{j}_2 \leqslant n-1. \tag{2.12}$$

When $H_0 \in \Phi_{0\Lambda}$, then we are able again to define an analogous sweeping procedure as the one we obtained by Lemma 2.2 for the β_n's, such that every sweeping factor α_n (or $\alpha^n_{j_1 j_2}$) is defined recurrently only in terms of $\alpha^n_{\bar{j}_1\bar{j}_2}$ with $\bar{j}_2 \leqslant \bar{j}_2 - 2$ and the ratios of the splitting constants δ_n's. This sweeping procedure inside B_0^{n+1} is estabished by the first part of Lemma 2.5 below. The reader can visualize it in the example of Fig. 2 (case $n = 21$). The second part of this lemma contains absolute bounds of these sweeping factors α_n analogous to those of the β_n's in Lemma 2.3. These absolute bounds express the fact that α_n is proportional, up to a factor smaller than 1, to the number $\mathscr{T}_n$ of swept terms inside B_0^{n+1}, where

$$\mathscr{T}_n = (n-1)/2, \tag{2.13}$$

and they are important tools in the proofs of Sec. III.

The proof of Lemma 2.5 is also given in III (cf. Proposition 3.1 of III).

Lemma 2.5 (Sweeping procedure inside B_0^{n+1}): Let $H \in \Phi_{0\Lambda}$; $\forall n \geqslant 5$ and for every fixed partition $w(J) = (j_1,j_2)$ with $[(n+1)/2] \leqslant j_2 \leqslant n-1$, the following assertions can be made.

(i) The sweeping factors $\alpha_n, \alpha^n_{j_1 j_2}$ of (2.11) and (2.12) are given recurrently as follows:

$$\alpha_n \equiv \alpha^n_{1,n-1}, \tag{2.14a}$$

$$\alpha^n_{(n+1)/2,(n-1)/2} = 1, \quad \text{if } (n+1)/2 = \text{odd}, \tag{2.14b}$$

$$\alpha_{(n-1)/2,(n+1)/2} = 1 + (n-1)/(n+3), \quad \text{if } (n+1)/2 = \text{even}, \tag{2.14c}$$

and

$$\alpha^n_{j_1 j_2} = 1 + \frac{j_2(j_2-1)}{(j_1+2)(j_1+1)}\,\frac{\delta_{j_1+2}}{\delta_{j_2+1}}\,\frac{z(\beta_{j_1+2})}{z(\beta_{j_2+1})}\,\alpha^n_{j_1+2,j_2-2} + \frac{j_1 s(j_1-1)}{j_2+1}, \quad \text{for } (n+1)/2 < j_2 \leqslant n-1, \tag{2.15}$$

with

$$s(j_1-1) = \begin{cases} 1, & \text{if } j_1 > 1, \\ 0, & \text{if } j_1 = 1. \end{cases}$$

(ii) The following bounds hold:

$$\overline{X}_n(\Lambda)\mathscr{T}_n \leqslant \alpha_n \leqslant \mathscr{T}_n X_n(\Lambda). \tag{2.16}$$

The quantities $X_n(\Lambda)$, $\overline{X}_n(\Lambda)$ are given recurrently (by explicit formulas presented in III) and they satisfy the following properties: They decrease slowly with n (cf. III for details); and

(a) $0 < \overline{X}_n(\Lambda) < X_n(\Lambda) \leqslant 1;$ (2.17a)

(b) $\lim_{\Lambda \to 0} X_n(\Lambda) = \lim_{\Lambda \to 0} \overline{X}_n(\Lambda) = X_n^{(0)} > 0;$ (2.17b)

(c) $\lim_{n\to\infty} X_n(\Lambda) = \lim_{n\to\infty} \overline{X}_n(\Lambda) = X_\infty(\Lambda) > 0;$ (2.17c)

(d) $\exists\, n_0 > N_0$: $\forall n \geqslant n_0$, $\overline{X}_n^2 \geqslant Y_{n+2}$. (2.17d)

(iii) Let $H_{(1)}, H_{(2)} \in \Phi_{0\Lambda}$ with $\delta_{(1)} > \delta_{(2)} \in \overline{\mathscr{C}}'_\Lambda$; then $\forall n \geqslant 5$,

$$\alpha_{n(2)} - \alpha_{n(1)} \geqslant 0, \quad \text{when } 0 < \Lambda \leqslant 0.01. \tag{2.17e}$$

Finally, we present some auxiliary results concerning the signs and relations satisfied by the global terms $A_0^{n+1}, B_0^{n+1}, C_0^{n+1}$ of the mapping $\mathscr{M}_0$ (1.3), which are useful for the main theorems in Sec. III. We notice that the second part of the statement below contains the "double splitting" and "tree dominance" properties (the crucial properties for the conservation of the norms) obtained "experimentally" at zero external momenta during the Φ iter-

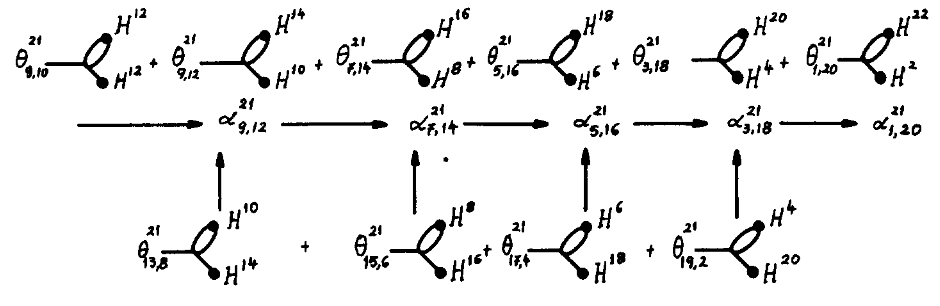

FIG. 2. Sweeping factors of B^{22}.

ation presented in I. One should expect to see them appearing in the definition of $\Phi_{0\Lambda}$. By the following proposition we show precisely that these properties are consequences of the two properties $\Phi_{0\Lambda}$ (i) and $\Phi_{0\Lambda}$ (ii), and that we do not need them as supplementary conditions on the subset $\Phi_{0\Lambda}$.

Proposition 2.1: Let $H_0 \in \Phi_{0\Lambda}$; the global terms of $\mathscr{M}_0$, $A_0^{n+1}(H_0), B_0^{n+1}(H_0), C_0^{n+1}(H_0)$ (1.3) satisfy the following properties $\forall n = 1,3,\ldots$.

$$\text{(i)}\ (-1)^{(n-1)/2} C_0^{n+1} > 0, \quad (-1)^{(n-1)/2} B_0^{n+1} < 0,$$

$$(-1)^{(n-1)/2} A_0^{n+1} > 0. \tag{2.18}$$

$$\text{(ii)}\ B_0^{n+1} = \vartheta_n \delta_n C_0^{n+1}, \tag{2.19}$$

where

$\vartheta_3 = \frac{3}{2}$, $\vartheta_5 = 1 + \frac{1}{2}\delta_3/\delta_5$,

and $\forall n \geqslant 7$, $\vartheta_n = \alpha_n/(n-1)$ with $0 < \vartheta_n \leqslant \frac{1}{2}$. (2.19a)

(iii) Double splitting:

$$\text{(a)}\ A_0^2 = \hat{\delta}_3(\Lambda), \quad \text{where } \hat{\delta}_3(\Lambda) = \Lambda \delta_3(\Lambda)(1 + \delta_1 \Lambda)^2; \tag{2.20}$$

(b) $\forall n \geqslant 3$, $\exists\, 0 < \hat{\delta}_n < +\infty$ such that

$$A_0^{n+1} = \hat{\delta}_{n+2} C_0^{n+1}, \tag{2.21}$$

with

$$\begin{aligned} &\hat{\delta}_5 = \tfrac{2}{3}\delta_3\delta_5, \\ &\hat{\delta}_7 = \delta_5\delta_7\beta_7/15, \\ &\hat{\delta}_n = \delta_n\delta_{n-2}\beta_n/3(n-2)(n-3), \quad \forall n \geqslant 9. \end{aligned} \tag{2.21a}$$

(iv) Tree dominance: $\forall n \geqslant 3$, $\exists\, 0 \leqslant \varepsilon_n(\Lambda) < 1$ such that

$$H_0^{n+1} = (1 - \varepsilon_n(\Lambda)) C_0^{n+1}, \tag{2.22}$$

with

$$\begin{aligned} &1 - \varepsilon_3(\Lambda) = \delta_3/6\Lambda(1 + \delta_1\Lambda), \\ &1 - \varepsilon_5(\Lambda) = \delta_5/15\Lambda(1 + \delta_1\Lambda), \\ &1 - \varepsilon_n(\Lambda) = \delta_n/3\Lambda n(n-1)(1 + \delta_1\Lambda). \end{aligned} \tag{2.22a}$$

Proof: (i) The sign properties are a direct consequence of the hypothesis $H_0 \in \Phi_{0\Lambda}$.

(ii) The "proportionality" of $B_0^{n+1} \sim C_0^{n+1}$, in the explicit form (2.19), is obtained by application of (i), of (1.4) and (2.11), and of Lemma 2.4 above.

(iii) Taking into account the hypothesis $H_0 \in \Phi_{0\Lambda}$ (sign properties) and (1.5) we write

$$H_0^4 = -\delta_3(\Lambda)(1 + \delta_1\Lambda)^2, \quad H_0^6 = 4\delta_3\delta_5[H_0^2]^3, \tag{2.23a}$$

and $\forall n \geqslant 7$,

$$H_0^{n+1} = \delta_n\delta_{n-2}\beta_n\beta_{n-2}H_0^{n-3}[H_0^2]^2. \tag{2.23b}$$

For $n = 3$ we then put $\hat{\delta}_3 = \Lambda\delta_3(1 + \delta_1\Lambda)^2$, which proves (2.20). Moreover, by the definitions (1.3) and (i) above we have

$$C_0^4 = -6\Lambda[H_0^2]^3, \tag{2.24a}$$

and $n \geqslant 7$,

$$C_0^{n-1} = -6\Lambda\theta_{n-4,1,1}^{n-2}\beta_{n-2}H_0^{n-3}[H_0^2]^2 \tag{2.24b}$$

[cf. (1.3a) for $\theta_{n-4,1,1}^{n-2}$]. Using (2.23) and (2.24) in the right-hand side of the definition

$$\hat{\delta}_n(\Lambda) = -\Lambda H^{n+1}/C^{n-1} \tag{2.25}$$

yields the equalities (2.21a). Q.E.D.

Now taking into account the properties $0 < \delta_n \leqslant \delta_\infty^\Lambda$ ($\delta \in \overline{\mathscr{C}}_\Lambda$) and $1 \leqslant \beta_n \leqslant \overline{\mathscr{T}}_n < n^2$ (by Lemma 2.3) we obtain that $\forall n$, $0 < \hat{\delta}_n < +\infty$. Q.E.D.

(iv) In an analogous way, by the hypothesis $H_0 \in \Phi_{0\Lambda}$ (signs and splitting formulas) and using the above equalities (2.24) we obtain (2.22) and (2.22a). Application of the absolute upper and lower bounds $\mathscr{C}_\Lambda$ (ii) of the splitting sequences $\delta \in \overline{\mathscr{C}}_\Lambda$ on the right-hand side of (2.22a) yields that $\forall n \geqslant 3$,

$0 < 1 - \varepsilon_n \leqslant 1.$ Q.E.D.

B. Two- (or one-) dimensional problem

We start this part of the section concerning the two- (or one-)dimensional problem by introducing the following sequences of triplets of ratios of ΦC's which characterize every coherent sequence of ΦC's. These parameters are the fundamental quantities that "translate" the essential difference between zero and more than one dimension, i.e., the nontriviality of Φ convolutions, which are equal to 1 in zero dimensions.

Definition 2(d) (The triplet of Φ parameters $\{{}^\Phi a_n, {}^\Phi b_n, {}^\Phi c_n\}$ $\forall \Phi \in \mathscr{F}$ [cf. (1.2b)]): For every coherent sequence of ΦC's, $\{\Phi^{(\bar{n},n)}H^{n+1}\} \in \mathscr{F}$, we define

$${}^\Phi a_n \overset{\text{def}}{\equiv} [\Phi^{(\bar{n},n)}(H) N_3^{(n+2)}\tilde{]}_0 / [\Phi^{(\bar{n},n)}(H)\tilde{]}_0; \tag{2.26a}$$

$${}^\Phi b_{j_1 j_2}^n \overset{\text{def}}{\equiv} [\Phi^{(\bar{n},n)}(H) N_2^{(j_2)} N_1^{(j_1)}\tilde{]}_0 / [\Phi^{(\bar{n},n)}(H)\tilde{]}_0, \quad \forall w(J), \tag{2.26b}$$

with ${}^\Phi b_n \equiv b_{1,n-1}^n$, ${}^\Phi b_1 \equiv 0$;

$${}^\Phi c_{i_1 i_2 i_3}^n \overset{\text{def}}{\equiv} \left[\Phi^{(\bar{n},n)}(H) \prod_{l=1}^{3} N_1^{(i_l)}\right]_0 \{[\Phi^{(\bar{n},n)}(H)\tilde{]}_0\}^{-1}, \quad \forall w(i_1 i_2 i_3), \tag{2.26c}$$

with ${}^\Phi c_n \equiv {}^\Phi c_{n-2,1,1}^n$, ${}^\Phi c_1 \equiv 0$. The subscript 0 means at zero external momenta.

Notice that *a priori* the above ΦC's are not completely reduced with respect to all corresponding functions H^{i_l+1} (resp. bubble vertices of $\tilde{G}_\Phi$); but the notation [$\tilde{]}$ used in the right-hand side of Eqs. (2.26) means at least reduced with respect to the "last" function (bubble vertex) H^{n+1} or H^{i_l+1}, appearing in the definition of the corresponding ΦC (cf. Sec. II of I for a review of the notation). We call $\{{}^\Phi a_n, {}^\Phi b_n, {}^\Phi c_n\}$ the triplet sequence of Φ parameters associated with a given coherent sequence of ΦC's, $\Phi \in \mathscr{F}$.

The completely reduced ΦC's (resp. the associated triplets of Φ parameters) satisfy some positivity and convergence properties (resp. lower and upper absolute bounds) that are useful for the proof of all statements below. We state these properties in the following proposition.

Proposition 2.2: Every completely reduced ΦC $\tilde{\Phi}_0^{(\bar{n},n)}$ satisfies the following properties.

$$\text{(i)}\ (-1)^{|V_n|}\tilde{\Phi}_0^{(\bar{n},n)} > 0; \tag{2.27}$$

$$\text{(ii)}\ |\tilde{\Phi}_0^{(\bar{n},n)}| < +\infty; \tag{2.28}$$

(iii) $\forall n = 1,3,5,\ldots$ [resp. $\forall w(j_1 j_2), w(I)$ of n] $\exists$ positive finite parameters

$\{\underline{a}_n,\underline{b}_n,\underline{c}_n\}, \quad \{\bar{a}_n,\bar{b}_n,\bar{c}_n\}$
(resp. $\{\underline{a}_n,\underline{b}^{\,n}_{j_1j_2},\underline{c}^n_{i_1i_2i_3}\}, \ \{\bar{a}_n,\bar{b}^{\,n}_{j_1j_2},\bar{c}^n_{i_1i_2i_3}\}$)
associated with $n(w(J),w(I))$ and such that

$$\sup_{\Phi} {}^{\Phi}a_n \leqslant \bar{a}_n, \quad \inf_{\Phi} {}^{\Phi}a_n \geqslant \underline{a}_n > 0, \tag{2.29a}$$

$$\sup_{\Phi} {}^{\Phi}b_n \leqslant \bar{b}_n, \quad \inf_{\Phi} {}^{\Phi}b_n \geqslant \underline{b}_n$$

$$(\text{resp. } \sup_{\Phi} {}^{\Phi}b^{\,n}_{j_1j_2} \leqslant \bar{b}^{\,n}_{j_1j_2}, \quad \inf_{\Phi} {}^{\Phi}b^{\,n}_{j_1j_2} \geqslant \underline{b}^{\,n}_{j_1j_2}), \tag{2.29b}$$

$$\sup_{\Phi} {}^{\Phi}c_n \leqslant \bar{c}_n, \quad \inf_{\Phi} {}^{\Phi}c_n \geqslant \underline{c}_n$$

$$(\text{resp. } \sup_{\Phi} {}^{\Phi}c^n_{i_1i_2i_3} \leqslant \bar{c}^n_{i_1i_2i_3}, \ \inf_{\Phi} {}^{\Phi}c^n_{i_1i_2i_3} \geqslant \underline{c}^n_{i_1i_2i_3}). \tag{2.29c}$$

Proof: (i) From the positivity of the integrand (product of free propagators) and by application of a fundamental statement on the spherical integrals, we obtain the positive sign of every multiple integration corresponding to a completely reduced ΦC multiplied by the sign of the dynamical vertices $(-1)^{|V_n|}$. Parts (ii) and (iii) are trivially obtained in view of the absolute convergence of the integrations (cf. Proposition 2.1 of I).

Definition 2(a2) (The subset $\widetilde{\mathscr{C}}_\Lambda$): We consider the family of sets $\mathscr{C}_\Lambda$ of Definition 2(a), and by analogy with Definition 2(a1) of $\widetilde{\mathscr{C}}_\Lambda$ we specify the subset $\widetilde{\mathscr{C}}_\Lambda(\tilde{\delta}^\Lambda_\infty,\tilde{I}_n,\widetilde{S}_n) \subset \mathscr{B}_\delta$. We suppose that a finite positive constant $\tilde{\delta}^\Lambda_\infty$ is given; then the associated sequence of pairs $(\tilde{I}_n,\widetilde{S}_n)$ is defined by the analogs of (2.4):

$$\begin{aligned}\tilde{I}_3 &= [1 + 9\Lambda\bar{b}_3(1+6\Lambda^2\bar{a}_1)]^{-1},\\ \widetilde{S}_1 &= 4[1-24\Lambda^2S_3\bar{a}_1]^{-1/2}[1+(1-24\Lambda^2S_3\bar{a}_1)^{1/2}];\end{aligned} \tag{2.30a}$$

$$\begin{aligned}\widetilde{S}_3 &= (1+6\Lambda^2\widetilde{S}_1)/(1+9\Lambda\underline{b}_3 - 60\Lambda^2\bar{a}_3), \quad \widetilde{S}_5 = 1,\\ \tilde{I}_5 &= (1-3\Lambda\bar{b}_5\widetilde{S}_3)\{1+15\Lambda\bar{b}_5(1+6\Lambda^2\bar{a}_1)\}^{-1};\end{aligned} \tag{2.30b}$$

$$\begin{aligned}\tilde{I}_n &= 2\{2+3\Lambda n(n-1)\bar{b}_n(1+6\Lambda^2\bar{a}_1)\}^{-1},\\ \widetilde{S}_n &= \tilde{\delta}^\Lambda_\infty(1+6\Lambda^2\bar{a}_1)/[\tilde{\delta}^\Lambda_\infty + 3\Lambda n(n-1)], \quad \forall n \geqslant 7.\end{aligned} \tag{2.30c}$$

We then state without proof the analog of Lemma 2.1. The proof goes exactly in the same way as in Lemma 2.1 by defining the corresponding "minimal" sequence, $\tilde{\delta}_n = 3\Lambda n(n-1)\tilde{I}_n$.

Lemma 2.1′: The subset $\widetilde{\mathscr{C}}_\Lambda(\tilde{\delta}^\Lambda_\infty,\tilde{I}_n,\widetilde{S}_n)$ of Definition 2(a2) is nonempty for every fixed Λ satisfying $0<\Lambda \lesssim 0.1(\underline{b}_3/\bar{a}_3)$.

Now we proceed to the precise description of the two-dimensional analog of the subset $\Phi_{0\Lambda}$.

Definition 2(e) (The subset $\Phi_\Lambda \subset \mathscr{B}$): Let $H\in\mathscr{B}$; we shall say that H belongs to $\Phi_\Lambda \subset \mathscr{B}$ iff for every coherent sequence of ΦC's $\{\Phi^{(\bar{n},n)}(H)H^{n+1}\}_n \in \mathscr{F}$ the following properties hold.

(1) *Signs, splitting at zero external momenta:* There exist ${}^{\Phi}\delta(\Lambda)\in\widetilde{\mathscr{C}}_\Lambda$ and ${}^{\Phi}H_0\in\mathscr{B}_0$ such that $\forall n = 1,3,5,\ldots,$

$$\{\Phi^{(\bar{n},n)}(H)H^{n+1}\}_0 = {}^{\Phi}H_0^{n+1}[\Phi^{(\bar{n},n)}(H)]_0, \tag{2.31}$$

with

$$\begin{aligned}{}^{\Phi}H_0^2 &= 1 + {}^{\Phi}\delta_1\Lambda, \quad {}^{\Phi}H_0^4 = -{}^{\Phi}\delta_3[{}^{\Phi}H_0^2]^2,\\ {}^{\Phi}H_0^6 &= -4{}^{\Phi}\delta_5{}^{\Phi}H_0^4{}^{\Phi}H_0^2,\end{aligned} \tag{2.31a}$$

and $\forall n \geqslant 7$,

$${}^{\Phi}H_0^{n+1} = -{}^{\Phi}\delta_n{}^{\Phi}\beta_n{}^{\Phi}H_0^{n-1}{}^{\Phi}H_0^2.$$

Here the sweeping factors ${}^{\Phi}\beta_n$ are defined by Eq. (1.8) [or (3.10) of I], and the superscript Φ denotes the corresponding (first- or second-type) coherent sequence.

(2) *Positivity of the $H^2(q,\Lambda)$ and "negativity" of the $H^4(q,\Lambda)$ functions:*

$$\text{(a)} \ H^2(q\Lambda)\Delta_F(q) > 1, \quad \forall q\in\mathscr{E}^2_{(q)}, \tag{2.32a}$$

$$\text{(b)} \ H^4(q,\Lambda) < 0, \quad \forall q\in\mathscr{E}^6_{(q)}. \tag{2.32b}$$

Remark 1: By comparison of Φ_Λ with the analogous Definition 2(b) of $\Phi_{0\Lambda}$, we notice that in the above definition both signs and splitting properties are expressed in a unified way. Moreover, the supplementary condition appearing now in the structure of the splitting (and which is trivially absent in the definition of $\Phi_{0\Lambda}$), i.e., the factorization of (at least partially) reduced ΦC's, stems from the dimensional character of the space $\mathscr{B}$ in opposition to the zero-dimensional $\mathscr{B}_0$.

Remark 2: The "dimensionality" of $\mathscr{B}$ is also the reason for the presence of two supplementary bits of information in Φ_Λ. They deal with the positive and negative sign of the two-point function and the four-point function, respectively, at every value of the external momenta ($q\in\mathscr{E}^2_{(q)}$ and $q\in\mathscr{E}^6_{(q)}$, respectively). Both properties have been revealed by the Φ iteration as we analyzed in Sec. II of I.

Now if we suppose that $^*\delta\in\widetilde{\mathscr{C}}_\Lambda$, then a sequence $^*H\in\mathscr{B}_0$ constructed by the analog of (2.10) (cf. Lemma 2.4) belongs also trivially to Φ_Λ (and to $\mathscr{B}$!). So we automatically have shown the following lemma to be true.

Lemma 2.4′: The subset Φ_Λ is nonempty.

By the following statements we present recursive definitions, signs, and bounds for the sweeping factors ${}^{\Phi}\beta_n, {}^{\Phi}\alpha_n$, and the global terms for the "zero-dimensional-type" equations. These properties are extensively used in the proofs of Sec. IV.

Lemma 2.6: Let $H\in\Phi_\Lambda$; for every coherent sequence of ΦC's the corresponding sweeping factors ${}^{\Phi}\beta_n, {}^{\Phi}\beta^n_{i_1i_2i_3}$ introduced by Eqs. (1.8) satisfy the following recurrence relations and absolute bounds:

(i) ${}^{\Phi}\beta_3 = 1 = {}^{\Phi}\beta_5$,

$${}^{\Phi}\beta_n = {}^{\Phi}\beta^n_{n-2,1,1}, \quad \forall n \geqslant 7$$

and for every partition $w_n(i_1i_2i_3)$ with $i_1 \geqslant i_2 \geqslant i_3$,

$$
{}^{\Phi}\beta^{n}_{i_1i_2i_3} = 1 + \zeta(i_1i_2)\frac{\theta^{n}_{i_1-2,i_2+2,i_3}}{\theta^{n}_{i_1i_2i_3}}\frac{{}^{\Phi}\delta_{i_2+2}}{{}^{\Phi}\delta_{i_1}}\frac{z({}^{\Phi}\beta_{i_2+2})}{z({}^{\Phi}\beta_{i_1})}\frac{{}^{\Phi}c^{n}_{i_1-2,i_2+2,i_3}}{{}^{\Phi}c^{n}_{i_1i_2i_3}}\,{}^{\Phi}\beta^{n}_{i_1-2,i_2+2,i_3}
$$
$$
+\zeta(i_2i_3)\frac{\theta^{n}_{i_1,i_2-2,i_3+2}}{\theta^{n}_{i_1i_2i_3}}\frac{{}^{\Phi}\delta_{1_3+2}}{{}^{\Phi}\delta_{i_2}}\frac{z({}^{\Phi}\beta_{i_3+2})}{z({}^{\Phi}\beta_{i_2})}\frac{{}^{\Phi}c^{n}_{i_1,i_2-2,i_3+2}}{{}^{\Phi}c^{n}_{i_1i_2i_3}}\,{}^{\Phi}\beta^{n}_{i_1,i_2-2,i_3+2}
$$
$$
+\zeta(i_1i_3)\frac{\theta^{n}_{i_1-2,i_2,i_3+2}}{\theta^{n}_{i_1i_2i_3}}\frac{{}^{\Phi}\delta_{i_3+2}}{{}^{\Phi}\delta_{i_1}}\frac{z(\beta_{i_3+2})}{z({}^{\Phi}\beta_{i_1})}\frac{{}^{\Phi}c^{n}_{i_1-2,i_2,i_3+2}}{{}^{\Phi}c^{n}_{i_1i_2i_3}}\,{}^{\Phi}\beta^{n}_{i_1-2,i_2,i_3+2}, \tag{2.33}
$$

where $z(\beta_i)$, $\zeta(i_l i_{l'})$, $l = 1,2,3, = l'(l \neq l')$ are defined by (2.6a)–(2.6c), respectively;

(ii) $(\tilde{Y}_n(\Lambda)/\underline{c}_n)\mathscr{T}_n \leqslant {}^{\Phi}\beta_n \leqslant \mathscr{T}_n(\tilde{Y}_n(\Lambda)/\bar{c}_n)$. (2.34)

Here the quantities $\tilde{Y}_n(\Lambda), \tilde{Y}_n(\Lambda)$ are independent on the particular ΦC, and they satisfy the analogous properties (2.9a)–(2.9c) of $Y_n(\Lambda), \bar{Y}_n(\Lambda)$. The factor $\mathscr{T}_n$ is again the number of different possible partitions $w(i_1i_2i_3)$ of n (or different terms inside ${}^{\Phi}C_0^{n+1}$), and it is explicitly calculated in Appendix B of III.

The proof of this statement goes exactly through the same arguments we have used for the proof of Lemma 2.3, and so we do not present it. We only notice the difference between (2.33) and (2.6) due to the presence of the constants ${}^{\Phi}c^{n}_{i_1i_2i_3}$, which precisely express the nontrivial character of the convergent integrals in the two- (or one-)dimensional momentum space [cf. Definition 2(d)].

Definition 2(f) (The sweeping factors ${}^{\Phi}\alpha_n$ for ${}^{\Phi}B_0^{n+1}$): By the analog of (2.11) we define the sweeping factors ${}^{\Phi}\alpha_n$, corresponding to the sum ${}^{\Phi}B_0^{n+1}$ [cf. Eq. (b)] for every coherent sequence $\Phi \in \mathscr{F}$ and $\forall H_0 \in \mathscr{B}_0$:

$$
|{}^{\Phi}B_0^{n+1}| = 3\Lambda{}^{\Phi}\alpha_n n|{}^{\Phi}H_0^{n+1}|\,|{}^{\Phi}H_0^2|\{\Phi^{(\bar{n},n)}(H_0)N_2^{(j_2)}N_1^{(j_1)}\}_0. \tag{2.35}
$$

Moreover, a definition analogous to (2.12) holds for the corresponding sweeping factor ${}^{\Phi}\alpha^{n}_{j_1j_2}$ associated with a fixed partition $w(j_1j_2)$. Using exactly analogous arguments as for the demonstration of Lemma 2.5 one proves an explicit recursive procedure for the ${}^{\Phi}\alpha^{n}_{j_1j_2}$'s and the corresponding upper and lower bounds. We also present these results without proof (cf. III) in the following lemma. We notice again the presence of the parameters ${}^{\Phi}b_n$, ${}^{\Phi}b^{n}_{j_1j_2}$ [cf. Definition 2(d)].

Lemma 2.7: Let $H \in \Phi_\Lambda$; then $\forall n \geqslant 3$ and for every fixed partition $w(J)$, the sweeping factors ${}^{\Phi}\alpha_n, {}^{\Phi}\alpha^{n}_{j_1j_2}$ [Definition 2(f)] corresponding to a given coherent sequence of ΦC's satisfy

(i) ${}^{\Phi}\alpha^{3}_{1,2} \equiv {}^{\Phi}\alpha_3 = 1$ and ${}^{\Phi}\alpha_n \equiv {}^{\Phi}\alpha^{n}_{1,n-1}$, (2.36a)

$$
{}^{\Phi}\alpha^{n}_{(n+1)/2,(n-1)/2} = 1, \quad \text{if} \quad (n+1)/2 = \text{odd}, \tag{2.36b}
$$

$$
{}^{\Phi}\alpha^{n}_{(n-1)/2,(n+1)/2} = 1 + \frac{(n-1)}{(n+3)}\frac{{}^{\Phi}b^{n}_{(n+3)/2,(n-3)/2}}{{}^{\Phi}b^{n}_{(n-1)/2,(n+1)/2}}, \quad \text{if } (n+1)/2 = \text{even}, \tag{2.36c}
$$

and for $(n+1)/2 < j_2 \leqslant n-1$,

$$
{}^{\Phi}\alpha^{n}_{j_1j_2} = 1 + \frac{j_2(j_2-1)}{(j_1+2)(j_1+1)}\frac{z({}^{\Phi}\beta_{j_1+2})}{z({}^{\Phi}\beta_{j_2+1})}\frac{{}^{\Phi}\delta_{j_1+2}}{{}^{\Phi}\delta_{j_2+1}}\frac{{}^{\Phi}b_{j_1+2,j_2-2}}{{}^{\Phi}b_{j_1j_2}}\,{}^{\Phi}\alpha^{n}_{j_1+2,j_2-2}
$$
$$
+\frac{j_1 s(j_1-1)}{j_2+1}\frac{{}^{\Phi}b^{n}_{j_2+1,j_1-1}}{{}^{\Phi}b^{n}_{j_1j_2}}; \tag{2.37}
$$

(ii) $\forall n \geqslant 7\ \exists$ finite positive quantities $\tilde{X}_n(\Lambda), \bar{\bar{X}}_n(\Lambda)$ such that

$$
\bar{\bar{X}}_n\mathscr{T}_n/\underline{b}_n \leqslant {}^{\Phi}\alpha_n \leqslant \mathscr{T}_n\tilde{X}_n/\bar{b}_n \quad [\text{with } \mathscr{T}_n = (n-1)/2]. \tag{2.38}
$$

Here again $\bar{\bar{X}}_n(\Lambda), \tilde{X}_n(\Lambda)$ satisfy exactly analogous properties with $X_n, \bar{X}_n$ of Lemma 2.5.

We close this section with the presentation of the analog of Proposition 2.1. All these properties—*signs, double splitting,* and *tree dominance*—concerning the global terms ${}^{\Phi}A_0^{n+1}, {}^{\Phi}B_0^{n+1}, {}^{\Phi}C_0^{n+1}$ when $H \in \Phi_\Lambda$ (for a given coherent sequence of ΦC's) can be easily proved in the same way as we explained for the zero-dimensional case. So we state them without proof in the proposition below.

Proposition 2.3: Let $H \in \Phi_\Lambda$; for a given coherent sequence of ΦC's, $\Phi \in \mathscr{F}$, the corresponding "global" terms ${}^{\Phi}A_0^{n+1}, {}^{\Phi}B_0^{n+1}, {}^{\Phi}C_0^{n+1}$ defined in Eqs. (a)–(c) satisfy the following properties:

(i)

$$
\begin{aligned}
&(-1)^{|V_n|}(-1)^{(n-1)/2}{}^{\Phi}C_0^{n+1} > 0,\\
&(-1)^{|V_n|}(-1)^{(n-1)/2}{}^{\Phi}B_0^{n+1} < 0,\\
&(-1)^{|V_n|}(-1)^{(n-1)/2}{}^{\Phi}A_0^{n+1} > 0;
\end{aligned} \tag{2.39}
$$

(ii)

$$
{}^{\Phi}B_0^{n+1} = -{}^{\Phi}\vartheta_n{}^{\Phi}\delta_n{}^{\Phi}C_0^{n+1}, \tag{2.40}
$$

where

$$
{}^{\Phi}\vartheta_3 = \tfrac{3}{2}{}^{\Phi}b_3/{}^{\Phi}c_3, \quad {}^{\Phi}\vartheta_5 = {}^{\Phi}\alpha_5{}^{\Phi}b_5/{}^{\Phi}c_5, \tag{2.40a}
$$

and $\forall n \geqslant 7$,

$$
{}^{\Phi}\vartheta_n = {}^{\Phi}\alpha_n{}^{\Phi}b_n/(n-1){}^{\Phi}c_n, \text{ with } 0 < {}^{\Phi}\vartheta_n \leqslant \tilde{X}_n/2\underline{c}_n; \tag{2.40b}
$$

(iii) (double splitting) $\forall n \geqslant 1$,

$$
{}^{\Phi}A_0^{n+1} = {}^{\Phi}\hat{\delta}_{n+2}{}^{\Phi}C_0^{n+1} \quad (\text{notice } {}^{\Phi}C_0^2 = 1), \tag{2.41}
$$

with

$$ {}^{\Phi}\hat{\delta}_3 = \Lambda^{\Phi}\delta_3(1 + {}^{\Phi}\delta_1\Lambda)^{2\Phi}a_1, \quad {}^{\Phi}\hat{\delta}_5 = \tfrac{2}{3}{}^{\Phi}\delta_3{}^{\Phi}\delta_5{}^{\Phi}a_3/{}^{\Phi}c_3, $$

$$ {}^{\Phi}\hat{\delta}_7 = {}^{\Phi}\delta_5{}^{\Phi}\delta_7{}^{\Phi}\beta_7{}^{\Phi}a_5/15{}^{\Phi}c_5, \tag{2.41a} $$

and

$$ \hat{\delta}_{n+2} = {}^{\Phi}\delta_n{}^{\Phi}\delta_{n+2}{}^{\Phi}\beta_{n+2}{}^{\Phi}a_n/3n(n-1){}^{\Phi}c_n, \quad \forall n \geqslant 9; \tag{2.41b} $$

(iv) (tree dominance) $\forall n \geqslant 3\ \exists\ 0 \leqslant {}^{\Phi}\varepsilon_n < 1$,

$$ \Phi^{(\bar{n},n)}(H)H^{n+1}I_0 = (1 - {}^{\Phi}\varepsilon_n){}^{\Phi}C_0^{n+1}, \tag{2.42} $$

with

$$ 1 - {}^{\Phi}\varepsilon_3 = {}^{\Phi}\delta_3/(H\,{}^{\Phi}\delta_1\Lambda)6\Lambda{}^{\Phi}c_3, $$

$$ 1 - {}^{\Phi}\varepsilon_5 = {}^{\Phi}\delta_5/(1 + {}^{\Phi}\delta_1\Lambda)15\Lambda{}^{\Phi}c_5, \tag{2.42a} $$

$$ (1 - {}^{\Phi}\varepsilon_n) = {}^{\Phi}\delta_n/(1 + {}^{\Phi}\delta_1\Lambda)3\Lambda n(n-1){}^{\Phi}c_n. $$

III. THE ZERO-DIMENSIONAL SOLUTION

In this section we construct a nontrivial solution $\overline{H}_0$ for the zero-dimensional system (1.2) (i.e., a fixed point of the mapping $\mathscr{M}_0$) that satisfies all good properties of signs and splitting, which we have obtained from the Φ iteration in I (Ref. 3) and which we formulated in the preceding section in terms of Definition 2(b) of the subset $\Phi_{0\Lambda} \subset \mathscr{B}_0$.

The most important ingredient of the method developed in this section in order to obtain the complete answer to the zero-dimensional problem is the subset $\overline{\mathscr{C}}_\Lambda \subset \mathscr{B}_\delta$ that we introduced in Sec. II [cf. Definition 2(a1)], which contains all the fine characteristics, i.e., limit values, and absolute and relative bounds of the splitting sequences. This fact becomes evident in the study that follows which contains three essential steps.

(1) We first establish an equivalence (cf. Theorem 3.1 below) between the existence of a unique nontrivial solution of (1.2), which belongs to $\Phi_{0\Lambda}$, and the existence and uniqueness of a nontrivial solution of a system described by (3.1a)–(3.1d), defined in the space $\mathscr{B}_\delta$ of the splitting sequences, that belongs to $\overline{\mathscr{C}}_\Lambda \subset \mathscr{B}_\delta$.

(2) The problem is then transformed by (1) into a fixed-point problem in the space $\mathscr{B}_\delta$ that we solve (under appropriate conditions on Λ) by Theorems 3.2 and 3.3 below, in two steps.

(a) We show the stability of $\overline{\mathscr{C}}_\Lambda$ (and $\overline{\mathscr{C}}_\Lambda^{\prime\prime}$), under the action of the mapping $\mathscr{M}_\delta$ described by the system (3.1) mentioned above, when Λ satisfies $0 < \Lambda \lesssim 0.1$ (and $0 < \Lambda \lesssim 0.01$). This constitutes the most crucial step. Moreover, the existence of a solution to (3.1) is ensured by application of the Leray–Shauder theorem.

(b) The contractivity of $\mathscr{M}_\delta$ (application of the contractive mapping principle) is proved inside $\overline{\mathscr{C}}_\Lambda^{\prime\prime}$, and so the uniqueness the solution is also obtained inside this subset. We only note that for (b) we use essentially (a) (absolute bounds norm), but now a supplementary condition is imposed on Λ $(0 < \Lambda \lesssim 0.01)$.

(3) Finally, by Corollary 3.1, which combines steps (1) and (2), we obtain the solution of (1.2).

Theorem 3.1: There is an equivalence between the following hypotheses 1 and 2 $\forall \Lambda$ satisfying $0 < \Lambda \lesssim 0.1$.

Hypothesis 1: The system (3.1) below has a unique nontrivial solution $\{\bar{\delta}_n(\Lambda)\}_n$ that belongs to $\overline{\mathscr{C}}_\Lambda$ [Definitions 2(a) and (2a1)]:

$$ \delta_1(\Lambda) = \frac{1 - 2\delta_3\Lambda - [1 - 4\delta_3\Lambda]^{1/2}}{2\delta_3\Lambda^2}, \tag{3.1a} $$

$$ \delta_3(\Lambda) = \frac{6\Lambda(1 + \delta_1\Lambda)}{1 + 6\Lambda(1 + \delta_1\Lambda)[\frac{3}{2} - \frac{2}{3}\delta_5]}, \tag{3.1b} $$

$$ \delta_5(\Lambda) = \frac{15\Lambda(1 + \delta_1\Lambda)}{1 + 15\Lambda(1 + \delta_1\Lambda)\Delta_5}, $$

$$ \text{with } \Delta_5(\Lambda) \equiv \vartheta_5 - \frac{\delta_7\beta_7}{15}, \tag{3.1c} $$

and $\forall n \geqslant 7$,

$$ \delta_n(\Lambda) = \frac{3\Lambda n(n-1)(1 + \delta_1\Lambda)}{1 + 3\Lambda n(n-1)(1 + \delta_1\Lambda)\Delta_n(\Lambda)}, $$

$$ \text{with } \Delta_n(\Lambda) \equiv \frac{\alpha_n}{n-1} - \frac{\delta_{n+2}\beta_{n+2}}{3n(n-1)}. \tag{3.1d} $$

(We omit very often the argument Λ for simplicity.) Here $\vartheta_5, \beta_n, \alpha_n$ are defined by Proposition 2.1 and Lemmas 2.2 and 2.5, respectively.

Hypothesis 2: The system (1.2) has one and only one nontrivial solution $\overline{H}$: $\mathscr{M}_0(\overline{H}) = \overline{H}$, which belongs to $\Phi_{0\Lambda}$.

Proof: Let us suppose that $\bar{\delta} \equiv \{\bar{\delta}_1, \bar{\delta}_3, ..., \bar{\delta}_n\}$ is a solution of the system (3.1) and that $\bar{\delta} \in \overline{\mathscr{C}}_\Lambda$. We define the following sequence:

$$ \overline{H} = \{\overline{H}^{n+1}(\Lambda)\}_{n=1,3,5,...}, $$

with

$$ \overline{H}^2 = 1 + \bar{\delta}_1\Lambda, \quad \overline{H}^4 = -\bar{\delta}_3[\overline{H}^2]^2, \quad \overline{H}^6 = -4\bar{\delta}_5\overline{H}^4\overline{H}^2, \tag{3.2} $$

and by recursion, $\forall n \geqslant 7$,

$$ \overline{H}^{n+1} = -\bar{\delta}_n\bar{\beta}_n\overline{H}^{n-1}\overline{H}^2. $$

In the above definition the sequences $\bar{\beta}_n(\bar{\delta}_n)$ are given recurrently in terms of all $\bar{\delta}_{\bar{n}}$ with $\bar{n} \leqslant n-2$ by Lemma 2.2.

(1a) We notice that the sequence $\overline{H}$ of (3.2) coincides with Definition (2.10). In view of the assumption $\bar{\delta} \in \overline{\mathscr{C}}_\Lambda$ and following Lemma 2.4, that means $\overline{H} \in \Phi_{0\Lambda}$. Q.E.D.

(1b) We shall prove that $\mathscr{M}_0$ keeps invariant $\overline{H}$ [or that $\overline{H}$ satisfies the system (1.2)], i.e.,

$$ \forall n = 1,3,5,..., \quad \overline{H}^{(n+1)\prime} = \overline{H}^{n+1}. \tag{3.3} $$

For $n = 1$, following Eq. (1.2a) and using the above result 1(a), we can apply Proposition 2.1 [in particular, Eq. (2.20)] and write

$$ H^{2\prime} = 1 + \Lambda\bar{\delta}_3(1 + \bar{\delta}_1\Lambda)^2. \tag{3.4} $$

We now replace the left-hand side of (3.3) for $(n = 1)$ by (3.4), and the corresponding right-hand side by Eqs. (3.2) of $n = 1$. We obtain

$$ 1 + \Lambda\bar{\delta}_3(1 + \bar{\delta}_1\Lambda)^3 = 1 + \bar{\delta}_1\Lambda. \tag{3.5} $$

The analytic solution, when $\Lambda \to 0$, of this second-order equation, with respect to $\bar{\delta}_1$ is the corresponding negative root,

$$ \bar{\delta}_1 = (1 - 2\bar{\delta}_3\Lambda - [1 - 4\bar{\delta}_3\Lambda]^{1/2})/2\bar{\delta}_3\Lambda^2. \tag{3.6} $$

But this equation holds in view of hypothesis 1 [cf. Eq. (3.1a)]. Q.E.D.

For $n \geqslant 3$ we obtain in an analogous way, by using [in view of the result (1a)] Proposition 2.1, Parts (i)–(iii), in the right-hand side of Eq. (1.2b),

$$H^{(n+1)\prime} = [(1 - \bar{\vartheta}_n \bar{\delta}_n + \hat{\delta}_{n+2})/(1 - \bar{\varepsilon}_n)]\bar{H}^{n+1}. \tag{3.7}$$

We recall that following Proposition 2.1, the above functionals $\bar{\vartheta}_n, \hat{\delta}_{n+2}, 1 - \bar{\varepsilon}_n$ are explicitly given in terms of the $\bar{\delta}_n$'s by the formulas (2.19a), (2.21a), and (2.22a), respectively (cf. Proposition 2.1).

Comparison of (3.7) with (3.3) yields that, for every $n \geqslant 3$, requiring stability is equivalent to ensuring

$$1 - \bar{\varepsilon}_n = 1 - \bar{\vartheta}_n \bar{\delta}_n + \hat{\delta}_{n+2}. \tag{3.8}$$

By insertion of the explicit formulas of $\bar{\vartheta}_n, \hat{\delta}_{n+2}, 1 - \bar{\varepsilon}_n$ into (3.8) $\forall n \geqslant 3$ and by solving each time with respect to the corresponding $\bar{\delta}_n$, we obtain after some simple manipulations Eqs. (3.1b)–(3.1d), respectively. In other words, the stability condition is ensured iff $\bar{\delta}$ satisfies the system (3.1). But this is exactly the hypothesis 1; so $\bar{H}$ of (3.2) is a solution of (1.2), or a fixed point of $\mathscr{M}_0$. Q.E.D.

1(c) We now show the uniqueness of this solution $\bar{H}$. Let us suppose that there exists another solution $\bar{\bar{H}}$ of (1.2) such that it also belongs to $\Phi_{0\Lambda}$. Then there would exist another sequence $\bar{\bar{\delta}} \equiv \{\bar{\bar{\delta}}_n\}_n$ in $\mathscr{B}_\delta$ such that

$$\bar{\bar{\delta}}_1 = \frac{\bar{\bar{H}}^2 - 1}{\Lambda}, \quad \bar{\bar{\delta}}_3 = -\frac{\bar{\bar{H}}^4}{[\bar{\bar{H}}^2]^2}, \quad \bar{\bar{\delta}}_5 = -\frac{\bar{\bar{H}}^6}{4\bar{\bar{H}}^4\bar{\bar{H}}^2},$$
$$\bar{\bar{\delta}}_n = -\bar{\bar{H}}^{n+1}/\beta_n \bar{\bar{H}}^{n-1}\bar{\bar{H}}^2, \quad \forall n \geqslant 7, \tag{3.9}$$

with

$$\bar{\bar{\beta}}_n = |\bar{\bar{C}}^{n+1}|/3\Lambda n(n-1)|\bar{\bar{H}}^{n-1}|\,\bar{\bar{H}}^2|^2. \tag{3.10}$$

Moreover, this sequence $\bar{\bar{\delta}} \in \overline{\mathscr{C}}_\Lambda$. On the other hand, using the stability hypothesis $\bar{\bar{H}}^{(n+1)\prime} = \bar{\bar{H}}^{n+1}$ we obtain by analogous arguments

$$1 + \bar{\bar{\delta}}_3\Lambda(1 + \bar{\bar{\delta}}_1\Lambda)^2 = 1 + \bar{\bar{\delta}}_1\Lambda$$

and

$$1 - \bar{\bar{\vartheta}}_n \bar{\bar{\delta}}_n + \hat{\delta}_{n+2} = 1 - \bar{\bar{\varepsilon}}_n, \quad \forall n \geqslant 3. \tag{3.11}$$

Here $\bar{\bar{\vartheta}}_n = \bar{\vartheta}_n(\bar{\bar{\delta}}_n)$, $\hat{\delta}_{n+2} \equiv \hat{\delta}_{n+2}(\bar{\bar{\delta}}_n)$, $1 - \bar{\bar{\varepsilon}}_n \equiv (1 - \bar{\varepsilon}_n)(\bar{\bar{\delta}}_n)$ are functionals of $\bar{\bar{\delta}}_n, \bar{\bar{\beta}}_n$ defined by the analogs of (2.19a), (2.21a), and (2.22a), respectively, so that Eqs. (3.11) are equivalent to the system (3.1). In other words, $\bar{\bar{\delta}}$ is also a solution of (3.1), which belongs to $\overline{\mathscr{C}}_\Lambda$. This result is in contradiction to hypothesis 1, which states that $\bar{\delta}$ is a unique solution of (3.1) inside $\overline{\mathscr{C}}_\Lambda$. Consequently the uniqueness of $\bar{H} \in \Phi_{0\Lambda}$ is proved.

(2) We now show the converse situation, i.e., from hypothesis 2 we obtain hypothesis 1. We suppose that $\bar{H}$ is a unique nontrivial solution of (1.2) in $\Phi_{0\Lambda}$ and prove that there exists a unique nontrivial solution of (3.1) in $\overline{\mathscr{C}}_\Lambda$. We define the following sequence $\tilde{\delta} \equiv \{\tilde{\delta}_n\}_{n=1,3,5,\ldots}$:

$$\tilde{\delta}_1\Lambda = \bar{H}^2 - 1, \quad \tilde{\delta}_3 = -\frac{\bar{H}^4}{[H^2]^2}, \quad \tilde{\delta}_5 = -\frac{\bar{H}^6}{4\bar{H}^4\bar{H}^2},$$
$$\tilde{\delta}_n = -\frac{\bar{H}^{n+1}}{\bar{H}^{n-1}\bar{H}^2\tilde{\beta}_n}, \quad \forall n \geqslant 7, \tag{3.12}$$

with

$$\tilde{\beta}_n = \frac{|\bar{C}^{n+1}|}{3\Lambda n(n-1)|\bar{H}^{n-1}|\,|\bar{H}^2|}. \tag{3.12a}$$

2(a) By the hypothesis $\bar{H} \in \Phi_{0\Lambda}$ and the above construction (3.12), it follows automatically that $\tilde{\delta} \in \overline{\mathscr{C}}_\Lambda$.

2(b) Using the stability hypothesis

$$\bar{H}^{(n+1)\prime} = \bar{H}^{n+1}, \tag{3.13}$$

we proceed in a way analogous to that presented in (1b) to find that $\tilde{\delta}_n$, $\forall n = 1,3,5,\ldots$, satisfies the corresponding equation of the system (3.1). Q.E.D.

2(c) By an argument similar to the one we used for the opposite case [cf. (1c)] we show also the uniqueness of this solution. Q.E.D.

This completes the proof of Theorem 3.1.

Definition 3(a) (The mapping $\mathscr{M}_\delta$ and the norm in $\mathscr{B}_\delta$): Let us consider the space $\mathscr{B}_\delta$ of the splitting sequences δ introduced by Definition 1(a) iv in the Introduction. The system (3.1) defines a nonlinear mapping $\mathscr{M}_\delta$ of $\mathscr{B}_\delta$ onto itself: $\mathscr{B}_\delta \xrightarrow{\mathscr{M}_\delta} \mathscr{B}_\delta$. In the following we shall seek a fixed point of $\mathscr{M}_\delta$ inside $\overline{\mathscr{C}}_\Lambda \subset \mathscr{B}_\delta$. For this reason we first need to define a precise appropriate norm such that every $\delta \in \overline{\mathscr{C}}_\Lambda$ has a finite norm and $\mathscr{B}_\delta$ becomes a Banach space. We introduce the following norm $\mathscr{N}_\delta$ in $\mathscr{B}_\delta$ for Λ fixed in (0;0.1]:

$$\|\delta\| = \sup_n N_n^{-1}|\delta_n(\Lambda)|, \tag{3.14}$$

with

$$N_1 = 2, \quad N_3 = 1, \quad N_5 = 5,$$
$$N_n = [n(n-1)]^2/5, \quad \forall n = 7,9,\ldots. \tag{3.14a}$$

By inspecting Definition 2(a1) of $\overline{\mathscr{C}}_\Lambda$ one easily understands that the above norm (3.14) is inspired by the absolute upper bounds in $\overline{\mathscr{C}}_\Lambda$. Moreover, one can directly verify that $\mathscr{B}_\delta$ is complete with respect to this norm. So by these considerations we are allowed to state without proof the following lemma.

Lemma 3.1: (i) The norm $\mathscr{N}_\delta$ of the definition (3.14) is finite $\forall \delta \in \overline{\mathscr{C}}_\Lambda$.

(ii) $\mathscr{B}_\delta$ is a Banach space in the norm $\mathscr{N}_\delta$.

Before we give the main theorem of this section we present an auxiliary lemma that describes properties concerning the functionals $\Delta_n(\Lambda)$, which we have introduced above by the definition (3.1d) of $\mathscr{M}_\delta$ $\forall n \geqslant 5$. These quantities satisfy, on one hand, absolute upper and lower bounds (resp. an "opposite" ordering $\mathscr{O}$) and, on the other hand, relative bounds analogous to that of the factors X_n, which "balance" the numerical weight of the sweeping factors $\alpha_n(\Lambda)$ (cf. Lemma 2.5). All these properties are necessary for the proof of Theorem 3.2 below, i.e., the stability of $\overline{\mathscr{C}}_\Lambda$ under $\mathscr{M}_\delta$ (resp. $\overline{\mathscr{C}}_\Lambda^{\prime\prime}$).

Lemma 3.2: Let $\delta \in \overline{\mathscr{C}}_\Lambda$, $\delta_{(1)} > \delta_{(2)} \in \overline{\mathscr{C}}_\Lambda^{\prime\prime}$; then $\forall n \geqslant 5$ the functionals $\Delta_n(\Lambda)$ of (3.1d) satisfy the following properties when $0 < \Lambda \lesssim 0.1$:

(i) $\lim_{\Lambda \to 0} \Delta_n(\Lambda) = \Delta_n^{(0)} = X_n^{(0)}/2$

[with $X_n^{(0)}$ given by (2.17b)], (3.15)

$$\Delta_5 \leqslant 1 + \bar{S}_3/5\bar{I}_5, \quad \Delta_n(\Lambda) \leqslant \tfrac{1}{2}, \quad \forall n \geqslant 7, \tag{3.16}$$

$$\Delta_5 \geqslant 1 + (\bar{I}_3/5\bar{S}_5)(1 - 14\Lambda) - \tfrac{42}{5}\Lambda, \quad \Delta_n(\Lambda) \geqslant 1/\delta_\infty^\Lambda, \quad \forall n \geqslant 7; \tag{3.17}$$

(ii) $\forall n \geqslant 9 \quad \exists \; 0 \leqslant \nu_n \leqslant \bar{\nu}_n(\Lambda) \leqslant 2/(n-2), \quad 0 \leqslant \bar{\omega}_n(\Lambda) \leqslant \omega_n(\Lambda) < \infty,$

$$\Delta_{n-2}/\Delta_n \leqslant \begin{cases} (1 - \nu_n(\Lambda))(n-1)/(n-3), & \text{if } n \leqslant N_\Lambda \equiv 4/3\Lambda + 1, \\ 1 + \omega_n(\Lambda)/(n-2)(n-3), & \text{if } n \geqslant N_\Lambda, \end{cases} \tag{3.18}$$

$$\Delta_{n-2}/\Delta_n \geqslant \begin{cases} (1 - \bar{\nu}_n(\Lambda))(n-1)/(n-3), & \text{if } n \leqslant N_\Lambda, \\ 1 + \bar{\omega}_n(\Lambda)/(n-2)(n-3), & \text{if } n \geqslant N_\Lambda; \end{cases} \tag{3.19}$$

(iii) for $0 < \Lambda \leqslant 0.01$,

$$\Delta_{n(2)} - \Delta_{n(1)} \geqslant 0, \qquad \forall n \geqslant 3, \tag{3.20}$$

$$\Delta_{n-2(2)} - \Delta_{n-2(1)} \geqslant \Delta_{n(2)} - \Delta_{n(1)}, \quad \forall n \geqslant 7. \tag{3.21}$$

The proof of this statement is given in III (Proposition 3.2).

Let us now proceed to the construction of the solution of the system (3.1) (i.e., the fixed point of $\mathscr{M}_\delta$). We present it in the form of two theorems. The first one contains the stability of the subset $\overline{\mathscr{C}}_\Lambda$ under $\mathscr{M}_\delta$ and the existence of a solution of $\mathscr{M}_\delta$ inside $\overline{\mathscr{C}}_\Lambda$ by applying the Leray–Shauder theorem. In the second one we proceed to the proof of the contractivity of $\mathscr{M}_\delta$ inside $\overline{\mathscr{C}}'_\Lambda$ by verification of the contractive mapping principle. In this way we show the uniqueness of the solution inside $\overline{\mathscr{C}}'_\Lambda$ and propose an iterative construction of it by contraction.

Theorem 3.2: (i) $\overline{\mathscr{C}}_\Lambda$ is a closed, convex, compact subset of $\mathscr{B}_\delta$.

(ii) $\overline{\mathscr{C}}_\Lambda$ is stable under the mapping $\mathscr{M}_\delta$ when Λ satisfies $0 < \Lambda \lesssim 0.1$.

(iii) There exists one fixed point of $\mathscr{M}_\delta$ inside $\overline{\mathscr{C}}_\Lambda$ when $0 < \Lambda \lesssim 0.1$.

(iv) The subset $\overline{\mathscr{C}}'_\Lambda \subset \overline{\mathscr{C}}_\Lambda$ is stable under $\mathscr{M}_\delta$ when $0 < \Lambda \lesssim 0.01$.

Proof: (i) By the norm definition (3.14) of $\mathscr{N}_\delta$ one directly verifies that every limit point of $\overline{\mathscr{C}}_\Lambda$ is contained inside it. Moreover, the convexity of $\overline{\mathscr{C}}_\Lambda$ can be trivially obtained in view of Definition 2(a).

In order to show the compactness of $\mathscr{C}_\Lambda$ we shall prove that the following criterion of Atkinson, Johnson, and Warnock[7] is verified: A subset $\mathscr{C} \subset K \subset \mathscr{B}$ (where K is a finite ball) of a Banach space $\mathscr{B}$ of sequences $\{a_n\}$ with norm $\|a\| = \sup_n |a_n|$ is compact in the topology induced by this norm if $\forall \varepsilon > 0 \;\exists$ an integer $n_0(\varepsilon)$ such that $\forall \{a_n\} \in \mathscr{C}$,

$$\sup_{n > n_0(\varepsilon)} |a_n| \leqslant \varepsilon. \tag{3.22}$$

Let $\delta \in \mathscr{C}_\Lambda$. We first note that following the norm definition we have

$$\|\delta\| = \sup_n N_n^{-1} |\delta_n| \leqslant 1, \tag{3.23}$$

and so

$$\mathscr{C}_\Lambda \subset K \equiv \{\delta \in \mathscr{B}_\delta : \|\delta\| = 1\}. \tag{3.24}$$

Moreover, for every $n \geqslant 7$ we obtain

$$N_n^{-1} \delta_n(\Lambda) \leqslant 5\delta_\infty^\Lambda / n^2(n-1)^2. \tag{3.25}$$

Manifestly, for every $\varepsilon > 0$ we can choose

$$n_0(\varepsilon) \equiv [5\delta_\infty^\Lambda / \varepsilon]^{1/4}, \tag{3.26}$$

so that $\forall n \geqslant n_0(\varepsilon) + 2$,

$$\sup_{n > n_0(\varepsilon)+2} N_n^{-1} \delta_n \leqslant \frac{5\delta_\infty^\Lambda}{(n_0(\varepsilon)+2)^2 (n_0(\varepsilon)+1)^2} \leqslant \frac{5\delta_\infty^\Lambda}{[n_0(\varepsilon)]^4} = \varepsilon. \tag{3.27}$$

So the above criterion is ensured in the norm $\mathscr{N}_\delta$, and $\mathscr{C}_\Lambda$ is a compact subset of $\mathscr{B}_\delta$ in the topology induced by $\mathscr{N}_\delta$. Q.E.D.

(ii) Under the assumption that an arbitrary sequence $\delta \in \mathscr{C}_\Lambda$, we show that the image $\mathscr{M}_\delta(\delta)$ also belongs to $\overline{\mathscr{C}}_\Lambda$. In other words, we ensure that $\forall n = 1,3,\ldots,\; \delta'_n$ defined by (3.1a)–(3.1d) satisfies the corresponding properties $\mathscr{C}_\Lambda$(i)–$\mathscr{C}_\Lambda$(iii) of Definition 2(a).

$\mathscr{C}_\Lambda$(i)a: In view of the hypothesis $\delta \in \overline{\mathscr{C}}_\Lambda$ and by expanding the square root $(1 - 4\delta_3\Lambda)^{1/2}$ of (3.1a), we have that for $n = 1$ [using $\lim_{\Lambda \to 0} \delta_3(\Lambda)/\Lambda = 6$],

$$\lim_{\Lambda \to 0} \frac{\delta'_1(\Lambda)}{\Lambda} = \lim_{\Lambda \to 0} \frac{\delta_3(\Lambda)}{\Lambda} \left[1 + \frac{4^2}{2} \frac{1 \cdot 3 \cdot 4\delta_3\Lambda}{2 \cdot 4 \cdot 6} + \cdots \right] = 6. \tag{3.28}$$

Q.E.D.

Moreover, for $n = 3$, using $\lim_{\Lambda \to 0} \delta_5(\Lambda) = 0$ inside (3.1b), we obtain directly

$$\lim_{\Lambda \to 0} \delta'_3(\Lambda)/\Lambda = 6. \tag{3.29}$$

Q.E.D.

Now in a similar way for every n odd $\geqslant 5$, we take into account the hypothesis $\delta \in \mathscr{C}_\Lambda$. So by Lemma 3.2 we have [cf. the property (3.15)]

$$\lim_{\Lambda \to 0} \Delta_n(\Lambda) = \Delta_n^{(0)} > 0, \tag{3.30}$$

and using the latter inside formula (3.1d) of $\delta'_n(\Lambda)$ we obtain

$$\lim_{\Lambda \to 0} \delta'_n(\Lambda)/\Lambda = 3n(n-1), \quad \forall n = 7,9,11,\ldots . \tag{3.31}$$

Q.E.D.

$\mathscr{C}_\Lambda$(i)b: The upper bound of $\delta'_n(\Lambda)$ at $n \to \infty$ is obtained again from formula (3.1d) by using the universal lower bound $\Delta_n(\Lambda) \geqslant 1/\delta^\Lambda_\infty$ given by Lemma 3.2, $\forall$ odd $n \geqslant 7$:

$$\lim_{n\to\infty} \delta'_n(\Lambda) = \lim_{n\to\infty} \left(\frac{1}{3\Lambda n(n-1)(1+\delta_1\Lambda)} + \Delta_n \right)^{-1} \leqslant \delta^\Lambda_\infty . \tag{3.32}$$

Q.E.D.

$\mathscr{C}_\Lambda$(ii): We then prove the absolute bounds. For $n = 1$, we require by (3.1a)

$$(1 - 2\delta_3\Lambda - (1-4\delta_3\Lambda)^{1/2})/2\delta_3\Lambda^2 \leqslant 6\Lambda. \tag{3.33}$$

In view of the hypothesis $\delta \in \mathscr{C}_\Lambda$ and when $0 < \Lambda \lesssim 0.1$, we first have

$$1 - 4\delta_3\Lambda \geqslant 1 - 24\Lambda^2 > 0 \tag{3.33a}$$

so that the square root of (3.33) can be isolated, and the square of both (positive) members of the inequality is taken. After some elementary algebraic manipulations and using formulas (2.2a) and (2.4a) of δ_3, we obtain instead of (3.33) the following stronger condition to be satisfied:

$$9 - 72\Lambda - 108\Lambda^3 - 216\Lambda^5 > 0. \tag{3.34}$$

The latter holds under the condition $0 < \Lambda \lesssim 0.1$. Q.E.D.

To obtain the lower bound of $\delta'_1(\Lambda)$ we expand the square root $(1-4\delta_3\Lambda)^{1/2}$ and obtain

$$\delta'_1(\Lambda) \geqslant \delta_3[1 + \delta_3\Lambda + 5\delta_3^2\Lambda^2 + \cdots] \geqslant \delta_3(\Lambda). \tag{3.35}$$

Using the lower bound (2.4b) of $\delta_3(\Lambda)$ ($\delta \in \overline{\mathscr{C}}_\Lambda$), that means

$$\delta'_1(\Lambda) \geqslant 6\Lambda I_3. \tag{3.35a}$$

Q.E.D.

For $n = 3$, the proof of the absolute bounds of $\delta'_3(\Lambda)$ is easily obtained from (3.1b) in view of the inequalities (2.2a) and the definition (2.4) for $\delta_1(\Lambda)$ and $\delta_5(\Lambda)$.

For $n = 5$, the upper (resp. lower) bound of δ'_5 is ensured by insertion of the inequality (3.17) [resp. (3.16)] of Lemma 3.2 into (3.1c).

$\mathscr{C}_\Lambda$(ii)b: The cases $\forall n \geqslant 7$ are obtained in a way exactly analogous to that above. We apply the upper and lower bounds of the functionals Δ_n given by properties (3.16) and (3.17) of Lemma 3.2 (in view of the hypothesis $\delta \in \mathscr{C}_\Lambda$) and the correnponding bounds (2.2a) of $\delta_1(\Lambda)$, in the general formula (3.1d) of δ'_n.

$\mathscr{C}_\Lambda$(iii): For the proof of relative bounds we use (3.1d) of δ'_n, δ'_{n-2} and write, after some simple calculus for the ratios of the splitting constants $\forall n \geqslant 9$,

$$\frac{\delta'_n}{\delta'_{n-2}} = \frac{n(n-1)}{(n-2)(n-3)}(1 - \gamma_n(\Lambda)), \tag{3.36}$$

where we identify

$$\gamma_n(\Lambda) \equiv \frac{3\Lambda(1+\delta_1\Lambda)[(4n-6)\Delta_n - (n-2)(n-3)\Delta_n(\Delta_{n-2}/\Delta_n - 1)]}{1 + 3\Lambda n(n-1)(1+\delta_1\Lambda)\Delta_n}, \tag{3.36a}$$

and for $n = 7$, respectively:

$$\delta'_7/\delta'_5 = \tfrac{42}{5}(1 - \gamma_7(\Lambda)), \tag{3.37}$$

with

$$\gamma_7(\Lambda) = \frac{3\Lambda(1+\delta_1\Lambda)[37\Delta_5 - 42\Delta_5(1-\Delta_7/\Delta_5)]}{1 + 126\Lambda(1+\delta_1\Lambda)\Delta_7}. \tag{3.37a}$$

We then define

$$\hat{\gamma}_n(\Lambda) = \operatorname*{Inf}_{\delta \in \mathscr{C}_\Lambda} \gamma_n(\Lambda), \quad \bar{\gamma}_n(\Lambda) = \sup_{\delta \in \mathscr{C}_\Lambda} \gamma_n(\Lambda), \tag{3.38}$$

and in view of the hypothesis $\delta \in \mathscr{C}_\Lambda$, we apply properties (3.16), (3.17), (3.19), and (3.20) of Lemma 3.2 concerning the absolute and relative bounds of Δ_n's for $n \leqslant 4/3\Lambda + 1 \equiv N_\Lambda$. After some elementary estimations,

$$\hat{\gamma}_n(\Lambda) = \frac{3\Lambda(n-1)[2 + \nu_n(\Lambda)(n-2)](1+6\Lambda^2 I_3)}{\delta^\Lambda_\infty + 3\Lambda n(n-1)(1+6\Lambda^2 I_3)}, \quad \forall N_\Lambda \geqslant n \geqslant 9, \tag{3.39a}$$

$$\hat{\gamma}_7(\Lambda) = 3\Lambda(1+6\Lambda^2 I_3)[1 + 16\Lambda(1+6\Lambda^2 I_3)]^{-1}, \quad \text{for } n = 7, \tag{3.39b}$$

$$\bar{\gamma}_n(\Lambda) = \frac{3\Lambda(n-1)[2 + \bar{\nu}_n(\Lambda)(n-2)](1+6\Lambda^2)}{2 + 3\Lambda n(n-1)(1+6\Lambda^2)}, \quad \forall n \geqslant 9, \tag{3.40a}$$

$$\bar{\gamma}_7(\Lambda) = \frac{16.5\Lambda(1+6\Lambda^2)}{1 + 21\Lambda(1+6\Lambda^2)}, \quad n = 7. \tag{3.40b}$$

Now taking into account the bounds $0 \leqslant \nu_n(\Lambda) \leqslant \bar{\nu}_n(\Lambda) \leqslant 2/(n-2)$, we obtain from (3.40a), for the region $3\Lambda(n-1) \leqslant 4$,

$$\bar{\gamma}_n(\Lambda) \leqslant 4/([2/3\Lambda(n-1)(1+6\Lambda^2)] + n) \leqslant 4/n. \tag{3.41}$$

The last result completes the proof of (2.3a) in Definition 2(a). Q.E.D.

As long as n increases, $\hat{\gamma}_n(\Lambda)$ also increases slowly, so that in the region $3\Lambda(n-1) \geqslant 4$ we pass smoothly to the regime of *slower increase* behavior of the splitting sequences

$\delta\in\mathscr{C}_\Lambda$. More precisely, taking into account the lower bound (3.17) and the upper relative bound (3.19) of Lemma 3.2 for the *slower decrease* behavior of the sequences Δ_n, we have from (3.36a), $\forall n \geqslant 4/3\Lambda + 1$,

$$\hat{\gamma}_n(\Lambda) = \frac{4(n-1)-(2+\omega_n)}{n(n-1)\{1+[\Delta_n(1+\delta_1\Lambda)3\Lambda n(n-1)]^{-1}\}}$$
$$\geqslant \frac{4}{n}\left[1-\frac{\delta_\infty^\Lambda+2+\omega_n}{4n}+\frac{(2+\omega_n)(\delta_\infty^\Lambda-4)}{16n(n-1)}\right]. \tag{3.42}$$

Insertion of (3.42) into (3.36) yields, after some estimations,

$$\delta'_n/\delta'_{n-2} \leqslant 1+\mu_n(\Lambda)/n^2,$$
$$\mu_n(\Lambda) \equiv (\delta_\infty^\Lambda+2+\omega_n)(1+4/(n-2)). \tag{3.43}$$

The above result ensures the upper bound for the detailed slow increase properties (2.3b) of $\mathscr{C}_\Lambda$ in the region $n \geqslant N_\Lambda \equiv 4/3\Lambda + 1$. The corresponding lower bound is obtained in an analogous way. The proof of the stability of $\overline{\mathscr{C}}_\Lambda$ is thus completed.

(iii) Using the definitions (3.1a)–(3.1d) of the mapping we can write, for every pair $\delta_{(1)},\delta_{(2)}\in\mathscr{C}_\Lambda$, and $\forall n = 1,3,...$ [cf. also the explicit formulas (3.45), (3.47), (3.49), and (3.51) below] under the condition $0<\Lambda \lesssim 0.1$,

$$|\delta'_{n(1)}-\delta'_{(2)}| \leqslant R_n(\Lambda)\|\delta_{(1)}-\delta_{(2)}\|. \tag{3.44}$$

Here $R_n(\Lambda)$ is a continuous function of n and Λ independent on the particular pair of sequences $\delta_{(1)},\delta_{(2)}\in\mathscr{C}_\Lambda$. In other words, the continuity (in the $\mathscr{N}_\delta$ norm topology) of $\mathscr{M}_\delta$ is obtained. Taking into account the last result and the already proven parts (i) and (ii) of this theorem, we are allowed to apply the Leray–Schauder–Tychonov theorem[7] and obtain the existence of at least one fixed point of $\mathscr{M}_\delta$ inside $\overline{\mathscr{C}}_\Lambda$. Q.E.D.

(iv) Let $\delta_{(1)},\delta_{(2)}\in\overline{\mathscr{C}}_\Lambda^{\prime\prime}$ with $\delta_{(1)} > \delta_{(2)}$. We shall show that this ordering is conserved by $\mathscr{M}_\delta$, i.e., $\delta'_{(1)} > \delta'_{(2)}$.

(a) For $n=1$, we obtain from (3.1a)

$$\delta'_{1(1)}-\delta'_{1(2)} = \frac{8(\delta_{3(1)}-\delta_{3(2)})}{[1+[1-4\delta_{3(1)}\Lambda]^{1/2}][1+[1-4\delta_{3(2)}\Lambda]^{1/2}]\{[1-4\delta_{3(1)}\Lambda]^{1/2}+[1-4\delta_{3(2)}\Lambda]^{1/2}\}}. \tag{3.45}$$

In view of the hypothesis $\delta_{3(1)} \geqslant \delta_{3(2)}$ and the fact that the denominator of (3.45) is also a purely positive quantity [cf. also (3.33a)] when $0<\Lambda\lesssim 0.1$, we conclude that

$$\delta'_{1(1)}-\delta'_{1(2)} \geqslant 0. \tag{3.46}$$

Q.E.D.

(b) For $n=3$, formula (3.1b) yields, after some elementary algebra,

$$\delta'_{3(1)}-\delta'_{3(2)} = 6\Lambda\frac{\Lambda(\delta_{1(1)}-\delta_{1(2)})+4\Lambda(\delta_{5(1)}-\delta_{5(2)})(1+\delta_{1(1)}\Lambda)(1+\delta_{1(2)}\Lambda)}{[1+6\Lambda(1+\delta_{1(1)}\Lambda)[\frac{3}{2}-\frac{2}{3}\delta_{5(1)}]][1+6\Lambda(1+\delta_{1(2)}\Lambda)[\frac{3}{2}-\frac{2}{3}\delta_{5(2)}]]}. \tag{3.47}$$

Using the hypotheses $\delta_{1(1)}-\delta_{1(2)}\geqslant 0$, $\delta_{5(1)}-\delta_{5(2)}\geqslant 0$, and the condition $0<\Lambda\leqslant 0.1$, the latter yields

$$\delta'_{3(1)}-\delta'_{3(2)} \geqslant 0. \tag{3.48}$$

Q.E.D.

(c) For $n=5$, we obtain in an analogous way by formula (3.1c)

$$\delta'_{5(1)}-\delta'_{5(2)} = 15\Lambda\left\{\frac{(\delta_{1(1)}-\delta_{1(2)})\Lambda+15\Lambda(1+\delta_{1(1)}\Lambda)(1+\delta_{1(2)}\Lambda)(\Delta_{5(2)}-\Delta_{5(1)})}{[1+15\Lambda(1+\delta_{1(1)}\Lambda)\Delta_{5(1)}][1+15\Lambda(1+\delta_{1(2)}\Lambda)\Delta_{5(2)}]}\right\}. \tag{3.49}$$

The denominator of the latter is again a purely positive quantity thanks to the property (3.17) (Lemma 3.2). Moreover, by property (iii), $\Delta_{5(2)}-\Delta_{5(1)}\geqslant 0$ of Lemma 3.2, and the hypothesis $\delta_{1(1)}-\delta_{1(2)}\geqslant 0$, the non-negativity of the numerator is ensured so that

$$\delta'_{5(1)}-\delta'_{5(2)} \geqslant 0. \tag{3.50}$$

Q.E.D.

By analogous considerations, using $\Delta_{n(2)}-\Delta_{n(1)}\geqslant 0$ of Lemma 3.2 and the hypothesis $\delta_{1(1)}-\delta_{1(2)}\geqslant 0$ inside the formula [obtained by (3.1d)]

$$\delta'_{n(1)}-\delta'_{n(2)} = 3\Lambda n(n-1)\left\{\frac{(\delta_{1(1)}-\delta_{1(2)})\Lambda+3\Lambda n(n-1)(1+\delta_{1(1)}\Lambda)(1+\delta_{1(2)}\Lambda)(\Delta_{n(2)}-\Delta_{n(1)})}{[1+3\Lambda n(n-1)(1+\delta_{1(1)}\Lambda)\Delta_{n(1)}][1+3\Lambda n(n-1)(1+\delta_{1(2)}\Lambda)\Delta_{n(2)}]}\right\}, \tag{3.51}$$

$\forall n\geqslant 7$, we also obtain

$$\delta'_{n(1)}-\delta'_{n(2)} \geqslant 0. \tag{3.52}$$

Q.E.D.

To complete the proof of property (iv), we need to show first that

$$\frac{\delta'_{n(1)}}{\delta'_{n(2)}} \geqslant \frac{\delta'_{n-2(1)}}{\delta'_{n-2(2)}}, \tag{3.53}$$

which is equivalent to the left-hand side of the inequality ($\mathscr{C}_\Lambda^{\prime\prime}$a), ($\mathscr{C}_\Lambda^{\prime\prime}$b), and then prove the upper bounds of the right-hand side of ($\mathscr{C}'_\Lambda$a). Using the proven properties $\mathscr{C}_\Lambda$ (iii) [cf. (3.36)], the relation (3.53) is translated to the requirement

$$(1-\gamma_{n(1)})\geqslant(1-\gamma_{n(2)}) \quad \text{or} \quad \gamma_{n(2)}\geqslant\gamma_{n(1)}. \tag{3.54}$$

By application of property (iii) (3.20) of Lemma 3.2, i.e., $\Delta_{n(2)}\geqslant\Delta_{n(1)}$, inside the definitions (3.36a) and (3.37a) for $\gamma_{n(2)}$ and $\gamma_{n(1)}$, one finally verifies that

$$\gamma_{n(2)}-\gamma_{n(1)}\geqslant 0, \quad \text{for } 0<\Lambda\lesssim 0.01.$$

Q.E.D.

On the other hand, using the equalities (3.36) and (3.51) for δ_n, δ_{n-2}, we write, $\forall n\geqslant 9$,

$$\frac{\delta'_{n(1)}-\delta'_{n(2)}}{\delta'_{n-2(1)}-\delta'_{n-2(2)}}=\left[\frac{n(n-1)}{(n-2)(n-3)}\right]^2(1-\gamma_{n(1)})(1-\gamma_{n(2)})\frac{(\Delta_{n(2)}-\Delta_{n(1)})}{(\Delta_{n-2(2)}-\Delta_{n-2(1)})}$$
$$\times\left\{\frac{1+(\delta_{1(1)}-\delta_{1(2)})[3n(n-1)(1+\delta_{1(1)}\Lambda)(1+\delta_{1(2)}\Lambda)(\Delta_{n(2)}-\Delta_{n(1)})]^{-1}}{1+(\delta_{1(1)}-\delta_{1(2)})[3(n-2)(n-3)(1+\delta_{1(1)}\Lambda)(1+\delta_{1(2)}\Lambda)(\Delta_{n-2(2)}-\Delta_{n-2(1)})]}\right\}. \tag{3.55}$$

Following Lemma 3.2, formula (3.21), we have the inequality

$$\Delta_{n(2)}-\Delta_{n(1)}\leqslant\Delta_{n-2(2)}-\Delta_{n-2(1)}.$$

Taking into account the latter and the fact that by $\mathscr{C}_\Lambda$ (iii)

$$0<(1-\gamma_{n(1)})(1-\gamma_{n(2)})<1,$$

we obtain the upper bound $[n(n-1)]^2/[(n-2)(n-3)]^2$ for the right-hand side of (3.55). Q.E.D.

Analogous considerations yield the bound $(\frac{42}{5})^2$ for $n=7$.

These results complete the proof of the stability of $\overline{\mathscr{C}}''_\Lambda$ under $\mathscr{M}_\delta$ and the proof of Theorem 3.2. We now proceed to the proof of the uniqueness of the solution inside $\overline{\mathscr{C}}''_\Lambda$.

Theorem 3.3: The mapping $\mathscr{M}_\delta$ [(3.1a)–(3.1d)] is a contraction in the norm $\mathscr{N}_\delta$ inside $\mathscr{C}''_\Lambda$ when Λ satisfies

$$0<\Lambda\lesssim 0.01. \tag{3.56}$$

Proof: We suppose that $\delta_{1(1)}>\delta_{1(2)}$ and $\delta_{1(1)}>\delta_{1(2)}$, $\delta_{3(1)}>\delta_{3(2)}$. By the definition of $\mathscr{M}_\delta$ and Theorem 3.2 one directly verifies that these two initial conditions imply $\delta'_{n(1)}>\delta'_{n(2)}$, $\forall n=1,3,5,\ldots$.

For $n=1$, we use (3.45) (cf. the proof of Theorem 3.2). By application of (2.2a), (2.4a), and the hypothesis $\delta_{(1)}>\delta_{(2)}\in\mathscr{C}''_\Lambda$, we take the lower bound of the denominator. By application of the norm definition (3.14a) we evaluate an upper bound of the numerator and obtain, finally,

$$|\delta'_{1(1)}-\delta'_{1(2)}|/N_1\leqslant K_1(\Lambda)\|\delta_{(1)}-\delta_{(2)}\|, \tag{3.57}$$

where

$$K_1(\Lambda)=\bar{S}_1(\Lambda)/2 \quad \text{and} \quad K_1(\Lambda)<1, \tag{3.57a}$$

even for $\Lambda\leqslant 0.1$.

For $n=3$, we estimate the lower bound of the denominator in the right-hand side of (3.47), and by application of the norm definition in the numerator we finally have (recall $\Delta_3=\frac{3}{2}-\frac{2}{3}\delta_5$)

$$|\delta'_{3(1)}-\delta'_{3(2)}|/N_3\leqslant K_3(\Lambda)\|\delta_{(1)}-\delta_{(2)}\|, \tag{3.58}$$

where

$$K_3(\Lambda)\equiv 12\Lambda^2\frac{1+10(1+6\Lambda^2\bar{S}_1)^2}{[1+9\Lambda-60\Lambda^2]^2}<1, \quad \text{for } 0<\Lambda\lesssim 0.1. \tag{3.59}$$

For $n=5$, we apply the hypothesis $\delta_{(1)}>\delta_{(2)}\in\overline{\mathscr{C}}''_\Lambda$ and the definitions $\Delta_5=1+\frac{1}{2}(\delta_3/\delta_5)-\delta_7\beta_7/15$ and $\bar{S}_5$ (2.4b) in the formula (3.49). After some elementary estimations, and by the norm definition, we write

$$|\delta'_{5(1)}-\delta'_{5(2)}|/N_5\leqslant K_5(\Lambda)\|\delta_{(1)}-\delta_{(2)}\|, \tag{3.60}$$

where

$$K_5(\Lambda)=\frac{15\Lambda}{\bar{S}_5^2}\left\{2\Lambda+I_5^{-1}\left[\frac{\bar{S}_3}{5\bar{I}_5}+\frac{1}{15}\left(1+\frac{1}{3}\frac{\bar{S}_3}{\bar{I}_5}\right)\left(\frac{42}{5}\right)^2\right]\right\}. \tag{3.61}$$

The last quantity can be smaller than 1 only if $0<\Lambda\lesssim 0.01$. To be precise,

$$K_5(\Lambda)\leqslant K_5(\Lambda=0.01)=0.8704<1. \tag{3.62}$$

For $n\geqslant 7$, we apply a recursion. We suppose that, $\forall 7\leqslant\bar{n}\leqslant n-2$,

$$|\delta'_{\bar{n}(1)}-\delta'_{\bar{n}(2)}|/N_{\bar{n}}\leqslant K_5(\Lambda)\|\delta_{(1)}-\delta_{(2)}\|. \tag{3.63}$$

The first step of this recurrence is directly verified by application of the properties $\delta'_{(1)}>\delta'_{(2)}\in\mathscr{C}''_\Lambda$ (proved by Theorem 3.2), ($\mathscr{C}''_\Lambda$a) (the right-hand side) for $n=7$, and (3.60), namely,

$$\frac{|\delta'_{7(1)}-\delta'_{7(2)}|}{N_7}\leqslant\left(\frac{42}{5}\right)^2\frac{N_5}{N_7}K_5(\Lambda)\|\delta_{(1)}-\delta_{(2)}\|. \tag{3.64}$$

Evidently,

$$(\tfrac{42}{5})^2N_5/N_7=1 \tag{3.65}$$

[by (3.14a) of $\mathscr{N}_\delta$]. To show the statement for $\bar{n}=n$ we apply again the bound ($\mathscr{C}''_\Lambda$a) of $\overline{\mathscr{C}}''_\Lambda$ and obtain

$$\frac{|\delta'_{n(1)}-\delta'_{n(2)}|}{N_n}\leqslant\frac{n^2(n-1)^2}{N_n}N_{n-2}K_5(\Lambda)\|\delta_{(1)}-\delta_{(2)}\|. \tag{3.66}$$

The latter yields directly the proof of the recursion in view of the norm definition. Q.E.D.

Now combining the above results (3.57a), (3.58)–(3.60), (3.62), and (3.63), we conclude that $\forall n=1,3,5,\ldots$, $\exists\, K(\Lambda)\equiv K_5(\Lambda)$,

$$|\delta'_{n(1)}-\delta'_{n(2)}|/N_n\leqslant K(\Lambda)\|\delta_{(1)}-\delta_{(2)}\|. \tag{3.67}$$

Following the norm definition this means that

$$\|\delta'_{(1)}-\delta'_{(2)}\|\leqslant K_5(\Lambda)\|\delta_{(1)}-\delta_{(2)}\|, \tag{3.68}$$

with $K_5(\Lambda) < 1$ if $0 < \Lambda \lesssim 0.01$, in view of (3.62). In other words, the contractivity criterion is satisfied and the contractive mapping principle[7] can be applied inside $\overline{\mathscr{C}}_{\Lambda,\mathscr{O}}$.

Q.E.D.

It follows that $\exists$ a unique fixed point of $\mathscr{M}_\delta$ inside $\overline{\mathscr{C}}^{\mathscr{O}}_\Lambda$ and that it can be constructed numerically starting from a given point of $\overline{\mathscr{C}}_\Lambda$ at fixed Λ in the interval $0 < \Lambda \lesssim 0.01$.

Finally, we apply Theorems 3.1–3.3 to obtain the complete answer to the zero-dimensional problem given by the following corollary.

Corollary 3.1: (i) Under the condition $0 < \Lambda \lesssim 0.1$ $\exists$ at least one nontrivial solution $\overline{H}_0$ of the zero-dimensional system (1.2) of the equations of motion for the Schwinger functions, a solution that satisfies the sign and splitting properties characterizing $\Phi_{0\Lambda}$. This solution is explicitly given in terms of the nontrivial solution $\bar{\delta}$ of the system (3.1) for the splitting constants (found in Theorem 3.2) by the following recursive definition:

$$\begin{aligned} \overline{H}_0^2 &= 1 + \bar{\delta}_1\Lambda, \\ \overline{H}_0^4 &= -\bar{\delta}_3[\overline{H}_0^2]^2, \\ \overline{H}_0^6 &= -4\bar{\delta}_5\overline{H}_0^4\overline{H}_0^2, \\ \overline{H}_0^{n+1} &= -\bar{\delta}_n\bar{\beta}_n\overline{H}_0^{n-1}\overline{H}_0^2, \quad \forall n \geqslant 7. \end{aligned} \tag{3.69}$$

Here the sequences $\bar{\beta}_n$ are recursively defined by Lemma 2.2 in terms of all $\bar{\delta}_{\bar{n}}$'s with $\bar{n} \leqslant n-2$.

(ii) When $0 < \Lambda \lesssim 0.01$, the solution $\bar{\delta}$ is uniquely defined inside $\overline{\mathscr{C}}^{\mathscr{O}}_\Lambda$ and can be constructed iteratively starting from the minimal sequence δ^* defined precisely in Lemma 2.1.

The solution $\overline{H}_0$ is therefore uniquely defined in the corresponding subset of $\Phi_{0\Lambda}$ by (3.69).

Remark 1: The above solution $\overline{H}_0$ and the technique developed for the proofs will be used in Sec. IV in order to find the solution of the zero-dimensional-type (δ,Φ) systems for $\Phi^4_{1,2}$ and to establish a convergent iteration for the construction of a unique nontrivial solution in two (and one) dimensions.

Remark 2: The method we presented above will be applied identically in a forthcoming work[8] to ensure the solution of the zero-dimensional type (δ,Φ) systems, for the construction of a nontrivial solution of the corresponding Φ^4_4 (and, *a fortiori*, Φ^4_3) equations of motion for the Schwinger functions.

Remark 3: As we already mentioned in I[3] and the Introduction, the above solution coincides with the one obtained by the generating functional method. Voros[9] has obtained this coincidence numerically and the exact results are presented in Appendix A of III.[6]

IV. THE TWO- (OR ONE-) DIMENSIONAL PROBLEM

In this part of the paper we study the Φ^4_2 equations of motion. Owing to the similarity between the one- and two-dimensional cases (nontrivial Φ convolutions, renormalization operator equal to the identity), everything that is established in what follows is also valid for the one-dimensional equations.

We consider the system of equations (1.7) from the Introduction. We shall construct a nontrivial solution $\{\overline{H}\}$ of this system by a fixed-point method and through the introduction of a precise iterative procedure. The solution we find belongs to the subset Φ_Λ [cf. Definition 2(e)]. This means that it satisfies all the signs and splitting properties at zero external momenta characterizing the subset Φ_Λ. These properties have already been studied in Sec. III for zero dimensions, and they originally appeared experimentally during the Φ iteration presented in I.[3] The main steps towards this purpose are as follows.

(1) We first introduce an appropriate norm in $\mathscr{B}$. Then we show the stability of Φ_Λ under the action of the mapping $\mathscr{M}$ [defined by the system (1.2)] by solving a "zero-dimensional-type" problem. Therefore we extensively apply the arguments of Sec. III, namely, the proofs (a) for the equivalence Theorem 3.1 concerning the relation (1.2) $\rightleftarrows$ (3.1), and (b) for the solution of (3.1) in $\mathscr{B}_\delta$ space. Moreover, the closedness of $\Phi_\Lambda \subset \mathscr{B}$ is ensured in the norm we mentioned above (cf. Theorem 4.1 which follows).

(2) By proving the contractivity of $\mathscr{M}$ inside Φ_Λ, we find a unique nontrivial fixed point $\overline{H}$ (cf. Theorem 4.2 below), and we propose a precise iteration in order to construct it.

Finally, we present a more direct proof of the convergence of the Φ iteration to the solution (cf. Sec. II of I for the corresponding analysis of our original proof). This allows us to obtain a different possible iterative construction of $\overline{H}$.

Before giving the precise norm definition, we start with the proof of the conservation of the zero-momentum dominance when $\mathscr{M}$ acts on every $H \in \Phi_\Lambda$. These properties express, in fact, upper bounds for arbitrary values of the external momenta of the Green's functions $H^{n+1}(q,\Lambda)$ and their ΦC's in terms of their corresponding numerical values (for fixed Λ) at zero external momenta. These zero-momentum dominance bounds have been revealed by the Φ iteration, as we explained in Sec. II of I, and they contribute to the finer definition of the norm in $\mathscr{B}$ space presented below.

Proposition 4.1 (Zero momentum dominance): Let $H \in \Phi_\Lambda$. For all $n \geqslant 1$ and for every element $\Phi(H) \in \mathscr{F}$ there exist positive sequences $M_n[\Phi^{(\bar{n},n)}(H)H^{n+1}](\Lambda)$ (resp. $\overline{M}_n[\Phi^{(\bar{n},n)}(H)H^{n+1}]$) such that the following bounds are conserved, respectively:

$$|\Phi^{(\bar{n},n)}(H)H^{n+1}| \leqslant M_n[\Phi^{(\bar{n},n)}(H)]\,|\Phi^{(\bar{n},n)}(H)H^{n+1}|_0, \tag{4.1a}$$

$$\begin{aligned} &|\Phi^{(\bar{n},n)}(H)(H^{n+1}_{(1)} - H^{n+1}_{(2)})| \\ &\quad \leqslant \overline{M}_n[\Phi^{(\bar{n},n)}(H)]\,|\Phi^{(\bar{n},n)}(H)(H^{n+1}_{(1)} - H^{n+1}_{(2)})|_0. \end{aligned} \tag{4.1b}$$

Moreover, these quantities $M_n, \overline{M}_n$ are bounded uniformly as

$$(\overline{M}_1)M_1[\Phi^{(\bar{n},1)}] \leqslant 2, \tag{4.2a}$$

$$\begin{aligned} &(\overline{M}_n)M_n[\Phi^{(\bar{n},n)}] \\ &\quad \leqslant n![3 + (\tilde{\delta}^\Lambda_\infty)^2/48 + 3\pi\tilde{\delta}^\Lambda_\infty]^{(n-3)/2} \equiv \widehat{M}_n, \quad \forall n \geqslant 3 \end{aligned} \tag{4.2b}$$

(by the abbreviated notation we express the fact that the same uniform bounds also hold for the quantities $\overline{M}_n[\Phi^{(\bar{n},n)}(H)H^{n+1}]$), and they satisfy the following self-consistency conditions:

$$(\overline{M}_n)M_n[\Phi^{(\bar n,n)}H^{n+1}] = M_n\left[\Phi^{(\bar n)}N_1^{(n)}H^{n+1}\prod_{l=2}^{3}N_1^{(i_l)}H^{i_l+1}\right] \quad (n+i_2+i_3=\bar n), \tag{4.2c}$$

$$(\overline{M}_{n+2})M_{n+2}[\Phi^{(\bar n,n)}N_3^{(n+2)}H^{n+3}] \leqslant M_{n-2}\left[\Phi^{(\bar n,n)}N_1^{(n-2)}H^{n-1}\prod_{l=2}^{3}H^2(q_l)\Delta_F(q_l)\right], \tag{4.2d}$$

$$(\overline{M}_n)M_n[\Phi^{(\bar n,n)}N_2^{(n)}H^{n+1}N_1H^2] \leqslant M_{n-2}\left[\Phi^{(\bar n,n)}N_1^{(n-2)}H^{n-1}\prod_{l=2}^{3}H^2(q_l)\Delta_F(q_l)\right]. \tag{4.2e}$$

Proof: For the bounds (4.1) when $n=1,3$, one trivially verifies that for $0<\Lambda\lesssim 0.01$ the definition (1.7) of $\mathscr{M}$ yields

$$|\Phi^{(\bar n,1)}H^{2\prime}|\leqslant(1+N_{3(0)}\Lambda\delta_1(\Lambda)M_3(\Lambda))|\Phi^{(\bar n,1)}H^2|_0 < 2|\Phi^{(\bar n,1)}H^2|_0; \tag{4.3a}$$

here $N_{3(0)}\equiv[N_3]_0$ and

$$|\Phi^{(\bar n,3)}H^{4\prime}|\leqslant 6|\Phi^{(\bar n,3)}H^4|_0. \tag{4.3b}$$

In order to show the analogous bounds for $n\geqslant 5$, together with the conditions (4.2c)–(4.2e), we proceed recursively and apply the consistency conditions to the terms $A^{n+1}, B^{n+1}, C^{n+1}$ of the mapping (1.7). We finally find

$$|\Phi^{(\bar n,n)}H^{(n+1)\prime}|\leqslant\frac{3\Lambda n(n-1)}{{}^{\Phi}\delta_n(\Lambda)}M_{n-2}[\Phi^{(\bar n,n)}N_1^{(n-2)}[N_1^{(1)}]^2](1+6\Lambda^2N_{3(0)}) \times\left\{1+\frac{{}^{\Phi}\delta_n{}^{\Phi}\delta_{n+2}{}^{\Phi}\beta_{n+2}N_{3(0)}}{3n(n-1)}+\frac{N_{2(0)}{}^{\Phi}\alpha_n\delta_n}{n-1}\right\}|{}^{\Phi}H_0^{n+1}|\,|\Phi^{(\bar n,n)}(H)\tilde{]}_0, \tag{4.4}$$

and, respectively,

$$|\Phi^{(\bar n,n)}N_3H^{(n+3)\prime}|\leqslant\frac{3\Lambda(n-2)(n-3)}{{}^{\Phi}\delta_{n-2}}M_{n-4}M_1^{\Phi}(1+6\Lambda^2N_{3(0)})\left\{\frac{{}^{\Phi}\delta_{n-2}}{{}^{\Phi}\delta_n}\frac{{}^{\Phi}\delta_n}{{}^{\Phi}\delta_{n+2}}\frac{(n+1)(n+2)}{n(n-1)}\frac{n(n-1)}{(n-2)(n-3)}\right\} \times\left\{1+\frac{{}^{\Phi}\delta_{n+4}{}^{\Phi}\beta_{n+4}{}^{\Phi}\delta_{n+2}N_{3(0)}}{3(n+1)(n+2)}+\frac{N_{2(0)}\delta_{n+2}{}^{\Phi}\alpha_{n+2}}{n+2}\right\}|{}^{\Phi}H_0^{n+3}|\,|\Phi^{(\bar n,n)}N_3^{(n+2)}\tilde{]}_0. \tag{4.5}$$

(Notice that $N_{2(0)}\equiv[N_2\tilde{]}_0=\pi$.) The quantities that multiply $|\Phi^{(\bar n,n)}H^{n+1}|_0$ (resp. $|\Phi^{(\bar n,n)}N_3^{(n+2)}H^{n+3}|_0$) in the right-hand side of (4.4) [resp. (4.5)] are manifestly bounded by the constant

$$n![3+(\tilde\delta_\infty^\Lambda)^2N_{3(0)}/48+3\pi\tilde\delta_\infty^\Lambda]^{(n-3)/2}$$
$$(\text{resp. } (n-2)![3+(\delta_\infty^\Lambda)^2N_{3(0)}/48+3\pi\delta_\infty^\Lambda]^{(n-5)/2}),$$

in view of the recurrence hypothesis and the assumption $H\in\Phi_\Lambda$ (application of Lemmas 2.6 and 2.7). Analogous arguments can be used for the property (4.1b). Then by taking the supremum (when $\delta\in\widetilde{\mathscr{C}}_\Lambda$) between the factors of (4.4) and (4.5) (bounded as we mentioned above), we define the corresponding quantity M_n (resp. M_{n+2}), which verifies (4.2b) [resp. (4.1d)]. This completes the proof of the zero momentum dominance property.

Definition 4(a): Let $H\in\mathscr{B}$. We define the following norm $\widetilde{\mathscr{N}}$ by

$$\|H\| = \sup_{n,\Phi\in\mathscr{T}} M_n^{\Phi^{-1}}|\tilde\Phi^{(\bar n,n)}|_0^{-1} \times\prod_{l=1}^{m_\Phi}|||H_0^{i_l+1}|||^{-1}\,|||H_0^{n+1}|||^{-1} \times|\Phi^{(\bar n,n)}(H)H^{n+1}|, \tag{4.6}$$

with

$$|||H_0^2|||=2,\quad |||H_0^{n+1}|||=n![\tilde\delta_\infty^\Lambda]^{(n-3)/2},\quad \forall n\geqslant 3, \tag{4.6a}$$

and

$$M_n^\Phi\equiv\begin{cases}\widehat{M}_n, & \text{if } \Phi^{(\bar n,n)}H^{n+1}=\Phi^{(\bar n n+2)}N_1^{(n)}N_1N_1,\\ \widehat{M}_{n-4}, & \text{if } \Phi^{(\bar n,n)}=\Phi^{(\bar n n-2)}N_3^{(n)}H^{n+1},\\ \widehat{M}_{n-2}, & \text{if } \Phi^{(\bar n,n)}=\Phi^{(\bar n,n)}N_2^{(n)}N_1^{(1)},\quad M_1^\Phi\equiv 1\end{cases} \tag{4.6b}$$

[for $\widehat{M}_n$ cf. (4.2b)]. Here $\tilde\Phi^{(\bar n,n)}$ means the completely reduced ΦC, and the product extends to the indices i_l corresponding to all functions H^{i_l+1} (except H^{n+1} itself) that define the ΦC $\Phi^{(\bar n,n)}(H)$. Moreover, the identity ΦC for the H^2 function must always be understood as coinciding with the free propagator $N_1^{(1)}=\Delta_F(q)$.

In view of Proposition (2.2), Definitions 2(a2) and 2(e), and Proposition 4.1 above, one verifies that when $H\in\Phi_\Lambda$, then $\widetilde{\mathscr{N}}$ is finite. Moreover, we can easily show that $\mathscr{B}$ is complete in this norm, which means that $\mathscr{B}$ is a Banach space in the norm $\widetilde{\mathscr{N}}$.

We now state the stability of Φ_Λ under $\mathscr{M}$ and its closedness. Both properties are necessary for the existence proof of a unique fixed point of $\mathscr{M}$ in Φ_Λ.

Theorem 4.1: (i) The subset $\Phi_\Lambda\subset\mathscr{B}$ is stable under the nonlinear mapping $\mathscr{M}$ (1.7) when $0<\Lambda\lesssim 0.1(\underline{b}_3/\bar a_3)$. [Notice that $\underline{b}_3\sim N_{2(0)}=\pi$, $\bar a_3\sim N_{3(0)}\sim 4.8\pi^2$, and $\Lambda_{\max}\equiv 0.1(\underline{b}_3/\bar a_3)\sim 0.0066$.]

(ii) The subset Φ_Λ is a closed subset of $\mathscr{B}$ in the norm $\widetilde{\mathscr{N}}$ [cf. (4.6)].

Proof of Theorem 4.1: (i) Let $H\in\Phi_\Lambda$. We verify that $\mathscr{M}(H)\in\Phi_\Lambda$ also.

(i)1 Following Definition 2(e), this means that for an arbitrary coherent sequence of ΦC's $\{\Phi^{(\bar n,n)}(H)H^{n+1}\}_n$

$\in\mathscr{F}$, $\exists$ a sequence ${}^{\delta}\delta'(\Lambda)\in\widetilde{\mathscr{C}}_{\Lambda}$ and, equivalently, a sequence ${}^{\Phi}H_0'\in\mathscr{B}_0$ such that, $\forall n = 1,3,...,$

$$[\Phi^{(\bar{n},n)}H^{(n+1)\prime}]_0 = {}^{\Phi}H_0^{(n+1)\prime}[\Phi^{(\bar{n},n)}(H)]_0, \quad (4.7)$$

where $\{{}^{\Phi}H_0^{(n+1)\prime}\}_n$ is defined recurrently in terms of ${}^{\Phi}\delta'(\Lambda),{}^{\Phi}\beta_n'$ by (2.31a). In fact we prove a stronger result, stated in the following lemma.

Lemma 4.1: (a) There is an equivalence between the following statements.

Hypothesis 1: For all $n = 1,3,5,...$ and for every element of $\mathscr{F}$,

$${}^{\Phi}H_0^{(n+1)\prime} = {}^{\Phi'}H_0^{n+1}, \quad \text{with } {}^{\Phi}\delta' = {}^{\Phi'}\delta. \quad (4.8)$$

Here the superscript $\Phi'(n)$ denotes the coherent sequence of first-type ΦC's associated with n following the definition $\Phi^{(\bar{n},n)}\equiv\Phi^{(n)}N_2^{(n)}N_1^{(1)}$ [cf. (2.13) of I].

Hypothesis 2: Every system in $\mathscr{B}_\delta$ of the family $\{\delta,\Phi\}$ defined as

$${}^{\Phi}\delta_1(\Lambda) = \frac{1-2{}^{\Phi}\delta_3\Lambda a_1 - [1-4{}^{\Phi}\delta_3\Lambda a_1]^{1/2}}{2{}^{\Phi}\delta_3 a_1\Lambda^2} \quad (a_1 = {}^{\Phi}a_1 = N_{3(0)}), \quad (4.9a)$$

$${}^{\Phi}\delta_3(\Lambda) = \frac{6\Lambda(1+{}^{\Phi}\delta_1\Lambda){}^{\Phi}C_3}{1+6\Lambda(1+{}^{\Phi}\delta_1\Lambda)\left[\frac{3}{2}{}^{\Phi}b_3 - \frac{2}{3}{}^{\Phi}\delta_5{}^{\Phi}a_3\right]}, \quad (4.9b)$$

$${}^{\Phi}\delta_5(\Lambda) = \frac{15\Lambda(1+{}^{\Phi}\delta_1\Lambda){}^{\Phi}C_5}{1+15\Lambda(1+{}^{\Phi}\delta_1\Lambda)[{}^{\Phi}\alpha_5{}^{\Phi}b_5 - {}^{\Phi}\delta_7{}^{\Phi}\beta_7{}^{\Phi}a_5/15]}, \quad (4.9c)$$

$${}^{\Phi}\delta_n(\Lambda) = \frac{3\Lambda n(n-1)(1+{}^{\Phi}\delta_1\Lambda){}^{\Phi}C_n}{1+3\Lambda n(n-1)(1+{}^{\Phi}\delta_1\Lambda){}^{\Phi}\Delta_n(\Lambda)}, \quad \forall n\geqslant 7,$$

$${}^{\Phi}\Delta_n \overset{\text{def}}{\equiv} \frac{{}^{\Phi}\alpha_n{}^{\Phi}b_n}{n-1} - \frac{{}^{\Phi}\delta_{n+2}{}^{\Phi}\beta_{n+2}{}^{\Phi}a_n}{3n(n-1)}, \quad (4.9d)$$

has one and only one nontrivial solution ${}^{\Phi}\bar{\delta}\underset{\mathscr{M}_\delta}{\in}\widetilde{\mathscr{C}}_{\Lambda}$.

(b) Every nonlinear mapping ${}^{\Phi}\mathscr{M}_\delta$: $\mathscr{B}_\delta\to\mathscr{B}_\delta$ defined by each system of the family $\{\delta,\Phi\}$ above has a unique fixed point in $\widetilde{\mathscr{C}}_{\Lambda}$ if

$$0<\Lambda\lesssim 0.1(b_3/\bar{a}_3). \quad (4.10)$$

Proof: (a) To show the equivalence between hypotheses 1 and 2 one has only to repeat arguments analogous to the ones we have explicitly presented to prove the equivalence Theorem 3.1 concerning the zero-dimensional problem. We only note the slight difference in the form between every (δ,Φ) and (3.1) due to the presence in (δ,Φ) of the triplets of parameters associated with every coherent sequence of ΦC's, $\{{}^{\Phi}a_n,{}^{\Phi}b_n,{}^{\Phi}c_n\}$, and which in the case of (3.1) were trivially equal to the identity.

(b) By application of the analog of the proof of Theorem 3.2, we first ensure the stability of the subset $\widetilde{\mathscr{C}}_{\Lambda}\subset\mathscr{B}_\delta$ under the mapping ${}^{\Phi}\mathscr{M}_\delta$. Then by using an appropriate norm definition [cf. the analogous (3.14)] in $\mathscr{B}_\delta$, we verify the contraction mapping principle so that a unique fixed point ${}^{\Phi}\bar{\delta}\in\widetilde{\mathscr{C}}_{\Lambda}$ is ensured. We do not give here the detailed proofs because one can repeat all the appropriate estimations extensively presented in Sec. III in order to obtain the solution of (3.1). We only remark that concerning the stability obtained for $0<\Lambda\lesssim 0.1(\underline{b}_3/\bar{a}_3)$ one needs to prove a statement analogous to Lemma 3.2 that describes bounds and decrease properties of the functionals ${}^{\Phi}\Delta_n$ of (4.19d) (cf. Proposition 3.3 of III). This completes the proof of Lemma 4.1.

Proof of Theorem 4.1 continued: Now using the above result we have ${}^{\Phi}\delta' = {}^{\Phi'}\delta$; so the condition (4.7) for the conservation of signs and splitting [property 1 of Definition 2(e) of Φ_Λ] is satisfied in its stronger form: When $H\in\Phi_\Lambda$, for every $\Phi\in\mathscr{F}$ the sequence ${}^{\Phi}\bar{\delta}\in\widetilde{\mathscr{C}}_{\Lambda}$ is uniquely defined [as a unique solution in $\widetilde{\mathscr{C}}_{\Lambda}$ of the corresponding system (δ,Φ)]. For every coherent sequence of ΦC's in $\mathscr{F}$, the associated sequence $\{{}^{\Phi}\overline{H}_0^{(n+1)\prime}\}\in\mathscr{B}_0$ is defined as the corresponding sequence $\{{}^{\Phi'}H_0^{n+1}\}\in\mathscr{B}_0$, where $\Phi'(n)$ is defined in hypothesis 1 of Lemma 4.1 above.

(i)2 By application of the hypothesis $H\in\Phi_\Lambda$, i.e., properties $H^4(q,\Lambda)<0$ and $H^2(q,\Lambda)\Delta_F(q)>1$, to Eqs. (1.7) of the mapping $\mathscr{M}$, we write

$$H^{2\prime}(q,\Lambda)\Delta_F(q) = 1-\Lambda[N_3H^4](q\Lambda)\Delta_F(q)>1 \quad (4.11)$$

and

$$(-1)H^{4\prime}(q,\Lambda) \geqslant 6\Lambda\prod_{l=1}^{3}H^2(q_l)\Delta_F(q_l) \times\left\{1-\frac{1}{2}\sum_{\omega_3(J)}\frac{|N_2^{(2)}H^4(q_{j_2})|\,|N_1^{(1)}H^2(q_{j_1})|}{\Pi_{l=1}^3H^2(q_l)\Delta_F(q_l)} - \frac{|N_3H^6|}{6\Pi_{l=1}^3H^2(q_l)\Delta_F(q_l)}\right\}. \quad (4.12)$$

Using the zero-momentum dominance (Proposition 4.1) and the splitting properties in Φ_Λ we obtain

$$|N_3H^6|\leqslant 2N_{3(0)}\delta_5(\Lambda)24\Lambda\prod_{l=1}^{3}H^2(q_l)\Delta_F(q_l) \quad (\text{with } N_{3(0)}\sim 4.8\pi^2), \quad (4.13)$$

$$|N_2H^4|\leqslant 3N_{2(0)}\delta_3(\Lambda)\prod_{l=1}^{2}H^2(q_l)\Delta_F(q_l) \quad (\text{with } N_{2(0)}=\pi). \quad (4.14)$$

So the quantity inside the bracket of the right-hand side of (4.12) has a positive lower bound

$$(1-18\Lambda\pi-120\Lambda^2N_{3(0)})>0, \quad (4.15)$$

under the condition $0<\Lambda\lesssim 0.01$. This result ensures the negative sign property of $H^{4\prime}(q,\Lambda)$, $\forall q\in\mathscr{E}_{(q)}^6$, and, together with the result (4.11), completes the proof of the stability of Φ_Λ under the action of $\mathscr{M}$. Q.E.D.

(ii) Let $\overline{H}\in\mathscr{B}$ be a limit point of an arbitrary sequence $H_{(\nu)}\in\Phi_\Lambda$. That means $\forall\varepsilon>0$ $\exists$ $N(\varepsilon)>0$; $\forall\nu>N(\varepsilon)$ and $\forall n = 1,3,5,...,$ in view of the formulas 4.1 of $\mathscr{N}$, we have

$$\widehat{M}_n^{-1}|||H_0^{n+1}|||^{-1}\times\prod_{l=1}^{m_\Phi}|||H^{i_l+1}|||^{-1}|\widetilde{\Phi}^{(\bar{n},n)}|_0^{-1}|\Phi^{(\bar{n},n)}(H_\nu)H_\nu^{n+1} - \Phi^{(\bar{n}n)}(\overline{H})\overline{H}^{n+1}|<\varepsilon, \quad (4.16a)$$

or

$$| \, |\Phi^{(\bar{n},n)}(H_\nu)H_\nu^{n+1}| - |\Phi^{(\bar{n},n)}(\overline{H})\overline{H}^{n+1}| \, |$$
$$< \varepsilon |||H_0^{n+1}||| \, \widehat{M}_n \prod_l |||H^{i_l+1}||| \, |\widetilde{\Phi}^{(\bar{n},n)}|_0, \qquad (4.16b)$$

for every coherent sequence $\Phi \in \mathscr{F}$. We demonstrate that $\overline{H} \in \Phi_\Lambda$ also. We first show that at zero external momenta $\overline{H}$ satisfies the signs and splitting properties [property 1 of Definition 2(e) of Φ_Λ] by using a recursion. We suppose that the given coherent sequence $\Phi \in \mathscr{F}$ is reduced with respect to all Green's functions H^{i_l+1}, $l = 1,\ldots,m_\Phi$, apart from H^{n+1} itself. By hypothesis $H_{(\nu)} \in \Phi_\Lambda$, and so (4.16b) implies that $\forall \nu > N(\varepsilon)$,

$$| \, |{}^\Phi H_{0(\nu)}^{n+1}| - |\Phi^{(\bar{n},n)}\overline{H}^{n+1}|_0 / |\widetilde{\Phi}^{(\bar{n},n)}|_0| < \varepsilon |||H_0^{n+1}||| \widehat{M}_n. \qquad (4.17)$$

This means that we can write

$$|\Phi^{(\bar{n},n)}\overline{H}^{n+1}|_0 = |\widetilde{\Phi}^{(\bar{n},n)}|_0 \, |{}^\Phi \overline{H}_0^{n+1}|, \qquad (4.18)$$

with

$$|{}^\Phi \overline{H}_0^{n+1}| \stackrel{\text{def}}{\equiv} \lim_{\nu \to \infty} |{}^\Phi H_{0(\nu)}^{n+1}|. \qquad (4.18a)$$

To show that ${}^\Phi \overline{H}_0^{n+1}$ has the same sign as ${}^\Phi H_{0(\nu)}^{n+1}$, it is sufficient to suppose that they have opposite signs. Then in view of (4.18), (4.17) yields

$$|{}^\Phi H_{(\nu)}^{n+1} + {}^\Phi \overline{H}_0^{n+1}| < \varepsilon, \qquad (4.19)$$

which is an absurd result because $\varepsilon \to 0$, and the left-hand side is a purely positive quantity. So we conclude that

$${}^\Phi \overline{H}_0^{n+1} = \lim_{\nu \to \infty} {}^\Phi H_{0(\nu)}^{n+1}, \qquad (4.20)$$

with

$${}^\Phi \overline{H}_0^2 = 1 + {}^\Phi \bar{\delta}_1 \Lambda, \quad {}^\Phi \overline{H}_0^4 = -{}^\Phi \bar{\delta}_3(\Lambda)[{}^\Phi \overline{H}_0^2]^2,$$
$${}^\Phi H_0^6 = -4{}^\Phi \bar{\delta}_5 {}^\Phi \overline{H}_0^4 {}^\Phi \overline{H}_0^2, \qquad (4.21a)$$

and, recursively,

$${}^\Phi \overline{H}_0^{n+1} = -{}^\Phi \bar{\delta}_n {}^\Phi \bar{\beta}_n {}^\Phi \overline{H}_0^{n-1} {}^\Phi \overline{H}_0^2,$$

with

$${}^\Phi \bar{\beta}_n = \lim_{n \to \infty} {}^\Phi \beta_{n(\nu)}, \qquad (4.21b)$$

and ${}^\Phi \bar{\delta}$ the solution of (δ,Φ) found above in (i)1.

This result constitutes the first step of the following recurrence hypothesis. For a given $\Phi \in \mathscr{F}$ we suppose that (4.20) and (4.21) hold for all ΦC's not reduced with respect to all $H_{(\nu)}^{i_l+1}$ for $1 \leqslant l \leqslant \overline{m} - 1$, and partially reduced $\forall \overline{m} \leqslant l \leqslant m_\Phi$. This means that

$${}^\Phi \overline{H}_0^{i_l+1} = [\Phi^{(\bar{n},n)}(H)\overline{H}^{i_l+1}]_0 / [\widetilde{\Phi}^{(\bar{n},n)}(H)]_0,$$
$$\forall 1 < l \leqslant \overline{m} - 1, \qquad (4.22)$$

with the analog of (4.20) and (4.21) satisfied. Then it is easy to prove (using analogous arguments as above) that the recursion (4.22) and the corresponding formulas of (4.20) and (4.21) are verified when $l = \overline{m}$ and $\widetilde{\Phi}^{(\bar{n},n)}(H)$ reduced for $\overline{m} + 1 \leqslant l \leqslant m_\Phi$. By repeated application of this result we obtain the sign and splitting properties for the $\Phi^{(\bar{n},n)}(H)H^{i_{m_\Phi}+1}$, and this allows us to conclude that property 1 of Φ_Λ is verified by $\overline{H}$. Q.E.D.

In order to show that $\overline{H}$ verifies $\overline{H}^2(q,\Lambda)\Delta_F(q) \geqslant 1$ and $\overline{H}^4(q,\Lambda) < 0$ we use again a *reductio ad absurdum* argument. For example, in view of (4.11) we write for the four-point function

$$|H_{(\nu)}^4(q,\Lambda) - \overline{H}^4(q,\Lambda)| < \varepsilon. \qquad (4.23)$$

If we suppose that $\overline{H}^4(q,\Lambda) \geqslant 0$, then (4.23) yields

$$|H_\nu^4(q,\Lambda)| + |\overline{H}^4(q,\Lambda)| < \varepsilon, \qquad (4.24)$$

which is impossible in view of $\varepsilon \to 0$ and $|H_\nu^4(q,\Lambda)| > 0$. It follows that

$$\overline{H}^4(q,\Lambda) < 0. \qquad \text{Q.E.D.}$$

With the last result we completed the proof of the property $\overline{H} \in \Phi_\Lambda$ and, automatically, the demonstration of the closedness of Φ_Λ in $\mathscr{B}$. We now prove the contraction theorem and the iterative construction of the solution.

Theorem 4.2: (i) When Λ satisfies $0 < \Lambda \leqslant 0.1(b_3/\bar{a}_3)$, then there exists a unique fixed point $\overline{H}$ of the mapping $\mathscr{M}$ (1.7) in Φ_Λ.

(ii) The Φ iteration defined in Sec. II of I (Ref. 3) converges to this solution.

Proof: (i) We shall show that the condition $0 < \Lambda \lesssim 0.1(b_3/\bar{a}_3)$ imposed on Λ for the stability of Φ_Λ (cf. Theorem 4.1 above) is sufficient in order that the contractive mapping criterion be satisfied, for every pair of sequences $H_{(1)}, H_{(2)} \in \Phi_\Lambda$. In other words, we ensure that $\exists$ $0 < K(\Lambda) < 1$ such that

$$\|\mathscr{M}(H_{(1)}) - \mathscr{M}(H_{(2)})\|$$
$$\leqslant K(\Lambda)\|H_{(1)} - H_{(2)}\|, \quad \forall H_{(1)}, H_{(2)} \in \Phi_\Lambda. \qquad (4.25)$$

Let us consider an arbitrary coherent sequence of ΦC's $\{\Phi^{(\bar{n},n)}(H)\}_n \in \mathscr{F}$. For $n = 1$, using (1.7a) in view of the hypothesis and the norm definition (4.6), we have

$$|\Phi^{(\bar{n},1)}(H)(H_{(1)}^{2\prime} - H_{(2)}^{2\prime})|$$
$$\leqslant \Lambda N_{3(0)} |||H_0^4||| \, \|H_{(1)} - H_{(2)}\| \, |\Phi^{(\bar{n},1)}(H)|_0. \qquad (4.26)$$

We then define

$$K_1(\Lambda) = \Lambda N_{3(0)} |||H_0^4|||/4, \qquad (4.27)$$

which verifies $K_1(\Lambda) < 1$ if $\Lambda \lesssim 0.01$ (notice that $N_{3(0)} \sim 4.8\pi^2$, $|||H_0^4||| = 3!$), and such that

$$\frac{|\Phi^{(\bar{n},1)}(H)(H_{(1)}^{2\prime} - H_{(2)}^{2\prime})|}{[\widetilde{\Phi}_0]_0 \widehat{M}_1 |||H_0^2||| \Pi |||H_0^{i_l+1}|||} \leqslant K_1(\Lambda)\|H_{(1)} - H_{(2)}\|. \qquad (4.28)$$

For $n \geqslant 3$ we proceed in an analogous way. In view of (1.7b) we write

$$|\Phi^{(\bar{n},n)}(H)(H_{(1)}^{(n+1)\prime} - H_{(2)}^{(n+1)\prime})|$$
$$\leqslant \Lambda |\Phi^{(\bar{n},n)}(H) N_3 (H_{(1)}^{n+3} - H_{(2)}^{n+3})|$$
$$+ 3\Lambda \sum_{w_n(J)} \{ |\Phi^{(\bar{n},n)}(H)|N_2^{(j_2)}(H_{(1)}^{j_2+1} - H_{(2)}^{j_2+2})| \, |N_1^{(j_1)}H^{j_1+1}| + \Phi^{(\bar{n},n)}|N_2^{(j_2)}H_{(2)}^{j_2+2}| \, |N_1(H_{(1)}^{j_1+1} - H_{(2)}^{j_1+1})|$$

$$+6\Lambda \sum_{\omega_n(I)} |\Phi^{(\bar{n}n)}| N_{(1)}^{(i_1)}(H_{(1)}^{i_1+1} - H_{(2)}^{i_1+1})| \prod_{l=2}^{3} |N_1^{(i_l)} H_{(1)}^{i_l+1}| \,| + |\Phi^{(\bar{n},n)} N_1^{(i_l)} H_{(2)}^{i_1+1}|$$

$$\times |N_1^{(i_2)}(H_{(1)}^{i_2+1} - H_{(2)}^{i_2+1})| \, |N_1^{(i_3)} H_{(1)}^{i_3+1}| \,| + |\Phi^{(\bar{n},n)}(H) \prod_{l=1}^{2} |N_1^{(i_l)} H_{(2)}^{i_l+1}| \, |N_1(H_{(1)}^{i_3+1} - H_{(2)}^{i_3+1})| \,|. \tag{4.29}$$

By application of the hypothesis $H_{(1)}, H_{(2)} \in \Phi_\Lambda$, Proposition 4.1, and the norm $\widetilde{\mathscr{N}}$ (4.6), we obtain from (4.29)

$$|\Phi^{(\bar{n},n)}(H_{(1)}^{(n+1)\prime} - H_{(2)}^{(n+1)\prime})|$$
$$\leqslant \widehat{M}_{n-2} \widehat{M}_1^2 \|H_{(1)} - H_{(2)}\| \, |\widetilde{\Phi}_0^{(\bar{n},n)}|$$
$$\times \prod_{l=1}^{m_\Phi} |||H_0^{i_l+1}|||\{\Lambda N_{3(0)} |||H_0^{n+3}||| + 6\Lambda n \mathscr{T}_n N_{2(0)} |||H_0^{n+1}||| \, |||H_0^2||| + 9\Lambda n(n-1)\widetilde{\mathscr{T}}_n |||H_0^{n-1}||| \, |||H_0^2|||^2\}. \tag{4.30}$$

Now using the definitions, (4.6a) and (4.6b) of $\widetilde{\mathscr{N}}$ inside the brackets on the right-hand side of (4.30), together with Lemmas 2.6 and 2.7 for the upper bounds $\widetilde{\mathscr{T}}_n, \widetilde{\mathscr{T}}_n$ of β_n, α_n, respectively, and the values $N_{2(0)} = \pi$, $\tilde{\delta}_\infty^\Lambda \geqslant 2$, we finally can define a quantity $K(\Lambda)$,

$$K(\Lambda) \equiv \frac{\Lambda[2N_{3(0)}\tilde{\delta}_\infty^\Lambda + 6\pi + 3/(4\tilde{\delta}_\infty^\Lambda)]}{3 + (\delta_\infty^\Lambda)^2 N_{3(0)} + 3\pi\tilde{\delta}_\infty^\Lambda}, \tag{4.31}$$

which manifestly satisfies $0 < K(\Lambda) < 1$ for $0 < \Lambda \lesssim 0.1$, and is such that for every n,

$$\frac{|\Phi^{(\bar{n},n)}(H)(H_{(1)}^{(n+1)\prime} - H_{(2)}^{(n+1)\prime})|}{\Pi_{l=1}^{m_\Phi} |||H_0^{i_l+1}||| \widehat{M}_n ||| \, |\widetilde{\Phi}_0^{(n,n)}|} \leqslant K(\Lambda)\|H_{(1)} - H_{(2)}\|. \tag{4.32}$$

The above bound is uniform not only with respect to n but also with respect to every coherent sequence of the family $\mathscr{F}\{\Phi^{(\bar{n},n)}\}$. So by taking the supremum of (4.32) over n and Φ we finally obtain, in view of the definition of $\widetilde{\mathscr{N}}$, the condition (4.25). It then follows by application of the contractive mapping principle in Banach spaces that there exists a unique fixed point $\overline{H}$ of $\mathscr{M}$ in Φ_Λ. Q.E.D.

(ii) Taking into account the first-order results of the Φ iteration (cf. I, Sec. II) one can establish a recurrence hypothesis for all $1 \leqslant \bar{\nu} \leqslant \nu - 1$ and $1 \leqslant n \leqslant 3^{\nu-1}$, which states that all properties of Φ_Λ are satisfied together with the zero-momentum dominance and, therefore, the absolute upper bounds given by the norm $\widetilde{\mathscr{N}}$. Notice that for every $n > 3^{\nu-1}$ we have $H_{(\bar{\nu})}^{n+1} = 0$, and so $\{H_{(\bar{\nu})}\}_{\bar{\nu}<\nu-1} \notin \mathscr{B}$; this also means that $\{H_{(\bar{\nu})}\} \notin \Phi_\Lambda$. Then, using arguments analogous to the ones we presented for the proof of the stability of Φ_Λ ($\mathscr{C}_\Lambda$ in Sec. III) and Proposition 4.1, we obtain the validity of the above properties for $\bar{\nu} = \nu$ and for all H_ν^{n+1} with $1 \leqslant n \leqslant 3^\nu$. Then taking the limit $\nu \to \infty$, and so also $n \to \infty$, we obtain that $\{H_\infty\} \in \mathscr{B}$ and also that $H_\infty \in \Phi_\Lambda$. This means that in view of the result (i) H_∞ automatically coincides with the solution, i.e., $H_\infty = \overline{H}$. Q.E.D.

V. A SIMPLER AND UNIFIED CONTRACTIVITY PROOF

Despite the nontriviality of Φ convolutions in the dimensional cases, there exists an analogy in the nature of combinatorial properties of both mappings $\mathscr{M}_0$ and $\mathscr{M}$ in zero and two (one) dimensions. This fact implies that one can find appropriate Banach spaces $\widehat{\mathscr{B}}_0$ and $\widehat{\mathscr{B}}$, respectively, and show that the mappings $\widehat{\mathscr{M}}_0$ and $\widehat{\mathscr{M}}$ are contractive inside the corresponding subsets $\widehat{\Phi}_{0\Lambda}, \widehat{\Phi}_\Lambda$ by using exactly the same norm. This is the purpose of this section. We start with the precise definitions of the new spaces and subsets. We then show some conservation properties, introduce the new norm, and finally proceed to the proof of fixed-point theorem.

Definition 5(a) ($\widehat{\mathscr{B}}_0, \widehat{\mathscr{B}}$): We introduce the space $\widehat{\mathscr{B}}_0$ of double sequences,

$$H_0(\Lambda) \equiv \{\{H_{0\nu}^{n+1}(\Lambda)\}_n\}_\nu,$$
$$\Lambda \in \mathbb{R}, \quad \nu \in \mathbb{N}, \quad n = 1,3,5,..., \tag{5.1}$$

such that $\forall H_0 \in \widehat{\mathscr{B}}_0$, $\exists\, 0 < c_0(H) < \infty$, $0 < c_1(H) < \infty$,

$$|H_{0\nu}^{n+1}(\Lambda)| \leqslant n!\, c_0(H)^{(n-1)/2}[c_1(H)]^{n4^\nu}. \tag{5.1a}$$

Respectively, we define the space $\widehat{\mathscr{B}}$ of double sequences,

$$H(q,\Lambda) \equiv \{\{H^{n+1}(q,\Lambda)\}_n\}_\nu,$$
$$\Lambda \in \mathbb{R}, \quad \nu \in \mathbb{N}, \quad n = 1,3,5,..., \quad q \in \mathscr{E}_q^{rn}, \quad r = 1,2, \tag{5.2}$$

such that $\forall H \in \widehat{\mathscr{B}}$, $\exists\, 0 < c_0(H) < \infty$, $0 < c_2(H) < \infty$,

$$|H_\nu^{n+1}(q,\Lambda)|$$
$$\leqslant n![c_0(H)]^{(n-1)/2}[c_2(H)]^{(n\times 4^\nu)} M_{n+1}(q).$$

Here

$$M_{n+1}(q) = \begin{cases} q^2 + 1, & \text{if } n = 1, \\ 1, & \text{if } n \geqslant 3. \end{cases} \tag{5.2b}$$

Evidently

$$\widehat{\mathscr{B}}_0 \subset \widehat{\mathscr{B}}. \tag{5.3}$$

Due to the stability of $\mathscr{C}_\Lambda$ (cf. Sec. III, Theorem 3.2) and in view of the equivalence Theorem 3.1, one trivially obtains the stability of $\Phi_{0\Lambda}$. Furthermore, from the nontriviality of Φ_Λ (cf. Lemma 2.4') and the stability of it (Theorem 4.1), we can directly show that for every point $H \in \Phi_\Lambda$ there exists an integer $\nu(H)$ that corresponds to the νth order of an iteration of $\mathscr{M}_0$ starting from a precise point $\overline{H} \in \Phi_\Lambda$. In other words, we ensure that the following lemma is true.

Lemma 5.1: For all $H_0 \in \Phi_{0\Lambda}$ $\exists$ $\overline{H}_0 \in \Phi_{0\Lambda}$ and $\nu(H) \in \mathbb{N}$ such that

$$H_0 \equiv \mathscr{M}_0^{(\nu)} \overline{H}_0 \equiv H_\nu(\overline{H}_0), \tag{5.4}$$

where $\mathscr{M}_0^{(\nu)}$ means the νth-order application of the mapping $\mathscr{M}_0$ on $\overline{H}_0$.

Moreover one can easily obtain the following bounds.

Lemma 5.2: For all $H_0 \in \Phi_\Lambda$,

$$|H_0^{n+1}| \leqslant |||H^{n+1}|||, \quad \forall n = 1,3,5,..., \tag{5.5}$$

with

$$|||H^2||| = 2,$$
$$|||H^{n+1}||| = c_0^{(n-3)/2} n!, \quad c_0 = 4\delta_\infty^\Lambda. \tag{5.5a}$$

These considerations suggest the definition of a new subset $\widehat{\Phi}_{0\Lambda} \subset \widehat{\mathscr{B}}$ of double sequences, as follows.

Definition 5(b) (The subset $\widehat{\Phi}_{0\Lambda} \subset \widehat{\mathscr{B}}_0$):

$$\widehat{\Phi}_{0\Lambda} = \{H_0 \equiv \{H_{0\nu}\}_\nu \in \widehat{\mathscr{B}}_0 \colon H_{0\nu} \in \Phi_{0\Lambda}$$
$$\text{and } \mathscr{M}(H_{0\nu}) = H_{\nu+1}, \ \forall \nu \in \mathbb{N}\}. \tag{5.6}$$

Notice that following Lemma 5.1, the index ν in (5.6) has the precise meaning of the order of an iteration of $\mathscr{M}_0$ inside $\Phi_{0\Lambda}$. In an analogous way, taking into account the nontriviality and stability of $\Phi_\Lambda \subset \mathscr{B}$, one shows that the following is also true.

Lemma 5.3: (i) For all $H \in \Phi_\Lambda$ $\exists$ $\overline{H} \in \Phi_\Lambda$ and $\nu(H) \in \mathbb{N}$,

$$H \equiv H_\nu(\overline{H}) \equiv \mathscr{M}^{(\nu)} \overline{H}. \tag{5.7}$$

(ii) For all $H \in \Phi_\Lambda$,

$$|H^{n+1}(q,\Lambda)| \leqslant N_{n+1}(H) |||H^{n+1}||| M_{n+1}(q), \tag{5.8}$$

where

$$N_{n+1}(H) \leqslant \overline{N}_{n+1} \equiv K_0^{(n \times 4^\nu)}, \quad K_0 = 12\sqrt{c_0 \overline{N}_3}$$

[for c_0, cf. (5.5a)], and

$$\overline{N}_3 = \sup_q [N_3]. \tag{5.8a}$$

Moreover, from the above results we can define a subset $\widehat{\Phi}_\Lambda \subset \widehat{\mathscr{B}}$ that is the analog of $\widehat{\Phi}_{0\Lambda}$.

Definition 5(c) (The subset $\widehat{\Phi}_\Lambda \subset \widehat{\mathscr{B}}$):

$$\widehat{\Phi}_\Lambda = \{H \equiv \{H_\nu\}_\nu \in \widehat{\mathscr{B}} \colon H_\nu \in \Phi_\Lambda,$$
$$\mathscr{M}(H_\nu) = H_{\nu+1}, \ \forall \nu \in \mathbb{N}\}. \tag{5.9}$$

Here also the index ν has the precise meaning of an iteration of $\mathscr{M}$ inside Φ_Λ. One can verify the nonemptiness and stability of $\widehat{\Phi}_{0\Lambda}$ (resp. $\widehat{\Phi}_\Lambda$), so we state without proof the following lemma.

Lemma 5.4: (i) The subset $\widehat{\Phi}_{0\Lambda} \subset \widehat{\mathscr{B}}_0$ [Definition 5(b)] is nonempty and remains stable under the action of $\widehat{\mathscr{M}}_0$ when $0 < \Lambda \lesssim 0.1$.

(ii) The subset $\widehat{\Phi}_\Lambda \subset \widehat{\mathscr{B}}$ [Definition 5(c)] is nonempty and remains stable under the action of $\widehat{\mathscr{M}}$ when $0 < \Lambda \lesssim 0.006$.

We now introduce the following norm $\widehat{\mathscr{N}}$ in $\widehat{\mathscr{B}}$ [and consequently, by (5.3), also in $\widehat{\mathscr{B}}_0$].

Definition 5(d):

$$\|H\| = \sup_{n,\nu} |||H^{n+1}|||^{-1} M_{n+1}^{-1}(q) \overline{N}_{n+1}^{-1} |H_\nu^{n+1}| \tag{5.10}$$

[for $\overline{N}$ cf. (5.8a)]. Notice that in the zero-dimensional case $M_{n+1}(q) = 1$, $\forall n = 1,3$ [cf. (5.2b) and $q = 0$]. One easily verifies that $\forall H \in \widehat{\Phi}_{0\Lambda}$ or $H \in \widehat{\Phi}_\Lambda$ $\widehat{\mathscr{N}}$ is finite, and that both $\widehat{\mathscr{B}}_0$ and $\widehat{\mathscr{B}}$ are complete with respect to this norm. So the contractive mapping principle for Banach spaces can be verified. More precisely, we show the following theorem.

Theorem 5.1: (i) $\widehat{\Phi}_{0\Lambda}$ (resp. $\widehat{\Phi}_\Lambda$) is a closed subset of $\widehat{\mathscr{B}}_0$ (resp. $\widehat{\mathscr{B}}$) in the norm $\widehat{\mathscr{N}}$.

(ii) When $0 < \Lambda \lesssim 0.1$, there exists a unique fixed point of $\mathscr{M}_0$ inside $\Phi_{0\Lambda}$.

(iii) When $0 < \Lambda \lesssim 0.006$, there exists a unique fixed point of $\mathscr{M}$ inside Φ_Λ.

Proof: For both statements (i) and (ii) one has only to follow closely the arguments in the proof of Sec. IV for the contractivity of $\mathscr{M}$ in Φ_Λ, using the norm $\widehat{\mathscr{N}}$ (5.10). Nevertheless it is worthwhile to note that for the contractivity criterion for both $\widehat{\mathscr{M}}_0$ inside $\widehat{\Phi}_{0\Lambda}$ and $\widehat{\mathscr{M}}$ inside $\widehat{\Phi}_\Lambda$ one finds a weaker condition on Λ; to be precise, we only require $0 < \Lambda < 1$. The conditions appearing in the theorem are those imposed by the stability of $\Phi_{0\Lambda}$ (resp. Φ_Λ) or $\widehat{\Phi}_{0\Lambda}$ (resp. $\widehat{\Phi}_\Lambda$) under $\mathscr{M}_0$ (resp. $\mathscr{M}$).

VI. FINAL REMARKS

(1) The last result allows us to conclude that the construction of the solution is provided in a natural way by the Φ iteration despite the fact that at any finite order of this procedure the corresponding sequence $\{H_\nu^{n+1}\}$ is truncated and does not belong to $\mathscr{B}$.

(2) Another possibility for an iterative construction of the solution is to start the iteration from the solution $\{^{\Phi_0}H_0\} \in \mathscr{B}_0$ of the zero-dimensional-type system, where Φ_0 is the identity coherent sequence of ΦC's in $\mathscr{F}$. Evidently $^{\Phi}H_0 \in \Phi_\Lambda$; so by the result (i) of Theorem 4.2 [resp. (iii) of Theorem 5.1] we obtain $\overline{H}$ by contraction.

ACKNOWLEDGMENTS

The realization of this work would not have been possible without the constant encouragement by C. Manolessou, Professor L. Michel, Professor J. Iliopoulos, and Professor Ph. Blanchard. The author is grateful to all of them for their moral support in particularly difficult moments.

The study concerning the zero-dimensional problem originated from a suggestion of J. Frölich to whom the author is indebted. She has greatly profited from the criticisms and interesting remarks of Ph. Blanchard, S. Albeverio, J. Ginibre, A. Voros, and B. Grammaticos.

Section V originated from a discussion with J. Ginibre. The author is grateful to him both for his question on the possibility of a unique contractivity proof in zero to four dimensions and for his constant interest in her work.

The author also wishes to acknowledge the interest of Professor A. Wightman.

The author is thankful to IHES where the material presentation of this paper has been assured.

[1] M. Manolessou, Ann. Phys. (NY) **152**, 327 (1984).
[2] M. Manolessou and B. Ducomet, Ann. Inst. H. Poincaré **40**, 389 (1984).
[3] M. Manolessou, J. Math. Phys. **29**, 2092 (1988).
[4] K. Osterwalder and R. Shrader, Commun. Math. Phys. **42**, 281 (1975).
[5] M. Manolessou, Bielefeld University Research Center preprint BIBOS 52 a/b/c, June, 1985.
[6] M. Manolessou, "The Φ^4 equations of motion. III. The sweeping factors," IHES preprint P/87/50, July 1986.
[7] D. Atkinson, P. W. Johnson, and R. L. Warnock, Commun. Math. Phys. **33**, 221 (1973).
[8] M. Reed and B. Simon, *Functional Analysis*, Vol. I of *Methods of Modern Mathematical Physics* (Academic, New York, 1972), Theorem V.18.
[9] M. Manolessou, The Φ_4^4 equation of motion. The existence of a nontrivial solution (unpublished).
[10] A. Voros (private communication).

Conservation of momentum for systems of charged particles

Barton L. Willis
Department of Mathematics, University of North Carolina at Charlotte, Charlotte, North Carolina 28223

(Received 13 January 1988; accepted for publication 7 September 1988)

It is shown that the interaction part of the symmetric energy-momentum tensor for a system of charged particles is integrable over any forward light cone. A covariant definition of the mechanical momentum is then given and conservation of momentum is shown.

I. INTRODUCTION

The symmetric electromagnetic energy-momentum tensor T is quadratic in the field strengths. Consequently, the energy-momentum tensor for a system of particles is the sum of self-energy terms and interaction energy terms. If the charge on the ith particle is e_i, then the self-energy terms are proportional to e_i^2; the interaction energy terms are proportional to $e_i e_j$, with $i \neq j$. The usual definition of the field momentum[1-5]

$$P^\alpha(\sigma) = \frac{1}{c}\int_\sigma T^{\alpha\beta}\, d\sigma_\beta, \tag{1.1}$$

where σ is a spacelike hyperplane, suffers from several problems. The most important one is that the integral diverges for point particles. If T is replaced by the interaction part of the energy-momentum tensor $T_{\rm int}$, the integral still diverges because of the null fields that vanish at spatial infinity like r^{-1}. I will show that one must integrate the interaction energy-momentum tensor over a forward light cone to obtain a finite field momentum. This integral is finite because the null fields are asymptotically perpendicular to any forward light cone at spatial infinity. I prove this in Sec. III. The field momentum, defined in this way, is a tensor because a forward light cone is a Lorentz invariant set. On the other hand, if T is integrated over a spacelike hyperplane, then the field momentum is a tensor only if the current vanishes.[6]

This integrability result would not be worth much unless it is possible to use this new definition of the field momentum to prove conservation of momentum. I do this in Sec. IV by generalizing the nonrelativistic conservation law given by[7]

$$\frac{d}{dt}(P_{(\rm mech)} + P_{(\rm field)})_\alpha = \int_S \sum_{\beta=1}^{3} T_{\alpha\beta}\, da^\beta, \tag{1.2}$$

to one that is covariant and is valid for systems of particles. The right-hand side of Eq. (1.2) is identified with the rate of change of $P_{(\rm rad)}$, the field momentum escaping from a surface S that encloses the charges.

There are several problems associated with generalizing Eq. (1.2) to the relativistic multiparticle case. First, of course, is the problem that the usual definition of the field momentum gives a divergent result for point particles. Second, $P_{(\rm mech)}$, in Eq. (1.2), refers to the sum of each particles momentum at a particular instant of time; however, the classical concept of absolute simultaneity has no analog in relativistic mechanics. Goldstein[8] gives a thoughtful discussion of the problems associated with describing multiparticle relativistic systems.

A covariant formulation of the conservation law Eq. (1.2) should define $P_{(\rm mech)}$ and $P_{(\rm field)}$ in a similar fashion; that is, they should be defined on the same domain. In Sec. IV, I define the mechanical momentum as a function on all of Minkowski space, instead of as a function of the proper times of each particle. The time derivative in Eq. (1.2) generalizes to a directional derivative in an arbitrary timelike direction.

Finally, observe that if the surface S in Eq. (1.2) is pushed out towards infinity, then Larmor's formula[9] states that $P_{(\rm rad)}$ is given by

$$\frac{d}{dt}P_{(\rm rad)} = \frac{2e^2}{3c^5}|a|^2 v, \tag{1.3}$$

where v and a are the velocity and the acceleration of the particle. Larmor's formula is valid only for a single particle. In the multiparticle case, the radiation field will contain interaction terms not considered in Eq. (1.3).

The following notation will be used: The pseudometric tensor g is defined by $g^{\alpha\beta} = \operatorname{diag}\{1, -1, -1, -1\}$, all four-velocities have unit length, and the forward and backward light cones with apex at z will be denoted by $L^+(z)$ and $L^-(z)$, respectively.

II. PRELIMINARIES

Without loss of generality, I consider two charged particles whose world lines W and W^* are parametrized by $Z(\cdot)$ and $Z^*(\cdot)$, respectively. The electric charges associated with W and W^* will be denoted by e and e^*, respectively, and the electromagnetic fields by F and F^*, respectively. The interaction part of the symmetric energy-momentum tensor is

$$T^{\alpha\beta}_{(\rm int)} = (1/4\pi)[F^{\alpha\gamma}F^{*\beta}_{\gamma} + F^{*\alpha\gamma}F_\gamma{}^\beta + \tfrac{1}{2}g^{\alpha\beta}F^{\mu\nu}F^*_{\mu\nu}]. \tag{2.1}$$

One may verify that $T_{(\rm int)}$ is symmetric and that its divergence is given by

$$T^{\alpha\beta}_{(\rm int),\beta} = -(1/c)(F^{\alpha\gamma}j^*_\gamma + F^{*\alpha\gamma}j_\gamma), \tag{2.2}$$

where j and j^* are the currents associated with the world lines W and W^*, respectively. In particular, $T^{\alpha\beta}_{(\rm int),\beta}(x) = 0$ for $x \notin W, W^*$. The field F is given by

$$F^{\alpha\beta}(x) = e(\Gamma^\alpha M^\beta - \Gamma^\beta M^\alpha), \tag{2.3a}$$

where

$$\Gamma^\alpha = X^\alpha - Z^\alpha(\tau_0), \tag{2.3b}$$

$$\rho = \Gamma\cdot\dot{Z}(\tau_0), \tag{2.3c}$$

$$Q = \Gamma\cdot\ddot{Z}(\tau_0), \tag{2.3d}$$

$$M^\alpha = (1/\rho^2)\ddot{Z}^\alpha + [(1-Q)/\rho^3]\dot{Z}^\alpha. \tag{2.3e}$$

The retarded proper time τ_0 is defined implicitly as a function of X by

$$(X - Z(\tau_0))^2 = 0 \quad \text{and} \quad X^0 - Z^0(\tau_0) \geqslant 0. \tag{2.3f}$$

The quantities Γ^*, M^*, ρ^*, Q^*, and τ_0^* are defined similarly.

Driver[10] showed that existence of τ_0 is not automatic. In this paper, I consider only world lines for which τ_0 and τ_0^* exist on an appropriate set of Minkowski space. Since the integral of $T_{(\text{int})}$ is over a forward light cone, I need only existence on some forward light cone. The next proposition shows it is sufficient to assume that W and W^* intersect $L^\pm(x_0)$.

Proposition 1: Let W be a world line parametrized by $Z(\cdot)$ that intersects $L^\pm(x_0)$ at $Z(\tau^\pm)$. Then for every $X \in L^+(x_0)$ there is a unique τ_0 with $\tau_0 \in [\tau^-,\tau^+]$ so that $X - Z(\tau_0)$ is null and $X^0 - Z^0(\tau_0) \geqslant 0$.

Proof: Without loss of generality, let $x_0 = 0$. Let $X \in L^+(0)$, and let $\tau_0 \in [\tau^-,\tau^+]$ be the largest real number so that $X^0 \geqslant Z^0(\tau_0)$. The vectors $X, Z(\tau^\pm)$ are null vectors. Furthermore, $X^0, Z^0(\tau^+) > 0$, and $Z^0(\tau^-) < 0$. I show that $X - Z(\tau^-)$ is timelike and $X - Z(\tau_0)$ is spacelike:

$$\begin{aligned}[X - Z(\tau^-)]^2 &= X^2 - 2X\cdot Z(\tau^-) + Z^2(\tau^-) \\ &= -2X\cdot Z(\tau^-) \\ &\geqslant 0.\end{aligned} \tag{2.4}$$

The last inequality follows from the Cauchy–Schwartz inequality in Minkowski space.[11] If $\tau_0 = \tau^+$, then a similar calculation shows that $X - Z(\tau_0)$ is spacelike. If $\tau_0 < \tau^+$, then $X^0 - Z^0(\tau_0) = 0$, so once again $X - Z(\tau_0)$ is spacelike. Therefore, $[X - Z(\tau^-)]^2 \geqslant 0$ and $[X - Z(\tau_0)]^2 \leqslant 0$. Thus $[X - Z(\tau)]^2$ has at least one zero in the interval $[\tau^-,\tau_0]$. Numerous authors have shown uniqueness.[12]

From now on, I will assume that the world lines W and W^* intersect $L^\pm(x_0)$.

Finally, I compute the surface element $d\sigma_\beta$ for the light cone $L^+(x_0)$. A parametrization of $L^+(x_0)$ is

$$X^\alpha = x_0^\alpha + \langle\sqrt{u^2+v^2+w^2},u,v,w\rangle, \quad u,v,w \in R, \tag{2.5a}$$

$$= x_0^\alpha + \langle r,u,v,w\rangle, \tag{2.5b}$$

where $r = \sqrt{u^2+v^2+w^2}$. The Jacobian pseudotensor J^α is given by

$$J^\alpha = \epsilon^{\alpha\beta\gamma\delta}X_{\beta,u}X_{\delta,w}X_{\gamma,v} \tag{2.6a}$$

$$= (1/r)\langle r,u,v,w\rangle \tag{2.6b}$$

$$= (1/r)(X^\alpha - x_0^\alpha). \tag{2.6c}$$

The terms in Eq. (2.6a) have been ordered so that $J^0 > 0$. In Sec. III, I show that $T_{(\text{int})}$ is integrable over $L^+(x_0)$.

III. INTEGRABILITY

I now show that

$$P^\alpha(x_0) = \frac{1}{c}\int_{L^+(x_0)} T^{\alpha\beta}_{(\text{int})}\, d\sigma_\beta \tag{3.1}$$

is finite. Again, I assume $x_0 = 0$. To prove integrability towards spatial infinity, I need only to show that $T^{\alpha\beta}_{(\text{int})}J_\beta = O(r^{-4})$ as $r\to\infty$. In terms of Γ, Γ^*, M, M^*, and the field point X, the integrand $T^{\alpha\beta}_{(\text{int})}J_\beta$ is given by

$$\begin{aligned}T^{\alpha\beta}_{(\text{int})}J_\beta &= (ee^*/4\pi r)[(\Gamma^*\cdot M)(X\cdot M^*) - (M\cdot M^*)(\Gamma^*\cdot X)]\Gamma^\alpha + (ee^*/4\pi r)[(\Gamma\cdot M^*)(\Gamma^*\cdot X) \\ &\quad - (\Gamma^*\cdot\Gamma)(M^*\cdot X)]M^\alpha + (ee^*/4\pi r)[(\Gamma\cdot M^*)(X\cdot M) - (M^*\cdot M)(X\cdot\Gamma)]\Gamma^{*\alpha} \\ &\quad + (ee^*/4\pi r)[(\Gamma^*\cdot M)(\Gamma\cdot X) - (\Gamma^*\cdot\Gamma)(M\cdot X)]M^{*\alpha} \\ &\quad + (ee^*/4\pi r)[(\Gamma\cdot\Gamma^*)(M\cdot M^*) - (\Gamma\cdot M^*)(\Gamma^*\cdot M)]X^\alpha.\end{aligned} \tag{3.2}$$

For $X \in L^+(0)$, the scalars in Eq. (3.2) simplify to

$$\Gamma\cdot\Gamma^* = -\tfrac{1}{2}(Z - Z^*)^2, \tag{3.3a}$$

$$\Gamma\cdot X = -\tfrac{1}{2}Z^2, \tag{3.3b}$$

$$\Gamma^*\cdot X = -\tfrac{1}{2}Z^{*2}, \tag{3.3c}$$

$$\Gamma\cdot M^* = 1/\rho^{*2} + (Z^* - Z)\cdot M^*, \tag{3.3d}$$

$$\Gamma^*\cdot M = 1/\rho^2 - (Z^* - Z)\cdot M, \tag{3.3e}$$

$$X\cdot M = 1/\rho^2 + Z\cdot M, \tag{3.3f}$$

$$X\cdot M^* = 1/\rho^{*2} + Z^*\cdot M. \tag{3.3g}$$

The Cauchy–Schwartz inequality implies that there are constants a, b, c, and d so that for $X \in L^+(0)$, $X \neq 0$,

$$0 < ar + b \leqslant \rho \leqslant cr + d. \tag{3.4}$$

And similarly for ρ^*. Using Eqs. (3.2a)–(3.2g) in the expression for $T^{\alpha\beta}_{(\text{int})}J_\beta$, along with the inequality (3.4), one may verify that each term of $T^{\alpha\beta}_{(\text{int})}J_\beta$ is $O(r^{-4})$ as $r\to\infty$. Thus $T^{\alpha\beta}_{(\text{int})}J_\beta$ is integrable towards spatial infinity.

I now show the singularities of $T_{(\text{int})}$ at $Z(\tau^\pm)$ are integrable. For the singularity at $Z(\tau^+)$, Γ has an order-1 zero at $Z(\tau^+)$ and $M = O(r^{-3})$ near $Z(\tau^+)$. Thus each term of Eq. (3.2) is $O(r^{-2})$ near $Z(\tau^+)$. These are integrable singularities in R^3. A similar argument shows integrability of the $Z(\tau^-)$ singularity.

I emphasize that Eqs. (3.3a)–(3.3g) are not true for X in some spacelike hyperplane. Consequently, $\int_\sigma T^{\alpha\beta}_{(\text{int})}\, d\sigma_\beta$ diverges for point particles if σ is a spacelike hyperplane because of the $O(r^{-1})$ behavior of the null fields at infinity. The assumption that the surface is a forward light cone is crucial to the proof.

IV. CONSERVATION OF MOMENTUM

I now show conservation of momentum. For any x_0 for which Z and Z^* intersect $L^{\pm}(x_0)$, I define the mechanical momentum $P_{(\mathrm{mech})}$ by

$$P_{(\mathrm{mech})}(x_0) = mc\dot{Z}(\tau^+) + m^*c\dot{Z}^*(\tau^{*+}), \tag{4.1}$$

where m and m^* are the rest masses associated with the world lines W and W^*, respectively, and $\dot{Z}(\tau^+)$ and $\dot{Z}^*(\tau^{*+})$ are the velocities of the particles where W and W^* intersect $L^+(x_0)$. This definition of $P_{(\mathrm{mech})}$ is covariant because all observers agree on the points of intersection of $L^+(x_0)$ and W and W^*. The conservation law is given in the following proposition.

Proposition 2: Let x_0 be any point so that W and W^* intersect $L^{\pm}(x_0)$. Let $P_{(\mathrm{int})}$ be defined by Eq. (3.1), where the normal to the surface $L^+(x_0)$ is chosen to be in $L^+(x_0)$. Then for any world lines that are solutions to the equations of motion,

$$mc\ddot{Z}^{\alpha} = (e/c)F^{*\alpha\beta}\dot{Z}_{\beta},$$

$$m^*c\ddot{Z}^{*\alpha} = (e^*/c)F^{\alpha\beta}\dot{Z}^*_{\beta},$$

the directional derivative of the $P_{(\mathrm{mech})} + P_{(\mathrm{int})}$ in a timelike direction l is given by

$$[P^{\alpha}_{(\mathrm{mech})} + P^{\alpha}_{(\mathrm{int})}]^{,\beta}(x_0)l_{\beta} = -\lim_{\epsilon\to 0}\lim_{R\to\infty}\int_{S_R} T^{\alpha\beta}\,d\sigma_{\beta}, \tag{4.2}$$

where S_R is the portion of the hyperplane $X^0 = R$ that lies between the cones $L^+(x_0+\epsilon l)$ and $L^+(x_0)$. Choose the normal to S_R to have a positive zeroth component.

Proof: The divergence of $P_{(\mathrm{mech})}$ is given by

$$P^{\alpha,\beta}_{(\mathrm{mech})}(x_0) = mc\ddot{Z}^{\alpha}(\tau^+)(\Gamma^{\beta}/\rho) + m^*c\ddot{Z}^{*\alpha}(\tau^{*+})(\Gamma^{*\beta}/\rho^*), \tag{4.3}$$

where Γ, Γ^*, ρ, and ρ^* are defined in Eqs. (2.3b) and (2.3c) with X replaced by x_0. I now compute the directional derivative of $P_{(\mathrm{int})}$ in some arbitrary timelike direction l. Thus

$$P^{\alpha,\beta}_{(\mathrm{int})}(x_0)l_{\beta} = \frac{1}{c}\lim_{\epsilon\to 0}\frac{1}{\epsilon}\left[\int_{L^+(x_0+\epsilon l)} T^{\alpha\lambda}_{(\mathrm{int})}\,d\sigma_{\lambda} - \int_{L^+(x_0)} T^{\alpha\lambda}_{(\mathrm{int})}\,d\sigma_{\lambda}\right]. \tag{4.4}$$

Surround the world lines W and W^* by tubes T_{Δ} and T^*_{Δ}, respectively, of radius Δ. I may now write Eq. (4.4) as

$$P^{\alpha,\beta}_{(\mathrm{int})}(x_0)l_{\beta} = \frac{1}{c}\lim_{\epsilon\to 0}\lim_{\Delta\to 0}\lim_{R\to\infty}\frac{1}{\epsilon}\times\left[\int_{L^+(x_0+\epsilon l)} - \int_{L^+(x_0)} - \int_{T_{\Delta}} - \int_{T^*_{\Delta}} + \int_{S_R} + \int_{T_{\Delta}} + \int_{T^*_{\Delta}} - \int_{S_R} T^{\alpha\lambda}_{(\mathrm{int})}\,d\sigma_{\lambda}\right]. \tag{4.5}$$

The first four integrals combine to an integral over a closed and bounded surface. Since $T^{\alpha\lambda}_{(\mathrm{int}),\lambda} = 0$ on the interior of this surface, Gauss's theorem implies that these integrals combine to zero. I now compute

$$\lim_{\Delta\to 0}\int_{T_{\Delta}} T^{\alpha\lambda}_{(\mathrm{int})}\,d\sigma_{\lambda}. \tag{4.6}$$

I parametrize the tube T_{Δ} by

$$X^{\alpha} = Z^{\alpha}(\tau) + \Delta L^{\alpha\beta}(\tau)n_{\beta}(\theta,\phi), \tag{4.7a}$$

$$n^{\beta} = \langle 1,\sin\phi\cos\theta,\sin\phi\sin\theta,\cos\phi\rangle, \tag{4.7b}$$

where $L^{\alpha\beta}(\tau)$ is a Lorentz boost chosen so that

$$L^{\alpha\beta}\dot{Z}_{\beta} = \delta^{\alpha 0}. \tag{4.7c}$$

Note that L is symmetric. The tube T_{Δ} is a surface of constant ρ. Its surface element is $d\sigma^{\alpha} = J^{\alpha}\,d\tau\,d\theta\,d\phi$ with J^{α} given by

$$J^{\alpha} = \Delta^2\sin\phi\dot{Z}^{\alpha} - \Delta(1-Q)\sin\phi\Gamma^{\alpha}, \tag{4.8}$$

with Γ and Q defined in Eqs. (2.3b) and (2.3d). To evaluate the integral in Eq. (4.6), integrals of products of Γ's are needed. Defining $d\Omega \equiv \sin\phi\,d\phi\,d\theta$, I find[13]

$$\int \Gamma^{\alpha}\,d\Omega = 4\pi\dot{Z}^{\alpha}, \tag{4.9a}$$

$$\int \Gamma^{\alpha}\Gamma^{\beta}\,d\Omega = \frac{16\pi}{3}\dot{Z}^{\alpha}\dot{Z}^{\beta} - \frac{4\pi}{3}g^{\alpha\beta}, \tag{4.9b}$$

$$\int \Gamma^{\alpha}\Gamma^{\beta}\Gamma^{\gamma}\,d\Omega = 8\pi\dot{Z}^{\alpha}\dot{Z}^{\beta}\dot{Z}^{\gamma} - \frac{2\pi}{3}g^{[\alpha\beta}\dot{Z}^{\gamma]}, \tag{4.9c}$$

$$\int \Gamma^{\alpha}\Gamma^{\beta}\Gamma^{\gamma}\Gamma^{\delta}\,d\Omega = \frac{64\pi}{5}\dot{Z}^{\alpha}\dot{Z}^{\beta}\dot{Z}^{\gamma}\dot{Z}^{\delta} - \frac{2\pi}{5}g^{[\alpha\beta}\dot{Z}^{\gamma}\dot{Z}^{\delta]} + \frac{\pi}{30}g^{[\alpha\beta}g^{\gamma\delta]}. \tag{4.9d}$$

The square brackets in the indices means symmetrization. These formulas along with the explicit formulas for F and F^* given by Eqs. (2.3a)–(2.3e) allow one to show that

$$\lim_{\Delta\to 0}\int_{T_{\Delta}} T^{\alpha\lambda}d\sigma_{\lambda} = -\frac{e}{c}\int_{\tau_1}^{\tau_2} F^{*\alpha\lambda}\dot{Z}_{\lambda}\,d\tau, \tag{4.10a}$$

where τ_1 and τ_2 are defined by $Z(\tau_2)\in L^+(x_0+\epsilon l)$ and $Z(\tau_1)\in L^+(x_0)$; τ^*_1 and τ^*_2 are defined similarly. Furthermore,

$$\lim_{\Delta\to 0}\int_{T^*_{\Delta}} T^{\alpha\lambda}\,d\sigma_{\lambda} = -\frac{e^*}{c}\int_{\tau^*_1}^{\tau^*_2} F^{\alpha\lambda}\dot{Z}_{\lambda}\,d\tau. \tag{4.10b}$$

The integrals in Eqs. (4.10a) and (4.10b) can be identified as line integrals along each world line. However, $\tau_2 - \tau_1 = \Gamma\cdot l + O(\epsilon)$ and $\tau^*_2 - \tau^*_1 = \Gamma^*\cdot l + O(\epsilon)$, where $\Gamma \equiv x_0 - Z(\tau^+)$ and $\Gamma^* \equiv x_0 - Z^*(\tau^{*+})$. Therefore,

$$\lim_{\epsilon\to 0}\lim_{\Delta\to 0}\int_{T^*_{\Delta}} + \int_{T_{\Delta}} T^{\alpha\lambda}_{(\mathrm{int})}\,d\sigma_{\lambda} = -\frac{e}{c}(\Gamma\cdot l)F^{*\alpha\lambda}\dot{Z}_{\lambda} - \frac{e^*}{c}(\Gamma^*\cdot l)F^{\alpha\lambda}\dot{Z}^*_{\lambda} \tag{4.11a}$$

$$= -P^{\alpha,\lambda}_{(\mathrm{mech})}(x_0)l_{\lambda}. \tag{4.11b}$$

In deriving Eq. (4.11b), the equations of motion have been used. Therefore, the Proposition has been proved.

V. DISCUSSION

The standard definition of the electromagnetic momentum suffers from several problems; it is not a tensor unless the current is zero, and it diverges for point particles. I solved these problems by integrating the energy-momentum tensor over a Lorentz invariant set. A definition of the mechanical momentum of a system of particles has also caused problems because of the lack of any notion of simultaneity in relativity theory. This problem was solved by viewing the mechanical momentum as a function defined over all of Minkowski space rather than as a function of the proper times of each particle.

[1]J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1975), 2nd ed., p. 793, Eq. (17.40).

[2]F. Rohrlich, *Classical Charged Particles* (Addison–Wesley, Reading, MA, 1965), pp. 89 and 90, Eq.(4-123).

[3]A. O. Barut, *Electrodynamics and Classical Theory of Fields and Particles* (Dover, New York, 1980), pp. 105–111, Eq. (3.76).

[4]L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Pergamon, Oxford, 1975), 4th Revised English ed., p. 79, Eq. (32.11).

[5]E. J. Konopinski, *Electromagnetic Fields and Relativistic Particles* (McGraw-Hill, New York, 1981), p. 435.

[6]Reference 2, p. 280.

[7]See, for example, Ref. 1, p. 239, Eq. (6.122); Ref. 4, p. 76, Eqs. (31.4)–(31.6); or Ref. 5, p. 157, Eqs. (6.15) and (6.16). References 4 and 5 imply that by moving the surface S out to infinity the right-hand side of Eq. (1.2) vanishes. In general, this is false because the null fields spatial infinity vanish like r^{-1}.

[8]H. Goldstein, *Classical Mechanics* (Addison–Wesley, Reading, MA, 1980), 2nd ed., pp. 331 and 332.

[9]Reference 3, p. 183, Eq. (5.55).

[10]R. D. Driver, Ann. Phys. (NY) **21**, 122 (1963).

[11]Reference 3, pp. 9 and 10.

[12]See, for example, Ref. 10.

[13]Two quick checks of Eqs. (4.9a)–(4.9d) are possible. First, one may show that Eq. (4.9d) implies Eq. (4.9c) which implies Eq. (4.9b) which implies Eq. (4.9a) by contracting each side with $\dot{Z}$. Second, since Γ is null, contraction of any two indices should give zero.

Decomposition of the scalar superfield in ten dimensions

P. Kwon and M. Villasante
Department of Physics, University of California, Los Angeles, California 90024

(Received 23 June 1988; accepted for publication 31 August 1988)

Explicit expressions for the irreducible superfields included in the ten-dimensional scalar superfield are derived by using a method based in a Cartan subalgebra.

I. INTRODUCTION

The systematic study of off-shell field representations of the supersymmetry algebra (local as well as global) requires knowledge of the irreducible representations of the super-Poincaré algebra. This analysis provides the auxiliary field structure that has been traditionally found by guessing it.[1] Superconformal methods provide a unified framework for $N=1$ and $N=2$ supersymmetries in four dimensions,[2] but a systematic analysis in higher dimensions is lacking due to the growing complexity of the superconformal algebra.[3]

The super-Poincaré algebra, on the other hand, retains its simple form in any number d of dimensions and the representations can be studied by similar methods for any d, with only the amount of technical difficulty involved being different. The methods alluded to above involve knowledge of the Casimir operators of the algebra as well as their eigenvalues, which makes possible the decomposition of a general superfield into irreducible ones. This procedure was pioneered by Sokatchev who applied it to the $N=1$ case in four dimensions.[4] Subsequently, it was successfully applied to the $N=2$ four-dimensional case.[5] More recently, the set of Casimirs for the $N=1$ super-Poincaré algebra in d dimensions SP_d has been given for any d in the massive ($P^2 \neq 0$) case.[6] The Casimir operators for SP_d are the square of the momentum P^2 and the Casimirs of $\mathrm{SO}(d-1)$. As a highly nontrivial example, the scalar superfield in 11 dimensions was decomposed.[6] However, the ten-dimensional case was found to be exceptional in the sense that the quadratic Casimir of $\mathrm{SO}(d-1)$ is degenerate in the irreducible pieces of the scalar superfield, and cannot be used to separate them.[7] Up to now this difficulty had prevented explicit expressions for the irreducible parts in terms of ordinary fields from being given. In principle, solutions can be given by applying projection operators, which are known, to general superfields, but these solutions are utterly impractical.

Some of these problems are present in the massless ten-dimensional case where they were solved and the scalar superfield was decomposed, after finding the Casimirs of SP_d for massless representations.[8] The irreducible superfields were constructed from the eigenstates of the Cartan subalgebra of the appropriate little algebra of SP_d, which turned out to be an $\mathrm{SO}(d-2)$ algebra.[8]

Here, we finally present the irreducible pieces contained in the ten-dimensional scalar superfield following this latter method which, however, presented special difficulties in this case.

We will structure the paper as follows. In Sec. II, we review some basic facts about the super-Poincaré algebra SP_d. In Sec. III, we describe, in general, the Cartan subalgebra method to construct irreducible superfields. This method is applied to the four-dimensional case in Sec. IV, in order to show that the solutions obtained by this method coincide with the ones previously obtained by other methods. Finally, the main object of this paper, the solution to the ten-dimensional problem, is detailed in Sec. V.

II. SOME FACTS ABOUT THE SUPER-POINCARÉ ALGEBRA SP_d

The super-Poincaré algebra in d dimensions SP_d is the grading of the Poincaré algebra P_d with a Majorana spinor Q, which has $2^{[d/2]}$ components. We can represent it in a superspace (x^A, θ^α), where the θ^α are anticommuting coordinates arranged in a Majorana spinor. Then Q and the covariant derivative D, which must anticommute with Q, can be expressed by

$$Q = i\left(\frac{\partial}{\partial\bar\theta} + \frac{1}{2}\not{P}\theta\right), \quad D = i\left(\frac{\partial}{\partial\bar\theta} - \frac{1}{2}\not{P}\theta\right),$$
$$\{Q,D\} = 0. \tag{2.1}$$

The generalized angular momentum operators[6]

$$U_{AB} = \Pi_A{}^E \Pi_B{}^F J_{EF} - (i/4P^2)\bar{Q}\not{P}\Gamma_{AB}Q,$$
$$\Pi_A{}^B = \delta_A{}^B - (1/P^2)P_A P^B, \tag{2.2}$$

satisfy

$$[U_{AB}, P^C] = [U_{AB}, Q] = 0,$$
$$P^A U_{AB} = 0, \tag{2.3}$$

and

$$[U_{AB}, U^{CD}] = -4i\Pi_{[A}{}^{[C} U_{B]}{}^{D]},$$
$$A,B,C,D = 0,1,\dots,d-1. \tag{2.4}$$

The algebra (2.4) becomes, in the rest frame, the set of commutation relations of $\mathrm{SO}(d-1)$,

$$[U_{ij}, U^{kl}] = -4i\delta_{[i}{}^{[k} U_{j]}{}^{l]},$$
$$i,j,k,l = 1,\dots,d-1. \tag{2.5}$$

So the operators U_{AB} are the generators of the even part of the little algebra and for the scalar superfield they are simply given by[6]

$$U_{AB} = -(i/4P^2)\bar{D}\not{P}\Gamma_{AB}D \tag{2.6}$$

in terms of the covariant derivatives.

In ten dimensions Q decomposes into two mutually anticommuting Majorana–Weyl spinors $Q^{(+)}, Q^{(-)}$ (the same for D and θ).[7] To keep both $Q^{(+)}$ and $Q^{(-)}$ amounts to working with an $N=2$ extended super-Poincaré algebra. We will consider only the super-Poincaré algebra SP_{10}^+ which in-

 0022-2488/89/010201-12$02.50

cludes $Q^{(+)}$, but not $Q^{(-)}$. That is to say that we are going to include only $\theta^{(-)}$ as the anticommuting coordinate in our superspace, and therefore deal only with chiral superfields $\phi(x,\theta^{(-)})$.[7] In that case

$$
\begin{aligned}
Q^{(+)} &= \Pi^{(+)}Q = i\left(\frac{\partial}{\partial\bar{\theta}^{(-)}} + \frac{1}{2}\not{P}\theta^{(-)}\right),\\
D^{(+)} &= \Pi^{(+)}D = i\left(\frac{\partial}{\partial\bar{\theta}^{(-)}} - \frac{1}{2}\not{P}\theta^{(-)}\right),\\
\theta^{(-)} &= \Pi^{(-)}\theta, \quad \Pi^{(\pm)} = \tfrac{1}{2}(I \pm \Gamma_{(11)}),
\end{aligned}
\tag{2.7}
$$

and instead of (2.2) we have

$$
U_{AB}^{(+)} = \Pi_A{}^E\Pi_B{}^F J_{EF} - (i/4P^2)\bar{Q}^{(+)}\not{P}\Gamma_{AB}Q^{(+)}. \tag{2.8}
$$

For the chiral scalar superfield we obtain the analog of (2.6) (Ref. 7),

$$
U_{AB}^{(+)} = -(i/4P^2)\bar{D}^{(+)}\not{P}\Gamma_{AB}D^{(+)}. \tag{2.9}
$$

Both (2.8) and (2.9) satisfy the algebra (2.4). In Ref. 7, it was shown by the Casimir operator approach that (2.9) implies that the chiral scalar superfield contains three irreducible representations, labeled by the SO(9) representations [2], $[\frac{3}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}]$, and [111]. From our experience with the massless case,[8] we know that the fermionic representation $[\frac{3}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}]$ will be easy to isolate and the nontrivial task will be the separation of the two bosonic representations [2] and [111]. As mentioned before, the straightforward procedure of constructing projection operators made out of Casimir operators and their eigenvalues (which are known) is not useful, because the operator involved (the quartic Casimir in this case) is too complicated to handle.[7]

III. THE CARTAN SUBALGEBRA METHOD

In this section we will describe the Cartan subalgebra approach to construct irreducible superfields. In this approach irreducible superfields are obtained systematically from the eigenstate associated to the highest weight vector of the corresponding irreducible representation.

The generators of the Cartan subalgebra of the SO$(d-1)$ little algebra (2.5) are

$$
H_I = U_{2I-1,2I}, \quad I = 1,\ldots,n, \tag{3.1}
$$

the number $n = [(d-1)/2]$ being the rank of SO$(d-1)$. An eigenstate $|\psi\rangle$ of these generators will satisfy

$$
\mathbf{H}|\psi\rangle = \mathbf{w}|\psi\rangle \tag{3.2}
$$

where $\mathbf{H}$ is the vector operator $\mathbf{H} = (H_1,\ldots,H_n)$ and $\mathbf{w} = (w_1,\ldots,w_n)$ is the weight vector associated to $|\psi\rangle$.

A complete basis of such eigenstates $|\psi\rangle$ for an irreducible representation can be generated from any one of them by the raising and lowering operators corresponding to the root vectors of SO$(d-1)$. The complete set of weight vectors $\mathbf{w}$, each associated with one eigenstate, form the weight diagram of the irreducible representation. The complete weight diagram of any irreducible representation can be generated from the highest weight by using Dynkin's algorithm[9] supplemented by a formula to compute the degeneracy of each weight.[10] In this procedure every weight vector is obtained by subtracting a certain sum of root vectors from the highest weight. Now, given that there is a correspondence between weights $\mathbf{w}$ and eigenstates $|\psi\rangle$ and between roots $\boldsymbol{\alpha}$ and raising/lowering operators $E(\boldsymbol{\alpha})$, we know how to generate all the eigenstates $|\psi\rangle$ from the one associated to the highest weight $\boldsymbol{\Lambda}$ by applying appropriate sequences of $E(\boldsymbol{\alpha})$.

In constructing the irreducible superfields, we will generate all the states of the representation from the one associated with the highest weight, following the procedure described above. Thus we need the weight diagrams of the SO(9) irreducible representations with highest weights $[2] = (2,0,0,0)$ and $[111] = (1,1,1,0)$. These weight diagrams are shown in Figs. 1 and 2, respectively.

The roots of SO$(d-1)$ are

$$
\eta\hat{e}_I + \eta'\hat{e}_J, \quad I,J = 1,\ldots,n, \quad I<J; \quad \eta,\eta' = + \text{ or } -, \tag{3.3a}
$$

when $d-1$ is even, and we have the additional set

$$
\eta\hat{e}_I, \quad I = 1,\ldots,n, \tag{3.3b}
$$

when $d-1$ is odd. More important are the simple positive roots

$$
\boldsymbol{\alpha}_a = \hat{e}_a - \hat{e}_{a+1}, \quad a = 1,\ldots,n-1, \tag{3.4a}
$$

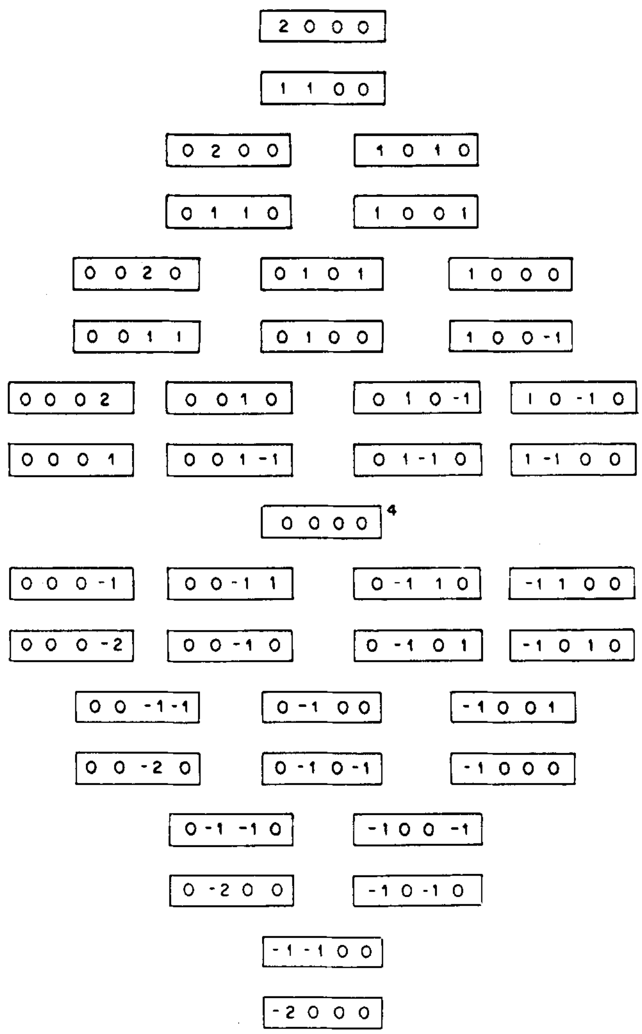

FIG. 1. Weight system of the representation [2] of SO(9) with the Dynkin layers displayed. The exponent on a weight denotes its multiplicity.

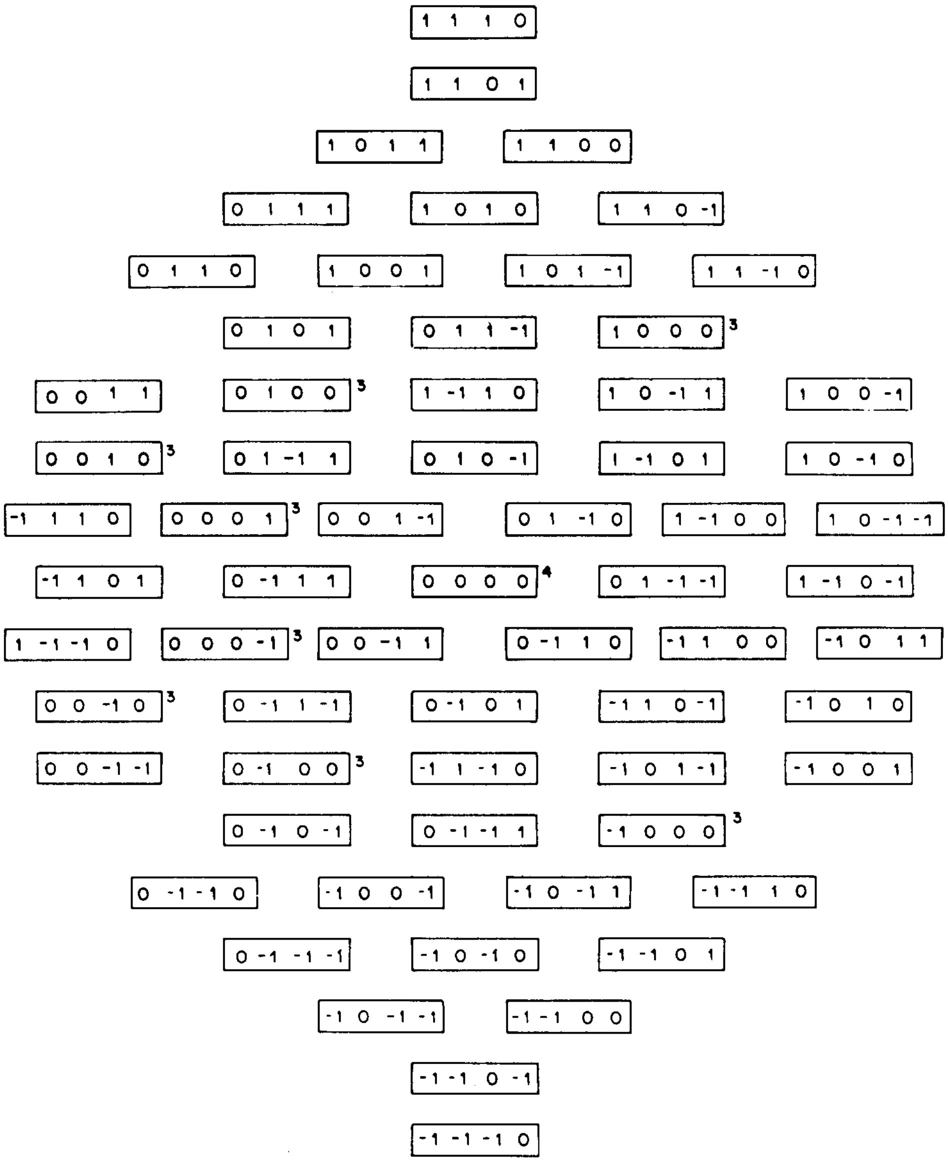

FIG. 2. Weight system of the representation [111] of SO(9) displaying the Dynkin layers and weight multiplicities.

supplemented by

$$\alpha_n = \hat{e}_{n-1} + \hat{e}_n, \quad \text{for } d-1 \text{ even}, \tag{3.4b}$$

or

$$\alpha_n = \hat{e}_n, \quad \text{for } d-1 \text{ odd}. \tag{3.4c}$$

Now let us construct the operators associated to the roots. Hereafter all calculations will be carried out in the collinear frame where the momentum has the form

$$P_A = P_0(1,0,...,0,z), \quad z = \sqrt{1-(M/P_0)^2}, \quad 0 \leqslant z < 1. \tag{3.5}$$

In (3.5), M^2 is the eigenvalue of P^2 and the massless limit is easily obtained by taking $z \to 1$.

In this frame the components of U_{AB} that satisfy the SO$(d-1)$ algebra (2.5) are

$$L_{ij} = U_{ij} = -(i/4P^2)\overline{D}\not{P}\Gamma_{ij}D,$$
$$L_{i,d-1} = U_{i,d-1} + zU_{0i} = -(i/4P^2)\overline{D}\not{P}\Gamma_i\widetilde{\Gamma}_{d-1}D$$
$$(i,j = 1,...,d-2), \tag{3.6}$$

where

$$\widetilde{\Gamma}_{d-1} = \Gamma_{d-1} - z\Gamma_0. \tag{3.7}$$

For later convenience we also define

$$\widetilde{\Gamma}_{-(d-1)} = (1/P_0)\not{P} = \Gamma_0 - z\Gamma_{d-1}. \tag{3.8}$$

These $\widetilde{\Gamma}$ matrices anticommute:

$$\{\widetilde{\Gamma}_{d-1}, \widetilde{\Gamma}_{-(d-1)}\} = 0. \tag{3.9}$$

The operators associated with the roots in (3.3a) and (3.3b) are constructed from the L_{AB} in (3.6) according to[11]

$$E(\eta\hat{e}_I + \eta'\hat{e}_J) = \tfrac{1}{2}(L_{2I-1,2J-1} + i\eta L_{2I,2J-1} + i\eta' L_{2I-1,2J} - \eta\eta' L_{2I,2J})$$
$$= -(i/4P^2)\overline{D}\not{P}\widetilde{\Gamma}_I{}^{(\eta)}\widetilde{\Gamma}_J{}^{(\eta')}D, \tag{3.10}$$

$$E(\eta\hat{e}_I) = (1/\sqrt{2})(L_{2I-1,d-1} + i\eta L_{2I,d-1})$$
$$= -(i/4P^2)\overline{D}\not{P}\widetilde{\Gamma}_I{}^{(\eta)}\widetilde{\Gamma}_{d-1}D,$$

where

$$\tilde{\Gamma}_I^{(\eta)} = (1/\sqrt{2})(\Gamma_{2I-1} + i\eta\Gamma_{2I}) \quad (\eta = \pm 1). \quad (3.11)$$

Indeed, the E operators in (3.10) satisfy

$$\begin{aligned} &[H_I, E(\eta\hat{e}_I + \eta'\hat{e}_J)] = \eta E(\eta\hat{e}_I + \eta'\hat{e}_J), \\ &[H_J, E(\eta\hat{e}_I + \eta'\hat{e}_J)] = \eta' E(\eta\hat{e}_I + \eta'\hat{e}_J), \\ &[H_I, E(\eta\hat{e}_I)] = \eta E(\eta\hat{e}_I), \quad \eta,\eta' = +,-, \end{aligned} \quad (3.12)$$

from which we see that the E operators (3.10) raise or lower the components of the weight vectors by one unit.

Since we are interested in the decomposition of real superfields we will need the properties of these operators under complex conjugation. From

$$\Gamma_A{}^\dagger = \Gamma_0 \Gamma_A \Gamma_0, \quad (3.13)$$

we can easily derive

$$\begin{aligned} &\tilde{\Gamma}_I^{(\eta)\dagger} = -\tilde{\Gamma}_I^{(-\eta)}, \\ &E(\eta\hat{e}_I + \eta'\hat{e}_J)^* = -E(-\eta\hat{e}_I - \eta'\hat{e}_J), \\ &E(\eta\hat{e}_I)^* = -E(-\eta\hat{e}_I), \end{aligned} \quad (3.14)$$

after making use of the facts that $\bar{D}\Gamma_A D$ is proportional to P_A and that we are working in the collinear frame (3.5). By using recurrently the E operators we can generate all the eigenstates of the Cartan subalgebra from any one of them. Then one can generate the irreducible superfield by applying all possible powers of Q and forming a linear combination whose coefficients are ordinary fields (functions of the space-time variables). Therefore, from the knowledge of just *one* eigenstate of the Cartan subalgebra, one can construct the complete irreducible superfield associated with it. But, besides the E operators that move us around inside a given representation, there are other operators that allow us to jump from one irreducible representation to another. For instance, an operator that is an odd power of the covariant derivatives D will take us from a representation with bosonic superweight to one with a fermionic superweight and vice versa; and once we have a particular state of a different representation, we can generate all the remaining ones by adequate use of the E operators. All that is required is knowledge of the weight system of the representation, so that one does not miss any state or count a state more than once. And it is well known how to generate the weight system of a representation from its highest weight.[9]

Furthermore, when the number of E operators does not saturate the number of operators quadratic in D, one can construct operators that take us from a representation with bosonic (fermionic) superweight to a different one with also a bosonic (fermionic) superweight. This will always be the case when there are more than one representation of a certain type (bosonic or fermionic) included in the general scalar superfield. Again, once we have a particular state of the new representation, the rest can be generated by use of the E operators. This process can continue indefinitely and the end result is that we are able to generate *all* the irreducible superfields included in the general scalar superfield from *one* state of any of the corresponding irreducible representations.

In practice, we will start by constructing the representation with greatest superweight and proceed in a descending fashion. Also, as illustrated in the massless case, we can treat several states on equal footing rather than generate everything from one particular state.

In the next section we apply this Cartan subalgebra method to the four-dimensional case in order to illustrate how it works, while we delay until Sec. V the solution of the full-fledge ten-dimensional problem.

IV. DECOMPOSITION OF THE FOUR-DIMENSIONAL SCALAR SUPERFIELD

The results we are going to obtain here have been derived in the literature by a variety of methods. It is well known that the general scalar (real) superfield in four dimensions contains three superspins: $Y = \frac{1}{2}$ and twice $Y = 0$.[4] That is to say, the irreducible representations of SO(3) whose highest weights are $[\frac{1}{2}]$, $[0]$, and $[0]$, respectively. In keeping with the program described in the previous section, we start with the representation $[\frac{1}{2}]$.

The Cartan subalgebra here contains only one generator which, according to (3.1), is

$$H = -(i/4P^2)\bar{D}\not{P}\Gamma_{1,2}D. \quad (4.1)$$

The eigenstates of this operator corresponding to the two weights of the representation $[\frac{1}{2}]$ can be found by considering the identity

$$\begin{aligned} He^{a\bar{\theta}\not{P}\Gamma_{1,2}\theta} &= (i/2)a(\mathrm{Tr}\, I)e^{a\bar{\theta}\not{P}\Gamma_{1,2}\theta} \\ &\quad - (i/4)(4a^2 + \tfrac{1}{4})\bar{\theta}\not{P}\Gamma_{1,2}\theta e^{a\bar{\theta}\not{P}\Gamma_{1,2}\theta}. \end{aligned} \quad (4.2)$$

If we choose $a = \mp(i/4)$, (4.2) becomes

$$H\chi_\pm = \pm\tfrac{1}{2}\chi_\pm, \quad (4.3)$$

with

$$\chi_\pm = \exp[\mp(i/4)\bar{\theta}\not{P}\Gamma_{1,2}\theta]. \quad (4.4)$$

These eigenstates corresponding to the two weights of the representation $[\frac{1}{2}]$ also satisfy

$$\begin{aligned} D^\alpha\chi_\pm &= -i(\Pi_\pm\not{P}\theta)^\alpha\chi_\pm, \\ Q^\alpha\chi_\pm &= i(\Pi_\mp\not{P}\theta)^\alpha\chi_\pm. \end{aligned} \quad (4.5)$$

In (4.5), $\Pi_\pm$ are the projection operators

$$\Pi_\pm = \tfrac{1}{2}(1 \pm i\Gamma_{1,2}). \quad (4.6)$$

Consequently the projected operators $D_\pm = \Pi_\pm D$ and $Q_\pm = \Pi_\pm Q$ satisfy

$$\begin{aligned} D_\pm{}^\alpha\chi_\pm &= -i(\not{P}\theta_\pm)^\alpha\chi_\pm, \\ Q_\mp{}^\alpha\chi_\pm &= i(\not{P}\theta_\mp)^\alpha\chi_\pm, \quad (\theta_\pm = \Pi_\pm\theta), \end{aligned} \quad (4.7)$$

and

$$D_\mp{}^\alpha\chi_\pm = 0, \quad Q_\pm{}^\alpha\chi_\pm = 0. \quad (4.8)$$

The important commutation relations for these projected objects are

$$\{Q_\pm{}^\alpha, \theta_\pm{}^\beta\} = 0 \text{ and } \{D_\pm{}^\alpha, \theta_\mp{}^\beta\} = i(\Pi_\pm C^{-1})^{\alpha\beta}. \quad (4.9)$$

On the other hand, from (3.10) we get the raising and lowering operators,

$$E_\pm = -(i/4P^2)\bar{D}\not{P}\tilde{\Gamma}_\pm\tilde{\Gamma}_3 D = \pm(i/4P_0)\bar{D}\Gamma_5\tilde{\Gamma}_\pm D$$
$$= \pm(i/4P_0)\bar{D}_\mp\Gamma_5\tilde{\Gamma}_\pm D_\mp \quad (4.10)$$
$$(\Gamma_5 = i\Gamma_0\Gamma_1\Gamma_2\Gamma_3, \quad \Gamma_5{}^2 = 1),$$

where the $\tilde{\Gamma}$ matrices as defined in (3.7), (3.11) are

$$\tilde{\Gamma}_\pm = (1/\sqrt{2})(\Gamma_1 \pm i\Gamma_2), \quad \tilde{\Gamma}_3 = \Gamma_3 - z\Gamma_0. \quad (4.11)$$

They satisfy

$$\tilde{\Gamma}_\pm\Pi_\mp = \Pi_\pm\tilde{\Gamma}_\pm = \tilde{\Gamma}_\pm. \quad (4.12)$$

These E operators when applied to $\chi_\pm$ give

$$E_\mp\chi_\pm = \mp(i/4)(P^2/P_0)\bar{\theta}\Gamma_5\tilde{\Gamma}_\mp\theta\chi_\pm. \quad (4.13)$$

Now everything is ready to construct the superfield corresponding to the representation $[\frac{1}{2}]$. Since $\chi_\pm$ satisfy (4.8), this superfield can be written

$$\phi^+_{[1/2]} = \sum_{m=0}^{2} Q_-{}^{\alpha_1}\cdots Q_-{}^{\alpha_m}\big[\chi_+F^{(1)}_{\alpha_1\cdots\alpha_m}(x) + E_-\chi_+G^{(1)}_{\alpha_1\cdots\alpha_m}(x)\big], \quad (4.14\text{a})$$

or alternatively

$$\phi^-_{[1/2]} = \sum_{m=0}^{2} Q_+{}^{\alpha_1}\cdots Q_+{}^{\alpha_m}\big[\chi_-F^{(2)}_{\alpha_1\cdots\alpha_m}(x) + E_+\chi_-G^{(2)}_{\alpha_1\cdots\alpha_m}(x)\big]; \quad (4.14\text{b})$$

however, in this case we can avoid the use of the E operators by using the pair of states χ_+ and χ_- instead

$$\phi_{[1/2]} = \sum_{m=0}^{2}\big[Q_-{}^{\alpha_1}\cdots Q_-{}^{\alpha_m}\chi_+F^+_{\alpha_1\cdots\alpha_m}(x) + Q_+{}^{\alpha_1}\cdots Q_+{}^{\alpha_m}\chi_-F^-_{\alpha_1\cdots\alpha_m}(x)\big]. \quad (4.14\text{c})$$

All these expressions are equivalent up to field redefinitions and reflect the freedom that we have in choosing a basis for the two-dimensional representation $[\frac{1}{2}]$ of SO(3). They are in general complex, so that we must take the real part to obtain the irreducible component of the real scalar superfield. The θ expansion of (4.14a) and (4.14b) can be readily obtained by using (4.13), (4.7), and (4.4), but we will study the expansion (4.14c) instead.

Under complex conjugation we have

$$\chi_\pm{}^* = \chi_\mp, \quad Q_\alpha{}^* = (C\Gamma_0)_{\alpha\beta}Q^\beta, \quad \Pi_\pm{}^* = C\Pi_\mp C^{-1}, \quad (4.15)$$

which implies

$$(Q_-{}^{\alpha_1}\cdots Q_-{}^{\alpha_m})^* = Q^*_{-\alpha_m}\cdots Q^*_{-\alpha_1}$$
$$= (C\Gamma_0)_{\alpha_1\beta_1}\cdots(C\Gamma_0)_{\alpha_m\beta_m}Q_+{}^{\beta_m}\cdots Q_+{}^{\beta_1}, \quad (4.16)$$

and the same result if we interchange $+ \leftrightarrow -$. Then the reality condition requires

$$F^\pm_{\alpha_1\cdots\alpha_m} = \epsilon(m)(C\Gamma_0)_{\beta_1\alpha_1}\cdots(C\Gamma_0)_{\beta_m\alpha_m}F^{\mp *\beta_1\cdots\beta_m},$$
$$F^{*\beta_1\cdots\beta_m} = (F_{\beta_1\cdots\beta_m})^* \quad (4.17)$$

with

$$\epsilon(m) = (-)^{(m/2)(m-1)}.$$

By using also (3.14), one can find the relations between the F and G fields in (4.14a) and (4.14b) imposed by the reality of the respective superfields though we will not write them.

Before we look at the standard expansion in powers of θ, we are going to obtain the expansion in Grassmann–Hermite functions of $\phi_{[1/2]}$ in (4.14c). We have the identities

$$Q_\pm{}^\alpha = i\Pi_\pm{}^\alpha{}_\beta\chi_\pm\frac{\partial}{\partial\bar{\theta}_\beta}\chi_\mp,$$
$$D_\pm{}^\alpha = i\Pi_\pm{}^\alpha{}_\beta\chi_\mp\frac{\partial}{\partial\bar{\theta}_\beta}\chi_\pm, \quad (4.18)$$

which allow us to write

$$\phi_{[1/2]} = \sum_{m=0}^{2}\big[\chi_+H_+{}^{\alpha_1\cdots\alpha_m}(\theta)\psi^+_{\alpha_1\cdots\alpha_m}(x) + \chi_-H_-{}^{\alpha_1\cdots\alpha_m}(\theta)\psi^-_{\alpha_1\cdots\alpha_m}(x)\big], \quad (4.19)$$

where the Grassmann–Hermite polynomials are given by

$$H_\pm{}^{\alpha_1\cdots\alpha_m}(\theta) = [\chi_\mp]^2\frac{\partial}{\partial\bar{\theta}_{\alpha_1}}\cdots\frac{\partial}{\partial\bar{\theta}_{\alpha_m}}[\chi_\pm]^2 \quad (4.20)$$

and the multispinors $\psi^\pm_{\alpha_1\cdots\alpha_m}(x)$ satisfy

$$\Pi_\pm{}^{\alpha_r}{}_\beta\psi^\pm_{\alpha_1\cdots\alpha_r\cdots\alpha_m}(x) = 0. \quad (4.21)$$

If we want to obtain the usual expansion of the superfield as a power series in θ, it is convenient to use the scaled supercharge,

$$q = -(i\not{P}/P^2)Q \quad (4.22)$$

which satisfies [see (4.7)–(4.9)]

$$q_\mp{}^\alpha\chi_\pm = \theta_\mp{}^\alpha\chi_\pm, \quad q_\pm{}^\alpha\chi_\pm = 0, \quad \{q_\pm{}^\alpha, \theta_\pm{}^\beta\} = 0. \quad (4.23)$$

Using (4.23) in (4.14c), we obtain, after redefining the fields $F^+_{\alpha_1\cdots\alpha_m}$ and $F^-_{\alpha_1\cdots\alpha_m}$,

$$\phi_{[1/2]} = \sum_{m=0}^{2}\big[q_-{}^{\alpha_1}\cdots q_-{}^{\alpha_m}\chi_+F^+_{\alpha_1\cdots\alpha_m} + q_+{}^{\alpha_1}\cdots q_+{}^{\alpha_m}\chi_-F^-_{\alpha_1\cdots\alpha_m}\big]$$
$$= \sum_{m=0}^{2}\big[\theta_-{}^{\alpha_1}\cdots\theta_-{}^{\alpha_m}\chi_+F^+_{\alpha_1\cdots\alpha_m} + \theta_+{}^{\alpha_1}\cdots\theta_+{}^{\alpha_m}\chi_-F^-_{\alpha_1\cdots\alpha_m}\big], \quad (4.24)$$

and if we now expand (4.4),

$$\chi_\pm = 1 \mp (i/4)\bar{\theta}\not{P}\Gamma_{1,2}\theta - \tfrac{1}{32}\bar{\theta}\not{P}\Gamma_{1,2}\theta\bar{\theta}\not{P}\Gamma_{1,2}\theta, \quad (4.25)$$

and make use of the identities of the Appendix A, we arrive at the simple form

$$\Phi_{[1/2]} = B_+ + \bar{\theta}\chi + \bar{\theta}\Gamma_5\Gamma^A\theta V_A + \tfrac{1}{4}\bar{\theta}\theta\bar{\theta}\not{P}\chi - \tfrac{1}{32}(\bar{\theta}\theta)^2P^2B_+, \quad (4.26)$$

with

$$B_\pm = F^+ \pm F^-,$$
$$\chi^\alpha = (\Pi_+ C^{-1})^{\alpha\beta} F^+{}_\beta + (\Pi_- C^{-1})^{\alpha\beta} F^-_\beta,$$
$$V_A = \tfrac{1}{4}(P_0\eta_{A,3} - P_3\eta_{A,0})B_- - (1/4\sqrt{2})(\eta_{A,1} + i\eta_{A,2})(\Gamma_5\tilde{\Gamma}_- C^{-1})^{\alpha\beta} F^+_{\alpha\beta} - (1/4\sqrt{2})(\eta_{A,1} - i\eta_{A,2})(\Gamma_5\tilde{\Gamma}_+ C^{-1})^{\alpha\beta} F^-{}_{\alpha\beta}, \tag{4.27}$$

where η_{AB} is the Minkowski metric. Recalling that we are working in the collinear frame (3.5), it is easy to see that V_A is a transverse vector field,

$$P^A V_A = 0. \tag{4.28}$$

Equation (4.26) with the transversality condition (4.28) shows that $\phi_{[1/2]}$ is nothing but the "transverse vector superfield" characterized by the constraint[12]

$$\bar{D}^{(+)} D^{(-)} \phi = 0, \quad D^{(+)} = \tfrac{1}{2}(I \pm \Gamma_5) D, \tag{4.29}$$

and is indeed the correct result. The transverse vector superfield carries a linear multiplet and the supersymmetry transformations are

$$\delta B_+ = \bar{\epsilon}\chi, \quad \delta\chi = -\tfrac{1}{2}\not{P}\epsilon B_+ + 2\Gamma_5\Gamma^A \epsilon V_A,$$
$$\delta V_A = -\tfrac{1}{8}\bar{\epsilon}\Gamma_5[\not{P},\Gamma_A]\chi, \tag{4.30}$$

which close the algebra off shell as well as on shell.

The two remaining irreducible pieces are degenerate since they correspond to the same representation of SO(3), the trivial one [0]. From (4.8) and the fact that

$$\{D_+, D_+\} = \{D_-, D_-\} = 0, \tag{4.31}$$

we can modify the expressions (4.14a) and (4.14b),

$$\phi^+_{[0]} = \sum_{m=0}^{2} Q_-{}^{\alpha_1} \cdots Q_-{}^{\alpha_m} D_+{}^\beta \chi_+ F^+_{\alpha_1 \cdots \alpha_m;\beta} \tag{4.32a}$$

and

$$\phi^-_{[0]} = \sum_{m=0}^{2} Q_+{}^{\alpha_1} \cdots Q_+{}^{\alpha_m} D_-{}^\beta \chi_- F^-_{\alpha_1 \cdots \alpha_m;\beta} \tag{4.32b}$$

The expansions (4.32a) and (4.32b) correspond to $H = 0$ as is clear from (4.3) and the commutation relations,

$$[H, D_\pm] = \mp \tfrac{1}{2} D_\pm . \tag{4.33}$$

If $\phi^+_{[0]}$ and $\phi^-_{[0]}$ are the superspin 0 pieces of a real scalar superfield, they must be complex conjugate to each other,

$$\phi^+_{[0]} = (\phi^-_{[0]})^*$$
$$\Leftrightarrow F^\pm_{\alpha_1 \cdots \alpha_m;\beta} = \epsilon(m+1)(C\Gamma_0)_{\gamma_1\alpha_1} \cdots (C\Gamma_0)_{\gamma_m\alpha_m} \times (C\Gamma_0)_{\delta\beta} F^{\mp * \gamma_1 \cdots \gamma_m;\delta}. \tag{4.34}$$

We can write, after appropriate field redefinitions, the superfield $\phi^+_{[0]}$ as an expansion in Grassmann–Hermite polynomials,

$$\phi^+_{[0]} = \sum_{m=0}^{2} \chi_+ H_+{}^{\alpha_1 \cdots \alpha_m \beta}(\theta) \psi_{\alpha_1 \cdots \alpha_m;\beta}(x) \tag{4.35}$$

and the multispinors $\psi_{\alpha_1 \cdots \alpha_m;\beta}(x)$ satisfy

$$\Pi_+{}^{\alpha_r}{}_\gamma \psi_{\alpha_1 \cdots \alpha_r \cdots \alpha_m;\beta}(x) = 0,$$
$$\Pi_-{}^\beta{}_\gamma \psi_{\alpha_1 \cdots \alpha_m;\beta}(x) = 0. \tag{4.36}$$

Clearly $\psi_{\alpha;\beta}(x)$ carries four bosonic degrees of freedom while $\psi_\beta(x)$ and $\psi_{\alpha_1\alpha_2;\beta}(x)$ carry two fermionic degrees of freedom each. These correspond to the degrees of freedom of the chiral superfield. The superfield $\phi^+_{[0]}$ is neither chiral nor antichiral though. This is of course no problem, because the way to separate the two superspin 0 states is not unique at all. The splitting into a chiral plus an antichiral superfield is just a particular example. The fact that we have two irreducible pieces corresponding to the same representation is what allows them not to be real and complex conjugate to each other. This does not occur in the ten-dimensional case where all the irreducible pieces included in the real scalar superfield must necessarily be real.

The remaining irreducible piece $\phi_{[1/2]}$ is unique, however (up to field redefinitions of course), and agrees with the one obtained previously in the literature.[12]

Another point worth a comment is the fact that $\phi_{[1/2]}$ ($\phi_{[0]}$) is bosonic (fermionic) at the D level, i.e., it contains only even (odd) powers of D in its expansion, in spite of having a fermionic (bosonic) superweight: $[\frac{1}{2}]$ ([0]). This is peculiar to four dimensions and has its origin in Eq. (4.2), which is quite general.[6,8] The vanishing of the second term implies $a = \mp (i/4)$ and the coefficient of the first term becomes $\pm \frac{1}{8}\,\mathrm{Tr}\, I$, which is a half-integer only if the dimension of the Dirac algebra ($\mathrm{Tr}\, I$) is 4. Since the dimension of the Dirac algebra grows exponentially, $\pm \frac{1}{8}\,\mathrm{Tr}\, I$ will be an integer in more than five space-time dimensions.

V. DECOMPOSITION OF THE CHIRAL SCALAR SUPERFIELD IN TEN DIMENSIONS

Now we turn to the problem of finding expressions for the irreducible components of the chiral scalar superfield in ten dimensions $\phi(x,\theta^{(-)})$. In this section we will always deal with the Weyl projections $D^{(+)}$, $Q^{(+)}$, and $\theta^{(-)}$; we will drop these labels and write simply D, Q, θ in order to simplify the notation, but they are understood to be always Weyl projected on top of any other projections that are going to appear. Other clarifications about the notation of this section are given in Appendix B. The SO(9) generators are given by Eqs. (3.1), (3.6), (3.10), and (2.9) with $n = 4$. In particular, the generators of the Cartan subalgebra are

$$H_I = -(i/4P^2)\bar{D}\not{P}\Gamma_{2I-1,2I} D, \quad I = 1,2,3,4. \tag{5.1}$$

Just like in the previous section, some eigenfunctions of H_I are easy to find in the form of Grassmann Gaussians,

$$\chi_J{}^{(\pm)} = \exp[\mp (i/4)\bar{\theta}\not{P}\Gamma_{2J-1,2J}\theta], \quad J = 1,2,3,4, \tag{5.2}$$

which satisfy

$$H_I \chi_J{}^{(\pm)} = \pm 2\delta_{I,J}\chi_J^{(\pm)}. \tag{5.3}$$

Obviously, the $\chi_J^{(\pm)}$ in (5.2) are the eigenstates corresponding to the weights $(\pm 2,0,0,0)$, $(0,\pm 2,0,0)$, $(0,0,\pm 2,0)$, and $(0,0,0,\pm 2)$ of the representation [2]. In particular, the one corresponding to the highest weight $(2,0,0,0)$ is

$$\chi_1{}^{(+)} = \exp[-(i/4)\bar{\theta}\not{P}\Gamma_{1,2}\theta]. \tag{5.4}$$

Now we can imitate the steps of the previous section. First,

$$D^\alpha \chi_1^{(\pm)} = -i(\Pi_1{}^\pm \not{P}\theta)^\alpha \chi_1^{(\pm)},$$
$$Q^\alpha \chi_1^{(\pm)} = i(\Pi_1{}^\mp \not{P}\theta)^\alpha \chi_1^{(\pm)}, \tag{5.5}$$

with

$$\Pi_1{}^\pm = \tfrac{1}{2}(1 \pm i\Gamma_{1,2}).$$

Second, the projected operators $D_1{}^\pm = \Pi_1{}^\pm D$, $Q_1{}^\pm = \Pi_1{}^\pm Q$ satisfy

$$D_1{}^{\pm\alpha}\chi_1^{(\pm)} = -i(\not{P}\theta_1{}^\pm)^\alpha \chi_1^{(\pm)},$$
$$Q_1{}^{\mp\alpha}\chi_1^{(\pm)} = i(\not{P}\theta_1{}^\mp)^\alpha \chi_1^{(\pm)}, \tag{5.6}$$
$$D_1{}^\mp \chi_1^{(\pm)} = 0,\quad Q_1{}^\pm \chi_1^{(\pm)} = 0,\quad \theta_1{}^\pm = \Pi_1{}^\pm\theta,$$

as well as the anticommutation relations

$$\{D_1{}^\pm, D_1{}^\pm\} = 0,\quad \{Q_1{}^\pm, Q_1{}^\pm\} = 0. \tag{5.7}$$

The superweights [2] and [111] are bosonic. This means that the representations [2] and [111] are bosonic at the D level and therefore contain only even powers of the covariant derivatives D^α (Ref. 8). On the other hand, the representation described by the fermionic superweight $[\frac{3}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}]$ will be fermionic at the D level, meaning that it contains only odd powers of the covariant derivatives D^α in its expansion. Thus the expression suggested by (5.6) and (5.7),

$$\phi^+_{[2]\oplus[111]} = \sum_{m=0}^{8}\sum_{p=0}^{4} Q_1{}^{-\alpha_1}\cdots Q_1{}^{-\alpha_m}$$
$$\times D_1{}^{+\beta_1}\cdots D_1{}^{+\beta_{2p}}\chi_1^{(+)} A^+_{\alpha_1\cdots\alpha_m;\beta_1\cdots\beta_{2p}}(x) \tag{5.8a}$$

describes the reducible representation $[2]\oplus[111]$. The same is true for

$$\phi^-_{[2]\oplus[111]} = \sum_{m=0}^{8}\sum_{p=0}^{4} Q_1{}^{+\alpha_1}\cdots Q_1{}^{+\alpha_m}$$
$$\times D_1{}^{-\beta_1}\cdots D_1{}^{-\beta_{2p}}\chi_1^{(-)} A^-_{\alpha_1\cdots\alpha_m;\beta_1\cdots\beta_{2p}}(x). \tag{5.8b}$$

Both (5.8a) and (5.8b) describe the same reducible representation $[2]\oplus[111]$ and one must consider one or the other but not both, since to do so would be redundant. The origin of this redundancy lies again, of course, in the freedom that we have in selecting the states of the representation. Moreover, the expressions (5.8a) and (5.8b) are complex and both the real and the imaginary parts describe the same representation $[2]\oplus[111]$. Given that we are interested in the decomposition of the real scalar superfield, we must take the real part of either (5.8a) or (5.8b) in order to obtain a superfield that describes the $[2]\oplus[111]$ part included in it.

In turn,

$$\phi^+_{[(3/2)(1/2)(1/2)(1/2)]}$$
$$= \sum_{m=0}^{8}\sum_{p=1}^{4} Q_1{}^{-\alpha_1}\cdots Q_1{}^{-\alpha_m}$$
$$\times D_1{}^{+\beta_1}\cdots D_1{}^{+\beta_{2p-1}}\chi_1^{(+)} B^+_{\alpha_1\cdots\alpha_m;\beta_1\cdots\beta_{2p-1}}(x) \tag{5.9a}$$

or

$$\phi^-_{[(3/2)(1/2)(1/2)(1/2)]}$$
$$= \sum_{m=0}^{8}\sum_{p=1}^{4} Q_1{}^{+\alpha_1}\cdots Q_1{}^{+\alpha_m}$$
$$\times D_1{}^{-\beta_1}\cdots D_1{}^{-\beta_{2p-1}}\chi_1^{(-)} B^-_{\alpha_1\cdots\alpha_m;\beta_1\cdots\beta_{2p-1}}(x) \tag{5.9b}$$

describe the irreducible representation $[\frac{3}{2}\frac{1}{2}\frac{1}{2}\frac{1}{2}]$, again the real part of either one being the piece of the real scalar superfield corresponding to this representation.

The special difficulty here that was not present in the massless case is the presence of two bosonic representations in (5.8a) and (5.8b) that must be separated in order to isolate the irreducible parts.

But, now, we are in a position where we can solve this problem since in (5.3) we already have eight of the 44 states of the representation [2] and we can obtain the remaining 36 by appropriate use of the E operators in (3.10). First, in order to simplify our expressions, let us define the modified E operators:

(a) for $I<J$,

$$\widetilde{E}_{\eta I,\xi J} = E(\eta\hat{e}_I + \xi\hat{e}_J),\quad I<J; \tag{5.10a}$$

(b) for $I>J$,

$$\widetilde{E}_{\eta I,\xi J} = -\widetilde{E}_{\xi J,\eta I},\quad I>J; \tag{5.10b}$$

(c) $\widetilde{E}_{\eta I,9} = E(\eta\hat{e}_I) = -\widetilde{E}_{9,\eta I};$ (5.10c)

(d) for $I=J$,

$$\widetilde{E}_{\eta I,\eta I} = (\widetilde{E}_{\eta I,9})^2,$$
$$\widetilde{E}_{\eta I,-\eta I} = -i\eta H_I, \tag{5.10d}$$
$$\widetilde{E}_{9,9} = 0.$$

In Eqs. (5.10a)–(5.10d), $\eta,\xi = +$ or $-$, $I,J = 1,\ldots,4$, the operators H_I are the ones in (5.1), while $E(\eta\hat{e}_I + \xi\hat{e}_J)$ and $E(\eta\hat{e}_I)$ are the ones in (3.10).

With these elements we can give the 44 states of the representation [2] of SO(9) as follows: (i) the eight states corresponding to the weights $(\pm 2,0,0,0),\ldots,(0,0,0,\pm 2)$ as given by (5.2),

$$\chi_J{}^{(\eta)} = -(1/2i)\widetilde{E}_{\eta J,-\eta J}\chi_J^{(\eta)}; \tag{5.11a}$$

(ii) the 24 states corresponding to the weights $(\pm 1,\pm 1,0,0),\ldots,(0,0,\pm 1,\pm 1)$,

$$\widetilde{E}_{\eta I,-\xi J}\chi_J^{(\xi)} + \widetilde{E}_{\xi J,-\eta I}\chi_I^{(\eta)},\quad I<J; \tag{5.11b}$$

(iii) the eight states associated with the weights $(\pm 1,0,0,0),\ldots,(0,0,0,\pm 1)$,

$$\widetilde{E}_{9,\eta I}\chi_I^{(-\eta)}; \tag{5.11c}$$

and (iv) finally, the four states associated with the fourfold degenerate weight (0,0,0,0) are

$$\widetilde{E}_{-J,-J}\chi_J^{(+)} + \widetilde{E}_{J,J}\chi_J^{(-)}. \tag{5.11d}$$

(This choice, of course, is not unique just as in the four-dimensional case of the previous section. Any other suitable choice of states will give expressions that can be transformed into ours by adequate field redefinitions.)

The irreducible superfield $\phi_{[2]}$ is then the sum of all 44 states in (5.11a)–(5.11d) and all the supersymmetric partners obtained by applying all possible powers of Q to those, with ordinary fields as coefficients

$$\phi_{[2]}=\sum_{m=0}^{8}\sum_{J=1}^{4}\sum_{\xi=+,-}Q_J^{(-\xi)\alpha_1}\cdots Q_J^{(-\xi)\alpha_m}$$
$$\times\Big\{\sum_{I=1}^{4}{}'\sum_{\eta=+,-}\widetilde{E}_{\eta I,-\xi J}\chi_J^{(\xi)}F_{-\eta I,-\xi J;\alpha_1\cdots\alpha_m}(x)$$
$$+\widetilde{E}_{\xi J,-\xi J}\chi_J^{(\xi)}F_{-\xi J,-\xi J;\alpha_1\cdots\alpha_m}(x)$$
$$+\widetilde{E}_{9,-\xi J}\chi_J^{(\xi)}F_{9,-\xi J;\alpha_1\cdots\alpha_m}(x)$$
$$+\widetilde{E}_{-\xi J,-\xi J}\chi_J^{(\xi)}F_{\xi J,-\xi J;\alpha_1\cdots\alpha_m}(x)\Big\},\qquad(5.12)$$

where

$$F_{\eta I,\xi J;\alpha_1\cdots\alpha_m}(x)=F_{\xi J,\eta I;\alpha_1\cdots\alpha_m}(x),$$
$$F_{\xi I,-\xi J;\alpha_1\cdots\alpha_m}(x)=F_{-\xi J,\xi J;\alpha_1\cdots\alpha_m}(x),$$

and Σ_I' means that the value $I=J$ is excluded from the sum. The projected charges $Q_J^{(\eta)}$ are

$$Q_J^{(\eta)}=\Pi_J^{(\eta)}Q,\qquad(5.13)$$

with

$$\Pi_J^{(\eta)}=\tfrac{1}{2}(I+i\eta\Gamma_{2J-1,2J}).\qquad(5.14)$$

The reason only $Q_J^{(-\xi)}$ appears in (5.12) is because we have

$$Q_J^{(\eta)}\chi_J^{(\eta)}=0\qquad(5.15)$$

in analogy with (5.6). For the covariant derivative, we have instead

$$D_J^{(-\eta)}\chi_J^{(\eta)}=0,\quad D_J^{(\eta)}=\Pi_J^{(\eta)}D.\qquad(5.16)$$

The projection operators $\Pi_J^{(\eta)}$ are related to the $\widetilde{\Gamma}_J^{(\eta)}$ matrices in (3.11) by

$$\widetilde{\Gamma}_J^{(\eta)}\widetilde{\Gamma}_J^{(-\eta)}=-2\Pi_J^{(\eta)}\qquad(5.17)$$

and furthermore, we have the relations

$$\Pi_J^{(\eta)}\widetilde{\Gamma}_J^{(\eta)}=\widetilde{\Gamma}_J^{(\eta)}\Pi_J^{(-\eta)}=\widetilde{\Gamma}_J^{(\eta)},$$
$$\widetilde{\Gamma}_J^{(\eta)2}=0.\qquad(5.18)$$

We can improve the appearance of (5.12) by introducing a new index k with values 1,...,9 whose 1 through 8 values are related to ηJ by[8]

$$k=\begin{cases}J, & \text{for } +J,\quad J=1,\ldots,4,\\ 8-J+1, & \text{for } -J,\quad J=1,\ldots,4,\end{cases}\qquad(5.19)$$

which gives Table I in Appendix B.

In this way χ_k is given by

$$\chi_k=\begin{cases}\chi_J^{(+)}, & \text{for } k=J=1,\ldots,4,\\ \chi_J^{(-)}, & \text{for } k=8-J+1=5,\ldots,8,\end{cases}\qquad(5.20)$$

and similarly for Q_k and Π_k.

Actually, instead of χ_k, it is better to use

$$\widehat{\chi}_k=\chi_{8-k+1},\quad k=1,\ldots,8,\qquad(5.21a)$$

supplemented by

$$\widehat{\chi}_9=0.\qquad(5.21b)$$

Lastly, it is convenient to redefine the E operators,

$$\widehat{E}_{k_1,k_2}=\widetilde{E}_{8-k_1+1,k_2},\quad k_1,k_2=1,\ldots,8,$$
$$\widehat{E}_{9,k}=\widetilde{E}_{9,k},\quad k=1,\ldots,8.\qquad(5.22)$$

With all these provisions, Eq. (5.12) simplifies to

$$\phi_{[2]}=\sum_{m=0}^{8}\sum_{k_1,k_2=1}^{9}Q_{k_2}^{\alpha_1}\cdots Q_{k_2}^{\alpha_m}\widehat{E}_{k_1k_2}\widehat{\chi}_{k_2}F_{k_1k_2;\alpha_1\cdots\alpha_m}(x),\qquad(5.23)$$

where the fields $F_{k_1k_2;\alpha_1\cdots\alpha_m}(x)$ are symmetric in the k indices and totally antisymmetric in the spinorial indices $\alpha_1,\ldots,\alpha_m$.

Notice that the undefined objects $F_{9,9;\alpha_1\cdots\alpha_m}(x)$ do not really appear in (5.23) because of (5.10d) or (5.21b). Thus the number of field components is

$$44\times\sum_{m=0}^{8}\binom{8}{m}=44\times 2^8,$$

as it should, since the dimension of the SO(9) representation [2] is 44. Furthermore, the fields $F_{k_1k_2;\alpha_1,\ldots,\alpha_m}(x)$ show in a natural way why the irreducible superfield in (5.23) is described by the representation [2] of SO(9) which corresponds to a traceless symmetric second rank tensor. Equation (5.23), however, is nothing but a compact way of writing (5.12) through the relabeling of (5.19) and (5.20) and the redefinitions (5.21) and (5.22).

The other bosonic representation can also be generated from the eight eigenstates $\chi_J^{(\pm)}$ of the representation [2] in (5.2). This is possible due to the existence of quadratic operators in the covariant derivatives D that are not operators of the SO(9) little algebra in (3.6). This did not happen in the massless case where the SO(8) little algebra saturated all the possible independent operators quadratic in the covariant derivatives.[8] After we have used one of these operators on any of the states of [2] in (5.2) to obtain a state of [111], we can generate the rest of the representation [111] by repeated application of the E operators of the SO(9) little algebra. It is easy to construct such operators M_{k_1,k_2,k_3}, $k_1,k_2,k_3=1,\ldots,9$ as follows: (a) if k_1,k_2,k_3 are all different,

$$M_{k_1k_2k_3}=\overline{D}\widetilde{\Gamma}_{k_1}\widetilde{\Gamma}_{k_2}\widetilde{\Gamma}_{k_3}D;\qquad(5.24a)$$

(b) if $k_3=9$, $k_1=k_2=k\neq 9$,

$$M_{\eta I,\eta I,9}=\sum_{J=1}^{4}\widetilde{E}_{\eta I,9}M_{\eta I,J,-J},\quad I=1,\ldots,4,\quad \eta=+\text{ or }-;\qquad(5.24b)$$

(c) if none of the above, then

$$M_{k_1k_2k_3}=0.\qquad(5.24c)$$

These operators have the properties

$$[H_I,M_{\eta I,\xi J,\xi K}]=\eta M_{\eta I,\xi J,\xi K},\quad I,J,K\text{ all different},$$
$$[H_I,M_{\eta I,\xi J,9}]=\eta M_{\eta I,\xi J,9},\quad I\neq J,$$
$$[H_I,M_{\eta I,\eta I,9}]=2\eta M_{\eta I,\eta I,9},$$
$$[H_I,M_{\eta I,-\eta I,k}]=0,\qquad(5.25)$$

which are similar to those of the E operators and precisely what we want.

With these operators $M_{k_1k_2k_3}$ we can construct the states of the representation [111] (given in the weight diagram of Fig. 2) which we will choose as follows (again this choice is not unique): (i) the 32 states corresponding to the weights $(\pm 1, \pm 1, \pm 1, 0),\ldots,(0, \pm 1, \pm 1, \pm 1)$,

$$M_{\eta I,\zeta J,\xi K}\chi_I^{(-\eta)} + M_{\xi K,\eta I,\zeta J}\chi_K^{(-\xi)} + M_{\zeta J,\xi K,\eta I}\chi_J^{(-\zeta)},$$
$$I,J,K \text{ all different and } \xi,\eta,\zeta = +,-; \tag{5.26a}$$

(ii) the 24 states corresponding to the weights $(\pm 1, \pm 1, 0, 0),\ldots,(0,0,\pm 1, \pm 1)$,

$$M_{\xi K,9,\eta I}\chi_K^{(-\xi)} - M_{\eta I,9,\xi K}\chi_I^{(-\eta)}, \quad I \neq K, \quad \eta,\xi = +,-; \tag{5.26b}$$

(iii) the $3\times 8 = 24$ states corresponding to the threefold degenerate weights $(\pm 1,0,0,0)^3,\ldots,(0,0,0,\pm 1)^3$,

$$M_{\xi K,\eta I,-\eta I}\chi_K^{(-\xi)}, \quad I \neq K; \tag{5.26c}$$

(iv) the four states corresponding to the fourfold degenerate weight $(0,0,0,0)^4$,

$$M_{\xi K,\xi K,9}\chi_K^{(-\xi)} - M_{-\xi K,-\xi K,9}\chi_K^{(\xi)}. \tag{5.26d}$$

Now, in order to obtain $\phi_{[111]}$ we apply all possible powers of Q to those states and add them up using ordinary fields as coefficients:

$$\phi_{[111]} = \sum_{m=0}^{8}\sum_{I=1}^{4}\sum_{\eta=+,-} Q_I^{(\eta)\alpha_1}\cdots Q_I^{(\eta)\alpha_m}\Big\{\sum_{J,K}{}'' \sum_{\zeta,\xi=+,-} M_{\eta I,\zeta J,\xi K}\chi_I^{(-\eta)}G_{\eta I,\zeta J,\xi K;\alpha_1\cdots\alpha_m}(x)$$
$$+ 2\sum_J{}' \sum_{\zeta=+,-} M_{\eta I,\zeta J,9}\chi_I^{(-\eta)}G_{\eta I,\zeta J,9;\alpha_1\cdots\alpha_m}(x) + \sum_J{}' \sum_{\zeta=+,-} M_{\eta I,\zeta J,-\zeta J}\chi_I^{(-\eta)}G_{\eta I,\zeta J,-\zeta J;\alpha_1\cdots\alpha_m}(x)$$
$$+ 2M_{\eta I,\eta I,9}\chi_I^{(-\eta)}G_{\eta I,-\eta I,9;\alpha_1\cdots\alpha_m}(x)\Big\}, \tag{5.27}$$

where the fields $G_{\eta I,\zeta J,\xi K;\alpha_1\cdots\alpha_m}(x)$ are totally antisymmetric in the indices $\eta I,\zeta J,\xi K$, while $G_{\eta I,\zeta J,9;\alpha_1\cdots\alpha_m}(x)$ and $G_{\eta I,-\eta I,9;\alpha_1\cdots\alpha_m}(x)$ must be antisymmetric in the first two, and

$$\sum_J{}'$$

means that $J = I$ is excluded while

$$\sum_{J,K}{}''$$

means that $J \neq K$ and both different from I.

Expression (5.27) can be greatly simplified if we incorporate all the G fields into a tensor $G_{k_1k_2k_3;\alpha_1\cdots\alpha_m}(x)$, $k_1,k_2,k_3 = 1,\ldots,9$ that is completely antisymmetric in the k indices by using the relabeling in (5.19). Also note that

$$M_{\eta I,-\eta I,k}\chi_I^{(-\eta)} = \bar{D}\tilde{\Gamma}_I^{(\eta)}\tilde{\Gamma}_I^{(-\eta)}\tilde{\Gamma}_k D\chi_I^{(-\eta)} = 0 \tag{5.28}$$

as can be seen from (5.16), (5.17), (5.18), and (3.11) together with the fact $\bar{D}\Gamma_A D = -8P_A$, which in our collinear frame implies

$$\bar{D}\tilde{\Gamma}_k D = 0, \quad k = 1,\ldots,9. \tag{5.29}$$

Thus we can write

$$\phi_{[111]} = \sum_{m=0}^{8}\sum_{k_1,k_2,k_3=1}^{9} Q_{k_1}^{\alpha_1}\cdots Q_{k_1}^{\alpha_m}$$
$$\times \hat{M}_{k_1k_2k_3}\hat{\chi}_{k_1}G_{k_1k_2k_3;\alpha_1\cdots\alpha_m}(x), \tag{5.30}$$

where $\hat{\chi}_k$ has been defined in (5.21a) and (5.21b) and $\hat{M}_{k_1k_2k_3}$ is equal to $M_{k_1k_2k_3}$ except for

$$\hat{M}_{\xi K,9,-\xi K} = -\hat{M}_{\xi K,-\xi K,9} = -M_{\xi K,\xi K,9}. \tag{5.31}$$

Again the fields $G_{k_1k_2k_3;\alpha_1\cdots\alpha_m}(x)$ show naturally why the superfield in (5.30) is described by the representation [111] of SO(9) that corresponds to a totally antisymmetric tensor of third rank. Anyhow, (5.30) is just a concise way of writing (5.27) and contains no extra information. Now let us pause to look at the restrictions on the F and G fields in (5.12) and (5.27) implied by the reality of the superfields $\phi_{[2]}$ and $\phi_{[111]}$, respectively. From

$$\Pi_J^{(\xi)*} = C\Pi_J^{(-\xi)}C^{-1} \tag{5.32}$$

we obtain

$$Q_J^{(\xi)*}{}_\alpha = (C\Gamma_0)_{\alpha\beta}Q_J^{(\xi)\beta}. \tag{5.33}$$

This, together with (3.14) and

$$\tilde{E}^*_{\eta I,-\eta I} = \tilde{E}_{\eta I,-\eta I} = -\tilde{E}_{-\eta I,\eta I},$$
$$\tilde{E}^*_{\eta I,\eta I} = \tilde{E}_{-\eta I,-\eta I}, \tag{5.34}$$

implies the relations

$$\begin{aligned}
&F_{-\eta I,-\xi J;\beta_1\cdots\beta_m}\\
&\quad = -\epsilon(m)(C\Gamma_0)_{\alpha_1\beta_1}\cdots(C\Gamma_0)_{\alpha_m\beta_m}F^{*\ \alpha_1\cdots\alpha_m}_{\eta I,\xi J},\\
&F_{9,-\xi J;\beta_1\cdots\beta_m}\\
&\quad = -\epsilon(m)(C\Gamma_0)_{\alpha_1\beta_1}\cdots(C\Gamma_0)_{\alpha_m\beta_m}F^{*\ \alpha_1\cdots\alpha_m}_{9,\xi J},\\
&F_{-\xi J,-\xi J;\beta_1\cdots\beta_m}\\
&\quad = -\epsilon(m)(C\Gamma_0)_{\alpha_1\beta_1}\cdots(C\Gamma_0)_{\alpha_m\beta_m}F^{*\ \alpha_1\cdots\alpha_m}_{\xi J,\xi J},\\
&F_{\xi J,-\xi J;\beta_1\cdots\beta_m}\\
&\quad = \epsilon(m)(C\Gamma_0)_{\alpha_1\beta_1}\cdots(C\Gamma_0)_{\alpha_m\beta_m}F^{*\ \alpha_1\cdots\alpha_m}_{-\xi J,\xi J},
\end{aligned} \tag{5.35}$$

where, of course, $F^{*\ \alpha_1\cdots\alpha_m}_{k_1k_2} = (F_{k_1k_2;\alpha_1\cdots\alpha_m})^*$. For the M operators in (5.24), we have

$$\begin{aligned}
M^*_{\eta I,\zeta J,\xi K} &= -M_{-\eta I,-\zeta J,-\xi K},\\
M^*_{\eta I,\zeta J,9} &= -M_{-\eta I,-\zeta J,9},\\
M^*_{\eta I,-\eta I,\xi K} &= -M_{-\eta I,\eta I,-\xi K} = M_{\eta I,-\eta I,-\xi K},\\
M^*_{\eta I,\eta I,9} &= -M_{-\eta I,-\eta I,9},
\end{aligned} \tag{5.36}$$

which give the relations for the G fields in (5.27),

$$
\begin{aligned}
&G_{\eta J,\zeta J,\xi K;\beta_1\cdots\beta_m} \\
&\quad = -\epsilon(m)(C\Gamma_0)_{\alpha_1\beta_1}\cdots(C\Gamma_0)_{\alpha_m\beta_m} \\
&\qquad \times G^{*}_{-\eta I,-\zeta J,-\xi K}{}^{\alpha_1\cdots\alpha_m}, \\
&G_{\eta I,\zeta J,9;\beta_1\cdots\beta_m} \\
&\quad = -\epsilon(m)(C\Gamma_0)_{\alpha_1\beta_1}\cdots(C\Gamma_0)_{\alpha_m\beta_m} G^{*}_{-\eta I,-\zeta J,9}{}^{\alpha_1\cdots\alpha_m},
\end{aligned} \tag{5.37}
$$

with

$$
G^{*}_{k_1k_2k_3}{}^{\alpha_1\cdots\alpha_m} = (G_{k_1k_2k_3;\alpha_1\cdots\alpha_m})^{*}.
$$

Next we will derive expressions in terms of Grassmann–Hermite functions for these superfields. The starting point is the pair of identities

$$
\begin{aligned}
Q_J^{(\zeta)} &= i\Pi_J^{(\zeta)}\left(\frac{\partial}{\partial\bar\theta} + \frac{1}{2}\not P\theta\right) = i\Pi_J^{(\zeta)}\chi_J^{(\zeta)}\frac{\partial}{\partial\bar\theta}\chi_J^{(-\zeta)}, \\
D_J^{(\zeta)} &= i\Pi_J^{(\zeta)}\left(\frac{\partial}{\partial\bar\theta} - \frac{1}{2}\not P\theta\right) = i\Pi_J^{(\zeta)}\chi_J^{(-\zeta)}\frac{\partial}{\partial\bar\theta}\chi_J^{(\zeta)},
\end{aligned} \tag{5.38}
$$

which trivially imply (5.15) and (5.16). Then, going back to (5.12), we obtain

$$
\begin{aligned}
\phi_{[2]} &= \sum_{m=0}^{8}\sum_{p=0}^{2}\sum_{J=1}^{4}\sum_{\zeta=+,-}\chi_J^{(\zeta)} \\
&\quad \times H_{\zeta J}^{\alpha_1\cdots\alpha_m\beta_1\cdots\beta_{2p}}(\theta)\psi_{\zeta J;\alpha_1\cdots\alpha_m;\beta_1\cdots\beta_{2p}}(x) \\
&= \sum_{m=0}^{8}\sum_{p=0}^{2}\sum_{k=1}^{8}\chi_k H_k{}^{\alpha_1\cdots\alpha_m\beta_1\cdots\beta_{2p}}(\theta) \\
&\quad \times \psi_{k;\alpha_1\cdots\alpha_m;\beta_1\cdots\beta_{2p}}(x),
\end{aligned} \tag{5.39}
$$

where the Grassmann–Hermite polynomials are given here by

$$
H_{\zeta J}^{\gamma_1\cdots\gamma_m}(\theta) = [\chi_J^{(-\zeta)}]^2\frac{\partial}{\partial\bar\theta_{\gamma_1}}\cdots\frac{\partial}{\partial\bar\theta_{\gamma_m}}[\chi_J^{(\zeta)}]^2, \tag{5.40}
$$

while the ψ fields are related to the F fields in (5.12) by

$$
\begin{aligned}
&\psi_{-\zeta J;\alpha_1\cdots\alpha_m}(x) \\
&\quad = (-2i)i^m\,\Pi_J^{(\zeta)\gamma_1}{}_{\alpha_1}\cdots\Pi_J^{(\zeta)\gamma_m}{}_{\alpha_m}F_{\zeta J,\zeta J;\gamma_1\cdots\gamma_m}(x), \\
&\psi_{-\zeta J;\alpha_1\cdots\alpha_m;\beta_1\beta_2}(x) \\
&\quad = (i^{m+1}/4P^2)\Pi_J^{(\zeta)\gamma_1}{}_{\alpha_1}\cdots\Pi_J^{(\zeta)\gamma_m}{}_{\alpha_m} \\
&\qquad \times\Bigg[\sum_I{}'\sum_{\eta=+,-}(C\not P\tilde\Gamma_I^{(\eta)}\tilde\Gamma_J^{(\zeta)})_{\beta_1\beta_2}F_{-\eta I,\zeta J;\gamma_1\cdots\gamma_m}(x) \\
&\qquad + (C\not P\tilde\Gamma_J^{(\zeta)}\tilde\Gamma_9)_{\beta_1\beta_2}F_{9,\zeta J;\gamma_1\cdots\gamma_m}(x)\Bigg],
\end{aligned} \tag{5.41}
$$

$$
\begin{aligned}
&\psi_{-\zeta J;\alpha_1\cdots\alpha_m;\beta_1\cdots\beta_4}(x) \\
&\quad = i^m(i/4P^2)^2\Pi_J^{(\zeta)\gamma_1}{}_{\alpha_1}\cdots\Pi_J^{(\zeta)\gamma_m}{}_{\alpha_m} \\
&\qquad \times(C\not P\tilde\Gamma_9\tilde\Gamma_J^{(\zeta)})_{\beta_1\beta_2}(C\not P\tilde\Gamma_9\tilde\Gamma_J^{(\zeta)})_{\beta_3\beta_4}F_{-\zeta J,\zeta J;\gamma_1\cdots\gamma_m}(x).
\end{aligned}
$$

These multispinors satisfy

$$
\begin{aligned}
&\Pi_J^{(\zeta)\alpha_r}{}_\gamma\psi_{\zeta J;\alpha_1\cdots\alpha_r\cdots\alpha_m;\beta_1\cdots\beta_{2p}}(x) = 0, \\
&\Pi_J^{(-\zeta)\beta_r}{}_\gamma\psi_{\zeta J;\alpha_1\cdots\alpha_m;\beta_1\cdots\beta_r\cdots\beta_{2p}}(x) = 0,
\end{aligned} \tag{5.42}
$$

For $\phi_{[111]}$, one obtains instead

$$
\begin{aligned}
\phi_{[111]} &= \sum_{m=0}^{8}\sum_{p=1}^{2}\sum_{I=1}^{4}\sum_{\eta=+,-}\chi_I^{(\eta)} \\
&\quad \times H_{\eta I}{}^{\alpha_1\cdots\alpha_m\beta_1\cdots\beta_{2p}}(\theta)\phi_{\eta I;\alpha_1\cdots\alpha_m;\beta_1\cdots\beta_{2p}}(x) \\
&= \sum_{m=0}^{8}\sum_{p=1}^{2}\sum_{k=1}^{9}\chi_k H_k{}^{\alpha_1\cdots\alpha_m\beta_1\cdots\beta_{2p}}(\theta) \\
&\quad \times\phi_{k;\alpha_1\cdots\alpha_m;\beta_1\cdots\beta_{2p}}(x),
\end{aligned} \tag{5.43}
$$

with

$$
\begin{aligned}
&\phi_{-\eta I;\alpha_1\cdots\alpha_m;\beta_1\beta_2}(x) \\
&\quad = i^{m+2}\Pi_I^{(\eta)\gamma_1}{}_{\alpha_1}\cdots\Pi_I^{(\eta)\gamma_m}{}_{\alpha_m} \\
&\qquad \times\Bigg\{\sum_{J=1}^{4}{}'\sum_{K=1}^{4}{}'\sum_{\zeta,\xi=+,-}(C\tilde\Gamma_I^{(\eta)}\tilde\Gamma_J^{(\zeta)}\tilde\Gamma_K^{(\xi)})_{\beta_1\beta_2} \\
&\qquad \times G_{\eta I,\zeta J,\xi K;\gamma_1\cdots\gamma_m}(x) \\
&\qquad + 2\sum_{J=1}^{4}{}'\sum_{\zeta=+,-}(C\tilde\Gamma_I^{(\eta)}\tilde\Gamma_J^{(\zeta)}\tilde\Gamma_9)_{\beta_1\beta_2} \\
&\qquad \times G_{\eta I,\zeta J,9;\gamma_1\cdots\gamma_m}(x)\Bigg\},
\end{aligned} \tag{5.44}
$$

$$
\begin{aligned}
&\phi_{-\eta I;\alpha_1\cdots\alpha_m;\beta_1\cdots\beta_4}(x) \\
&\quad = 2i^m\Pi_I^{(\eta)\gamma_1}{}_{\alpha_1}\cdots\Pi_I^{(\eta)\gamma_m}{}_{\alpha_m} \\
&\qquad \times\sum_{J=1}^{4}{}'\left(-\frac{i}{4P^2}C\not P\tilde\Gamma_I^{(\eta)}\tilde\Gamma_9\right)_{\beta_1\beta_2} \\
&\qquad \times(C\tilde\Gamma_I^{(\eta)}\tilde\Gamma_J^{(+)}\tilde\Gamma_J^{(-)})_{\beta_3\beta_4}G_{\eta I,-\eta I,9;\gamma_1\cdots\gamma_m}(x),
\end{aligned}
$$

where the primes mean that the value I is not allowed for the running index.

The ϕ fields in (5.44) also obey the restrictions in (5.42). The main difference between (5.39) and (5.43) is the term $p = 0$, which is absent in the second.

For the fermionic representation $\phi_{[(3/2)(1/2)(1/2)(1/2)]}$, the expansion we derive from (5.9a) is

$$
\begin{aligned}
&\phi_{[(3/2)(1/2)(1/2)(1/2)]} \\
&\quad = \mathrm{Re}\sum_{m=0}^{8}\sum_{p=1}^{4}\chi_1{}^{(+)}H_{+1}^{\alpha_1\cdots\alpha_m\beta_1\cdots\beta_{2p-1}}(\theta) \\
&\qquad \times\psi_{\alpha_1\cdots\alpha_m;\beta_1\cdots\beta_{2p-1}}(x),
\end{aligned} \tag{5.45}
$$

where the multispinors $\psi_{\alpha_1\cdots\alpha_m;\beta_1\cdots\beta_{2p-1}}(x)$ are given in terms of the B^+ fields in (5.9a) by

$$
\begin{aligned}
&\psi_{\alpha_1\cdots\alpha_m;\beta_1\cdots\beta_{2p-1}}(x) \\
&\quad = i^{m+2p-1}\Pi_1^{(-)\gamma_1}{}_{\alpha_1}\cdots\Pi_1{}^{(-)\gamma_m}{}_{\alpha_m} \\
&\qquad \times\Pi_1^{(+)\delta_1}{}_{\beta_1}\cdots\Pi_1^{(+)\delta_{2p-1}}{}_{\beta_{2p-1}}B^{+}_{\gamma_1\cdots\gamma_m;\delta_1\cdots\delta_{2p-1}}(x)
\end{aligned} \tag{5.46}
$$

and, of course, satisfy

$$
\begin{aligned}
&\Pi_1^{(+)\alpha_r}{}_\gamma\psi_{\alpha_1\cdots\alpha_r\cdots\alpha_m;\beta_1\cdots\beta_{2p-1}}(x) \\
&\quad = \Pi_1^{(-)\beta_r}{}_\gamma\psi_{\alpha_1\cdots\alpha_m;\beta_1\cdots\beta_r\cdots\beta_{2p-1}}(x) = 0.
\end{aligned} \tag{5.47}
$$

We will finish by looking at the massless limit of these superfields in order to see how they relate to the massless irreducible superfields of Ref. 8. To do so, we introduce the light cone projectors

$$\Pi_\pm = \tfrac{1}{2}(I \pm \Gamma_{0,9}) \quad (5.48)$$

which commute with all the projection operators we have introduced so far. Then we have the identity

$$Q_I^{(\eta)\alpha_1}\cdots Q_I^{(\eta)\alpha_m} = \sum_{l=0}^{m}\binom{m}{l} Q_{I,-}^{(\eta)[\alpha_1}\cdots Q_{I,-}^{(\eta)\alpha_l}Q_{I,+}^{(\eta)\alpha_{l+1}}\cdots Q_{I,+}^{(\eta)\alpha_m]},$$

$$Q_{I,\pm}^{(\eta)} = \Pi_\pm Q_I^{(\eta)}, \quad (5.49)$$

where the terms with $l > 4$ or $m - l > 4$ vanish identically. We also have the factorization

$$\chi_J^{(\eta)}(\theta) = \chi_J^{(\eta)}(\theta_+)\chi_J^{(\eta)}(\theta_-). \quad (5.50)$$

Equations (5.49) and (5.50) imply the identity that amounts to an addition formula for Grassmann–Hermite polynomials:

$$H_{\zeta J}^{\alpha_1\cdots\alpha_m}(\theta) = \sum_{l=0}^{m}\binom{m}{l} H_{\zeta J}{}^{[\alpha_1\cdots\alpha_l}(\theta_+)H_{\zeta J}{}^{\alpha_{l+1}\cdots\alpha_m]}(\theta_-), \quad (5.51)$$

where again the terms with $l > 4$ or $m - l > 4$ vanish identically. To take the massless limit $z \to 1$, we must take

$$\theta_+ \to 0 \quad (5.52a)$$

so that

$$\chi_J^{(\zeta)}(\theta_+) \to 1, \quad H_{\zeta J}{}^{\alpha_1\cdots\alpha_l}(\theta_+) \to 0, \quad l \text{ odd}, \quad (5.52b)$$

and then it is clear from (5.50), (5.51), and (5.52) that after some adequate field redefinitions (5.39) will take the form

$$\phi_{[2]} \to \sum_{k=1}^{8}\sum_{l=0}^{4}\chi_k(\theta_-)H_k^{\beta_1\cdots\beta_l}(\theta_-)\psi_{k;\beta_1\cdots\beta_l}(x), \quad (5.53)$$

which is nothing but the massless irreducible superfield $\phi_{[1]}$ obtained in Ref. 8.

The superfield $\phi_{[111]}$ vanishes in this massless limit. This is best seen by realizing that all the M operators in (5.26) vanish identically in that limit.

Expectedly the remaining superfield $\phi_{[(3/2)(1/2)(1/2)(1/2)]}$ becomes, in the limit, the second irreducible massless piece $\phi_{[(1/2)(1/2)(1/2)(1/2)]}$

$$\phi_{[(3/2)(1/2)(1/2)(1/2)]} \to \mathrm{Re}\sum_{m=0}^{4}\sum_{p=1}^{2}\chi_1^{(+)}(\theta_-) \times H^{\alpha_1\cdots\alpha_m\beta_1\cdots\beta_{2p-1}}(\theta_-)\psi_{\alpha_1\cdots\alpha_m;\beta_1\cdots\beta_{2p-1}}(x). \quad (5.54)$$

VI. CONCLUSION

We have found the expansions of the superfields $\phi_{[2]}$, $\phi_{[111]}$, and $\phi_{[(3/2)(1/2)(1/2)(1/2)]}$ that are the irreducible pieces of the scalar superfield in ten dimensions. This has been done by finding a complete set of eigenstates of the Cartan subalgebra corresponding to the representations [2] and [111]. Thus for [2] we have the 44 states in (5.11a)–(5.11d) which lead to the expressions for $\phi_{[2]}$ given in (5.12) and (5.23) as well as the Grassmann–Hermite polynomial expansion in (5.39). Similarly for [111] we have the states listed in (5.26a)–(5.26d), the expressions for $\phi_{[111]}$ (5.27) and (5.30), and the Grassmann–Hermite polynomial expansion in (5.43). The remaining one, $\phi_{[(3/2)(1/2)(1/2)(1/2)]}$, is easier to obtain because it is the only one containing odd powers of the covariant derivatives D in its expansion given in (5.9a) and (5.9b) while in (5.45) we displayed its expansion in Grassmann–Hermite polynomials. Then $\phi_{[2]}$ and $\phi_{[(3/2)(1/2)(1/2)(1/2)]}$ become in the massless limit the massless irreducible pieces $\phi_{[1]}$ and $\phi_{[(1/2)(1/2)(1/2)(1/2)]}$, respectively. Each of these three irreducible superfields $\phi_{[2]}$, $\phi_{[111]}$, and $\phi_{[(3/2)(1/2)(1/2)(1/2)]}$ includes the auxiliary field structure needed for the off-shell completion of a certain massless multiplet. However, $\phi_{[2]}$ contains more than the off-shell supergravity multiplet, even though its massless limit $\phi_{[1]}$ contains just the on-shell supergravity multiplet. This is clear from its physical field content.[7]

Our expressions are not covariant since they have been obtained in the collinear frame. As we have pointed out, the covariant problem (the eigenvalue equations for the Casimir operators) appears so far to be quite intractable. The covariantization of our expressions is certainly a simpler task. Once achieved one should be able to sort out our auxiliary field components into auxiliary fields. The solution to the four-dimensional problem that we have provided illustrates the subtleties involved.

ACKNOWLEDGMENT

We appreciate very helpful discussions with R. J. Finkelstein.

APPENDIX A: IDENTITIES IN FOUR DIMENSIONS

Since $\theta^\alpha\theta^\beta$ is antisymmetric in α,β, one has

$$\theta^\alpha\theta^\beta = -\tfrac{1}{4}[(C^{-1})^{\alpha\beta}\bar\theta\theta + (\Gamma_5 C^{-1})^{\alpha\beta}\bar\theta\Gamma_5\theta - (\Gamma_5\Gamma_A C^{-1})^{\alpha\beta}\bar\theta\Gamma_5\Gamma^A\theta],$$

and upon projection

$$\theta_\pm{}^\alpha\theta_\pm{}^\beta = -\tfrac{1}{4}(\Gamma_5\tilde\Gamma_\pm C^{-1})^{\alpha\beta}\bar\theta\Gamma_5\tilde\Gamma_\mp\theta.$$

On the other hand, the following identities are derived by Fierz transformations:

$$\not{P}\Gamma_5\tilde\Gamma_-\theta\bar\theta\not{P}\Gamma_{1,2}\theta = -iP^2\theta_-\bar\theta\Gamma_5\tilde\Gamma_-\theta,$$

$$\theta_-\bar\theta\not{P}\Gamma_{1,2}\theta = (i/2)\not{P}\Gamma_5\tilde\Gamma_-\theta_+\bar\theta\Gamma_5\tilde\Gamma_+\theta,$$

$$\theta\bar\theta\Gamma_5\tilde\Gamma_\pm\theta = -\Gamma_5\tilde\Gamma_\pm\theta\bar\theta\theta,$$

$$(\bar\theta\not{P}\Gamma_{1,2}\theta)^2 = -P^2\bar\theta\Gamma_5\tilde\Gamma_+\theta\bar\theta\Gamma_5\tilde\Gamma_-\theta = P^2(\bar\theta\theta)^2.$$

TABLE I. Index relabeling in (5.19).

k	ηI
1	+1
2	+2
3	+3
4	+4
5	−4
6	−3
7	−2
8	−1

APPENDIX B: NOTATION USED IN SEC. V

The letters η, ζ, and ξ are used to denote a sign $+$ or $-$. Indices labeled by the uppercase latin letters I, J, K run from 1 to 4 [rank of SO(9)]. Thus we can construct indices with negative and positive integer values just by multiplication: ηI will run from -4 to $+4$ skipping 0. Such an index can be relabeled into the values 1,...,8 of an index, which we denote always by the lowercase latin letter k, with range $k = 1,...,9$, according to Table I.

[1]S. Ferrara and P. van Nieuwenhuizen, Phys. Lett. B **74**, 333 (1978); K. S. Stelle and P. C. West, *ibid.* **74**, 330 (1978); M. F. Sohnius and P. C. West, *ibid.* **105**, 353 (1981).

[2]For a review of how to obtain $N = 1$ Poincaré supersymmetry from superconformal symmetry, see T. Kugo and S. Uehara, Nucl. Phys. B **226**, 49 (1983); while for the $N = 2$ case, see B. de Wit, J. W. van Holten, and A. van Proeyen, *ibid.* **167**, 186 (1980); **172**, 543 (1980).

[3]J. W. van Holten and A. van Proeyen, J. Phys. A **15**, 3763 (1982).

[4]E. Sokatchev, Nucl. Phys. B **99**, 96 (1975). The Casimir operators and its eigenvalues had been given before in A. Salam and J. Strathdee, *ibid.* **76**, 477 (1974).

[5]J. G. Taylor, Nucl. Phys. B **169**, 484 (1980); J. Kim, J. Math. Phys. **25**, 2037 (1984).

[6]R. Finkelstein and M. Villasante, J. Math. Phys. **27**, 1595 (1986).

[7]M. Villasante, Lett. Math. Phys. **11**, 351 (1986).

[8]P. Kwon and M. Villasante, J. Math. Phys. **29**, 560 (1988).

[9]E. B. Dynkin, Trudy Moskov. Mat. Obsc. **1**, 39 (1952); Am. Math. Soc. Transl. **6**, 245 (1965).

[10]N. Jacobson, *Lie Algebras* (Dover, New York, 1962), p. 247.

[11]H. Georgi, *Lie Algebras in Particle Physics* (Benjamin/Cummings, Reading, MA, 1982).

[12]J. Strathdee, "Introduction to supersymmetry," in *Supergravity '81*, edited by S. Ferrara and J. G. Taylor (Cambridge U.P., Cambridge, 1982), p. 7.

Relativistic potential scattering and phase shift analysis

R. Arshansky and L. P. Horwitz
School of Physics and Astronomy, Raymond and Beverly Sackler Faculty of Exact Sciences, Ramat Aviv 69 978, Tel Aviv, Israel

(Received 4 November 1987; accepted for publication 20 July 1988)

It is shown that two-body relativistic scattering cross sections can be represented in terms of phase shift analysis in essentially the same form as that of nonrelativistic scattering theory. The representation is covariant; the variable that corresponds to the orbital quantum number (and approaches it in the nonrelativistic limit) is relativistically *invariant*. The Levinson theorem is valid, and provides a link between the bound states, which have support in an O(2,1) invariant subspace of the full spacelike region of relative coordinates, and the scattering states that contain resonant behavior. These scattering states consequently have support in the same restricted subspace, and the same procedure may be used for the separation of variables. As an example, the resonances of an O(3,1) invariant "square well" are discussed.

I. INTRODUCTION

We have recently shown[1] that the two-body relativistic bound state for O(3,1) invariant direct action potentials is represented by a wave function with support in an O(2,1) invariant subspace of the full spacelike region of Minkowski space. In this region, one finds a lower mass ground state than one obtains from solutions of the problem with support in the full spacelike region, in particular, for a potential that is the invariant generalization of the nonrelativistic Coulomb potential (and coincides with it in the nonrelativistic limit). The selection of this O(2,1) invariant region, which we shall call the RMS (restricted Minkowski space), corresponds to a choice of synchronization, relating the pairs of events associated with the two particles that are taken to occur at the same value of the historical time parameter τ describing the dynamical evolution of the system.[2] We assumed[1] that this type of synchronization is characteristic of the bound state, and hence the entire physical Coulomb spectrum is given by the spectrum of this operator as constructed in the RMS. Furthermore, we assumed that it is applicable for bound states in general, and treated the space-time oscillator and relativistic square well in this way. The results have smooth extrapolation to the nonrelativistic limit, with both the spectra and wave functions in agreement with the results of the nonrelativistic Schrödinger theory. The classical relativistic bound state problem can be formulated in the same framework,[3] and it is found that the solutions that tend smoothly to the nonrelativistic orbits also have support in the RMS [i.e., for Casimir functions $\mathbf{L}^2 - \mathbf{A}^2 \geqslant 0$, for O(3,1), and $N^2 = L_3{}^2 - A_1{}^2 - A_2{}^2 \geqslant 0$, for O(2,1)]; this statement is also true for the classical scattering problem.[3]

In fact, as we shall argue, the Levinson theorem is valid in the relativistic case as well, and, since the relative space-time coordinates for the bound states occur with support in the RMS, this implies that resonant scattering (corresponding to passage through $\pi/2$ of the phase shifts) also has support in the RMS. Hence the scattering states that include resonant structure have synchronization characterized by the RMS.

In this paper, we shall develop the partial wave expansion for the scattering wave functions with support in the RMS; this sector is conserved and disjoint from other possible sectors, such as the complementary spacelike region.

Because of the equivalence between mass and energy in a relativistic theory in the unconstrained framework that we are using, it is formally possible for the initial and final masses of the individual particles to change as a result of the scattering interaction. In nature, one observes that the familiar particles have more or less definite masses, which are, in some cases, quite sharp, and it is an obvious question to ask whether an S matrix, which is used to measure transition probabilities from one definite mass state to another, in a theory in which all mass transitions are, in principle, possible, could be unitary. We shall show that the mass change of individual constituents controls the distribution of the geometrical (boost) parameter β in the final state, and that unitarity follows independently of this distribution (the different distributions of β correspond precisely to different synchronizations of the pair of events generating the particle world lines). There is, in this sense, no qualitative difference between elastic and (two-body) inelastic unitarity.

The partial wave expansion that we shall obtain contains phase shifts labeled by a quantum number l, which determines the value of the O(3,1) Casimir operator intrinsic to the RMS and corresponds to the nonrelativistic orbital angular momentum quantum number. The structure of the result, with this correspondence, is precisely of the form of the nonrelativistic theory, up to kinematical factors.

In a straightforward extension of the ideas of nonrelativistic scattering theory,[4] one obtains a cross section with the dimension of volume. The hyperarea in four-dimensional space-time perpendicular to the space direction of the incoming beam is three dimensional (L^2T). This cross section would include the scattering of an ensemble of processes that involve all possible distributions of β related to the synchronization of the pair of scattering events and to the mass change of the particles after scattering. What we have done in the present analysis is to restrict our consideration to scattering in a small neighborhood of a prescribed distribution of β corresponding, as we shall see, to a definite mass shift. The restriction of the incident current to an interval $d\beta$ requires the Jacobian factor $dx^0/d\beta$, reducing the dimensionality of the incident flux to that of current per unit area. The cross section that we shall obtain for scattering in the neighbor-

 0022-2488/89/010213-06$02.50

hood of a definite mass shift is therefore that of a two-dimensional area, as measured in laboratory experiments.

II. REPRESENTATION FOR THE ASYMPTOTIC WAVE FUNCTION

The general form of the two-body (two event) state with support for relative coordinates in the RMS was discussed in connection with the bound state problem in Ref. 1. Since the RMS is O(2,1) invariant, and not O(3,1) invariant, the representations for the full Lorentz group were obtained by labeling the spacelike direction stabilized by O(2,1) by a unit vector n_μ, and inducing representations over the spacelike hyperboloidal orbit of this unit vector. To do this, a set of standard RMS coordinates y_μ, referred by convention to the direction n_μ as its z axis, was constructed, and wave functions labeled by n_μ, on this set of accompanying space-time coordinates, provide representations for O(3,1) of this type. Under Lorentz transformation, $n \to \Lambda n$, and (as in the representation of a relativistic particle with spin) the wave function carries a representation of O(2,1) that moves along an orbit labeled by n_μ, accompanied by a Wigner "rotation," in this case, an O(2,1) transformation. Furthermore, we analyzed the structure of this motion of the representation along the orbit into irreducible representations of $O(3) \subset O(3,1)$, with quantum numbers (L,q), and showed that one obtains in this way the principal series of Gel'fand. The bound states of the two-body problem can be described equally well for any spacelike choice of n_μ, and the mass levels are completely degenerate with respect to this choice. The unitarity of the Gel'fand representation involves an integration on the whole manifold of the Lorentz group; the scalar product for the theory therefore contains an integration over the hyperboloidal orbit with measure $d^4n\,\delta(n^2-1)$ as well as on the accompanying coordinates y_μ.

In the problem of two-body scattering, the direction of the beam selects a definite spacelike direction. For the component of the general wave function for which n_μ is oriented along this direction, one can argue [for an O(3,1) symmetric potential] that the scattered wave will be maximally symmetric around this axis. The maximally symmetric state is the one for which the Gel'fand representation contains only the value of L corresponding to the lowest weight state of the tower, and we shall assume that the scattering matrix (which is diagonal in n_μ) is completely described, to a good approximation, by such a state. The result we shall obtain agrees in form with the well-known nonrelativistic partial wave expansion, which has been useful in describing experimental data. Alternative choices of n_μ necessarily result in states with less than maximal symmetry; while we do not rule out their occurrence, they evidently do not contribute in an important way to the effectiveness of the partial wave representation for real scattering experiments. We shall therefore restrict our attention to this special class of scattering states in this paper.

The full evolution operator for a two-body system with O(3,1) invariant potential is given by

$$K = p_1{}^2/2M_1 + p_2{}^2/2M_2 + V(\rho) = P^2/2M + p^2/2m + V(\rho), \tag{2.1}$$

where [$p^2 = p^\mu p_\mu$; we use $g_{\mu\nu} = (-1,1,1,1)$, $\hbar = c = 1$]

$$p^\mu = p_1{}^\mu + p_2{}^\mu, \quad p^\mu = (M_2 p_1{}^\mu - M_1 p_2{}^\mu)/(M_1+M_2),$$
$$\rho = \sqrt{(x_1 - x_2)^2}. \tag{2.2}$$

The general form of the τ-independent two-body wave function in relative coordinates in the continuous spectrum $\{\kappa/2m\}$ of $K_{\rm rel}$, where

$$K_{\rm rel} = p^2/2m + V(\rho), \tag{2.3}$$

corresponding to a definite κ, c_2 (the value of the second Casimir operator $L\cdot A$ for the full Lorentz group), and a definite direction n^μ is[1]

$$\psi_{n_\mu}{}^{c_2}(y) = \left[\frac{1}{\sqrt{\rho \sin\theta \cosh\beta}}\right] \sum_{l,n,k,L,q} A^{Lq}{}_{lnk} \hat{R}{}^\kappa_l(\rho) \hat{\Theta}{}^n_l(\theta) \times \Xi^{c_2,L}_{n,k}(u) P^L_{q-M_k}(z) e^{-iq\gamma} \hat{\chi}_{n+k}{}^{-n}(\beta,\varphi), \tag{2.4}$$

where the measure on the Hilbert space $\mathscr{H}_n$, to which this function belongs, is $d\mu = \rho^3 \sin^2\theta \cosh\beta\, d\beta\, d\theta\, d\varphi$. The integer parameter n, determining the Casimir operator for the O(2,1) little group, plays the role of the magnetic quantum number in the corresponding nonrelativistic problem; it fixes the relation between the value c_1 of the first Casimir operator $\mathbf{L}^2 - \mathbf{A}^2$ and c_2 according to ($\hat{n} = n + \frac{1}{2}$)

$$-c_1 = 1 - \hat{n}^2 + c_2{}^2/\hat{n}^2. \tag{2.5}$$

The integer quantum number l, determining the value of the formal first Casimir operator of O(3,1) in the RMS, plays the role of the orbital angular momentum in the corresponding nonrelativistic problem.

The variables $u = \tanh\alpha$, $z = \sin\omega$, and γ, describe the orientation of the spacelike vector n^μ, and $\rho,\theta,\beta,\varphi$ correspond to the relative coordinates y^μ in the RMS defined by n^μ.[1] The equations for the functions $R^\kappa_l(\rho)$, $\Theta^n_l(\theta)$ satisfying the differential equations in ρ,θ after separation of variables become identical to the radial and angular part of the corresponding nonrelativistic problem when the factors $1/\sqrt{\rho}$ and $1/\sqrt{\sin\theta}$ are extracted; we have denoted the corresponding functions as $\hat{R}{}^\kappa_l(\rho)$, $\hat{\Theta}{}^n_l(\theta)$ and provided these factors explicitly. We have also extracted the factor $1/\sqrt{\cosh\beta}$ from the functions $\chi^{-n}{}_{n+k}(\beta,\varphi)$, which are the irreducible representations of O(2,1).

In the configuration we shall consider here, for which n_μ is directed parallel to the incoming beam, the stationary wave function should have maximal symmetry around this axis, which we take to be the z axis (the accompanying coordinates $\{y^\mu\}$ then coincide with the laboratory space-time coordinates $\{x^\mu\}$, and the parameters α, ω, and γ are zero). In this case, only the lowest weight state in the Gel'fand representation contributes, i.e., $L = \frac{1}{2}$ and $n = k = 0$ [for states of definite L, $0 \leqslant k \leqslant L - n - \frac{1}{2}$ (see Ref. 1)]. We are therefore left with the simple form

$$\psi(x) = \sum_{l=0}^{\infty} \left[\frac{A_l \hat{R}_l{}^\kappa(\rho) P_l(\cos\theta) e^{i\varphi/2}}{\sqrt{\rho \sin\theta \cosh\beta}}\right]. \tag{2.6}$$

The coefficients A_l can be determined, as for the corresponding nonrelativistic problem, by requiring ψ to take the asymptotic value for the incoming wave ($\tau \to -\infty$ for a wave packet on κ),

$$\psi_{\rm inc} \sim e^{i\kappa\rho\cos\theta} e^{i\varphi/2}/\sqrt{\rho\cosh\beta\sin\theta}\,, \tag{2.7}$$

where we have used the fact that $\hat{R}\,^{\kappa}_{l}(\rho)$ is a solution of the nonrelativistic radial equation in the independent variable ρ, and the spacelike interval $\rho \to \infty$ in this limit. Equation (2.7) *with* the denominator, is a solution of the free generalized eigenvalue problem, i.e., it is a generalized eigenstate for the momenta

$$\begin{aligned}
p_0 &= -i\left(\frac{\partial}{\partial x^0}\right) = -i\Big[-\sin\theta\sinh\beta\frac{\partial}{\partial\rho} \\
&\quad -\left(\frac{1}{\rho}\right)\cos\theta\sinh\beta\frac{\partial}{\partial\theta} \\
&\quad +\left(\frac{\cosh\beta}{\rho\sin\theta}\right)\frac{\partial}{\partial\beta}\Big], \\
p_1 &= -i\left(\frac{\partial}{\partial x^1}\right) = -i\Big\{\cos\varphi\Big[\sin\theta\cosh\beta\,\frac{\partial}{\partial\rho} \\
&\quad +\left(\frac{1}{\rho}\right)\cos\theta\cosh\beta\frac{\partial}{\partial\theta} \\
&\quad -\left(\frac{\sinh\beta}{\rho\sin\theta}\right)\frac{\partial}{\partial\beta}\Big] \\
&\quad -\left(\frac{\sin\varphi}{\rho\cos\theta\cosh\beta}\right)\frac{\partial}{\partial\varphi}\Big\}, \\
p_2 &= -i\left(\frac{\partial}{\partial x^2}\right) = -i\Big\{\sin\varphi\Big[\sin\theta\cosh\beta\frac{\partial}{\partial\rho} \\
&\quad +\left(\frac{1}{\rho}\right)\cos\theta\cosh\beta\frac{\partial}{\partial\theta} \\
&\quad -\left(\frac{\sinh\beta}{\rho\sin\theta}\right)\frac{\partial}{\partial\beta}\Big] \\
&\quad +\left(\frac{\cos\varphi}{\rho\sin\theta\cosh\beta}\right)\frac{\partial}{\partial\varphi}\Big\}, \\
p_3 &= -i\left(\frac{\partial}{\partial x^3}\right) = -i\left[\cos\theta\frac{\partial}{\partial\rho} - \left(\frac{1}{\rho}\right)\sin\theta\frac{\partial}{\partial\theta}\right],
\end{aligned} \tag{2.8}$$

with eigenvalue κ for p_3 (for $\rho \to \infty$); asymptotically, the eigenvalues for the other components vanish so that $p_\mu \sim (0,0,0,\kappa)$.

We remark that (2.7) also leads to the correct current for the incident particle beam. The conserved relative current for the (τ-independent) stationary scattering states is[5]

$$j_\mu = -(i/2m)\psi^*(x)\overleftrightarrow{\partial}_\mu\psi(x), \tag{2.9}$$

where we have recognized that it is the *relative* current that is relevant for the counting of scattering events, and that the τ integration required for the construction of a conserved current[5] serves, in the free particle case, to link the κ^2 (mass squared) values of the two wave-function factors with a δ function. In an interval $d\kappa^2/2\pi$, one obtains (2.9) (with the ψ's at equal mass); with (2.7), one obtains (the other components vanish)

$$j_z = (\kappa/m)(1/\rho\sin\theta\cosh\beta). \tag{2.10}$$

As we shall show, the relation between the parameters β and θ determines the synchronization between the pair of events being considered, and this determines the mass change during the scattering. We shall therefore be interested in a definite value of β, in an interval $d\beta$, for the current. The measure for j_z, as a density, is $d^3x\,dx^0$; in an interval $d\beta$, we have for the particle current per unit area,

$$j_z\left(\frac{dx^0}{d\beta}\right)d\beta = \frac{\kappa}{m}d\beta. \tag{2.11}$$

Since $\hat{R}\,^{\kappa}_{l}(\rho)$ is a solution of a radial equation of Schrödinger type, the τ-independent interacting wave function has the form

$$\psi^{(+)}(x) \sim \left[(1/\sqrt{\rho\cosh\beta\sin\theta})e^{i\varphi/2}\right] \times\{e^{i\kappa\rho\cos\theta} + (1/\rho)f(\theta)e^{i\kappa\rho}\}, \tag{2.12}$$

where, in the Legendre expansion of $f(\theta)$ the coefficients (following the usual arguments for asymptotic values of ρ) are related to a set of phase shifts $\delta_l(\kappa)$ according to

$$f(\theta) = \frac{1}{2i\kappa}\sum_{l=0}^{\infty}(2l+1)(S_l - 1)P_l(\cos\theta), \tag{2.13}$$

and

$$S_l = e^{2i\delta_l}, \tag{2.14}$$

where S_l is the l component of the S matrix. The numbers δ_l are the same as the nonrelativistic phase shifts since the radial equation [for $V(\rho)$ of the same form] for $\hat{R}(\rho)$ is identical to that of the nonrelativistic problem. We shall show, in fact, that the differential cross section has the same form, as a function of θ, as for the corresponding nonrelativistic problem as well.

III. DIFFERENTIAL CROSS SECTION

To calculate the differential cross section, we compute the asymptotic outgoing current associated with the scattered part of the wave function (2.12), i.e., the second term. The partial derivatives entering the components of (2.9) in terms of the angular and hyperangular parameters of the RMS are given in (2.8). Derivatives with respect to θ, β, and φ contain factors $1/\rho$ and therefore vanish asymptotically. The contributions of the ρ derivatives lead to

$$j^{0(\rm scatt)}\rho^3\sin^2\theta\cosh\beta\,d\theta\,d\varphi\,d\beta = \sin\theta\sinh\beta(\kappa/m)|f(\theta)|^2\sin\theta\,d\theta\,d\varphi\,d\beta, \tag{3.1}$$

$$j^{1(\rm scatt)}\rho^3\sin^2\theta\cosh\beta\,d\theta\,d\varphi\,d\beta = \cos\varphi\sin\theta\cosh\beta(\kappa/m)|f(\theta)|^2\sin\theta\,d\theta\,d\varphi\,d\beta, \tag{3.2}$$

$$j^{2(\rm scatt)}\rho^3\sin^2\theta\cosh\beta\,d\theta\,d\varphi\,d\beta = \sin\theta\sin\varphi\cosh\beta\,(\kappa/m)|f(\theta)|^2\sin\theta\,d\theta\,d\varphi\,d\beta, \tag{3.3}$$

$$j^{3(\rm scatt)}\rho^3\sin^2\theta\cosh\beta\,d\theta\,d\varphi\,d\beta = \cos\theta(\kappa/m)|f(\theta)|^2\sin\theta\,d\varphi\,d\beta, \tag{3.4}$$

where we have multiplied by the four-volume element divided by $d\rho$, the infinitesimal volume element lying on the constant ρ hypersurface. This volume element consists of the measure dx^0 that we have discussed above, and a two-dimensional space surface element which is of the form of ρ^2 times a solid angle. The quantities (3.1)–(3.4) therefore correspond to the number of particles per unit time scattered through this surface element in each of the four space-time directions. We are concerned with the flow of particles through this surface element in the scattering directions specified by θ, φ, and β, i.e., parallel to the unit vector[1]

$$x^\mu/\rho = (\sin\theta\sinh\beta, \sin\theta\cosh\beta\cos\varphi, \sin\theta\cosh\beta\sin\varphi, \cos\theta). \tag{3.5}$$

We obtain for the number of particles per unit time through this surface element

$$(x^\mu/\rho) j_\mu^{(\mathrm{scatt})} = (\kappa/m)|f(\theta)|^2 \sin\theta\, d\theta\, d\varphi\, d\beta. \tag{3.6}$$

Dividing by the incident flux, we obtain the differential cross section (we have integrated over the azimuthal angle φ)

$$d\sigma(\theta) = 2\pi|f(\theta)|^2\, d\Omega(\theta), \tag{3.7}$$

where

$$d\Omega(\theta) = \sin\theta\, d\theta. \tag{3.8}$$

We note that the differential cross section we have obtained has the dimension of area.

IV. KINEMATICAL RESTRICTIONS FOR ELASTIC SCATTERING

In this section we obtain a relation for β for elastic scattering, i.e., where there is no mass change of the individual particles involved in the scattering process. It follows from (2.2) that

$$p_1^\mu = (M_1/M)P^\mu + p^\mu, \tag{4.1}$$

and hence

$$p_1^2 = -(M_1^2/M^2)s + p^2 + (2M_1/M)P^\mu p_\mu, \tag{4.2}$$

where the Mandelstam variable $s = -P^2$ is absolutely conserved, and p^2 is conserved asymptotically when there is no change in the effective free evolution operator in initial and final states.[6] We shall assume that this is the case in our consideration of elastic processes. Using the asymptotic relation

$$K \sim -(s/2M) + p^2/2m \tag{4.3}$$

and taking $K \cong -M/2$ (for $p_i^2 \cong -M_i^2$, close to their "on shell" values) we obtain, for the asymptotic value of p^2,

$$\kappa^2 = p^2 \sim (m/M)(s - M^2). \tag{4.4}$$

Equation (4.2) then becomes

$$p_1^2 \cong s(M_1/M^2)(M_2 - M_1) - M_1M_2 + (2M_1/M)P^\mu p_\mu. \tag{4.5}$$

Applying the operator [the last equation of (2.8)]

$$-i\frac{\partial}{\partial x^3} = -i\left[\cos\theta\frac{\partial}{\partial\rho} - \left(\frac{1}{\rho}\right)\sin\theta\frac{\partial}{\partial\theta}\right] \tag{4.6}$$

to the incident wave function given by Eq. (2.7), we obtain the generalized eigenvalue κ for p_z (for $\rho \to \infty$; all other components of p^μ vanish asymptotically) and hence for the negative mass squared of the incident particle we obtain

$$p_1^2 \cong s(M_1/M^2)(M_2 - M_1) - M_1M_2 + (2M_1/M)P_z\kappa. \tag{4.7}$$

We remark that the kinematical structure of the state we are considering is not consistent with scattering in the center-of-mass frame (unless $M_1 = M_2$). The result that we shall obtain can, of course, be transformed to the center-of-mass or any other frame as we will explain below. The fourth component of the relative momentum is

$$p^0 = (M_2 p_1^0 - M_1 p_2^0)/(M_1 + M_2). \tag{4.8}$$

In the center-of-mass frame, $\mathbf{p}_1 = -\mathbf{p}_2$, and on asymptotic mass shells, $p_i^0 = \sqrt{p_i^2 + M_i^2}$; hence, unless $M_1 = M_2$, p^0 could not be zero. If $M_1 = M_2$, and $p_1^2 \cong -M_1^2$, then it follows from (4.7) that $P_z = 0$, i.e., our kinematics corresponds to the center-of-mass frame in this case. However, if $M_1 \neq M_2$, then P_z is determined by (4.7) and is not zero.[7]

It is consistent to take $P_x = P_y = 0$, and we shall do so for simplicity in the following. In the final state, we recognize that the relative momentum associated with a measured asymptotic event is parallel to the relative space-time coordinate and hence[1]

$$p'^\mu \cong \kappa(\sin\theta\sinh\beta, \cos\varphi\sin\theta\cosh\beta, \sin\theta\sin\varphi\cosh\beta, \cos\theta). \tag{4.9}$$

Equation (4.2) (for $P^0 = \sqrt{s + P_z^2}$) becomes

$$p_1'^2 \cong s(M_1/M^2)(M_2 - M_1) - M_1M_2 + (2M_1/M)\kappa(-\sqrt{s+P_z^2}\sin\theta\sinh\beta + P_z\cos\theta). \tag{4.10}$$

Assuming that $p_1'^2 = p_1^2$, it follows from (4.7) and (4.10) that

$$P_z(1 - \cos\theta) = -\sqrt{s + P_z^2}\sin\theta\sinh\beta \tag{4.11}$$

or

$$\sinh\beta = -(P_z/\sqrt{s + P_z^2})\tan(\theta/2). \tag{4.12}$$

This result determines β as a function of s and θ, since, according to (4.4),

$$\kappa^2 = (m/M)(s - M^2) \tag{4.13}$$

and P_z is then determined as a function of s by (4.7) for $p_1^2 \cong -M_1^2$.

If $p_1'^2 \neq p_1^2$, i.e., in case of a mass change for the individual particles after scattering, (4.7) and (4.10) would yield a formula of the type (4.11) with an additional term proportional to $\Delta p_1^2/\kappa$; the hyperbolic angle β is still determined but at a value different from (4.12). We consider in the following the case of elastic scattering only.

In the frame in which $p^0 = 0$ and the initial relative momentum is oriented along the z direction, we may write the final state momenta in terms of the magnitude $|\mathbf{p}'|$, φ, and a laboratory angle $\hat{\theta}$:

$$p_1' = |\mathbf{p}'|\sin\hat{\theta}\cos\varphi,$$
$$p_2' = |\mathbf{p}'|\sin\hat{\theta}\sin\varphi, \tag{4.14}$$
$$p_3' = |\mathbf{p}'|\cos\hat{\theta}.$$

It follows from (4.9) that (note that $p^{\mu\prime}p_\mu' = \kappa^2$)

$$|\mathbf{p}'|^2 = \kappa^2 + \kappa^2\sin^2\theta\sinh^2\beta = \kappa^2\{1 + [P_z^2/(s + P_z^2)](1 - \cos\theta)^2\}, \tag{4.15}$$

where the last is obtained from (4.11), and hence, with the last of (4.14),

$$\cos \hat{\theta} = \cos \theta / \sqrt{1 + (P_z{}^2/(s + P_z{}^2))(1 - \cos \theta)^2}. \tag{4.16}$$

The cross section (3.7) may now be expressed in terms of measurable laboratory parameters, $|\mathbf{p}'|$, $\hat{\theta}$,

$$d\hat{\sigma}(\hat{\theta}) \equiv d\sigma(\theta(\hat{\theta})) = 2\pi |f(\theta(\hat{\theta}))|^2 \, d\Omega(\theta(\hat{\theta})). \tag{4.17}$$

Since the number of particles through the surface element we are considering is, as may be seen from the construction (3.6), invariant, one may transform, alternatively, to obtain the cross section in the center-of-mass frame. Using the relative velocity $v = P_z/P^0$ to construct this transformation on (4.9), one obtains [with (4.11)] the relation

$$\cos \theta_{\rm cm} = (s \cos \theta + P_z{}^2)/(s + P_z{}^2). \tag{4.18}$$

For measurements in the center-of-mass system, one uses $\theta(\theta_{\rm cm})$ in place of $\theta(\hat{\theta})$ in (4.17); note that

$$|\mathbf{p}_{\rm cm}| = \kappa \sqrt{(s + P_z{}^2)/s}. \tag{4.19}$$

V. UNITARITY AND THE LEVINSON THEOREM

We shall demonstrate unitarity on the scattering states by comparing the total probability to find a particle in the scattering state $\psi^{(+)}(x)$ given by (2.12) with the total probability contained in the incoming wave (this procedure actually verifies the unitarity of the *wave operator* on scattering states, which is sufficient). The total probability for the incoming wave (2.7) is

$$\int \rho^3 \sin^2 \theta \cosh \beta \, d\rho \, d\theta \, d\varphi \, |\psi_{\rm inc}(x)|^2 = \int_0^R d\rho = R, \tag{5.1}$$

for an interval $d\beta$, as for our computation of the current in (2.11). In the computation of the norm of $\psi^{(+)}(x)$, in an interval $d\beta$, one finds, with the help of the orthogonality relations for Legendre polynomials (the measure on θ, after cancellation with the square root factor in the wave function, is $\sin \theta \, d\theta$), a volume equal to (5.1), and hence we have demonstrated unitarity in the interval $d\beta$ for each β.

As we have remarked, the value of β (for each θ) is associated with the mass change of the individual particles. Our result shows that the S matrix is unitary for every possible choice of synchronization determined by β, i.e., for each possible mass change.

It follows, furthermore, from (2.13) and (2.14) that the optical theorem holds in the usual form.

In the course of our calculation we have used the RMS based on a spacelike direction determined by the direction of the incident beam. Only under this condition can we argue that the two-body state has maximum symmetry. The unitarity that we have demonstrated is therefore applicable with respect to all outgoing waves related (by the wave operator for this orientation of n^μ) to this direction of the incoming beam. The formal description of a complete scattering system would necessarily include incoming plane waves of any possible direction and, therefore, the complete set of all RMS(n^μ). What we have shown is that unitarity is valid for each of these orientations. A more complete treatment of the general structure of formal scattering theory will be given elsewhere (in this more general case, complete unitarity would include integration over the hypersurface defined by $n^2 = 1$, as in the unitarity of the Gel'fand representation).

We now turn to the Levinson theorem. The analytic properties and interpretation of $S_l(\kappa)$ follow from the radial equation and the asymptotic form (2.12) for the outgoing waves. These properties are therefore identical to those of nonrelativistic scattering theory. Following Levinson[8] (see Ref. 9, p. 72), we identify the part of the wave function with asymptotic behavior $\sim \exp(+i(\kappa\rho - \pi/2))$; the limit of this function for $\rho \to 0$ (on the light cone) is called $D_l(\kappa)$; then, $D_l^*(\kappa) = (-1)^l D_l(-\kappa)$,

$$S_l(\kappa) = D_l^*(\kappa)/D_l(\kappa), \tag{5.2}$$

and it follows from integration on κ from $-\infty$ to $+\infty$ $[\delta_l(\kappa) = -\delta_l(-\kappa)]$ that

$$\delta_l(\infty) - \delta_l(0) = -\pi N_b^l, \tag{5.3}$$

where N_b^l is the number of bound states for a given l.

As we have mentioned, this connection between the scattering phase shifts and the bound states is one of the reasons for restricting our study of the scattering states to the RMS. Moreover, we see that the bound states with support in the RMS associated with a given direction of n^μ are directly associated with scattering in the corresponding RMS.

VI. EXAMPLE

Let us consider the example of a 3 + 1 relativistic square well. The potential has the form

$$V(\rho) = \begin{cases} 0, & \rho > a, \\ -U, & \rho < a. \end{cases} \tag{6.1}$$

This potential has value $-U$ inside the single sheeted spacelike hyperboloid $\rho \leqslant a$. The solutions to the radial equation are [we have absorbed the factor A_l in (2.6) in the coefficients here]

$$\hat{R}_l(\rho) = C_l j_l(\kappa_1 \rho), \quad \rho \leqslant a, \tag{6.2}$$

$$\hat{R}_l(\rho) = A_l j_l(\kappa_0 \rho) + B_l n_l(\kappa_0 \rho), \quad \rho > a, \tag{6.3}$$

where

$$\kappa_0 = \sqrt{2mK_a/\hbar^2}, \quad \kappa_1 = \sqrt{2m(K_a + U)/\hbar^2}. \tag{6.4}$$

Here, as for the corresponding nonrelativistic problem,

$$\begin{aligned} A_l &= (2l+1) i^l \cos \delta_l \, e^{i\delta_l}, \\ B_l &= -(2l+1) i^l \sin \delta_l \, e^{i\delta_l}, \end{aligned} \tag{6.5}$$

and it follows from the continuity conditions that

$$C_l = i^l e^{i\delta_l} \cos \delta_l [j_l(\tau) - \tan \delta_l n_l(\tau)]/j_l(\tau), \tag{6.6}$$

and

$$\kappa_1 j_l'(\theta)/j_l(\theta) = \kappa_0 (j_l'(\tau) - \tan \delta_l n_l'(\tau))/(j_l(\tau) - \tan \delta_l n_l(\tau)), \tag{6.7}$$

where (the symbols defined in this section should not be confused with some of those used earlier)

$$\theta = \kappa_1 a, \quad \tau = \kappa_0 a. \tag{6.8}$$

Solving (6.7) for $\tan \delta_l$, one obtains

$$\tan\delta_l = \left[\left(\frac{n_l(\tau)}{j_l(\tau)}\right) - \left(\frac{\kappa_0}{\kappa_1}\right)\left(\frac{n_l'(\tau)}{j_l(\tau)}\right)\left(\frac{j_l(\theta)}{j_l'(\theta)}\right)\right] \times\left[1+\left(\frac{\kappa_0}{\kappa_1}\right)\left(\frac{j_l'(\tau)}{j_l(\tau)}\right)\left(\frac{j_l(\theta)}{j_l'(\theta)}\right)\right]^{-1}. \tag{6.9}$$

In the limit of large U, i.e., for a very deep well (but, as we shall point out, for $U \lesssim M/2$),

$$(j_l(\theta)/j_l'(\theta)) \sim \tan(\theta - l\pi/2). \tag{6.10}$$

Resonances occur for

$$\theta = a\kappa_1 = (2n+l+1)(\pi/2), \tag{6.11}$$

and, at these values,

$$\tan\delta_l = -[n_l'(\tau)/j_l'(\tau)]. \tag{6.12}$$

Note that for $K \cong -M/2$, the invariant scattering energy squared (Mandelstam variable) is

$$s = M^2 + 2MK_a, \tag{6.13}$$

so that small K_a corresponds to low energy scattering, which we shall study here (the energies in the well are, however, large for large U). Since in this case τ is small, we may use the asymptotic forms

$$j_l(\tau) \sim [\tau^l/(2l+1)!!][1-(\tau^2/2)/(2l+3)], \tag{6.14}$$

$$n_l(\tau) \sim -[(2l+1)!!/(2l+1)\tau^{l+1}], \tag{6.15}$$

where we have retained the next to leading order in (6.14) in case $l=0$. From these, we obtain

$$\tan\delta_l \sim -((l+1)/l) \times\{[(2l+1)!!]^2/(2l+1)\}(1/\tau^{2l+1}) \quad (l \neq 0)$$

and

$$\tan\delta_0 \sim 2/\tau^3. \tag{6.16}$$

Clearly, for τ small, $\tan\delta_l$ is large, and hence δ_l is close to $\pi/2$. Hence the conditions (6.11) lead to "virtual" resonances.

The resonance conditions (6.11) and the relativistic kinematic relation (6.13) provide the resonant energy conditions

$$s = M^2 - 2MU + \left(\frac{M}{m}\right)\left(\frac{\hbar^2}{a^2}\right)(2n+l+1)^2\left(\frac{\pi}{2}\right)^2. \tag{6.17}$$

We have used the condition that the last two terms in this equation combine to a small number. It was pointed out in Ref. 1 that the validity of the estimate $K \cong -M/2$, to ensure the positivity of s (for small n), requires that

$$U \lesssim M/2. \tag{6.18}$$

For U not much smaller than this upper bound, the first two terms in (6.17) can become small compared to the third term, i.e., for

$$n \gg \sqrt{m(M-2U)}\,(2/\pi\hbar). \tag{6.19}$$

In this case, the center-of-mass energy goes approximately linearly with l and n (a result found in the phenomenological application of Regge pole theory):

$$\sqrt{s} \sim (\hbar\pi/2a)\sqrt{(M/m)}\,(2n+l+1). \tag{6.20}$$

For U close to the bound (6.18), it follows from (6.4) that

$$\kappa_1 > \sqrt{(M_1M_2/\hbar^2)} \tag{6.21}$$

since K_a is positive. The range of our approximation is therefore relativistic.

VII. CONCLUSIONS

We have argued that the Levinson theorem implies that resonant scattering is represented by wave functions with support in a restricted O(2,1) invariant region of the full spacelike domain of relative motion (RMS) in a two-body scattering system, since the bound states have support in the RMS. Furthermore, the most symmetric scattering amplitude is associated with an RMS oriented in the direction of the incident beam.

The partial wave expansion obtained in this way is represented as a Legendre series with coefficients determined by a set of phase shifts δ_l, where l is the quantum number, with integer values $l = 0,1,2,3...$ corresponding to the O(3,1) Casimir operator [which is well-defined in the O(2,1) invariant RMS]. In the nonrelativistic limit, O(3,1)$\to$O(3) (note that we are discussing here the *relative* motion), and l smoothly goes to the nonrelativistic quantum number labeling its Casimir operator.[1]

For any fixed value of $\Delta p_1^{\,2}$, the mass change of an individual particle in the scattering process, the boost variable is determined as a specific function of the scattering angle θ. We studied, in particular, the elastic case, and showed that the resulting (two-dimensional) cross section has a form similar to that of the nonrelativistic cross section, but with relativistic kinematic corrections associated with the angle of observation θ. We have shown, furthermore, that the scattering process is unitary for every possible value of mass change of the particles.

As an illustration we studied the problem of scattering for the relativistic square well potential. In the nonrelativistic limit the resonance structure agrees with that of the usual nonrelativistic problem, but in the case in which n is large and for any l, where the wave number may be large compared to the Compton wave number of the geometric mean of the constituent masses, we find equal spacing in l and n.

ACKNOWLEDGMENT

This work was supported in part by the U.S.-Israel Binational Science Foundation, Jerusalem, Israel.

[1] R. Arshansky and L. P. Horwitz, Tel Aviv University preprints TAUP 1467-86 and TAUP 1481-86, 1986. See also, R. Arshansky and L. P. Horwitz, Phys. Lett. A **128**, 123 (1988). A brief summary of the work of this paper appears in R. I. Arshansky and L. P. Horwitz, Phys. Lett. A **131**, 222 (1988).

[2] L. P. Horwitz and C. Piron, Helv. Phys. Acta **46**, 316 (1973).

[3] R. Arshansky, Tel Aviv University preprint TAUP 1479-86, 1986.

[4] L. P. Horwitz and Y. Lavie, Phys. Rev. D **26**, 819 (1982).

[5] R. Arshansky, L. P. Horwitz, and Y. Lavie, Found. Phys. **13**, 1167 (1983).

[6] This is true if the normal unmodified wave operator exists. In this case, it follows from (2.1) that $\Delta p_1^{\,2}/M_1 + \Delta p_2^{\,2}/M_2 = 0$. There may be cases of scattering with an effective potential for which only a nontrivially modified wave operator exists, and this relation must be altered.

[7] If the term $P_z\ \kappa$ in (4.7) vanishes for $M_1 \neq M_2$, it follows (for $p_1^{\,2} = -M_1^{\,2}$) that $s = M^2$; this implies $\mathbf{p}_1$ and $\mathbf{p}_2$ are zero, and there is no scattering.

[8] N. Levinson, Dan. Vid. Selsk. Mat. Fys. Medd. **25**, No. 9 (1949).

[9] See, for example, A. I. Baz, Ya. B. Zel'dovich, and A. M. Perelomov, *Scattering, Reactions and Decay* (Israel Program for Scientific Translations, Jerusalem, 1969).

On the boson–quasifermion realization of the particle–hole SO($2\Omega+1$) algebra

Abraham Klein
Department of Physics, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Eugene R. Marshalek
Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556

(Received 19 July 1988; accepted for publication 24 August 1988)

The shell-model algebra SO($2\Omega + 1$) generated by all bilinear and linear combinations of fermion creation and annihilation operators acting on a Fock space of Ω orbitals may be mapped into an "ideal space" in which particle–hole pairs are described by boson operators, while excess particles or holes are described by fermionlike degrees of freedom called "quasifermions." After a review of the derivation of the nonunitary generalized Dyson realization, it is shown how this realization can be unitarized in a very simple way with the help of a recently developed technique that utilizes the Casimir invariants of a "core subalgebra."

I. INTRODUCTION

It is well-known that classical Lie algebras can be realized in terms of bilinear combinations of either fermion or boson operators.[1] The former arise naturally in the context of the nuclear shell model. In recent years it has proven fruitful to map fermion shell-model algebras onto boson realizations,[2] or in some cases, onto what we call boson–quasifermion realizations, as discussed below. The boson realizations, however, are not necessarily bilinear, but may involve higher-degree boson polynomials, and, in some cases, infinite expansions in boson polynomials. In this way it becomes possible to convert the original many-fermion problem into an equivalent many-boson problem, albeit one with additional constraints necessary to fulfill the Pauli principle. One advantage of such a metamorphosis is that it facilitates the development of new kinds of many-body approximations that would be very difficult to implement in the original fermion picture. Recently, boson mapping theory has received stimulation from two sources: the challenge of accounting for the remarkable success of the phenomenological *interacting boson model* (IBM) in nuclear physics,[3] and the rapid development of the Sp($6,R$) collective model.[4]

The traditional approach to boson mappings in nuclear physics primarily utilized elementary algebraic techniques for linear vector spaces.[5] In recent years, however, the power and simplicity of Lie-algebraic techniques has increasingly been brought to bear on the problem. This has been especially important for the derivation of unitary realizations. The progress thus far has been engendered primarily through the efforts of three groups,[6–8] the last of which developed the *vector coherent-state* method into a powerful general tool for inducing matrix representations of semisimple Lie groups and their algebras from ladder representations of certain subgroups. A feature common to all of these methods is the decomposition of the algebra into a set of raising and lowering operators and a "core subalgebra," usually a maximal compact subalgebra.

In the present paper, we apply the approach of Refs. 7 to what we call the particle–hole SO($2\Omega + 1$) algebra, defined on a Fock space of Ω single-particle levels. It is well-known that the set of all bilinear combinations of fermion creation and destruction operators for Ω levels spans the algebra of SO(2Ω), which, augmented by the single-fermion operators themselves, becomes SO($2\Omega + 1$). The methods of Ref. 8 have recently been applied to the SO(2Ω) mapping in which all fermion pairs are replaced by bosons,[9] while the approach of Ref. 7 has been applied to the SO($2\Omega + 1$) mapping in which all fermion pairs are bosonized, while single-fermion degrees of freedom are mapped onto quasifermions.[10] In the particle–hole version discussed in this paper, the single-particle levels are segregated into "particle" and "hole" species, and the generators of the core subalgebra are chosen as hole or particle scattering operators. The fermion space is then mapped into what is commonly dubbed the *ideal space*, in which the degrees of freedom of the particle–hole pairs are replaced by perfect boson operators, while excess particles or holes are represented by fermionlike operators we call *quasifermions*, which commute with the bosons. This formulation provides a convenient starting point for describing the interplay between the collective excitations of a closed shell, which are bosonlike superpositions of particle–hole pair excitations, and excess valence particles or holes. In this way, one can provide a microscopic foundation for phenomenological nuclear collective models such as the particle-vibrator[11] and cluster-vibration models.[12]

The boson–quasifermion mappings of the particle–hole algebra were first derived independently by Marshalek[13] and by Geyer and Hahne[14] using traditional techniques. While both obtained the nonunitary generalized Dyson realization, Marshalek derived in addition the unitary generalized Holstein–Primakoff (GHP) realization. Afterwards, Yamamura showed that the same mappings can be derived by means of a Dirac-bracket quantization of the time-dependent Hartree–Fock self-consistent field equations.[15] The main purpose of this paper is to provide a new and, in our opinion, much simpler derivation of the unitarized (GHP) mapping along the lines of Refs. 7. An ancillary purpose is to publicize more widely the technique of realizing Lie algebras in terms of bosons and quasifermions, which thus far has been implemented only in a narrow area of nuclear physics. In addition, we also disclose details of the algebraic derivation of the corresponding Dyson mapping that were omitted in previous work. In this connection, it will be emphasized that the mapped operators need not satisfy a closed algebra, but, as in this case, may consist of a nonclosed set containing a closed subalgebra together with a system of irreducible tensors of that subalgebra. Some other examples of this already exist in the literature, all of them involving what is called a "quantized Bogoliubov–Valatin transformation."[16,17]

In Sec. II, we review the properties of the fermion particle–hole $SO(2\Omega+1)$ algebra and its fermion carrier space, and show that one may proceed through the group chain $U_h(\Omega_h)\times U_p(\Omega_p)\subset U(\Omega)\subset SO(2\Omega)\subset SO(2\Omega+1)$, the first subgroup corresponding to the core subalgebra, with Ω_h being the number of hole levels and Ω_p the number of particle levels. We then describe the nature of the ideal space and the rationale for introducing quasifermions, and, finally, derive the generalized Dyson images of elementary fermion-pair and single-fermion operators with the aid of the commutation rules. In Sec. III, we construct the *physical subspace* of the ideal space, which is the image of the fermion Hilbert space under the Dyson mapping. With the help of an identity derived from the quadratic Casimir invariants of the core subalgebra, we then prove that the physical subspace carries the spinor representations of $SO(2\Omega+1)$. In Sec. IV, we carry out the unitarization of the Dyson representation with the help of a small number of identities derived from the Casimir invariants. It is shown that in the unitary representation each generator can be represented in two forms that are equivalent within the physical subspace. A unique third form is then derived in the guise of an infinite expansion of the GHP type.

II. MAPPING FROM THE FERMION TO THE IDEAL SPACE

A. The particle–hole shell-model algebra in the fermion space

Let c_I, $c^I\equiv c_I{}^\dagger$, $I=1,\ldots,\Omega$, denote a set of Ω fermion destruction and creation operators. We partition this into a subset of Ω_h destruction and creation operators c_μ, $c^\mu\equiv c_\mu{}^\dagger$, $\mu=1,\ldots,\Omega_h$, associated with what we call *hole* states, and a set of Ω_p destruction and creation operators $c_i,c^i\equiv c_i{}^\dagger$, $i=\Omega_{h+1},\ldots,\Omega_h+\Omega_p$, associated with what we call *particle* states, with the equality $\Omega=\Omega_h+\Omega_p$. Throughout the paper, we adhere to the following conventions: greek indices denote hole states, lowercase roman indices denote particle states, and uppercase roman indices can be replaced consistently throughout an equation by either hole or particle values, i.e., there are really two equations, one for holes and one for particles. Unless stated otherwise, the summation convention for repeated indices is assumed to hold. Thus the fermion anticommutation relations are given by

$$\{c_I,c^J\}=\delta_I^J,\quad \{c^I,c^J\}=\{c_J,c_I\}=0, \tag{2.1a}$$

$$\{c_\mu,c^i\}=\{c_i,c^\mu\}=\{c^i,c^\mu\}=\{c_\mu,c_i\}=0, \tag{2.1b}$$

where $\{A,B\}$ denotes the anticommutator of A and B.

As is well-known, the set of all bilinear fermion operators

$$\{c^Ic^J,c_Jc_I,c^ic^\mu,c_\mu c_i,\tfrac{1}{2}(c^Ic_J-c_Jc^I),c^ic_\mu,c^\mu c_i\} \tag{2.2}$$

spans the Lie algebra corresponding to the group $SO(2\Omega)$, while the set obtained by adjoining the linear fermion operators $i(c_I+c^I)/2$, $i(c_I-c^I)/2$ spans the algebra of $SO(2\Omega+1)$.[18] We now introduce the following notation:

$$A^I_J\equiv c^Ic_J, \tag{2.3a}$$

$$R^{i\mu}\equiv c^ic^\mu,\quad R_{i\mu}\equiv c_\mu c_i, \tag{2.3b}$$

$$A^I\equiv c^I,\quad A_I\equiv c_I, \tag{2.3c}$$

$$R^{IJ}\equiv c^Ic^J,\quad R_{IJ}\equiv c_Jc_I,\quad A^i_\mu\equiv c^ic_\mu,\quad A^\mu_i\equiv c^\mu c_i. \tag{2.3d}$$

These operators have the following behavior under Hermitian conjugation:

$$A^J_I=A^{I\dagger}_J, \tag{2.4a}$$

$$R^{i\mu}=R_{i\mu}{}^\dagger,\quad A^I=A_I{}^\dagger,\quad R^{IJ}=R_{IJ}{}^\dagger,\quad A^\mu_i=A^{i\dagger}_\mu, \tag{2.4b}$$

as well as the antisymmetry properties

$$R^{\mu i}=-R^{i\mu}, \tag{2.5a}$$

$$R^{JI}=-R^{IJ},\quad R_{JI}=-R_{IJ}. \tag{2.5b}$$

The set of operators (2.3a) consists of the subsets A^μ_ν and A^i_j, the former generating the unitary subalgebra $U_h(\Omega_h)$ based on the hole levels and the latter the unitary subalgebra $U_p(\Omega_p)$ based on the particle levels. The commutation rules, which follow from (2.1), may be summarized by

$$\begin{aligned}[][A^I_J,A^K_L]&=\delta^K_JA^I_L-\delta^I_LA^K_J,\\ [A^i_j,A^\mu_\nu]&=0,\end{aligned} \tag{2.6a}$$

which is the subalgebra of the group $U_h(\Omega_h)\times U_p(\Omega_p)$. It is this subalgebra that will be chosen as the core subalgebra in the sense of Refs. 8. The remaining generators in (2.3) are the ladder operators, consisting of the particle–hole creation and destruction operators (2.3b), the one-particle transfer operators (2.3c), and the two-particle transfer operators (2.3d). Of the remaining commutation relations, we list explicitly those that will be needed for later developments. First, there are the commutators of the particle–hole operators with the generators of the core subalgebra given by

$$[A^\lambda_\nu,R^{i\mu}]=\delta^\mu_\nu R^{i\lambda},\quad [R_{i\mu},A^\nu_\lambda]=\delta^\nu_\mu R_{i\lambda}, \tag{2.6b}$$

$$[A^k_j,R^{i\mu}]=\delta^i_jR^{k\mu},\quad [R_{i\mu},A^j_k]=\delta^j_iR_{k\mu}, \tag{2.6c}$$

and then the mutual commutators of the particle–hole operators, given by

$$[R^{i\mu},R^{j\nu}]=[R_{j\nu},R_{i\mu}]=0, \tag{2.6d}$$

$$[R_{j\nu},R^{i\mu}]=\delta^i_j\delta^\mu_\nu-\delta^i_jA^\mu_\nu-\delta^\mu_\nu A^i_j. \tag{2.6e}$$

The commutators (2.6a)–(2.6e) constitute a closed subalgebra equivalent to that of the group $U(\Omega)$ (to obtain a Lie algebra one should use $A^I_J-\frac{1}{2}\delta^I_J$ as generator in place of A^I_J). This is the shell-model algebra for describing the properties of a closed-shell nucleus. Next, we add the pair-transfer operators (2.3d), which, together with the generators of

U(Ω), form the subalgebra SO(2Ω). Their commutators with the U(Ω) generators are given by

$$[A^K_L,R^{IJ}]=\delta^J_L R^{IK}+\delta^I_L R^{KJ},$$
$$[R_{IJ},A^L_K]=\delta^L_J R_{IK}+\delta^L_I R_{KJ},$$
$$[A^\mu_\nu,R^{ij}]=[R_{ij},A^\nu_\mu]=[A^i_j,R^{\mu\nu}]=[R_{\mu\nu},A^j_i]=0,$$
$$[A^\kappa_\lambda,A^i_\mu]=-\delta^\kappa_\mu A^i_\lambda,\qquad [A^\mu_i,A^\lambda_\kappa]=-\delta^\mu_\kappa A^\lambda_i,$$
$$[A^j_k,A^i_\mu]=\delta^i_k A^j_\mu,\quad [A^\mu_i,A^k_j]=\delta^k_i A^\mu_j; \tag{2.6f}$$

$$[R_{i\lambda},R^{\mu\nu}]=\delta^\mu_\lambda A^\nu_i-\delta^\nu_\lambda A^\mu_i,$$
$$[R_{\mu\nu},R^{i\lambda}]=\delta^\lambda_\mu A^i_\nu-\delta^\lambda_\nu A^i_\mu,$$
$$[R_{k\mu},R^{ij}]=\delta^j_k A^i_\mu-\delta^i_k A^j_\mu,$$
$$[R_{ij},R^{k\mu}]=\delta^k_j A^\mu_i-\delta^k_i A^\mu_j,$$
$$[R^{i\lambda},R^{\mu\nu}]=[R_{\mu\nu},R_{i\lambda}]=[R^{k\mu},R^{ij}]=[R_{ij},R_{k\mu}]=0,$$
$$[R^{j\nu},A^i_\mu]=\delta^\nu_\mu R^{ij},\quad [A^\mu_i,R_{j\nu}]=\delta^\mu_\nu R_{ij}. \tag{2.6g}$$

The *nonvanishing* mutual commutators of the pair-transfer operators are given by

$$[R_{IJ},R^{KL}]=\delta^K_I\delta^L_J-\delta^K_J\delta^L_I+\delta^K_J A^L_I+\delta^L_I A^K_J$$
$$-\delta^K_I A^L_J-\delta^L_J A^K_I,$$
$$[A^i_\mu,A^\nu_j]=\delta^\nu_\mu A^i_j-\delta^i_j A^\nu_\mu,$$
$$[A^i_\mu,R^{\nu\lambda}]=\delta^\nu_\mu R^{i\lambda}-\delta^\lambda_\mu R^{i\nu}, \tag{2.6h}$$
$$[R_{\nu\lambda},A^\mu_i]=\delta^\mu_\nu R_{i\lambda}-\delta^\mu_\lambda R_{i\nu},$$
$$[A^\mu_i,R^{jk}]=\delta^k_i R^{j\mu}-\delta^j_i R^{k\mu},$$
$$[R_{jk},A^i_\mu]=\delta^i_k R_{j\mu}-\delta^i_j R_{k\mu}$$

(all other commutators from the set $R^{IJ},R_{IJ},A^i_\mu,A^\mu_i$ vanish).

Finally, the SO(2Ω) algebra is extended to SO(2Ω + 1) by including the single-fermion operators, whose commutators with the U(Ω) generators are given by

$$[A^I_J,A^K]=\delta^K_J A^I,\quad [A_K,A^J_I]=\delta^J_K A_I,$$
$$[A^i_j,A^\mu]=[A_\mu,A^j_i]=[A^\mu_\nu,A^i]=[A_i,A^\nu_\mu]=0; \tag{2.6i}$$

$$[R_{i\nu},A^\mu]=-\delta^\mu_\nu A_i,\quad [A_\mu,R^{i\nu}]=-\delta^\nu_\mu A^i,$$
$$[R_{j\mu},A^i]=\delta^i_j A_\mu,\quad [A_i,R^{j\mu}]=\delta^j_i A^\mu, \tag{2.6j}$$
$$[R^{i\mu},A^I]=0,\quad [A_I,R_{i\mu}]=0.$$

The nonvanishing commutators of the one-particle transfer operators with the two-particle transfer operators are the following:

$$[A^I,R_{JK}]=\delta^I_K A_J-\delta^I_J A_K,\quad [R^{JK},A_I]=\delta^K_I A^J-\delta^J_I A^K,$$
$$[A^i_\nu,A^\mu]=\delta^\mu_\nu A^i,\quad [A_\mu,A^\nu_i]=\delta^\nu_\mu A_i, \tag{2.6k}$$
$$[A^\mu_j,A^i]=\delta^i_j A^\mu,\quad [A_i,A^j_\mu]=\delta^j_i A_\mu$$

(all other commutators of the A^I,A_I with R^{JK},R_{JK},A^i_μ and A^μ_i vanish).

To complete the algebra, we give the mutual commutators of the one-particle transfer operators:

$$[A_J,A^I]=\delta^I_J-2A^I_J; \tag{2.6l}$$
$$[A^i,A^\mu]=2R^{i\mu},\quad [A_\mu,A_i]=2R_{i\mu}; \tag{2.6m}$$
$$[A^I,A^J]=2R^{IJ},\quad [A_J,A_I]=2R_{IJ},$$
$$[A^\mu,A_i]=2A^\mu_i,\quad [A^i,A_\mu]=2A^i_\mu. \tag{2.6n}$$

Note that the commutation relations (2.6l)–(2.6n) are just equivalent to the fermion anticommutation rules (2.1).

An orthonormal basis for the 2^Ω-dimensional fermion vector space $\mathfrak{H}_F$ is provided by the set consisting of the normalized vacuum $|0\rangle$ (the closed-shell system), satisfying

$$c_i|0\rangle=c_\mu|0\rangle=0, \tag{2.7}$$

for all i and μ, together with the vectors

$$|i_1\mu_1\cdots i_{N_B}\mu_{N_B}\rangle=\prod_{n=1}^{N_B}c^{i_n}c^{\mu_n}|0\rangle, \tag{2.8a}$$

$$|\nu_1\cdots\nu_{n_h}i_1\mu_1\cdots i_{N_B}\mu_{N_B}\rangle=\prod_{n=1}^{n_h}c^{\nu_n}|i_1\mu_1\cdots i_{N_B}\mu_{N_B}\rangle, \tag{2.8b}$$

$$|j_1\cdots j_{n_p}i_1\mu_1\cdots i_{N_B}\mu_{N_B}\rangle=\prod_{n=1}^{n_p}c^{j_n}|i_1\mu_1\cdots i_{N_B}\mu_{N_B}\rangle, \tag{2.8c}$$

which span the subspaces with an equal number N_B of particles and holes, and those with an excess n_h of holes and n_p of particles. In Eqs. (2.8b) and (2.8c) it is understood that in the case $N_B=0$, the ket on the right-hand side (rhs) becomes the vacuum $|0\rangle$. The fermion space $\mathfrak{H}_F$ carries the solitary spinor irreducible representation (irrep) of SO(2Ω + 1), while the subspaces with even and odd particle numbers separately carry the two $2^{\Omega-1}$-dimensional spinor irreps of the subgroup SO(2Ω). The fermion space also carries Ω + 1 antisymmetric irreps of the subgroup U(Ω), each of which may be labeled by N_p-N_h, the difference in the number of particles and holes (see below). These irreps may be further decomposed into a total of $(\Omega_h+1)(\Omega_p+1)$ antisymmetric irreps of the subgroup $U_h(\Omega_h)\times U_p(\Omega_p)$, each labeled by the number of particles and the number of holes (see below).

Of central importance in our later analysis are the Casimir invariants of $U_h(\Omega_h)$ and $U_p(\Omega_p)$. In terms of the generators, the corresponding Casimir operators of order k are given by[19]

$$\mathscr{C}_h^{(k)}=A^{\mu_1}_{\mu_2}A^{\mu_2}_{\mu_3}\cdots A^{\mu_k}_{\mu_1},\quad \mathscr{C}_p^{(k)}=A^{i_1}_{i_2}A^{i_2}_{i_3}\cdots A^{i_k}_{i_1}. \tag{2.9}$$

Upon inserting the fermion realization (2.3a) and performing a trivial rearrangement, one obtains for (2.9) the diagonal forms

$$\mathscr{C}_h^{(k)}=\hat N_h(\Omega_h-\hat N_h+1)^{k-1},$$
$$\mathscr{C}_p^{(k)}=\hat N_p(\Omega_p-\hat N_p+1)^{k-1}, \tag{2.10}$$

where $\hat N_h$ and $\hat N_p$ are the hole and particle number operators

$$\hat N_h=c^\mu c_\mu,\quad \hat N_p=c^i c_i. \tag{2.11}$$

Each of the number operators (2.11) commutes with all the generators of $U_h(\Omega_h)\times U_p(\Omega_p)$, thereby providing labels for the antisymmetric irreps. The difference $\hat{N}_p - \hat{N}_h$ commutes with all the generators of $U(\Omega)$, thereby providing a label for its irreps.

We also note that the quadratic Casimir operator for $SO(2\Omega)$ is given by

$$C_{SO(2\Omega)} = [A^I_J - \tfrac{1}{2}\delta^I_J][A^J_I - \tfrac{1}{2}\delta^J_I] + \{A^i_\mu, A^\mu_i\} + \{R^{i\mu},R_{i\mu}\} + \tfrac{1}{2}\{R^{IJ},R_{IJ}\}, \tag{2.12}$$

while that for $SO(2\Omega+1)$ is given by[18]

$$C_{SO(2\Omega+1)} = C_{SO(2\Omega)} + \tfrac{1}{2}\{A^I,A_I\}. \tag{2.13}$$

For the fermion realization (2.3), these reduce to pure numbers,

$$C_{SO(2\Omega)} = \tfrac{1}{2}\Omega(\Omega - \tfrac{1}{2}), \quad C_{SO(2\Omega+1)} = \tfrac{1}{2}\Omega(\Omega + \tfrac{1}{2}), \tag{2.14}$$

which are just the eigenvalues for the spinor irreps.

B. The nature of the ideal space

As mentioned earlier, our aim is to map the algebra discussed in the previous section together with the finite-dimensional fermion carrier space $\mathfrak{H}_F$ into a subspace of a Hilbert space, historically referred to as the *ideal space*,[20] and denoted here by $\mathfrak{J}$. The space $\mathfrak{J}$ is generated by the boson destruction and creation operators $B_{i\mu}$, $B^{i\mu}\equiv B_{i\mu}{}^\dagger$, respectively, together with the set of what we call *quasifermion* destruction and annihilation operators, a_μ, $a^\mu\equiv a_\mu{}^\dagger$, associated with hole levels, and $a_i, a^i\equiv a_i{}^\dagger$, associated with particle levels. The bosons $B_{i\mu}$ and $B^{i\mu}$, which replace the degrees of freedom of particle–hole pairs $c_\mu c_i$ and $c^i c^\mu$, respectively, obey the Heisenberg–Weyl algebra

$$[B_{j\nu},B^{i\mu}] = \delta^i_j\delta^\mu_\nu, \qquad [B^{i\mu},B^{j\nu}] = [B_{j\nu},B_{i\mu}] = 0. \tag{2.15}$$

The quasifermions, which are to represent *valence* particles or holes, are assumed to commute with all the boson operators,

$$[a^I,B_{i\mu}] = [B^{i\mu},a_I] = [a^I,B^{i\mu}] = [B_{i\mu},a_I] = 0, \tag{2.16}$$

and also obey an algebra to be discussed presently.

A possible definition of the ideal space, which has been used in previous work,[14,20] is $\mathfrak{J} = \mathfrak{H}_F \otimes \mathfrak{H}_B$, the tensor product of the fermion space discussed in Sec. II A, with $\mathfrak{H}_B$, the boson space generated by the $B_{i\mu}, B^{i\mu}$ acting on the boson vacuum. In such a space, a particle–hole excitation could be redundantly represented by either $B^{i\mu}$ or $c^i c^\mu$. Since the aim of the mapping is to replace the fermion particle–hole pairs by the bosons, the *physical subspace*, i.e., the replica of $\mathfrak{H}_F$ in $\mathfrak{J}$, must be chosen so that $c_\mu c_i = 0$ in this subspace. In Ref. 14, this is achieved by banishing all vectors containing one or more fermion particle–hole pairs to the *unphysical subspace*, defined as the orthogonal complement of the physical subspace. However, one can also take the somewhat different viewpoint of Ref. 13 discussed below. As a prelude, we define the projectors Q_h and Q_p by

$$Q_h \equiv 1 - c^\mu(1+\hat{N}_h)^{-1}c_\mu, \quad Q_p \equiv 1 - c^i(1+\hat{N}_p)^{-1}c_i \tag{2.17}$$

(summation convention!), where $\hat{N}_h$ and $\hat{N}_p$ are the hole and particle number operators (2.11). The Hermitian operators (2.17) satisfy

$$Q_h{}^2 = Q_h, \quad Q_p{}^2 = Q_p, \quad Q_hQ_p = Q_pQ_h; \tag{2.18}$$

Q_h is the projector to the subspace of $\mathfrak{J}$ having *zero holes*, while Q_p is the projector to the subspace having *zero particles*. Now, a disadvantage of the fermion operators c_I, c^I is that they can connect physical and unphysical vectors. However, such is not the case with the operators a_I, a^I defined by

$$a^\mu \equiv c^\mu Q_p = Q_p c^\mu, \quad a_\mu \equiv Q_p c_\mu = c_\mu Q_p, \\ a^i \equiv c^i Q_h = Q_h c^i, \quad a_i \equiv Q_h c_i = c_i Q_h. \tag{2.19}$$

From the property

$$Q_h c^\mu = c_\mu Q_h = 0, \quad Q_p c^i = c_i Q_p = 0, \tag{2.20}$$

it follows immediately that

$$a_i a_\mu = a_\mu a_i = a^i a^\mu = a^\mu a^i = a_i a^\mu = a_\mu a^i = 0. \tag{2.21}$$

Thus, when acting in the physical subspace, the operators a_I, a^I behave like c_I, c^I, respectively, except that the former do not have the undesirable property of connecting physical and unphysical vectors. The operators (2.19), which are the restrictions of the fermion operators to a subspace containing no fermion particle–hole pairs, are examples of what we call *quasifermion* operators. They do not obey the usual fermion anticommutation rules but rather the following anticommutation rules, which are readily obtained from Eqs. (2.1), (2.18), and (2.19),

$$\{a_\nu,a^\mu\} = \delta^\mu_\nu Q_p, \quad \{a_j,a^i\} = \delta^i_j Q_h, \tag{2.22a}$$

$$\{a^I,a^J\} = \{a_J,a_I\} = 0. \tag{2.22b}$$

Equations (2.21) and (2.22) imply that the quasihole operators a^μ, a_μ behave like ordinary fermions in the subspace having zero fermion particles, as do the quasiparticle operators a^i, a_i in the subspace having zero fermion holes.

If one adopts the boson–fermion definition of the ideal space and writes all mapped operators as functions of boson and fermion operators, then it is necessary to append the projectors Q_h and Q_p as indicated in (2.19) to obtain images that leave the physical subspace invariant. That is the approach taken in Ref. 14. Alternatively, as in Ref. 13, one can simply replace the fermions by quasifermions, thereby avoiding the need to carry along the projectors explicitly. Indeed, one can push this a little further and define the ideal space as $\mathfrak{J} = \mathfrak{H}_{QF} \otimes \mathfrak{H}_B$, where $\mathfrak{H}_{QF}$ is the vector space generated by the quasifermion operators and thus *ipso facto* cannot contain any particle–hole pairs. This definition obviates the need to relegate vectors containing particle–hole pairs to the unphysical subspace. The quasifermions can then be regarded as fundamental entities postulated to obey an abstract algebra consisting of the anticommutation rules (2.22), where Q_h and Q_p are commuting projectors [as in Eq. (2.18)] defined with the following properties:

$$Q_h a^\mu = a_\mu Q_h = 0, \quad Q_p a^i = a_i Q_p = 0,$$
$$Q_p a^\mu = a^\mu Q_p = a^\mu, \quad a_\mu Q_p = Q_p a_\mu = a_\mu, \tag{2.23}$$
$$Q_h a^i = a^i Q_h = a^i, \quad a_i Q_h = Q_h a_i = a_i,$$

which, incidentally, are entirely compatible with the fermion interpretation (2.19). The postulated properties are then easily shown to imply Eqs. (2.21). It is this second interpretation that we prefer to adopt throughout most of the paper for reasons of economy, although the reader is free to resurrect at any time the underlying fermion interpretation of the quasifermions if desired. In connection with the second interpretation, we have three remaining comments. First, the projectors Q_h and Q_p should now be represented as follows:

$$Q_h = 1 - a^\mu (1+\hat{n}_h)^{-1} a_\mu = 1 - a^\mu (1+\hat{n})^{-1} a_\mu,$$
$$Q_p = 1 - a^i (1+\hat{n}_p)^{-1} a_i = 1 - a^i (1+\hat{n})^{-1} a_i, \tag{2.24}$$

where $\hat{n}_h$ and $\hat{n}_p$ are the quasihole and quasiparticle number operators, while $\hat{n}$ is their sum:

$$\hat{n}_h \equiv a^\mu a_\mu, \quad \hat{n}_p \equiv a^i a_i, \quad \hat{n} \equiv \hat{n}_h + \hat{n}_p. \tag{2.25}$$

In the second form of Eqs. (2.24), the identities

$$a^\mu f(\hat{n}_h) = a^\mu f(\hat{n}), \quad a^i f(\hat{n}_p) = a^i f(\hat{n}),$$
$$f(\hat{n}_h) a_\mu = f(\hat{n}) a_\mu, \quad f(\hat{n}_p) a_i = f(\hat{n}) a_i, \tag{2.26}$$

valid for an arbitrary function f, were invoked. The validity of (2.26) is obvious from the constraint (2.21). It is straightforward to check that the representation (2.24) is compatible with the properties (2.23).

The second comment is that, as is easily checked, the quasifermion algebra preserves the following properties associated with the fermion algebra:

$$[a^I a_J, a^K a_L] = \delta^K_J a^I a_L - \delta^I_L a^K a_J, \tag{2.27a}$$
$$[a^I a_J, a^K] = \delta^K_J a^I, \quad [a_K, a^J a_I] = \delta^J_K a_I. \tag{2.27b}$$

Finally, in the ideal space $\mathfrak{I} = \mathfrak{H}_{\mathrm{QF}} \otimes \mathfrak{H}_{\mathrm{B}}$, one may introduce the orthonormal basis spanned by the vectors

$$\prod_I (a^I)^{n_I} \prod_{i\mu} (n_{i\mu}!)^{-1/2} (B^{i\mu})^{n_{i\mu}} |0), \tag{2.28}$$

where $|0)$ is the normalized vacuum state (tensor product of the quasifermion and boson vacua), satisfying

$$a_\mu |0) = a_i |0) = B_{i\mu} |0) = 0. \tag{2.29}$$

Later, it will be shown that the physical subspace of $\mathfrak{I}$ is spanned by antisymmetric combinations of the vectors (2.28).

C. The generalized Dyson mapping

Our ultimate aim is to find an injective linear mapping $\mathfrak{H}_{\mathrm{F}} \to \mathfrak{I}_p$, where $\mathfrak{I}_P \subset \mathfrak{I}$ denotes the 2^Ω-dimensional physical subspace, which is the carrier of the spinor irrep of $\mathrm{SO}(2\Omega+1)$. Thus the mapping of the fermion generators onto operators in $\mathfrak{I}$ must preserve the algebra (2.6) in the physical subspace. The generalized Dyson mapping is a particular example of such a mapping, but one that does not preserve the property of Hermitian conjugation, and, therefore, provides a nonunitary representation. As we shall show, in practice it is only necessary to satisfy a certain subset of the commutation relations (2.6) over the whole ideal space (i.e., on a dense subset of $\mathfrak{I}$). Along with the requirement that the fermion vacuum be mapped into the vacuum of the ideal space, $|0\rangle \to |0)$, this turns out to be sufficient to define the physical subspace $\mathfrak{I}_P$ and the mapping of all operators. It will then be shown that the remaining commutation relations (2.6) are automatically fulfilled in $\mathfrak{I}_P$.

For any fermion operator F, let $(F)_{\mathrm{D}}$ denote the image under the Dyson mapping. It is convenient to adopt for the Dyson images of the generators the same notation as in Eq. (2.3), i.e., henceforth let

$$A^I_J \equiv (c^I c_J)_{\mathrm{D}}, \tag{2.30a}$$
$$R^{i\mu} \equiv (c^i c^\mu)_{\mathrm{D}}, \quad R_{i\mu} \equiv (c_\mu c_i)_{\mathrm{D}}, \tag{2.30b}$$
$$A^I \equiv (c^I)_{\mathrm{D}}, \quad A_I \equiv (c_I)_{\mathrm{D}}, \tag{2.30c}$$
$$R^{IJ} \equiv (c^I c^J)_{\mathrm{D}}, \quad R_{IJ} \equiv (c_J c_I)_{\mathrm{D}},$$
$$A^i_\mu \equiv (c^i c_\mu)_D, \quad A^\mu_i \equiv (c^\mu c_i)_{\mathrm{D}}. \tag{2.30d}$$

The first task is to find the mapping of the $\mathrm{U}_h(\Omega_h) \times \mathrm{U}_p(\Omega_p)$ core subalgebra (2.6a). This can be accomplished with the aid of the additional requirement that under the core subalgebra the boson operators $B^{i\mu}, B_{i\mu}$ transform like the particle–hole operators (2.30b) and the quasifermion operators a^I, a_I transform like the one-particle transfer operators (2.30c). That is, in accord with Eqs. (2.6b), (2.6c), and (2.6i) it is required that

$$[A^\lambda_\nu, B^{i\mu}] = \delta^\mu_\nu B^{i\lambda}, \quad [B_{i\mu}, A^\nu_\lambda] = \delta^\nu_\mu B_{i\lambda}, \tag{2.31a}$$
$$[A^k_j, B^{i\mu}] = \delta^i_j B^{k\mu}, \quad [B_{i\mu}, A^j_k] = \delta^j_i B_{k\mu}, \tag{2.31b}$$
$$[A^I_J, a^K] = \delta^K_J a^I, \quad [a_K, A^J_I] = \delta^J_K a_I, \tag{2.31c}$$
$$[A^j_i, a^\mu] = [a_\mu, A^j_i] = [A^\mu_\nu, a^i] = [a_i, A^\nu_\mu] = 0. \tag{2.31d}$$

We note that a similar requirement is common to the methods of Refs. 6–8. From Eqs. (2.15) and (2.27), the solution of Eqs. (2.31) is *uniquely* given by

$$A^\mu_\nu = B^{i\mu} B_{i\nu} + a^\mu a_\nu, \quad A^i_j = B^{i\mu} B_{j\mu} + a^i a_j, \tag{2.32}$$

which also satisfies the subalgebra (2.6a), the Hermitian conjugation (2.4a), and, from (2.29), the condition

$$(c^I c_J)_{\mathrm{D}} |0) = A^I_J |0) = 0, \tag{2.33}$$

which is essential for the correspondence of the fermion and ideal-space vacuum states.

Before proceeding to the construction of the particle–hole operators (2.30b), we digress briefly to consider the Casimir invariants of the core subalgebra in the ideal space. These are obtained by substituting the generators (2.32) into Eqs. (2.9). These Casimir invariants, unlike their counterparts in the fermion space, are not in general diagonal, but, as will be seen later, their restrictions to the physical subspace are diagonal. The quadratic Casimir operators, which are of particular importance, are given by

$$\mathscr{C}_h^{(2)} = \Omega_h (\hat{N}_{\mathrm{B}} + \hat{n}_h) + \overline{\mathscr{C}}_h^{(2)},$$
$$\mathscr{C}_p^{(2)} = \Omega_p (\hat{N}_{\mathrm{B}} + \hat{n}_p) + \overline{\mathscr{C}}_p^{(2)}, \tag{2.34}$$

where $\hat{N}_{\mathrm{B}}$ is the boson number operator

$$\hat{N}_{\mathrm{B}} = B^{i\mu} B_{i\mu}, \tag{2.35}$$

and $\overline{\mathscr{C}}_h{}^{(2)}$, $\overline{\mathscr{C}}_p{}^{(2)}$ are the two-body parts of the respective operators, given by

$$\overline{\mathscr{C}}_h{}^{(2)} = \overline{\mathscr{C}}_{\mathrm{B}}{}^{(2)} + a^\mu a^\nu a_\mu a_\nu + 2B^{i\mu}B_{i\nu}a^\nu a_\mu ,$$
$$\overline{\mathscr{C}}_p{}^{(2)} = \overline{\mathscr{C}}_{\mathrm{B}}{}^{(2)} + a^i a^j a_i a_j + 2B^{i\mu}B_{j\mu}a^j a_i , \tag{2.36}$$

with $\overline{\mathscr{C}}_{\mathrm{B}}{}^{(2)}$, the common boson part, given by

$$\overline{\mathscr{C}}_{\mathrm{B}}{}^{(2)} = B^{i\mu}B^{j\nu}B_{j\mu}B_{i\nu}. \tag{2.37}$$

Given the realization (2.32) of the core subalgebra, the images of the particle–hole pairs (2.30b) are required to satisfy the commutation rules (2.6b)–(2.6e). The relations (2.6b) and (2.6c) only require $R^{i\mu}$ and $R_{i\mu}$ to be irreducible tensors under $\mathrm{U}_h(\Omega_h)\times\mathrm{U}_p(\Omega_p)$ that transform like $B^{i\mu}$ and $B_{i\mu}$, respectively. Such tensors can be systematically generated by the well-known procedure of calculating commutators of $B^{i\mu}$ and $B_{i\mu}$ with the Casimir operators or combinations thereof. Thus, for example, from Eqs. (2.36) and (2.37) one obtains the following additional tensors with the required transformation properties:

$$\tfrac{1}{2}[B^{i\mu},\overline{\mathscr{C}}_{\mathrm{B}}{}^{(2)}] = -B^{i\nu}B^{j\mu}B_{j\nu}, \tag{2.38a}$$

$$\tfrac{1}{2}[B^{i\mu},\overline{\mathscr{C}}_h{}^{(2)} - \overline{\mathscr{C}}_{\mathrm{B}}{}^{(2)}] = -B^{i\nu}a^\mu a_\nu,$$
$$\tfrac{1}{2}[B^{i\mu},\overline{\mathscr{C}}_p{}^{(2)} - \overline{\mathscr{C}}_{\mathrm{B}}{}^{(2)}] = -B^{j\mu}a^i a_j. \tag{2.38b}$$

While it is possible to satisfy the commutation relations (2.6d) and (2.6e) with an infinite sum of such tensors, the situation is greatly simplified if one observes from the boson commutation rules (2.15) and the sum of Eqs. (2.38) that

$$[B_{j\nu},B^{i\mu} + \tfrac{1}{2}[B^{i\mu},\overline{\mathscr{C}}_h{}^{(2)} + \overline{\mathscr{C}}_p{}^{(2)} - \overline{\mathscr{C}}_{\mathrm{B}}{}^{(2)}]]$$
$$= \delta^i_j\delta^\mu_\nu - \delta^i_j A^\mu_\nu - \delta^\mu_\nu A^i_j. \tag{2.39}$$

The rhs of (2.39) is identical to that of Eq. (2.6e). Therefore, the commutation relation (2.6e), as well as (2.6d), can be satisfied by choosing

$$R_{i\mu} = B_{i\mu}, \tag{2.40a}$$

$$R^{i\mu} = B^{i\mu} + \tfrac{1}{2}[B^{i\mu},\overline{\mathscr{C}}_h{}^{(2)} + \overline{\mathscr{C}}_p{}^{(2)} - \overline{\mathscr{C}}_{\mathrm{B}}{}^{(2)}]$$
$$= B^{i\mu} - B^{i\nu}B^{j\mu}B_{j\nu} - B^{i\nu}a^\mu a_\nu - B^{j\mu}a^i a_j. \tag{2.40b}$$

To be sure, the choice (2.40) fails to satisfy the condition that $R_{i\mu}$ and $R^{i\mu}$ be Hermitian conjugates [Eq. (2.4b)], but that only means that the realization is nonunitary; it can and will be unitarized by a similarity transformation later. It may be worthwile to point out that a Hermitian conjugate (H.c.) realization also exists, in which $R^{i\mu}$ is represented by the boson creation operator $B^{i\mu}$, while $R_{i\mu}$ is represented by the H.c. of the rhs of (2.40b) [for physical applications, the choice (2.40) is more convenient]. A nonunitary realization of a Lie algebra in which either the raising or lowering operators are chosen as perfect boson operators is what is usually called a generalized Dyson representation.[5] We also should point out that one is free to multiply the rhs of (2.40a) by an arbitary scale factor while multiplying that of (2.40b) by its inverse, without disturbing the commutation rule (2.6e). The advantage of the choice (2.40) is that it normalizes the image vector of a one-particle–one-hole state: $(c^i c^\mu)_{\mathrm{D}}|0) = R^{i\mu}|0) = B^{i\mu}|0)$. At this point, we have obtained a realization of the subalgebra $\mathrm{U}(\Omega)$, valid over the full ideal space.

We can now discuss the mapping of the single-fermion operators (2.30c), which is determined by the commutation rules (2.6i) and (2.6j). According to the former, A^I and A_I must transform like c^I and c_I, respectively, under the core subalgebra. The simplest such operators are a^μ and $B^{i\mu}a_i$, which transform like c^μ, and a^i and $B^{i\mu}a_\mu$, which transform like c^i, as well as the corresponding H.c. operators. More complex operators of this type can be constructed by taking commutators with Casimir invariants, as, for example,

$$\tfrac{1}{2}[a^\mu,\overline{\mathscr{C}}_h{}^{(2)}] = -a^\nu A^\mu_\nu,\quad \tfrac{1}{2}[a^i,\overline{\mathscr{C}}_p{}^{(2)}] = -a^j A^i_j. \tag{2.41}$$

Without loss of generality, one may write

$$A_\mu = \varphi^\nu_\mu a_\nu - a^i B_{i\nu}\gamma^\nu_\mu,\quad A_i = \varphi^j_i a_j + a^\mu B_{j\mu}\gamma_i{}^j, \tag{2.42a}$$

$$A^\mu = a^\nu\Phi^\mu_\nu - \Gamma^\mu_\nu B^{i\nu}a_i,\quad A^i = a^j\Phi^i_j + \Gamma^i_j B^{j\mu}a_\mu, \tag{2.42b}$$

where the operator coefficients φ^I_J,γ^I_J, Φ^I_J, and Γ^I_J, denoted collectively by V^I_J, are each required by Eqs. (2.6i) to satisfy

$$[A^K_L,V^I_J] = \delta^I_L V^K_J - \delta^K_J V^I_L,$$
$$[A^i_j,V^\mu_\nu] = [A^\mu_\nu,V_j{}^i] = 0. \tag{2.43}$$

Thus Eqs. (2.6i) require that these coefficients transform under the core subalgebra like the generators themselves, i.e., they are vector operators in the terminology of Okubo.[21]

More stringent conditions on the coefficients are provided by Eqs. (2.6j), which, from Eqs. (2.40), are given by

$$[B_{i\nu},A_\mu] = 0,\quad [B_{j\mu},A_i] = 0; \tag{2.44a}$$

$$[B_{i\nu},A^\mu] = -\delta^\mu_\nu A_i,\quad [B_{j\mu},A^i] = \delta^i_j A_\mu; \tag{2.44b}$$

$$[A_\mu,B^{i\nu} - B^{i\lambda}B^{j\nu}B_{j\lambda} - B^{i\lambda}a^\nu a_\lambda - B^{j\nu}a^i a_j] = -\delta^\nu_\mu A^i,$$
$$[A_i,B^{j\mu} - B^{j\nu}B^{k\mu}B_{k\nu} - B^{j\nu}a^\mu a_\nu - B^{k\mu}a^j a_k] = \delta^j_i A_\mu; \tag{2.44c}$$

$$[A^\mu,B^{i\nu} - B^{i\lambda}B^{j\nu}B_{j\lambda} - B^{i\lambda}a^\nu a_\lambda - B^{j\nu}a^i a_j] = 0,$$
$$[A^i,B^{j\mu} - B^{j\nu}B^{k\mu}B_{k\nu} - B^{j\nu}a^\mu a_\nu - B^{k\mu}a^j a_k] = 0. \tag{2.44d}$$

First of all, Eqs. (2.44a) require that

$$[B_{k\mu},\varphi^I_J] = 0,\quad [B_{k\mu},\gamma^I_J] = 0. \tag{2.45}$$

This means that the coefficients φ^I_J and γ^I_J must be independent of the boson operators; they are vector operators constructed solely from the quasifermions. The task at hand, then, is to deduce the most general form of these vector operators. Now, the vector operators ρ^I_J defined by

$$\rho^I_J \equiv a^I a_J \tag{2.46}$$

generate the quasihole unitary group $u_{qh}(\Omega_h)$ (greek indices) and the quasiparticle unitary group $u_{qp}(\Omega_p)$ (roman indices). Since the quasifermions obey the commutators (2.27), these unitary groups have properties analogous to the corresponding fermion unitary groups. This has the following two implications. First, the Casimir invariants of the quasifermion groups are given by a formula analogous to (2.10); i.e., the quasifermion Casimir invariants are functions of the number operators $\hat{n}_h$ or $\hat{n}_p$ [Eq. (2.25)]. Second, as proved by Klein[22] and Okubo,[21] an arbitrary (quasifermion) vector operator can be written as a (finite) linear combination of the operators $(\rho^n)^I_J$, defined recursively by

$$(\rho^0)^I_J = \delta^I_J, \quad (\rho^n)^I_J = (\rho^{n-1})^K_J \rho^I_K \qquad (n = 1,\ldots,\Omega_h \text{ or } \Omega_p). \tag{2.47}$$

It is straightforward to prove inductively that ρ^n can be rearranged into the simple form

$$(\rho^n)^\mu_\nu = f_n(\hat{n}_h)\delta^\mu_\nu + g_n(\hat{n}_h)a^\mu a_\nu, \qquad (\rho^n)^i_j = f_n(\hat{n}_p)\delta^i_j + g_n(\hat{n}_p)a^i a_j, \tag{2.48}$$

where f_n and $g_n = 1 - f_n$ are certain polynomials in the number operators that need not be further specified. Thus an arbitrary vector operator V^I_J is a linear combination of the operators (2.48), with coefficients that are functions of the number operators $\hat{n}_h$ and $\hat{n}_p$, and, therefore, has the form

$$V^I_J = x(\hat{n}_h,\hat{n}_p)\delta^I_J + y(\hat{n}_h,\hat{n}_p)a^I a_J, \tag{2.49}$$

where x and y are arbitrary functions of the number operators. In applying Eq. (2.49) to the coefficients φ^I_J and γ^I_J of Eqs. (2.42a), two additional simplifications occur in the sums. An example of the first is as follows:

$$\begin{aligned}\varphi^\nu_\mu a_\nu &= x(\hat{n}_h,\hat{n}_p)a_\mu + y(\hat{n}_h,\hat{n}_p)a^\nu a_\mu a_\nu \\ &= x(\hat{n}_h,\hat{n}_p)a_\mu - y(\hat{n}_h,\hat{n}_p)\hat{n}_h a_\mu \\ &= x(\hat{n}_h,0)a_\mu - y(\hat{n}_h,0)\hat{n}_h a_\mu \equiv f(\hat{n}_h)a_\mu,\end{aligned} \tag{2.50}$$

where $f(\hat{n}_h)$, being a combination of arbitrary functions, is also arbitrary. In the third step, use has been made of the condition that $\mathfrak{J}$ contains no $a^i a^\mu$ pairs [Eq. (2.21)]. The same condition can also be used to obtain the following simplification:

$$\begin{aligned}a^i\gamma^\nu_\mu &= a^i[x(\hat{n}_h,\hat{n}_p)\delta^\nu_\mu + y(\hat{n}_h,\hat{n}_p)a^\nu a_\mu] \\ &= a^i x(0,\hat{n}_p)\delta^\nu_\mu \equiv a^i g(\hat{n}_p)\delta^\nu_\mu,\end{aligned} \tag{2.51}$$

where $g(\hat{n}_p)$ is an arbitrary function. In this way, one obtains finally

$$A_\mu = f(\hat{n}_h)a_\mu - a^i B_{i\mu} g(\hat{n}_p), \qquad A_i = f'(\hat{n}_p)a_i + a^\mu B_{i\mu} g'(\hat{n}_h), \tag{2.52}$$

where f, g, f', and g' are arbitrary functions that will be determined later.

The raising operators A^μ and A^i can be obtained from Eqs. (2.44b) and (2.44c) after substitution of (2.52). Omitting the fine details, which are similar to those just discussed, we note that (2.44b) gives the following results for the coefficients of (2.42b):

$$\Phi^\mu_\nu = \delta^\mu_\nu h'(\hat{n}_h) - A^\mu_\nu g'(\hat{n}_h), \qquad \Phi^i_j = \delta^i_j h(\hat{n}_p) - A^i_j g(\hat{n}_p), \tag{2.53a}$$

$$\Gamma^\mu_\nu = f'(\hat{n}_h)\delta^\mu_\nu, \quad \Gamma^i_j = f(\hat{n}_p)\delta^i_j, \tag{2.53b}$$

which involve two additional arbitrary functions, $h'(\hat{n}_h)$ and $h(\hat{n}_p)$. However, Eq. (2.44c) subsequently determines these as follows:

$$h'(\hat{n}_h) = g'(\hat{n}_h), \quad h(\hat{n}_p) = g(\hat{n}_p). \tag{2.54}$$

The explicit expressions for the raising operators then are given by

$$A^\mu = (a^\mu - a^\nu A^\mu_\nu)g'(\hat{n}_h) - f'(\hat{n}_p)B^{i\mu}a_i, \qquad A^i = (a^i - a^j A^i_j)g(\hat{n}_p) + f(\hat{n}_h)B^{i\mu}a_\mu, \tag{2.55}$$

which involve the same four arbitrary functions as in (2.52). It can then be checked straightforwardly that Eqs. (2.52) and (2.55) satisfy identically the remaining commutation relations (2.44d). Thus (2.52) and (2.55) are the most general Dyson mappings of the single-fermion operators consistent with the required transformation properties under $U(\Omega)$.

The remaining question is how to determine the four "arbitrary" functions of the number operators. The simplest choice is to require that in the purely quasihole subspace of $\mathfrak{J}$, A^μ and A_μ should be equivalent to the corresponding operators a^μ and a_μ, while in the purely quasiparticle subspace, A^i and A_i should be equivalent to the corresponding operators a^i and a_i. This means that successive application of these operators on the vacuum $|0)$ can only create normalized vectors. Now, from Eqs. (2.32), the operators (2.55) can be rewritten as follows:

$$\begin{aligned}A^\mu &= [a^\mu(1+\hat{n}_h) - a^\nu(A_B)^\mu_\nu]g'(\hat{n}_h) - f'(\hat{n}_p)B^{i\mu}a_i, \\ A^i &= [a^i(1+\hat{n}_p) - a^j(A_B)^i_j]g(\hat{n}_p) + f(\hat{n}_h)B^{i\mu}a_\mu,\end{aligned} \tag{2.56}$$

where

$$(A_B)^\mu_\nu \equiv B^{i\mu}B_{i\nu}, \quad (A_B)^i_j \equiv B^{i\mu}B_{j\mu} \tag{2.57}$$

are the generators of the boson unitary groups $\mathscr{U}_{hB}(\Omega_h)$ and $\mathscr{U}_{pB}(\Omega_p)$. It is then immediately seen that our "normalization" condition requires that

$$(1+\hat{n}_h)g'(\hat{n}_h) = (1+\hat{n}_p)g(\hat{n}_p) = 1. \tag{2.58a}$$

From the expression (2.52) for the annihilation operators, the normalization condition immediately yields

$$f(\hat{n}_h) = f'(\hat{n}_p) = 1. \tag{2.58b}$$

With the functions of the number operators determined, the final expressions for the Dyson images of the single-fermion operators are given by

$$\begin{aligned}A^\mu &= (a^\mu - a^\nu A^\mu_\nu)(1+\hat{n}_h)^{-1} - B^{i\mu}a_i \\ &= (a^\mu - a^\nu A^\mu_\nu)(1+\hat{n})^{-1} - B^{i\mu}a_i, \\ A^i &= (a^i - a^j A^i_j)(1+\hat{n}_p)^{-1} + B^{i\mu}a_\mu \\ &= (a^i - a^j A^i_j)(1+\hat{n})^{-1} + B^{i\mu}a_\mu;\end{aligned} \tag{2.59a}$$

$$\begin{aligned}A_\mu &= a_\mu - B_{i\mu}a^i(1+\hat{n}_p)^{-1} = a_\mu - B_{i\mu}a^i(1+\hat{n})^{-1}, \\ A_i &= a_i + B_{i\mu}a^\mu(1+\hat{n}_h)^{-1} = a_i + B_{i\mu}a^\mu(1+\hat{n})^{-1};\end{aligned} \tag{2.59b}$$

where the identities (2.26) were used to replace $(1+\hat{n}_h)^{-1}$ or $(1+\hat{n}_p)^{-1}$ by $(1+\hat{n})^{-1}$ in the second form of each equation. For later use, we also note from Eq. (2.41), that the creation operators can be rewritten as follows:

$$\begin{aligned}A^\mu &= (a^\mu + \tfrac{1}{2}[a^\mu, \mathscr{C}_h^{(2)}])(1+\hat{n})^{-1} - B^{i\mu}a_i, \\ A^i &= (a^i + \tfrac{1}{2}[a^i, \overline{\mathscr{C}}_p^{(2)}])(1+\hat{n})^{-1} + B^{i\mu}a_\mu.\end{aligned} \tag{2.60}$$

Equipped with the Dyson images $R^{i\mu}$, $R_{i\mu}$, A^{I}, and A_I, one can construct the Dyson mappings of all other fermion operators. In particular, for the elementary pair-transfer operators (2.30d), one may take

$$R^{IJ} = \tfrac{1}{2}[A^{I}, A^{J}], \quad R_{IJ} = \tfrac{1}{2}[A_J, A_I], \tag{2.61a}$$

$$A^{i}_{\mu} = \tfrac{1}{2}[A^{i}, A_{\mu}], \quad A^{\mu}_{i} = [A^{\mu}, A_i], \tag{2.61b}$$

which is tantamount to fulfilling the commutation relations (2.6n) by definition. In addition, it is easily seen that because A_I and A^{I} satisfy all the transformation requirements under $U(\Omega)$, so must the expressions (2.61) for the pair-transfer operators, i.e., the commutation rules (2.6f) and (2.6g) are fulfilled. Moreover, (2.61a) takes care of the antisymmetry requirement (2.5b).

In summary, we have obtained in the ideal space a Dyson realization of the full $U(\Omega)$ algebra as well as of the core subalgebra $U_h(\Omega_h) \times U_p(\Omega_p)$. With the inclusion of a convenient normalization requirement, the realizations of one-particle transfer operators with the correct tensor properties under $U(\Omega)$ are also uniquely determined. This also holds for higher tensors, such as the two-particle transfer operators. But so far, nothing has been said about the mutual commutators (2.6h) of the two-particle transfer operators, the commutators (2.6k) of the one- with the two-particle transfer operators [which, because of (2.61), become triple commutators of one-particle transfer operators], and finally the commutators (2.6l) and (2.6m). These additional commutation rules must be satisfied if one is to have a realization of the full $SO(2\Omega+1)$ algebra. Now, it is a straightforward exercise to check that, in general, they are not identically satisfied, i.e., not over the full ideal space. However, as we shall show in Sec. III, they are satisfied in the finite-dimensional physical subspace, which is all that is needed.

III. THE SPINOR IRREP OF SO(2Ω+1) IN THE IDEAL SPACE

A. Construction of the physical subspace

In addition to Eq. (2.33), the Dyson mapping also satisfies the conditions

$$(c_\mu c_i)_D |0) = R_{i\mu}|0) = 0, \quad (c_I)_D|0) = A_I|0) = 0, \tag{3.1}$$

which permit the correspondence $|0\rangle \to |0)$ of the fermion and ideal-space vacuum states. Therefore, the image in $\mathfrak{J}$ of the fermion basis (2.8) can be obtained by mapping the vacua and replacing the fermion operators with the corresponding Dyson operators. We wish to show that the image vectors form an orthogonal (though not normalized) set of antisymmetric vectors; these span a 2^{Ω}-dimensional subspace of $\mathfrak{J}$, which we call the physical subspace $\mathfrak{J}_P$. It will subsequently be shown that $\mathfrak{J}_P$ carries the spinor irrep of $SO(2\Omega+1)$.

We discuss first the mapping of the vectors (2.8a) having equal numbers of particles and holes. By directly operating with (2.40b) and (2.59a) it is easily verified that

$$|i\mu)_D \equiv R^{i\mu}|0) = A^{i}A^{\mu}|0) = -A^{\mu}A^{i}|0) = B^{i\mu}|0), \tag{3.2}$$

and

$$|i_1\mu_1 i_2\mu_2)_D \equiv R^{i_1\mu_1}R^{i_2\mu_2}|0) = (B^{i_1\mu_1}B^{i_2\mu_2} - B^{i_1\mu_2}B^{i_2\mu_1})|0), \tag{3.3}$$

where (3.3) is antisymmetric to the exchange of a pair of hole or a pair of particle indices. More generally, it can now be proved that

$$|i_1\mu_1 \cdots i_{N_B}\mu_{N_B})_D \equiv \prod_{n=1}^{N_B} R^{i_n\mu_n}|0) = \prod_{n=1}^{N_B} A^{i_n}A^{\mu_n}|0) = \sum_{\mathscr{P}} (-1)^{\mathscr{P}} \mathscr{P} \prod_{n=1}^{N_B} B^{i_n\mu_n}|0), \tag{3.4}$$

where the sum runs over all signed permutations $\mathscr{P}$ of either hole indices, or, equivalently, particle indices. Thus we wish to prove that repeated operation with the operators $R^{i\mu}$ on the vacuum generates a boson vector that is completely antisymmetric under the exchange of any pair of hole indices or particle indices for $N_B > 1$. The proof follows easily from the observation that the factors $R^{i\mu}$ all commute [Eq. (2.6d)]. Thus any pair of factors whose indices are to be exchanged can be commuted through to the vacuum, whereupon, application of Eq. (3.3), shows that the vector is antisymmetric to that exchange. Moreover, according to Eq. (3.2), each $R^{i\mu}$ acting on the vacuum can be converted to the product $A^{i}A^{\mu}$, which also commutes with all the remaining $R^{i\mu}$. Thus the process can be continued until each $R^{i\mu}$ is converted to the corresponding product $A^{i}A^{\mu}$. Finally, from the structure of the operators $R^{i\mu}$ [Eq. (2.40b)], it can be seen that (3.4) gives the correct sign phase of the antisymmetrized product.

Next, we consider the mapping of the vectors having an excess number of holes or particles. By operating directly on the vacuum one finds

$$|\nu i\mu)_D \equiv A^{\nu}R^{i\mu}|0) = R^{i\mu}A^{\nu}|0) = (a^{\nu}B_{i\mu} - a^{\mu}B^{i\nu})|0), \tag{3.5a}$$

$$|ji\mu)_D \equiv A^{j}R^{i\mu}|0) = R^{i\mu}A^{j}|0) = (a^{j}B^{i\mu} - a^{i}B^{j\mu})|0), \tag{3.5b}$$

and

$$|\nu_1\nu_2)_D \equiv A^{\nu_1}A^{\nu_2}|0) = a^{\nu_1}a^{\nu_2}|0) = -a^{\nu_2}a^{\nu_1}|0) = -A^{\nu_2}A^{\nu_1}|0), \tag{3.6a}$$

$$|j_1 j_2)_D \equiv A^{j_1}A^{j_2}|0) = a^{j_1}a^{j_2}|0) = -a^{j_2}a^{j_1}|0) = -A^{j_2}A^{j_1}|0). \tag{3.6b}$$

The vectors (3.5a) and (3.6a) are antisymmetric to the exchange of the two-hole indices, while (3.5b) and (3.6b) are antisymmetric to the exchange of the two-particle indices. More generally, for the images of the vectors (2.8b) and (2.8c) we can then prove that

$$|\nu_1\cdots\nu_{n_h} i_1\mu_1\cdots i_{N_B}\mu_{N_B})_D \equiv \prod_{n=1}^{n_h} A^{\nu_n}|i_1\mu_1\cdots i_{N_B}\mu_{N_B})_D = (n_h!)^{-1}\sum_{\mathscr{P}_h}(-1)^{\mathscr{P}_h}\mathscr{P}_h \prod_{n=1}^{n_h} a^{\nu_n}\prod_{m=1}^{N_B} B^{i_m\mu_m}|0), \tag{3.7a}$$

$$|j_1\cdots j_{n_p} i_1\mu_1\cdots i_{N_B}\mu_{N_B})_D \equiv \prod_{n=1}^{n_p} A^{j_n}|i_1\mu_1\cdots i_{N_B}\mu_{N_B})_D = (n_p!)^{-1}\sum_{\mathscr{P}_p}(-1)^{\mathscr{P}_p}\mathscr{P}_p \prod_{n=1}^{n_p} a^{j_n}\prod_{m=1}^{N_B} B^{i_m\mu_m}|0), \tag{3.7b}$$

where $\mathscr{P}_h$ refers to a permutation of hole and $\mathscr{P}_p$ of particle indices, and it is understood that in case the boson number $N_B = 0$, the ket $|i_1\mu_1\cdots i_{N_B}\mu_{N_B})_D$ on the rhs becomes the vacuum $|0)$. In fact, in this case, as is easily seen from Eqs. (2.59), the vectors (3.7) are pure products of quasifermion creation operators acting on the vacuum, which are automatically antisymmetric. According to (3.4), the vectors (3.7) are antisymmetric to exchanges of indices among the bosons alone. Furthermore, they are also antisymmetric to exchanges of indices among the quasifermions alone, which follows from (3.6) and the commutation of the operators A^I with the $R^{i\mu}$. The only point left to prove is the antisymmetry to exchanges of indices between the quasifermions and the bosons when $N_B \neq 0$. But this readily follows from the antisymmetry in Eqs. (3.5), together with the mutual commutation of the set of operators $R^{i\mu}$, A^I, so that any pair $A^I R^{i\mu}$ can be commuted through to the vacuum. Finally, the phase of the antisymmetrized vectors is correctly given by (3.7) as can be seen from the forms of the operators A^i and $R^{i\mu}$.

According to (3.4) and (3.7) all the antisymmetric vectors can be expressed as products of the A^I acting on the vacuum. Thus the correspondence $|0\rangle \rightarrow |0)$, together with the Dyson mapping of one-particle transfer operators $c^I \rightarrow A^I$ establishes a one-to-one correspondence between the orthonormal fermion basis vectors (2.8) and the vectors (3.4) and (3.7). The latter, which clearly form an orthogonal but not orthonormal set, by definition span the physical subspace $\mathfrak{F}_P$. The normalization will be taken care of later. Denoting the injective map by V, which corresponds to what is historically called the Usui operator,[23] one has

$$V|\,\rangle = |\,)_D, \tag{3.8}$$

where $|\,\rangle$ denotes any vector in $\mathfrak{H}_F$ and $|\,)_D$ any vector in $\mathfrak{F}_P$. The inverse V^{-1} may be defined so that for any vector $|u)$ in the orthogonal complement of $\mathfrak{F}_P$, $V^{-1}|u) = 0$. Then V satisfies

$$V^{-1}V = 1_F, \quad VV^{-1} = P, \quad PV = V, \tag{3.9}$$

where 1_F is the identity operator in $\mathfrak{H}_F$ and P is the projector to $\mathfrak{F}_P$.

From the structure of the basis vectors (3.4) and (3.7), the vacuum conditions (2.33) and (3.1), and the fulfillment on $\mathfrak{F}$ of the U(Ω) subalgebra as well as the tensor properties of the one-particle transfer operators A^I and A_I, it follows that $\mathfrak{F}_P$ is the direct sum of all the antisymmetric U(Ω) irreps contained in $\mathfrak{H}_F$ and therefore of those of the core subalgebra, a point which is central to the discussion in the next subsection. However, it cannot yet be asserted that $\mathfrak{F}_P$ carries the spinor representation of SO($2\Omega + 1$), since it has not yet been shown that the remaining commutators that are not fulfilled on the whole space $\mathfrak{F}$ are, in fact, fulfilled in $\mathfrak{F}_P$. This will be done in Sec. III C after development of some results that will also play an important role in the unitarization. As a first step in this direction we sketch the proof of the following relations (more details can be found in Ref. 14):

$$PXP = XP, \quad \text{for } X = B_{i\mu}, A_I, a_I, \tag{3.10a}$$

$$PXP = PX, \quad \text{for } X = B^{i\mu}, a^I, \tag{3.10b}$$

$$PXP = XP = PX, \quad \text{for } X = A^J_{\ I}, R^{i\mu}, A^I. \tag{3.10c}$$

What these relations signify is that the operators X leave invariant the following subspaces of $\mathfrak{F}$: the physical subspace $\mathfrak{F}_P$ in (3.10a), its orthogonal complement, the unphysical subspace in (3.10b), and both subspaces in (3.10c), i.e., $\mathfrak{F}$ is reduced by X. The invariance of $\mathfrak{F}_P$ is immediate for the raising operators $R^{i\mu}$ and A^I from the construction of the basis vectors (3.4) and (3.7); this invariance also holds for $A^I_{\ J}$, $R_{i\mu} = B_{i\mu}$, and A_I, because of their commutation relations with $R^{i\mu}$ and A^I and the property of annihilating the vacuum. The H.c. of Eq. (3.10a) then implies (3.10b) for $X = B^{i\mu}$ and $A^J_{\ I}$, which also shows that the latter satisfies (3.10c). The invariance of the physical subspace under a_I follows from the fact that these operators annihilate pure boson states while acting like A_I on physical vectors with one or more quasifermions. This establishes (3.10a), and, by Hermitian conjugation, (3.10b) as well. Finally, $PXP = XP$ for $X = (R^{i\mu})^\dagger, (A^I)^\dagger$ since these are combinations of operators that leave $\mathfrak{F}_P$ invariant. Hermitian conjugation then completes the proof of (3.10c). Finally, the operators X listed as satisfying (3.10a) and (3.10b) do not also satisfy (3.10c), as can readily be proven by counterexample.

B. Casimir-operator identities

Our method for unitarizing the Dyson mapping utilizes a few mathematical identities that can be derived from the quadratic Casimir operators. Since one of these identities is also useful for completing our proof that $\mathfrak{F}_P$ carries the spinor representation of SO($2\Omega + 1$), we digress briefly to derive them.

As shown by Eqs. (2.34)–(2.37), the Casimir operators of the core subalgebra are not diagonal on $\mathfrak{F}$, but when restricted to $\mathfrak{F}_P$, the situation is different. Since $\mathfrak{F}_P$ carries all the antisymmetric irreps of the core subalgebra, the relations (2.10) are valid in $\mathfrak{F}_P$. In particular, for the quadratic Casimir operators, one has the identities

$$\begin{aligned} \mathscr{C}_h^{(2)}\overleftrightarrow{P} &= (\hat{N}_B + \hat{n}_h)(\Omega_h - \hat{N}_B - \hat{n}_h + 1)\overleftrightarrow{P}, \\ \mathscr{C}_p^{(2)}\overleftrightarrow{P} &= (\hat{N}_B + \hat{n}_p)(\Omega_p - \hat{N}_B - \hat{n}_p + 1)\overleftrightarrow{P}, \end{aligned} \tag{3.11}$$

where the relations

$$\hat{N}_h = \mathrm{Tr}(A^{\mu}_{\mu}) = \hat{N}_{\mathrm{B}} + \hat{n}_h, \quad \hat{N}_p = \mathrm{Tr}(A^{i}_{i}) = \hat{N}_{\mathrm{B}} + \hat{n}_p, \tag{3.12}$$

were used, and the notation $\overleftrightarrow{P}$ indicates that the attached projector P to the physical subspace could also be commuted to the far left in accord with Eq. (3.10c). For the two-body part of (3.11) one obtains from Eq. (2.34) the results

$$\begin{aligned} \overline{\mathscr{C}}_h{}^{(2)}\overleftrightarrow{P} &= -(\hat{N}_{\mathrm{B}} + \hat{n}_h)(\hat{N}_{\mathrm{B}} + \hat{n}_h - 1)\overleftrightarrow{P}, \\ \overline{\mathscr{C}}_p{}^{(2)}\overleftrightarrow{P} &= -(\hat{N}_{\mathrm{B}} + \hat{n}_p)(\hat{N}_{\mathrm{B}} + \hat{n}_p - 1)\overleftrightarrow{P}, \end{aligned} \tag{3.13}$$

which, in turn, imply that

$$\overline{\mathscr{C}}_{\mathrm{B}}{}^{(2)}\overleftrightarrow{P} = -\hat{N}_{\mathrm{B}}(\hat{N}_{\mathrm{B}} - 1)\overleftrightarrow{P}. \tag{3.14}$$

According to (3.13), the relation $\overline{\mathscr{C}}_{\mathrm{B}}{}^{(2)} = -\hat{N}_{\mathrm{B}}(\hat{N}_{\mathrm{B}} - 1)$ holds in the subspace with $n_h = 0$ and also in the subspace with $n_p = 0$, and, therefore, must hold in the union of these subspaces, which is just $\mathfrak{F}_P$.

A chain of useful identities may be derived by taking commutators of both sides of Eqs. (3.13) and (3.14) with various operators, always taking Eqs. (3.10) into account. For example, with the aid of Eqs. (2.38), one obtains for the commutators with $B^{i\mu}$ the following identities:

$$P\tfrac{1}{2}[B^{i\mu},\overline{\mathscr{C}}_{\mathrm{B}}{}^{(2)}] = -PB^{i\nu}B^{j\mu}B_{j\nu} = PB^{i\mu}\hat{N}_{\mathrm{B}}, \tag{3.15a}$$

$$P\tfrac{1}{2}[B^{i\mu},\overline{\mathscr{C}}_h{}^{(2)} - \overline{\mathscr{C}}_{\mathrm{B}}{}^{(2)}] = -PB^{i\nu}a^{\mu}a_{\nu} = PB^{i\mu}\hat{n}_h, \tag{3.15b}$$

$$P\tfrac{1}{2}[B^{i\mu},\overline{\mathscr{C}}_p{}^{(2)} - \overline{\mathscr{C}}_{\mathrm{B}}{}^{(2)}] = -PB^{j\mu}a^{i}a_{j} = PB^{i\mu}\hat{n}_p. \tag{3.15c}$$

By adding these three equations together with $PB^{i\mu}$ and taking note of Eq. (2.40b), one also obtains

$$PR^{i\mu} = PB^{i\mu}(1 + \hat{N}_{\mathrm{B}} + \hat{n}). \tag{3.16}$$

Adding (3.15a) separately to (3.15b) and (3.15c) yields

$$\begin{aligned} -PB^{i\nu}A^{\mu}_{\nu} &= PB^{i\mu}(\hat{N}_{\mathrm{B}} + \hat{n}_h), \\ -PB^{j\mu}A^{i}_{j} &= PB^{i\mu}(\hat{N}_{\mathrm{B}} + \hat{n}_p). \end{aligned} \tag{3.17}$$

From the commutator of both sides of the first of these equations with a^{ν} and the second with a^{j}, followed by Hermitian conjugation (for later convenience), one obtains the Pauli principle between the bosons and quasifermions in the form

$$\begin{aligned} (a_{\nu}B_{i\mu} + a_{\mu}B_{i\nu})P &= 0, \\ (a_{j}B_{i\mu} + a_{i}B_{j\nu})P &= 0, \end{aligned} \tag{3.18}$$

which also implies

$$\begin{aligned} (a^{\mu}a_{\nu}B_{i\mu} + \hat{n}_h B_{i\nu})P &= 0, \\ (a^{i}a_{j}B_{i\mu} + \hat{n}_p B_{j\nu})P &= 0. \end{aligned} \tag{3.19}$$

Equations (3.18) and (3.19) are important for the next subsection.

Next, by taking commutators of (3.13) with the quasifermions, recalling Eqs. (2.41), one obtains the final identities

$$\begin{aligned} P\tfrac{1}{2}[a^{\mu},\overline{\mathscr{C}}_h{}^{(2)}] &= -Pa^{\nu}A^{\mu}_{\nu} = Pa^{\mu}(\hat{N}_{\mathrm{B}} + \hat{n}_h), \\ P\tfrac{1}{2}[a^{i},\overline{\mathscr{C}}_p{}^{(2)}] &= -Pa^{j}A^{i}_{j} = Pa^{i}(\hat{N}_{\mathrm{B}} + \hat{n}_p). \end{aligned} \tag{3.20}$$

C. The Pauli principle in the physical subspace

We now return to the problem of showing that the remaining commutation relations (2.6h) and (2.6k)–(2.6m), which are not identically satisfied over the whose space $\mathfrak{F}$, are, in fact, satisfied in $\mathfrak{F}_P$. This would mean that the physical subspace is a faithful replica of the fermion Fock space and that it carries the spinor irrep of $\mathrm{SO}(2\Omega + 1)$. Of physical importance is that all aspects of the Pauli exclusion principle would then be satisfied in $\mathfrak{F}_P$.

First, we can prove the following relations:

$$(A^{i}A^{\mu} - R^{i\mu})\overleftrightarrow{P} = 0, \quad (A^{\mu}A^{i} + R^{i\mu})\overleftrightarrow{P} = 0, \tag{3.21a}$$

$$(A_{\mu}A_{i} - R_{i\mu})P = 0, \quad (A_{i}A_{\mu} + R_{i\mu})P = 0, \tag{3.21b}$$

$$(A^{I}A_{J} - A^{I}_{J})P = 0, \quad (A_{J}A^{I} + A^{I}_{J} - \delta^{I}_{J})P = 0, \tag{3.21c}$$

$$(A^{I}A^{J} - R^{IJ})\overleftrightarrow{P} = 0, \quad (A_{J}A_{I} - R_{IJ})P = 0, \tag{3.21d}$$

$$(A^{i}A_{\mu} - A^{i}_{\mu})P = 0, \quad (A^{\mu}A_{i} - A^{\mu}_{i})P = 0. \tag{3.21e}$$

Each of these equations can be proved by straightforward substitution of the Dyson operators into the left-hand side (lhs) followed by invocation of one of the identities (3.18) or (3.19). The validity of Eqs. (3.21a), however, also follows immediately from the discussion in Sec. III A [see Eqs. (3.2) and (3.4)]. It should also be noted that because of Eqs. (2.61), the lhs of each of Eqs. (3.21d) and (3.21e) is equivalent to the anticommutator of the corresponding one-particle transfer operators. As a matter of fact, the discussion in connection with Eqs. (3.6) and (3.7) already implies the anticommutation relation $(A^{I}A^{J} + A^{J}A^{I})\overleftrightarrow{P} = 0$ [Eq. (3.23c) below], which also establishes the first of Eqs. (3.21d). For the rest, it suffices to give one sample calculation, say for the second of Eqs. (3.21c), which is one of the more complicated ones. Having taken into account Eqs. (2.59), (2.57), and (2.21) at the outset one may proceed as follows:

$$\begin{aligned} A_{j}A^{i}P &= [a_{j}a^{i} - a_{j}a^{k}(1 + \hat{n}_p)^{-1}B^{i\mu}B_{k\mu} + a^{\mu}(1 + \hat{n}_h)^{-1}a_{\nu}B_{j\mu}B^{i\nu}]P \\ &= [\delta_{j}{}^{i}Q_h - a^{i}a_{j} - Q_h(1 + \hat{n}_p)^{-1}B^{i\mu}B_{j\mu} + a^{k}a_{j}(1 + \hat{n}_p)^{-1}B^{i\mu}B_{k\mu} \\ &\quad + \delta_{j}{}^{i}a^{\mu}(1 + \hat{n}_h)^{-1}a_{\mu} + a^{\mu}(1 + \hat{n}_h)^{-1}a_{\nu}B^{i\nu}B_{j\mu}]P \\ &= [\delta_{j}{}^{i}Q_h - a^{i}a_{j} - Q_h(1 + \hat{n}_p)^{-1}B^{i\mu}B_{j\mu} - \hat{n}_p(1 + \hat{n}_p)^{-1}B^{i\mu}B_{j\mu} \end{aligned}$$

$$+\delta^i_j a^\mu(1+\hat n_h)^{-1}a_\mu - a^\mu(1+\hat n_h)^{-1}a_\mu B^{i\nu}B_{j\nu}]P$$

$$=[\delta^i_j Q_h - a^i a_j - Q_h(1+\hat n_p)^{-1}B^{i\mu}B_{j\mu} - \hat n_p(1+\hat n_p)^{-1}B^{i\mu}B_{j\mu} + (1-Q_h)(\delta^i_j - B^{i\mu}B_{j\mu})]P$$

$$=(\delta^i_j - a^i a_j - B^{i\mu}B_{j\mu})P = (\delta^i_j - A^i_j)P. \tag{3.22}$$

In the second step the operators were arranged in normal order, with the quasifermion anticommutators (2.22a) taken into account; in the third step the identity (3.19) was used; in the fourth step the expression (2.24) for Q_h was used; and in the fifth step the identity $(1-Q_h)(1+\hat n_p)^{-1}=(1-Q_h)$, which is obvious from the definition of Q_h, was used to obtain the final simplification.

By adding together Eqs. (3.21a)–(3.21c) pairwise, and from (2.61), (3.21d), and (3.21e) directly, one sees that all the fermion anticommutation rules are obeyed in the physical subspace:

$$\{A^i,A^\mu\}\overleftrightarrow{P} = \{A_\mu,A_i\}P = 0, \tag{3.23a}$$

$$\{A^I,A_J\}P = \delta^I_J P, \tag{3.23b}$$

$$\{A^I,A^J\}\overleftrightarrow{P} = \{A_J,A_I\}P = \{A^i,A_\mu\}P = \{A^\mu,A_i\}P = 0. \tag{3.23c}$$

Turning now to the remaining commutation relations, we first note that subtraction of each of the pairs of equations (3.21a)–(3.21c) generates precisely the commutation rules (2.61) and (2.6m) in the physical subspace:

$$[A_J,A^I]P = (\delta^I_J - 2A^I_J)P, \tag{3.24a}$$

$$[A^i,A^\mu]\overleftrightarrow{P} = 2R^{i\mu}\overleftrightarrow{P},\quad [A_\mu,A_i]P = 2R_{i\mu}P. \tag{3.24b}$$

Next, the commutation rules (2.6k), which are really triple commutators of one-particle transfer operators because of Eqs. (2.61), can readily be shown to hold in the physical subspace with the use of the commutation rules already established and the Jacobi identity. As a typical example, we consider the first of Eqs. (2.6k); from the Jacobi identity and Eqs. (3.10) we obtain

$$\begin{aligned}[A^I,R_{JK}] &= \tfrac12[A^I,[A_K,A_J]]P\\ &= \tfrac12[[A_J,A^I]P,A_KP] - \tfrac12[[A_K,A^I]P,A_JP]\\ &= -[A^I_JP,A_KP] + [A^I_KP,A_JP]\\ &= -[A^I_J,A_K]P + [A^I_K,A_J]P\\ &= (\delta^I_K A_J - \delta^I_J A_K)P,\end{aligned} \tag{3.25}$$

where (3.24a) was used in the third step and the second of Eqs. (2.6i) in the last step. Finally, the commutation rules (2.6h), which can be expressed in terms of quadruple commutators of one-particle transfer operators, can be verified to hold in $\mathfrak{F}_P$ with the aid of the Jacobi identity and the projections of Eqs. (2.6k) into $\mathfrak{F}_P$.

As an alternative to direct verification of the commutation relations (2.6h) and (2.6k), one may argue that the realization of the fermion anticommutation rules (3.23) in $\mathfrak{F}_P$, along with the basis vectors constructed in Sec. III A and the other established commutation relations, implies an isomorphism between the physical subspace and the fermion Fock space. Thus any fermion operator X transforms under the mapping V defined by Eq. (3.8) as

$$VXV^{-1} = PX_{\rm D}P, \tag{3.26}$$

where $X_{\rm D}$ is a Dyson operator that is not necessarily unique, as indicated by Eqs. (3.21). Depending on the operator X, the projector P may be dropped from the left or the right in accord with the rules (3.10). By choosing X as any one of the commutators (2.6h) or (2.6k), one may then immediately establish the corresponding commutation relation in $\mathfrak{F}_P$. For that matter, any operator identity in $\mathfrak{H}_{\rm F}$ may be so transformed into a corresponding "Pauli principle" identity in $\mathfrak{F}_P$. For example, the identity

$$c^Ic_Jc^Kc_L + c^Kc_Jc^Ic_L - \delta^K_Jc^Ic_L - \delta^I_Jc^Kc_L = 0$$

is transformed into

$$(A^I_JA^K_L + A^K_JA^I_L - \delta^K_JA^I_L - \delta^I_JA^K_L)\overleftrightarrow{P} = 0,$$

an identity first derived by Okubo for the antisymmetric representations of unitary groups.[21] The transformation can also be used to prove that the SO$(2\Omega+1)$ Casimir operator (2.13) has the eigenvalue (2.14) in $\mathfrak{F}_P$. Such Pauli constraints can be useful in practical applications for keeping approximate wave functions confined to the physical subspace.

IV. UNITARY REPRESENTATION IN THE IDEAL SPACE

A. Unitarizing the Dyson realization

As is characteristic of Dyson representations, ours fails to preserve all Hermitian conjugations; in particular, (2.4b) is violated. This is reflected in the fact that the basis vectors (3.4) and (3.7) are not in general normalized although mutually orthogonal. The representation can be unitarized either by finding a transformation that normalizes the basis vectors, as in Ref. 13, or directly transforms the generators so that Hermitian conjugation is fully restored. The second alternative chosen here has the advantage that the construction of the basis vectors, which may be inconvenient for some algebras, could be bypassed if desired.

In accord with the recent literature,[6–8] we seek a positive-definite similarity transformation S, such that for any generator X

$$(X)_U \equiv S(X)_{\rm D}S^{-1},\quad S(X^\dagger)_{\rm D}S^{-1} = (X^\dagger)_U = (X)_U{}^\dagger, \tag{4.1}$$

where $(\ )_{\rm D}$ denotes the Dyson image and $(\ )_U$ the corresponding image in the unitary representation. Equations (4.1) imply that

$$(X)_{\rm D}{}^\dagger M = M(X^\dagger)_{\rm D}, \tag{4.2}$$

where

$$M \equiv S^\dagger S = S^2, \tag{4.3}$$

based on the self-fulfilling ansatz $S^\dagger = S$ and thus $M^\dagger = M$. Equation (4.2) implies that M commutes with all generators that satisfy the Hermitian conjugation $(X)_D{}^\dagger = (X^\dagger)_D$, which can only be the generators A^I_J of the core subalgebra. It therefore follows that M is a function of the Casimir invariants of the core subalgebra and, therefore, of the number operators, at least in $\mathfrak{F}_P$. We exploit (4.2) in the form

$$M^{-1}(X_D)^\dagger M = (X^\dagger)_D. \tag{4.4}$$

Given a positive-definite operator M satisfying (4.4), one may choose $S = M^{1/2}$ and obtain the unitarized image from the relation

$$(X^\dagger)_U = S^{-1}(X_D)^\dagger S, \tag{4.5}$$

implied by Eqs. (4.1). In place of Eq. (4.4), one may, of course, solve its Hermitian conjugate, and in place of (4.5) use the first of Eqs. (4.1). This will usually give rise to a different unitarized operator, but the two operators are, in fact, equivalent when acting in the physical subspace, as will be seen below.

If S is a function of the number operators, then so is S^{-1}, and they must then satisfy

$$PSP = SP = PS, \quad PS^{-1}P = S^{-1}P = PS^{-1}, \tag{4.6}$$

which follows from the invariance of $\mathfrak{F}_P$ under functions of the number operators and the self-adjointness of S and S^{-1}. Now, the mapping operator V [Eqs. (3.8)] is not isometric, but we can define an isometric mapping operator U_M, called the Marumori operator,[20] by

$$U_M \equiv SV, \tag{4.7}$$

with the properties

$$U_M{}^\dagger U_M = 1_F, \quad U_M U_M{}^\dagger = P, \quad PU_M = U_M, \tag{4.8}$$

which are implied by (3.9) and the fact that S normalizes the basis for $\mathfrak{F}_P$ [see Eq. (4.20) below]. For any fermion operator X, the ideal-space image under U_M, denoted by X_M and called the Marumori image of X, is given by

$$X_M \equiv U_M X U_M{}^\dagger = P(X)_U P. \tag{4.9}$$

The operator $(X)_U$ in (4.9) may not be unique, and the projector P can possibly be dropped on one side, depending on X.

With the preliminaries taken care of, we turn to the solution of Eq. (4.4), first of all for the simplest X_D, namely, $R_{i\mu} = B_{i\mu}$. Since it is only necessary to satisfy (4.4) in the physical subspace, we multiply both sides from the left by the projector P, and utilize the Casimir identity (3.16) to obtain

$$\begin{aligned} PM^{-1}(\hat{N}_B,\hat{n})B^{i\mu}M(\hat{N}_B,\hat{n}) \\ = PR^{i\mu} = PB^{i\mu}(1+\hat{N}_B+\hat{n}). \end{aligned} \tag{4.10}$$

With the aid of the identity

$$f(\hat{N}_B,\hat{n})B^{i\mu} = B^{i\mu}f(\hat{N}_B+1,\hat{n}), \tag{4.11}$$

valid for an arbitrary function f, (4.10) can be rewritten in the form

$$\begin{aligned} PB^{i\mu}M^{-1}(\hat{N}_B+1,\hat{n})M(\hat{N}_B,\hat{n}) \\ = PR^{i\mu} = PB^{i\mu}(1+\hat{N}_B+\hat{n}), \end{aligned} \tag{4.12}$$

which is fulfilled by requiring that $S = M^{1/2}$ satisfy

$$S^{-1}(\hat{N}_B+1,\hat{n})S(\hat{N}_B,\hat{n}) = (1+\hat{N}_B+\hat{n})^{1/2}. \tag{4.13}$$

From (4.5), (4.9), (4.12), and (4.13), the Marumori image of the particle–hole creation operator can be evaluated as follows:

$$\begin{aligned} \tilde{R}^{i\mu} &\equiv (c^i c^\mu)_M \\ &= PS^{-1}(\hat{N}_B,\hat{n})R_{i\mu}{}^\dagger S(\hat{N}_B,\hat{n}) \\ &= PS^{-1}(\hat{N}_B,\hat{n})B^{i\mu}S(\hat{N}_B,\hat{n}) \\ &= PB^{i\mu}S^{-1}(\hat{N}_B+1,\hat{n})S(\hat{N}_B,\hat{n}) \\ &= PB^{i\mu}(1+\hat{N}_B+\hat{n})^{1/2} = PR^{i\mu}(1+\hat{N}_B+\hat{n})^{-1/2}, \end{aligned} \tag{4.14a}$$

$$\begin{aligned} \tilde{R}_{i\mu} \equiv (c_\mu c_i)_M = (\tilde{R}^{i\mu})^\dagger &= (1+\hat{N}_B+\hat{n})^{1/2}B_{i\mu}P \\ &= (1+\hat{N}_B+\hat{n})^{-1/2}(R^{i\mu})^\dagger P. \end{aligned} \tag{4.14b}$$

As anticipated, each Marumori operator can be written in two ways, involving representations of $(c^i c^\mu)_U$ that are different in the whole space $\mathfrak{F}$, but are equivalent in $\mathfrak{F}_P$.

Next, we solve (4.4) for the cases when $X_D = A_\mu$ and A_i. It is obviously sufficient to provide the details only for the first. Upon substituting from Eqs. (2.59), using the H.c. of the identity (2.26), the Casimir identity (3.20) on the rhs, the identity (4.11), and the identities

$$\begin{aligned} f(\hat{N}_B,\hat{n})a^I &= a^I f(\hat{N}_B,\hat{n}+1), \\ f(\hat{N}_B,\hat{n})a_I &= a_I f(\hat{N}_B,\hat{n}-1), \end{aligned} \tag{4.15}$$

valid for arbitrary functions f, one obtains the equation

$$\begin{aligned} P[&a^\mu M^{-1}(\hat{N}_B,\hat{n}+1)M(\hat{N}_B,\hat{n}) \\ &-(1+\hat{n})^{-1}M^{-1}(\hat{N}_B,\hat{n})M(\hat{N}_B-1,\hat{n}+1)B^{i\mu}a_i] \\ &= P[a^\mu(1+\hat{N}_B+\hat{n})(1+\hat{n})^{-1} - B^{i\mu}a_i]. \end{aligned} \tag{4.16}$$

This is satisfied provided that

$$S^{-1}(\hat{N}_B,\hat{n}+1)S(\hat{N}_B,\hat{n}) = [(1+\hat{N}_B+\hat{n})/(1+\hat{n})]^{1/2}, \tag{4.17a}$$

$$S^{-1}(\hat{N}_B,\hat{n})S(\hat{N}_B-1,\hat{n}+1) = (1+\hat{n})^{1/2}. \tag{4.17b}$$

Actually, (4.17b) is not independent but can be derived from (4.13) and (4.17a). The Marumori image of the one-particle transfer operator obtained with the aid of Eqs. (2.59b), (4.5), (4.9), and (4.15) is given by

$$\begin{aligned} \tilde{A}^\mu \equiv (c^\mu)_M &= PS^{-1}(\hat{N}_B,\hat{n})A_\mu{}^\dagger S(\hat{N}_B,\hat{n}) \\ &= P[a^\mu S^{-1}(\hat{N}_B,\hat{n}+1)S(\hat{N}_B,\hat{n}) \\ &\quad -(1+\hat{n})^{-1}S^{-1}(\hat{N}_B,\hat{n}) \\ &\quad \times S(\hat{N}_B-1,\hat{n}+1)a_i B^{i\mu}]. \end{aligned} \tag{4.18}$$

After substituting (4.17), one finally obtains

$$\widetilde{A}^{\mu} \equiv (c^{\mu})_{\rm M} = P(a^{\mu}[(1+\widehat{N}_{\rm B}+\hat{n})/(1+\hat{n})]^{1/2} - (1+\hat{n})^{-1/2}a_i B^{i\mu})$$
$$= P((a^{\mu} - a^{\nu}A^{\mu}_{\nu})[(1+\widehat{N}_{\rm B}+\hat{n})(1+\hat{n})]^{-1/2} - (1+\hat{n})^{-1/2}a_i B^{i\mu}), \quad (4.19a)$$

$$\widetilde{A}^{i} \equiv (c^{i})_{\rm M} = P(a^{i}[(1+\widehat{N}_{\rm B}+\hat{n})/(1+\hat{n})]^{1/2} + (1+\hat{n})^{-1/2}a_{\mu} B^{i\mu})$$
$$= P((a^{i} - a^{j}A^{i}_{j})[(1+\widehat{N}_{\rm B}+\hat{n})(1+\hat{n})]^{-1/2} + (1+\hat{n})^{-1/2}a_{\mu} B^{i\mu}), \quad (4.19b)$$

$$\widetilde{A}_{I} = (c_{I})_{\rm M} = (\widetilde{A}^{I})^{\dagger}, \quad (4.19c)$$

where the second form of each equation is obtained from the identities (3.20) and (2.26).

Since the generators of the core subalgebra are invariant under the transformation S, their Marumori representation is trivially given by

$$\widetilde{A}^{I}_{J} \equiv (c^{I}c_{J}) = A^{I}_{J}\overleftrightarrow{P}. \quad (4.20)$$

Finally, we can check that operation with $S(\widehat{N}_{\rm B},\hat{n})$ on the basis vectors (3.4) and (3.7) gives the proper normalization. From the recursion relations provided by Eqs. (4.13) and (4.17a), it is easy to deduce that

$$S(N_{\rm B},n) = (n!/(N_{\rm B}+n)!)^{1/2}, \quad (4.21)$$

provided that $S(0,0) = 1$, which is justified by the stipulation that the vacuum $|0)$ is normalized. Equation (4.21) indeed yields the correct normalization.

B. The generalized Holstein–Primakoff (GHP) representation

The GHP representation is yet another form of the unitary representation that corresponds formally to an infinite power series, which is useful for perturbative applications. It is easy to derive the GHP expressions from those given in the previous subsection with the aid of the Casimir identities. First, we recursively define the operators A^{n} by

$$(A^{0})^{I}_{J} \equiv I^{I}_{J} = \delta^{I}_{J}, \quad (A^{n})^{I}_{J} = (A^{n-1})^{K}_{J}A^{I}_{K}. \quad (4.22)$$

We also define $A_{\rm B}{}^{n}$ as the pure boson part of (4.22), where $(A_{\rm B})^{I}_{J}$ is defined by Eqs. (2.57). Iterated multiplication of (3.15a) from the right with matrix elements of $A_{\rm B}$ followed by invocation of (3.15a) yields the identities

$$PB^{i\mu}(-\widehat{N}_{\rm B})^{n} = PB^{i\nu}(A_{\rm B}{}^{n})^{\mu}_{\nu} = PB^{j\mu}(A_{\rm B}{}^{n})^{i}_{j} \quad (n=0,1,\ldots). \quad (4.23)$$

In fact, the last equality holds even without the projector, and follows from a simple rearrangement, which is based on the commutation of $(A_{\rm B})^{\mu}_{\nu}$ and $(A_{\rm B})^{i}_{j}$ (Ref. 24). In an analogous way, one may derive from Eqs. (3.17) the identities

$$\begin{aligned} PB^{i\mu}(-1)^{n}(\widehat{N}_{\rm B}+\hat{n}_{h})^{n} &= PB^{i\nu}(A^{n})^{\mu}_{\nu}, \\ PB^{i\mu}(-1)^{n}(\widehat{N}_{\rm B}+\hat{n}_{p})^{n} &= PB^{j\mu}(A^{n})^{i}_{j}. \end{aligned} \quad (4.24)$$

For any holomorphic function $g(z) = \Sigma^{\infty}_{n=0}c_n z^n$, one may define a corresponding vector operator $[g(A)]^{I}_{J} = \Sigma^{\infty}_{n=0}c_n(A^{n})^{I}_{J}$. In particular, corresponding to the square-root function $g(z) = (1+z)^{1/2}$, Eqs. (4.23) and (4.24) yield the formal identities

$$PB^{i\mu}(1+\widehat{N}_{\rm B})^{1/2} = PB^{i\nu}[(I-A_{\rm B})^{1/2}]^{\mu}_{\nu} = PB^{j\mu}[(I-A_{\rm B})^{1/2}]^{i}_{j} \equiv P[B(I-A_{\rm B})^{1/2}]^{i\mu}, \quad (4.25a)$$

$$PB^{i\mu}(1+\widehat{N}_{\rm B}+\hat{n}_{h})^{1/2} = PB^{i\nu}[(I-A)^{1/2}]^{\mu}_{\nu}, \quad (4.25b)$$

$$PB^{i\mu}(1+\widehat{N}_{\rm B}+\hat{n}_{p})^{1/2} = PB^{j\mu}[(I-A)^{1/2}]^{i}_{j}, \quad (4.25c)$$

where I is the identity matrix defined in (4.22). Next, by applying the identity

$$f(\widehat{N}_{\rm B},\hat{n}) = f(\widehat{N}_{\rm B},\hat{n}_{h}) + f(\widehat{N}_{\rm B},\hat{n}_{p}) - f(\widehat{N}_{\rm B},0), \quad (4.26)$$

which holds in $\mathfrak{F}$ because of Eqs. (2.21), to the first form of $\widetilde{R}^{i\mu}$ [Eq. (4.14a)], making use of Eqs. (4.25), one obtains

$$\widetilde{R}^{i\mu} = P[B^{i\mu}(1+\widehat{N}_{\rm B}+\hat{n}_{h})^{1/2} + B^{i\mu}(1+\widehat{N}_{\rm B}+\hat{n}_{p})^{1/2} - B^{i\mu}(1+\widehat{N}_{\rm B})^{1/2}]$$
$$= P\{B^{i\nu}[(I-A)^{1/2}]^{\mu}_{\nu} + B^{j\mu}[(I-A)^{1/2}]^{i}_{j} - [B(I-A_{\rm B})^{1/2}]^{i\mu}\}, \quad (4.27a)$$

$$R_{i\mu} = (R^{i\mu})^{\dagger}. \quad (4.27b)$$

The second expression in braces is the GHP representation for the particle–hole creation operator. It can readily be shown that the second form of (4.14a) leads again to (4.27), so that the GHP representation is unique.

The GHP representation can also be extended to the one-particle transfer operators. As the first step, we obtain from Eqs. (3.20) and (2.26) the identity

$$Pa^{I}(-1)^{n}(\widehat{N}_{\rm B}+\hat{n})^{n} = Pa^{J}(A^{n})^{I}_{J}, \quad (4.28)$$

which implies the formal relation

$$Pa^{I}(1+\widehat{N}_{\rm B}+\hat{n})^{1/2} = Pa^{J}[(I-A)^{1/2}]^{I}_{J}. \quad (4.29)$$

One more identity remains to be derived. From the quasifermion algebra, it is trivial to see that $a^{K}\rho^{J}_{I} = a^{J}\rho^{K}_{I}$, where ρ^{J}_{I} is defined by Eq. (2.46). Setting $K = I$ and summing over repeated indices then gives $a^{I}\rho^{J}_{I} = -a^{J}\hat{n}$, taking Eqs. (2.26) into account. By recursion, it is then easy to obtain the relation $a^{I}(\rho^{n})^{J}_{I} = a^{J}(-\hat{n})^{n}$, which applied to the square-root function yields the desired identity

$$a^{I}[(I-\rho)^{-1/2}]^{J}_{I} = a^{J}(1+\hat{n})^{-1/2}. \quad (4.30)$$

Substitution of Eqs. (4.29) and (4.30) into the first form of each of Eqs. (4.19) then yields

$$\widetilde{A}^{\mu} = P\{a^{\lambda}[(I-\rho)^{-1/2}]^{\nu}_{\lambda}[(I-A)^{1/2}]^{\mu}_{\nu} - B^{i\mu}[(I-\rho)^{-1/2}]^{j}_{i}a_{j}\}, \quad (4.31a)$$

$$\widetilde{A}^{i} = P\{a^{k}[(I-\rho)^{-1/2}]^{j}_{k}[(I-A)^{1/2}]^{i}_{j} + B^{i\mu}[(I-\rho)^{-1/2}]^{\nu}_{\mu}a_{\nu}\}, \quad (4.31b)$$

$$\widetilde{A}_{I} = (\widetilde{A}^{I})^{\dagger}, \quad (4.31c)$$

where the expressions in braces correspond to the GHP representation. It can easily be shown that use of the second form of Eqs. (4.19) leads to exactly the same expressions, so the GHP representation is completely unique.

V. FINAL REMARKS

The original SU(2) Holstein–Primakoff[25] expansion has a parameter of smallness, namely, $J^{-1/2}$, where J is the total angular-momentum quantum number that labels the irreps. As pointed out by Okubo,[18] the GHP expansion lacks a manifest expansion parameter and therefore does not converge. In physical applications this is not a problem since at some stage one introduces new bosons that are coupled to good angular momentum, or are obtained by a small-oscillation theory such as the random-phase approximation (RPA). Such procedures introduce small parameters analogous to that in the Holstein–Primakoff expansion which are suitable for perturbation theory. If one insists on using the above GHP operators in a global way, convergence difficulties can be avoided by the expedient of introducing a small parameter, for example, using $(I-\epsilon A)^{1/2}$, and at the end taking the limit $\epsilon \to 1$.[13]

As a final comment, we point out that the logical development used here could be modified as follows. Instead of postulating the properties of the quasifermions, one could *derive* them from the requirement that the U(Ω) commutation rules and the single-fermion anticommutation rules be satisfied (in the physical subspace).

ACKNOWLEDGMENTS

Both authors are grateful to the Alexander von Humboldt Stiftung for U.S. Senior Scientist awards during the preliminary stages of this work, and to their respective hosts, Professor Dr. Reiner Dreizler and Professor Dr. Walter Greiner, at the Institut für Theoretische Physik in Frankfurt.

This work was supported in part by the Department of Energy (A. K.) and the National Science Foundation (E. R. M.).

[1] B. G. Wybourne, *Classical Groups for Physicists* (Wiley–Interscience, New York, 1974); A. Nowicki, Fortschr. Phys. **34**, 649 (1986).
[2] A. Klein and E. R. Marshalek, Rev. Mod. Phys. (to be published).
[3] F. Iachello and A. Arima, *The Interacting Boson Model* (Cambridge U.P., Cambridge, 1987).
[4] D. J. Rowe, Rep. Prog. Phys. **48**, 1419 (1985).
[5] D. Janssen, F. Dönau, S. Frauendorf, and R. V. Jolos, Nucl. Phys. A **172**, 145 (1971).
[6] J. Deneen and C. Quesne, J. Math. Phys. **23**, 878, 2004 (1982); **25**, 1638 (1984); **26**, 2705 (1985); O. Castaños, E. Chacòn, M. Moshinsky, and C. Quesne, *ibid.* **26**, 2107 (1985).
[7] A. Klein, T. D. Cohen, and C. T. Li, Ann. Phys. (NY) **141**, 382 (1982); A. Klein and Q. Y. Zhang, J. Math. Phys. **27**, 1987 (1986); D. Bonatsos, A. Klein, and Q. Y. Zhang, Phys. Lett. B **175**, 249 (1986); D. Bonatsos, A. Klein, and Q. Y. Zhang, Phys. Rev. C **34**, 686 (1986).
[8] D. J. Rowe, J. Math. Phys. **25**, 2662 (1984); D. J. Rowe, G. Rosensteel, and R. Gilmore, *ibid.* **26**, 2787 (1985); K. T. Hecht, *The Vector Coherent State Method and Its Application to Problems of Higher Symmetries* (Springer, Berlin, 1987); D. J. Rowe, R. Le Blanc, and K. T. Hecht, J. Math. Phys. **29**, 287 (1988).
[9] D. J. Rowe and J. Carvalho, Phys. Lett. B **175**, 243 (1986).
[10] A. Klein and E. R. Marshalek, Z. Phys. A **329**, 441 (1988).
[11] A. Bohr and B. R. Mottelson, *Nuclear Structure* (Benjamin, Reading, MA, 1975), Vol. 2, pp. 416–446.
[12] V. Paar, in *Future Directions in Studies of Nuclei Far from Stability*, edited by J. H. Hamilton, E. H. Spejewski, C. R. Bingham, and E. F. Zganjar (North-Holland, Amsterdam, 1980).
[13] E. R. Marshalek, Phys. Lett. B **92**, 245 (1980); Nucl. Phys. A **357**, 398 (1981).
[14] H. B. Geyer and F. J. W. Hahne, Phys. Lett. B **90**, 6 (1980); H. B. Geyer and F. J. W. Hahne, S. Afr. Tydskr. Fis. **6**, 67 (1983).
[15] A. Kuriyama and M. Yamamura, Prog. Theor. Phys. **66**, 2130, 2147 (1981).
[16] T. Suzuki and K. Matsuyanagi, Prog. Theor. Phys. **56**, 1156 (1977).
[17] F. J. W. Hahne and A. Klein (unpublished).
[18] S. Okubo, Phys. Rev. C **10**, 2048 (1974).
[19] A. M. Perelomov and V. S. Popov, Sov. J. Nucl. Phys. **3**, 676 (1966).
[20] T. Marumori, M. Yamamura, and A. Tokunaga, Prog. Theor. Phys. **31**, 1009 (1964); T. Marumori, M. Yamamura, A. Tokunaga, and K. Takada, *ibid.* **32**, 726 (1964); M. Yamamura, *ibid.* **33**, 199 (1965).
[21] S. Okubo, J. Math. Phys. **16**, 528 (1975).
[22] A. Klein, J. Math. Phys. **4**, 1283 (1963).
[23] T. Usui, Prog. Theor. Phys. **23**, 787 (1960).
[24] E. R. Marshalek and J. Weneser, Phys. Rev. C **2**, 1682 (1970).
[25] T. Holstein and H. Primakoff, Phys. Rev. **58**, 1698 (1950).

Electromagnetic fields in an expanding universe

P. A. Hogan
Mathematical Physics Department, University College, Dublin, Ireland
G. F. R. Ellis
International School for Advanced Studies, Strada Costieri 11, 34014 Trieste, Italy

(Received 17 May 1988; accepted for publication 10 August 1988)

The asymptotic form of the electromagnetic field due to a bounded distribution of charge current in an open, expanding Friedmann–Lemaître–Robertson–Walker universe is studied. The technique used is to first describe a mechanism for passing from a solution of Maxwell's vacuum field equations on Minkowskian space-time to a solution of Maxwell's field equations in a region free of charge current on the cosmological background. This is tested on the field of an accelerating point charge and then applied to the rigorous treatment of the asymptotic electromagnetic field of a bounded charge-current distribution in Minkowskian space-time given by Goldberg and Kerr [J. Math. Phys. **5**, 172 (1964)]. A "peeling expansion" of the electromagnetic field in the expanding universe is obtained in inverse powers of a parameter that is proportional to the area distance along the generators of future null cones with vertices on the world line of a fundamental observer. The algebraic character of the two leading coefficients in the expansion is the same as that of the two leading coefficients in the Goldberg–Kerr expansion in Minkowskian space-time. In addition, bounds can be calculated, at any instant in the history of a fundamental observer, on all the coefficients in the peeling expansion, as a consequence of the evaluation of such bounds by Goldberg and Kerr in the case treated by them.

I. INTRODUCTION

In the pioneering paper by Hawking[1] on the asymptotic gravitational field of a bounded system in a Friedmann–Lemaître–Robertson–Walker (FLRW) universe, the field described by the Weyl tensor exhibits an unconventional "peeling-off" behavior. The field components die away along future-pointing hypersurface-orthogonal null geodesics in inverse *half-integral* powers of the affine parameter distance along these geodesics. This is in striking contrast to the behavior of the asymptotic gravitational field due to bounded sources in a vacuum whose field components die away in inverse *integral* powers of the affine parameter distance (see, for example, Refs. 2–5) or the luminosity distance.[6,7] In an attempt to get a clear picture of the asymptotic behavior of fields due to bounded sources in cosmological models we study in this paper the asymptotic electromagnetic field due to an arbitrary but bounded distribution of charge current in an open expanding FLRW universe.

A mathematically rigorous study of the asymptotic electromagnetic field due to a bounded distribution of charge current in Minkowskian space-time is given in an important paper by Goldberg and Kerr,[8] which is briefly summarized in Sec. IV below. They establish the asymptotic expansion of the electromagnetic field in inverse integral powers of the affine parameter distance along the null geodesic generators of future-pointing null cones with vertices on an arbitrary timelike line within the history of the charge-current distribution. Explicit bounds on the coefficients of the inverse integral powers of the affine parameter distance, together with the algebraic classification of the two leading coefficients, are given. The coefficients are calculated from integrals of the source distribution (and its multipole moments) over a compact region of space-time. Hence bounds for them would be expected to exist. An advantage in having them explicitly arises when we translate this result into the FLRW universe. We can use the bounds evaluated by Goldberg and Kerr to prove that the corresponding asymptotic expansion of the electromagnetic field of a bounded charge-current distribution in an open expanding FLRW universe exists. This is the main result of the present paper (given in Sec. IV below). We also show that the coefficients of the two leading terms in this asymptotic expansion have similar algebraic properties to the coefficients of the two leading terms in the asymptotic expansion in Minkowskian space-time studied by Goldberg and Kerr. An important by-product of our work is that the natural expansion parameter to use (from our point of view) along future-pointing hypersurface-orthogonal null geodesics in the FLRW universe turns out to be proportional to the "area distance," rather than to the affine parameter distance. Using the inverse of this former distance as expansion parameter, the asymptotic series displays the conventional peeling-off property involving *integral* powers of the expansion parameter.

The electromagnetic field of a charged particle in an open, expanding FLRW universe is a special case of the bounded charge-current distribution mentioned above. The electromagnetic field in this case can be discussed in closed form; so we examine it first, in Sec. II, and use it later as an illustration of the mechanism described in Sec. III for passing from a solution of Maxwell's vacuum field equations on Minkowskian space-time to a solution of Maxwell's source-free field equations on the FLRW space-time. The relation of this work to the geometric optics approximation is discussed in a future paper.[9]

 0022-2488/89/010233-08$02.50

II. CHARGED PARTICLE IN AN EXPANDING UNIVERSE

The open, expanding FLRW dust-filled universe, with surfaces of homogeneity of constant negative curvature, has a line element that can be written in the form

$$ds^2 = \Omega^2\left[\frac{(dx^1)^2 + (dx^2)^2 + (dx^3)^2}{\{1 - \frac{1}{4}((x^1)^2 + (x^2)^2 + (x^3)^2)\}^2} - dt^2\right]$$
$$= \Omega^2(t) ds_0^2, \tag{2.1}$$

with

$$\Omega(t) = A(\cosh t - 1), \quad A = \text{const} > 0. \tag{2.2}$$

We consider solutions of Maxwell's equations on the cosmological background (2.1) having as source a timelike world line in this space-time. As Maxwell's equations are conformally invariant, we can equivalently consider their particle-like solutions on the space-time with line element ds_0^2 given by (2.1).

All the results of this section can be extended to space-times with conformal line elements

$$ds_0^2 = \frac{(dx^1)^2 + (dx^2)^2 + (dx^3)^2}{\{1 + (k/4)(x^1)^2 + (x^2)^2 + (x^3)^2)\}^2} - dt^2, \tag{2.3}$$

with $k = 0, \pm 1$, and thence to the cosmological models with line elements

$$ds^2 = \Omega^2(t) ds_0^2, \tag{2.4}$$

with $\Omega(t) = \frac{1}{2}At^2$ when $k = 0$ in (2.3) (this being the Einstein–de Sitter open universe), with $\Omega(t)$ given by (2.2) when $k = -1$, and with $\Omega(t) = A(1 - \cos t)$ when $k = +1$. In all cases A is a positive constant. The case $k = 0$ involves Maxwell's equations on Minkowskian space-time with the line element given by (2.3) with $k = 0$, while the case $k = +1$ is a closed universe. Since the Maxwell fields of charged particles on Minkowskian space-time are the well-known Liénard–Wiechert solutions, we will not discuss the case $k = 0$ here. As the charged particle fields given below do not involve asymptotic expansions, they can easily be extended to the case $k = +1$. However, their generalization to the fields of bounded charge-current distributions, given in Sec. IV below, involves asymptotic expansions. We will therefore exclude the closed universe from our discussions.

The line element (2.1) can be written in the null form

$$ds^2 = \Omega^2(v + R) ds_0^2, \tag{2.5}$$

with

$$ds_0^2 = \frac{f^2(dx^2 + dy^2)}{(1 + \frac{1}{4}(x^2 + y^2))^2} - 2\, dv\, dR - dv^2, \tag{2.6}$$

where $f = \sinh R$, while $\Omega(v + R)$ is given by (2.2) with $t = v + R$. We note that here $R = 0$ is a timelike geodesic of (2.6) with v as proper time along it, and the $v = \text{const}$ are the future light cones based on this world line. We shall assume the coordinate ranges $0 \leqslant R + \infty$, $-\infty < x < +\infty$, $-\infty < y < +\infty$, and $-\infty < v < +\infty$. The hypersurface-orthogonal future-pointing null geodesics tangent to $\partial/\partial R$ have the expansion

$$\rho = \frac{d}{dR}(\log f) = \coth R, \tag{2.7}$$

and R is an affine parameter along them.

The space-time with line element (2.6) is conformally flat. We can express this fact in the form

$$\frac{q^2(dx^2 + dy^2)}{(1 + \frac{1}{4}(x^2 + y^2))^2} - 2\, d\lambda\, dq - d\lambda^2$$
$$= \Phi^2\left[\frac{f^2(dx^2 + dy^2)}{(1 + \frac{1}{4}(x^2 + y^2))^2} - 2\, dv\, dR - dv^2\right]. \tag{2.8}$$

The line element on the left-hand side of this equation is that of Minkowskian space-time with λ labeling null cones, q measuring distance from the world line $q = 0$, and x,y the coordinates on the two-spheres $\lambda,q = \text{const}$. The term Φ is a function of R,v, and the coordinate transformation leading to the equality (2.8) is given by

$$\lambda = \frac{(a_0 - c_0 k_0)\sinh \frac{1}{2}v + c_0 \cosh \frac{1}{2}v}{\cosh(v/2) - k_0 \sinh(v/2)}, \tag{2.9}$$

$$\lambda + 2q = \frac{(a_0 - c_0 k_0)\sinh(R + v/2) + c_0 \cosh(R + v/2)}{\cosh(R + v/2) - k_0 \sinh(R + v/2)}, \tag{2.10}$$

with

$$\Phi^{-1} = \frac{2}{a_0}\left(\cosh\frac{v}{2} - k_0 \sinh\frac{v}{2}\right)$$
$$\times\left(\cosh\left(R + \frac{v}{2}\right) - k_0 \sinh\left(R + \frac{v}{2}\right)\right), \tag{2.11}$$

and a_0, c_0, k_0 are constants. If $a_0 = 1$ and $c_0 = k_0 = 0$, this becomes a transformation given by Infeld and Schild,[10]

$$\lambda = \tanh(v/2), \quad \lambda + 2q = \tanh(R + v/2). \tag{2.12}$$

However, the bounds these equations imply on λ,q for the assumed ranges of v,R are too severe for our purposes. We can improve on this by requiring first that $q \to +\infty$ as $R \to +\infty$ for any v,λ. Then (2.9) and (2.10) imply $k_0 = 1$. Now a translation of v can have the same effect as putting $a_0 = 2$, while a translation of λ allows us to put $c_0 = 1$. Then (2.9)–(2.11) result in

$$\lambda = e^v, \quad \lambda + 2q = e^{v + 2R}, \quad \Phi = e^{v + R}. \tag{2.13}$$

This transformation has been used already by Walker.[11]

In the flat space-time with the line element given by the left-hand side of (2.8), $q = 0$ is a timelike geodesic with λ proper time along it. We have $0 \leqslant q < +\infty$ and $-\infty < \lambda < +\infty$ in general. By (2.13) the line $q = 0$ appears to be mapped to the line $R = 0$ in the space-time (2.6). However, since $-\infty < v < +\infty$, we see that only half of the line $q = 0$, corresponding to $0 \leqslant \lambda < +\infty$, is in fact mapped to the whole line $R = 0$.

The potential one-form

$$A = -e_0 \rho\, dv, \tag{2.14}$$

with $e_0 = \text{const}$ and ρ given by (2.7), is a solution of Maxwell's vacuum field equations on (2.6). The corresponding Maxwell field is given by the two-form

$$F = (e_0/f^2) dR \wedge dv. \tag{2.15}$$

This is singular on the timelike geodesic $R = 0$ of (2.6), which we thus consider to be the world line of the charge e_0. On account of the conformal invariance of Maxwell's vacuum field equations, this is also a Maxwell field on the FLRW space-time with line element (2.5). We can regard A and F

above as the potential one-form and Maxwell two-form on the FLRW space-time with line element (2.5) because the Hodge dual of the two-form F in (2.15) has the same value,

$$e_0(1+\tfrac{1}{4}(x^2+y^2))^{-2}\,dx\wedge dy,$$

when calculated with respect to the metric given by (2.6) and when calculated with respect to the conformally related metric given by (2.5). We notice from (2.6) that $f=\sinh R$ is the area distance along the null geodesics tangent to $\partial/\partial R$, and is thus proportional (with proportionality factor Ω) to the area distance along the null geodesics tangent to $\partial/\partial R$ in the FLRW space-time with line element (2.5).

The Maxwell field (2.15) can be generalized to the field of a charge e_0 with an arbitrary timelike world line as follows. We begin by generalizing (2.6) to a form in which $R=0$ is not necessarily a geodesic. This generalization is given by

$$ds_1^2=f^2p^{-2}(dX^2+dY^2)-2\,du\,dr -(1-2hff'+2hf^2g)du^2. \qquad (2.16)$$

In general (2.16) is conformally related to (2.6) (see the Appendix). However, only if $r=0$ is a timelike geodesic of (2.16) are the null hypersurfaces $u=$ const and the null geodesics tangent to $\partial/\partial r$, the same as the null hypersurfaces $v=$ const and the null geodesics tangent to $\partial/\partial R$ in (2.6). Also, in (2.16) $f=\sinh r$ and $f'=df/dr$. We will take $g=+1$, but in fact (2.16) is conformally flat and thus conformally related to (2.6), for any $g(u)$. In addition,

$$p=v^4(u)(1+\tfrac{1}{4}(X^2+Y^2))-v^3(u)(1-\tfrac{1}{4}(X^2+Y^2)) -v'(u)X-v^2(u)Y, \qquad (2.17)$$

with the four-velocity components $v^i(u)$, $i=1,2,3,4$, arbitrary functions of u except for the condition

$$(v^1)^2+(v^2)^2+(v^3)^2-(v^4)^2=-1. \qquad (2.18)$$

Finally

$$h=\frac{\partial}{\partial u}(\log p). \qquad (2.19)$$

The line $r=0$ in (2.16) has the unit timelike tangent or four-velocity $\partial/\partial u$, and the magnitude of the four-acceleration squared is

$$p^2\left[\left(\frac{\partial h}{\partial X}\right)^2+\left(\frac{\partial h}{\partial Y}\right)^2\right]+h^2 =(\dot v^1)^2+(\dot v^2)^2+(\dot v^3)^2-(\dot v^4)^2, \qquad (2.20)$$

where $\dot v^i=dv^i/du$, $i=1,2,3,4$. If this four-acceleration vanishes we can choose $v^i=\delta^i_4$, and (2.16) reduces to (2.6). The coordinate transformation

$$\sigma=e^u,\quad \sigma+2w=e^{u+2r} \qquad (2.21)$$

applied to (2.16) results in

$$w^2P^{-2}(dX^2+dY^2)-2\,dw\,d\sigma-(1-2Hw)d\sigma^2 =\Phi^2\,ds_1^2, \qquad (2.22)$$

where

$$P(\sigma,X,Y) =V^4(\sigma)(1+\tfrac{1}{4}(X^2+Y^2)) -V^3(\sigma)(1-\tfrac{1}{4}(X^2+Y^2))-V^1(\sigma)X-V^2(\sigma)Y, \qquad (2.23)$$

with

$$(V^1)^2+(V^2)^2+(V^3)^2-(V^4)^2=-1, \qquad (2.24)$$

and

$$P(e^u,X,Y)=p(u,X,Y), \qquad (2.25)$$

with

$$V^i(e^u)=v^i(u), \qquad (2.26)$$

so that p in (2.25) coincides with p in (2.17). Also

$$H=\frac{\partial}{\partial\sigma}(\log P)=e^{-u}h, \qquad (2.27)$$

with h given by (2.19). The function $\Phi=e^{u+r}$ in (2.22), and the left-hand side of (2.22) is a form of the line element of Minkowskian space-time given by Newman and Unti.[12]

The potential one-form

$$A=-e_0(\theta-h)du \qquad (2.28)$$

is a solution of Maxwell's source-free equations on the space-time with line element (2.16); consequently, on account of the conformal invariance of Maxwell's equations, it is a solution of Maxwell's equations on the FLRW space-time (2.1). In (2.28) $\theta=f^{-1}f'$ is the expansion of the null geodesics, of the space-time (2.16), tangent to $\partial/\partial r$. The coordinate r in (2.16) is an affine parameter along these null geodesics. We note in passing that the Liénard–Wiechert solution of Maxwell's equations on Minkowskian space-time can be put in the form (2.28)—a result that is implicit in the work of Robinson and Trautman.[13]

The electromagnetic field corresponding to (2.28) is given by the two-form

$$F=\frac{e_0}{f^2}\,\omega^3\wedge\omega^4 -\frac{e_0}{f}\left\{p\frac{\partial h}{\partial X}\,\omega^1\wedge\omega^4+p\frac{\partial h}{\partial Y}\,\omega^2\wedge\omega^4\right\}, \qquad (2.29)$$

with

$$\omega^1=\omega_1=fp^{-1}\,dX=E_{(1)i}\,dz^i, \qquad (2.30a)$$

$$\omega^2=\omega_2=fp^{-1}\,dY=E_{(2)i}\,dz^i, \qquad (2.30b)$$

$$\omega^3=-\omega_4=-dr-(c/2)du=L_i\,dz^i, \qquad (2.30c)$$

$$\omega^4=-\omega_3=-du=K_i\,dz^i, \qquad (2.30d)$$

where $z^i=(X,Y,r,u)$, $c=1-2hff'+2hf^2$, and the final equalities in (2.30) define a half-null tetrad $E^{\,i}_{(1)}$, $E^{\,i}_{(2)}$, $K^{\,i}$, $L^{\,i}$, which we will make use of later.

The electromagnetic field (2.29) is singular on the accelerated world line $r=0$ in the space-time (2.16). We consider this the history of the charge e_0. In (2.29) we have a "peeling expansion" in integral powers of f^{-1}, which terminates. We see that f is the area distance along the null geodesics tangent to $\partial/\partial r$ in the space-time (2.16). It is therefore proportional to the area distance along these null geodesics in the FLRW space-time. For large values of the affine parameter r we see that the field (2.29) becomes an infinite series in e^{-r}. Hence the conventional peeling-off property does not hold if the affine parameter distance is used, in contradistinction to the case of a Liénard–Wiechart field in Minkowskian space-time [cf. (3.15) below].

The case $k = +1$ of (2.3) can be dealt with in a fashion similar to the case $k = -1$ given above. Now the electromagnetic field of the charged particle has the same form as (2.29) but with $f = \sin r$. Thus the field is singular at $r = 0$ and at the conjugate point $r = \pi$. Near such a place of refocusing a peeling expansion will break down, and a near-field approximation will take over.

The f^{-1} part of (2.29) contains an arbitrary dependence on the null coordinate u through the function h, which is typical of information-carrying waves (see, for example, Trautman[14]). This is the part of (2.29) describing the electromagnetic radiation in the field of the accelerated charge. The arbitrary dependence on u follows from the arbitrariness of the timelike world line $r = 0$. The function h is given by (2.19) with (2.17) and involves the unspecified four-velocity of $r = 0$ and its four-acceleration. Thus h depends upon three independent arbitrary functions of u and their first derivatives.

An alternative approach to the electromagnetic field of a charged particle in space-times with line elements of the form (2.3) has been given by Katz,[15] while an extensive study of Maxwell fields in general on Robertson–Walker universes has been carried out by Infeld and Schild.[10,16] None of these authors has made use of the geometry of the null cones emanating from the world line of a charged particle as we have done here, and which we exploit, following the example of Goldberg and Kerr,[8] in dealing with extended sources in Sec. IV. For further discussions on the conformal invariance of Maxwell's equations see Fulton, Rohrlich, and Witten,[17] and Penrose and Rindler.[18]

III. MAPPINGS OF ELECTROMAGNETIC FIELDS

We will set up a mechanism for mapping a solution of Maxwell's vacuum field equations on Minkowskian space-time to a solution of the source-free Maxwell equations on the FLRW space-time (2.1). In doing this we shall exploit the conformal invariance of Maxwell's equations.

Let $\{X^i\}$, $i = 1,2,3,4$, be rectangular Cartesian coordinates and time in Minkowskian space-time. In these coordinates the metric tensor of Minkowskian space-time has components $\eta_{ij} = \mathrm{diag}(1,1,1,-1)$. Consider the Newman–Unti[12] transformation

$$X^i = x^i(\sigma) + wk^i, \tag{3.1}$$

with

$$k^i = P^{-1}(X,Y,1-\tfrac{1}{4}(X^2+Y^2),1+\tfrac{1}{4}(X^2+Y^2)), \tag{3.2}$$

and $P(\sigma,X,Y)$ given implicitly by

$$\eta_{ij}k^iV^j = -1, \tag{3.3}$$

where $V^j(\sigma) = dx^j/d\sigma$ is the four-velocity of the timelike line $w = 0$, or $X^i = x^i(\sigma)$, and thus

$$\eta_{ij}V^iV^j = -1. \tag{3.4}$$

The function P appearing here is thus the same function as that given in (2.23) above. If the transformation (3.1) is combined with the transformation (2.21) linking σ,w with the coordinates u,r, we obtain

$$dX^i = e^u l^i\, du + \Phi^2 e^{-u} k^i (dr + (c/2)du) + \Phi f p^{-1}(e^i_{(1)}\, dX + e^i_{(2)} dY). \tag{3.5}$$

Here $f = \sinh r$ and $\Phi = e^{u+r}$ as before, p is related to P according to (2.25), $c = 1 - 2hff' + 2hf^2$ with h constructed from P via (2.27), and the vectors k^i, l^i, $e^i_{(A)}$, $A = 1,2$, are given by (3.2) and

$$l^i = V^i - \tfrac{1}{2}k^i, \quad e^i_{(1)} = P\frac{\partial k^i}{\partial X}, \quad e^i_{(2)} = P\frac{\partial k^i}{\partial Y}. \tag{3.6}$$

These vectors in Minkowskian space-time are parallel transported along the null geodesics tangent to $\partial/\partial w = k^i(\partial/\partial X^i)$. The coordinate w is an affine parameter along these null geodesics. We have

$$\eta_{ij}k^il^j = -1, \quad \eta_{ij}e^i_{(A)}e^j_{(B)} = \delta_{A,B}, \tag{3.7}$$

with all other scalar products among k^i, l^i, $e^i_{(A)}$ (with respect to the Minkowskian metric η_{ij}) vanishing. It follows now directly from (3.5) that

$$\eta_{ij}\, dX^i\, dX^j = \Phi^2\, ds_1^2, \tag{3.8}$$

with ds_1^2 given by (2.16) and $g = +1$. Using (3.5) and the one-forms $\{\omega^a\}$, $a = 1,2,3,4$, defined by (2.30), we find

$$dX^i \wedge dX^j = 2\Phi e^u l^{[i}e^{j]}_{(A)}\,\omega^A \wedge \omega^4 + 2\Phi^2\{k^{[i}l^{j]}\omega^3\wedge\omega^4 + e^{[i}_{(1)}e^{j]}_{(2)}\,\omega^1\wedge\omega^2\} + 2\Phi^3 e^{-u}k^{[i}e^{j]}_{(A)}\,\omega^A\wedge\omega^3. \tag{3.9}$$

Here square brackets denote skew symmetrization. When we multiply the left-hand side of (3.9) by a bivector on Minkowskian space-time, having components $F_{ij} = -F_{ji}$ in coordinates $\{X^i\}$, we obtain a two-form defined on the flat space-time with line element given by the right-hand side of (3.8). If the bivector is a solution of Maxwell's vacuum field equations on Minkowskian space-time, then the two-form will be a solution of Maxwell's source-free equations on the space-time with line element ds_1^2. Since this space-time is conformally related to the FLRW space-time, the two-form will be a solution of Maxwell's source-free equations on the FLRW background. We note three important properties of the two-forms (3.9).

(1) The one-forms $\{\omega^a\}$ defined in (2.30) are covariantly constant along the null geodesics of ds_1^2 tangent to $\partial/\partial r$, and thus the two-forms $\{\omega^a\wedge\omega^b\}$ appearing in (3.9) are covariantly constant along these curves.

(2) The bivectors $l^{[i}e^{j]}_{(A)}$, $k^{[i}l^{j]}$, $e^{[i}_{(1)}e^{j]}_{(2)}$, and $k^{[i}e^{j]}_{(A)}$ on Minkowskian space-time, in coordinates $\{X^i\}$, are covariantly constant along the null geodesics tangent to $k^i(\partial/\partial X^i) = \partial/\partial w$.

(3) In an instantaneous rest frame of $w = 0$ at fixed σ [and therefore fixed u, on account of (2.21)] we have

$$V^i = \delta^i_4, \quad P = 1 + \tfrac{1}{4}(X^2+Y^2), \tag{3.10}$$

and in this frame

$$|k^i| \leqslant 1, \quad |l^i| \leqslant \tfrac{1}{2}, \quad |e^i_{(A)}| \leqslant 1. \tag{3.11}$$

It follows from (2) that if the bivectors mentioned there are contracted with bivectors on Minkowskian space-time whose components are given in coordinates $\{X^i\}$ and which are covariantly constant along $\partial/\partial w$, then they will give us scalar functions, on the flat space-time with line element $\Phi^2\, ds_1^2$, which are independent of r. We finally note from (2.21) that we can write

$$w = \Phi f, \tag{3.12}$$

with $f = \sinh r$ and $\Phi = e^{u+r}$, thus emphasizing that the half-line $w = 0$, $0 \leqslant \sigma < +\infty$, in Minkowskian space-time is mapped via (2.21) to the line $r = 0$, $-\infty < u < +\infty$, in the space-time line element ds_1^2 given by (2.16) with $g = +1$.

We will now use the charged particle example of the previous section to illustrate the mechanism for passing from a solution of Maxwell's vacuum field equations on Minkowskian space-time to a solution of Maxwell's source-free equations on the space-time with line element ds_1^2, and thus to a solution on the FLRW space-time. This mechanism will be applied to extended sources in Sec. IV.

The Liénard–Wiechert potential one-form of a particle with charge e_0 having world line $w = 0$ in Minkowskian space-time is given, using coordinates $\{X^i\}$, by

$$A = e_0 (V_i\, dX^i / w), \tag{3.13}$$

with V^i the four-velocity of the charge appearing in (3.3) and (3.4) above. Substituting for dX^i from (3.5) and using the scalar products of V^i with k^i, l^i, $e^i_{(A)}$ given by (3.3), (3.4), and (3.6), we easily find that (3.13) may be rewritten as

$$A = -e_0(\theta - h)du, \tag{3.14}$$

modulo an exact differential. Here $\theta = f^{-1} f'$, and h is given by (2.19). We see that we have here recovered (2.28). The electromagnetic field corresponding to (3.13) has components F_{ij}, in coordinates $\{X^i\}$, and

$$F_{ij} = w^{-1} N_{ij} + w^{-2}\, \mathrm{III}_{ij}, \tag{3.15}$$

with

$$N_{ij} = e_0 (k_j D V_i - k_i D V_j), \tag{3.16}$$

$$\mathrm{III}_{ij} = e_0 (k_j V_i - k_i V_j), \tag{3.17}$$

where

$$D = \frac{\partial}{\partial \sigma} + \eta_{ij} a^i k^j, \tag{3.18}$$

and $a^i = dV^i/d\sigma$ is the four-acceleration of the charge. We note that N_{ij} and III_{ij} are covariantly constant along $\partial/\partial w$ and satisfy the algebraic conditions

$$N_{ij} k^j = 0, \quad \mathrm{III}_{ij} k^j = e_0 k_i. \tag{3.19}$$

Multiplying (3.9) by N_{ij} and using the scalar products (3.3), (3.4), and (3.7) we find

$$N_{ij}\, dX^i \wedge dX^j = 2\Phi e^u N_{ij} l^i e^j_{(A)} \omega^A \wedge \omega^4, \tag{3.20}$$

and thus using (3.12) we have

$$w^{-1} N_{ij}\, dX^i \wedge dX^j = 2 f^{-1} e^u N_{ij} l^i e^j_{(A)} \omega^A \wedge \omega^4. \tag{3.21}$$

From P in (2.23) and k^i in (3.2) it is straightforward to show that

$$\frac{\partial}{\partial \sigma} (\log P) = -\eta_{ij} a^i k^j. \tag{3.22}$$

Using this and the scalar products (3.3) and (3.7) we can write

$$N_{ij} l^i e^j_{(A)} = -e_0 P \frac{\partial H}{\partial x^A} = -e_0 p e^{-u} \frac{\partial h}{\partial x^A}, \tag{3.23}$$

with the last equality obtained from (2.25) and (2.27), and with $x^A = (X, Y)$. Thus (3.21) becomes

$$\frac{1}{2} w^{-1} N_{ij}\, dX^i \wedge dX^j = -e_0 f^{-1} p \frac{\partial h}{\partial x^A} \omega^A \wedge \omega^4. \tag{3.24}$$

A similar calculation involving multiplying (3.9) by III_{ij} in (3.17) yields

$$\tfrac{1}{2} w^{-2} \mathrm{III}_{ij}\, dX^i \wedge dX^j = e_0 f^{-2} \omega^3 \wedge \omega^4. \tag{3.25}$$

Putting (3.23) and (3.25) together we recover the Maxwell field (2.29) on the FLRW space-time due to an accelerating charge e_0.

IV. CHARGE-CURRENT DISTRIBUTION

The procedure, described and illustrated in the previous section, for mapping a solution of Maxwell's vacuum field equations on Minkowskian space-time to a solution of Maxwell's source-free equations on the FLRW background will now be applied to the asymptotic electromagnetic field outside a bounded charge-current distribution given by Goldberg and Kerr.[8]

The Goldberg–Kerr result is in two parts. The first part consists in showing that the electromagnetic field F_{ij}, in coordinates $\{X^i\}$ on Minkowskian space-time, due to a distribution of the four-current J^i confined to a timelike world tube, takes the form

$$F_{ij} = w^{-1} N_{ij} + w^{-2} \mathrm{III}_{ij} + O(w^{-3}), \tag{4.1}$$

where $w = 0$, given by $X^i = x^i(\sigma)$ [cf. (3.1)], is an arbitrary timelike world line satisfying the condition that $J^i(X) = 0$ for all X^i such that

$$\eta_{ij}(X^i - x^i(\sigma))(X^j - x^j(\sigma)) \geqslant \max(0, a^2 - w^2), \tag{4.2}$$

where $a(\sigma)$ is some positive function, and

$$N_{ij} = 2 k_{[j} D \mathscr{J}_{i]}, \tag{4.3}$$

$$\mathrm{III}_{ij} = 2 D \mathscr{J}_{[j;i]} + 2 V_{[i} \mathscr{J}_{j]} + 2 k_{[i} B_{j]}. \tag{4.4}$$

Here k^i, V^i are those vectors appearing in (3.2)–(3.4), D is given by (3.18),

$$B_i = V_j (\kappa + 2D) \mathscr{J}_{i:}{}^{j} - \mathscr{J}_i + \tfrac{1}{2}(\kappa + D) D \mathscr{J}_{i:j}{}^{j} + a_j \mathscr{J}_{i:}{}^{j}, \tag{4.5}$$

$$\mathscr{J}_i = \int_{\mathscr{H}} J_i\, d\Omega, \tag{4.6}$$

and

$$\mathscr{J}_{i:j_1 j_2 \cdots j_n} = \int_{\mathscr{H}} J_i \zeta_{j_1} \zeta_{j_2} \cdots \zeta_{j_n}\, d\Omega \quad (n \geqslant 1). \tag{4.7}$$

In (4.5) $\kappa = \eta_{ij} a^i k^j$, and in (4.6) and (4.7) $\mathscr{H}$ is the null hypersurface $\{(\xi^i): \eta_{ij} k^i \zeta^j = 0, \zeta^j = \xi^j - x^j(\sigma)\}$ while $d\Omega$ is the invariant measure on $\mathscr{H}$. The moments (4.7) depend only on σ and the direction of k^i, and thus are parallel transported along k^i. All the other vectors appearing in (4.3)–(4.6) are parallel along k^i, and consequently the scalar κ

appearing in (4.5) and (3.18) is constant along k^i. Thus it follows that N_{ij} and III_{ij} in (4.3) and (4.4) are parallel propagated along k^i. Goldberg and Kerr prove that N_{ij} and III_{ij} above satisfy the algebraic conditions

$$N_{ij}k^j = 0, \quad \mathrm{III}_{ij}k^j = -A_0 k_i, \tag{4.8}$$

with

$$A_0 = D\mathscr{F}_{i:}{}^i + V_i\mathscr{F}^i. \tag{4.9}$$

We notice that in the charged particle limit in which

$$\mathscr{F}_i = e_0 V_i, \quad \mathscr{F}_{i:j_1j_2\cdots j_n} = 0 \quad (n \geqslant 1), \tag{4.10}$$

Eqs. (4.3) and (4.4) reduce to Eqs. (3.16) and (3.17), respectively, while the algebraic conditions (4.8) become the algebraic conditions (3.19).

The second part of the Goldberg–Kerr analysis establishes that in an instantaneous rest frame at $\sigma = \text{const}$,

$$F_{ij} = \sum_{s=0}^{n-1} w^{-s-1}\,{}_sf_{ij} + w^{-n-1}\,{}_nJ_{ij} \quad (n \geqslant 1), \tag{4.11}$$

and the remainder ${}_nJ_{ij}$ is bounded according to

$$|{}_nJ_{ij}| \leqslant \frac{8\pi}{n!}\sum_{r=0}^{n} M_r A^n_r\left(1 - \frac{a}{w}\right)^{-n-1}\frac{a^{n+r+3}}{n+r+3}. \tag{4.12}$$

In (4.11) ${}_0f_{ij} = N_{ij}$ and ${}_1f_{ij} = \mathrm{III}_{ij}$ given by (4.3) and (4.4) and calculated in the instantaneous rest frame at $\sigma = \text{constant}$. In (4.12) $a(\sigma)$ is given by (4.2) and is now constant, while the $M_r(\sigma)$, $r = 0,1,2,\ldots,n$, are constants depending on σ which are bounds on the absolute values of the rth derivative of $J_{[i,j]}(X)$, for X^i in the compact domain that is the complement of the region specified by (4.2) with $\sigma = \text{const}$. The quantities A^n_r, $r = 0,1,2,\ldots n$, are constants which are solutions of the recurrence formula

$$A^{n+1}_r = (n+1)A^n_r + 2A^n_{r-1} \quad (r \geqslant 1), \tag{4.13}$$

with $A^n_0 = n!$ and $A^n_n = 2^n$. Thus ${}_nJ_{ij}$ is asymptotically $O(1)$ for $n = 1,2,3,\ldots$, and in this sense the existence of the asymptotic expansion (4.1) is ensured.

For the accuracy indicated in (4.1) we can put $n = 2$ in (4.11) so that it reads

$$F_{ij} = w^{-1}N_{ij} + w^{-2}\mathrm{III}_{ij} + w^{-3}\,{}_2J_{ij}, \tag{4.14}$$

with the proviso that this equation holds only in an instantaneous rest frame when $\sigma = \text{const}$. Now ${}_2J_{ij}$ is bounded according to (4.12) with $n = 2$.

With $\Phi = e^{u+r}$ and $f = \sinh r$ we can write

$$\Phi = e^u f\{1 + (1 + f^{-2})^{1/2}\}. \tag{4.15}$$

Now multiplying (4.11) by (3.9) and using (3.12) we find that, for large $r > 0$,

$$\tfrac{1}{2}F_{ij}\,dX^i\wedge dX^j = f^{-1}\underset{\circ}{N} + f^{-2}\underset{\circ}{\mathrm{III}} + O(f^{-3}), \tag{4.16}$$

with

$$\underset{\circ}{N} = \underset{\circ}{N}_A\,\omega^A\wedge\omega^4, \quad \underset{\circ}{N}_A = e^u N_{ij}l^i e^j_{(A)}, \tag{4.17}$$

and

$$\underset{\circ}{\mathrm{III}} = \mathrm{III}_{ij}k^i l^j\omega^3\wedge\omega^4 + \mathrm{III}_{ij}e^i_{(1)}e^j_{(2)}\omega^1\wedge\omega^2. \tag{4.18}$$

In terms of the vectors $E^i_{(A)}$, K^i, L^i, introduced in (2.30), on the space-time with line element ds_1^2 given by (2.16), with $g = +1$, the components $\underset{\circ}{N}$, $\underset{\circ}{\mathrm{III}}$ [in coordinates $Z^i\,(X,Y,r,u)$] are given by

$$\underset{\circ}{N}_{ij} = 2\underset{\circ}{N}_A E^A_{[i}K_{j]}, \tag{4.19}$$

and

$$\underset{\circ}{\mathrm{III}}_{ij} = 2\underset{\circ}{\mathrm{III}}^{(1)}L_{[i}K_{j]} + 2\underset{\circ}{\mathrm{III}}^{(2)}E^{(1)}_{[i}E^{(2)}_{j]}, \tag{4.20}$$

where $\underset{\circ}{\mathrm{III}}^{(1)}$ and $\underset{\circ}{\mathrm{III}}^{(2)}$ are the coefficients of $\omega^3\wedge\omega^4$ and $\omega^1\wedge\omega^2$, respectively, in (4.18). Now it follows immediately that N_{ij} and III_{ij} satisfy

$$\underset{\circ}{N}_{ij}K^j = 0, \quad \underset{\circ}{\mathrm{III}}_{ij}K^j = \underset{\circ}{\mathrm{III}}^{(1)}K_i\,. \tag{4.21}$$

Thus the coefficients of the two leading terms in the expansion (4.16) of the asymptotic Maxwell field on the space-time with line element ds_1^2 (and thence on the FLRW space-time) satisfy algebraic conditions similar to those satisfied in (4.8) by the corresponding terms in the Goldberg–Kerr asymptotic expansion.

Turning next to the Goldberg–Kerr expansion (4.14), with a remainder ${}_2J_{ij}$, which holds at $\sigma = \text{const}$, we find that this translates, using (3.9), into the following expansion at $u = \text{const}$:

$$\begin{aligned}
\tfrac{1}{2}F_{ij}\,dX^i\wedge dX^j
&= f^{-1}e^u N_{ij}l^i e^j_{(A)}\,\omega^A\wedge\omega^4\\
&\quad + f^{-2}(\mathrm{III}_{ij}k^i l^j\omega^3\wedge\omega^4 + \mathrm{III}_{ij}e^i_{(1)}e^j_{(2)}\omega^1\wedge\omega^2)\\
&\quad + f^{-3}([1 + (1 + f^{-2})^{1/2}]^{-1}\mathrm{III}_{ij}l^i e^j_{(A)}\omega^A\wedge\omega^4\\
&\quad + e^{-u}\,{}_2J_{ij}k^i e^j_{(A)}\omega^A\wedge\omega^3)\\
&\quad + f^{-4}e^{-u}[1 + (1 + f^{-2})^{1/2}]^{-1}({}_2J_{ij}k^i l^j\omega^3\wedge\omega^4\\
&\quad + {}_2J_{ij}e^i_{(1)}e^j_{(2)}\omega^1\wedge\omega^2)\\
&\quad + f^{-5}e^{-u}[1 + (1 + f^{-2})^{1/2}]^{-2}\,{}_2J_{ij}l^i e^j_{(A)}\omega^A\wedge\omega^4.
\end{aligned} \tag{4.22}$$

Using the bounds on $k^i, l^i, e^i_{(A)}$ given in (3.11) and those on ${}_2J_{ij}$ (and ${}_1J_{ij}$) given by (4.12), it is straightforward now to calculate bounds on the coefficients of $\omega^a\wedge\omega^b$ here, and thus to see that when $u = \text{const}$,

$$\tfrac{1}{2}F_{ij}\,dX^i\wedge dX^j = f^{-1}\underset{\circ}{N} + f^{-2}\underset{\circ}{\mathrm{III}} + f^{-3}\,{}_2\underset{\circ}{J}, \tag{4.23}$$

with ${}_2\underset{\circ}{J} = O(1)$ asymptotically. Hence in this sense the existence of the asymptotic expansion (4.16) is guaranteed.

We have established in (4.16) a peeling expansion in integral powers of f^{-1}, and f is the area distance along the null geodesics tangent to $\partial/\partial r$ in the space-time (2.16). In this space-time, using coordinates $Z^i = (X,Y,r,u)$, a half-null tetrad basis of vector fields $E^i_{(1)}, E^i_{(2)}, L^i, K^i$ is defined via the one-forms (2.30). If $F'_{ij}(Z)$ denotes the components of the Maxwell field (4.16) in the coordinates $\{Z^i\}$, then (4.16) can be expressed in Newman–Penrose[3] notation as

$$\begin{aligned}
\phi_2 &= \sqrt{2}\,F'_{ij}L^i(E^j_{(1)} + i\,E^j_{(2)})\\
&= \sqrt{2}(\underset{\circ}{N}_1 + i\underset{\circ}{N}_2)f^{-1} + O(f^{-3}),
\end{aligned} \tag{4.24a}$$

$$\begin{aligned}
\phi_1 &= -F'_{ij}(K^iL^j + iE^i_{(1)}E^j_{(2)})\\
&= -(\underset{\circ}{\mathrm{III}}^{(1)} + i\,\underset{\circ}{\mathrm{III}}^{(2)})f^{-2} + O(f^{-3}),
\end{aligned} \tag{4.24b}$$

$$\phi_0 = \sqrt{2}\,F'_{ij}K^i(E^j_{(1)} - i\,E^j_{(2)}) = O(f^{-3}). \tag{4.24c}$$

On account of properties (1) and (2) above of the two-forms (3.9) [see the remark following Eq. (3.11)] and the fact that the half-null tetrad $E^i_{(1)}, E^i_{(2)}, L^i, K^i$ is parallel transported along the intergral curves of $K^i\,\partial/\partial Z^i = \partial/\partial r$, the coefficients of the inverse powers of f in (4.24) are functions only of (X,Y,u).

V. CONCLUSIONS

The use of an arbitrary timelike world line $r=0$ on which to base the coordinate system $Z^i(X,Y,r,u)$, in the space-time with line element ds_1^2 given by (2.16), is very general but does not relate in a simple way to the usual FLRW coordinates and observations. The FLRW universe contains a privileged set of fundamental observers with geodesic world lines given by the t lines of (2.1), and $r=0$ in (2.16) does not correspond to one of them. As a consequence of this the future null cones $u=\text{const}$ of (2.16), with vertices on $r=0$, do not correspond to the future null cones with vertices on the world line of a fundamental observer in the FLRW space-time. Hence red shifts and area distances, for example, in (2.16) are not related in a simple way to the standard expressions for these in the FLRW space-time. In this regard we note that the conformal factor relating (2.16) to the FLRW space-time (2.1) is not simply Ω in (2.2), if $r=0$ is arbitrary, but $\Omega\Psi^{-1}$, with Ψ given by Eqs. (A2) and (A9) of the Appendix.

To relate our results easily to observations of a fundamental observer in a FLRW universe, one should choose the central world line $r=0$ in a particular way, namely, as one of the fundamental world lines of the universe. Making $r=0$ a timelike geodesic in (2.16) means that $h=0, p=1+\frac{1}{4}(X^2+Y^2)$, and so $ds_1^2=ds_0^2$ with ds_0^2 given by (2.6), or (2.3) with $k=-1$. Now the expansion parameter f^{-1} in (4.16) or (4.24) is such that Ωf is the area distance along the null geodesics tangent to $\partial/\partial r$ in the FLRW universe. In (4.24) we may regard $F'_{ij}(Z)$ as the components of a Maxwell field *on the FLRW space-time* in coordinates $Z^i=(X,Y,r,u)$, because the covariant Maxwell tensor is conformally invariant. However, the Newman–Penrose components of the Maxwell field on the FLRW space-time are obtained from (4.24), *when $r=0$ is a timelike geodesic*, by multiplying the components given in (4.24) by Ω^{-2}. This is because one passes from (2.16), with $r=0$ geodesic, to the FLRW space-time by replacing the half-null tetrad $E^i_{(1)}$, $E^i_{(2)}, L^i, K^i$ by $\Omega^{-1}E^i_{(1)}, \Omega^{-1}E^i_{(2)}, \Omega^{-1}L^i, \Omega^{-1}K^i$. If the resulting Newman–Penrose components of the Maxwell field on the FLRW space-time are denoted $\hat{\phi}_2, \hat{\phi}_1, \hat{\phi}_0$, then the peeling expansion in this space-time can be written

$$\Omega^2\hat{\phi}_2=\sqrt{2}(\overset{\circ}{N}_1+i\overset{\circ}{N}_2)f^{-1}+O(f^{-3}), \tag{5.1a}$$

$$\Omega^2\hat{\phi}_1=-(\overset{\circ}{\mathrm{III}}{}^{(1)}+i\,\overset{\circ}{\mathrm{III}}{}^{(2)})f^{-2}+O(f^{-3}), \tag{5.1b}$$

$$\Omega^2\hat{\phi}_0=O(f^{-3}). \tag{5.1c}$$

The Goldberg–Kerr condition on the localization of the distribution of the four-current J^i, expressed by (4.2), means that J^i should vanish in the complement of the region inside the future and past null cones with vertex on $w=0$ at each fixed σ, and inside the timelike world tube with center on $w=0$ and radius $a(\sigma)$ (measured in the instantaneous rest frame of $w=0$ at $\sigma=\text{const}$). This domain, mapped to the FLRW space-time, is topologically unchanged. The "radius" of the world tube is merely rescaled.

In the paper to follow,[9] the coordinates (x,y,R,v) used in (2.5) and (2.6) are employed to examine the relation of the geometric optics approximation to solutions of Maxwell's equations in these cosmological space-times.

APPENDIX: CONFORMALLY RELATED SPACE-TIMES

We will demonstrate here the conformal relationship between the line element (2.6) and the line element (2.16) with $g=+1$. Beginning with (2.6), which reads

$$ds_0^2=\frac{\sinh^2 R(dx^2+dy^2)}{(1+\frac{1}{4}(x^2+y^2))^2}-2\,dv\,dR-dv^2, \tag{A1}$$

our objective is to demonstrate that

$$ds_0^2=\Psi^2\,ds_1^2, \tag{A2}$$

with ds_1^2 given by (2.16) with $g=+1$, and to find the conformal factor.

We start by making the transformation

$$\bar{v}=e^v,\quad \bar{v}+2\bar{R}=e^{v+2R} \tag{A3}$$

on (A1) to obtain

$$ds_0^2=(\bar{v}^2+2\bar{v}\bar{R})^{-1}\{\bar{R}^2(dx^2+dy^2)/(1+\tfrac{1}{4}(x^2+y^2))^2 - 2\,d\bar{v}\,d\bar{R}-d\bar{v}^2\}. \tag{A4}$$

Now make the transformation

$$(x,y,\bar{R},\bar{v})\rightarrow(X,Y,w,\sigma), \tag{A5}$$

given by

$$\bar{v}\delta^i_4+\bar{R}\bar{K}^i(x,y)=x^i(\sigma)+wk^i(\sigma,X,Y), \tag{A6}$$

$i=1,2,3,4$. We are putting two Newman–Unti[12] transformations together here, with

$$\bar{K}^i=(1+\tfrac{1}{4}(x^2+y^2))^{-1}\times\{x,y,1-\tfrac{1}{4}(x^2+y^2),1+\tfrac{1}{4}(x^2+y^2)\}, \tag{A7}$$

and k^i given by (3.2). Now follow this with the transformation

$$\sigma=e^u,\quad \sigma+2w=e^{u+2r}, \tag{A8}$$

and we finally obtain

$$ds_0^2=e^{2u+2r}/(\bar{v}^2+2\bar{v}\bar{R})\,ds_1^2, \tag{A9}$$

with ds_1^2 given by (2.16) with $g=+1$. Here $\bar{v}$, $\bar{R}$ are given in terms of X, Y, r, u via (A6) and (A8). We have thus established (A2).

If $r=0$ is a geodesic of (2.16), then in (2.16) $h=0$. It thus follows from (2.27) that

$$H=\frac{\partial}{\partial\sigma}(\log P)=0, \tag{A10}$$

where P occurs in (2.23) and (3.2). This implies that $w=0$, or $X^i=x^i(\sigma)$, is a timelike geodesic, and so we can write $x^i(\sigma)=\sigma\delta^i_4$ (since σ is proper time along $w=0$) in (A6). We can then conclude from (A6) that

$$\bar{v}=\sigma,\quad \bar{R}=w,\quad x=X,\quad y=Y, \tag{A11}$$

in this case, and so now the conformal factor in (A9) becomes unity.

[1]S. W. Hawking, J. Math. Phys. 9, 598 (1968).

[2]R. K. Sachs, "Distance and asymptotic behavior of waves in general rela-

tivity," in *Recent Developments in General Relativity* (Pergamon, New York, 1962), p. 395.
[3]E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[4]E. T. Newman and T. W. J. Unti, J. Math. Phys. **3**, 891 (1962).
[5]P. A. Hogan and A. Trautman, "On gravitational radiation from bounded sources," in *Gravitation and Geometry*, edited by W. Rindler and A. Trautman (Bibliopolis, Naples, 1987), p. 215.
[6]H. Bondi, M. G. J. van der Burg, and A. W. K. Metzner, Proc. R. Soc. London Ser. A **269**, 21 (1962).
[7]R. K. Sachs, Proc. R. Soc. London Ser. A **270**, 103 (1962).
[8]J. N. Goldberg and R. P. Kerr, J. Math. Phys. **5**, 172 (1964).
[9]P. A. Hogan and G. F. R. Ellis, "Geometric optics in expanding universes" (unpublished).
[10]L. Infeld and A. Schild, Phys. Rev. **70**, 410 (1946).
[11]A. G. Walker, Mon. Not. R. Astron. Soc. **95**, 263 (1935).
[12]E. T. Newman and T. W. J. Unti, J. Math. Phys. **4**, 1467 (1963).
[13]I. Robinson and A. Trautman, Proc. R. Soc. London Ser. A **265**, 463 (1962).
[14]A. Trautman, "On the propagation of information by waves," in *Recent Developments in General Relativity* (Pergamon, New York, 1962), p. 459.
[15]A. Katz, J. Math. Phys. **9**, 1694 (1968).
[16]L. Infeld and A. Schild, Phys. Rev. **68**, 250 (1945).
[17]T. Fulton, F. Rohrlich, and L. Witten, Rev. Mod. Phys. **34**, 442 (1962).
[18]R. Penrose and W. Rindler, *Spinors and Space-Time*, Vol. 1 (Cambridge U.P., Cambridge, 1984).

ERRATUM

Erratum: New integrable nonlinear integrodifferential equations and related solvable finite-dimensional dynamical systems [J. Math. Phys. 29, 49 (1988)]

Y. Matsuno
Department of Physics, Faculty of Liberal Arts, Yamaguchi University, Yamaguchi 753, Japan

(Received 9 August 1988; accepted for publication 7 September 1988)

The determinantal expression (2.9b) with (2.9c) is incorrect for $N \geqslant 4$. It should be replaced by the following expression:

$$f = \prod_{j=1}^{N}\left[i\theta_j + \frac{a_j}{a_j - 1}\right] + \sum_{n=1}^{[N/2]} \frac{1}{n!2^n} \sum_{j_1, j_2, \ldots, j_{2n}}^{(N)} B_{j_1 j_2} B_{j_3 j_4} \cdots B_{j_{2n-1} j_{2n}} \times \prod_{\substack{k=1 \\ (k \neq j_1, j_2, \ldots, j_{2n})}}^{N} \left[i\theta_k + \frac{a_k}{a_k - 1}\right],$$

with

$$\theta_j = x - a_j t - x_{0j} \quad (j = 1,2,\ldots,N),$$

$$B_{jk} = \frac{2(a_j + a_k)a_j a_k}{(a_j - a_k)^2} \quad (j \neq k, \quad j,k = 1,2,\ldots,N),$$

where the notation

$$\sum_{j_1, j_2, \ldots, j_{2n}}^{(N)}$$

means the summation over all possible combinations of $j_1, j_2, \ldots, j_{2n}$ that are taken from $1,2,\ldots,N$, and $[N/2]$ implies the greatest integer not exceeding $N/2$.

The author thanks Professor R. Hirota for pointing out the error.

 0022-2488/89/010241-01$02.50