# Functions of infinite generalized cyclic matrices

Philippe Audit
*Laboratoire PMTM, Université Paris–Nord, 93430 Villetaneuse, France*

(Received 12 April 1984; accepted for publication 28 September 1984)

Analytic expressions are derived for the elements or the trace of generalized cyclic matrix functions. Application to the representation of the linear difference operators with constant coefficients and periodic boundary conditions is considered, in order to express functions of such operators in closed form.

## I. INTRODUCTION

Much attention is paid nowadays to physical theories based on lattice models. Not to mention solid state theory, problems in $d$ dimensions ($d > 3$, up to infinity) are currently studied in lattice statistics, and increasing use is also made of lattices in field theory to approximate the continuum. The most natural tool for handling such problems is offered by generalized cyclic matrices (i.e., multiple direct products of cylic matrices), and in fact the solutions of many lattice problems are just given by some element or the trace of a generalized cyclic matrix function. Although some interesting methods of calculation for special cases have been proposed[1–4] earlier, to my knowledge, no general theory of such functions had been presented yet.

The lattices to be considered have two fundamental properties: infinite size and periodicity. Both properties have to be fully exploited to guarantee an exact solution to lattice problems. The well-known property of cyclic matrices to admit the primitive roots of unity as eigenvalues allows us, in some simple cases, to diagonalize the finite matrix variable by means of a finite unitary matrix, and then to take the limit over an infinite number of lattice sites. Although this procedure may be correct for some matrix functions, for more reliability and generality, the thermodynamical limit in lattice statistics or the continuum limit in lattice field theory should be obtained from the infinite matrix properties, as will be shown in Sec. II of this paper. In some cases, much simpler analytical expressions of the matrix functions' elements are obtained by taking advantage of the orthogonal properties of Chebyshev polynomials; application of those results to express the resolvent or other functions of a linear difference operator are considered in the last section.

## II. PERIODIC LATTICES AND CYCLIC MATRICES

Let us consider a $d$-dimensional simple cubic lattice $Z^d$, of side $\nu$, with periodic boundary conditions. A finite $s\times s$ matrix $\mathsf{a}_{\mathbf{k}}$ is assigned to each one of the $N = \nu^d$ sites located at $\mathbf{k} = (k_1, k_2, \ldots, k_d)$, $k_i \in z$. Next we introduce the following $sN \times sN$ generalized cyclic matrix

$$\mathsf{A} = \sum_{\mathbf{k}\in K} \mathsf{m}_{k_1} \otimes \mathsf{m}_{k_2} \otimes \cdots \otimes \mathsf{m}_{k_d} \otimes \mathsf{a}_{\mathbf{k}}. \tag{2.1}$$

Here, $\mathsf{m}_i$ is the topological $\nu\times\nu$ cyclic matrix of order $i$, defined by

$$\langle 0|\mathsf{m}_i|l\rangle = \langle 0|\mathsf{m}_i|\nu - l\rangle = \delta(i,l), \tag{2.2}$$

where $\delta(i,l)$ is the Krönecker symbol and $l$ is the column's index varying from 0 to $\nu - 1$. The matrix $\mathsf{m}_i$ has a single nonzero element in each row, and as for any cyclic matrix we shall always refer to the entries of the line 0, the entries of the other lines being obtained by cyclic permutation. The summation in Eq. (2.1) is made over a finite subset of neighbors of the site **0** at the origin: $K = \{\mathbf{k}:|\mathbf{k}|\leqslant n\} \subset Z^d$.

For application in solid state physics, the points of the lattice can be thought of as equilibrium locations of the atoms in a crystal. Then the matrix A can stand for a real space dynamical matrix, the representation of a tight-binding or a Heisenberg Hamiltonian, etc.; so that the entries of $\mathsf{a}_{\mathbf{k}}$ can represent couplings between sites **0** and **k**, e.g., force constants, hopping integrals between orbitals, exchange integrals between localized spins, etc. The value of $s$ is fixed by the nature of the interaction and the number of interacting species associated with each lattice site; $n$ is the range of interaction.

For application to classical fields the lattice $Z^d$ is just an approximation to the continuum $R^d$, where the partial differential operators take the discretized form of lattice operators. In this case the $a_k$'s are scalars, whose values are determined by the derivative's order and the lattice spacings in each coordinate direction; a precise definition will be given in Sec. V. [See Eqs. (5.7) and (5.8).] In this case the value of $n$ depends on the differential operator's order.

The $l_i$ entries of the $s\nu^d \times s\nu^d$ generalized cyclic matrix A are $s\nu^{d-1} \times s\nu^{d-1}$ generalized cyclic matrices, whose entries $l_j$ are $s\nu^{d-2} \times s\nu^{d-2}$ generalized cyclic matrices, etc. It is then convenient to use the components of the vector $\mathbf{l} = (l_1, l_2, \ldots, l_d)$ to label the $s\times s$ noncyclic matrices, which are the entries of A, in the form $\langle 0\cdots 0|\mathsf{A}|l_1\cdots l_d\rangle$ or more concisely $\langle \mathbf{0}|\mathsf{A}|\mathbf{l}\rangle$.

It is worth pointing out that the definition (2.1) where the sum is taken over all the elements of $K$, concerns the simple hypercubical lattice; but restrictions on the $k_i$'s allowed values lead to the whole class of hypercubical lattices. For example, in the case $d = 3$, a fcc lattice with second-neighbor interactions corresponds to the set $K = \{(0, 0, 0), (\pm 1, \pm 1, 0), (0, \pm 1, \pm 1), (\pm 1, 0, \pm 1), (\pm 2, 0, 0), (0, \pm 2, 0), (0, 0, \pm 2)\}$.

Now we summarize the topological matrices' properties to be used in this article. The $\nu\times\nu$ matrices $\mathsf{m}_i$ are commutative and satisfy the following identities:

$$\mathsf{m}_0 = \mathsf{m}_\nu = \mathsf{1}_\nu, \tag{2.3}$$

$$\mathsf{m}_i\mathsf{m}_j = \mathsf{m}_j\mathsf{m}_i = \mathsf{m}_{i+j}, \tag{2.4}$$

$$(\mathsf{m}_i)^p = \mathsf{m}_{ip}. \tag{2.5}$$

Hereafter, we start by considering the general or *anisotropic case*, where the components $\mathsf{a}_{\mathbf{k}}$ of the matrix A in Eq.

 0022-2488/85/030361-04$02.50 © 1985 American Institute of Physics 

(2.1) vary with the bond direction $\mathbf{k}$ (an example is offered by the hopping integrals between $p$ or $d$ orbitals, in the matrix representation of a tight-binding Hamiltonian). In Sec. IV we study the *isotropic case*, where $\mathsf{a}_{\mathbf{k}} = \mathsf{a}_{|\mathbf{k}|}$ (corresponding to $s$ orbitals in the previous example); use of the obtained results is made in Sec. V to treat scalar difference operators.

## III. ANISOTROPIC CASE

To proceed with the derivation of the elements of some matrix function $f(\mathsf{A})$, we need to estimate first the matrix $\mathsf{A}^p$, for a positive integer $p$. Repeated use of the rules (2.4) and (2.5) gives

$$\mathsf{A}^p = \sum_{\Sigma p_{\mathbf{k}} = p} \mathsf{m}_{\sum_{\mathbf{k}} k_1 p_{\mathbf{k}}} \otimes \cdots \otimes \mathsf{m}_{\sum_{\mathbf{k}} k_d p_{\mathbf{k}}} \otimes \overline{\prod_{\mathbf{k}\in K} (\mathsf{a}_{\mathbf{k}})^{p_{\mathbf{k}}}}, \tag{3.1}$$

where, by convention, the overlined factor represents the sum of products in all possible orders. For instance,

$$\overline{\mathsf{a}_1^2 \mathsf{a}_2} = \mathsf{a}_1^2 \mathsf{a}_2 + \mathsf{a}_2 \mathsf{a}_1^2 + \mathsf{a}_1 \mathsf{a}_2 \mathsf{a}_1. \tag{3.2}$$

Recalling the relation (2.2) and the direct product properties, the matrix element $\mathbf{l}$ can be written as

$$\langle \mathbf{0}|\mathsf{A}^p|\mathbf{l}\rangle = \sum_{\Sigma p_{\mathbf{k}} = p} \delta\Big(\sum \mathbf{k} p_{\mathbf{k}}, \mathbf{l}\Big) \overline{\prod_{\mathbf{k}\in K} (\mathsf{a}_{\mathbf{k}})^{p_k}}. \tag{3.3}$$

To proceed, we replace the Krönecker symbol by its Fourier representation

$$(2\pi)^{-d} \int_{-\pi}^{\pi} \exp\Big[ -i\Big(\sum_{\mathbf{k}\in K} p_{\mathbf{k}} \mathbf{k} - \mathbf{l}\Big)\cdot\boldsymbol{\theta}\Big]\, d\boldsymbol{\theta} = \delta\Big(\sum_{\mathbf{k}} \mathbf{k} p_{\mathbf{k}}, \mathbf{l}\Big), \tag{3.4}$$

in terms of the $d$-dimensional vector

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d), \tag{3.5}$$

and obtain

$$\begin{aligned} \langle \mathbf{0}|\mathsf{A}^p|\mathbf{l}\rangle &= (2\pi)^{-d} \int_{-\pi}^{\pi} \sum_{\Sigma p_{\mathbf{k}} = p} \overline{\prod_{\mathbf{k}\in K} (\mathsf{a}_{\mathbf{k}} e^{-i\mathbf{k}\cdot\boldsymbol{\theta}})^{p_{\mathbf{k}}}} \\ &\quad \times e^{i\mathbf{l}\cdot\boldsymbol{\theta}}\, d\boldsymbol{\theta} \\ &= (2\pi)^{-d} \int_{-\pi}^{\pi} (\mathsf{a}(\boldsymbol{\theta}))^p e^{i\mathbf{l}\cdot\boldsymbol{\theta}}\, d\boldsymbol{\theta} \qquad (p < \nu/n), \end{aligned} \tag{3.6}$$

where

$$\mathsf{a}(\boldsymbol{\theta}) = \sum_{\mathbf{k}\in K} \mathsf{a}_{\mathbf{k}} e^{-i\mathbf{k}\cdot\boldsymbol{\theta}}. \tag{3.7}$$

Equation (3.6) is valid provided that $p < \nu/n$; otherwise in the summation (3.1) some topological cyclic matrix would become a unit matrix, because, according to Eqs. (2.3)–(2.5), if $p = \nu/n$ we have $(\mathsf{m}_n)^p = 1_\nu$.

Hence, for any function expressible as a power series

$$f(\mathsf{A}) = \sum_{p=0}^{\infty} \alpha^p \mathsf{A}^p,$$

the matrices $\mathsf{A}^p$ can be evaluated using (3.6) only if $\nu \to \infty$, a necessary condition to assure the validity of (3.6). Then the following result holds.

**Theorem 1:** Let $\mathsf{A}$ be the $s\nu^d \times s\nu^d$ matrix (2.1) and $\mathsf{a}(\boldsymbol{\theta})$ the $s \times s$ matrix (3.7); then, in the limit $\nu \to \infty$, the block element $\mathbf{l}$ of the matrix function $f(\mathsf{A})$ is given by the Fourier transform of the matrix $f(\mathsf{a})$:

$$\lim_{\nu\to\infty} \langle \mathbf{0}|f(\mathsf{A})|\mathbf{l}\rangle = (2\pi)^{-d} \int_{-\pi}^{\pi} f(\mathsf{a}(\boldsymbol{\theta})) e^{i\mathbf{l}\cdot\boldsymbol{\theta}}\, d\boldsymbol{\theta}. \tag{3.8}$$

Upon setting $\mathbf{l} = \mathbf{0}$, we obtain corollary 1.

*Corollary 1:* The trace of the matrix $f(\mathsf{A})$ in the limit $\nu \to \infty$ is given by the Fourier transform of the trace of the matrix $f(\mathsf{a})$:

$$\lim_{\nu\to\infty} \nu^{-d} \operatorname{Tr} f(\mathsf{A}) = (2\pi)^{-d} \int_{-\pi}^{\pi} \operatorname{Tr} f(\mathsf{a}(\boldsymbol{\theta}))\, d\boldsymbol{\theta}. \tag{3.9}$$

Note that the traces in the above expression are those of an infinite and finite matrix, respectively. Having considered an infinite lattice from start, the results are expressed in terms of the *continuous* variable $\boldsymbol{\theta}$. This vector can be considered as the continuous limit of the reciprocal vector associated with a finite lattice, taking only discrete values inside the Brillouin zone, to use the solid state physics' language. But it is necessary to now make an important and somewhat subtle remark, in order to enlighten a point which is rather obscure in the classical treatments of the translational symmetry, based on Bloch's theorem (i.e., their ability to deal with any function defined on an infinite lattice). As stressed above, in the power series expansion of the function $f(\mathsf{A})$, the elements $\langle \mathbf{0}|\mathsf{A}^p|\mathbf{l}\rangle$ are dependent on $\nu$ for large $p$ values, and it is in the limit $\nu \to \infty$ only that this dependence vanishes, yielding the simple results of Eqs. (3.8) and (3.9). Here the advantage of the lattice's infinite size has been properly taken into account.

Use of Theorem 1 can be made to evaluate the elements of some generalized cyclic matrix functions of physical interest. All results are valid in the limit $N \to \infty$, as understood. The inverse of the matrix (2.1) follows readily from (3.8) as

$$\langle \mathbf{0}|\mathsf{A}^{-1}|\mathbf{l}\rangle = (2\pi)^{-d} \int_{-\pi}^{\pi} (\mathsf{a}(\boldsymbol{\theta}))^{-1} e^{i\mathbf{l}\cdot\boldsymbol{\theta}}\, d\boldsymbol{\theta}. \tag{3.10}$$

The resolvent matrix, defined by

$$\mathsf{G}(z) = (z - \mathsf{A})^{-1}, \tag{3.11}$$

is also straightforward to derive from (3.8) as

$$\langle \mathbf{0}|\mathsf{G}(z)|\mathbf{l}\rangle = (2\pi)^{-d} \int_{-\pi}^{\pi} (z - \mathsf{a}(\boldsymbol{\theta}))^{-1} e^{i\mathbf{l}\cdot\boldsymbol{\theta}}\, d\boldsymbol{\theta}. \tag{3.12}$$

Paramount interest is offered by the spectral density, which is obtained from the relation

$$n(E) = (\pi N)^{-1} \operatorname{Tr} \operatorname{Im} \mathsf{G}(E - i0^+), \tag{3.13}$$

by using Eq. (3.9) as

$$n(E) = 2^{-d} \pi^{-(d+1)} \sum_i \int_{-\pi}^{\pi} \operatorname{Tr} \operatorname{Im}(E - i0^+ - \mathsf{a}(\boldsymbol{\theta}))\, d\boldsymbol{\theta}, \tag{3.14}$$

which can be put in the form

$$n(E) = (2\pi)^{-d} \sum_i \int_{-\pi}^{\pi} \delta(E - \lambda_i(\boldsymbol{\theta}))\, d\boldsymbol{\theta}, \tag{3.15}$$

where $\lambda_i(\boldsymbol{\theta})$ are the eigenvalues of the $\mathsf{a}(\boldsymbol{\theta})$ matrix.

The determinant of the matrix (2.1) follows readily from the well-known relation

$$\log\det(\mathsf{A}) = \mathrm{Tr}\log(\mathsf{A}). \tag{3.16}$$

Making use of Eq. (3.9) we have

$$N^{-1}\log\det \mathsf{A} = (2\pi)^{-d}\int_{-\pi}^{\pi}\log(\det \mathsf{a}(\boldsymbol{\theta}))d\,\boldsymbol{\theta}. \tag{3.17}$$

## IV. ISOTROPIC CASE

Now we consider the most common situation, where the interaction between the lattice sites $\mathbf{k}$ and $\mathbf{0}$, expressed by the matrix $\mathsf{a}_{\mathbf{k}}$, is only $|\mathbf{k}|$-dependent, i.e., isotropic in space. New interesting analytical expressions for the matrix function elements will be derived in this case. So let us now discuss matrices in the form

$$\mathsf{B} = \sum_{\mathbf{k}\in K}(1+\delta_{k_1})^{-1}\mathsf{M}_{k_1}\otimes\cdots\otimes(1+\delta_{k_k})^{-1}\mathsf{M}_{k_d}\otimes\mathsf{a}_{\mathbf{k}}, \tag{4.1}$$

where $\delta_{k_i}$ denotes the Krönecker symbol $\delta_{k_i} = \delta(0,k_i)$; now the components of the $\mathbf{k}$ vectors are non-negative integers $k\in N^d$ and the $\nu\times\nu$ symmetric topological matrix of order $i$ is defined by

$$\mathsf{M}_i = \mathsf{m}_i + \mathsf{m}_{-i}, \tag{4.2}$$

and obeys the relation[1]

$$C_p(\mathsf{M}_1) = \mathsf{M}_p, \tag{4.3}$$

where $C_p(x)$ is the $p$th-order Chebyshev polynomial of the second kind. Those polynomials satisfy the identity[5]

$$C_p(2\cos\theta) = 2\cos p\theta, \tag{4.4}$$

and they are orthogonal with respect to an inner product defined by

$$\langle f(x), C_p(x)\rangle = (2\pi(1+\delta_p))^{-1}\int_{-1}^{1}(1-x^2)^{-1/2}f(2x)C_p(2x)dx \tag{4.5}$$

$$= (\pi(1+\delta_p))^{-1}\int_0^{\pi}f(2\cos\theta)\cos p\theta\, d\theta. \tag{4.6}$$

Now taking advantage of the relation (4.4), we define from $\mathsf{B}$, according to the prescription (3.7) and the relation $\mathsf{a}_{\mathbf{k}} = \mathsf{a}_{|\mathbf{k}|}$, the following matrix:

$$\begin{aligned}\mathsf{b}(\boldsymbol{\theta}) &= \sum_{\mathbf{k}\in K}(1+\delta_{k_1})^{-1}(e^{ik_1\theta_1}+e^{-ik_1\theta_1})\\ &\quad\cdots(1+\delta_{k_d})^{-1}(e^{ik_d\theta_d}+e^{-ik_d\theta_d})\mathsf{a}_{\mathbf{k}}\\ &= \sum_{\mathbf{k}\in K}(1+\delta_{k_1})^{-1}C_{k_1}(2\cos\theta_1)\\ &\quad\cdots(1+\delta_{k_d})^{-1}C_{k_d}(2\cos\theta_d)\mathsf{a}_{\mathbf{k}}.\end{aligned} \tag{4.7}$$

Then applying Theorem 1, we have

$$\langle\mathbf{0}|f(\mathsf{B})|\mathbf{l}\rangle = (2\pi)^{-d}\int_{-\pi}^{\pi}f(\mathsf{b}(\boldsymbol{\theta}))e^{i\mathbf{l}\cdot\boldsymbol{\theta}}\,d\,\boldsymbol{\theta}, \tag{4.8}$$

but $\mathsf{b}(\boldsymbol{\theta})$ being an even function in $\theta_i$ $(\forall_i)$ the last expression can be rewritten as $2^{-d}$

$$\begin{aligned}\langle\mathbf{0}|f(\mathsf{B})|\mathbf{l}\rangle &= \pi^{-d}\int_0^{\pi}f(b(\boldsymbol{\theta}))\cos l_1\theta_1\cdots\cos l_d\theta_d\; d\,\boldsymbol{\theta}\\ &= (2\pi)^{-d}\int_0^{\pi}f(b(\boldsymbol{\theta}))C_{l_1}(2\cos\theta_1)\\ &\quad\cdots C_{l_d}(2\cos\theta)d\,\boldsymbol{\theta}\end{aligned} \tag{4.9}$$

and by generalization of the inner product (4.5) to $d$ dimensions, making a change of variables, we have the following theorem.

**Theorem 2:** Let $\mathsf{B}$ be the symmetric $s\nu^d\times s\nu^d$ matrix (4.1); and for $\mathbf{x} = (x_1, x_2, \ldots, x_d)\in R^d$, let $\mathsf{b}(\mathbf{x})$ be the $s\times s$ matrix given by

$$\begin{aligned}\mathsf{b}(\mathbf{x}) &= \sum_{\mathbf{k}\in K}(1+\delta_{k_1})^{-1}C_{k_1}(x_1)\\ &\quad\cdots(1+\delta_{k_d})^{-1}C_{k_d}(x_d)\mathsf{a}_{\mathbf{k}},\end{aligned} \tag{4.10}$$

then the block element $\mathbf{l}$ of the matrix function $f(\mathsf{B})$ is given in the limit $\nu\to\infty$ by the scalar product:

$$\lim_{\nu\to\infty}\langle\mathbf{0}|f(\mathsf{B})|\mathbf{l}\rangle = \left\langle f(\mathsf{b}(\mathbf{x})),\ \prod_{j=1}^{d}(1+\delta_{l_j})C_{l_j}(x_j)\right\rangle. \tag{4.11}$$

Then obviously we have Corollary 2.

*Corollary 2:* The trace of the matrix function $f(\mathsf{B})$ is given by

$$\lim_{\nu\to\infty}\nu^{-d}\,\mathrm{Tr}f(\mathsf{B}) = 2^d\left\langle\mathrm{Tr}f(\mathsf{b}(\mathbf{x})),\ \prod_{j=1}^{d}C_0(x_j)\right\rangle. \tag{4.12}$$

Functions of matrices are generally defined as infinite power series expansions in terms of the matrix variable. Consideration of the orthogonal properties of Chebyshev polynomials allows much easier calculation of matrix elements by using Theorem 2; especially taking into account the property

$$\langle x^p, C_q(x)\rangle = 0, \qquad \text{for } q>p \tag{4.13}$$

reduces calculations to the evaluation of matrix polynomials, whose degree does not exceed some finite values fixed by $\mathbf{l}$ and $K$.

## V. CYCLIC MATRIX REPRESENTATION OF LATTICE OPERATORS

To define variables at the sites of a lattice, actually existing or acting as an approximation to the continuum $R^d$, yields a discretization of partial differential equations and associated operators. We will discuss now the representation of $d$-dimensional linear equations with constant coefficients and periodic boundary conditions in terms of generalized cyclic matrices, and the possibility to obtain closed expressions for the associated lattice resolvent operators by virtue of the above properties of functions of infinite generalized cyclic matrices.

Let $\psi(\mathbf{x})$ be some discontinuous function defined on the (discrete) sites $\mathbf{x}$ of an infinite $d$-dimensional toroidal lattice $T^d$, with a spacing $h = (h_1, \ldots, h_d)$ which is finite though small. Alternatively the value of $\psi(\mathbf{x})$, $\mathbf{x} = \Sigma_{i=1}^{d} l_i h_i \in T^d$, $l_i\in Z$ can be labeled $\psi(l_1, \ldots, l_d)$. A $d$-dimensional difference equation obeyed by the variable $\psi$ is obtained by replacing each derivative in the partial differential equation by a finite-difference approximation, so that an infinite set of simultaneous algebraic equations involving the values of $\psi$ at the lattice points can be written down in accordance with the following prescriptions.

The forward lattice first derivative in the $\alpha$th-coordinate direction is defined by[6]

$$\partial_\alpha^{+}\psi(l_1, \ldots, l_d) = h_\alpha^{-1}(\psi(l_1, \ldots, l_\alpha+1, \ldots, l_d) - \psi(l_1, \ldots, l_\alpha, \ldots, l_d)); \tag{5.1}$$

the matrix representation, in the $\psi$ basis set, of this operator is the direct product of $d$ infinite matrices

$$\partial_\alpha^+ = h_\alpha^{-1}(1 \otimes \cdots \otimes (\mathsf{m}_1 - 1) \otimes \cdots \otimes 1). \tag{5.2}$$

Similarly, the backward lattice first derivative

$$\partial_\alpha^- \psi(l_1, \ldots, l_d) = h_\alpha^{-1}(\psi(l_1, \ldots, l_\alpha, \ldots, l_d) - \psi(l_1, \ldots, l_\alpha - 1, \ldots, l_d)), \tag{5.3}$$

has the matrix representation

$$\partial_\alpha^- = h_\alpha^{-1}(1 \otimes \cdots \otimes (1 - \mathsf{m}_{-1}) \otimes \cdots \otimes 1). \tag{5.4}$$

Likewise, a symmetric lattice gradient can be defined by

$$\partial_\alpha^0 \psi(l_1, \ldots, l_d) = (2h_\alpha)^{-1}(\psi(l_1, \ldots, l_\alpha + 1, \ldots, l_d) - \psi(l_1, \ldots, l_\alpha - 1, \ldots, l_d)), \tag{5.5}$$

and represented by

$$\partial_\alpha^0 = (2h_\alpha)^{-1}(1 \otimes \cdots \otimes (\mathsf{m}_1 - \mathsf{m}_{-1}) \otimes \cdots \otimes 1). \tag{5.6}$$

Then from the previous expressions higher-order difference derivative operators easily follow, namely

$$(\partial_\alpha^+)^p (\partial_\beta^-)^q = h_\alpha^{-p} h_\beta^{-q} 1 \otimes \cdots \otimes (\mathsf{m}_1 - 1)^p \otimes \cdots \otimes (1 - \mathsf{m}_{-1})^q \otimes \cdots \otimes 1, \tag{5.7}$$

$$(\partial_\alpha^+)^p (\partial_\alpha^-)^q = h_\alpha^{-(p+q)} 1 \otimes \cdots \otimes (\mathsf{m}_1 - 1)^p \times (1 - \mathsf{m}_{-1})^q \otimes \cdots \otimes 1. \tag{5.8}$$

Using those basic formulas any linear lattice operator with constant coefficients can be readily obtained. For example, the lattice Laplace operator which is defined to be

$$\Delta = \sum_{\alpha=1}^{d} \partial_\alpha^+ \partial_\alpha^-, \tag{5.9}$$

has the matrix representation

$$\Delta = h_1^{-2}(\mathsf{M}_1 - 21) \otimes 1 \cdots \otimes 1 + \cdots + h_d^{-2} 1 \otimes \cdots \otimes (\mathsf{M}_1 - 21). \tag{5.10}$$

Similarly the general two-dimensional second-order partial differential equation

$$a_{11}\frac{\partial^2\psi}{\partial x_1^2} + a_{12}\frac{\partial^2\psi}{\partial x_1 \partial x_2} + a_{22}\frac{\partial^2\psi}{\partial x_2^2} + b_1\frac{\partial\psi}{\partial x_1} + b_2\frac{\partial\psi}{\partial x_2} + c\psi = 0, \tag{5.11}$$

can be approximated by the set of algebraic equations written as

$$(a_{11}h_1^{-2}(\mathsf{M}_1 - 21) \otimes 1 + a_{12}(h_1h_2)^{-1}(\mathsf{m}_1 - 1) \otimes (1 - \mathsf{m}_{-1}) + a_{22}h_2^{-2}1 \otimes (\mathsf{M}_1 - 21) + b_1h_1^{-1}(\mathsf{m}_1 - 1) \otimes 1 + b_2h_2^{-2}1 \otimes (\mathsf{m}_1 - 1) + c1 \otimes 1)\psi = 0, \tag{5.12}$$

in matrix notation.

Now the previous prescriptions yield the generalized cyclic matrix representation of any linear operator with constant coefficients $\mathsf{L}$, it is then straightforward to obtain the lattice resolvent operator $\mathsf{G} = \mathsf{L}^{-1}$ in closed form by using Theorem 1 or 2, according as $\mathsf{L}$ is nonsymmetric or symmetric.

As an example we consider the two-dimensional biharmonic equation[7]

$$\nabla^4\psi + E\psi = 0; \tag{5.13}$$

from Eq. (5.10), assuming unit lattice spacing, we have for the matrix representation of the lattice biharmonic operator

$$\mathsf{L}_B = (\mathsf{M}_1 - 21)^2 \otimes 1 + 2(\mathsf{M}_1 - 21) \otimes (\mathsf{M}_1 - 21) + 1 \otimes (\mathsf{M}_1 - 21)^2 + E1 \otimes 1. \tag{5.14}$$

So that application of Theorem 2 gives the resolvent in the form

$$\langle 00|\mathsf{L}_B^{-1}|l_1l_2\rangle = (1 + \delta_{l_1})(1 + \delta_{l_2})\langle((x_1 - 2)^2 + 2(x_1 - 2)(x_2 - 2) + (x_2 - 2)^2 + E)^{-1}, C_{l_1}(x_1)C_{l_2}(x_2)\rangle, \tag{5.15}$$

which, by using Eq. (4.6), takes the form

$$\langle 00|\mathsf{L}_B^{-1}|l_1l_2\rangle = \pi^{-2}\int_0^\pi\int_0^\pi \frac{\cos l_1\theta_1 \cos l_2\theta_2 \, d\theta_1 \, d\theta_2}{(2\cos\theta_1 - 2)^2 + 2(2\cos\theta_1 - 2)(2\cos\theta_2 - 2) + (2\cos\theta_2 - 2)^2 + E} = \pi^{-2}\int_0^\pi\int_0^\pi \frac{\cos l_1\theta_1 \cos l_2\theta_2 \, d\theta_1 \, d\theta_2}{16(\sin^2(\theta_1/2) + \sin^2(\theta_2/2)) + E}. \tag{5.16}$$

For applications in quantum and statistical mechanics one often has to evaluate the exponential of some differential operator $\mathsf{H}$; if $\mathsf{H}$ can be put in cyclic matrix form, closed expressions are readily derivable for the matrix elements of $\exp(\beta\mathsf{H})$($\beta$ is a scalar). A trivial example is offered by the case $\mathsf{H} = \Delta$, which corresponds to the $d$-dimensional propagator for a free particle.[8] By virtue of Eqs. (5.10), (4.10), (4.11), and (4.6) the matrix elements can be written

$$\langle \mathbf{0}|\exp(\beta\Delta/2)|\mathbf{1}\rangle = \pi^{-d}\int_0^\pi \cdots \int_0^\pi \exp\left(\beta\sum_{i=1}^{d} h_i^{-2}(\cos\theta_i - 1)\right) \times \cos l_1\theta_1 \cdots \cos l_d\theta_d \, d\theta_1 \cdots d\theta_d = \pi^{-d}\prod_{i=1}^{d}\int_0^\pi \exp(\beta h_i^{-2}(\cos\theta_i - 1))\cos l_i\theta_i \, d\theta_i = \prod_{i=1}^{d}\exp(-\beta h_i^{-2})I_{l_i}(\beta h_i^{-2}), \tag{5.17}$$

which is simply the product of $d$ one-dimensional propagators expressed in terms of the modified Bessel[5] functions $I_p(z)$.


[1] P. O. Löwdin, R. Pauncz, and J. de Heer, J. Math. Phys. **1**, 461 (1960).
[2] T. L. Gilbert, J. Math. Phys. **3**, 107 (1962).
[3] P. B. Abraham and G. Weiss, J. Math. Phys. **3**, 340 (1962).
[4] J. L. Calais and K. Appel, J. Math. Phys. **5**, 1001 (1964).
[5] M. Abramowitz and J. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1970).
[6] G. D. Smith, *Numerical Solution of Partial Differential Equations: Finite Difference Methods* (Oxford U. P., London, 1978).
[7] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), p. 1786.
[8] L. S. Schulman, *Techniques and Applications of Path Integration* (Wiley, New York, 1981), p. 4.

# On the integrability of certain symmetric representations of the Lie algebra of $\overline{SO_0(4,1)}$

A. Bohm and P. Moylan
*Center for Particle Theory, The University of Texas at Austin, Austin, Texas 78712*

(Received 6 March 1984; accepted for publication 20 July 1984)

A proof of the existence of an essentially self-adjoint extension of a symmetric $\overline{SO_0(4,1)}$ Nelson operator, which is constructed out of the generators of a positive mass, arbitrary spin unitary irreducible representation of the Poincaré group, is presented. Our analysis of $\overline{SO_0(4,1)}$ and its Lie algebra provides us with an example of an observation of Harish-Chandra: There exist subspaces of the space of differentiable vectors of a representation of a noncompact group which are invariant under the Lie algebra, but the closures of the subspaces are not invariant under the group. The chief results of this paper should hold true for $\overline{SO_0(n,1)}$. In particular, we should have a realization of an arbitrary principal series irreducible unitary representation of $SO_0(n,1)$ on the direct sum of two identical unitary irreducible representation spaces of the motion group in an $n$-dimensional Minkowski space, which has one timelike dimension.

## I. INTRODUCTION

It was first pointed out by Harish-Chandra that there are subspaces of the set of differentiable vectors of a representation of a noncompact Lie group which are invariant under the Lie algebra, but the closures of the subspaces are not invariant under the group.[1] This unexpected behavior seems largely to have remained unnoticed by most theoretical physicists; they usually assume that if we are given a representation of a Lie algebra by symmetric operators in a linear space $\phi$, which is dense in a Hilbert space $\mathscr{H}$ (dense with respect to the topology given by the inner product), then there exists a unitary representation of the associated Lie group in $\mathscr{H}$. We present here an example of Harish-Chandra's remark that should be of interest to both mathematicians and mathematical physicists: We describe representations of the Lie algebra of the Poincaré group $\overline{\mathscr{P}}$ on the space of differentiable vectors $\Phi(m,s;\pm)$ of an irreducible unitary Poincaré group representation,[2] and we construct on this space a representation of the Lie algebra of $\overline{SO_0(4,1)}$—the simply connected covering group of the de Sitter group—by symmetric operators. We show that the closure of the space $\Phi(m,s;\pm)$ does not furnish us with a representation of $\overline{SO_0(4,1)}$. However, we are able to show that the direct sum of two identical spaces of differentiable vectors, $\Phi(m,s;\pm)\oplus\Phi(m,s;\pm)$ for irreducible unitary Poincaré group representations does furnish us with a representation of $\overline{SO_0(4,1)}$ on its closure, the direct sum of two identical unitary irreducible representation (UIR) spaces of $\overline{\mathscr{P}}$: $\mathscr{H}(m,s;\pm)\oplus\mathscr{H}(m,s;\pm)$. [$\mathscr{H}(m,s,\pm)$ denotes the space of a positive mass,[2] arbitrary spin, positive or negative energy UIR of $\overline{\mathscr{P}}$.] We are able to realize all of the principal series UIR's of $\overline{SO_0(4,1)}$ in this manner.

Before we state our results for the general case we describe those representations of $\overline{SO_0(4,1)}$ which are realized as equivalent multiplier representations on either the three-sphere or on the two-sheeted three-dimensional (compactified) hyperboloid,[3,4] and prove our claims for those cases. First, however, we discuss some important results on infinite-dimensional group representations, which we will need later on.

The description of the $\overline{SO_0(4,1)}$ multiplier representations on the three-sphere or on the compactified hyperboloid shows the geometrical origin behind the noninvariance of the closure of the space of differentiable vectors of a Poincaré group representation under the $\overline{SO_0(4,1)}$ representation: The vector fields defined on each branch of the hyperboloid, which are associated with the $\overline{SO_0(4,1)}$ Lie algebra generators, are not complete,[5] and the action of $\overline{SO_0(4,1)}$ on each branch of the (noncompactified) hyperboloid must be defined as the action of a local Lie group.[6]

We now make a few remarks concerning notation. Lowercase Roman indices generally run from 1, 2, 3 or 0, 1, 2, 3, 4, and Greek ones run from 0, 1, 2, 3. The metric tensor of Minkowski space $M_{3,1}$ is $\eta_{\mu\nu}=\mathrm{diag}(1,-1,-1,-1)$ and $n_{ab}=\mathrm{diag}(1,-1,-1,-1,-1)$ is the metric tensor of a $(4+1)$-dimensional Minkowski space, $M_{4,1}$. The translation generators in $M_{4,1}$, are denoted by a four-vector operator $P^\mu=(E^{\mathrm{op}},\mathbf{P}^{\mathrm{op}})$ and the contravariant four-momentum vector is $p^\mu=(E,\mathbf{p})$. The position operator acting on $\mathscr{L}^2(\mathbb{R}^4)$, where $\mathbb{R}^4$ is the character space of the four-dimensional abelian translation group (momentum space), is given by

$$Q^\mu=-i\frac{\partial}{\partial p_\mu}=-\left(i\frac{\partial}{\partial E},\frac{1}{i}\nabla_\rho\right).$$

## II. THE SPACE OF DIFFERENTIABLE VECTORS

Let $G$ be a Lie group and denote its Lie algebra by $\mathscr{G}$. Let $(U(G),\mathscr{H})$ be a bounded continuous representation of $G$ in a Hilbert space $\mathscr{H}$. We say that $v\in\mathscr{H}$ is a $C^\infty$ vector for the representation $U$ if $g\to U(g)v$ is of class $C^\infty$ on $G$ ($g\in G$). Obviously the $C^\infty$ vectors form a vector subspace of $\mathscr{H}$; we denote the subspace by $\mathscr{D}^\infty(U(G))$, or simply by $\mathscr{D}^\infty$ when no confusion arises.

 0022-2488/85/030365-10$02.50 © 1985 American Institute of Physics 

We have the following well-known result.[7]

**Theorem 2.1:** Let $(U(G,\mathscr{H}))$ be a bounded continuous representation of the Lie group $G$ on a Hilbert space $\mathscr{H}$. For $X\in\mathscr{G}$ define the linear map $dU(X)$ on $\mathscr{D}^\infty(U(G))$ into $\mathscr{H}$ by

$$i\,dU(X)v=\lim_{t\to 0}\frac{U(e^{tX})v-v}{t}.$$

Then each $dU(X)$ leaves $\mathscr{D}^\infty(U(G))$ stable and $i\,dU(X)$ is a representation of $\mathscr{G}$ on $\mathscr{D}^\infty(U(G))$.

By the theorem, since $i\,dU$ furnishes us with a representation of $\mathscr{G}$ on $\mathscr{D}^\infty(U(G))$, $i\,dU$ extends to a representation of the universal enveloping algebra $\mathscr{E}(\mathscr{G})$ on $\mathscr{D}^\infty(U(G))$ with $i\,dU(1)=I$.[8]

We now state the following important result due to Goodman.[9]

*Lemma 2.1:* Let $X_1\cdots X_n$ be a basis for $\mathscr{G}$; suppose that $v\in\mathscr{H}$ lies in the domain of $J_\alpha^m=dU(X_\alpha)^m=dU(X_\alpha)dU(X_\alpha)\cdots dU(X_\alpha)$ for $m=1, 2, \ldots$ and $1\leqslant\alpha\leqslant n$, then $v\in\mathscr{D}^\infty$.

The following result characterizes $\mathscr{D}^\infty(U(G))$ as the largest subspace of $\mathscr{H}$ which is stabilized by $dU(\mathscr{G})$. This theorem will be very important later on. The characterization was stated to our knowledge first by Nagel,[10] but not proved. However, it certainly seems to be known by other researchers in the field,[11] and is a straightforward consequence of Goodman's result.

**Theorem 2.2:** $\mathscr{D}^\infty$ is the largest subspace of $\mathscr{H}$ such that

(i) $\mathscr{D}^\infty\subset\bigcap_{\alpha=1}^{n}\mathscr{D}(J_\alpha)$,

(ii) $J_\alpha\,\mathscr{D}^\infty\subset\mathscr{D}^\infty$ for each $\alpha(\alpha=1,\ldots n)$.

Since the proof does not seem to appear elsewhere, we present it in Appendix A.

We remark that $\mathscr{D}^\infty$ is dense in the representation space $\mathscr{H}$. This follows from Gårding's theorem on the density of the Gårding subspace $\mathfrak{G}$ (Ref. 12) and also Nelson's theorem that $\mathfrak{G}\subset\mathscr{D}^\infty$.[13]

For symmetric elements of $\mathscr{E}(\mathscr{G})$ and $U$ unitary we obtain the following.

**Theorem 2.3:** Let the representation $U(G)$ on $\mathscr{H}$ be unitary. Then the representation $dU(\mathscr{G})$ of $\mathscr{G}$ on $\mathscr{D}^\infty$ is symmetric, i.e., all of the operators $dU(X)$ $(X\in\mathscr{G})$ are symmetric operators in $\mathscr{H}$. If $M\in\mathscr{E}(\mathscr{G})$ is a real symmetric polynomial function in the variables $X_1\cdots X_n$,

$$M=M(X_1\cdots X_n)=\sum_{\alpha_1\cdots\alpha_n}C_{\alpha_1\cdots\alpha_n}\,\mathbb{P}(X_{\alpha_1}\cdots X_{\alpha_n})\quad(C_{\alpha_1\cdots\alpha_n}\in\mathbb{R}),$$

where

$$\mathbb{P}(X_{\alpha_1}\cdots X_{\alpha_n})=\frac{1}{n!}\sum_{\sigma(\alpha(1))\cdots\sigma(\alpha(n))}X_{\sigma(\alpha(1))}\cdots X_{\sigma(\alpha(n))}$$

and the representation $dU(\mathscr{G})$ on $\mathscr{D}^\infty$ is symmetric, then

$$(w|dU^\dagger(M)v)=(w|dU(M)v).$$

That is to say, $dU(M)$ is a symmetric operator.

*Note:* A Lie algebra representation is always a map from the Lie algebra into skew-symmetric operators, if the corresponding representation of the group is unitary. Thus, we have the appearance of the factor $i$ in the above. We do not always make this trivial distinction between the "mathematical" and "physical" representations, and sometimes we call a mapping of the Lie algebra into symmetric operators a representation, as we did in Theorem 2.3.

## III. MULTIPLIER REPRESENTATIONS OF $SO_0(4,1)$ ON $\mathscr{L}_2(T^3)$

We now describe certain multiplier representations of $\overline{SO_0(4,1)}$. First we recall some facts about local (global) actions of Lie groups on manifolds and multiplier representations.[14,15]

*Definition 3.1:* Let $G$ be an $m$-dimensional Lie group, $V$ a neighborhood of the identity in $G$, and let $U$ be an open set in $\mathbb{R}^n$ containing zero. A local action $(\Phi,V)$ of the Lie group $G$ on $U\subset\mathbb{R}^n$ is an analytic mapping $\Phi$: $V\times U\to\mathbb{R}^n$ such that

(i) for all $x\in U$, $\Phi(e,x)=ex=x$ ($e$ is identity in $G$);

(ii) for every $g$, $h\in V$ with $\Phi(g,x)=gx\in U(x\in U)$, $\Phi(h,\Phi(g,x))=\Phi(hg,x)=(hg)x$.

From (i) and (ii) it follows that $\Phi_g=\Phi(g,)\!:x\to gx$ is locally one-to-one for fixed $g$.[16]

Let $\exp Xt$, $X\in\mathscr{G}$ (Lie algebra of $G$) be a one-parameter group in $G$. If $x_0\in U$ we call the curve $x(t)=(\exp Xt)x_0=\Phi(\exp Xt,x_0)$ the trajectory of $x^0$ under $\exp Xt$.

Next we define a local multiplier representation $T_\nu$ of $G$ on the set $\mathscr{A}=\mathscr{A}(U)$ of all complex-valued functions on $\mathbb{R}^n$, which are analytic in a neighborhood containing zero, which is contained in a set $U$ on which a local group action is defined.

*Definition 3.2:* Let $G$ be a Lie group, $(\Phi,V)$ a local action of $G$ on $U\subset\mathbb{R}^n$ ($U$ an open subset of $\mathbb{R}^n$ containing zero). A local multiplier representation $T_\nu$ of $G$ on $\mathscr{A}$ with multiplier $\nu$, consists of a mapping $T_\nu(g)$ of $\mathscr{A}$ onto $\mathscr{A}$ defined for $g\in V$, $f\in\mathscr{A}$, by

$$[T_\nu(g)f](x)=\nu(g,x)f(gx),\quad x\in U,\tag{3.1}$$

where $\nu(g,x)$ is a complex-valued analytic function of $g$ and $x$ and satisfies

(1) $\nu(e,x)=1$, $\forall x\in U$,

(2) $\nu(g_1g_2,x)=\nu(g_1,x)\nu(g_2,g_1x)$, $g_1,g_2,g_1g_2\in V$, $x\in U$.

From (2) follows a local homomorphism condition

$$[T_\nu(g_1g_2)f](x)=[T_\nu(g_1)T_\nu(g_2)f](x)\quad(g_1,g_2,g_1g_2\in V).$$

Now we define the so-called generalized Lie derivative for a local multiplier representation.[14]

*Definition 3.3:* The generalized Lie derivative $D_Xf$ of a function $f:U\to\mathbb{C}$, $f\in\mathscr{A}$ under the one-parameter group $\exp Xt:U\to\mathbb{R}^n$ ($t$ sufficiently small so that $\exp Xt\in V$) is the analytic function

$$[D_Xf](x)=\frac{d}{dt}[T_\nu(\exp Xt)f]\,|_{t=0}(x)$$
$$=\sum_{j=1}^{m}\sum_{i=1}^{n}X_j\frac{\partial\Phi_i}{\partial g_j}(g(t),x)\,|_{g=e}\,\frac{\partial f}{\partial x_i}(x)$$
$$+\sum_{j=1}^{m}X_j\,P_j(x)f(x).\tag{3.2}$$

Here the $g_j$'s are the coordinates in a parameter space of the group: $g=(g_1,g_2,\ldots,g_m)\in V$, and the functions $P_j(x)$ are defined by the differential equations

$$\sum_{j=1}^{m} X_j\, P_j(x) = \frac{d}{dt}\, \nu\,(\exp Xt,x)\,|_{t=0} \tag{3.3}$$

[$X=(X_1,X_2,\ldots X_m)$ with $g_j(t)=\exp(X_j t)$]. This set of all generalized Lie derivatives form a Lie algebra which is a homomorphic image of $\mathcal{G}$. We call such an action of $\mathcal{G}$ on $U$ a generalized infinitesimal action. The operations, are, of course, addition and Lie product of Lie derivatives. The generalization to these local multiplier representations of the converse of Lie's second fundamental theorem[17] is the following.[18]

**Theorem 3.1:** Let $\chi^{\text{ext}}(U)$ denote the space of all analytic differential operators on $U$. Let $\phi:\mathcal{G}\to\chi^{\text{ext}}(U)$ be a generalized infinitesimal action of $\mathcal{G}$ on $U$, an open set in $\mathbf{R}^n$ containing zero. Let all $\phi(X)$ ($X\in\mathcal{G}$) be of the form

$$\phi(X) = \sum_{j=1}^{m} \alpha_j D_j(x),$$

where

$$X = \sum_{j=1}^{m} \alpha_j I_j$$

(the $I_j$ are a basis for $\mathcal{G}$) and where

$$D_j(x) = \sum_{i=1}^{n} P_{ji}(x)\frac{\partial}{\partial x_i} + P_j(x) = \phi(I_j)$$

are a set of $m$ linearly independent differential operators defined and analytic in $U\subset\mathbf{R}^n$, which are the images of a basis for the Lie algebra $\mathcal{G}$. Then the set of all the $\phi(X)$'s is the algebra of generalized Lie derivatives for a local multiplier representation $T_\nu$ of a group $G$ whose Lie algebra is $\mathcal{G}$. If $G$ is simply connected, the local multiplier representation of $G$ is unique except for the possible choice of $V\subset G$.

The action of $G$ on $x^0\in U$ is obtained by solving the differential equations

$$\frac{d}{dt}\, x_i(t) = \sum_{j=1}^{m} \alpha_j\, P_{ji}(x(t)),$$

$$x_i(0) = x_i^0 \quad (i=1,\ldots,n),$$

$$\frac{d}{dt}\,\nu(\exp Xt,x_0) = \nu(\exp Xt,x_0)\sum_{j=1}^{m}\alpha_j\, P_j(x(t))$$

$$(\nu(e,x^0)=1),$$

where $x(t)=\exp Xt\, x_0$ ($X\in\mathcal{G}$). Specifically its action is given by

$$[\,T_\nu(\exp Xt)f\,](x_0) = \nu(\exp Xt,x_0)f(\exp Xt\, x_0),$$

where $\nu(\exp Xt,x_0)$ and $x(t)$ are the solutions of the above differential equations.

Obviously the local definitions which we have made in the above can be applied to a manifold $M$ if we consider $U$ to be the image of a subset of $M$ in a chart on $M$.

*Definition 3.4:* An infinitesimal $\mathcal{G}$ action on a manifold $M$ is defined to be a homomorphism $\phi$ of $\mathcal{G}$ into the Lie algebra of all differential vector fields on $M$.[19]

We define a global $G$ action on a manifold $M$ as follows[19]:

*Definition 3.5:* Let $G$ be a Lie group and let $M$ be a manifold. Let $\mathcal{D}(M)$ denote the group of all diffeomorphisms of $M$ into itself. A global $G$ action $\Phi$ on $M$ is a map $\Phi$: $G\times M\to M$ with $g\to\Phi_g$ a homeomorphism of $G$ into $\mathcal{D}(M)$. If the map $\Phi$ is one-to-one the action is said to be essential. [$\Phi_g(x)=gx\in M$.]

We would like to know when an infinitesimal $\mathcal{G}$-action on a manifold $M$ generates a unique global action of $G$ on the manifold. For $M$ compact we have the following theorem due to Palais.[20]

**Theorem 3.2:** If $G$ is simply connected and $M$ is a compact Hausdorff manifold, then every infinitesimal $\mathcal{G}$-action $\phi$ on $M$, generates a unique global action $\Phi$ of $G$ on $M$.

Next we define a multiplier representation $T_\nu$ of $G$ on $M$.

*Definition 3.6:* Let $G$ be a Lie group and let $\Phi$ be a global action of $G$ on a manifold $M$. Let $\mathcal{L}^2(M,\mu)$ be the Hilbert space completion of the space of all complex-valued functions on $M$ which are square integrable with respect to a measure $\mu$ defined on $M$. A multiplier representation $T_\nu$ of $G$ with multiplier $\nu$ is a bounded, continuous representation $T_\nu(G)$ of $G$ on $\mathcal{L}^2(M,\mu)$ with $T_\nu(g):\mathcal{L}^2(M,\mu)\to\mathcal{L}^2(M,\mu)$ defined for $g\in G$ and $f\in\mathcal{L}^2(M,\mu)$ by

$$[\,T_\nu(g)f\,](x) = \nu(g,x)f(gx) \quad (x\in M, gx=\Phi(g,x)), \tag{3.4}$$

where $\Phi$ is a global action of $G$, on $M$, and $\nu(g,x)$ is a complex-valued, a.e. (almost everywhere) continuously differentiable function in $x$ and $g$ such that

(i) $\nu(e,x)=1, \quad \forall x\in M,$

(ii) $\nu(g_1g_2,x)=\nu(g_1,x)\nu(g_2,g_1x),$

$g_1,g_2\in G,\ x\in M.$

Just as in the local case we may show that the homomorphism property

$$[\,T_\nu(g_1g_2)f\,](x) = [\,T_\nu(g_1)(T_\nu(g_2))f\,](x) \tag{3.5}$$

follows from (ii) combined with the action $\Phi$ of $G$ on $M$.

The analog of the Lie derivative of an analytic function $f$ on $M$ [$f\in\mathcal{A}(M)$, where $\mathcal{A}(M)$ denotes the space of all analytic functions on $M$] is the generalized Lie derivative.

*Definition 3.7:* The generalized Lie derivative $D_X f$ of $f\in\mathcal{A}(M)$ with respect to the vector field $\phi(X)$ ($X=\sum_{j=1}^m \alpha_j I_j\in\mathcal{G}$) associated with the one-parameter group $\Phi_{g(t)}=\Phi_{\exp Xt}:M\to M$ is the analytic function

$$(D_X f)(x) = \frac{d}{dt}\,[\,T_\nu(\exp Xt)f\,](x)\,|_{t=0}.$$

Computation in a Euclidean coordinate system $\psi_i(p)=(x_1(p),x_2(p),\ldots,x_n(p))$ ($\psi$ is a differentiable map from an open subset $U$ of $M$ into $\mathbf{R}^n$, whose $n$ components $\psi_i$ we denote by $x_i$) yields

$$(D_X f\circ\psi^{-1})(x) = \sum_{j=1}^{m}\sum_{i=1}^{n} X_j \left.\frac{\partial\Phi_i(x,g(t))}{\partial g_j}\right|_{g=e} \frac{\partial(f\circ\psi^{-1})}{\partial x_i}(x) + \sum_{j=1}^{m} X_j\, P_j(x)(f\circ\psi^{-1})(x), \tag{3.6}$$

where $\Phi_i(x,g)=\psi_i(gx)$ and where the functions $P_j(x)$ are defined, as in the local case, by the differential equation

$$\sum_{j=1}^{m} X_j P_j(x) = \frac{d}{dt} v(\exp Xt, x)|_{t=0}.$$

The set of all generalized Lie derivatives of a multiplier representation of a Lie group form a Lie algebra which is a homomorphic image of $\mathcal{G}$. The operations are, of course, addition and Lie product of Lie derivatives.

Next we describe certain representations of the de Sitter group and its Lie algebra on subspaces of spin-zero unitary representations of the Poincaré group. The Poincaré group is the semidirect product of the Lorentz group $SO_0(3,1)$ with the abelian four-dimensional translation group $T_4$,

$$\mathcal{P} = SO_0(3,1) \circledS T_4.$$

SO(3,1) is the component connected to the identity of the group of all real linear transformations of $\mathbf{R}^4$ which preserve the quadratic form

$$x_0^2 - x_1^2 - x_2^2 - x_3^2$$

on $\mathbf{R}^4$. The simply connected covering group of $\mathcal{P}$ we denote by $\overline{\mathcal{P}}$ and is equal to $\overline{SO_0(3,1)} \circledS T_4$, where $\overline{SO_0(3,1)} = SL(2,C)$ is the simply connected covering group of $SO_0(3,1)$. Positive mass, integer, or semi-integer spin UIR's of $\overline{\mathcal{P}}$ were constructed by Wigner.[21] We denote the UIR spaces of these representations by $\mathcal{H}(m,s;+)$ or $\mathcal{H}(m,s;-)$, the + or − signs referring to UIR's with different signs of the eigenvalue of the generator of translations in the $x_0$ direction. Realizations of these representations for spin zero are provided by Hilbert spaces of $\mathcal{L}^2$ functions on the positive and negative branches of the momentum hyperboloid (the character space of $T_4$ for these representations):

$$\begin{aligned} &\text{(a)}\ p_0^2 - |\mathbf{p}|^2 = m^2, \quad p_0 > 0, \\ &\text{(b)}\ p_0^2 - |\mathbf{p}|^2 = m^2, \quad p_0 < 0, \end{aligned} \tag{3.7}$$

where $(p_0,\mathbf{p}) = (p_0,p_1,p_2,p_3)$ denotes the coordinates in the character space. The two branches are depicted in Fig. 1, where it is shown how they form the cone in projective four-momentum space.[22] The following measure on $T_3^+$ is invariant under the Poincaré group:

$$d\mu_{\mathcal{P}}^+(p) = |1/2p_0| dp_1\, dp_2\, dp_3, \quad p_0 > 0. \tag{3.8}$$

We may construct on $T_3^+$ the Hilbert space $\mathcal{L}^2(T_3^+, d\mu_{\mathcal{P}}^+)$. We have the well-known result[23]

$$H(m,0;+) = \mathcal{L}^2(T_3^+, d\mu_{\mathcal{P}}^+).$$

Likewise on $T_3^-$ we have the Poincaré group invariant measure

$$d\mu_{\mathcal{P}}^-(p) = |1/2p_0| dp_1\, dp_2\, dp_3, \quad p_0 > 0 \tag{3.9}$$

and the Hilbert space

$$\mathcal{L}^2(T_3^-, d\mu_{\mathcal{P}}^-) \cong H(m,0;-).$$

For convenience we write these Hilbert spaces simply as $\mathcal{L}^2(T_3^+)$ or $\mathcal{L}^2(T_3^-)$ when no confusion arises.

Now let us denote the space of all bounded linear operators on $\mathcal{L}^2(T_3^+)$ [or $\mathcal{L}^2(T_3^-)$] by $\mathcal{B}(\mathcal{L}^2(T_3^+))$ [or $\mathcal{B}(\mathcal{L}^2(T_3^-))$].

Then the action of $\overline{\mathcal{P}}$ on $\mathcal{L}^2(T_3^+)$ [or $\mathcal{L}^2(T_3^-)$] is given by the mapping $U: \overline{\mathcal{P}} \to \mathcal{B}(\mathcal{L}^2(T_3^+))$ [ or $\mathcal{B}(\mathcal{L}^2(T_3^-))$] with

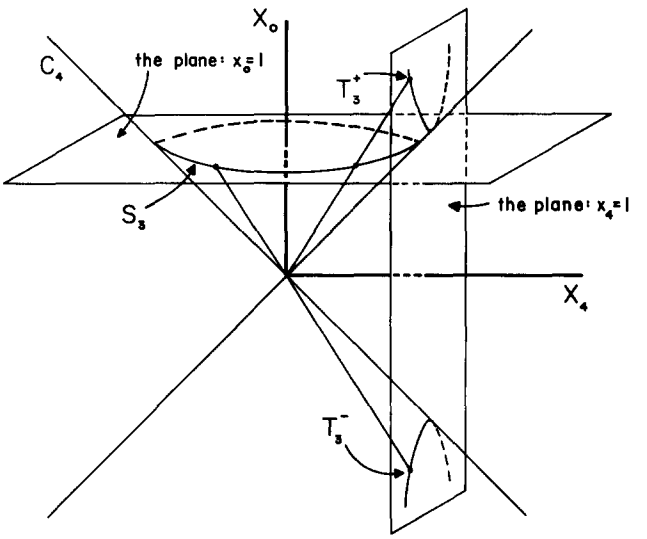



FIG. 1. The three-sphere of radius $m$ and the hyperboloid of radius $m$ in projective four-space ($\mathbf{R}^5$). The cone $C_4 = \{x | x_0^2 - x_1^2 - x_2^2 - x_3^2 - x_4^2 = 0\}$ intersects the plane $x_0 = m$ in the sphere $S_3$ and $C_4$ intersects the plane $x_4 = m$ in the two-sheeted hyperboloid $T_3 = T_3^+ \cup T_3^-$. $p_\mu = (x_0/mx_4, x_i/mx_4)$ and $\hat{p}_\mu = (x_4/mx_0, x_i/mx_0)$ relate usual coordinates ($p_\mu$ and $\hat{p}_\mu$) on $T_3$ and $S_3$ to projective coordinates $x_a$.

$$[U((\Lambda,a))f](p) = e^{-ia\cdot p} f(\Lambda^{-1} p), \tag{3.10}$$

where $p \to \Lambda^{-1} p$ is the usual linear action of SO(3,1) on $\mathbf{R}^4$ [$\Lambda \in SO(3,1)$] and $a \in T_4$.

This representation is easily shown to be continuous. For computation of Lie derivatives it is necessary to transfer these actions of $\overline{\mathcal{P}}$ onto certain images of $\mathcal{L}^2(T_3^+)$ [or $\mathcal{L}^2(T_3^-)$] associated with the charts on $T_3$ obtained by the projection $\pi$ of $T_3^+$ or $T_3^-$ onto $\mathbf{R}^3$, as shown in Fig. 2. $\pi^{(+)}$ [or $\pi^{(-)}$] is defined by

$$\pi^{(+)}(p_0,p_i) = p_i,$$
$$\pi^{(+)-1}(p_i) = ((\mathbf{p}^2 + m^2)^{1/2}, p_i), \tag{3.11}$$
$$\pi^{(-)}(p_0,p_i) = p_i,$$
$$\pi^{(-)-1}(p_i) = (-(\mathbf{p}^2 + m^2)^{1/2}, p_i). \tag{3.12}$$

With these projections we define the isometric mappings

$$\Pi^{(+)}: \mathcal{L}^2(T_3^+) \to \mathcal{L}^2(\mathbf{R}^3, d\mu_{\mathcal{P}}^+)$$
$$\ni (\Pi^{(+)} f)(p_i) = f(\pi^{(+)-1}(p_i)), \tag{3.13}$$

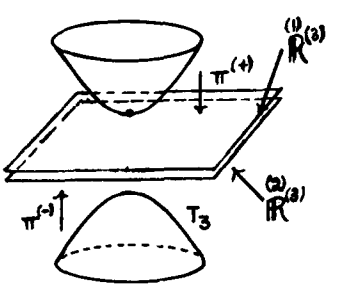



FIG. 2. Projection of $T_3^+$ onto $\mathbf{R}^{3}$ and $T_3^-$ onto $\mathbf{R}^{3}$.

$$\Pi^{(-)}:\mathscr{L}^2(T_3^-)\to\mathscr{L}^2(\mathbb{R}^3,d\mu_{\bar{\mathscr{P}}})$$
$$\ni(\Pi^{(-)}f)(p_i)=f(\pi^{(-)-1}(p_i)),$$

where $\mathscr{L}^2(\mathbb{R}^3,d\mu_{\mathscr{P}}^{\pm})$ [or $\mathscr{L}^2(\mathbb{R}^3,d\mu_{\bar{\mathscr{P}}})$] is the Hilbert space completion of the space of all $f:\mathbb{R}^3\to\mathbb{C}$ which satisfy

$$(f,f)=\int\frac{d^3p}{2(p^2+m^2)^{1/2}}\bar{f}(p)f(p)<\infty.$$

Using $\Pi^{(+)}$ (or $\Pi^{(-)}$) we can define equivalent actions of $\bar{\mathscr{P}}$ on $\mathscr{L}^2(\mathbb{R}^3,d\mu_{\mathscr{P}}^{\pm})$ [or $\mathscr{L}^2(\mathbb{R}^3,d\mu_{\bar{\mathscr{P}}})$] by

$$U^{\pi^{(+)}}=\Pi^{(+)}U\Pi^{(+)-1}$$

or

$$U^{\pi^{(-)}}=\Pi^{(-)}U\Pi^{(-)-1}.$$

Using this equation and Eq. (3.10) we compute the Lie derivatives of the one-parameter subgroups of SO(3,1) rotations and the translations in the charts on $T_3$ defined by $\pi^{(+)}$ or $\pi^{(-)}$. The well-known results are[24,25]

$$\begin{aligned}&[P_i f](p_i)=p_i f(p_i),[P_0 f](p_i)\\&\quad=\pm\sqrt{m^2+\mathbf{p}^2}f(p_i)\in\mathscr{L}^2(\mathbb{R}_3,d\mu_{\mathscr{P}}^{\pm}),\\&[M_{lm}f](p_i)=i\left(p_l\frac{\partial}{\partial p^m}-p_m\frac{\partial}{\partial p^l}\right)f(p_i)\\&[M_{ok}f](p_i)=ip_0\frac{\partial}{\partial p^k}f(p_i)\end{aligned}\Bigg\},\ l,m,k=1,2,3.\tag{3.14}$$

Here $f\in\mathscr{L}^2(\mathbb{R}^3,d\mu_{\mathscr{P}}^{\pm})$ [or $f\in\mathscr{L}^2(\mathbb{R}^3,d\mu_{\bar{\mathscr{P}}})$] is any differentiable vector for the Poincaré group representations $U^{\pi^{(+)}}$ (or $U^{\pi^{(-)}}$). The $+$ sign for the eigenvalue of $P_0$ is for $f\in\mathscr{L}^2(\mathbb{R}^3,d\mu_{\mathscr{P}}^{\pm})$ and the $-$ sign is for $f\in\mathscr{L}^2(\mathbb{R}^3,d\mu_{\bar{\mathscr{P}}})$.

Now we define some local actions of $\overline{\mathrm{SO}_0(4,1)}$ on certain subsets of $\mathbb{R}^3$. Here, O(4,1) is the set of all real linear transformations of $\mathbb{R}^5$ which preserve the form

$$x_0^2-x_1^2-x_2^2-x_3^2-x_4^2.$$

$\overline{\mathrm{SO}_0(4,1)}$ is the simply connected covering group of the component connected to the identity, $\mathrm{SO}_0(4,1)$ of O(4,1).

The linear transformations of $\mathbb{R}^5$ which preserve the cone $C_4$ (see Fig. 1) induce through projection, a local action on $T^3$ as follows: Suppose $x^a$ ($a=0,1,2,3,4$) is a lightlike vector in five dimensions, $x^a x_a=x_0^2-x_1^2-x_2^2-x_3^2-x_4^2=0$. All transformations $\Lambda$ of O(4,1) [and hence of $\overline{\mathrm{SO}_0(4,1)}$] preserve the length of this vector, i.e., $\Lambda^a_{\ b}x^b=y^a(y_a y^a=0)$. To each vector $x^a$, we may associate a vector $p^\mu$ on $T_3$ as $(p^\mu/m)=(x^\mu/x^4)$. Let $U^+$ be an open subset of $T_3^+$ containing the point (1,0,0,0). Choose $V$ to be a neighborhood of the identity in $\overline{\mathrm{SO}_0(4,1)}$ so that for all $p^\mu\in U^+$

$$\begin{aligned}&p^{\mu'}=m\frac{y^\mu}{y^4}=m\frac{(\Lambda x)^\mu}{(\Lambda x)^4}\in T_3^+\\&\left(x=(x^\mu,x^4)\quad\text{with}\ \frac{p^\mu}{m}=\frac{x^\mu}{x^4}\right).\end{aligned}\tag{3.15}$$

Equation (3.15) together with these remarks define an action of $\overline{\mathrm{SO}_0(4,1)}$ on $T_3^+$. In a similar way we may define an action of $\overline{\mathrm{SO}_0(4,1)}$ on $T_3^-$ through the use of Eq. (3.15). [Let $U^-$ be an open subset of $T_3^-$ containing the point $(-1,0,0,0)$.] Through the use of the projections $\pi^{(+)}$ and $\pi^{(-)}$ these local actions induce local actions as defined in Definition 3.1 of $\overline{\mathrm{SO}_0(4,1)}$ on $\pi^{(+)}(U^+)$ and $\pi^{(-)}(U^-)$, which are subsets of $\mathbb{R}^3$.

Next let us define certain generalized infinitesimal actions of the Lie algebra of $\mathrm{SO}_0(4,1)$ on $\mathscr{A}(\pi^{(+)}(U^+))$ and $\mathscr{A}(\pi^{(-)}(U^-))$. The actions are constructed out of the expressions for the generators of $\bar{\mathscr{P}}$ given in Eq. (3.14) and are given by the set of all

$$X=\sum_{\mu,\nu=0}^{3}\alpha^{\mu\nu}M_{\mu\nu}+\Sigma\,\alpha^\nu\frac{1}{\lambda}B_\nu\quad(\alpha^{\mu\nu},\alpha^\nu\in\mathbb{R})$$

with

$$M_{\mu\nu}=(M_{ij},M_{ok})$$

and

$$\frac{1}{\lambda}B_\mu^{\pm}=\frac{1}{\lambda}\left(P_\mu+\frac{\lambda}{2m}\{P^\rho,M_{\rho\mu}\}\right)\quad(0<\lambda\in\mathbb{R}),$$

where the $\pm$ distinguish between the two possibilities for $p_0$ [$p_0=+(m^2+p^2)^{1/2}$ for $\pi^{(+)}(U^+)$ and $p_0=-(m^2+p^2)^{1/2}$ for $\pi^{(-)}(U^-)$]. We argue below that these operators satisfy the commutation relations of the Lie algebra of $\overline{\mathrm{SO}_0(4,1)}$. Thus by Theorem 3.1 these generalized infinitesimal actions of $\mathscr{G}$ on $U^+$ (or $U^-$) generate local multiplier representations $T^{(+)}_{(m^2/\lambda^2)^{1/2}}$ (or $T^{(-)}_{(m^2/\lambda^2)^{1/2}}$ of $\overline{\mathrm{SO}_0(4,1)}$ on $\mathscr{A}(\pi^{(+)}(U^+))$ and $\mathscr{A}(\pi^{(-)}(U^-))$. Their forms are explicitly given by

$$T^{(+)}_{(m^2/\lambda^2)^{1/2}}=\Pi^{(+)}T_{(m^2/\lambda^2)^{1/2}}\Pi^{(+)-1}$$

or

$$T^{(-)}_{(m^2/\lambda^2)^{1/2}}=\Pi^{(-)}T_{(m^2/\lambda^2)^{1/2}}\Pi^{(-)-1},$$

where $T_{(m^2/\lambda^2)^{1/2}}$ is given by $\overline{\mathrm{SO}_0(4,1)}\ni A\to T_{(m^2/\lambda^2)^{1/2}}(A)$ with

$$\begin{aligned}&(T_{(m^2/\lambda^2)^{1/2}}(A)f)(p)\\&\quad=|\mu(\bar{A}^{-1},p)|^{-3/2-i(m^2/\lambda^2)^{1/2}}f(m\bar{A}^{-1}p/m),\end{aligned}\tag{3.16}$$

where $f\in\mathscr{L}^2(T_3^+)$ [or $f\in\mathscr{L}^2\mathscr{L}^2(T_3^-)$] is the image of an $\hat{f}\in\mathscr{A}(\pi^{(+)}(U^+))$ [or $\hat{f}\in\mathscr{A}(\pi^{(-)}(U^-))$] under $\Pi^{(+)}$ (or $\Pi^{(-)}$), and

$$\begin{aligned}\left(\bar{A}^{-1}\frac{p}{m}\right)^\mu&=\frac{a_4^\mu(\bar{A}^{-1})+\Sigma\,a_\nu^\mu(\bar{A}^{-1})(p^\nu/m)}{a_4^4(\bar{A}^{-1})+\Sigma\,a_\nu^4(\bar{A}^{-1})(p^\nu/m)}\\&=\frac{1}{m}\Phi(\bar{A},p),\end{aligned}\tag{3.17}$$

$$\mu(\bar{A}^{-1},p)=a_4^4(\bar{A}^{-1})+\sum_\mu a_\mu^4(\bar{A}^{-1})\frac{p^\mu}{m}$$
$$(\nu(\bar{A},p)=|\mu(\bar{A}^{-1},p)|^{-3/2-i(m^2/\lambda^2)^{1/2}})$$

with

$$\bar{A}=\begin{vmatrix}a^0_{\ 0}&a^0_{\ 1}&a^0_{\ 2}&a^0_{\ 3}&a^0_{\ 4}\\a^1_{\ 0}&a^1_{\ 1}&a^1_{\ 2}&a^1_{\ 3}&a^1_{\ 4}\\a^2_{\ 0}&a^2_{\ 1}&a^2_{\ 2}&a^2_{\ 3}&a^2_{\ 4}\\a^3_{\ 0}&a^3_{\ 1}&a^3_{\ 2}&a^3_{\ 3}&a^3_{\ 4}\\a^4_{\ 0}&a^4_{\ 1}&a^4_{\ 2}&a^4_{\ 3}&a^4_{\ 4}\end{vmatrix}\in P(V)\subset\mathrm{SO}_0(4,1).$$

[$\bar{A}$ is an element of $\overline{\mathrm{SO}_0(4,1)}$ corresponding to the element $A$ of $\mathrm{SO}_0(4,1)$ under the covering projection

$P$: $\overline{SO_0(4,1)} \to SO_0(4,1)$.] The reason why these local multiplier representations cannot be extended to global multiplier representations on a larger subset of $\mathbb{R}^3$ is made clear from an analysis of Eq. (3.17) and Fig. 1. For certain $A$'s in $\overline{SO_0(4,1)}$ but not in $V$, points in $U^+$ or $U^-$ can, under the action of the mapping defined by (3.17), be moved from one branch of the hyperboloid into the other. As seen from Fig. 1, any $A$ which moves points of the left hemisphere of $S_3$ into points of its right hemisphere will interchange points in different branches of the hyperboloid.

Despite the above result, the generalized infinitesimal action of the Lie algebra of SO(4,1) on $\mathscr{A}(\pi^{(+)}(U^+))$ and $\mathscr{A}(\pi^{(-)}(U^-))$ can be extended to representations on functions analytic on all of $\mathbb{R}^3$. In fact, it even gives us a representation of the Lie algebra of $\overline{SO_0(4,1)}$ on the space of differentiable vectors for the associated spin zero, positive mass UIR's of the Poincaré group. We denote these spaces of differentiable vectors by $\mathscr{D}^\infty_+$ and $\mathscr{D}^\infty_-$.

**Theorem 3.3:** The following expressions are the generators for representations of the Lie algebra of $SO_0(4,1)$ by symmetric operators on $\mathscr{D}^\infty_+$ or $\mathscr{D}^\infty_-$

$$M_{\mu\nu}, \tag{3.18}$$

$$\frac{1}{\lambda} B_\mu^\pm = \frac{1}{\lambda}\left(P_\mu + \frac{\lambda}{2m}\{P^\rho, M_{\rho\mu}\}\right), \quad 0<\lambda\in R. \tag{3.19}$$

For $p_0>0$ the representation is on $\mathscr{D}^\infty_+$, and for $p_0<0$ it is on $\mathscr{D}^\infty_-$.

*Proof:* We verify straightforwardly, using the commutation relations of the Lie algebra generators of $\overline{\mathscr{P}}$ that $M_{\mu\nu}$ and $(1/\lambda)B_\mu^\pm$ satisfy the commutation relations of the Lie algebra of SO(4,1)

$$\left[\frac{1}{\lambda} B_\mu^\pm, \frac{1}{\lambda} B_\nu^\pm\right] = iM_{\mu\nu},$$

$$\left[M_{\mu\nu}, \frac{1}{\lambda} B_\rho^\pm\right] = i\left(g_{\nu\rho}\frac{1}{\lambda} B_\mu^\pm - g_{\mu\rho}\frac{1}{\lambda} B_\nu^\pm\right). \tag{3.20}$$

Since we have on $\mathscr{D}^\infty_+$ a representation of $\mathscr{E}(\overline{\mathscr{P}})$, the generators $(1/\lambda)B_\mu^+$ and $M_{\mu\nu}$ leave invariant $D^\infty_+$, and (by Theorem 2.3) are symmetric operators on $D^\infty_+ \subset \mathscr{L}^2(T_3^+)$. Thus the generators $(1/\lambda)B_\mu^+$ and $M_{\mu\nu}$ are a basis for an $SO_0(4,1)$ Lie subalgebra of $dU$ ($\mathscr{E}(\overline{\mathscr{P}})$) on $D^\infty_+$ and they generate a symmetric representation of the Lie algebra of $SO_0(4,1)$ in $\mathscr{L}^2(T_3^+)$. The same arguments show that $(1/\lambda)B_\mu^-$ and $M_{\mu\nu}$ generate a representation of the Lie algebra of $SO_0(4,1)$ on $\mathscr{D}^\infty_- \subset H(m,0,-) = \mathscr{L}^2(T_3^-)$.

Does there exist a unitary representation of $\overline{SO_0(4,1)}$ in either $\mathscr{L}^2(T_3^+)$ or $\mathscr{L}^2(T_3^-)$ such that its Lie algebra representation is the one generated by the symmetric operators $(1/\lambda)B_\mu^+$ or $(1/\lambda)B_\mu^-$ and $M_{\mu\nu}$ which are defined on either of the dense subspaces $\mathscr{D}^+$ or $\mathscr{D}^-$, i.e., a $U(\overline{SO_0(4,1)}; +)$ [or a $U(\overline{SO_0(4,1)}; -)$] such that

$$i(1/\lambda)B_\mu^\pm v = \lim_{t\to 0}(t^{-1}[U(e^{tI_{5\mu}}, \pm)v - v]) \tag{3.21}$$

and

$$iM_{\mu\nu}v = \lim_{t\to 0}(t^{-1}[U(e^{tI_{\mu\nu}})v - v]) \tag{3.22}$$

for $v$ either in $\mathscr{D}^\infty_+$ or $\mathscr{D}^\infty_-$? The answer is no. A very abstract proof of this statement is presented in Ref. 2 (p. 235); it uses the reduction of a UIR of $\overline{SO_0(4,1)}$ in a noncompact "generalized basis" which "diagonalizes" the noncompact $\overline{SO_0(3,1)}$ subgroup. Through the use of Theorem 3.1 we can provide here the geometrical reason for this fact. (We treat the case $p_0>0$.) We can choose $U\subset\mathbb{R}^3$ so that the expressions in (3.18) and (3.19) generate a generalized infinitesimal action on $U$. By Theorem 3.1 this action is the algebra of generalized Lie derivatives for a (unique) local multiplier representation $T_\nu$ of $\overline{SO_0(4,1)}$ and by Definition 3.3 this local multiplier representation is the one described by Eq. (3.16). {Using (3.16) and the mapping $\Pi$ [Eq. (3.13)] compute $(D_x f)(x)$ to obtain (3.18) and (3.19).} However, for the reasons stated above it cannot be extended to a global multiplier representation on $\mathbb{R}^3$, with an action on $T_3^+$ given by Eq. (3.17) for arbitrary $U$'s. Likewise for the case $p_0<0$ we can show there exists no unitary representation of $SO_0(4,1)$ on $\mathscr{L}^2(T_3^-)$ with the generators given by expressions (3.18) and (3.19), and an action given by (3.17).

From this result we know there cannot exist a $\mathscr{D}$ with $\mathscr{D}^\infty_+ \subset \mathscr{D} \subset \mathscr{L}^2(T_3^+)$ [or a $\mathscr{D}$ with $\mathscr{D}^\infty_- \subset \mathscr{D} \subset \mathscr{L}^2(T_3^-)$] with the property that an extension of the Nelson operator on $\mathscr{D}^\infty_+$ (or $\mathscr{D}^\infty_-$),

$$\mathscr{N} = \frac{1}{\lambda^2} B_0^{\pm 2} + \frac{1}{\lambda^2}\sum_{i=1}^3 B_i^{\pm 2} + \frac{1}{2}\sum_{i,j=1}^3 M_{ij}M_{ij} + \sum_{i=1}^3 M_{oi}M_{oi},$$

is essentially self-adjoint on $\mathscr{D}$. This follows from a theorem of Nelson.[26] Otherwise we would have a group representation of $\overline{SO_0(4,1)}$ on $\mathscr{L}^2(T_3^+)$ or $\mathscr{L}^2(T_3^-)$.

From Theorem 3.2 we expect that it might be possible to obtain a global action of $\overline{SO_0(4,1)}$ on some compactification of $T_3^+$ and $T_3^-$. To this end we define a global multiplier representation of $\overline{SO_0(4,1)}$ on a compact manifold $T_3^\infty$, which contains $T_3^+$ and $T_3^-$. This $T_3^\infty$ is obtained from $T_3 = T_3^+ \cup T_3^-$ by the adjunction of a surface at infinity of codimension one such that the map

$$\tau: T_3^\infty \to S_3: \tau(p) = [mp_0^{-1}, (-mp_i/p_0)] \quad (p^\mu \in T_3^\infty) \tag{3.23}$$

establishes a homeomorphism of $T_3^\infty$ onto $S_3$. Its inverse is given by

$$\tau^{-1}(u) = [m^2 u_0^{-1}, (-mu_i/u_0)] \quad (u^\mu \in S_3). \tag{3.24}$$

A neighborhood of a point of infinity of $T_3^\infty$ is chosen so that $\tau$ is to be a homeomorphism. Equation (3.15) then determines a global action of $SO_0(4,1)$ on $T_3^\infty$. For the measure on $T_3$ we take the measure on $T_3^+$ and $T_3^-$ given by Eqs. (3.8) and (3.9). The Hilbert space is the completion of the set of all $f: T_3 \to \mathbb{C}$ such that

$$(f,f) = \int_{T_3^+} \bar{f}(p)f(p)d\mu^+ + \int_{T_3^-} \bar{f}(p)f(p)d\mu^- < \infty. \tag{3.25}$$

(Since an $f:T_3^\infty \to \mathbb{C}$ differs from an $f:T_3 \to \mathbb{C}$ by a set of measure zero it suffices to consider functions on $T_3$.) With these definitions we have

**Theorem 3.4:** A multiplier representation $T_{(m^2/\lambda^2)^{1/2}}$ of $\overline{SO_0(4,1)}$ on the compact manifold $T_3^\infty$ is provided by the operators $T_{(m^2/\lambda^2)^{1/2}}(A)$ for each $A \in \overline{SO_0(4,1)}$ with

$$[T_{(m^2/\lambda^2)^{1/2}}(A)f](p) = \frac{1}{[(\overline{A}^{-1})_4^4 + (\overline{A}^{-1})_\mu^4(p^\mu/m)]^{3/2+i(m^2/\lambda^2)^{1/2}}} \times f\left(m\frac{(\overline{A}^{-1})_4^\mu + (\overline{A}^{-1})_\nu^\mu(p^\nu/m)}{(\overline{A}^{-1})_4^4 + (\overline{A}^{-1})_\nu^4(p^\nu/m)}\right). \tag{3.26}$$

Furthermore, $T_{(m^2/\lambda^2)^{1/2}}$ is a unitary irreducible representation of $\overline{SO_0(4,1)}$ on $\mathcal{L}^2(T_3)$. The proof is presented in Appendix B. Note that the multiplier is undefined for certain $A$'s and $p$'s.

Using Definition 3.7 we may compute the generalized Lie derivatives of an analytic function $f \in \mathcal{A}(T_3)$ with respect to the one-parameter subgroups of $SO_0(4,1)$ rotations along the coordinate planes in the projective space of Fig. 1. In order to explicitly compute the generalized derivatives we use the coordinate systems defined by the projections $\pi^{(\pm)}$ of Fig. 2, along with Eq. (3.26). The results are[27]

$$(D_{I^{ij}}(f\circ\pi^{(\pm)-1}))(p_i) = -i\left(p_i\frac{\partial}{\partial p^j} - p_j\frac{\partial}{\partial p^i}\right)(f\circ\pi^{(\pm)-1})(p_i) = M_{ij}(f\circ\pi^{(\pm)-1})(p_i), \tag{3.27a}$$

$$(D_{I^{ok}}(f\circ\pi^{(\pm)-1}))(p_i) = -i\left(p_0\frac{\partial}{\partial p^k}\right)(f\circ\pi^{(\pm)-1})(p_i) = M_{ok}(f\circ\pi^{(\pm)-1})(p_i) \quad (i,j,k=1,2,3),$$

$$(D_{I^{5\mu}}(f\circ\pi^{(\pm)-1}))(p_i) = (1/\lambda)(P_\mu + (\lambda/2m)\{P^\rho, M_{\rho\mu}\})(f\circ\pi^{(\pm)-1})(p_i), \tag{3.27b}$$

where we have introduced the notation $\pi^{(\pm)}$ to stand for the two maps $\pi^{(+)}$ and $\pi^{(-)}$, so that Eqs. (3.27a) and (3.27b) are two sets of equations, one set for the chart on $T_3^+$ and the other for the chart on $T_3^-$. [We could also compute derivatives at points at infinity using the mapping (3.23) followed by a stereographic projection of $S^3$ onto $\mathbb{R}^3$, but this result is of no importance for us.]

Next we cast (3.27a) and (3.27b) into more useful forms. Consider the Hilbert space direct sum

$$\mathcal{H} = \mathcal{L}^2(T_3^+) \oplus \mathcal{L}^2(T_3^-) = \mathcal{L}^2(T_3). \tag{3.28}$$

Also consider the decompositions

$$\mathcal{A}(T_3) = \mathcal{A}(T_3^+) \oplus \mathcal{A}(T_3^-) \subset \mathcal{H} \tag{3.29}$$

and

$$\phi = \mathcal{D}_+^\infty \oplus \mathcal{D}_-^\infty \subset \mathcal{H}. \tag{3.30}$$

These decompositions are obvious since any function $f$ in $\mathcal{L}^2(T_3)$ is equivalent to the pair $(f_1, f_2)$ with $f_1 \in \mathcal{L}^2(T_3^+)$ and $f_2 \in \mathcal{L}^2(T_3^-)$. Using the decomposition (3.29) we can combine the two sets of equations of (3.27a) and (3.27b) into a convenient matrix form:

$$(D_{I^{\mu\nu}}(f\circ\pi^{(\pm)-1}))(p_i) = \begin{pmatrix} M_{\mu\nu} & 0 \\ 0 & M_{\mu\nu} \end{pmatrix}\begin{pmatrix} (f_1\circ\pi^{(+)-1})(p_i) \\ (f_2\circ\pi^{(-)-1})(p_i) \end{pmatrix}, \tag{3.27a'}$$

$$(D_{I^{5\mu}}(f\circ\pi^{(\pm)-1}))(p_i) = \begin{pmatrix} (1/\lambda)(P_\mu + (\lambda/2m)\{P^\rho, M_{\rho\mu}\}) & 0 \\ 0 & (1/\lambda)(P_\mu + (\lambda/2m)\{P^\rho, M_{\rho\mu}\}) \end{pmatrix}\begin{pmatrix} (f_1\circ\pi^{(+)-1})(p_i) \\ (f_2\circ\pi^{(-)-1})(p_i) \end{pmatrix}. \tag{3.27b'}$$

Viewing $f_1\circ\pi^{(+)-1}$ and $f_2\circ\pi^{(-)-1}$ as functions in $\mathcal{D}_+^\infty$ and $\mathcal{D}_-^\infty$, respectively, we obtain the following theorem whose proof follows from Theorem 3.4.

**Theorem 3.5:** The following expressions are the generators for a unitary representation of the Lie algebra of $\overline{SO_0(4,1)}$ by symmetric operators on $\phi = \mathcal{D}_+^\infty \oplus \mathcal{D}_-^\infty$:

$$\begin{pmatrix} M_{\mu\nu} & 0 \\ 0 & M_{\mu\nu} \end{pmatrix} \text{ and } \begin{pmatrix} (1/\lambda)B_\mu^+ & 0 \\ 0 & (1/\lambda)B_\mu^- \end{pmatrix} \tag{3.31}$$

with

$$\frac{1}{\lambda}B_\mu^\pm = \frac{1}{\lambda}\left(P_\mu + \frac{\lambda}{2m}\{P^\rho, M_{\rho\mu}\}\right).$$

Thus on $\phi$ we have a symmetric representation of the Lie algebra of $\overline{SO_0(4,1)}$ with

$$\begin{pmatrix} M_{\mu\nu} & 0 \\ 0 & M_{\mu\nu} \end{pmatrix} \text{ and } \begin{pmatrix} (1/\lambda)B_\mu^+ & 0 \\ 0 & (1/\lambda)B_\mu^- \end{pmatrix}$$

being the generators of the representation. Furthermore, this is a representation on $\phi$ of the algebra of the generalized Lie derivatives of the global $\overline{SO_0(4,1)}$ multiplier representation $T_{(m^2/\lambda^2)^{1/2}}$ on the compact manifold $T_3^\infty$.

With this we have shown that there exists a unitary representation $T_{(m^2/\lambda^2)^{1/2}}$ of $\overline{SO_0(4,1)}$ on $\mathcal{L}^2(T_3)$ with the property that

$$J_{\mu\nu} = \lim_{t\to 0} t^{-1}[T_{(m^2/\lambda^2)^{1/2}}(e^{tI_{\mu\nu}})f - f] = i\begin{pmatrix} M_{\mu\nu} & 0 \\ 0 & M_{\mu\nu} \end{pmatrix}\begin{pmatrix} f_1 \\ f_2 \end{pmatrix}, \tag{3.32a}$$

$$J_{5\mu} = \lim_{t\to 0} t^{-1}[T_{(m^2/\lambda^2)^{1/2}}(e^{tI_{5\mu}})f - f] = i\begin{pmatrix} (1/\lambda)B_\mu^+ & 0 \\ 0 & (1/\lambda)B_\mu^- \end{pmatrix}\begin{pmatrix} f_1 \\ f_2 \end{pmatrix}, \tag{3.32b}$$

for any $f = (f_1, f_2)$, with $f_1 \in \mathcal{D}_+^\infty$ and $f_2 \in \mathcal{D}_-^\infty$ differentiable vectors of the positive and negative mass unitary representations $\mathcal{H}(m,0;+)$ and $\mathcal{H}(m,0;-)$ of $\overline{\mathcal{P}}$.

Now we show that the Nelson operator constructed out of the

$$\begin{pmatrix} L_{\mu\nu} & 0 \\ 0 & L_{\mu\nu} \end{pmatrix} \text{ and } \begin{pmatrix} (1/\lambda)B_\mu^+ & 0 \\ 0 & (1/\lambda)B_\mu^- \end{pmatrix},$$

$$\mathcal{N}=\begin{bmatrix} \frac{1}{\lambda^2}B_0^{+2}+\frac{1}{\lambda^2}\sum_{i=1}^{3}B_i^+B_i^+ +\sum_{i=1}^{3}M_{oi}^2+\sum_{i,j=1}^{3}M_{ij}M_{ij} & 0 \\ 0 & \frac{1}{\lambda^2}B_0^{-2}+\frac{1}{\lambda^2}\sum_{i=1}^{3}B_i^-B_i^- +\sum_{i=1}^{3}M_{oi}^2+\sum_{i,j=1}^{3}M_{ij}M_{ij} \end{bmatrix}, \tag{3.33}$$

which is a symmetric operator defined on $\phi$, has an essentially self-adjoint extension to a larger domain ($\mathscr{D}_\infty$), namely the Nelson operator for the $\overline{SO_0(4,1)}$ multiplier representation, $T_{(m^2/\lambda^2)^{1/2}}$ defined on the space $\mathscr{D}^\infty$ of differentiable vectors for this $\overline{SO_0(4,1)}$ group representation. First observe that $\phi=\mathscr{D}_+^\infty\oplus\mathscr{D}_-^\infty$ satisfies conditions (i) and (ii) of Theorem 2.2 for the representation of the Lie algebra of $SO_0(4,1)$ considered in Theorem 3.5. [Clearly

$$\begin{pmatrix}(1/\lambda)B_\mu^+ & 0\\ 0 & (1/\lambda)B_\mu^-\end{pmatrix}\quad\text{and}\quad\begin{pmatrix}M_{\mu\nu} & 0\\ 0 & M_{\mu\nu}\end{pmatrix}$$

stabilize $\phi$, and $v\in\phi\Rightarrow v\in\cap_{a,b}\mathscr{D}(J_{ab})$ since

$$v\in\phi\Rightarrow\begin{pmatrix}(1/\lambda)B_\mu^+ & 0\\ 0 & (1/\lambda)B_\mu^-\end{pmatrix}v\in\mathscr{L}^2(T_3)$$

and

$$\begin{pmatrix}M_{\mu\nu} & 0\\ 0 & M_{\mu\nu}\end{pmatrix}v\in\mathscr{L}^2(T_3)$$

so that $J_{ab}v\in\mathscr{L}^2(T_3)$ for all $a,b=0,1,2,3,4$ by the above expressions for $J_{ab}$.] Therefore, $\phi=\mathscr{D}_+^\infty\oplus\mathscr{D}_-^\infty$ satisfies conditions (i) and (ii) of Theorem 2.2, so $\phi\subset\mathscr{D}^\infty$. We can define extensions of the operators (3.31) to be the $J_{\mu\nu}$'s of (3.32a) and the $J_{5\mu}$'s of (3.32b) defined on their maximal common invariant domain, $\mathscr{D}^\infty$. They are all symmetric on $\mathscr{D}^\infty$, since, according to 3.4, $T_{(m^2/\lambda^2)^{1/2}}$ is unitary (Theorem 2.3) so they are symmetric extensions of the expressions (3.32a) and (3.32b).

Now we claim $\mathcal{N}$ has an essentially self-adjoint extension $\mathcal{N}^{\text{ext}}$ to $\mathscr{D}^\infty$. We take the extension of the Nelson operator $\mathcal{N}$ on $\phi$ to be the Nelson operator constructed out of the infinitesimal generators (3.32a) and (3.32b) of the $SO_0(4,1)$ multiplier representation $T_{(m^2/\lambda^2)^{1/2}}$. Here $\mathcal{N}^{\text{ext}}$ agrees with $\mathcal{N}$ on $\phi$. We claim it is essentially self-adjoint (e.s.a.) on $\mathscr{D}^\infty$. To show this we prove the following result.

**Theorem 3.6:** Let $A$ be a symmetric operator with domain $D\subset\mathscr{H}$. Let $D_1\subset D$ be a dense linear subset of $\mathscr{H}$ and suppose that $A$ restricted to $D_1$ ($A\,|D_1$) is e.s.a. then $A$ is e.s.a. Although the theorem is well known, its proof is not so straightforward, so we present it in Appendix C.

Now the Nelson operator for a unitary group representation is e.s.a. on the space of analytic vectors $A$ (see Ref. 28) and $A\subset\mathscr{D}^\infty$ ($U(G)$) (see Ref. 29) so by Theorem 3.6

$$\mathcal{N}^{\text{ext}}=J_{o5}J^{o5}+J_{i5}J^{i5}+J_{oi}J^{oi}+\tfrac{1}{2}J_{ik}J^{ik}$$

on $\mathscr{D}^\infty=\mathscr{D}^\infty(T_{(m^2/\lambda^2)^{1/2}})$ is e.s.a.

Now let us carry the representation $T_{(m^2/\lambda^2)^{1/2}}$ of $\overline{SO_0(4,1)}$ over into $\mathscr{L}^2(T_3^+)\oplus\mathscr{L}^2(T_3^+)$ and obtain analogous results about integrability of the Lie algebra representation on a dense subspace of this space. For this purpose we define the unitary operator

$$\theta:\mathscr{L}^2(T_3^+)\oplus\mathscr{L}^2(T_3^-)\to\mathscr{L}^2(T_3^+)\oplus\mathscr{L}^2(T_3^+)$$

by

$$\mathscr{L}^2(T_3^+)\oplus\mathscr{L}^2(T_3^-)$$
$$\ni v(p)\to(\theta v)(p)$$
$$=(v_+(p),v_-(-p))\in\mathscr{L}^2(T_3^+)\oplus\mathscr{L}^2(T_3^+).$$

[$v_+(p)$ and $v_-(p)$ are the components of $v(p)$ in $\mathscr{L}^2(T_3^+)$ and $\mathscr{L}^2(T_3^-)$, respectively.] It is readily shown to be unitary, and we define "abstractly" a representation of $\overline{SO_0(4,1)}$ on $\mathscr{L}^2(T_3^+)\oplus\mathscr{L}^2(T_3^+)$ by $\hat{T}_{(m^2/\lambda^2)^{1/2}}=\theta T_{(m^2/\lambda^2)^{1/2}}\theta^{-1}$. Here, $\hat{T}_{(m^2/\lambda^2)^{1/2}}$ is constructed so that $\hat{T}_{(m^2/\lambda^2)^{1/2}}$ and $T_{(m^2/\lambda^2)^{1/2}}$ are unitarily equivalent. Because of this unitary equivalence, we have also an extension of a similar Lie algebra representation defined on the direct sum of the spaces of differentiable vectors of two positive energy spin-zero (positive mass) UIR's of $\overline{\mathscr{P}}$. This proves (for the case $s=0$) the result conjectured in Ref. 2.

Finally we remark that the multiplier representations of $\overline{SO_0(4,1)}$ on $\mathscr{L}^2(T_3)$, which we have constructed, are equivalent to certain multiplier representations of $\overline{SO_0(4,1)}$ on $\mathscr{L}^2(S_3)$. The unitary equivalence is established through the use of the mapping $\tau$ defined in (3.23). (See Refs. 3 and 4.)

## IV. THE INTEGRABILITY OF THE LIE ALGEBRA REPRESENTATION OF $\overline{SO_0(4,1)}$ ON $\Phi=\Phi(m,s;+)\oplus\Phi(m,s;-)\subset\mathscr{H}(m,s;+)\oplus\mathscr{H}(m,s;-)$

In this section we consider an arbitrary spin unitary irreducible representation of $\overline{\mathscr{P}}$—$\mathscr{H}(m,s;\pm)$—and state results analogous to those which we proved for the spin-zero case. First we note that there exists a generalization of Theorem 3.4 to arbitrary spin: There exists a continuous, unitary irreducible representation of $\overline{SO_0(4,1)}$ on the Hilbert space direct sum

$$\mathscr{H}(m,s;+)\oplus\mathscr{H}(m,s;-).$$

This generalization to arbitrary spin, which is described in Ref. 4, involves replacing complex-valued functions on $T_3$ with complex vector-valued functions on $T_3$, and a suitable modification of the multiplier $\mu(\bar{A},p)$ together with an internal $\overline{SO_0(4,1)}$ rotation of the finite-dimensional spaces in which the functions take their values[30]; so in principle, the analysis of the last section should carry over with little difference, except for modifications that come about from dealing with vector-valued functions on $T_3$.

The analog of Theorem 3.5 for arbitrary spin involves only the addition of a term $S_{\mu\nu}$ (Ref. 31) to the orbital angular momentum tensor $M_{\mu\nu}$. The $S_{\mu\nu}$'s generate various Lorentz transformations in the internal space of the vector-valued functions. Using this remark we may state the following ana-

log of Theorem 3.5 for arbitrary spin.

**Theorem 4.1:** The following expressions are the generators for a unitary representation of the Lie algebra of $\overline{SO_0(4,1)}$ by symmetric operators on $\Phi(m,s) = \Phi(m,s;+) \oplus \Phi(m,s;-)$

$$\begin{pmatrix} L_{\mu\nu} & 0 \\ 0 & L_{\mu\nu} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \frac{1}{\lambda} B_\mu^+ & 0 \\ 0 & \frac{1}{\lambda} B_\mu^- \end{pmatrix} \tag{4.1}$$

with

$$\frac{1}{\lambda} B_\mu^\pm = \frac{1}{\lambda}\left(P_\mu + \frac{\lambda}{2m}\{P^\rho, L_{\rho\mu}\}\right)$$

and $L_{\mu\nu} = M_{\mu\nu} + S_{\mu\nu}$. Therefore, on $\Phi$ we have a symmetric representation of the Lie algebra of $\overline{SO_0(4,1)}$ in which

$$\begin{pmatrix} L_{\mu\nu} & 0 \\ 0 & L_{\mu\nu} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \frac{1}{\lambda} B_\mu^+ & 0 \\ 0 & \frac{1}{\lambda} B_\mu^- \end{pmatrix}$$

are the generators of the representation.

In order to prove this theorem we must first determine the form of the generators of the Poincaré group representation acting on $\mathscr{H}(m,s;\pm)$. These expressions are well known and are given by[31] generators of translations $P_\mu$ and generators of Lorentz transformations $L_{\mu\nu} = M_{\mu\nu} + S_{\mu\nu}$, where $P_\mu$ and $M_{\mu\nu}$ are given by (3.14). Thus the $(1/\lambda) B_\mu^\pm$ and $L_{\mu\nu}$ of Eq. (4.1) are, as in Sec. III, bases for representations of the $\overline{SO_0(4,1)}$ Lie algebra of $\mathscr{E}(\overline{\mathscr{P}})$ in $\Phi(m,s;+)$ and $\Phi(m,s;-)$. Furthermore, they generate symmetric representations of $\overline{SO_0(4,1)}$ in $\Phi(m,s;+)$ and $\Phi(m,s;-)$ (by Theorem 2.3), and therefore the matrix expressions in Eq. (4.1) generate a symmetric representation of $\overline{SO_0(4,1)}$ in $\Phi(m,s)$. Finally we claim that the expressions in (4.1) are the generators of a unitary irreducible representation of $\overline{SO_0(4,1)}$ on $\mathscr{H}(m,s;+) \oplus \mathscr{H}(m,s;-)$. To prove this we must know the explicit form of the generalization of the multiplier representation of Theorem 3.4 and then, using this form, we must calculate the infinitesimal generators of the representation. The calculation, which is basically the same as the calculation of Eqs. (3.27a) and (3.27b) (except the functions now are vector-valued functions), is given in Ref. 4. We obtain for the infinitesimal generators exactly those matrix expressions in (4.1), and for the same reasons as in Sec. III, they can be applied to any $v \in \Phi(m,s)$. These remarks complete the proof.

The proof that the Nelson operator on $\Phi(m,s)$ constructed out of the $(1/\lambda) B_\mu^\pm$ and $L_{\mu\nu}$ in (4.1) has an e.s.a. extension to the space of differentiable vectors for an $\overline{SO_0(4,1)}$ representation is essentially identical to the proof for the spin-zero case. [We use the fact that there exists a unitary irreducible representation of $\overline{SO_0(4,1)}$ on $\mathscr{H}(m,s;+) \oplus \mathscr{H}(m,s;-)$; and the infinitesimal generators of the representation when restricted to $\Phi(m,s)$ are given by the matrix expressions of Eq. (4.1).]

Finally we note that it is possible to obtain realizations of the principal series representations of $\overline{SO_0(4,1)}$ on the Hilbert space direct sum of two identical irreducible unitary representations of $\overline{\mathscr{P}}$

$$\mathscr{H}(m,s;+) \oplus \mathscr{H}(m,s;+).$$

The proof of this fact involves an obvious generalization of the isometric isomorphism $\theta$ defined in Sec. III to vector-valued functions on $T_3$ (see Ref. 4). Using this fact we can establish that there exists an essentially self-adjoint extension of the operator which is the image under $\theta$ of the above Nelson operator on $\Phi(m,s)$, i.e., an extension of the Nelson operator on $\Phi(m,s;+) \oplus \Phi(m,s;-)$ constructed out of only $(1/\lambda) B_\mu^+$ and $L_{\mu\nu}$.


## ACKNOWLEDGMENTS

We wish to thank Dr. Fred Weissler for help with the proof of Theorem 3.6. One of the authors (PM) gratefully acknowledges that part of the work for this paper was completed during the tenure of a fellowship from Deutscher Akademischer Austauschdienst (DAAD).


## APPENDIX A: PROOF OF THEOREM 2.2

*Proof:* First we show $\mathscr{D}^\infty$ satisfies properties (i) and (ii). (ii) is true by Theorem 2.1. If $v \in \mathscr{D}^\infty$ then the map $f: t \to U(e^{tI_\alpha})v$ is $C^\infty$ for all the generators $I_\alpha$ of $\mathscr{G}$. But this implies $J_\alpha v$ exists since $iJ_\alpha v$ is the derivative of $f$ at $t = 0$. Hence,

$$v \in \bigcap_{\alpha=1}^{n} \mathscr{D}(J_\alpha).$$

Now suppose there exists a linear subspace $\mathscr{D}'$ of $\mathscr{H}$ which satisfies conditions (i) and (ii) of the theorem and $\mathscr{D}^\infty \subset \mathscr{D}'$. Let $v \in \mathscr{D}'$ then by (ii) $J_\alpha v \in \mathscr{D}'$ and so by induction, for each $\alpha$, $J_\alpha^m v \in \mathscr{D}'$ for $m = 1,2,\ldots$. But then, by Lemma 2.1 $v \in \mathscr{D}^\infty$. Consequently $\mathscr{D}' = \mathscr{D}^\infty$. Q. E. D.

## APPENDIX B: PROOF OF THEOREM 3.4

To show that $T_{(m^2/\lambda^2)^{1/2}}$ is a multiplier representation of $\overline{SO_0(4,1)}$ we must show $T_{(m^2/\lambda^2)^{1/2}}(A)$ is bounded and continuous and that

$$\mu(\overline{A}, p) = (\overline{A}^{-1})_4^4 + (\overline{A}^{-1})_\mu^4 (p^\mu/m)$$

is a continuously differentiable function of $A$ and $p^\mu$, which satisfies

(i) $\mu(e, p) = 1, \quad \forall\, p \in T_3^\infty$,

(ii) $\mu(\overline{A}_1\overline{A}_2, p) = \mu(\overline{A}_1, p)\mu(\overline{A}_2, \overline{A}_1 p), \quad p \in T_3^\infty$.

When we prove that $T_{(m^2/\lambda^2)^{1/2}}$ is unitary, the boundedness requirement will follow immediately. Continuity means if the sequence $A_n \in G$ converges to $A$, then $U(A_n)\psi \to U(A)\psi$ for every $\psi \in \mathscr{L}^2(T_3)$. This is clear since the multiplier is an a.e. continuously differentiable function of $A$ and $p^\mu$ and if $A_n \to A$, then $f(m\overline{A}_n^{-1} p/m) \to f(m\overline{A}^{-1} p/m)$. Condition (i) for $\mu(\overline{A}, p)$ is easily seen to be valid. Condition (ii) can be shown by a direct calculation.[4] Thus $T_{(m^2/\lambda^2)^{1/2}}$ is a multiplier representation of $\overline{SO_0(4,1)}$ on $\mathscr{L}^2(T_3)$.

To prove $T_{(m^2/\lambda^2)^{1/2}}$ is unitary we first observe that the measure on $T_3$, [see Eq. (3.25)], which we denote by $d\Omega$, transforms in the following way under $A \in \overline{SO_0(4,1)}$ (Refs. 3 and 4):

$$d\Omega' = [\,|a_4^4(\overline{A}^{-1}) + \Sigma a_\nu^4(\overline{A}^{-1})(p^\nu/m)|^3\,]^{-1}\, d\Omega. \quad \text{(B1)}$$

Consequently we have for $f, g \in \mathscr{L}^2(T_3)$

$$(T_{(m^2/\lambda^2)^{1/2}}(A)f, T_{(m^2/\lambda^2)^{1/2}}(A)g) = (f,g),$$

where we have used Eqs. (3.25), (3.26), and the above equation for the transformation property of the measure. Irreducibility of this representation is proved in Ref. 3 (Bander and Itzykson) and also Ref. 4 by demonstrating the unitary equivalence between $T_{(m^2/\lambda^2)^{1/2}}$ $\overline{SO_0(4,1)}$ and another representation of $\overline{SO_0(4,1)}$ on $\mathscr{L}^2(S_3)$. In the quoted reference, irreducibility for the $\overline{SO_0(4,1)}$ representation on $\mathscr{L}^2(S_3)$ is explicitly demonstrated.[32]

## APPENDIX C: PROOF OF THEOREM 3.6

*Proof:* We have (1) $A$ symm$\Rightarrow A^{**} = \overline{A}$ ($L$ symm$\Rightarrow L \subset L^* \Rightarrow D(L^*)$dense$\Rightarrow A^{**} = \overline{A}$ (Ref. 33)).

Also (2) $A$ symmetric$\Rightarrow\overline{A}$ symmetric [Proof: $A \subset A^* \Rightarrow A^{**} \subset A^*$ (take *)$\Rightarrow A^{**} \subset (A^{**})^*$ (take *). But $\overline{A} = A^{**}$ and $\overline{A}^* = (A^{**})^*$. $\therefore \overline{A} \subset \overline{A}^*$.] Further (3) $A$ e.s.a. $\Rightarrow A^*$ s.a. [Proof: $A^* = (A^{**})^* = (\overline{A}) = \overline{A}$, since the closure of an e.s.a. operator is self-adjoint and also we have used the fact that $A$ symm$\Rightarrow A^{**} = \overline{A}$.]

Now let $A$ be the operator in the hypothesis of the theorem; then

$$A_1 \subset A \text{ so } A^* \subset A_1^* = A_1^{**} \quad [\text{since } A_1^* \text{ is s.a. by (3)}].$$

Also

$$A_1^* = A_1^{***} \subset A^{**} = \overline{A} \text{ (Ref. 33)}$$

$$\therefore A_1^* \subset \overline{A}.$$

But $A_1^*$ is s.a., and $\overline{A}$ is symmetric by (2). So we will be finished provided we show for any two operators $L_1, L_2$ with $L_1$ s.a., $L_2$ symmetric and $L_1 \subset L_2$ that $L_2$ is s.a. (which means $L_1 = L_1$). To see this: $L_1 \subset L_2 \Rightarrow L_2^* \subset L_1^*$ but $L_2$ symmetric and $L_1$ s.a., so $L_2 \subset L_2^* \subset L_1^* = L_1$. $\therefore L_2 \subset L_1 \Rightarrow L_2 = L_1$. (Thus $\overline{A} = A_1^* = \overline{A}_1$, so $\overline{A}$ is s.a. so $A$ is e.s.a.) Q. E. D.


[1]Harish-Chandra, Trans. Am. Math. Soc. **75**, 185 (1953).

[2]A. Bohm, "Generalized Eigenvectors and Group Representations—The Connection between Representations of SO(4,1) and the Poincaré Group," in *Studies in Mathematical Physics*, edited by A. O. Barut (Reidel, New York, 1973).

[3]M. Bander and C. Itsykson, Rev. Mod. Phys. **38**, 330 (1966); M. Bander and C. Itzykson, Rev. Mod. Phys. **38**, 346 (1966); H. Abarbanel, in *Studies in Mathematical Physics*, edited by E. Lieb, B. Simon, and A. S. Wightman (Princeton U. P., Princeton, NJ, 1976).

[4]P. Moylan, J. Math. Phys. **24**, 2706 (1983).

[5]F. Warner, *Foundations of Differentiable Manifolds and Lie Groups* (Scott, Foresman, London, 1971), Chap. 1.

[6]W. Miller, Jr., *Lie Theory and Special Functions* (Academic, New York, 1968), p. 2.

[7]G. Warner, *Harmonic Analysis on Semi-Simple Lie Groups* I (Springer, New York, 1972), p. 254.

[8]N. Jacobson, *Lie Algebras* (Dover, New York, 1979), pp. 156–163.

[9]R. Goodman, "Analytic and Entire Vectors for Representations of Lie Groups," Trans. Am. Math. Soc. **143**, 55 (1969).

[10]B. Nagel, "Generalized Eigenvectors in Group Representations," in *Studies in Mathematical Physics*, edited by A. O. Barut (Reidel, New York, 1973), p. 140.

[11]A. Knapp, Lectures on Representation Theory of Lie Groups, Princeton, 1979 (unpublished).

[12]L. Gårding, Proc. Natl. Acad. Sci. USA **33**, 331 (1947).

[13]E. Nelson, Ann. Math. **70**, 572 (1959).

[14]Reference 6, Chap. 1.

[15]R. S. Palais, "A Global Formulation of the Lie Theory of Transformation Groups," Mem. Am. Math. Soc. **22**, 1 (1957).

[16]Reference 6, p. 13.

[17]S. Lie, *Theorie der Transformationsgruppen I, II, III*, with collaboration of F. Engel (Teubner, Leipzig, 1888, 1890, 1893).

[18]For the proof see Ref. 6, p. 19.

[19]Reference 15, p. 34. (What we call a global $G$ action or an infinitesimal $\mathscr{G}$ action is called in Ref. 15 a global $G$-transformation group or an infinitesimal $G$-transformation group, respectively.)

[20]Reference 15, p. 82, Corollary 2.

[21]E. P. Wigner, in *Istanbul Summer School of Theoretical Physics*, edited by F. Gursey (Gordon and Breach, New York, 1964).

[22]Reference 3, Bander and Itzykson; K. C. Hannabuss, Proc. Cambridge Philos. Soc. **70**, 283 (1971).

[23]V. Bargmann and E. P. Wigner, Proc. Natl. Acad. Sci. USA **34** (5), 211 (1946).

[24]A. Chakrabarti, J. Math. Phys. **7**, 426 (1966).

[25]L. Fonda and G. C. Ghirardi, *Symmetry Principles in Quantum Physics* (Dekker, New York, 1970), Chap. 5.

[26]E. Nelson, Ann. Math. **70**, 572 (1959).

[27]Reference 4.

[28]Reference 26.

[29]Reference 7, p. 278.

[30]P. Moylan, J. Math. Phys. **26**, 29 (1985). In this work two equivalent realizations of the arbitrary spin principal series unitary representations of $\overline{SO_0(4,1)}$ are given: One is the realization of the induced representations on $T_3$, described in Hannabuss, Ref. 22; the other is the covariant realization of Ref. 4. The equivalence of these two realizations is explicitly demonstrated, and the infinitesimal generators for both realizations are determined.

[31]H. Joos, Fortsch. Phys. **10**, 65 (1962).

[32]Reference 3, Bander and Itzykson, pp. 345–354 and 355–356. Reference 4, Appendix A.

[33]M. Reed and B. Simon, *Functional Analysis* (Academic, New York, 1972), Vol. I, p. 253.

# A Jacobson–Morozov lemma for sp(2$n$, R)

Hüseyin Koçak
*Lefschetz Center for Dynamical Systems, Division of Applied Mathematics, Brown University, Providence, Rhode Island 02912*

(Received 13 April 1984; accepted for publication 27 July 1984)

A variation of the lemma of Jacobson–Morozov on the imbedding of a nonzero nilpotent element of the real symplectic algebra into the split simple three-dimensional Lie algebra is proved. The proof is algorithmic and relies on our earlier work on the theory of normal forms for the real symplectic algebra.

## I. INTRODUCTION

Let $(\mathbb{R}^{2n}, \omega)$ be the $2n$-dimensional real vector space with a symplectic form $\omega$. Consider a real analytic Hamiltonian function $H: \mathbb{R}^{2n} \to R$ with power series $H = \oplus_{i=2}^{\infty} H_i$, where $H_i$ is a degree $i$ real homogeneous polynomial. When investigating the local behavior of the corresponding Hamiltonian vector field $X_H$, one of the most powerful techniques is to eliminate or to simplify the terms of $H$ through successive applications of canonical transformations, also known as the (nonlinear) normal form theory.

The linear vector field $X_{H_2}$ viewed as a first-order differential operator maps the space of $n$-degree homogenous polynomials into itself. Putting $H$ into a normal form, then, reduces to the problem of finding suitable generators for the complement of the image of this map. The linear Hamiltonian vector field $X_{H_2}$ belongs to the real symplectic algebra sp(2$n$, $\mathbb{R}$) and, in principle, the normal form problem can be solved using representation theory of sp(2$n$, $\mathbb{R}$) on the graded algebra of real homogenous polynomials.

Recently, Cushman and his collaborators[1] carried out this program successfully for systems with two degrees of freedom using representation theory of sp(4, $\mathbb{R}$). Following the customary first step in representation theory of Lie algebras,[2] they first imbed the nilpotent part of $X_{H_2}$, in a certain way, into the split simple three-dimensional Lie algebra sl(2, $\mathbb{R}$).

The purpose of this paper is to prove a variant of the classical Jacobson–Morozov lemma[3,4] by generalizing their imbedding for any element of sp(2$n$, $\mathbb{R}$) with nonzero nilpotent part. A similar lemma can be proved for other classical Lie algebras by using the results in Refs. 5 and 6.

## II. A JACOBSON–MOROZOV LEMMA

*Lemma:* Let $A$ be an element of sp(2$n$,$\mathbb{R}$) and $A = N + S$ be its Jordan–Chevalley decomposition, that is, $N$ is nilpotent, $S$ is semisimple with $NS = SN$. Suppose that $N \neq 0$. Then there exist two elements $X$ and $Y$ in sp(2$n$, $\mathbb{R}$) contained in the centralizer of $S$ satisfying the commutation relations

$$[X, Y] = 2Y, \quad [X, N] = -2N, \quad [Y, N] = X.$$

*Proof:* Using symplectic change of bases, that is, elements of the real symplectic group Sp(2$n$, $\mathbb{R}$) we can first put $A$ into a (linear) normal form. For our purpose a particularly suitable complete set of all possible normal forms of $A$ is given in List II of Ref. 6, which is also reproduced as List I in Ref. 7. The notation used below refers to this list.

It suffices to prove the lemma only for the elements of sp(2$n$, $\mathbb{R}$) which are in normal form and indecomposable. Notice, first, that in the list the nilpotent parts $N$ are in the classical Jordan normal form. Therefore, it follows from the proof of the Morozov lemma for gl($n$, $\mathbb{C}$) given by Jacobson[3] that the elements $N$, $X$, and $Y$ described below satisfy the desired commutation relations. It remains to show, then, that in the special basis chosen in the list, $X$ and similarly $Y$ satisfy the following matrix equations:

$$\omega X = -X^t \omega, \tag{1}$$

$$SX = XS. \tag{2}$$

To solve (1), let $\omega = (\omega^{\alpha,\beta})$ and $X = (X^{\alpha,\beta})$ be their block divisions according to the sizes of the diagonal blocks of $\omega$. It is evident from the list that

$$\omega^{\alpha,\alpha} = \epsilon^{\alpha,\alpha} \begin{bmatrix} & & & 1 \\ & & \iddots & -1 \\ & 1 & & \\ -1 & & & \end{bmatrix}, \quad \text{with } \epsilon^{\alpha,\alpha} = \pm 1,$$

$$\omega^{\alpha\beta} = 0, \quad \text{if } \alpha \neq \beta.$$

Suppose that $X^{\alpha,\beta}$ is of size $k \times l$. Then $X$ satisfies (1) if and only if

$$(-1)^{i+1} X^{\alpha,\beta}_{k-i+1,j} = (-1)^{j+1} X^{\beta,\alpha}_{l-j+1,i} \tag{3}$$

with $i = 1, \ldots, k$ and $j = 1, \ldots, l$.

Using (3) it is easy to verify that $X$ and $Y$ described below are infinitesimally symplectic.

*Case I:* Suppose that $A = N$ is as given in (1) of the list. Then

$$X_{i,i} = m - 2(i-1), \qquad i = 1, \ldots, m+1,$$

$$X_{i,j} = 0, \qquad \text{otherwise},$$

and

$$Y_{i,i+1} = -im + i(i-1), \qquad i = 1, \ldots, m,$$

$$Y_{i,j} = 0, \qquad \text{otherwise}.$$

*Case II:* Suppose that $A = N$ is as given in (2) of the list. Then

$$X_{i+n(m+1),i+n(m+1)} = m - 2(i-1),$$

$$i = 1, \ldots, m+1 \text{ and } n = 0, 1,$$

$$X_{i,j} = 0, \quad \text{otherwise},$$

and

 0022-2488/85/030375-02$02.50 © 1985 American Institute of Physics 

$$Y_{i+n(m+1),i+1+n(m+1)} = -im + i(i-1),$$
$$i = 1, \ldots, m \text{ and } n = 0,1,$$
$$Y_{i,j} = 0, \quad \text{otherwise.}$$

In the remaining cases semisimple parts of the indecomposable types in the list are no longer zero. Therefore, one needs to verify that $X$ and $Y$ given below commute with $S$. This is immediate, however, since the semisimple parts are piecewise multiples of the identity.

*Case III:* Suppose that $A = N + S$ is as given in (3), (4), or (5) of the list. Then $X$ and $Y$ are the same as in Case II above.

*Case IV:* Suppose that $A = N + S$ is as given in (6) of the list. Then

$$X_{i+n(m+1),i+n(m+1)} = m - 2(i-1),$$
$$i = 1, \ldots, m+1, \quad n = 0, 1, 2, 3,$$
$$X_{i,j} = 0, \quad \text{otherwise,}$$

and

$$Y_{i+n(m+1),i+1+n(m+1)} = -im + i(i-1),$$
$$i = 1, \ldots, m \text{ and } n = 0, 1, 2, 3,$$
$$Y_{i,j} = 0, \quad \text{otherwise.}$$

This concludes the proof of the lemma.

The imbedding above is algorithmic in the sense that following Burgoyne and Cushman[8] one can first put any element $A$ of $\mathrm{sp}(2n, \mathbb{R})$ into the normal form as a direct sum of the entries of the list and then use the infinitesimally symplectic matrices $X$ and $Y$ given in the proof above.

Notice that the bases in the list are not the standard one. Using the information in the two lists given in Ref. 6, however, the elements $N$, $X$, and $Y$ can easily be written in the standard symplectic basis.

## ACKNOWLEDGMENTS

I am grateful to Professor Richard Cushman for suggesting the problem.

Work was supported in part by the National Science Foundation under NSF-MCS 8205355.

[1] R. Cushman, A. Deprit, and R. Mosak, J. Math. Phys. **24**, 2103–2117 (1983).
[2] J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory* (Springer-Verlag, Berlin, 1980).
[3] *N. Jacobson, Lie Algebras* (Interscience, New York, 1960), pp. 83–98.
[4] S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces* (Academic, New York, 1978).
[5] N. Burgoyne and R. Cushman, J. Algebra **44**, 339–362 (1977).
[6] H. Kocak, J. Differential Equations **51**, 359–407 (1984).
[7] H. Kocak, J. Math. Phys. **23**, 2375–2380 (1982).
[8] N. Burgoyne and R. Cushman, *The 1976 NASA Conference on Geometric Control Theory* (Math-Sci. Press, Brookline, MA, 1977), pp. 483–529.

# Representations of Kac–Moody algebras by step algebras

Jouko Mickelsson
*Department of Mathematics, University of Jyväskylä, Seminaarinkatu 15, SF-40100 Jyväskylä 10, Finland*

(Received 18 May 1984; accepted for publication 3 August 1984)

Representations of a Kac–Moody algebra $\hat{\mathfrak{g}}$ associated to a semisimple Lie algebra $\mathfrak{g}$, which decompose to a direct sum of finite-dimensional representations of the subalgebra $\mathfrak{g}$, are constructed using a step algebra method. The cases $\mathfrak{g} = \mathfrak{su}(2)$, $\mathfrak{su}(3)$ are considered in detail.

## I. INTRODUCTION AND NOTATION

Kac–Moody algebras appear in several places in classical and quantum physics. It is known that certain classical field equations in $1+1$ space-time dimensions admit infinite-dimensional symmetry groups which have a Kac–Moody type of Lie algebra.[1] In $2+1$ dimensions the current algebra of a Yang–Mills theory with an anomalous term (topological mass) is a full Kac–Moody algebra (including the central term) provided that certain nontrivial boundary conditions for the gauge fields are satisfied.[2] In four space-time dimensions the self-duality equations for the gauge potentials possess a Kac–Moody symmetry.[3] There are also several models in $(1+1)$-dimensional quantum field theory which lead to Kac–Moody algebras.[4] I have given here only a few references to the fastly growing literature; more can be found for example in the review article by Dolan.[5]

Let $\mathfrak{g}$ be a finite-dimensional Lie algebra with the Killing form $\langle x, y\rangle := \operatorname{tr} \operatorname{ad} x \cdot \operatorname{ad} y$, $x, y \in \mathfrak{g}$ and $\operatorname{ad} x: \mathfrak{g} \to \mathfrak{g}$, $\operatorname{ad} x(y) := [x, y]$. Let $\hat{\mathfrak{g}}$ be the linear space of Laurent polynomials $p(z, z^{-1})$ with values in $\mathfrak{g}$. There is a natural grading by degree and we can write

$$\hat{\mathfrak{g}} = \bigoplus_{n \in \mathbf{Z}} \mathfrak{g}^{(n)}. \tag{1.1}$$

If $x \in \mathfrak{g}$ denote $x^{(n)} := x \cdot z^{(n)} \in \mathfrak{g}^{(n)}$. The Lie algebra structure on $\hat{\mathfrak{g}}$ is defined by

$$[x^{(n)}, y^{(m)}] := [x, y]^{(n+m)} + i\alpha \cdot n\delta_{n,-m}\langle x, y\rangle, \tag{1.2}$$

where $\alpha$ is a constant. If $G$ is a simply connected and connected real Lie group with the Lie algebra $\mathfrak{g}$, then $\hat{\mathfrak{g}}$ is the Lie algebra of a one-dimensional central extension $\hat{G}$ of the group of $G$-valued $C^\infty$-functions on the unit circle $S^1$ (with point-wise multiplication),

$$(f, a)(g, b) := (fg, abe^{i\omega(f,g)}), \tag{1.3}$$

with $(fg)(\phi) := f(\phi)g(\phi)$ and $\omega(f, g)$ is a two-cocycle. If we consider the group consisting of maps $D \to G$ instead of $S^1 \to G$, where $D$ is the unit disk with $S^1$ as the boundary, then $\omega$ can be defined by

$$\omega(f, g) := \alpha \int_D \langle f^{-1}\, df, g^{-1}\, dg\rangle. \tag{1.4}$$

If

$$f(z) = e^{tx(z)}, \quad g(z) = e^{sy(z)}, \quad z \in D, \tag{1.5}$$

where $x$ and $y$ are $\mathfrak{g}$-valued functions and $t, s$ real parameters, then

$$\frac{d^2}{dt\, ds}\omega(f, g)|_{t=s=0} = \alpha \int_D \langle dx, dy\rangle = \alpha \int_{S^1} \langle x, dy\rangle. \tag{1.6}$$

Inserting the Fourier components $x^{(n)}(\varphi) = x \exp in\varphi$, $y^{(m)}(\varphi) = y \exp im\varphi$ in (1.6), one gets the central term in (1.2).

There are two ways to construct representations of $\hat{\mathfrak{g}}$. First one can try to construct representations of $\hat{G}$ by global methods and then differentiate these to give representations of $\hat{\mathfrak{g}}$. Global methods have been used for example in Ref. 6. I shall follow a second route, namely I shall construct directly representations of $\hat{\mathfrak{g}}$ by algebraic methods. I am not going to solve the (very difficult) problem which of the representations of $\hat{\mathfrak{g}}$ correspond to unitary representations of $\hat{G}$.

Step algebra ( = algebra of raising and lowering operators) methods have been applied earlier to a wide variety of problems. In the beginning, step algebras were used for constructing irreducible finite-dimensional representations of classical Lie algebras in canonical chains of subalgebras.[7] It was later realized that step algebras could be constructed also for noncanonical chains.[8] In the next phase it was shown how to use these methods to classify irreducible infinite-dimensional $\mathfrak{k}$-finite representations of Lie algebras containing a semisimple subalgebra $\mathfrak{k}$ ($\mathfrak{k}$-finite means that the representation is a direct sum of finite-dimensional representations of $\mathfrak{k}$; irreducibility refers to the larger Lie algebra).[9] In this paper I want to show that step algebras are useful also in the representation theory of Kac–Moody algebras. The algebra $\hat{\mathfrak{g}}$ contains the subalgebra $\mathfrak{g}^{(0)}$ which can be identified with $\mathfrak{g}$. I shall discuss the $\mathfrak{g}$-finite representations of $\hat{\mathfrak{g}}$ for the semisimple Lie algebras $\mathfrak{g}$. The case $\mathfrak{g} = A_1 \cong \mathfrak{su}(2)$ is worked out in detail in Sec. II. Section III contains first some general results, valid for all semisimple $\mathfrak{g}$, which are then applied to the case $\mathfrak{g} = A_2$.

The representations which will be constructed are called the discrete series because of a certain parallelism with the theory of discrete series of semisimple Lie algebras; the term "discrete" has to be understood in a somewhat generalized sense since, in addition to a set of discrete parameters (dominant integral weight of $\mathfrak{g}$), the representations will be parametrized by a representation of an infinite-dimensional Heisenberg algebra. The same kind of Heisenberg algebras appear also in the theory of maximal weight representations of $\hat{\mathfrak{g}}$.[10,11]

Let $\mathfrak{h} \subset \mathfrak{g}$ be a Cartan subalgebra, $\Psi$ the system of roots for $(\mathfrak{g}, \mathfrak{h})$ and $\Delta = \{\alpha_1, \ldots, \alpha_l\} \subset \Psi$ a system of simple roots. Let $\mathfrak{h}^*$ be the dual of $\mathfrak{h}$ and $(\cdot, \cdot)$; $\mathfrak{h}^* \times \mathfrak{h}^* \to \mathbf{C}$ the dual of the Killing form of $\mathfrak{g}$ restricted to $\mathfrak{h}$. Put

$$\langle \alpha, \beta\rangle := 2((\alpha, \beta)/(\beta, \beta)). \tag{1.7}$$

Then the set of dominant integral weights is

 0022-2488/85/030377-06$02.50 © 1985 American Institute of Physics 

$$\Lambda^+ = \{\lambda \in \mathfrak{h}^* | \langle \lambda, \alpha \rangle \in \mathbf{Z}_+ \ \forall \ \alpha \in \Delta \}. \tag{1.8}$$

$\mathbf{Z}_+$ is the set of non-negative integers. If the $\hat{\mathfrak{g}}$-module $V$ is $\mathfrak{g}$-finite, then

$$V = \bigoplus_{\lambda \in \Lambda^+} V_\lambda, \tag{1.9}$$

where $V_\lambda$ is a direct sum of irreducible finite-dimensional $\mathfrak{g}$-modules with the highest weight $\lambda$. The $\mathfrak{g}$-submodule $V_\lambda \subset V$ is the minimal components of $V$, if $V_\mu = 0$ for $\mu < \lambda$ (here "$<$" refers to a certain lexicographical ordering on $\Lambda^+$) and $V_\lambda \neq 0$. My aim is to show that if the minimal $\lambda$-value is not "too small" then all the irreducible $\mathfrak{g}$-finite $\hat{\mathfrak{g}}$-modules with the minimal component $V_\lambda$ can be parametrized by representations of a certain infinite-dimensional Heisenberg algebra which is acting on the subspace $V_\lambda^+ \subset V_\lambda$ of highest weight vectors. This claim will be proven for the cases $\mathfrak{g} = A_1, A_2$ but, using the results of Sec. III, it should be rather straightforward work to prove similar theorems for other semisimple algebras $\mathfrak{g}$.

For different approaches to the representation theory of $\hat{\mathfrak{g}}$ see for example Refs. 11. The results in Ref. 11 seem to have little in common with the present paper. Except for the highest weight modules, people have earlier been interested mainly in modules which correspond to unitary representations of groups. Here the representations are (for the most part) nonunitary.

## II. THE CASE $\mathfrak{g} = \mathfrak{su}(2)$

Let $\{x, h, y\}$ be a basis of the (complex) Lie algebra $\mathfrak{g} = A_1 \cong \mathfrak{su}(2)$ such that

$$[x, y] = 2h, \quad [h, x] = x, \quad [h, y] = -y. \tag{2.1}$$

The element $h$ spans the Cartan subalgebra $\mathfrak{h} \subset \mathfrak{g}$. A basis of $\hat{\mathfrak{g}}$ is given by the elements $x^{(n)}, h^{(n)}, y^{(n)}$ $(n \in \mathbf{Z})$ with the commutators

$$\begin{aligned} &[x^{(n)}, y^{(m)}] = 2h^{(m+n)} + 4\alpha n \delta_{n,-m}, \\ &[h^{(n)}, x^{(m)}] = x^{(n+m)}, \quad [h^{(n)}, y^{(m)}] = -y^{(n+m)}, \\ &[h^{(n)}, h^{(m)}] = 2\alpha n \delta_{n,-m}. \end{aligned} \tag{2.2}$$

For brevity, let us write $x^{(0)} = x$, $y^{(0)} = y$, and $h^{(0)} = h$. Define the following set of elements in the enveloping algebra $U(\hat{\mathfrak{g}})$ of $\hat{\mathfrak{g}}$:

$$\begin{aligned} s_+^{(n)} &= x^{(n)}, \quad s_0^{(n)} = x^{(n)} y + 2h^{(n)} h, \\ s_-^{(n)} &= x^{(n)} y^2 + h^{(n)} y(4h - 2) - y^{(n)} h(4h - 2), \end{aligned} \tag{2.3}$$

where $n = \pm 1, \pm 2, \ldots$. All these elements $s$ have the property

$$xs \equiv 0 \bmod U(\hat{\mathfrak{g}})x. \tag{2.4}$$

In addition,

$$[h, s_+^{(n)}] = s_+^{(n)}, \quad [h, s_0^{(n)}] = 0, \quad [h, s_-^{(n)}] = -s_-^{(n)}. \tag{2.5}$$

Let $V$ be any $\hat{\mathfrak{g}}$-module and $v \in V$ such that $xv = 0$. From (2.4) follows that

$$xp(s_\pm^{(n)}, s_0^{(n)})v = 0 \tag{2.6}$$

for any polynomial $p$ of the elements (2.3). Let $S_0 = S_0(\hat{\mathfrak{g}}, \mathfrak{g})$ be the algebra generated by (2.3) and $h$,

$$S_0^{\mathfrak{h}} := \{s \in S_0 | hs = sh\} \subset S_0. \tag{2.7}$$

The set of dominant integral weights $\Lambda^+ \subset \mathfrak{h}^*$ is characterized by $\lambda(h) \in \frac{1}{2}\mathbf{Z}$, $\lambda(h) \geqslant 0$ (by agreeing that $x$ corresponds to the positive root). For brevity, let us write $\lambda$ for the value $\lambda(h)$. If $V$ is $\mathfrak{g}$-finite, then

$$V = \bigoplus_{\lambda = 0, \frac{1}{2}, 1, \ldots} V_\lambda. \tag{2.8}$$

Set

$$V_\lambda^+ = \{v \in V^\lambda | xv = 0\}, \quad V^+ = \{v \in V | xv = 0\}. \tag{2.9}$$

Clearly $S_0 V^+ \subset V^+$ and $S_0^{\mathfrak{h}} V_\lambda^+ \subset V_\lambda^+$. The annihilator in $U(\mathfrak{g})$ of a vector $0 \neq v \in V_\lambda^+$ is denoted by $J_\lambda$,

$$J_\lambda = U(\mathfrak{g})x + U(\mathfrak{g}) \cdot (h - \lambda) + U(\mathfrak{g}) \cdot y^{2\lambda + 1}. \tag{2.10}$$

Let

$$D_\lambda = S_0^{\mathfrak{h}} / S_0^{\mathfrak{h}} \cap U(\hat{\mathfrak{g}}) J_\lambda. \tag{2.11}$$

Each $V_\lambda^+ \subset V$ has a natural $D_\lambda$-module structure. In fact, from Ref. 12, Proposition 1.3, it follows that the map $V \mapsto V_\lambda^+$ gives a 1–1 correspondence between (equivalence classes of) irreducible $\mathfrak{g}$-finite $\hat{\mathfrak{g}}$-modules and irreducible $D_\lambda$-modules. In Ref. 12 the Lie algebras were assumed to be finite dimensional, but the proof does not really depend on this assumption; the only thing which matters is that the adjoint action of the semisimple subalgebra $\mathfrak{g} \subset \hat{\mathfrak{g}}$ on $\hat{\mathfrak{g}}$ splits into a direct sum of finite-dimensional representations of $\mathfrak{g}$. (In our case here, each summand is equivalent to the vector representation $\lambda = 1$.)

We call $V_\lambda$ the minimal component of $V$ if $V_\lambda \neq 0$ and $V_\mu = 0$ for $\mu < \lambda$. It is clear that $s_-^{(n)} V_\lambda^+ = 0$ for all $n$. Let $T \subset S_0$ be the subspace spanned by the elements $s_-^{(n)}$, and let

$$D_\lambda^{\mathfrak{h}} = S_0^{\mathfrak{h}} / S_0^{\mathfrak{h}} \cap (U(\hat{\mathfrak{g}}) J_\lambda + S_0 T). \tag{2.12}$$

Since $S_0 T$ annihilates $V_\lambda^+$, the subspace $V_\lambda^+ \subset V$ carries in a natural way a representation of the algebra $D_\lambda^{\mathfrak{h}}$. According to Ref. 12, Theorem 1.4, the mapping $V \mapsto V_\lambda^+$ gives a 1–1 correspondence between irreducible $\mathfrak{g}$-finite representations of $\hat{\mathfrak{g}}$, with the minimal component $V_\lambda$ and the irreducible nonzero representations of $D_\lambda^{\mathfrak{h}}$. All we have to do is to compute $D_\lambda^{\mathfrak{h}}$.

*Lemma 2.1:* The following equations hold mod $U(\hat{\mathfrak{g}})x$.

$$\begin{aligned} &s_-^{(n)} s_+^{(m)} h^2 (h - \tfrac{1}{2}) \\ &\quad = -\tfrac{1}{2}(h+1) s_+^{(n)} s_-^{(m)} + h(h+1)(h+\tfrac{1}{2}) s_+^{(m)} s_-^{(n)} \\ &\qquad - 2(h - \tfrac{1}{2})(h + \tfrac{1}{2}) s_0^{(n)} s_0^{(m)} + 4(1 - \delta_{n,-m}) \\ &\qquad \times (h+1)(h - \tfrac{1}{2})(h + \tfrac{1}{2}) h s_0^{(n+m)} \\ &\qquad + 8h^2 (h - \tfrac{1}{2})(h + \tfrac{1}{2})(h+1)(h + 2\alpha m + 1)\delta_{n,-m}, \end{aligned} \tag{2.13a}$$

$$\begin{aligned} s_-^{(n)} s_0^{(m)} (h - 1) &= s_0^{(m)} s_-^{(n)} h - s_0^{(n)} s_-^{(m)} + 2s_-^{(n+m)} \\ &\quad \times h(h-1)(1 - \delta_{n,-m}), \end{aligned} \tag{2.13b}$$

$$\begin{aligned} &(s_0^{(n)} s_0^{(m)} - s_0^{(m)} s_0^{(n)} - 16\alpha n \delta_{n,-m} h^2)(h - \tfrac{1}{2}) \\ &\quad = -\tfrac{1}{2} s_+^{(n)} s_-^{(m)} + \tfrac{1}{2} s_+^{(m)} s_-^{(n)}. \end{aligned} \tag{2.14}$$

*Proof:* By a direct computation. The computations can be greatly simplified using the following result. Let $U_1 \subset U(\hat{\mathfrak{g}})$ be the subspace spanned by the monomials

$$x^{(n_1)}\cdots x^{(n_p)}h^{(m_1)}\cdots h^{(m_q)}y^{(l_1)}\cdots y^{(l_r)}h^{s} \quad (n_i, m_i, l_i \neq 0), \tag{2.15}$$

with $n_1 \leqslant n_2 \leqslant \cdots \leqslant n_p$ and a similar ordering for the $m_i$'s and $l_i$'s. Then

$$U(\hat{\mathfrak{g}}) = U_1 \oplus U(\hat{\mathfrak{g}})x \oplus U_1 U(y)y. \tag{2.16}$$

Let $Q$ be the projection on the first summand. One can show (Ref. 13, Theorem 1, Ref. 14, pp. 46 and 47) that if $s_1, s_2 \in S_0$ then

$$s_1 \equiv s_2 \bmod x \Leftrightarrow Q(s_1) = Q(s_2). \tag{2.17}$$

**Theorem 2.2:** If $\lambda = 1, \frac{3}{2}, 2, \ldots$ then $D_\lambda^b$ is the infinite-dimensional Heisenberg algebra generated by $s_0^{(n)} (n = \pm 1, \pm 2, \ldots)$,

$$[s_0^{(n)}, s_0^{(m)}] = 16\alpha n \delta_{n,-m} \lambda^2. \tag{2.18}$$

*Proof:* Using the commutation relations (2.13a) and (2.13b), in any product containing $s_\pm$'s and $s_0$'s the elements $s_-$ can be commuted step by step to the right modulo $J_\lambda$; on the other hand elements in $S_0 T$ represent the class zero in the factor algebra $D_\lambda^b$. The commutation relations of the elements $s_0^{(n)}$ (more precisely, the corresponding classes in $D_\lambda^b$) follow now from (2.14). Thus $s_0^{(n)}, s_0^{(-n)}$ satisfy for each $0 < n$ the commutation relations of a Heisenberg algebra and these algebras commute for different values of $n \in \mathbf{N}$. Using the map $Q$ it is easy to show that $0 \neq \mathbf{1} \bmod (J_\lambda + T)$. From the standard properties of a Heisenberg algebra follows that there cannot be any other polynomial relations in $D_\lambda^b$, in addition to those obtained from the commutation relations (2.18), without the algebra being zero. □

By the theorem above, all the irreducible $\mathfrak{g}$-finite $\hat{\mathfrak{g}}$-modules with the minimal component $V_\lambda, \lambda = 1, \frac{3}{2}, 2, \ldots$, are parametrized by irreducible representations of a Heisenberg algebra. If $\lambda = 0, \frac{1}{2}$, the structure of $D_\lambda^b$ is much more complicated and I have no results about representations of $\hat{\mathfrak{g}}$ with a minimal component, $V_0$ or $V_{1/2}$.

If the $\hat{\mathfrak{g}}$-module $V$ corresponds to a unitary representation of the group of SU(2)-valued functions on the unit circle, then there should be an inner product in $V$ such that the following Hermiticity conditions hold:

$$x^{(n)\dagger} = y^{(-n)}, \quad y^{(n)\dagger} = x^{(-n)}, \quad h^{(n)\dagger} = h^{(-n)}. \tag{2.19}$$

From (2.19) it follows that

$$(v_1, s_+^{(n)\dagger} h(4h+2)v_2) = (v_1, -s_-^{(-n)}v_2), \tag{2.20a}$$

$$(v_1, s_-^{(n)\dagger} v_2) = (v_1, -h(4h+2)s_+^{(-n)}v_2), \tag{2.20b}$$

$$(v_1, s_0^{(n)\dagger} v_2) = (v_1, s_0^{(-n)}v_2), \tag{2.20c}$$

for $v_1, v_2 \in V^+$.

In particular, the minimal component $V_\lambda^+ (\lambda \geqslant 1)$ of $V$ carries a representation of the algebra $D_\lambda^b$ satisfying the Hermiticity conditions (2.20c). Suppose for example that $\alpha > 0$. then $s_0^{(n)}$, $0 < n$, plays the role of an annihilation operator and $s_0^{(-n)}$ is the corresponding creation operator. A reasonable first guess would be that $V$ is a Hermitian $\hat{\mathfrak{g}}$-module if $V_\lambda^+$ is a Hermitian $D_\lambda^b$-module [in the sense of (2.20c)]. However, things are not so simple as the following example shows. Take the Fock representation of $D_\lambda^b$, which is characterized by

$$V_\lambda^+ = D_\lambda^b v_0, \quad s_0^{(n)} v_0 = 0 \quad \text{for } n > 0, \tag{2.21}$$

and by $s_0^{(n)\dagger} = s_0^{(-n)}$. Now the square of the norm of the vector $s_+^{(n)} v_0$ is

$$\begin{aligned}(s_+^{(n)} v_0, s_+^{(n)} v_0) &= (v_0, s_+^{(n)\dagger} s_+^{(n)} v_0) \\ &= (v_0, s_-^{(-n)} s_+^{(n)} v_0) \cdot \frac{-1}{(\lambda+1)(4\lambda+6)} \\ &= 2\frac{\lambda + \frac{1}{2}}{\lambda^2(\lambda+1)(4\lambda+6)} (v_0, s_0^{(-n)} s_0^{(n)} v_0) \\ &\quad - 8 \cdot \frac{(\lambda+\frac{1}{2})(\lambda + 2\alpha n + 1)}{4\lambda+6} \\ &= -8 \cdot \frac{(\lambda+\frac{1}{2})(\lambda+2\alpha n+1)}{4\lambda+6}, \text{ if } n > 0,\end{aligned} \tag{2.22}$$

where the Eq. (2.13a) has been used. Therefore, $\|s_+^{(n)} v_0\|^2 < 0$ for $n > 0$.

## III. DISCRETE SERIES OF $\hat{\mathfrak{g}}$ FOR SEMISIMPLE $\mathfrak{g}$

Let $\mathfrak{g}$ be a semisimple Lie algebra. I shall use the notation of Sec. I. Let us define a lexicographical ordering in the set of weights

$$\Lambda = \{\lambda \in \mathfrak{h}^* | \langle \lambda, \alpha_i \rangle \in \mathbf{Z}, 1 \leqslant i \leqslant l\}.$$

When $\lambda, \lambda' \in \Lambda$ then $\lambda > \lambda'$ if $\lambda \neq \lambda'$ and the first nonzero integer in the set $\langle \lambda - \lambda', \alpha_1 \rangle, \langle \lambda - \lambda', \alpha_2 \rangle, \ldots, \langle \lambda - \lambda', \alpha_l \rangle$ is positive. Let $\{t_1^{(n)}, \ldots, t_N^{(n)}\}$ be a basis of $\mathfrak{g}^{(n)} \subset \hat{\mathfrak{g}}$ such that each $t_i^{(n)}$ is a weight vector,

$$[h, t_i^{(n)}] = \mu_i(h) t_i^{(n)}, \quad h \in \mathfrak{h}. \tag{3.1}$$

It is clear that each $\mu_i$ is a root for $(\mathfrak{g}, \mathfrak{h})$, $\mu_i \in \Psi$. Fix the indexing so that

$$\mu_1 \geqslant \mu_2 \geqslant \cdots \geqslant \mu_N. \tag{3.2}$$

Of course, only the root $\mu_i = 0$ can have a multiplicity bigger than one. The set $\{t_i^{(n)} | 0 \neq n \in \mathbf{Z}, 1 \leqslant i \leqslant N\}$ is a basis of the ad $\mathfrak{g}$-invariant complement $\mathfrak{p}$ of $\mathfrak{g}$ in $\hat{\mathfrak{g}}$. Define an ordering in this basis by $t_i^{(n)} < t_j^{(m)}$ if $i < j$ or $i = j$ and $n < m$. Let $U_1 \subset U(\hat{\mathfrak{g}})$ be the subspace spanned by the ordered monomials $t_{i_1}^{(n_1)} \cdots t_{i_k}^{(n_k)}$ multiplied by arbitrary elements of $U(\mathfrak{h})$. One can write

$$\mathfrak{g} = \mathfrak{g}_+ \oplus \mathfrak{h} \oplus \mathfrak{g}_-, \tag{3.3}$$

where $\mathfrak{g}_+$ (respectively, $\mathfrak{g}_-$) corresponds to the positive (respectively, negative) roots; here "positive" refers to the standard ordering determined by the set of simple roots $\Delta$. Now

$$U(\hat{\mathfrak{g}}) = U_1 \oplus U(\hat{\mathfrak{g}})\mathfrak{g}_+ \oplus U_1 U(\mathfrak{g}_-)\mathfrak{g}_-. \tag{3.4}$$

Let $Q'$ be the projection on the first summand and let $Q: U(\hat{\mathfrak{g}})/U(\hat{\mathfrak{g}})\mathfrak{g}_+ \to U_1$,

$$Q(u + U(\hat{\mathfrak{g}})\mathfrak{g}_+) := Q'(u). \tag{3.5}$$

Define

$$S'(\hat{\mathfrak{g}}, \mathfrak{g}) = \{u \in U(\hat{\mathfrak{g}}) | \mathfrak{g}_+ u \subset U(\hat{\mathfrak{g}})\mathfrak{g}_+\} \tag{3.6}$$

and the step algebra

$$S(\hat{\mathfrak{g}}, \mathfrak{g}) = S'(\hat{\mathfrak{g}}, \mathfrak{g})/U(\hat{\mathfrak{g}})\mathfrak{g}_+. \tag{3.7}$$

It is known that the mapping

$$Q: S(\hat{\mathfrak{g}}, \mathfrak{g}) \to U_1 \tag{3.8}$$

is injective (Ref. 13, Theorem 1; Ref. 14, pp. 46 and 47). Let $S_0(\hat{\mathfrak{g}},\mathfrak{g})$ be the subalgebra of $S(\hat{\mathfrak{g}},\mathfrak{g})$ generated by the elements $s$ such that $Q(s)\in \mathfrak{p}U(\mathfrak{h})$. It is known that if $Q(s)=t_i^{(n)}p(h)$, where $p(h)\in U(h)$, then

$$s=t_i^{(n)}p(h)+\sum_j t_j^{(n)}q_j,$$

where $q_j\in U(\mathfrak{g}_-+\mathfrak{h})$ and $q_j\neq 0$ only if $\mu_j-\mu_i$ is a positive root (Ref. 14, Proposition 1). For $\lambda\in\Lambda^+$ I denote by $J_\lambda \subset U(\mathfrak{g})$ the annihilator of the vector of highest weight in a finite-dimensional $\mathfrak{g}$-module with highest weight $\lambda$. Let $S_0^\downarrow$ be the subalgebra of $S_0=S_0(\hat{\mathfrak{g}},\mathfrak{g})$ generated by the elements $s_1,s_2$, where $s_1,s_2\in S_0$ and $s_2$ is a weight vector (with respect to the adjoint action of $\mathfrak{h}$) with a negative weight in the lexicographical ordering. Finally,

$$D_\lambda^\flat := S_0^\flat/S_0^\flat\cap(U(\hat{\mathfrak{g}})J_\lambda+S_0^\downarrow), \tag{3.9}$$

where $S_0^\flat=\{s\in S_0|[h,s]=0\ \ \forall h\in\mathfrak{h}\}$. If $V$ is a $\mathfrak{g}$-finite $\hat{\mathfrak{g}}$-module, then $V_\lambda\subset V$ is the minimal component if $0\neq V_\lambda$ is minimal with respect to the lexicographical ordering. As in the case $\mathfrak{g}=A_1$, the $\mathfrak{g}$-finite irreducible representations of $\hat{\mathfrak{g}}$ with a minimal component $V_\lambda$ will be described by the action of $D_\lambda^\flat$ on $V_\lambda^+$ (Ref. 12, Theorem 1.4).

Let $\mathscr{I}_\lambda=U(\hat{\mathfrak{g}})\cdot\{h-\lambda(h)|h\in\mathfrak{h}\}$ $(\lambda\in\mathfrak{h}^*)$ and let $\pi_\lambda: U(\hat{\mathfrak{g}})\to U(\hat{\mathfrak{g}})/\mathscr{I}_\lambda$ be the canonical projection, $Q_\lambda:=\pi_\lambda\circ Q$.

*Lemma 3.1:* Let $s\in S(\hat{\mathfrak{g}},\mathfrak{g})$. Then $Q_\lambda(s)=0$ iff $s\in U(\hat{\mathfrak{g}})J_\lambda/U(\hat{\mathfrak{g}})\mathfrak{g}_+$.

*Proof:* (Ref. 15, Lemma 4.4 or Ref. 16, Proposition II. 2.12). □

Let $\nu\in\Lambda^+$ and let $m(A\otimes B;\nu)$ denote the multiplicity of the irreducible representation with highest weight $\nu$, in the tensor product of representations $A$ and $B$ of $\mathfrak{g}$. If $T\subset U(\hat{\mathfrak{g}})$ is an ad $\mathfrak{h}$-invariant subspace, denote by $T(\lambda)$ the weight space corresponding to the weight $\lambda\in\mathfrak{h}^*$.

*Lemma 3.2:* Let $T\subset U_1$ be an ad $\mathfrak{g}$-invariant finite-dimensional subspace, $\lambda\in\Lambda^+$ and $\nu\in\Lambda$. Then the dimension $n(\lambda,\nu)$ of the subspace $Q_\lambda(\{s\in S(\hat{\mathfrak{g}},\mathfrak{g})|Q_\lambda(s)\in T(\nu)\})\subset T(\nu)$ is equal to $m(T\otimes\lambda,\lambda+\nu)$.

*Proof:* Let $v^+:=1+U(\hat{\mathfrak{g}})J_\lambda\in U(\hat{\mathfrak{g}})/U(\hat{\mathfrak{g}})J_\lambda$. Then $U(\hat{\mathfrak{g}})/U(\hat{\mathfrak{g}})J_\lambda$ is a $\mathfrak{g}$-finite (left) $\hat{\mathfrak{g}}$-module and $U(\mathfrak{g})v^+$ is in fact the $\mathfrak{g}$-module $\lambda$. Let $W:=T\cdot v^+$; this is a finite-dimensional $\mathfrak{g}$-submodule of $U(\hat{\mathfrak{g}})/U(\hat{\mathfrak{g}})J_\lambda$. By Ref. 16, Theorem II. 2.20 there exists $s_1,\ldots,s_p\in S(\hat{\mathfrak{g}},\mathfrak{g})$ such that $\{s_1v^+,\ldots,s_pv^+\}$ is a basis of $W_{\lambda+\nu}^+$. By Lemma 3.1, $\dim W_{\lambda+\nu}^+=n(\lambda,\nu)$. On the other hand, clearly $\dim W_{\lambda+\nu}^+=m(T\otimes\lambda;\lambda+\nu)$. □

Let $S_\nu^s(n,m)\subset S_0(\hat{\mathfrak{g}},\mathfrak{g})$ be the subspace spanned by vectors of the type $s_i^{(n)}s_j^{(m)}+s_i^{(m)}s_j^{(n)}$, where $Q(s_i^{(k)})\in t_i^{(k)}U(\mathfrak{h})(k=n,m)$ with $\mu_i\geqslant\mu_j$ and $\mu_i+\mu_j=\nu$. Let $S_\nu^a(n,m)\subset S_0$ be the space spanned by the antisymmetric combinations $s_i^{(n)}s_j^{(m)}-s_i^{(m)}s_j^{(n)}$. Denote by $R_-$ (respectively, $R_+$) the set of negative (respectively, positive) roots with respect to the lexicographical ordering; set $\overline{R}_\pm=R_\pm\cup\{0\}$. Let $\omega$ be the adjoint representation of $\mathfrak{g}$. Remember that $\mathfrak{g}\cong\mathfrak{g}^{(n)}$ for each $n$, as ad $\mathfrak{g}$-modules. Denote by $\omega_s^2$ the symmetric part of the tensor product $\omega\otimes\omega$ and let $\omega_a^2$ be the antisymmetric part.

**Theorem 3:3:** Let $\Lambda_0^+\subset\Lambda$ be a subset such that

(i) $\dim Q_\lambda(S_\nu^s(n,m))\geqslant m(\omega_s^2\otimes\lambda;\lambda+\nu)$,

and if $n\neq m$,

$\dim Q_\lambda(S_\nu^a(n,m))\geqslant m(\omega_a^2\otimes\lambda;\lambda+\nu)$,

$\forall\nu\in R_-+\overline{R}_+$,

(ii) if $\lambda\in\Lambda_0^+$, $\mu_i\in\overline{R}_+$ and $\lambda+\mu_i\in\Lambda^+$,

then $\lambda+\mu_i\in\Lambda_0^+$.

Then $D_\lambda^\flat$ is generated by the weight zero elements $s\in S_0$, $Q_\lambda(s)\in\mathfrak{p}(0)$.

*Proof:* Consider a typical monomial $s=s_1s_2\cdots s_f$ in $S_0(\hat{\mathfrak{g}},\mathfrak{g})$, $Q(s_k)\in t_{i_k}^{(n_k)}U(\mathfrak{h})$. I want to show by an induction on the length $f$ that $s$ is equal, mod $(U(\hat{\mathfrak{g}})J_\lambda+S_0^\downarrow)$, to a polynomial containing only elements $s_k$ of weight $\mu\in\overline{R}_+$. If in particular $s\in S_0^\flat$, then each $s_k$ has to be of weight zero. Clearly the assertion holds if $f=1$. Consider now the induction from $f-1$ to $f$. By the induction assumption we can assume that $\mu_{i_k}\geqslant 0$ for $k=2,3,\ldots,f$. Suppose that $\mu_{i_1}<0$. Denote

$$\lambda'=\lambda+\mu_{i_2}+\mu_{i_3}+\cdots+\mu_{i_f},\quad \nu=\mu_{i_1}+\mu_{i_2},$$

$$v=s_3s_4\cdots s_f\cdot(1+U(\hat{\mathfrak{g}})J_\lambda)\in U(\hat{\mathfrak{g}})/U(\hat{\mathfrak{g}})J_\lambda.$$

Using the $\mathfrak{g}$-module isomorphism (provided by the symmetrization mapping) between the symmetric algebra $A(\hat{\mathfrak{g}})$ on $\hat{\mathfrak{g}}$ and the enveloping algebra $U(\hat{\mathfrak{g}})$, the vector $s_1s_2v$ can be considered as a vector in the $\mathfrak{g}$-module $\mathfrak{g}^{(n_1)}\otimes\mathfrak{g}^{(n_2)}\otimes\lambda'$, modulo lower-order terms in $\mathfrak{p}$ arising from the commutators $[\mathfrak{g}^{(n_1)},\mathfrak{g}^{(n_2)}]$. Dividing this vector into a symmetric and an antisymmetric part in the upper indices, it follows that $s_1s_2v=w_av+w_sv$, mod lower-order terms, where $w_s\in S_\nu^s(n_1,n_2)$ and $w_a\in S_\nu^a(n_1,n_2)$ by (i); $\lambda'\in\Lambda_0^+$ by the assumption (ii). Now both $w_s$ and $w_a$ are linear combinations of terms of the type $s_1's_2'$ where $s_1'$, $s_2'$ are of first order in $\mathfrak{p}$, of weight $\gamma_1,\gamma_2$ (in this order) with $\gamma_1\geqslant\gamma_2$, $\gamma_1+\gamma_2=\nu$. Again by the induction assumption $s_2's_3s_4\cdots s_f$ can be written as a polynomial mod $(U(\hat{\mathfrak{g}})J_\lambda+S_0^\downarrow)$ of degree $\leqslant f-1$, containing only elements of non-negative weight. Since $\gamma_1>\mu_{i_1}$, we can show by a second induction on the weight of the first factor that $s$ is equal to polynomial, mod $(U(\hat{\mathfrak{g}})J_\lambda+S_0^\downarrow)$, where all the elements (including the first) are of non-negative weight. □

As an example I shall show how the general results above can be applied to the case $\mathfrak{g}=A_2$. Let $\{\alpha_1,\alpha_2\}$ be a set of simple roots for $\mathfrak{g}$. Then the set of nonzero roots is $\psi=\{\alpha_1,\alpha_2,\alpha_1+\alpha_2,-\alpha_1,-\alpha_2,-\alpha_1-\alpha_2\}$. In the lexicographical ordering "$<$" with respect to the coordinates $\langle\lambda,\alpha_1\rangle$, $\langle\lambda,\alpha_2\rangle$ of a weight $\lambda\in\Lambda$ we have

$$R_+=\{\alpha_1,\alpha_1+\alpha_2,-\alpha_2\},$$

$$R_-=\{-\alpha_1,-\alpha_1-\alpha_2,\alpha_2\}. \tag{3.10}$$

Let

$$\Lambda_0^+:=\{\lambda\in\Lambda^+|\langle\lambda,\alpha_1\rangle\geqslant 2\}.$$

Clearly $\Lambda_0^+$ satisfies the condition (ii) in Theorem 3.4. The adjoint representation is now $\omega=(\omega_1,\omega_2)=(1,1)$. Using the standard Young tableaux rules for tensor products one gets

$$\omega_s^2=(2,2)\oplus(1,1)\oplus(0,0),\quad \omega_a^2=(3,0)\oplus(0,3)\oplus(1,1). \tag{3.11}$$

If $\lambda_1\geqslant 2$,

$$\lambda\otimes\omega=(\lambda_1+1,\lambda_2+1)\oplus(\lambda_1-1,\lambda_2+2)$$
$$\oplus(\lambda_1-2,\lambda_2+1)\oplus(2-\delta_{0\lambda_2})\cdot\lambda$$
$$\oplus(1-\delta_{0\lambda_2})\cdot(\lambda_1+2,\lambda_2-1)\oplus(1-\delta_{0\lambda_2})$$
$$\cdot(\lambda_1-1,\lambda_2-1)$$
$$\oplus(1-\delta_{0\lambda_2}\delta_{1\lambda_2})\cdot(\lambda_1+1,\lambda_2-1). \tag{3.12}$$

All the multiplicities are $\leqslant 1$ except the multiplicity of $\lambda$, which is 1 or 2. By a straightforward computation of the relevant multiplicities one can show that the condition (i) of Theorem 3.3 is satisifed. As an example, let us consider the case $\nu=0$.

**A. $\lambda_2\geqslant 1$**

In this case $m(\omega_s^2\otimes\lambda;\lambda)=6$ and $m(\omega_a^2\otimes\lambda;\lambda)=4$. Either using the explicit formulas in the Appendix or Lemma 3.2 and the following multiplicities

$$m(\omega\otimes\lambda;\lambda-\alpha_1)=m(\omega\otimes\lambda;\lambda+\alpha_2)$$
$$=m(\omega\otimes\lambda;\lambda-\alpha_1-\alpha_2)=1, \tag{3.13}$$
$$m(\omega\otimes\lambda;\lambda)=2,$$
$$m(\omega\otimes\lambda;\lambda+\mu)=m(\omega\otimes(\lambda+\mu);\lambda),$$

it is seen that for each $0\neq n\in\mathbf{N}$ there exists elements $s_1^{(n)}$, $s_2^{(n)}$ and for each root $\mu\neq 0$ elements $s^{(n)}(\mu)$ in $S_0$ such that $0\neq Q_\lambda(s_i^{(n)})\in\mathfrak{g}^{(n)}(0)$ $(i=1,2)$, $0\neq Q_\lambda(s^{(n)}(\mu))\in\mathfrak{g}^{(n)}(\mu)$ and such that the following symmetrized and (when $n\neq m$) antisymmetrized combinations are linearly independent:

$$s^{(n)}(\mu)s^{(m)}(-\mu)+s^{(m)}(\mu)s^{(n)}(-\mu),\quad \mu\in R_+,$$
$$s_i^{(n)}s_j^{(m)}+s_i^{(m)}s_j^{(n)},\quad i\geqslant j,$$
$$s^{(n)}(\mu)s^{(m)}(-\mu)-s^{(m)}(\mu)s^{(n)}(-\mu),\quad \mu\in R_+,$$
$$s_1^{(n)}s_2^{(m)}-s_1^{(m)}s_2^{(n)}. \tag{3.14}$$

There are six symmetric and four antisymmetric combinations and so the condition (i) is satisfied.

**B. $\lambda_2=0$**

Now $m(\omega_s^2\otimes\lambda;\lambda)=3$ and $m(\omega_a^2\otimes\lambda;\lambda)=2$. Using the fact that $m(\omega\otimes\lambda;\lambda-\alpha_1)=m(\omega\otimes\lambda;\lambda+\alpha_2)=m(\omega\otimes\lambda;\lambda)$ and $m(\omega\otimes\lambda;\lambda+\mu)=m(\omega\otimes(\lambda+\mu);\lambda)$ one can see that the following combinations are linearly independent, mod $U(\hat{\mathfrak{g}})J_\lambda$:

$$s^{(n)}(\mu)s^{(m)}(-\mu)+s^{(m)}(\mu)s^{(n)}(-\mu),\quad \mu=\alpha_1 \text{ or } -\alpha_2,$$
$$s_1^{(n)}s_1^{(m)}+s_1^{(m)}s_1^{(n)}, \tag{3.15}$$
$$s^{(n)}(\mu)s^{(m)}(-\mu)-s^{(m)}(\mu)s^{(n)}(-\mu),$$
$$\mu=\alpha_1 \text{ or } -\alpha_2,\quad n\neq m.$$

Thus there are three independent symmetric elements and two (when $m\neq n$) antisymmetric elements.

The other cases ($\nu\neq 0$) are handled in a similar way. Thus,

**Theorem 3.4:** Let $\mathfrak{g}=A_2$ and $\Lambda_0^+=\{\lambda\in\Lambda^+\mid\langle\lambda,\alpha_1\rangle\geqslant 2\}$. Then for all $\lambda\in\Lambda_0^+$ the algebra $D_\lambda^b$ is generated by the elements $s\in S_0$ with $Q_\lambda(s)\in\mathfrak{p}(0)$.

*Remark*: If $\lambda_2\geqslant 1$, there are two independent sets of generators, namely $\{s_1^{(n)}\mid 0\neq n\in\mathbf{Z}\}$ and $\{s_2^{(n)}\mid 0\neq n\in\mathbf{Z}\}$ as given in the Appendix; if $\lambda_2=0$, $\{s_1^{(n)}\}$ is the set of generators.

**Theorem 3.5:** Let $\mathfrak{g}=A_2$ and $\lambda\in\Lambda_0^+$. Then the algebra $D_\lambda^b$ is an infinite-dimensional Heisenberg algebra with the commutation relations

$$[s_1^{(n)},s_2^{(m)}]=0,$$
$$[s_1^{(n)},s_1^{(m)}]=16n\alpha\delta_{n,-m}\lambda_1^2(\lambda_1+\lambda_2+1)^2, \tag{3.16}$$
$$[s_2^{(n)},s_2^{(m)}]=16n\alpha\delta_{n,-m}\lambda_2^2(\lambda_1+\lambda_2+1)^2,$$

where $0\neq n,m\in\mathbf{Z}$.

*Proof*: By a simple computation,

$$Q_\lambda([s_1^{(n)},s_1^{(m)}])=(e_{12}^{(n)}e_{21}^{(m)}-e_{12}^{(m)}e_{21}^{(n)})$$
$$\times\lambda_1(\lambda_1+\lambda_2+1)^2+9(e_{13}^{(n)}e_{31}^{(m)}-e_{13}^{(m)}e_{31}^{(n)})$$
$$\times\lambda_1(\lambda_1+1)(\lambda_1+\lambda_2+1)$$
$$+16n\alpha\delta_{n,-m}\lambda_1^2(\lambda_1+\lambda_2+1)^2. \tag{3.17}$$

Looking at the formulas for the generators in the Appendix one can see that the sum of the two first terms is a linear combination of $Q_\lambda(s^{(n)}(\alpha_1)s^{(m)}(-\alpha_1)-s^{(m)}(\alpha_1)s^{(n)}(-\alpha_1))$and $Q_\lambda(s^{(n)}(\alpha_1+\alpha_2)s^{(m)}(-\alpha_1-\alpha_2)-s^{(m)}(\alpha_1+\alpha_2)\times s^{(n)}(-\alpha_1-\alpha_2))$. Since $s^{(n)}(-\alpha_1)$, $s^{(n)}(-\alpha_1-\alpha_2)\in S_0^\perp$, the commutator in the algebra $D_\lambda^b$ is equal to the last term in (3.17). Using the injectivity of $Q_\lambda$ and the Poincaré–Birkhoff–Witt theorem one can show easily that there cannot be any other polynomial identities than those obtained from the commutation relations in Theorem 3.5. □

Combined with Ref. 12, Theorem 1.4, Theorem 3.5 gives a complete characterization of all irreducible $\mathfrak{g}$-finite $\hat{\mathfrak{g}}$-modules with a minimal component $V_\lambda$, $\lambda\in\Lambda_0^+$.

## APPENDIX COMPUTATION OF $S_0(\hat{g},g)$ FOR $g=A_2$

*Here we give a complete set of elements $s\in S_0(\hat{\mathfrak{g}},\mathfrak{g})$* of first order in $\mathfrak{p}$ for the case $\mathfrak{g}=A_2$. These generate, together with $U(\mathfrak{h})$, the algebra $S_0$. The generators have been obtained by a trial and error method. They could be computed also using the general formula (Lemma 8) in Ref. 17. Let us denote by $e_{ij}$ the generator of $\mathfrak{g}$ which is represented in the fundamental $3\times 3$-matrix respresentation by the matrix

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

and set $h_1=e_{11}-e_{22}$, $h_2=e_{22}-e_{33}$. Then

$$s^{(n)}(\alpha_1+\alpha_2)=e_{13}^{(n)},\quad s^{(n)}(\alpha_2)=e_{13}^{(n)}e_{21}-e_{23}^{(n)}h_1,$$
$$s^{(n)}(\alpha_1)=e_{13}^{(n)}e_{32}+e_{12}^{(n)}h_2,$$
$$s_1^{(n)}=(2h_1^{(n)}+h_2^{(n)})h_1(h_1+h_2+1)$$
$$+3e_{12}^{(n)}e_{21}(h_1+h_2+1)+3e_{13}^{(n)}e_{21}e_{32}$$
$$+3e_{13}^{(n)}e_{31}(h_1+1),$$
$$s_2^{(n)}=(h_1^{(n)}+2h_2^{(n)})h_2(h_1+h_2+1)$$
$$+3e_{23}^{(n)}e_{32}(h_1+h_2+1)-3e_{13}^{(n)}e_{21}e_{32}$$
$$+3e_{13}^{(n)}e_{31}h_2,$$

$$
\begin{aligned}
s^{(n)}(-\alpha_2) &= e_{32}^{(n)} h_2(h_2-1)(h_1+h_2+1) \\
&\quad - h_2^{(n)} e_{32}(h_2-1)(h_1+h_2+1) \\
&\quad - e_{23}^{(n)} e_{32}^2(h_1+h_2+1) + e_{12}^{(n)} e_{32} e_{21}(h_2-1) \\
&\quad - e_{12}^{(n)} e_{31}(h_2+1)(h_2-1) - e_{13}^{(n)} e_{31} e_{32}(h_2+1) \\
&\quad + e_{13}^{(n)} e_{32}^2 e_{21}, \\
s^{(n)}(-\alpha_1) &= e_{21}^{(n)} h_1(h_1-1)(h_1+h_2+1) \\
&\quad - h_1^{(n)} e_{21}(h_1-1)(h_1+h_2+1) \\
&\quad - e_{12}^{(n)} e_{21}^2(h_1+h_2+1) + e_{23}^{(n)} e_{21} e_{32}(h_1-1) \\
&\quad + e_{23}^{(n)} e_{31}(h_1-1)(h_1+1) - e_{13}^{(n)} e_{31} e_{21}(h_1+1) \\
&\quad - e_{13}^{(n)} e_{21}^2 e_{32}, \\
s^{(n)}(-\alpha_1-\alpha_2) &= e_{31}^{(n)} h_1 h_2(h_1+h_2)(h_1+h_2+1) \\
&\quad + e_{32}^{(n)} e_{21} h_2(h_1+h_2)(h_1+h_2+1) \\
&\quad - e_{21}^{(n)} e_{32} h_1(h_1+h_2)(h_1+h_2+1) \\
&\quad - h_2^{(n)}(e_{32}e_{21} + e_{31}h_1)h_2(h_1+h_2) \\
&\quad - h_1^{(n)}(e_{32}e_{21} - e_{31}(h_2+1))h_1(h_1+h_2) \\
&\quad + e_{12}^{(n)}(e_{32}e_{21}^2 - e_{31}e_{21}(h_2+2))(h_1+h_2) \\
&\quad - e_{23}^{(n)}(e_{32}^2 e_{21} + e_{31}e_{32}h_1)(h_1+h_2) \\
&\quad + e_{13}^{(n)}(e_{32}^2 e_{21}^2 + e_{31}e_{32}e_{21}(h_1-h_2-2) \\
&\quad - e_{31}^2 h_1(h_2+1)).
\end{aligned}
$$


[1] W. Kinnersley and D. M. Chitre, J. Math. Phys. **18**, 1529, 1538 (1977); **19**, 1926, 2037 (1978); I. Hauser and F. J. Ernst, J. Math. Phys. **21**, 1126 (1980); L. Dolan, Phys. Rev. Lett. **47**, 1371 (1981); M. Adler and P. V. M. Moerbeke, Adv. Math. **38**, 267 (1980); C. Devchand and D. B. Fairlie, Nucl. Phys. **194**, 232 (1982); J. Lepowsky and R. L. Wilson, Commun. Math. Phys. **62**, 43 (1978); Y.-S. Wu, Nucl. Phys. B **211**, 160 (1983).

[2] J. Mickelsson, Lett. Math. Phys. **7**, 45 (1983).

[3] L.-L. Chau, M.-L. Ge, and Y.-S. Wu, Phys. Rev. D **25**, 1086 (1982); L. Dolan, Phys. Lett. B **113**, 387 (1982); K. Ueno, and Y. Nakamura, Publ. Res. Inst. Math. Sci. **19**, 519 (1983).

[4] K. Bardacki and M. B. Halpern, Phys. Rev. D **3**, 2493 (1971); E. Witten, Commun. Math. Phys. **92**, 455 (1984).

[5] L. Dolan, "Kac–Moody Algebras and Exact Solvability in Hadronic Physics," Report RU83/B/63, The Rockefeller University, New York, 1983.

[6] A. M. Vershik, I. M. Gelfand, and M. I. Graev, Funct. Anal. Appl. **17**, 137 (1983); A. L. Carey, "Infinite Dimensional Groups and Quantum Field Theory," preprint, Department of Mathematics, Australian National University, Canberra, 1984.

[7] J. G. Nagel and M. Moshinsky, J. Math. Phys. **6**, 682 (1965); S. -C. Pang and K. T. Hecht, J. Math. Phys. **8**, 1233 (1967); M. K. F. Wong, J. Math. Phys. **8**, 1899 (1967); **10**, 1065 (1969); A. M. Bincer, J. Math. Phys. **19**, 1173 (1978).

[8] J. Mickelsson, J. Math. Phys. **12**, 2378 (1971); Rep. Math. Phys. **3**, 193 (1972); **4**, 319 (1973); **4**, 77 (1973); A. M. Bincer, J. Math. Phys. **23**, 347 (1982); M. Perroud, J. Math. Phys. **17**, 1998 (1976).

[9] J. Mickelsson, Ann. Acad. Sci. Fenn. Ser. A 1: Math. Phys. **1**, 285 (1975); Math. Scand. **41**, 63 (1977).

[10] A. Rocha-Caridi and N. R. Wallach, Trans. Am. Math. Soc. **277**, 133 (1983).

[11] I. B. Frenkel, J. Funct. Anal. **44**, 259 (1981); Lecture Notes in Math. **933**, 71 (1982); G. B. Segal, Commun. Math. Phys. **80**, 301 (1981); I. Frenkel and V. G. Kac, Invent. Math. **62**, 23 (1980); V. G. Kac and D. H. Peterson, Proc. Natl. Acad. Sci. USA **78**, 3308 (1981); V. G. Kac, *Infinite Dimensional Lie Algebras* (Birkhäuser, Cambridge, MA, 1983).

[12] J. Mickelsson, Rep. Math. Phys. **18**, 197 (1980).

[13] J. Mickelsson, Rep. Math. Phys. **4**, 307 (1973).

[14] A. van den Hombergh, Indag. Math. **37**, 42 (1975).

[15] J. Mickelsson, Ann. Acad. Sci. Fenn. Ser. A 1:Math. Phys. 598 (1975).

[16] A. van den Homberg, "Harish-Chandra Modules and Representations of Step Algebras," Ph. D. thesis, Department of Mathematics, Katholieke Universiteit te Nijmegen, Netherlands, 1976.

[17] M. D. Gould and S. A. Edwards, "Enveloping Algebra, Annihilators and Projection Techniques for Finite Dimensional Cyclic Modules of a Semi-Simple Lie Algebra," preprint, Department of Mathematical Physics, University of Adelaide, February, 1983.

# Spectral concentration for the Laplace operator in the exterior of a resonator

Claudio Fernández
*Facultad de Matemáticas, Pontificia Universidad Católica de Chile, Casilla 114-D, Santiago, Chile*

(Received 1 August 1984; accepted for publication 21 September 1984)

We consider the Laplace operator defined in the exterior of a resonator and we provide explicit estimates for the spectral concentration, in the sense of Kato, in terms of the width of the channel connecting the cavity with the exterior.

## I. INTRODUCTION

In this paper we use Kato's notion of spectral concentration[1,2] in a situation where the perturbed operators are defined in different Hilbert spaces. This case arises when one considers the negative Laplace operator $H = -\Delta_\Omega$, defined in an exterior region $\Omega$, which has an almost opened cavity. We deal only with the Dirichlet boundary condition. Results involving other boundary conditions, which make the operator $H$ self-adjoint, can be proven in a similar way. In Ref. 3 it is shown that $H$ has resonances in the sense of Lavine,[4] and explicit bounds for the sojourn times are provided in terms of the width of the channel connecting the cavity with the exterior. We use this result to estimate the spectral concentration of $H$ near the resonant energies.

## II. SPECTRAL CONCENTRATION

Let $\rho_0$ be a positive real number and for each $\rho\in[0,\rho_0]$, let $H_\rho$ be a self-adjoint operator defined on a Hilbert space $\mathscr{H}_\rho$. In order to state Kato's definition of spectral concentration, we make the additional assumption that $\mathscr{H}_0$ is a closed subspace of $\mathscr{H}_\rho$. Also, we denote by $\pi_\rho$ the projection of $H_\rho$ onto $\mathscr{H}_0$, and by $E_\rho(\lambda)$ the spectral measure associated to the operator $H_\rho$.

*Definition 1:* For $0<\rho\leqslant\rho_0$, let $S_\rho$ and $T$ be Borel subsets of $R$.

(a) We say that the part of the spectrum of $H_\rho$ in $T$, with respect to $\mathscr{H}_0$, is asymptotically concentrated in $S_\rho$ if s-lim $E_\rho(T-S_\rho)\pi_\rho = 0$, i.e., if $E_\rho(T-S_\rho)\phi\to 0$, as $\rho\to 0$, for all $\phi\in\mathscr{H}_0$. In this case we write $\sigma(H_\rho)\cap T\subsetneq S_\rho$.

(b) Suppose that as $\rho\to 0$, the operator $E_\rho(I)\pi_\rho$ converges strongly to $E_0(I)\pi_\rho$ [i.e., $(E_\rho(I)-E_0(I))\phi\to 0$, for all $\phi\in\mathscr{H}_0$], for any interval $I\subset R$. Then, we say that the part of the spectrum of $H_\rho$ in $S_\rho$, with respect to $\mathscr{H}_0$, is asymptotically the part of the spectrum of $H_0$ in $T$ if s-lim $E_\rho(S_\rho)\pi_\rho = E_0(T)$. In this case we write $\sigma(H_\rho)\cap S_\rho\approx\sigma(H_0)\cap T$.

Throughout the rest of this section we shall assume that the operators $H_\rho$ converge to $H_0$ in $\mathscr{H}_0$, in the sense that $E_\rho(I)\phi\to E_0(I)\phi$, for all $\phi\in\mathscr{H}_0$ and all intervals $I\subset R$. The following result establishes the relationship between parts (a) and (b) of Definition 1. The proof is the same as in the case where all the operators $H_\rho$ are defined on the same Hilbert space.[1,2]

*Proposition 1:* Suppose that $T=(a,b)$, where $a$ and $b$ belong to the resolvent set of $H_0$ and, for each $\rho$ with $0<\rho\leqslant\rho_0$, let $S_\rho\subset T$. Then $\sigma(H_\rho)\cap T\subsetneq S_\rho$ if and only if $\sigma(H_\rho)\cap S_\rho\approx\sigma(H_0)\cap T$.

Closely related to the spectral concentration of a family $\{H_\rho\}$ is the notion of pseudoeigenvector.

*Definition 2:* Let $\{H_\rho\}$ be as above and suppose that $\lambda_0$ is a simple, isolated eigenvalue of $H_0$ with eigenvector $\psi_0\in\mathscr{H}_0$. Two families $\{\psi_\rho\in\mathscr{H}_\rho/0<\rho\leqslant\rho_0\}$ and $\{\lambda_\rho\in R/0<\rho\leqslant\rho_0\}$ are called the pseudoeigenvector and pseudoeigenvalue of $\{H_\rho\}$, respectively, if

$$\text{(i) } \lim_{\rho\to 0}\|\psi_\rho-\psi_0\| = 0,$$

and

$$\text{(ii) } \epsilon_\rho = \|(H_\rho-\lambda_\rho)\psi_\rho\|\to 0, \quad \text{as } \rho\to 0.$$

The rate of convergence of $\epsilon_\rho$ to zero is called the order of the pseudoeigenpair $\{\psi_\rho\},\{\lambda_\rho\}$.

The next result, which we prove in our more general context, was originally established by Kato. Its proof is a slight modification of the argument given in Ref. 1.

**Theorem 1:** Let $H_\rho$ and $\lambda_0$ be as in Definition 2 and let $I$ be a closed interval such that $I\cap\sigma(H_0)=\{\lambda_0\}$. Then, there exists a function $g(\rho)$ such that $g(\rho)\to 0$ as $\rho\to 0$ and $\sigma(H_\rho)\cap I\subsetneq I_\rho=(\lambda_\rho-g(\rho),\lambda_\rho+g(\rho))$ if and only if $\{\lambda_\rho\}$ is a pseudoeigenvalue of $\{H_\rho\}$ of order $\epsilon_\rho$, where $\epsilon_\rho = o(g(\rho))$, as $\rho\to 0$.

*Proof:* Suppose that $\sigma(H_\rho)\cap I\subsetneq I_\rho$, and let $\psi\in\mathscr{H}_0$ be an eigenvector of $H_0$ with eigenvalue $\lambda_0$. By Proposition 1 it then follows that $\sigma(H_\rho)\cap I_\rho\approx\sigma(H_0)\cap I$, which means that $\lim_{\rho\to 0}E_\rho(I_\rho)\phi = E_0(\{\lambda_0\})\phi$, for all $\phi\in\mathscr{H}_0$. Therefore, $\psi_\rho = E_\rho(I_\rho)\psi\to\psi$ and

$$\|(H_\rho-\lambda_\rho)\psi_\rho\|^2 = \int_{I_\rho}(\lambda-\lambda_\rho)^2 d\langle\psi_\rho, E_\rho(\lambda)\psi_\rho\rangle$$
$$\leqslant(g(\rho))^2\|\psi_\rho\|^2.$$

Hence, $\{\lambda_\rho\}$ is a pseudoeigenvalue of order $\epsilon_\rho = 0(g(\rho))$.

In order to prove the converse, let $\{\psi_\rho\},\{\lambda_\rho\}$ be a pseudoeigenpair of $\{H_\rho\}$ of order $\{\epsilon_\rho\}$ and let $g(\rho)$ be such that $\epsilon_\rho = o(g(\rho))$, as $\rho\to 0$. Then,

$$\epsilon_\rho^2 = \int(\lambda-\lambda_\rho)^2\langle d\psi_\rho, E_\rho(\lambda)\psi_\rho\rangle$$
$$\geqslant(g(\rho))^2\int_{R-I_\rho} d\langle\psi_\rho, E_\rho(\lambda)\psi_\rho\rangle,$$

and we have $\|E_\rho(I-I_\rho)\psi_\rho\|\leqslant\|(I-E(I_\rho))\psi_\rho\|\leqslant\epsilon_\rho/g(\rho)$, which converges to zero. Since $\psi_\rho\to\psi$ as $\rho\to 0$, we conclude that $E_\rho(I-I_\rho)\psi\to 0$, and therefore $E_\rho(I-I_\rho)E_0\{\lambda_0\}\to 0$. On the other hand,

$$E_\rho(I-I_\rho)(I-E_0(\{\lambda_0\}))\leqslant E_\rho(I)(I-E_0(\{\lambda_0\}))$$
$$\to E_0(I)(I-E_0(\{\lambda_0\})) = 0.$$

We thus conclude that $E_\rho(I-I_\rho)$ converges strongly to zero, with respect to $\mathscr{H}_0$, i.e., $\sigma(H_\rho)\cap I\subsetneq I_\rho$. ■

 0022-2488/85/030383-02$02.50 © 1985 American Institute of Physics 

We remark that Theorem 1 can be extended to the case where $\lambda_0$ is an eigenvalue of $H_0$ of multiplicity greater than 1. This extended result can be proved using the same argument given above (see Ref. 2).

## III. THE LAPLACE OPERATOR IN THE EXTERIOR OF A RESONATOR

For $\rho>0$, $\Omega_\rho = C \cup Z_\rho \cup E$ will denote an exterior domain in $R^3$ with smooth boundary, which consists of a bounded set $C$ (the cavity), an exterior region $E$, and a channel $Z_\rho$ connecting the cavity with the exterior. We assume that there exists $\rho_0>0$, such that for $0<\rho\leqslant\rho_0$, $Z_\rho$ is contained in a cylinder of radius $\rho$ and height $h$. The self-adjoint operators acting on $\mathscr{H}_\rho = L^2(\Omega_\rho)$ and $\mathscr{H}_0 = L^2(C)$ given by the negative Laplacian with Dirichlet boundary condition will be denoted by $H_\rho$ and $H_0$, respectively.

It seems apparent that as the channel closes up, the part of $H_\rho$ in $L^2(C)$ converges, in some generalized sense, to the operator $H_0$. On the other hand, all the operators $H_\rho$ (for $\rho>0$) have continuous spectra consisting of the interval $[0, \infty)$, while the spectrum of $H_0$ is formed by a countable set of isolated eigenvalues. In what follows, we shall construct a pseudoeigenvector of the family $\{H_\rho\}$. By the results of the previous section we then conclude that as the channel closes up, the spectrum of $H_\rho$ concentrates, in the sense of Kato, around the eigenvalues of the cavity. First we show that as $\rho\to 0$, $E_\rho(I)\phi - E_0(I)\phi$, for all $\phi\in\mathscr{H}_0$. This follows as usual from the corresponding result for the resolvents of $H_\rho$ and $H_0$, which we prove now. Given $z$ in the resolvent set of $H_\rho$, we write $\delta = \delta(z) = \operatorname{dist}(z, \sigma(H_\rho))$, so that $\|(H_\rho - z)^{-1}\| = 1/\delta$.

*Lemma 1:* As $\rho$ converges to zero, the operator $\chi_{Z_\rho}(H_\rho - z)^{-1}$ converges to zero in norm.

*Proof:* Given $\phi\in\mathscr{H}_\rho$, set $\psi_\rho = (H_\rho - z)^{-1}\phi$, so that $\psi_\rho\in D(H_\rho)$ and $(-\Delta - z)\psi_\rho = \phi$. Then, by the Poincaré inequality,

$$\int_{Z_\rho}|\psi_\rho|^2\leqslant\rho^2\int_{Z_\rho}|\nabla\psi_\rho|^2\leqslant\rho^2\langle\psi_\rho, -\Delta\psi_\rho\rangle$$
$$=\rho^2(\langle\psi_\rho,\phi\rangle + z\|\psi_\rho\|^2).$$

But, since $\|\psi_\rho\|\leqslant 1/\delta\,\|\phi\|$, we obtain

$$\|\chi_{Z_\rho}(H_\rho - z)^{-1}\phi\|^2\leqslant\rho^2(\|\psi_\rho\|\|\phi\| + |z|\|\psi_\rho\|^2)$$
$$\leqslant\rho^2/\delta(1 + |z|/\delta)\|\phi\|^2.$$

Hence $\|\psi_{Z_\rho}(H_\rho - z)^{-1}\|^2\leqslant(\rho^2/\delta)(1 + |z|/\delta)$, which converges to zero. ■

**Theorem 2:** For any $\phi\in\mathscr{H}_0$, $(H_\rho - z)^{-1}\phi$ converges to $(H_0 - z)^{-1}\phi$, as $\rho$ converges to zero.

*Proof:* As in the proof of Lemma 1 we set $\psi_\rho = (H_\rho - z)^{-1}\phi$ (where $\phi$ is extended to be zero outside $C$). First, we show that $\chi_{Z_\rho\cup E}\psi_\rho\to 0$. Let $\alpha$ be a $C^\infty$ function, such that $\alpha = 0$ in $C$, $\alpha = 1$ in $E$, and $0\leqslant\alpha\leqslant 1$. Multiplying the equation $(-\Delta - z)\psi_\rho = \phi$ by $\alpha\bar{\psi}_\rho$ and integrating over $\Omega_\rho$, we obtain

$$\int\alpha|\nabla\psi_\rho|^2 + \int\bar{\psi}_\rho\nabla\alpha\cdot\nabla\psi_\rho - z\int\alpha|\psi_\rho|^2 = 0.$$

If we assume that $v = \operatorname{Im} z\neq 0$, then it follows that

$$|v|\int\alpha|\psi_\rho|^2\leqslant\left|\int\bar{\psi}_\rho\nabla\alpha\cdot\nabla\psi_\rho\right|\leqslant C\|\psi_\rho\|_{Z_\rho}\|\nabla\psi_\rho\|.$$

But, since $\|\nabla\psi_\rho\|^2 = \langle\psi_\rho,(-\Delta - z)\psi_\rho\rangle + z\|\psi_\rho\|^2\leqslant(1/\delta + |z|/\delta^2)\|\phi\|^2$, we conclude that

$$\int_{Z_\rho\cup E}\alpha|\psi_\rho|^2\leqslant C'\|\psi_\rho\|_{Z_\rho}\|\phi\|.$$

On the other hand, if $v = 0$, then $k = \operatorname{Re} z<0$ and (1) implies that

$$|k|\int\alpha|\psi_\rho|^2\leqslant -\operatorname{Re}\int\bar{\psi}_\rho\nabla\alpha\cdot\nabla\psi_\rho$$
$$=\frac{1}{2}\int\Delta\alpha|\psi_\rho|^2\leqslant C''\|\psi_\rho\|^2_{Z_\rho}.$$

By Lemma 1, it follows that $\chi_{Z_\rho\cup E}\psi_\rho\to 0$, as $\rho$ converges to zero.

Finally, since $(-\Delta - z)\psi_\rho = \phi$, we have

$$\int_{\Omega_\rho}|\nabla\psi_\rho|^2 - z\int_{\Omega_\rho}|\psi_\rho|^2 = \int_{\Omega_\rho}\bar{\psi}_\rho\phi.$$

Therefore, $\{\psi_\rho\}$ is bounded in $\mathscr{H}^1(C)$ and by compactness of the inclusion $\mathscr{H}^1(C)\to L^2(C)$ we can assume that it has a subsequence $\{\psi_{\rho_n}\}$ which converges to $\psi_0\in\mathscr{H}^1(C)$, weakly in $\mathscr{H}^1(C)$ and strongly in $L^2(C)$. A standard argument shows that $(-\Delta - z)\psi_0 = \phi$ weakly in $C$, and that $\psi_0 = 0$ on the boundary of $C$, in the $L^2$ sense (see Ref. 3). We conclude that $\psi_0 = (H_0 - z)^{-1}\phi$, as desired. ■

Finally, we construct a pseudoeigenpair for the family $\{H_\rho\}$. As in Ref. 3, let us consider the perturbed operator $A_\rho = H_\rho + V_\rho$, with $D(A_\rho) = D(H_\rho)$, where $V_\rho = (1/\rho^2)\chi_E$. For each $\rho>0$, $A_\rho$ is a positive self-adjoint operator, whose discrete spectrum consists of a finite number of eigenvalues below $1/\rho^2$, and with essential spectrum $[1/\rho^2, \infty)$. Moreover, as $\rho$ approaches zero, the eigenvalues and eigenvectors of $A_\rho$ converge to the corresponding eigenvalues and eigenvectors of $H_0 = -\Delta_C$ (see Ref. 3).

**Theorem 3:** Let $\phi_\rho\in\mathscr{H}_\rho$ and $\lambda_\rho>0$ be such that $A_\rho\phi_\rho = \lambda_\rho\phi_\rho$. Then $\{\psi_\rho\}$ is a pseudoeigenvector of $\{H_\rho\}$ with order

$$\epsilon_\rho = \|(H_\rho - \lambda_\rho)\phi_\rho\| = (\sqrt{\lambda_\rho}/c\rho)e^{-hc/\rho},$$

where $c = (1 - \lambda_\rho\rho^2)^{1/2}$.

*Proof:* We have that $\|(H_\rho - \lambda_\rho)\phi_\rho\| = \|V_\rho\phi_\rho\| = 1/\rho^2\|\phi_\rho\|_E$ and, therefore, the theorem follows from

$$\|\phi_\rho\|^2_E\leqslant(\lambda_\rho\rho^2/c)e^{-2hc/\rho},$$ which is proven in Ref. 3. ■


## ACKNOWLEDGMENT

The author's research is supported by the Dirección de Investigación, Universidad Católica de Chile.



[1] T. Kato, *Perturbation Theory for Lineal Operators* (Springer, New York, 1976), 2nd ed.

[2] M. Reed and B. Simon, *Methods of Modern Mathematical Physics IV, Analysis of Operators* (Academic, New York, 1979).

[3] C. Fernández, "Resonances in Scattering by a resonator," to appear in the Indiana Math. J.

[4] R. Lavine, "Spectral density and Sojourn times," in *Atomic Scattering Theory*, edited by J. Nutall (Univ. Western Ontario, London, Ontario, 1978).

# Consistent superspace integration

Alice Rogers
*Department of Mathematics, King's College, Strand, London WC2R 2LS, England*

(Received 25 May 1984; accepted for publication 20 July 1984)

The standard Berezin method for integration over odd variables is combined in a new way with De Witt's contour method for integration over even Grassmann variables to give a new method of superspace integration. It is shown that this integral, unlike the standard superspace integral, is invariant under coordinate transformations in superspace. The relation between the new method and the standard method is discussed.

## I. INTRODUCTION

In recent years integration over anticommuting variables has become an essential technique in many areas of quantum field theory. The integration is invariably carried out according to the Berezin prescription[1]

$$\int \theta \, d\theta = 1, \quad \int 1 \, d\theta = 0. \tag{1.1}$$

[Here, as usual, $\theta$ denotes an odd Grassman variable, or "anticommuting number" as it is often called. Because the square of $\theta$ is zero, the two rules in (1.1) cover all the functions usually considered. No "limits of integration" are included.] In some cases, such as in supersymmetric theories, odd and even variables intermingle, and so Berezin integration must be fitted in with ordinary integration. However, the standard method for doing this, as developed by Berezin and Leîtes,[2] Bernstein and Leîtes,[3,4] and Berezin[5] does not always give consistent results under change of coordinates.[3,6] In this paper some of the ideas of Berezin, Bernstein, and Leîtes are combined in a new way with De Witt's idea[7] that integration over even variables should be regarded as contour integration, to give a fully consistent method for integrating over odd and even variables. By "fully consistent" it is meant that the rule for transforming the integrand and the "range of integration" under a suitably smooth change of coordinates (including mixture of odd and even coordinates) should be such that the value of the integral is unchanged; thus the definition is coordinate independent. Coordinate independence is obviously essential if one wishes to integrate on general supermanifolds (made by patching regions with local odd and even coordinates). But even when considering the rigid "super-Minkowski" space of global supersymmetry, where symmetry transformations are represented as translations, a (more restricted) coordinate independence is required of integrals; even for this invariance the standard method of superspace integration is inadequate.

Several examples are given below, showing how the consistency breaks down; these make it clear that, while the Berezin method takes care of the odd variables, no account has been taken of the fact that when one includes odd Grassmann variables the even variables are not merely real (or complex) variables, but also have a nilpotent part, and the theory of ordinary integration over real (or complex) numbers must be extended to integration over the full even algebra. Previously it has been assumed that these parts can be ignored (some reasons for this are given below); it is the purpose of this paper to show that the idea (due to De Witt[7]) of regarding integrals over even Grassmann variables as contour integrals can (if carefully combined with Berezin integration for the odd variables) give a consistent method of superspace integration.

In Sec. II, after specifying some necessary notation and terminology, the standard method for superspace integration is described, and some examples given of where its consistency under change of variable breaks down. In Sec. III contour integrals over one even Grassmann variable are defined, and several useful properties proved. Section IV repeats the work of Sec. III for contour integrals over several even variables. Section V contains the main result of this paper: a method for combining the contour integrals with Berezin integration is described, and proved to give fully consistent results under change of variable. In Sec. VI the compatibility between the contour approach and the standard approach is discussed, with particular reference to supersymmetry and supergravity, while Sec. VII contains conclusions and discusses further possibilities.

## II. THE STANDARD METHOD AND EXAMPLES OF ITS BREAKDOWN

Before describing the standard method, some notation and terminology is necessary.

*Definition 2.1:* (a) For each positive integer $L$ let $B_L$ denote the real Grassman algebra over $L$ odd generators. (Here, and through the rest of this paper, attention is restricted to Grassmann algebras over the reals; the extension to complex Grassmann algebras is straightforward but not entirely trivial, and will be dealt with in a separate paper.)

(b) Let $\epsilon$: $B_L \to \mathbb{R}$ be the (unique) algebra homomorphism which sets all the odd generators to zero.

(c) For each pair of positive integers $m$ and $n$ let $B_L^{m,n}$ denote the Cartesian product of $m$ copies of the even part of $B_L$ and $n$ copies of the odd part. ($B_L^{m,n}$ is often referred to as "super-Euclidean space" or simply "superspace". In applications to supersymmetry, superspace has additional transformation properties under groups such as the Lorentz group, but these are not relevant here.) A typical element of $B_L^{m,n}$ is denoted $(x^1, ..., x^m; \theta^1, ..., \theta^n)$ or $(x;\theta)$.

(d) Let $\epsilon_{m,n}: B_L^{m,n} \to \mathbb{R}^m$ be defined by

$$\epsilon(x^1, ..., x^m;\theta^1, ..., \theta^n) = (\epsilon(x^1), ..., \epsilon(x^m)). \tag{2.1}$$

 0022-2488/85/030385-08$02.50 © 1985 American Institute of Physics 

Turning now to the standard method of superspace integration, suppose that $U$ is an open set in $\mathbf{R}^m$ and let $f$: $\epsilon_{m,n}^{-1}(U)\to B_L$ be sufficiently "analytic" for it to have a power series expansion in the $\theta^j$. Such a series will necessarily terminate. If its highest term is $f_n(x)\theta^n\theta^{n-1}\cdots\theta^1$ then one defines[2–5]

$$\int_U f(x,\theta)d^m x\, d^n\theta := \int_U f_n(x)dx, \tag{2.2}$$

where on the right-hand side $x$ is simply treated as a real variable, and the integral is evaluated in the usual manner. Under a suitably smooth and invertible change of coordinates

$$h:\epsilon_{m,n}^{-1}(U)\to B_L^{m,n},$$

with

$$h(x,\theta) = (y(x,\theta);\phi(x,\theta)), \tag{2.3}$$

the "volume form" $d^m x\, d^n\theta$ is defined to transform according to the rule

$$d^m x\, d^n\theta = \mathrm{Ber}(J(h)(y,\phi))d^m y\, d^n\phi, \tag{2.4}$$

where Ber denotes the superdeterminant (named after F. A. Berezin who first discovered it in 1971) while $J(h)(y,\phi)$ is the $(m,n)\times(m,n)$ Jacobian matrix

$$J(h) = \begin{pmatrix} \dfrac{\partial x^i}{\partial y^k} & \dfrac{\partial\theta^j}{\partial y^k} \\ \dfrac{\partial x^i}{\partial\phi^l} & \dfrac{\partial\theta^j}{\partial\phi^l} \end{pmatrix}. \tag{2.5}$$

(Derivatives of functions of Grassman variables are defined below, and in detail in Ref. 8. In this section an informal approach is sufficient.) The function $f$ transforms in the usual manner, while $U$ becomes $\epsilon_{m,n}\circ h\circ\epsilon_{m,n}^{-1}(U)$. Thus the integral is invariant under the change of coordinates if

$$\int_U f d^m x\, d^n\theta = \pm\int_{\epsilon_{m,n}\circ h\circ\epsilon_{m,n}^{-1}(U)} fJ\, d^m y\, d^n\phi \tag{2.6}$$

(where the sign is determined by the orientation of $h$). It is shown by Bernstein and Leĭtes[3] and by Pakhomov[9] that this is true if $f|_U$ has compact support, and is $C^\infty$. A similar argument is used by Regge.[10] An alternative proof, due to Fung, is quoted by Van Nieuwenhuizen in Ref. 11; this gives an incomplete proof for a wider class of functions. The proof is incomplete because it assumes (wrongly) that "ordinary" integration over even variables is consistently handled by ignoring the even nilpotent elements in the range of integration. In this paper we overcome this problem by introducing contour integrals as suggested by De Witt[7] and then Fung's proof can be adapted to give a complete proof (Sec. V). First we present some examples where the consistency breaks down, if one uses the rules described above.

*Example 2.2:* Integration over $B_L^{1,0}$ (one even Grassmann variable). Let $a, b\in\mathbf{R}$. Consider

$$\int_a^b x\, dx = \frac{1}{2}(b^2 - a^2). \tag{2.7}$$

Now apply the change of variable $h$: $\epsilon_{1,0}^{-1}(a,b)\to B_L^{1,0}$, where $h(x) = x+\alpha$, $\alpha$ being some fixed even nilpotent element in $B_L^{1,0}$. The transformed integral is then

$$\int_a^b (x-\alpha)dx = \left[\frac{x^2}{2} - \alpha x\right]_a^b = \frac{1}{2}(b^2 - a^2) - \alpha(b-a), \tag{2.8}$$

and thus the rule for transforming the integral has led to a change in its value.

*Example 2.3:* Integration on $B_L^{1,2}$. (This example is given in Ref. 3.) Consider

$$\int_a^b x\, dx\, d\theta^1\, d\theta^2 = 0. \tag{2.9}$$

Now apply the change of variables $h$: $B_L^{1,2}\to B_L^{1,2}$, where

$$h(x,\theta^1,\theta^2) = h(x+\theta^1\theta^2,\theta^1,\theta^2). \tag{2.10}$$

This is a bijective mapping; the Berezinian of the transformation is 1 and the transformation leaves the range of integration unchanged. The transformed integral is

$$\int_a^b (x-\theta^1\theta^2)dx\, d\theta^1\, d\theta^2 = (b-a). \tag{2.11}$$

Again the value of the integral has changed, this time by a pure real number.

*Example 2.4:* Integration over $B_L^{1,1}$. Consider

$$\int_a^b x\, dx\, d\theta = 0. \tag{2.12}$$

Now apply the change of variables $h$: $B_L^{1,1}\to B_L^{1,1}$, where

$$h(x,\theta) = (x+\eta\theta,\theta), \tag{2.13}$$

with $\eta$ some fixed odd Grassmann element. Again the Berezinian of the transformation is 1 and the transformation leaves the range of integration unchanged, and thus the transformed integral is

$$\int_a^b (x-\eta\theta)dx\, d\theta = -\eta(b-a). \tag{2.14}$$

Example 2.2 makes it glaringly obvious that the source of the problem is that the limits of integration are not changed, because the method pays no attention to the nilpotent even parts. The contour integral approach of De Witt[7] has no such problems. There is also an example in the physics literature, where shifts of the range of integration in the even nilpotent directions prove essential and are introduced by Hassoun *et al.*[12] The formula these authors give for constant even shifts is precisely that which the contour integrals to be defined in this paper give. It is perhaps worth mentioning the historical reasons for the seemingly curious treatment of the range of integration which the standard method uses. Initially odd variables were introduced at a more abstract level than that at which they are used in the physicists' superspace. If $U$ is an open set in $\mathbf{R}^m$ (or, more generally, a $C^\infty$ manifold) then the set of infinitely differentiable functions on $U$ is denoted $C^\infty(U)$. This set naturally has the structure of an algebra (one can add and multiply functions in an obvious manner). Initially odd variables were introduced by extending these function algebras without actually extending the underlying manifold[2,3]; in this approach the idea of shifting the range of integration in even nilpotent directions has no meaning. Subsequently, largely motivated by supersymmetry, a more down-to-earth approach to odd variables was introduced with $\mathbf{R}^m$ extended to $B_L^{m,n}$ and then

functions on this extended space being considered. The two approaches, algebraic and geometric, can be linked[8,14] and regarded as equivalent except that the algebras in the first approach must be extended by taking their tensor product with $B_L$. In the geometric approach a topology was introduced by De Witt[7] which takes an "all or nothing" approach to the nilpotent parts of the Grassmann algebra; a set $V$ in $B_L^{m,n}$ is open if and only if there exists an open set $U$ in $\mathbb{R}^m$ such that

$$V = \epsilon_{m,n}^{-1}(U), \tag{2.15}$$

so that while an open set may be bounded in the "real" direction, it contains the full range in all the nilpotent directions. In this topology the open set $V$ in $B_L^{m,n}$ may be unambiguously specified by the open set $U = \epsilon_{m,n}(V)$ in $\mathbb{R}^m$; in the geometric approach this is how the region of integration $U$ in (2.2) is interpreted, and thus it is obvious that shifts in nilpotent directions will not affect the range of integration.

In the rest of this paper a consistent theory of integration is used making use of the geometric approach to odd variables, and a finer (norm-induced) topology on $B_L^{m,n}$.

## III. CONTOUR INTEGRALS IN ONE-DIMENSIONAL EVEN SUPERSPACE

In the previous section we showed that the most commonly adopted approach to integration over even Grassmann variables was unsatisfactory, in that it did not lead to coordinate-independent (and thus well-defined) integrals. Now, for more than a century, there has existed a consistent theory of integration over one commuting algebra over the reals—the theory of integration of a complex variable. In this section, following the idea of De Witt,[7] we take a similar approach to integration in even superspace and find that many useful, indeed vital, properties of complex contour integration apply here. The definition and results are similar to those of De Witt[7]; the difference is that the detailed superspace formalism used here (developed in Ref. 8 and described below) gives $B_L$ a Banach algebra structure and puts the analysis on a firmer footing, especially when considering an algebra with an infinite number of generators. Before proceeding to the fundamental definition of a contour integral, a few definitions relating to the detailed structure of $B_L$ will be given.

*Definition 3.1:* (a) $B_L$ denotes the algebra over the reals with generators $1, \beta_1, \ldots, \beta_L$ and relations

$$\beta_i \beta_j = -\beta_j \beta_i, \quad 1 \leqslant i,j \leqslant L. \tag{3.1}$$

(b) Following Kostant,[3] if $L$ is a positive integer, $M_L$ denotes the set of sequences of integers $\mu = (\mu_1, \ldots, \mu_k)$, where $1 \leqslant \mu_1 < \mu_2 < \cdots < \mu_k \leqslant L$. $M_{L0}$ and $M_{L1}$ denote sequences in $M_L$ with even and odd numbers of terms, respectively. Also, $M_L$ includes the empty sequence, denoted $\phi$.

(c) A typical element of $B_L$ may be expressed as

$$b = \sum_{\mu \in M_L} b^\mu \beta_\mu, \tag{3.2}$$

where each $b_\mu$ is a uniquely determined real number and

$$\beta_\mu := \beta_{\mu_1} \beta_{\mu_2} \cdots \beta_{\mu_k}, \quad \beta_\phi = 1.$$

(d) The set of even elements in $B_L$ is denoted $B_{L0}$.

(e) A norm on $B_L$ is defined by

$$\|b\| = \sum_{\mu \in M_L} |b^\mu|. \tag{3.3}$$

It is proved in Ref. 8 that this norm gives $B_L$ the structure of a Banach algebra; throughout the rest of this paper the topology used on $B_L$ will be the topology induced by this norm (it is also the usual topology on $B_L$ regarded as a finite dimensional vector space). (An extension of these ideas to a "Grassmann algebra" with an infinite number of generators is described in Ref. 8.)

The definition of a contour integral on even superspace, which is fundamental to this paper, will now be given. Apart from some analytic details, it is essentially that due to De Witt.[7]

*Definition 3.2:* (a) Let $[a,b]$ be a finite closed interval in $\mathbb{R}$. The mapping $\gamma: [a,b] \to B_{L0}$ is called a path in $B_{L0}$ if it is a continuous and piecewise $C^1$ mapping of real Banach spaces. The path $\gamma$ is closed if $\gamma(a) = \gamma(b)$.

(b) Suppose that $\gamma: [a,b] \to B_{L0}$ is a path in $B_{L0}$ and that $U$ is open in $B_{L0}$ with $\gamma([a,b]) \subset U$. Also suppose $f: U \to B_L$ is a continuous map of Banach spaces. Then the integral of $f$ along the path $\gamma$ is defined to be

$$\int_\gamma f\,dx := \int_b^a f(\gamma(t))\gamma'(t)\,dt. \tag{3.4}$$

This integration has several useful properties which are precise analogs of the properties of complex contour integration; they do not depend on the specific algebra, but only on the properties of maps between Banach spaces and on the fact that we are considering a commutative algebra. The first of these theorems gives a consistent method for transforming the integral under change of variable. This underpins the method for consistent integration in full (even and odd) superspace developed in Sec. V. Before starting the theorem, a method for differentiating functions of even Grassmann variables is required, together with some properties of this form of differentiation.

*Definition 3.3:* Let $U$ be open in $B_{L0}$ and let $f: U \to B_L$. Then $f$ is said to be $G^1$ on $U$ if there exists $f_1: U \to B_L$ and $\eta: B_{L0} \to B_L$ such that given $a$, $a + h$ in $U$,

$$f(a+h) = f(a) + hf_1(a) + \eta(h)\|h\|, \tag{3.5}$$

with

$$\|\eta(h)\| \to 0 \quad \text{as} \quad \|h\| \to 0. \tag{3.6}$$

[Also one may define $G^P$ inductively, by saying that $f$ is $G^P$ if $f$ is $G^{p-1}$ and the $(p-1)$-th derivative of $f$ is $G^1$.]

Further details of this "superdifferentiation" and its properties may be found in Ref. 8. In particular the following "chain rule" is proved in this paper [Proposition 2.12(h)].

*Proposition 3.4:* Let $\gamma: [a,b] \to U$, where $U$ is open in $B_{L0}$ be a $C^1$ map of Banach spaces and let $h: U \to B_L$ be $G^1$. Then

$$\frac{d}{dt}(h \circ \gamma) = (h_1 \circ \gamma)\frac{d\gamma}{dt} \tag{3.7}$$

(where, as in Definition 3.3, $h_1$ denotes the superderivative of $h$).

**Theorem 3.5:** Let $U$ be open in $B_{L,0}$ and let $h: U \to B_{L,0}$

be an injective $G^1$ mapping. Also let $\gamma$: $[a,b]\to U$ be a path in $U$, and $f$: $U\to B_L$ be continuous. Then

$$\int_\gamma f\,dx = \int_{h\circ\gamma} f\circ h^{-1}\frac{1}{h_1}\,dx \tag{3.8}$$

(where $h_1$ is the superderivative of $h$ defined in Definition 3.3).

*Proof:*

$$\begin{aligned}\int_{h\circ\gamma} f\circ h^{-1}\frac{1}{h_1}\,dx &= \int_b^a \frac{(f\circ h^{-1})(h\circ\gamma(t))(h\circ\gamma)'(t)}{h_1(\gamma(t))}\,dt \\ &= \int_a^b f(\gamma(t))\frac{h_1(\gamma(t))\gamma'(t)}{h_1(\gamma(t))}\,dt \end{aligned}\tag{3.9}$$

(by Proposition 3.4).

Hence

$$\int_{h\circ\gamma} f\circ h^{-1}\frac{1}{h_1}\,dx = \int_\gamma f\,dx. \tag{3.10}$$

Another theorem for complex contour integrals which has its analog here is Cauchy's theorem which states that the integral of a suitably differentiable function around a closed path is zero; this allows one to integrate a $G^1$ function "between limits" without specifying the contour, and also to clarify the relationship between the contour approach and the standard approach to even superspace integration.

**Theorem 3.6** (Generalized Cauchy theorem): Let $\gamma$: $[a,b]\to B_{L0}$ be a closed path in $B_{L0}$. Let $S$ be a smooth surface in $B_{L0}$ bounded by $\gamma$. Also, suppose that $U$ is an open subset of $B_{L0}$ with $S\subset U$ and let $f$: $U\to B_L$. Then, if $f$ is $G^1$ on $U$, and $f_1$ is continuous,

$$\int_\gamma f\,dx = 0. \tag{3.11}$$

*Corollary 3.7:* Let $\gamma_1$: $[a,b]\to B_{L0}$, $\gamma_2$: $[c,d]\to B_{L0}$ be paths in $B_{L0}$ with $\gamma_1(a)=\gamma_2(c)=p$ (say) and $\gamma_1(b)=\gamma_2(d)=q$ (say). Let $S$ be a surface bounded by $\gamma_1$ and $\gamma_2$. Also suppose $U$ is an open subset of $B_{L0}$ with $S\subset U$, and let $f$: $U\to B_L$. Then (a), if $f$ is $G^1$ on $U$,

$$\int_{\gamma_1} f\,dx = \int_{\gamma_2} f\,dx. \tag{3.12}$$

(b) If $f$: $B_{L0}\to B_L$ is $G^1$ on all $B_{L0}$ one can write unambiguously

$$\int_p^q f\,dx = \int_\gamma f\,dx, \tag{3.13}$$

where $\gamma$: $[a',b']\to B_{L0}$ is any path in $B_{L0}$ with $\gamma(a')=p$ and $\gamma(b')=q$. Before proving Theorem 3.6, a lemma establishing the analog of the Cauchy–Riemann equations is required. A function from $B_{L0}$ into $B_L$ can be regarded simply as a function from $\mathbb{R}^{2^{L-1}}\to\mathbb{R}^{2^L}$. In the lemma are found conditions which the partial derivatives of the function must obey if the function is $G^1$ (superdifferentiable).

*Lemma 3.8:* Let $f$: $U\to B_L$ where $U$ is an open subset of $B_{L0}$. Let $\iota$ denote the natural identification of $B_{L0}$ and $\mathbb{R}^{2^{L-1}}$,

$$\iota\Big(\sum_{\mu\in M_{L0}} x^\mu\beta_\mu\Big) := (x_\phi, x_{(1,2)}, x_{(1,3)}, \ldots, x_{(L-1,L)}, \ldots). \tag{3.14}$$

[The expression on the right-hand side of (3.14) will be denoted $(x_\mu)$.] Then, if $f$ is $G^1$ with continuous derivative $f_1$ on $U$, (a) $f\circ\iota^{-1}$: $\mathbb{R}^{2^{L-1}}\to B_L$ is a $C^1$ function of Banach spaces, with continuous partial derivatives and (b) if $2^{L-1}$ functions $f^\rho$: $\iota(U)\to\mathbb{R}$ are defined by

$$\sum_{\rho\in M_{L0}} f^\rho(x_\mu)\beta_\rho = f\Big(\sum_{\mu\in M_{L0}} x^\mu\beta_\mu\Big), \tag{3.15}$$

then

$$\sum_{\rho\in M_{L0}} \beta_\rho\,\partial_\mu f^\rho = \beta_\mu f_1\circ\iota, \tag{3.16}$$

where $\partial_\mu$ denotes differentiation with respect to $x^\mu$. [These equations (3.16) will be referred to as the generalized Cauchy–Riemann equations.]

*Proof:* [Note that an unconventional norm on $\mathbb{R}^{2^{L-1}}$, that is $\|(x^\mu)\| := \Sigma_{\mu\in M_{L0}}|x^\mu|$, must be used. This is because of the norm (3.3) used on $B_L$, which gives $B_L$ a Banach algebra structure. Because the norm is equivalent to the usual one, the definition of derivative is unaffected.[15]] Choose $x$, $x+h\in U$. Then

$$f(x+h) = f(x) + hf_1(x) + \|h\|\eta(h), \tag{3.17}$$

where

$$\|\eta(h)\|\to 0 \quad \text{as} \|h\|\to 0.$$

Hence

$$\begin{aligned} f\circ\iota^{-1}(x^\mu+h^\mu) &= f\circ\iota^{-1}(x^\mu) + \sum_{\mu\in M_{L0}} h^\mu\beta_\mu f_1\circ\iota^{-1}(x^\mu) \\ &\quad + \|h\|\,\eta\circ\iota^{-1}(h^\mu). \end{aligned}\tag{3.18}$$

The result follows immediately.

*Proof of Theorem 3.6:*

$$\int_\gamma f\,dx = \int_a^b f(\gamma(t))\gamma'(t)\,dt \tag{3.19}$$

$$= \sum_{\mu\in M_L}\int_a^b f^\mu\circ\iota(\gamma(t))\beta_\mu\gamma'(t)\,dt \tag{3.20}$$

$$= \sum_{\mu\in M_L,\nu\in M_{L0}}\int_{\iota\circ\gamma([a,b])} f^\mu\,dx^\nu\beta_\mu\beta_\nu \tag{3.21}$$

$$= \sum_{\mu\in M_L\rho,\nu\in M_{L0}}\int_{\iota(S)}\Big(\beta_\nu\beta_\mu\frac{\partial f^\mu}{\partial x^\rho} - \beta_\rho\beta_\mu\frac{\partial f^\mu}{\partial x^\nu}\Big) \times dx^\rho\,dx^\nu, \tag{3.22}$$

using Stokes' theorem. Thus by Lemma 3.8,

$$\begin{aligned}\int_\gamma f\,dx &= \sum_{\rho\in M_{L0},\nu\in M_{L0}}\int(\beta_\nu\beta_\rho - \beta_\rho\beta_\nu)f_1\,dx^\rho\,dx^\nu \\ &= 0, \quad \text{since each } \nu\in M_{L0}. \end{aligned}\tag{3.23}$$

The proof of Corollary 3.7 requires a further lemma, this time on the reparametrization of paths.

*Lemma 3.9:* (a) Let $\gamma$: $[a,b]\to B_{L0}$ be a path in $B_{L0}$ and let $c$, $d\in\mathbb{R}$. Also let $\alpha$: $[c,d]\to[a,b]$ be $C^1$ with $\alpha(c)=a$,

$\alpha(d) = b$ and $\alpha'(t) > 0$ for all $t$ in $[c,d]$. Then

$$\int_\gamma f\,dx = \int_{\gamma\circ\alpha} f\,dx \tag{3.24}$$

[and so the integral is independent of the parametrization of the path, apart from considerations of orientation, which is the content of part (c) of this lemma].

(b) Let $\gamma_1$: $[a,b]\to B_{L0}$ and $\gamma_2$: $[c,d]\to B_{L0}$ be two paths in $B_{L0}$ with $\gamma_1(b) = \gamma_2(c)$. Also define $\gamma_1 + \gamma_2$ to be the path $\gamma_1 + \gamma_2$: $[a, b + d - c]\to B_L$ defined by

$$(\gamma_1 + \gamma_2)(t) = \begin{cases} \gamma_1(t), & a \leqslant t \leqslant b, \\ \gamma_2(t - b + c), & b < t \leqslant b + d - c. \end{cases} \tag{3.25}$$

Then, if $U$ is open in $B_{L0}$, $f$: $U\to B_L$ is continuous and $\gamma_1([a,b])\subset U$, $\gamma_2([c,d])\subset U$,

$$\int_{\gamma_1+\gamma_2} f\,dx = \int_{\gamma_1} f\,dx + \int_{\gamma_2} f\,dx. \tag{3.26}$$

(c) Let $\gamma$: $[a,b]\to B_{L0}$ be a path in $B_{L0}$. Define the curve $-\gamma$: $[a,b]\to B_L$ by

$$-\gamma(t) = \gamma(a + b - t). \tag{3.27}$$

Then, if $U$ is open in $B_{L0}$, with $\gamma([a,b])\subset U$, and $f$: $U\to B_L$ is continuous,

$$\int_{-\gamma} f\,dz = -\int_\gamma f\,dz. \tag{3.28}$$

The proof of this lemma is omitted because it is essentially the same as the proof of the equivalent result in complex analysis. Corollary 3.7 follows from this lemma and Theorem 3.6, with the closed path $\gamma$ of Theorem 3.6 chosen to be $\gamma_1 + (-\gamma_2)$.

## IV. CONTOUR INTEGRATION IN MULTIDIMENSIONAL EVEN SUPERSPACE

In this short section the definitions and results of the previous section are extended to multidimensional even superspace $B_L^{m,0}$. Proofs are omitted, since they differ from the one-dimensional proofs in length rather than in principle. One slight change is that the domain of "paths" is standardized to the unit cube, as is customary when working in higher dimensions; in view of the "reparametrization" invariance of Lemma 3.9, this is a convenience rather than a loss of generality.

*Definition 4.1:* (This extends Definition 3.2.)

(a) Let $I^m$ denote the unit cube in $\mathbb{R}^m$. The mapping $\gamma$: $I^m\to B_L^{m,0}$ is called an $m$-path in $B_L^{m,0}$ if $\gamma$ is continuous and piecewise $C^1$. A formal sum of such paths is called an $m$-chain. If $\partial\gamma = 0$ (where the boundary operator $\partial$ is defined in the usual manner) then $\gamma$ is said to be closed.

(b) If $\gamma$ is an $m$-path in $B_L^{m,0}$, $U$ an open set in $B_L^{m,0}$ with $\gamma(I^m)\subset U$ and $f$: $U\to B_L$ is a continuous map of Banach spaces, then

$$\int_\gamma f\,dx^1\cdots dx^m := \int_{I^m} f(\gamma(t_1,\ldots,t_m)) \times \det \partial_j\,\gamma_i\,dt_1\cdots dt_m, \tag{4.1}$$

where $\partial_j\,\gamma_i$ denotes differentiation of the $i$th component of $\gamma$ with respect to $t_j$.

*Definition 4.2* [8]: (This extends Definition 3.3.) Let $U$ be open in $B_L^{m,0}$ and let $f$: $U\to B_L$. Then $f$ is said to be $G^1$ on $U$ if there exist $m$ functions $G_k f$: $U\to B_L$ $(k = 1, \ldots, m)$ and a function $\eta$: $B_L^{m,0}\to B_L$ such that, given $a$, $a + h$ in $U$,

$$f(a + h) = f(a) + \sum_{k=1}^{m} h^k G_k(a) + \|h\|\eta(h) \tag{4.2}$$

and $\|\eta(h)\|\to 0$ as $\|h\|\to 0$.

*Proposition 4.3:* (This extends Proposition 3.4.) Let $\gamma$: $I^m\to U$, where $U$ is open in $B_L^{m,0}$, be a $C^1$ map of Banach spaces. Let $f$: $U\to B_L$ be $G^1$. Then

$$\partial_i(h\circ\gamma) = \sum_{k=1}^{m} \partial_i\,\gamma^k G_k h. \tag{4.3}$$

**Theorem 4.4:** (This extends Theorem 3.5.) Let $U$ be open in $B_L^{m,0}$ and let $h$: $U\to B_L^{m,0}$ be an injective $G^1$ mapping. Also let $\gamma$: $[a,b]\to U$ be an $m$-path in $U$ and $f$: $U\to B_L$ be continuous. Then

$$\int_\gamma f\,dx^1\cdots dx^m = \int_{h\circ\gamma} f\circ h^{-1}\,\frac{1}{\det(G_k h^j)}\,dx^1\cdots dx^m, \tag{4.4}$$

where the $m$ functions $h^i$: $U\to B_L$ are defined by

$$h(x^1,\ldots,x^m) = (h^1(x^1,\ldots,x^m),\ldots,h^m(x^1,\ldots,x^m)). \tag{4.5}$$

**Theorem 4.5:** (This extends Theorem 3.6.) Let $\gamma$: $I^m\to B_L^{m,0}$ be a closed path in $B_{L0}$. Let $\rho$ be an $m + 1$ chain in $B_L^{m,0}$ whose boundary is $\gamma$. Also suppose $U$ is an open set in $B_L^{m,0}$ which contains $\rho(I^{m+1})$. Then if $f$: $U\to B_L$ is $G^1$ on $U$,

$$\int_\gamma f\,dx^1\cdots dx^m = 0. \tag{4.6}$$

*Lemma 4.6:* (This extends Lemma 3.8.) Let $f$: $U\to B_L$ where $U$ is open in $B_L^{m,0}$ and let $\iota$ denote the natural identification of $B_L^{m,0}$ and $\mathbb{R}^{m(2^{L-1})}$, that is,

$$\iota(x^1,\ldots,x^m) := (x^{1\mu},\ldots,x^{m\mu}), \tag{4.7}$$

if $x^k = \Sigma_{\mu\in M_0} x^{k\mu}\beta_\mu$ for $k = 1, \ldots, m$. Then

$$\partial_{k\mu}(f\circ\iota) = \beta_\mu(G_k f\circ\iota). \tag{4.8}$$

## V. CONSISTENT INTEGRATION IN ODD AND EVEN SUPERSPACE

In this section it is shown how the definition of contour integrals in even superspace may be combined with the usual Berezin definition for integration over odd variables to give a method of integration over the full odd and even superspace, together with a fully consistent rule for change of variable. In order to integrate over the $(m,n)$-dimensional superspace $B_L^{m,n}$, we first require a definition of an $(m,n)$ path.

*Definition 5.1:* An $(m,n)$ path in $B_L^{m,n}$ is a map $\gamma$: $I^m\times B_L^{0,n}\to B_L^{m,n}$ such that (a) letting $\gamma[\theta]$ denote the mapping of $I^m\to B_L^{m,0}$, defined by

$$\gamma[\theta](t) = \mathrm{Pr}_{(m,0)}\,\gamma(t^1,\ldots,t^m;\theta^1,\ldots,\theta^n) \tag{5.1}$$

[where $(t^1, \ldots, t^m)\in I^m$ and $(\theta^1, \ldots, \theta^n)\in B_L^{0,n}$], $\gamma[\theta]$ is an $m$-path in $B_L^{m,0}$ (Definition 4.1) for each $\theta$ in $B_L^{0,n}$; and (b) letting $\gamma[t]$ denote the mapping of $B_L^{0,n}$ into $B_L^{0,n}$, defined by

$$\gamma[t](\theta) = \mathrm{Pr}_{(0,n)}\,\gamma(t^1,\ldots,t^m;\theta^1,\ldots,\theta^n) \tag{5.2}$$

(where $\mathrm{Pr}_{(0,n)}$ denotes projection of $B_L^{m,n}$ onto $B_L^{0,n}$), $\gamma[t]$ is bijective and $G^n$ with continuous $n$th derivatives for each $t$ in

$I^m$. Also $\gamma$ is piecewise $C^1$ with respect to the $t^i$, and $G^n$ with respect to the $\theta^j$. (The definition of differentiation with respect to odd variables is similar to Definition 4.2 and given in Ref. 8.)

This definition now gives one something to integrate over which, being a mapping into $B_L^{m,n}$, has a natural way of transforming under change of coordinate in $B_L^{m,n}$.

The key definition of this paper is the definition of integration over an $(m,n)$ path. This definition will now be given, and then it will be shown that integrals are invariant under the appropriate transformation rule for change of variable. In the next section we show how the "quasicontour" approach described here (which is essential for consistency) relates to the standard approach.

*Definition 5.2:* Let $U$ be open in $B_L^{m,n}$ and let $F$: $U \to B_L$ be $G^n$, so that the Taylor expansion of $f$ in powers of $\theta$ exists, and suppose that the coefficients in this expansion are continuous. Also let $\gamma$ be an $(m,n)$ path in $\mathrm{Pr}_{(m,0)}(U) \times B_L^{0,n}$. Then the integral of $f$ over $\gamma$ is defined as

$$\int_\gamma f\, d^m x\, d^n\theta = \int_{I^m} f(\gamma(t,\theta))B(\gamma)(t,\theta)\, d^m t\, d^n\theta, \tag{5.3}$$

where the integral with respect to $t$ is a standard Riemannian integral while the $\theta$ integration is done in the usual manner; $B(\gamma)$ is the superdeterminant of the $(m+n)\times(m+n)$ matrix $(M_{ij})$ with

$$\begin{aligned} M_{ij} &= \frac{\partial \gamma^i}{\partial t^j}, \quad 1 \leqslant i,j \leqslant m, \\ M_{m+i,j} &= \frac{\partial \gamma^{m+i}}{\partial t^j}, \quad 1 \leqslant i \leqslant n, \quad 1 \leqslant j \leqslant m, \\ M_{i,m+j} &= G_{m+j}\gamma^i, \quad 1 \leqslant i \leqslant m, \quad 1 \leqslant j \leqslant n, \\ M_{m+i,m+j} &= G_{m+j}\gamma^{m+i}, \quad 1 \leqslant i,j \leqslant n \end{aligned} \tag{5.4}$$

(where $G_{m+j}$ is the super derivative with respect to $\theta^j$).

It is of course necessary to prove that this definition has the correct behavior under change of variable. This requires two theorems: first, Theorem 5.3, which proves the transformation rule under change of coordinates in superspace (this is a very straightforward consequence of Definition 5.2) and, second, Theorem 5.4, which establishes the reparametrization invariance of the integrals.

**Theorem 5.3:** Let $U$ be open in $B_L^{m,n}$ and let $h$: $U \to B_L^{m,n}$ be injective and $G^n$ (with continuous $n$th derivatives). Also let $\gamma$ and $f$ be as in Definition 5.2. Then

$$\int_\gamma f\, d^m x\, d^n\theta = \int_{h\circ\gamma} f \circ h^{-1} J(h)\, d^m x\, d^n\theta, \tag{5.5}$$

where, if $h(x,\theta) = (h^1(x,\theta),\ldots,h^m(x,\theta); h^{m+1}(x,\theta), \ldots, h^{m+n}(x,\theta))$, $J(h)$ is the superdeterminant of the inverse of the matrix $(G_k h^i)$.

*Proof:* The following "chain rules" may easily be established:

$$\frac{\partial(h\circ\gamma)^i}{\partial t^j} = \sum_{k=1}^{m+n} (G_k h^i)\circ\gamma \times \frac{\partial \gamma^k}{\partial t^j}, \quad i = 1,\ldots,m+n, \quad j = 1,\ldots m, \tag{5.6}$$

$$G_{m+j}(h\circ\gamma)^i = \sum_{k=1}^{m+n} (G_k h^i)\circ\gamma \times G_{m+j}\gamma^k. \tag{5.7}$$

Thus

$$\begin{aligned} &\int_{h\circ\gamma} f\circ h^{-1} J(h)\, d^m x\, d^n\theta \\ &= \int_{I^m} f(\gamma(t,\theta))J(h)(\gamma(t,\theta))B(h\circ\gamma)(t,\theta)\, d^m t\, d^n\theta \\ &= \int_{I^m} f(\gamma(t,\theta))B(\gamma)(t,\theta)\, d^m t\, d^n\theta \end{aligned} \tag{5.8}$$

[using the chain rules (5.5) and (5.6)],

and the result is proved.

The second theorem in this section contains the real substance, because it is the reparametrization invariance which allows one to think of an integral as being over a specific subset of superspace—the image of the path $\gamma$—rather than over the path.

**Theorem 5.4:** Let $U$ be open in $B_L^{m,n}$ and let $A$ be a compact set in $\mathbb{R}^m$. Also let $\gamma$: $A \times B_L^{0,n} \to \mathrm{Pr}_{(m,0)}(U) \times B_L^{0,n}$ be an $(m,n)$ path. [Note that this is a slight generalization of the $(m,n)$ path of Definition 5.2. In general it is simpler to standardize the domain of $\gamma$; the present theorem shows this involves no loss of generality.] Now suppose $f$: $U \to B_L$ is as in Definition 5.2, and that $B$ is a compact subset of $\mathbb{R}^m$ and that $\alpha$: $B \times B_L^{0,n} \to A \times B_L^{0,n}$ satisfies the following conditions: $\alpha$ is $C^1$ with respect to the real variables, and $G^n$ with respect to the odd variables (with continuous $n$th derivatives). Also the Berezin determinant $B(\alpha)$ [defined as in Eq. (5.4), except that in this case $G_{m+j}\alpha^i$ must be zero for $1 \leqslant j \leqslant n$, $1 \leqslant i \leqslant M$] must be positive. Then

$$\int_\gamma f\, d^m x\, d^n\theta = \int_{\gamma\circ\alpha} f\, d^m x\, d^n\theta. \tag{5.9}$$

*Proof:* Following the idea of Fung (quoted in Ref. 11) we may decompose $\alpha$ into $\alpha = \alpha_1 \circ \alpha_2$, where

$$\alpha_2(t;\theta) = (\alpha^1(t),\ldots,\alpha^m(t);\theta^1,\ldots,\theta^n) \tag{5.10}$$

and

$$\alpha_1(t;\theta) = (t^1,\ldots,t^m;\alpha^{m+1}(\alpha_2^{-1}(t;\theta)),\ldots,\alpha^{m+n}(\alpha_2^{-1}(t;\theta))). \tag{5.11}$$

Then Eq. (5.9) is easily seen to hold for both types of reparametrization. For $\alpha_2$, it is simply the $m$-dimensional version of Lemma 3.9(a) while for $\alpha_1$ it is a standard result of Berezin integration.[6,9]

As well as giving a definition of an integral on a subset of superspace, this reparametrization invariance also allows one to patch together integrals to give integrals on supermanifolds; there are restrictions on the type of supermanifold, because the domain and range of $\gamma$ must include all of $B_L^{0,n}$. The author is investigating the possibility of reducing these restrictions, and also improving on the somewhat hybrid nature of an $(m,n)$ path.

## VI. APPLICATION TO SUPERSYMMETRY

Superspace techniques are frequently used in supersymmetric quantum field theories[16]; superspace extends real space-time (Minkowski space or a more general manifold) by adding odd dimensions, and making the space-time dimension even Grassmann rather than simply real. Supersym-

metry transformations are represented by translations or general coordinate transformations in superspace; the usual techniques of differential geometry are applied; in particular, supersymmetric invariants are constructed as integrals over superspace. Obviously, if an integral really is to lead to a supersymmetry invariant, the integral must be invariant under change of coordinates, and it has been shown in this paper that this is true for the "quasicontour" method of Sec. V, but not for the standard approach described in Sec. I. The space-time theory is recovered from the superspace version by using the augmentation map $\epsilon_{m,n}$ (Definition 2.1) to project our space-time; it is easy to combine the standard method of superspace integration with this projection to obtain from a superspace integral an integral over space-time; one simply carries out the $\theta$ integration. Because the standard method does not have full coordinate invariance, the space-time integrals obtained by carrying out the $\theta$ integration are not guaranteed to be fully invariant under supersymmetry transformations; when explicit calculations are made of the variations of such "invariants" under infinitesimal supersymmetry transformations, the variation is found to be equal to a surface integral which becomes zero if the usual sort of "dying away at large distances" conditions are put on the fields. In fact, one can use the quasicontour approach together with the generalized Cauchy theorem (Theorem 4.5), to work out the possible departure from true supersymmetry invariance of any space-time integral derived from a superspace integral, and thus to work out what boundary conditions must be placed on the fields if the integral is to be invariant under both infinitesimal and finite supersymmetry transformations. To begin with, one must define a method for obtaining a space-time integral from a quasicontour integral. To make contact with the standard approach one naturally derives from an $(m,n)$ path $\gamma$ and a function $f: B_L^{m,n} \to B_L$ the space-time integral

$$\int_{\epsilon_{m,n} \circ \gamma} f^n \Big|_{j_{m,n} \circ \epsilon_{m,n} \circ \gamma(t^m)} d^m t, \tag{6.1}$$

where $f^n$ is the coefficient of $\theta^1 \cdots \theta^n$ in the $\theta$ expansion of $f$ and the integration is carried out in the usual (Riemannian) way, and $j_{m,n}$ denotes the natural injection of $\mathbf{R}^m$ into $B_L^{m,n}$; if $\gamma$ satisfies

$$\gamma = j_{m,n} \circ \epsilon_{m,n} \circ \gamma, \tag{6.2}$$

then

$$\int_\gamma f\, d^m x\, d^n\theta = \int_{\epsilon_{m,n} \circ \gamma} f^n(\epsilon_m(x)) d^m x. \tag{6.3}$$

However if $h$ gives a coordinate transformation of $B_L^{m,n}$ (that is, $h: B_L^{m,n} \to B_L^{m,n}$ is bijective and $G^\infty$) then the path $h \circ \gamma$ will not necessarily satisfy (6.2) even if $\gamma$ does, and so the invariance of a quasicontour integral under transformations of superspace cannot, of course, guarantee the invariance of the associated space-time integral (6.1). The possible discrepancy is

$$\int_\gamma f\, d^m x\, d^n\theta - \int_\beta f\, d^m x\, d^n\theta, \tag{6.4}$$

where $\beta$ is any $m$, $n$ path, which satisfies

$$\epsilon_{m,n} \circ \beta = \epsilon_{m,n} \circ \gamma. \tag{6.5}$$

Now suppose $\sigma$ is an $(m,n)$ chain in $B_L^{m,n}$ such that $\gamma[\theta] - \beta[\theta] - \sigma[\theta]$ is closed for each $\theta$ in $B_L^{0,n}$. Then, by Theorem 4.5, if $f$ is sufficiently well-behaved,

$$\int_\gamma f\, d^m x\, d^n\theta - \int_\beta f\, d^m x\, d^n\theta = \int_\sigma f\, d^m x\, d^n\theta, \tag{6.6}$$

and so the maximum breakdown in supersymmetry invariance of the space-time integral (6.1) is $\int_\sigma f\, d^m x\, d^n\theta$. To see the nature of this quantity, we consider the simple case $m = 1$; then, if $\gamma, \beta$ satisfy (6.5) a possible choice of $(1,n)$ path $\sigma$ such that $\gamma[\theta] - \beta[\theta] - \sigma[\theta]$ is closed is

$$\begin{aligned} \sigma &= \sigma_0 + \sigma_1, \\ \text{where} \quad \sigma_0(t,\theta) &= \gamma(0,\theta) + t(\beta(0,\theta) - \gamma(0,\theta)), \\ \sigma_1(t,\theta) &= \beta(1,\theta) + t(\gamma(1,\theta) - \beta(1,\theta)). \end{aligned} \tag{6.7}$$

Thus

$$\begin{aligned} &\int_\sigma f\, d^m x\, d^n\theta \\ &\quad = \int \Big( \int_0^1 f(\gamma(0,\theta) + t(\beta(0,\theta) - \gamma(0,\theta))) \\ &\qquad \times (\beta(0,\theta) - \gamma(0,\theta)) - f(\beta(1,\theta) + t(\gamma(1,\theta) \\ &\qquad - \beta(1,\theta)))(\gamma(1,\theta) - \beta(1,\theta)) dt \Big) d^n\theta. \end{aligned} \tag{6.8}$$

Now, if $g$ is a $G^\infty$ function of $U \subset B_L^{m,n}$ into $B_L$,

$$g(x,\theta) = \sum_{\mu \in M_n} g^\mu(x)\theta_\mu,$$

and

$$g^\mu(x) = g^\mu(\epsilon(x)) + \sum_{i=1}^m (x^i - \epsilon(x^i))\partial_i g^\mu(\epsilon(x)) + \cdots \tag{6.9}$$

(cf. Ref. 8, Corollary 2.9, which includes a form for the remainder). Thus one sees that a sufficient condition on the field $f$ for the space-time integral (6.1) to be invariant under supersymmetry transformations is that $f$ and all its derivatives vanish on the boundary of $\gamma$. This is a slightly weaker condition than requiring $f$ to be of compact support; it also extends the result for infinitesimal transformations (familiar to anyone who has worked with supersymmetry transformations in component form, where supersymmetric invariants are always invariant only up to a surface integral) to finite transformations. Moreover, in any given situation one can calculate precisely the extent to which using the standard approach as opposed to the "quasicontour" approach affects the invariance of a superspace integral under coordinate transformations.

Although it might seem that by making quite reasonable assumptions on fields, the standard approach (relating easily to space-time integrals) can be used instead of the quasicontour approach, one must remember that quantization may involve very singular field configurations; a simple example (such as those in Sec. II) shows that for a discontinuous function the invariance may break down in a serious manner. However, although this might lead to some questioning of the status of the space-time Lagrangian, the quan-

tization can be carried out in superspace[17] where the approach developed in this paper does lead to full supersymmetry invariance.

## VII. CONCLUSION AND FURTHER POSSIBILITIES

In this paper there has been presented for the first time a fully consistent method of integration in superspace, pulling together several ideas in a new way. The method relies on the geometric approach to superspace, and the use of the finer of the two topologies which can be used on superspace. In many cases it gives the same result as the standard method; it also makes clear the cases where the standard method breaks down, and is of much wider applicability—for instance, it can be applied to a superspace with a boundary. A consistent integration method is clearly essential for putting the various applications of superspace integration in quantum field theory on sound foundations. It is also important for the extension to superspace of many standard techniques in differential geometry. One development which immediately springs to mind is a method for integration on supermanifolds; supermanifolds are made by patching together bits of $B_L^{m,n}$ much as conventional manifolds are made by patching together $\mathbf{R}^m$ or $C^m$. Bernstein and Leîtes[3,4] and Berezin[5] have described how, using "superforms" and superspace integration one can define integrals on supermanifolds; this approach becomes fully consistent when combined with the method of integration on $B_L^{m,n}$ defined in this paper. Berezin's method applies only to a certain class of supermanifold—basically, those where there is no patching in the $\theta$ directions. It is an open question whether or not a good definition of integration can be given for the wider class of supermanifolds that has been considered.[8]

One area of conventional differential geometry which has found wide application in theoretical physics is the theory of characteristic classes and their integral representations; a start on developing a similar theory for supermanifolds was made by Berezin[5]; it should be possible to make further progress using the fully consistent method of integration developed in this paper.

It should finally be added that the author feels that the hybrid theory presented here is a step on the way to a more complete understanding of odd and even integration.


## ACKNOWLEDGMENT

This research was supported by the Science and Engineering Research Council of Great Britain (S. E. R. C.).



[1]F. A. Berezin, *The Method of Second Quantisation* (Academic, New York, 1966).
[2]F. A. Berezin and D. A. Leîtes, Sov. Math. Dokl. **16**, 1218 (1975).
[3]I. N. Bernstein and D. A. Leîtes, Func. Anal. Appl. **11**, 45 (1977).
[4]I. N. Bernstein and D. A. Leîtes, Func. Anal. Appl. **11**, 219 (1977).
[5]F. A. Berezin, Sov. J. Nucl. Phys. **30**, 605 (1979).
[6]D. A. Leîtes, Russ. Math. Surveys **35:1**, 1 (1980).
[7]B. S. De Witt, *Supermanifolds* (Cambridge U. P., Cambridge, 1984).
[8]A. Rogers, J. Math. Phys. **21**, 1352 (1980).
[9]V. F. Pakhomov, Mat. Zametki **16**, 65 (1974).
[10]T. Regge, "Relativity, Groups and Topology," lectures given at the 1983 Les Houches Summer School.
[11]P. Van Nieuwenhuizen, Phys. Rep. **68**, 189 (1981).
[12]J. Hassoun, A. Restuccia, and J. G. Taylor, "Superfield Actions for $N = 2$ degenerate central charges," King's College London Preprint, June 1983.
[13]B. Kostant, "Graded Manifolds, Graded Lie Theory and Prequantisation," in *Lecture Notes in Mathematics*, Vol. 570 (Springer, Berlin, 1977).
[14]M. Batchelor, Trans. Am. Math. Soc. **258**, 257 (1980).
[15]J. Dieudonné, *Foundations of Modern Analysis* (Academic, New York, 1969).
[16]D. V. Volkov and V. P. Akulov, Phys. Lett. B **46**, 109 (1973); Abdus Salam and J. Strathdee, Nucl. Phys. B **76**, 477 (1974); J. Wess and B. Zumino, Phys. Lett. B **66**, 361 (1977).
[17]Superfield quantization techniques are described in S. J. Gates, M. T. Grisaru, M. Rocek, and W. Siegel, *Superspace or One Thousand and One Lessons in Supersymmetry* (Benjamin, New York, 1983).

# Application of linked Bäcklund transformations to nonlinear boundary value problems

C. Rogers
*Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada*

(Received 26 June 1984; accepted for publication 21 September 1984)

A class of nonlinear boundary value problems is reduced to linear canonical form by a combination of Bäcklund transformations.

## I. INTRODUCTION

The application of Bäcklund transformations to the solution of nonlinear boundary value problems is less well-developed than their role in the solution of privileged initial value problems via the inverse scattering transform. However, in a recent paper by Fokas and Yortsos,[1] application was made of coupled Bäcklund transformations to solve a nonlinear boundary value problem involving two-phase flow in an unbounded porous medium. Further developments in the solution of nonlinear boundary value problems via Bäcklund transformations are given in Refs. 2–4. Here, a class of nonlinear boundary value problems on a slab is linearized by combination of a reciprocal Bäcklund transformation and a version of the Cole–Hopf transformation applied to a Burgers' hierarchy.

## II. A CLASS OF RECIPROCAL TRANSFORMATIONS

Conservation laws of the form

$$\frac{\partial}{\partial t}\left\{T\left(\frac{\partial}{\partial x};\frac{\partial}{\partial t};u\right)\right\}+\frac{\partial}{\partial x}\left\{F\left(\frac{\partial}{\partial x};\frac{\partial}{\partial t};u\right)\right\}=0 \tag{1}$$

are considered where

$$T\left(\frac{\partial}{\partial x};\frac{\partial}{\partial t};u\right) := T(u, u_x, u_{xx}, \ldots; u_t, u_{tt}, \ldots), \tag{2}$$

$$F\left(\frac{\partial}{\partial x};\frac{\partial}{\partial t};u\right) := F(u, u_x, u_{xx}, \ldots; u_t, u_{tt}, \ldots). \tag{3}$$

The transformation $R$ is introduced according to

$$\left.\begin{aligned} dx^* &= [aT+b]dx-[aF+c]dt,\\ t^* &= \mathrm{et}+\mathrm{h(u)},\\ u^* &= 1/u, \end{aligned}\right\} \quad R \tag{4}$$

where $a, b, c, e \in \mathbb{R}$ and $a, e$ are nonzero.

It will be required that $R$ be involutory so that $R^2 = I$, whence

$$dx^{**} = dx, \tag{5}$$

$$t^{**} = t, \tag{6}$$

$$u^{**} = u. \tag{7}$$

Thus,

$$\begin{aligned} dx^{**} &= [aT^*+b]dx^*-[aF^*+c]dt^*\\ &= [aT^*+b]\{[aT+b]dx-[aF+c]dt\}\\ &\quad -[aF^*+c]\{edt+h'u_x\,dx+h'u_t\,dt\}\\ &= [(aT^*+b)(aT+b)-(aF^*+c)h'u_x]dx\\ &\quad -[(aT^*+b)(aF+c)+(aF^*+c)\\ &\quad \times(e+h'u_t)]dt = dx \end{aligned}$$

requires that

$$(aT^*+b)(aT+b)-(aF^*+c)h'u_x = 1,$$

$$(aT^*+b)(aF+c)+(aF^*+c)(e+h'u_t) = 0.$$

Hence, $T^*$ and $F^*$ are given by

$$aT^*+b = (e+h'u_t)/\Delta, \tag{8}$$

$$aF^*+c = -(aF+c)/\Delta, \tag{9}$$

where

$$\Delta := J(x^*, t^*; x, t) = (aT+b)(e+h'u_t)+(aF+c)h'u_x. \tag{10}$$

Condition (6) shows that

$$t^{**} = et^*+h(u^*) = e^2t+eh(u)+h(u^{-1}) = t,$$

whence

$$e^2 = 1, \tag{11}$$

$$eh(u)+h(u^{-1}) = 0. \tag{12}$$

Accordingly, either

$$e = +1, \quad h(u) = \Psi(\ln|u|), \quad \Psi \text{ odd}, \tag{13}$$

or

$$e = -1, \quad h(u) = \Theta(\ln|u|), \quad \Theta \text{ even}. \tag{14}$$

Furthermore, it is noted that

$$u^{**} = 1/u^* = u, \tag{15}$$

so that the last reciprocal condition (7) is met.

Thus, we obtain the following result.

**Theorem:** The conservation law

$$\frac{\partial T}{\partial t}+\frac{\partial F}{\partial x} = 0 \tag{16}$$

is transformed to the *reciprocally associated* conservation law

$$\frac{\partial T^*}{\partial t^*}+\frac{\partial F^*}{\partial x^*} = 0 \tag{17}$$

under the reciprocal transformation given by (4), where $h(u)$ is subject to conditions (13) and (14) and $T^*, F^*$ are given by (8)–(10).

The above result is an extension of the reciprocal result recently presented by Kingston and Rogers.[5] This has been applied to inverse scattering schemes by Rogers and Wong.[6] Here, reciprocal Bäcklund transformations are used in con-

 0022-2488/85/030393-03$02.50 © 1985 American Institute of Physics 

junction with a Cole–Hopf-type transformation to reduce a class of *nonlinear* boundary value problems to linear *canonical* form.

## III. THE NONLINEAR BOUNDARY VALUE PROBLEMS: APPLICATION OF A RECIPROCAL TRANSFORMATION

The class of nonlinear boundary value problems to be considered is given by

$$\frac{\partial u}{\partial t}+\frac{\partial}{\partial x}\left[u\sum_{i=1}^{N}\alpha_i(t)\Phi_i\right]=0,\quad 0<x<L,\quad t>0, \tag{18}$$

$$u\sum_{i=1}^{N}\alpha_i(t)\Phi_i=\Psi_1(t)\quad \text{at } x=0,\quad t>0, \tag{19}$$

$$u\sum_{i=1}^{N}\alpha_i(t)\Phi_i=\Psi_2(t)\quad \text{at } x=L,\quad t>0, \tag{20}$$

$$u=\bar{u}(x)\quad \text{at } t=0,\quad 0\leqslant x\leqslant L, \tag{21}$$

where the $\Phi_i$ are defined recursively according to

$$u\Phi_i=\Phi_{i-1}+\frac{\partial\Phi_{i-1}}{\partial x},\quad i=1,2,\ldots,N,\qquad \Phi_0=1. \tag{22}$$

It is noted that nonlinear boundary conditions (19) and (20) correspond to prescribed flux at $x=0$ and $x=L$, respectively.

Under the reciprocal transformation,

$$dx'=u\,dx-u\sum_{i=1}^{N}\alpha_i(t)\Phi_i\,dt,\quad t'=t,\qquad u'=1/u, \tag{23}$$

Eq. (18) becomes

$$\frac{\partial u'}{\partial t'}-\frac{\partial}{\partial x'}\left[\sum_{i=1}^{N}\alpha_i(t')\Phi_i'\right]=0, \tag{24}$$

where the $\Phi_i'$ are given recursively by

$$\Phi_i'=u'\Phi_{i-1}'+\frac{\partial\Phi_{i-1}'}{\partial x'},\quad i=1,2,\ldots,N,\qquad \Phi_0'=1. \tag{25}$$

Furthermore, (23) shows that

$$\frac{\partial x'}{\partial x}=u,$$

$$\frac{\partial x'}{\partial t}=-u\sum_{i=1}^{N}\alpha_i(t)\Phi_i=-\int_0^x\frac{\partial}{\partial x}\left[u\sum_{i=1}^{N}\alpha_i(t)\Phi_i\right]dx-\Psi_1(t)=\int_0^x\frac{\partial u}{\partial t}\,dx-\Psi_1(t),$$

whence

$$x'(x,t)=\int_0^x u\,dx+\Theta_1(0)-\Theta_1(t), \tag{26}$$

where $\dot{\Theta}_1:=\Psi_1$, and we have taken $x'(0,0)=0$.

Thus, the boundary condition (19) becomes

$$\sum_{i=1}^{N}\alpha_i(t')\Phi_i'=u'\Psi_1(t')\quad \text{at } x'=\Theta_1(0)-\Theta_1(t'),\quad t'>0, \tag{27}$$

under the reciprocal transformation (23).

The flux conditions at $x=0$ and $x=L$ show that

$$\left[u\sum_{i=1}^{N}\alpha_i(t)\Phi_i\right]_0^L=\int_0^L\frac{\partial}{\partial x}\left[u\sum_{i=1}^{N}\alpha_i(t)\Phi_i\right]dx=-\int_0^L\frac{\partial u}{\partial t}\,dx=-\frac{\partial}{\partial t}\int_0^L u\,dx=\Psi_2(t)-\Psi_1(t),$$

whence, on use of the initial condition (21),

$$\int_0^L u\,dx=\Theta_1(t)-\Theta_2(t)+\Theta_2(0)-\Theta_1(0)+\int_0^L\bar{u}\,dx, \tag{28}$$

where $\dot{\Theta}_2:=\Psi_2$.

Accordingly, the boundary condition (20) becomes, by virtue of (26) and (28),

$$\sum_{i=1}^{N}\alpha_i(t')\Phi_i'=u'\Psi_2'(t')\quad \text{at } x'=\Theta_2(0)-\Theta_2(t')+\int_0^L\bar{u}\,dx,\quad t'>0. \tag{29}$$

Thus, to summarize, the reciprocal boundary value problem is

$$\frac{\partial u'}{\partial t'}-\frac{\partial}{\partial x'}\left[\sum_{i=1}^{N}\alpha_i(t')\Phi_i'\right]=0,$$

$$\sum_{i=1}^{N}\alpha_i(t')\Phi_i'=u'\Psi_1(t')\quad \text{at } x'=\Theta_1(0)-\Theta_1(t'),\quad t'>0,$$

$$\sum_{i=1}^{N}\alpha_i(t')\Phi_i'=u'\Psi_2(t')\quad \text{at } x'=\Theta_2(0)-\Theta_2(t')+\int_0^L\bar{u}(\sigma)d\sigma,\quad t'>0, \tag{30}$$

$$u'=U'(x')\quad \text{at } t'=0,$$

where

$$x'|_{t=0}=\int_0^x\bar{u}(\sigma)d\sigma:=U(x), \tag{31}$$

$$U'(x')=1/\bar{u}(U^{-1}(x')), \tag{32}$$

and the $\Phi_i'$ are given by (25) and generate a Burgers' hierarchy.

## IV. REDUCTION OF THE BURGERS' HIERARCHY

We now introduce the Bäcklund transformation

$$u^*_{x^*}=u'u^*,\qquad u^*_{t^*}=\left(\sum_{i=1}^{N}\alpha_i\Phi_i'\right)u^*,\qquad x^*=x',\quad t^*=t'. \tag{33}$$

Under this transformation it is readily shown that

$$\Phi_i'=\frac{1}{u^*}\frac{\partial^i u^*}{\partial x^{*i}}, \tag{34}$$

and the nonlinear evolution equation

$$\frac{\partial u'}{\partial t'}-\frac{\partial}{\partial x'}\left[\sum_{i=1}^{N}\alpha_i(t')\Phi_i'\right]=0, \tag{35}$$

becomes

$$\frac{\partial u^*}{\partial t^*} = \sum_{i=1}^{N} \alpha_i(t^*) \frac{\partial^i u^*}{\partial x^{*i}} + T(t^*)u^*, \tag{36}$$

where $T(t^*)$ is arbitrary. Introduction of $\tilde{u}(x^*, t^*)$ according to

$$\tilde{u} = \exp\left\{ -\int_0^{t^*} T(\sigma)d\sigma \right\} u^*, \tag{37}$$

reduces (36) to

$$\frac{\partial \tilde{u}}{\partial t^*} = \sum_{i=1}^{N} \alpha_i(t^*) \frac{\partial^i \tilde{u}}{\partial x^{*i}}, \tag{38}$$

and (33) together with (37) reduce the nonlinear boundary value problem to the *linear* canonical form

$$\begin{aligned}
&\frac{\partial \tilde{u}}{\partial t^*} = \sum_{i=1}^{N} \alpha_i(t^*) \frac{\partial^i \tilde{u}}{\partial x^{*i}}, \quad t^* > 0, \\
&\frac{\partial \tilde{u}}{\partial t^*} - \Psi_1(t^*) \frac{\partial \tilde{u}}{\partial x^*} = 0 \\
&\quad \text{at } x^* = \Theta_1(0) - \Theta_1(t^*), \quad t^* > 0, \\
&\frac{\partial \tilde{u}}{\partial t^*} - \Psi_2(t^*) \frac{\partial \tilde{u}}{\partial x^*} = 0 \\
&\quad \text{at } x^* = \Theta_2(0) - \Theta_2(t^*) + \int_0^L \tilde{u}(\sigma)d\sigma, \\
&\tilde{u} = \exp\left\{ \int_{x_0^*}^{x^*} U'(\tau)d\tau \right\} \quad \text{at } t^* = 0.
\end{aligned} \tag{39}$$

Thus, it is seen that the class of nonlinear boundary value problems defined by (18)–(22) may be reduced to linear canonical form by combination of a reciprocal transformation and a Bäcklund transformation appropriate to the reduction of a Burgers' hierarchy. The result obtained in Ref. 1 may be retrieved as a special case of the above when reduction is to a linear boundary value problem for the classical $1 + 1$ heat equation. Application was made in that case to two-phase flow in a porous medium.


## ACKNOWLEDGMENT

Support under Natural Sciences and Engineering Research Council of Canada Grant No. A0879 is gratefully acknowledged.



[1] A. S. Fokas and Y. C. Yortsos, Soc. Ind. Appl. Math. J. Appl. Math. **42**, 318 (1982).
[2] C. Rogers, J. Phys. A: Math. Gen. **16**, L493 (1983).
[3] C. Rogers, H. Rasmussen, and M. P. Stallybrass, Int. J. Multiphase Flow **10**, 95 (1984).
[4] C. Rogers, M. P. Stallybrass, and D. L. Clements, Nonlinear Anal. Theory, Meth. Appl. **7**, 785 (1983).
[5] J. G. Kingston and C. Rogers, Phys. Lett. A **92**, 261 (1982).
[6] C. Rogers and P. Wong, Phys. Scr. **30**, 10 (1984).

# Some properties of hyperspherical harmonics

Zhen-Yi Wen[a] and John Avery
*Department of Physical Chemistry, H. C. Ørsted Institute, University of Copenhagen, Copenhagen 0, Denmark*

(Received 23 September 1983; accepted for publication 27 July 1984)

A general formula is given for the canonical decomposition of a homogeneous polynomial of order $\lambda$ in $m$ variables into a sum of harmonic polynomials. This formula, which involves successive applications of the generalized Laplace operator, is proved in the Appendix. It is shown that the group-theoretical method for constructing irreducible Cartesian tensors follows from the general formula for canonical decomposition. The relationship between harmonic polynomials and hyperspherical harmonics is discussed, and an addition theorem for hyperspherical harmonics is derived. An expansion of a many-dimensional plane wave in terms of Gegenbauer polynomials and Bessel functions is derived and used to construct bicenter expansions of arbitrary functions in many-dimensional spaces. Finally, a formula is derived for the $3\lambda$ coefficients of hyperspherical harmonics. These coefficients give the values of integrals involving the products of three harmonics.

## I. INTRODUCTION

During the last few years, the application of the hyperspherical expansion method to the quantum mechanical many-body problem has attracted a great deal of attention. In nuclear physics, the method has been developed to a high degree, mainly by Soviet authors.[1,2] In atomic physics hyperspherical analysis of two- or three-electron atoms has led to a new insight into electron correlation.[3–13] For more complex atoms, some impressive qualitative conclusions can be drawn from the hyperspherical coordinate method.[14] The calculations involved are extremely complicated,[15] but an interesting suggestion for simplifying them has been proposed.[16] Hyperspherical analysis has also been applied to molecular problems, such as the treatment of large-amplitude vibrations and reaction coordinates.[17,18]

The increasing interest in hyperspherical harmonics in theoretical physics and chemistry makes it worthwhile to undertake a mathematical investigation of some of their properties. Although many books and papers on this subject are available,[19–21] we believe that some results still require clarification. The aim of this paper is to clarify some of the concepts underlying the theory of hyperspherical harmonics and to explore relations among them. Many of the questions discussed in this paper are well known, but others are perhaps new.

## II. HARMONIC POLYNOMIALS AND HYPERSPHERICAL HARMONICS

Let us begin by recalling that a homogeneous polynomial which satisfies the generalized Laplace equation is called an harmonic polynomial. If we consider an $m$-dimensional space with Cartesian coordinates $x_1, x_2, \ldots, x_m$, then the generalized Laplace operator in this space is

$$\Delta \equiv \sum_{j=1}^{m} \frac{\partial^2}{\partial x_j^2}. \tag{1}$$

The following general formula allows us to express a homogeneous polynomial in terms of harmonic polynomials (this is called a canonical decomposition of the homogeneous polynomial): Let $f_\lambda(\mathbf{x})$ be a homogeneous polynomial of order $\lambda$ in the coordinates $x_1, x_2, \ldots, x_m$. Then

$$f_\lambda(\mathbf{x}) = \sum_{k=0}^{[\lambda/2]} r^{2k} h_{\lambda-2k}(\mathbf{x}), \tag{2}$$

where

$$h_{\lambda-2k}(\mathbf{x}) = \frac{(m+2\lambda-4k-2)}{2^k k!(m+2\lambda-2k-2)!!} \times \sum_{t=0}^{[(\lambda-2k)/2]} \frac{(-1)^t (m+2\lambda-4k-2t-4)!!}{2^t t!} \times r^{2t} \Delta^{k+t} f_\lambda(\mathbf{x}). \tag{3}$$

In Eq. (3), $\Delta^{k+t}$ is the generalized Laplace operator [Eq. (1)] applied $k+t$ times, and $r$ is the hyperradius defined by

$$r^2 = x_1^2 + x_2^2 + \cdots + x_m^2. \tag{4}$$

A special case of Eq. (3) ($k=0$) appears in Vilenkin's book.[20] A proof of the general formula is given in the Appendix. As a simple example of the canonical decomposition of a homogeneous polynomial by means of Eqs. (2) and (3), we might consider the case where $\lambda = 3$ and

$$f_\lambda(\mathbf{x}) = f_3(\mathbf{x}) = x_1^2 x_2. \tag{5}$$

Then Eq. (2) becomes

$$x_1^2 x_2 = h_3(\mathbf{x}) + r^2 h_1(\mathbf{x}), \tag{6}$$

where, from Eq. (3),

$$h_3(\mathbf{x}) = x_1^2 x_2 - r^2 x_2/(m+2) \tag{7}$$

and

$$h_1(x) = x_2/(m+2). \tag{8}$$

A number of harmonic polynomials are listed in Ref. 22, in which the polynomials are called $\lambda$-projections and derived by angular integration. If an harmonic polynomial is considered to be a component of a tensor, the tensor will be irreducible with respect to the $m$-dimensional rotation group

[a] Permanent address: Mining Institute of Xi'an, Xi'an, Shaanxi, The People's Republic of China.

 0022-2488/85/030396-08$02.50 © 1985 American Institute of Physics 

So($m$). Just as we do in three-dimensional space,[23] we may call the hyperspherical polynomial of the highest order $[h_\lambda(x)]$ an irreducible Cartesian tensor. It is a symmetrical and traceless tensor of order $\lambda$. There is a general method in Hamermesh's book[24] for constructing such a tensor. Let $\tilde{\Sigma}$ be an operator which symmetrizes the indices $i_1,i_2,\ldots,i_\lambda$ keeping only distinct permutations of the indices. Then if $\lambda = 2t$, the symmetrical traceless tensor can be written as

$$\begin{aligned} &m(m+2)\cdots(m+4t-4)(x_{i_1}x_{i_2}\cdots x_{i_{2t}}) \\ &\quad - m(m+2)\cdots(m+4t-6)r^2\tilde{\Sigma}(\delta_{i_1i_2}x_{i_3}\cdots x_{i_{2t}}) \\ &\quad + \cdots + (-1)^{t-1}m(m+2)\cdots(m+2t-2)r^{2t-2} \\ &\quad \times\tilde{\Sigma}(\delta_{i_1i_2}\cdots\delta_{i_{2t-3}i_{2t-2}}x_{i_{2t-1}}x_{i_{2t}}) \\ &\quad + (-1)^t m(m+2)\cdots(m+2t-4)r^{2t} \\ &\quad \times\tilde{\Sigma}(\delta_{i_1i_2}\cdots\delta_{i_{2t-1}i_{2t}}). \end{aligned} \tag{9a}$$

If $\lambda = 2t+1$, the symmetrical traceless tensor of order $\lambda$ becomes

$$\begin{aligned} &m(m+2)\cdots(m+4t-2)(x_{i_1}\cdots x_{i_{2t+1}}) \\ &\quad - m(m+2)\cdots(m+4t-4)r^2\tilde{\Sigma}(\delta_{i_1i_2}x_{i_3}\cdots x_{i_{2t+1}}) \\ &\quad + \cdots + (-1)^{t-1}m(m+2)\cdots(m+2t)r^{2t-2} \\ &\quad \times\tilde{\Sigma}(\delta_{i_1i_2}\cdots\delta_{i_{2t-2}i_{2t-1}}x_{i_{2t}}x_{i_{2t+1}}) \\ &\quad + (-1)^t m(m+2)\cdots(m+2t-2)r^{2t} \\ &\quad \times\tilde{\Sigma}(\delta_{i_1i_2}\cdots\delta_{i_{2t-1}i_{2t}}x_{i_{2t+1}}). \end{aligned} \tag{9b}$$

As a simple example of Eq. (9b), let us consider the case where $\lambda = 3$, $i_1 = 1$, $i_2 = 1$, and $i_3 = 2$. Then the symmetrical traceless tensor of (9b) becomes

$$\begin{aligned} &m(m+2)x_1^2x_2 - mr^2\tilde{\Sigma}(\delta_{1,1}x_2) \\ &\quad = m(m+2)(x_1^2x_2 - r^2x_2/(m+2)), \end{aligned} \tag{10}$$

which should be compared with Eq. (7).

We will now show that Eqs. (9a) and (9b) can be obtained from Eq. (3), although they seem very different at first glance! From the definition of the generalized Laplace operator, it follows that

$$\begin{aligned} &\Delta(x_{i_1}x_{i_2}\cdots x_{i_\lambda}) \\ &\quad = \frac{x_{i_1}x_{i_2}\cdots x_{i_\lambda}}{x_1^2}\left(\sum_{j=1}^{\lambda}\delta_{i_j1}\right)\left(-1+\sum_{j=1}^{\lambda}\delta_{i_j1}\right) \\ &\quad + \frac{x_{i_1}x_{i_2}\cdots x_{i_\lambda}}{x_2^2}\left(\sum_{j=1}^{\lambda}\delta_{i_j2}\right)\left(-1+\sum_{j=1}^{\lambda}\delta_{i_j2}\right) \\ &\quad + \cdots\frac{x_{i_1}x_{i_2}\cdots x_{i_\lambda}}{x_m^2}\left(\sum_{j=1}^{\lambda}\delta_{i_jm}\right)\left(-1+\sum_{j=1}^{\lambda}\delta_{i_jm}\right) \\ &\quad = 2\tilde{\Sigma}(\delta_{i_1i_2}x_{i_3}\cdots x_{i_\lambda}). \end{aligned} \tag{11}$$

Similarly, we have

$$\Delta^2(x_{i_1}\cdots x_{i_\lambda}) = 2^2 2!\tilde{\Sigma}(\delta_{i_1i_2}\delta_{i_3i_4}x_{i_5}\cdots x_{i_\lambda}), \tag{12}$$

where the factor 2! comes from the symmetrization of the indices of the two $\delta$-symbols. Since there are $t!$ different products for $t$ $\delta$-symbols when the indices are symmetrized, we have, in general,

$$\Delta^t(x_{i_1}\cdots x_{i_\lambda}) = 2^t t!\tilde{\Sigma}(\delta_{i_1i_2}\delta_{i_3i_4}\cdots\delta_{i_{2t-1}i_{2t}}x_{i_{2t+1}}\cdots x_{i_\lambda}). \tag{13}$$

If we substitute Eq. (13) into Eq. (3), and let $\lambda = 2t$ for even $\lambda$ or $\lambda = 2t+1$ for odd $\lambda$, we obtain Eqs. (9a) and (9b).

We can ask how many linearly independent harmonic polynomials of order $\lambda$ it is possible to construct in an $m$-dimensional space. The number of linearly independent functions of the form $x_{i_1}x_{i_2}\cdots x_{i_\lambda}$ that it is possible to construct is

$$N(\lambda) = \binom{\lambda+m-1}{\lambda}. \tag{14}$$

Since the harmonic polynomials are required to be traceless, there will be

$$N(\lambda-2) = \binom{\lambda+m-3}{\lambda-2} \tag{15}$$

relations of the form

$$h_{\lambda;n_1+2,n_2,\ldots,n_m} + h_{\lambda;n_1,n_2+2,\ldots,n_m} + \cdots + h_{\lambda;n_1,n_2,\ldots,n_m+2} = 0, \tag{16}$$

with

$$n_1 + n_2 + \cdots + n_m = \lambda - 2, \tag{17}$$

where the $n_j$'s are the powers of the $x_j$'s. Therefore the number of linearly independent harmonic polynomials of order $\lambda$ in $m$-dimensional space is

$$\begin{aligned} N(\lambda) - N(\lambda-2) &= \binom{\lambda+m-1}{\lambda} - \binom{\lambda+m-3}{\lambda-2} \\ &= \frac{(2\lambda+m-2)(\lambda+m-3)}{\lambda!(m-2)!}. \end{aligned} \tag{18}$$

A set of harmonic polynomials of order $\lambda$ in an $m$-dimensional space is related to a set of hyperspherical harmonics of order $\lambda$ by

$$Y_{\lambda;\mu}(\Omega) = \eta r^{-\lambda}h_{\lambda;\mu}(\mathbf{x}), \tag{19}$$

where $\mu = (\mu_1,\mu_2,\ldots,\mu_{m-2})$ is a set of indices which label the different linearly independent harmonic polynomials, and hence also the different hyperspherical harmonics. In Eq. (19), $\eta$ is a normalization constant and $\Omega$ stands for a set of $m-1$ angles in the space. Since the harmonic polynomials satisfy the generalized Laplace equation, we have

$$\Delta[h_{\lambda;\mu}(\mathbf{x})] = \Delta[r^\lambda Y_{\lambda\mu}(\Omega)] = 0. \tag{20}$$

One can express the $m$-dimensional Laplace operator in the form

$$\Delta = \frac{1}{r^{m-1}}\frac{\partial}{\partial r}r^{m-1}\frac{\partial}{\partial r} - \frac{\Lambda_m^2}{r^2}, \tag{21}$$

where $\Lambda_m^2$ is the generalized angular momentum operator

$$\Lambda_m^2 = -\sum_{i<j}^{m}\Lambda_{ij}^2 \tag{22}$$

and

$$\Lambda_{ij} = x_i\frac{\partial}{\partial x_j} - x_j\frac{\partial}{\partial x_i}. \tag{23}$$

From Eqs. (20) and (21) it follows that

$$\Lambda_m^2 Y_{\lambda;\mu}(\Omega) = \lambda(\lambda+m-2)Y_{\lambda;\mu}(\Omega). \tag{24}$$

Equation (24) can be regarded as a definition of hyperspherical harmonics. They are eigenfunctions of the generalized angular momentum operator $\Lambda_m^2$. Usually they are cho-

sen in such a way that they also fulfill an orthonormality relation of the form

$$\int d\Omega_m \; Y_{\lambda';\mu'}(\Omega) Y_{\lambda;\mu}(\Omega) = \delta_{\lambda'\lambda} \, \delta_{\mu'\mu}, \tag{25}$$

where $d\Omega_m$ is defined by

$$dx_1 \, dx_2 \cdots dx_m = r^{m-1} \, dr \, d\Omega_m. \tag{26}$$

Now suppose that we have found a set of hyperspherical harmonics $Y_{\lambda;\mu}(\Omega)$ which satisfy Eqs. (24) and (25) for a particular value of $\lambda$. Then any other set of functions $\mathcal{Y}_{\lambda;\tau}(\Omega)$ which are related to the set $Y_{\lambda;\mu}(\Omega)$ by a unitary transformation

$$\mathcal{Y}_{\lambda;\tau}(\Omega) = \sum_{\mu} Y_{\lambda;\mu}(\Omega) U_{\mu\tau} \tag{27}$$

will also satisfy (24) and (25). Obviously, there are infinitely many ways of carrying out such a transformation, since the only restriction on $U_{\mu\tau}$ is that it should be unitary. Hence there are infinitely many possible ways of constructing sets of hyperspherical harmonics, and further specification of $Y_{\lambda;\mu}(\Omega)$ will depend on the organization of the indices $\mu$ and on the definition of the hyperspherical angles.

As an example of a particular way of specifying a set of hyperspherical harmonics let us consider the case where the angles are defined by the equation[25]

$$\begin{aligned}
x_1 &= r \sin\theta_1 \sin\theta_2 \cdots \sin\theta_{m-3} \sin\theta_{m-2} \sin\phi, \\
x_2 &= r \sin\theta_1 \sin\theta_2 \cdots \sin\theta_{m-3} \sin\theta_{m-2} \cos\phi, \\
x_3 &= r \sin\theta_1 \sin\theta_2 \cdots \sin\theta_{m-3} \cos\theta_{m-2}, \\
x_4 &= r \sin\theta_1 \sin\theta_2 \cdots \cos\theta_{m-3}, \\
&\vdots \\
x_{m-1} &= r \sin\theta_1 \cos\theta_2, \\
x_m &= r \cos\theta_1.
\end{aligned} \tag{28}$$

In terms of the angles $\theta_1,\ldots,\theta_{m-2},\phi$ the element of the solid angle becomes

$$d\Omega_m = (\sin\theta_1)^{m-2} (\sin\theta_2)^{m-3} \cdots (\sin\theta_{m-3})^2 \times \sin\theta_{m-2} \, d\theta_1 \, d\theta_2 \cdots d\theta_{m-2} \, d\phi. \tag{29}$$

Integrating over the ranges $0 \leqslant \theta_j \leqslant \pi$ and $0 \leqslant \phi \leqslant 2\pi$, we obtain

$$\int d\Omega_m = \frac{2\pi^{m/2}}{\Gamma(m/2)} = \frac{N_m}{(m-2)!!}, \tag{30}$$

where

$$N_m \equiv \begin{cases} (2\pi)^{m/2}, & \text{if } m \text{ is even}, \\ \pi(2\pi)^{(m-1)/2}, & \text{if } m \text{ is odd}. \end{cases} \tag{31}$$

[Notice that (30) does not depend on the particular choice of angle (28), since $d\Omega_m$ can be defined by (26).] In terms of the angles defined by (28), a particular set of hyperspherical harmonics can be written in the form[20]

$$Y_{\lambda;\mu_1,\ldots,\mu_{m-2}}(\Omega) = \left[ \prod_{j=1}^{m-3} \Theta(\mu_{j-1},\mu_j;\theta_j) \right] Y_{\mu_{m-3},\mu_{m-2}}(\theta_{m-2},\phi), \tag{32}$$

$$\Theta(\mu_{j-1},\mu_j;\theta_j) = \left[ \frac{2^{(2\mu_j + m - j - 3)} \, \Gamma^2(\mu_j + (m-j-1)/2) \Gamma(\mu_{j-1} - \mu_j + 1)(2\mu_{j-1} + m - j - 1)}{\pi \Gamma(\mu_{j-1} - \mu_j + m - j - 1)} \right]^{1/2} \times (\sin\theta_j)^{\mu_j} \, C_{\mu_{j-1}-\mu_j}^{\mu_j + (m-j-1)/2}(\cos\theta_j). \tag{33}$$

In the above formulas the functions $C_{\mu_{j-1}-\mu_j}^{\mu_j + (m-j-1)/2}(\cos\theta_j)$ are Gegenbauer polynomials defined by

$$C_n^{(m-2)/2}(x) = \sum_{t=0}^{(n/2)} \frac{(-1)^t (m + 2n - 2t - 4)!!}{t! 2^{n-t} (m-4)!! (n-2t)!!} (2x)^{n-2t}, \tag{34}$$

while the functions $Y_{\mu_{m-3},\mu_{m-2}}(\theta_{m-2},\phi)$ are the familiar three-dimensional spherical harmonics. The set of indices satisfies

$$\lambda = \mu_0 \geqslant \mu_1 \geqslant \cdots \geqslant \mu_{m-2} \geqslant 0. \tag{35}$$

For example, we can find the four-dimensional spherical harmonics by means of the formula

$$\begin{aligned}
&Y_{\lambda;\mu_1,\mu_2}(\theta_1,\theta_2,\phi) \\
&\quad = 2^{\mu_1} \Gamma(\mu_1 + 1) \left[ \frac{2(\mu_0 + 1)(\mu_0 - \mu_1)!}{\pi(\mu_0 + \mu_1 + 1)!} \right]^{1/2} \\
&\qquad \times (\sin\theta_1)^{\mu_1} \, C_{\mu_0 - \mu_1}^{\mu_1 + 1}(\cos\theta_1) Y_{\mu_1,\mu_2}(\theta_2,\phi) \\
&\quad = i^{-\mu} \, [2(\mu 0 + 1)/\pi]^{1/2} \, H_{\mu_0/2,\mu 1}(\cos\theta_1) Y_{\mu_1,\mu_2}(\theta_2,\phi),
\end{aligned} \tag{36}$$

where $H_{\mu_0/2,\mu_1}(\cos\theta_1)$ is the function which is discussed in some detail by Bander and Itzykson[26] and Talman.[27]

The hyperspherical harmonics $Y_{\lambda;\mu_1,\ldots,\mu_{m-2}}(\Omega)$ defined by Eqs. (32) and (33) obey the orthonormality relation

$$\int d\Omega_m \; Y^*_{\lambda;\mu_1,\ldots,\mu_{m-2}}(\Omega) Y_{\lambda;\mu'_1,\ldots,\mu'_{m-2}}(\Omega) = \prod_{j=0}^{m-2} \delta_{\mu_j\mu'_j}, \tag{37}$$

and they are simultaneous eigenfunctions of the set of commuting Casimir operators $\Lambda^2_m, \Lambda^2_{m-1}, \ldots, \Lambda^2_3$ and $\partial^2/\partial\phi^2$, so that they obey

$$\Lambda^2_{m-j} \, Y_{\lambda;\mu}(\Omega) = \mu_j(\mu_j + m - j - 2) Y_{\lambda;\mu}, \quad j = 0,1,\ldots,m-3 \tag{38}$$

and

$$\frac{\partial}{\partial\phi} Y_{\lambda \pm \mu}(\Omega) = \pm i\mu_{m-2} \, Y_{\lambda;\mu}(\Omega). \tag{39}$$

Some authors define hyperspherical coordinates in ways which are different from Eq. (28). For example, in the papers of Delves[28] and Knirk[14] the hyperspherical coordinates are defined by a clustering approach in which the polar angles of individual particles are preserved, and thus their hyperangular functions contain indices representing the angular momenta of individual particles. From the standpoint of group theory both the hyperspherical harmonics defined by Eqs. (32) and (33) and those defined by Knirk and Delves can be seen to be basis functions of the same irreducible representations of SO($m$). They differ, however, in their sub-

group symmetry. The hyperspherical harmonics of Knirk and Delves are adapted to the subgroup chain

$$\mathrm{So}(m) \supset \mathrm{SO}(m-3) \oplus \mathrm{SO}(3) \supset \cdots \supset [\oplus\, \mathrm{SO}(3)]^{m/3}, \tag{40}$$

while the harmonics of Eqs. (32) and (33) are adapted to the canonical subgroup chain

$$\mathrm{SO}(m) \supset \mathrm{SO}(m-1) \supset \cdots \supset \mathrm{SO}(2). \tag{41}$$

## III. THE ADDITION THEOREM FOR HYPERSPHERICAL HARMONICS[29-31]

In Refs. 30 and 31 it is shown that in three- and four-dimensional spaces the multiplication rule for matrix representations of the rotation group leads to an addition theorem. Now let us try to carry through an analogous proof for $m$-dimensional hyperspherical harmonics. Since the functions $Y_{\lambda;\mu}(\Omega)$ form the basis of an irreducible unitary representation of SO($m$), it follows that if $R$ is an element of SO($m$) we can write

$$RY_{\lambda;\mu}(\Omega) = RY_{\lambda;\mu}(\mathbf{u}) \equiv Y_{\lambda;\mu}(R^{-1}\mathbf{u}) = \sum_{\mu'} Y_{\lambda;\mu'}(\mathbf{u})\, D^{\lambda}_{\mu',\mu}(R), \tag{42}$$

where

$$\mathbf{u} = \left(\frac{x_1}{r}, \frac{x_2}{r}, \ldots, \frac{x_m}{r}\right) \tag{43}$$

and where $D^{\lambda}_{\mu',\mu}(R)$ is the matrix representing $R$ in the irreducible representation $\lambda$.

Now suppose that the unit vector $\mathbf{u}$ points in the direction of the $x_m$ axis so that

$$\mathbf{u} = \mathbf{u}_m \equiv (0,0,\ldots,0,1). \tag{44}$$

Then from Eqs. (32) and (33) we have

$$Y_{\lambda;\mu}(\mathbf{u}_m) = 0, \quad \text{if } \mu \neq 0, \tag{45}$$

and

$$Y_{\lambda;\mu}(\mathbf{u}_m) = \Theta(\lambda,0,0)\sqrt{\frac{1}{4\pi}} \times \prod_{j=2}^{m-3}\left[\frac{2^{m-j-3}\,\Gamma^2[(m-j-1)/2](m-j-1)}{\pi(m-j-2)!}\right]^{1/2} = \left[\frac{(m-2)!!\Gamma(\lambda+m-2)(2\lambda+m-2)}{N_m \lambda!\Gamma(m-1)}\right]^{1/2}. \tag{46}$$

Setting $\mathbf{u} = u_m$ in Eq. (42) and making use of (45) and (46), we

$$Y_{\lambda;\mu}(R^{-1}u_m) = \left[\frac{(m-1)!!\Gamma(\lambda+m-2)(2\lambda+m-2)}{N_m \lambda!\Gamma(m-1)}\right]^{1/2} \times D^{\lambda}_{0,\mu}(R). \tag{47}$$

Because of the unitarity of $D^{\lambda}_{\mu',\mu}$ we can rewrite Eq. (47) in the form

$$D^{\lambda}_{\mu,0}(R)^* = \left[\frac{N_m \lambda!\Gamma(m-1)}{(m-2)!!\Gamma(m-2)(2\lambda+m-2)}\right]^{1/2} Y_{\lambda;\mu}(Ru_m). \tag{48}$$

As a special case of (48) we have

$$D^{\lambda}_{0,0}(R) = \frac{\lambda!\Gamma(m+2)}{\Gamma(\lambda+m-2)} C_{\lambda}^{(m-2)/2}(\cos\bar{\theta}_1), \tag{49}$$

where $\bar{\theta}_1$ is the first angular coordinate of the vector $R\mathbf{u}_m$, i.e., the angle between $R\mathbf{u}_m$ and $\mathbf{u}_m$.

From the definition of a representation, we have

$$D^{\lambda}_{\mu,\mu'}(R_2R_1) = \sum_{\mu''} D^{\lambda}_{\mu,\mu''}(R_2) D^{\lambda}_{\mu'',\mu'}(R_1). \tag{50}$$

Combining Eq. (50) with Eqs. (47)–(49), we obtain

$$\sum_{\mu''} Y^*_{\lambda;\mu''}(R_1\mathbf{u}_m) Y_{\lambda;\mu''}(R_2^{-1}\mathbf{u}_m) = \frac{(m-4)!!(2\lambda+m-2)}{N_m} C_{\lambda}^{(m-2)/2}(\cos\bar{\theta}_1'). \tag{51}$$

Here $\bar{\theta}_1$ is the angle between $R_2R_1\mathbf{u}_m$ and $\mathbf{u}_m$, or the angle between the vectors $R_1\mathbf{u}_m$ and $R_2^{-1}\mathbf{u}_m$ (see Fig. 1). If we let $\Omega$ and $\Omega'$ be the angular coordinates related, respectively, to $\mathbf{u}$ and $\mathbf{u}'$, then Eq. (51) becomes

$$\sum_{\mu} Y^*_{\lambda;\mu}(\Omega) Y_{\lambda;\mu}(\Omega') = \frac{(2\lambda+m-2)(m-4)!!}{N_m} C_{\lambda}^{(m-2)/2}(\mathbf{u}\cdot\mathbf{u}'), \tag{52}$$

where $N_m$ is defined by Eq. (31). This is the form of the addition theorem given by us in Ref. 22. It should be noted that although we have used explicit expressions for the hyperspherical harmonics to derive Eq. (52), the formula is correct in general, since it can be derived using only the unitarity of the representations and Eq. (24) (see Ref. 22).

From Eq. (52) it follows that if $f(\Omega)$ is an arbitrary function of the hyperangular coordinates $\Omega$ and if $O^{(m)}_{\lambda}$ is a projection operator which projects out the component of $f(\Omega)$, which is an eigenfunction of $\Lambda^2_m$ with the eigenvalue $\lambda(\lambda+m-2)$, then

$$O^{(m)}_{\lambda}[f(\Omega)] = \frac{(m-4)!!(2\lambda+m-2)}{N_m} \times \int d\Omega'_m\, C_{\lambda}^{(m-2)/2}(\mathbf{u}\cdot\mathbf{u}') f(\Omega'). \tag{53}$$

For example, if $f(\Omega) = 1$,

$$O^{(m)}_{\lambda}[1] = \frac{(m-4)!!(2\lambda+m-2)}{N_m} \times \int d\Omega'_m\, C_{\lambda}^{(m-2)/2}(\mathbf{u}\cdot\mathbf{u}'), \tag{54}$$

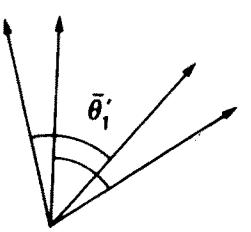



FIG. 1. Angles in Eq. (51).

while if $f(\Omega) = C_\lambda^{(m-2)/2}(\mathbf{u}''\cdot\mathbf{u})$,

$$O_\lambda^{(m)}[C_\lambda^{(m-2)/2}(\mathbf{u}''\cdot\mathbf{u})]$$
$$= \frac{(m-4)!!(2\lambda+m-2)}{N_m}\int d\Omega'_m\, C_\lambda^{(m-2)/2}(\mathbf{u}\cdot\mathbf{u}')$$
$$\times C_\lambda^{(m-2)/2}(\mathbf{u}''\cdot\mathbf{u}')$$
$$= \delta_{\lambda,\lambda'}\, C_\lambda^{(m-2)/2}(\mathbf{u}''\cdot\mathbf{u}). \tag{55}$$

## IV. BICENTER EXPANSIONS OF FUNCTIONS BY MEANS OF GEGENBAUER POLYNOMIALS

The formalism which we have been discussing can be used to construct bicenter expansions of functions in an $m$-dimensional space. In order to do this, we begin by expanding an $m$-dimensional plane wave in terms of Gegenbauer polynomials:

$$e^{i\mathbf{k}\cdot\mathbf{x}} = e^{i(k_1x_1 + k_2x_2 + \cdots + k_mx_m)}$$
$$= \sum_{\lambda=0}^{\infty} i^\lambda (m+2\lambda-2)(m-4)!!\, j_\lambda^m(kr)$$
$$\times C_\lambda^{(m-2)/2}(\mathbf{u}_k\cdot\mathbf{u}), \tag{56}$$

where $\mathbf{u}_k = \mathbf{k}/k$, $\mathbf{u} = \mathbf{x}/r$, and where the functions $j_\lambda^m(kr)$ are to be determined [the factor $(m+2\lambda-2)(m-4)!!$ is introduced for the sake of convenience]. Applying the operator $(\Delta_r + k^2)$ to both sides of Eq. (56), we obtain

$$(\Delta_r + k^2)e^{i\mathbf{k}\cdot\mathbf{x}} = (\Delta_r + k^2)\sum_{\lambda=0}^{\infty} i^\lambda (m+2\lambda-2)(m-4)!!$$
$$\times j_\lambda^m(kr)\, C_\lambda^{(m-2)/2}(\mathbf{u}_k\cdot u) = 0. \tag{57}$$

Combining Eqs. (57), (21), (24), and (52) we obtain an equation satisfied by $j_\lambda^m(\rho)$:

$$j_\lambda^{m''}(\rho) + \left(\frac{m-1}{r}\right) j_\lambda^{m'}(\rho) + \left[-\frac{\lambda(\lambda+m-2)}{r^2}\right] j_\lambda^m(\rho) = 0, \tag{58}$$

where $\rho = kr$. We now let

$$j_\lambda^m(\rho) = \rho^{-\nu} F(\rho) \quad \text{with } \nu = (m-2)/2. \tag{59}$$

Then Eq. (58) becomes

$$\rho^2 F''(\rho) + \rho F'(\rho) + [\rho^2 - (\lambda+\nu)^2] F(\rho) = 0,$$

which is the equation satisfied by the Bessel function of order $\lambda + \nu$. Thus we obtain

$$j_\lambda^m(\rho) = \frac{F(\rho)}{\rho^\nu} = \frac{J_{\lambda+\nu}(\rho)}{\rho^\nu}. \tag{60}$$

The many-dimensional plane wave of Eq. (58) is related to the $\delta$-function in the $m$-dimensional space by

$$\delta(\mathbf{x}) = \frac{1}{(2\pi)^m}\int d^m k\, e^{i\mathbf{k}\cdot\mathbf{x}}$$
$$= \frac{1}{(2\pi)^m}\sum_{\lambda=0}^{\infty} i^\lambda (m+2\lambda-2)(m-4)!!$$
$$\times \int d^m k\, j_\lambda^m(kr)\, C_\lambda^{(m-2)/2}(\mathbf{u}_k\cdot\mathbf{u}). \tag{61}$$

An arbitrary function $f(\mathbf{x}' - \mathbf{x})$ can be expressed as

$$f(\mathbf{x}' - \mathbf{x}) = \int d^m x''\, f(\mathbf{x}'')\delta(\mathbf{x}'' - \mathbf{x} + \mathbf{x}). \tag{62}$$

In the special case of an angle-independent function, $f(|\mathbf{x}' - \mathbf{x}|)$, (62) becomes

$$f(|\mathbf{x}' - \mathbf{x}|) = \int_0^\infty dr''\, r''^{m-1} f(r'')\int d\Omega_m\, \delta(\mathbf{x}'' - \mathbf{x}' + \mathbf{x}). \tag{63}$$

If we replace $\delta(\mathbf{x}'' - \mathbf{x}' + \mathbf{x})$ by (61) and make use of (54) and (55) we obtain

$$f(|\mathbf{x}' - \mathbf{x}|) = \sum_\lambda a_\lambda(r', r)\, C_\lambda^{(m-2)/2}(\mathbf{u}'\cdot\mathbf{u}), \tag{64}$$

where

$$a_\lambda(r', r) = \frac{N_m^2 (m+2\lambda-2)(m-4)!!}{(2\pi)^2}$$
$$\times \int_0^\infty dr''\, r''^{m-1} f(r'') J_{0\lambda\lambda}(r'', r', r) \tag{65}$$

and

$$J_{0\lambda\lambda} = \int_0^\infty dk\, k^{m-1} j_0^m(kr'')\, j_\lambda^m(kr')\, j_\lambda^m(kr). \tag{66}$$

The function $a_\lambda(r', r)$ can be expressed in a differential form by the following procedure. We rewrite $a_\lambda(r', r)$ as

$$a_\lambda(r', r) = N_m^2 (m+2\lambda-2)(m-4)!!$$
$$\times \int_0^\infty dk\, k^{m-1} F(k)\, j_\lambda^m(kr')\, j_\lambda^m(kr), \tag{67}$$

where

$$F(k) = \frac{1}{(2\pi)^m}\int_0^\infty dr''\, r''^{m-1} f(r'')\, j_0^m(kr''). \tag{68}$$

Substituting the explicit series for $j_\lambda^m(kr)$ into (69), we obtain

$$a_\lambda(r', r) = N_m^2 (m+2\lambda-2)(m-4)!!$$
$$\times \sum_{n=0}^{\infty} \frac{(-1)^n r^{2n+\lambda}}{(2n)!!(2n+2\lambda+m-2)!!}$$
$$\times \int_0^\infty dk\, k^{2m+\lambda+m-1} F(k)\, j_\lambda^m(kr'). \tag{69}$$

It is not difficult to prove by reduction that

$$j_\lambda^m(kr') = \left(-\frac{1}{k}\right)^\lambda r'^\lambda \left(\frac{1}{r'}\frac{\partial}{\partial r'}\right)^\lambda j_0^m(kr'). \tag{70}$$

Then the formula (69) becomes

$$a_\lambda(r', r) = N_m^2 (m+2\lambda-2)(m-4)!!$$
$$\times \sum_{n=0} \frac{(-1)^n r^{2n+\lambda}}{(2n)!!(2n+2\lambda+m-2)!!}$$
$$\times (-1)^\lambda r'^\lambda \left(\frac{1}{r'}\frac{\partial}{\partial r'}\right)^\lambda \int_0^\infty dk\, k^{2n+m-1}$$
$$\times F(k)\, j_0^m(kr'). \tag{71}$$

From formulas (58) and (56) it can be verified that

$$N_m \int_0^\infty dk\, k^{m-1} F(k)\, j_0^m(kr')$$
$$= \int d^m \mathbf{k}\, e^{-i\mathbf{k}\cdot\mathbf{x}'} F(k) = f(r').$$

Applying $\nabla_{r'}^2$ to both sides of the previous equation, we obtain

$$\nabla_{r'}^2 f(r') = -\int d^m \mathbf{k}\, k^2 e^{-i\mathbf{k}\cdot\mathbf{r}'} F(k)$$

$$= -N_m \int_0^\infty dk\, k^{2+m-1} j_0^m(kr') F(k).$$

Successive applications of $\nabla_{r'}^2$ $n$ leads to

$$\nabla_{r'}^{2n} f(r') = (-1)^n N_m \int_0^\infty dk\, k^{2n+m-1} j_0^m(kr') F(k). \tag{72}$$

By combining (71) and (72), we finally obtain

$$a_\lambda(r',r) = \sum_{n=0}^{\infty} f_{n\lambda}(r')\, r^{2n+\lambda}, \tag{73}$$

where

$$f_{n\lambda}(r') = \frac{(-1)(m+2\lambda-2)(m-4)!!N_m}{(2n)!!(2n+2\lambda+m-2)!!} \times r'^{\lambda}\left(\frac{1}{r'}\frac{\partial}{\partial r'}\right)^{\lambda} \nabla_{r'}^{2n} f(r'). \tag{74}$$

Formulas (64), (73), and (74) are the $m$-dimensional generalizations of the three-dimensional expressions obtained by one of us.[32]

## V. $3\lambda$ COEFFICIENTS FOR THE HYPERSPHERICAL HARMONICS

The Wigner coefficient and the $3j$ coefficient related to it for SO(3) and their great utility are well known. Wigner coefficients for some higher rotation groups, such as SO(4) and SO(6) (see Ref. 33), have been worked out and used in calculations. We now consider in general the $3\lambda$ coefficients of the hyperspherical harmonics for SO($m$). We define symmetrized "$3\lambda$" symbols as follows:

$$\int d\Omega\, Y_{\lambda_1;\mu^1}(\Omega) Y_{\lambda_2;\mu^2}(\Omega) Y_{\lambda_3;\mu^3}(\Omega) \equiv \begin{bmatrix} \lambda_1 & \lambda_2 & \lambda_3 \\ \mu^1 & \mu^2 & \mu^3 \end{bmatrix}, \tag{75}$$

where the superscripts have been attached to the sets of $\mu$'s, thus $\mu^i$ means $(\mu_1^i, \mu_2^i, \ldots, \mu_{m-2}^i)$. If we substitute Eq. (32) into (75) and introduce the notation

$$I\begin{pmatrix} v_j; & \mu_{j-1}^1 & \mu_{j-1}^2 & \mu_{j-1}^3 \\ & \mu_j^1 & \mu_j^2 & \mu_j^3 \end{pmatrix} = \int_0^\pi d\omega_j\, \Theta(\mu_{j-1}^1, \mu_j^1; \theta_j)\Theta(\mu_{j-1}^2, \mu_j^2; \theta_j) \times \Theta(\mu_{j-1}^3, \mu_j^3; \theta), \tag{76}$$

$d\omega_j = d\theta_j (\sin\theta_j)^{m-j-1}$, $j = 1,2,\ldots,m-3$,

then the "$3\lambda$" can be written as

$$\begin{bmatrix} \lambda_1 & \lambda_2 & \lambda_3 \\ \mu^1 & \mu^2 & \mu^3 \end{bmatrix} = \prod_{j=1}^{m-2} I\begin{pmatrix} v_j; & \mu_{j-1}^1 & \mu_{j-1}^2 & \mu_{j-1}^3 \\ & \mu_j^1 & \mu_j^2 & \mu_j^3 \end{pmatrix}, \tag{77}$$

where the integral with $j = m-2$ is defined by

$$I\begin{pmatrix} v_{m-2}; & \mu_{m-3}^1 & \mu_{m-3}^2 & \mu_{m-3}^3 \\ & \mu_{m-2}^1 & \mu_{m-2}^2 & \mu_{m-2}^3 \end{pmatrix} = \left[\frac{(2\mu_{m-3}^1+1)(2\mu_{m-3}^2+1)(2\mu_{m-3}^3+1)}{4\pi}\right]^{1/2} \times \begin{pmatrix} \mu_{m-3}^1 & \mu_{m-3}^2 & \mu_{m-3}^3 \\ 0 & 0 & 0 \end{pmatrix} \times \begin{pmatrix} \mu_{m-3}^1 & \mu_{m-3}^2 & \mu_{m-3}^3 \\ \mu_{m-2}^1 & \mu_{m-2}^2 & \mu_{m-2}^3 \end{pmatrix}, \tag{78}$$

and where ( ) are the $3j$ symbols for SO(3). In order to evaluate integral (55) we introduce the explicit expression (33) of $\Theta$ functions, i.e., we write

$$I\begin{pmatrix} v_j; & \mu_{j-1}^1 & \mu_{j-1}^2 & \mu_{j-1}^3 \\ & \mu_j^1 & \mu_j^2 & \mu_j^3 \end{pmatrix} = \int_0^\pi d\theta_j\, (\sin\theta_j)^{M_j+2v_j} \times \prod_{i=1}^{3} C_{\lambda_i;\mu_j^i} C^{\mu_j^i+v_j}_{\mu_{j-1}^i-\mu_j^i}(\cos\theta_j), \tag{79}$$

where

$$M_j = \mu_j^1 + \mu_j^2 + \mu_j^3, \quad v_j = (m-j-1)/2,$$

and

$$C_{\lambda_i;\mu_j^i} = \{2^{2\mu_j^i+2v_j-2}\Gamma^2(\mu_j^i+v_j) \times \Gamma(\mu_{j-1}^i-\mu_j^i+1)(2\mu_{j-1}^i+2v_j) \times [\pi\Gamma(\mu_{j-1}^i+\mu_j^i+2v_j)]^{-1}\}^{1/2}. \tag{80}$$

Gegenbauer[34] stated that

$$C_s^\alpha(x) = \sum_{t=0}^{[s/2]} C_t(s,\alpha,\beta)\, C^\beta_{s-2t}(x), \tag{81}$$

where

$$C_t(s,\alpha,\beta) = \frac{(s-2t+\beta)\Gamma(t+\alpha-\beta)\Gamma(\alpha+s-t)\Gamma(\beta)}{\Gamma(1+t)\Gamma(\alpha-\beta)\Gamma(s-t+\beta+1)\Gamma(\alpha)}. \tag{82}$$

See Hua[35] for a simple proof.

We have

$$I\begin{pmatrix} v_j; & \mu_{j-1}^1 & \mu_{j-1}^2 & \mu_{j-1}^3 \\ & \mu_j^1 & \mu_j^2 & \mu_j^3 \end{pmatrix} = \sum_{t_i=0}^{[(\mu_{j-1}^i-\mu_j^i)/2]} \prod_{i=1}^{3} C_{\lambda_i;\mu_j^i} C_{t_i}\left(\mu_{j-1}^i-\mu_j^i, \mu_j^i+v_j, \frac{1}{2}\right) \times \int_0^\pi d\theta_j (\sin\theta_j)^{M_j+2v_j} \prod_{i=1}^{3} P_{\mu_{j-1}^i-\mu_j^i-2t}(\cos\theta_j). \tag{83}$$

In obtaining the above equation we have used the well-known relation between Gegenbauer functions and Legendre functions

$$C_s^{1/2}(x) = P_s(x). \tag{84}$$

It is not difficult to evaluate the remaining integral in Eqs. (83) if we make use of the following relations:

$$P_{n_1}(x)P_{n_2}(x) = \sum_{n_3} (2n_3+1)\begin{pmatrix} n_1 & n_2 & n_3 \\ 0 & 0 & 0 \end{pmatrix}^2 P_{n_3}(x) \tag{85}$$

and[36]

$$\int_{-1}^{1} dx(1-x^2)^{\rho-1} P_\nu(x) = \frac{\pi[\Gamma(\rho)]^2}{\Gamma(\rho+\nu/2+1/2)\Gamma(\rho-\nu/2)\Gamma(\nu/2+1)\Gamma(-\nu/2+1/2)}, \quad 4 \quad \mathrm{Re}\,\rho>0. \tag{86}$$

Integrating (83) out we obtain

$$I\begin{pmatrix} v_j; & \mu_{j-1}^1 & \mu_{j-1}^2 & \mu_{j-1}^3 \\ & \mu_j^1 & \mu_j^2 & \mu_j^3 \end{pmatrix} = \sum_{t_i=0}^{[(1/2)(\mu_{j-1}^i-\mu_j^i)]} \prod_{i=1}^{3} F_1(v_j,\mu_{j-1}^i,\mu_j^i,t_i) \sum_{\tau_1\tau_2} (2\tau_1+1)(2\tau_2+1) F_2(v_j,M_i;\tau_2)$$

$$\times \begin{pmatrix} \mu_{j-1}^1-\mu_j^1-2t_1 & \mu_{j-1}^2-\mu_j^2-2t_2 & \tau_1 \\ 0 & 0 & 0 \end{pmatrix}^2 \begin{pmatrix} \mu_{j-1}^3-\mu_j^3-2t_3 & \tau_2 & \tau_1 \\ 0 & 0 & 0 \end{pmatrix}^2, \tag{87}$$

where

$$F_1(v_j,\mu_{j-1}^i,\mu_j^i;t_i) = \left[\frac{2^{2\mu_j^i+1}(\mu_{j-1}^i+v_j)\Gamma(\mu_{j-1}^i-\mu_j^i+1)}{(\mu_{j-1}^i+\mu_j^i+2v_j)\Gamma^2(\mu_j^i+v_j-\frac{1}{2})}\right]^{1/2}$$

$$\times \frac{(\mu_{j-1}^i-\mu_j^i-2t_i+\frac{1}{2})\Gamma(t_i+\mu_j^i+v_j-\frac{1}{2})\Gamma(\mu_{j-1}^i+v_j-t_i)}{\Gamma(t_i+1)\Gamma(\mu_{j-1}^i-\mu_j^i-t_i+\frac{3}{2})} \tag{88}$$

and

$$F_2(v_j,M_j;\tau_2) = \frac{\Gamma^2[(M_j+2v_j+1)/2]\Gamma[(1-\tau_2)/2]\sin[(1+\tau_2)\pi/2]}{\Gamma[(M_j+2v_j+2+\tau_2)/2]\Gamma[(M_j+2v_j+1-\tau_2)/2]\Gamma[(2+\tau_2)/2]}. \tag{89}$$

Here, $\tau_1$ and $\tau_2$ take on all the values allowed by the $3j$ symbols. For SO(4), it follows that

$$\begin{bmatrix} \lambda_1 & \lambda_2 & \lambda_3 \\ \mu^1 & \mu^2 & \mu^3 \end{bmatrix} = I\begin{pmatrix} 1; & \mu_0^1 & \mu_0^2 & \mu_0^3 \\ & \mu_1^1 & \mu_1^2 & \mu_1^3 \end{pmatrix} \cdot I\begin{pmatrix} \frac{1}{2}; & \mu_1^1 & \mu_1^2 & \mu_1^3 \\ & \mu_2^1 & \mu_2^2 & \mu_2^3 \end{pmatrix}. \tag{90}$$

The previous formulas produce the same values as those given by Shibura and Wulfman[37] within a common factor $(1/2\pi^2)^{1/2}$.


## ACKNOWLEDGMENTS

It is a pleasure to acknowledge stimulating conversations with Professor Jens Peder Dahl, Professor Steve Berry, Professor Brian Judd, Professor Hendrik J. Monkhorst, Professor Jan Linderberg, and Professor Metin Demiralp. We are also extremely grateful to Lektor Svend Erik Nielsen, Lektor Sten Rettrup, and Dr. Birthe L. Christensen-Dalsgaard for their adviee and interest. One of us (Z-Y. W.) would like to thank Professor Thor A. Bak and the Danish Ministry of Education for the opportunity to visit Copenhagen.


## APPENDIX: PROOF OF EQ. (3)

Suppose that $f_\lambda(x)$ is a homogeneous polynomial of degree $\lambda$ in an $m$-dimensional space. Then we have the canonical decomposition[14]

$$f_\lambda(x) = \sum_{k=0}^{[\lambda/2]} r^{2k} h_{\lambda-2k}(x), \tag{A1}$$

where $h_{\lambda-2k}(x)$ is a homogeneous harmonic polynomial of degree $\lambda-2k$. It can be verified by direct calculation that

$$\Delta(r^{2k} f_{\lambda-2k'}) = 2k(m+2\lambda+2k-4k'-2)r^{2k-2} f_{\lambda-2k'} + r^{2k}\Delta f_{\lambda-2k'}, \tag{A2}$$

where $f_{\lambda-2k'}$ is a homogeneous polynomial of degree $\lambda-2k'$. Thus we can prove from the induction of (A2) that

$$\Delta^{k'}(r^{2k} f_{\lambda-2k}) = \sum_{t=0}^{k'} 2^{k'-t}\binom{k'}{t} \frac{k!(n+2\lambda-2k-2t-2)!!}{(k-k'+t)!(n+2\lambda-2k-2k'-2)!!} \times r^{2k-2k'+2t}\Delta^t f_{\lambda-2k} \quad \text{for } k'<k. \tag{A3}$$

Applying the operator $\Delta$ to both sides of Eq. (A1) $k$ times and using Eq. (A3), we obtain

$$\Delta^k f_\lambda(x) = 2^k k! \frac{(n+2\lambda-2k-2)!!}{(n+2\lambda-4k-2)!!} h_{\lambda-2k}(x) + \sum_{k'=k+1}^{[\lambda/2]} \Delta^{k'}(r^{2k'} h_{\lambda-2k'}). \tag{A4}$$

It is obvious that $\Delta^k f_\lambda(x)$ is a homogeneous polynomial of degree $\lambda-2k$. Thus the harmonic polynomial of degree $\lambda-2k$ is

$$\begin{aligned} h_{\lambda-2k}(x) &= \frac{(n+2\lambda-4k-2)!!}{2^k k!(n+2\lambda-2k-2)!!} \times \sum_{t=0}^{[(\lambda-2k)/2]} \frac{(-1)^t(n+2\lambda-4k-2t-4)!!}{2^t t!(n+2\lambda-4k-4)!!} \times r^{2t}\Delta^{k+t} f_\lambda(x) \\ &= \frac{(n+2\lambda-4k-2)}{2^k k!(n+2\lambda-2k-2)!!} \times \sum_{t=0}^{[(\lambda-2k)/2]} \frac{(-1)^t(n+2\lambda-4k-2t-4)!!}{2^t t!} \times r^{2t}\Delta^{k+t} f_\lambda(x). \end{aligned} \tag{A5}$$

A special case of the above equation ($k = 0$) appears in Ref. 20.


[1]J. L. Ballot, M. Fabre, and De La Ripelle, Ann. Phys. **127**, 62 (1980).
[2]Yu. F. Smirnov and K. V. Shitikova, Sov. J. Part. Nucl. **8**, 344 (1977).
[3]J. H. Macek, J. Phys. B **1**, 831 (1968).
[4]R. J. White and F. H. Stillinger, Jr., J. Chem. Phys. **52**, 5800 (1970).
[5]U. Fano and C. D. Lin, in *Atomic Physics* (Plenum, New York, 1975), Vol. 4, p. 47.
[6]U. Fano, Rep. Prog. Phys. **46**, 99 (1983).
[7]C. D. Lin, Phys. Rev. A **25**, 76 (1982); **26**, 2305 (1982); **27**, 22 (1983).
[8]H. Klar and M. Klar, Phys. Rev. A **17**, 1007 (1977).
[9]G. O. Morrell and D. L. Knirk, Theor. Chim. Acta (Berlin), **37**, 345 (1975).
[10]R. C. Whitten and J. C. Sims, Phys. Rev. A **9**, 1586 (1974).
[11]V. D. Efros, A. M. Frolou, and M. I. Mukhtorova, J. Phys. B **15**, L819 (1982).
[12]M. S. Kildyushov and E. L. Surkov, Sov. J. Nucl. Phys. **22**, 333 (1975).
[13]C. W. Clark and C. H. Greene, Phys. Rev. A **21**, 1786 (1980).
[14]D. L. Knirk, J. Chem. Phys. **60**, 66 (1974).
[15]D. L. Knirk, J. Chem. Phys. **60**, 760 (1974).
[16]D. W. McColn and G. W. Erickson, Phys. Rev. A **26**, 1173 (1982).
[17]C. H. Greene and C. Jungen, in *Abstracts of Contributed Papers, II International Conference on the Physics of Electronic and Atomic Collisions*, Gatlinburg, TN, 1981, edited by S. Datz, p. 1019.
[18]Y. Öhrn and J. Linderberg, Mol. Phys. (to appear).
[19]A. Erdelyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953).
[20]N. J. Vilenkin, *Speical Functions and the Theory of Group Representations, Translations of Mathematical Monographs* (Am. Math. Soc., Providence, RI, 1968), Vol. 22.
[21]J. D. Louck, J. Mol. Spectrosc. **4**, 298 (1960).
[22]J. Avery and Z-Y. Wen, Int. J. Quant. Chem. **22**, 717 (1982).
[23]J. M. Normand and J. Raynal, J. Phys. A **15**, 1437 (1982).
[24]M. Hamermesh, *Group Theory and Its Applications in Physical Problems* (Addison–Wesley, Reading, MA, 1962).
[25]E. Kyriakopoulos, Phys. Rev. **174**, 1846 (1968).
[26]M. Bander and C. Itzykson, Rev. Mod. Phys. **38**, 33 (1966).
[27]J. D. Talman, *Special Functions* (Benjamin, New York, 1968).
[28]L. D. Delves, Nucl. Phys. **9**, 391 (1958–59).
[29]A. O. Barut and Y. Kitagawara, J. Phys. A **14**, 2581 (1981).
[30]A. R. Edmonds, *Augular Momentum in Quantum Mechanics* (Princeton, U.P., Princeton, NJ, 1957).
[31]J. S. Alper, J. Chem. Phys. **55**, 3770 (1971); K. G. Kay and J. S. Alper, J. Chem. Phys. **56**, 4243 (1972).
[32]J. Avery, Int. J. Quant. Chem. **16**, 1265 (1979).
[33]F. del Aguila, J. Math. Phys. **21**, 2327 (1980).
[34]L. Gegenbauer, Math. Naturwiss. Kl. **48**, 293 (1884).
[35]L. K. Hua, *Harmonic Analysis of Functions of Several Complex Variables in the Classical Domains* (English translation) (Am. Math. Soc., Providence, RI, 1963).
[36]I. S. Gradshteyn and I. M. Ryzhin, *Tables of Integrals, Series and Products* (Academic, New York, 1965), Formula 7.132,1.
[37]T. Shibura and C. E. Wulfman, Proc. R. Soc. London Ser. A **286**, 376 (1965).

# On the hyperbolic complex linear symmetry groups and their local gauge transformation actions

Zai-Zhe Zhong
*Department of Physics, Liaoning Normal University, Dalian, Liaoning, The People's Republic of China*

(Received 22 May 1984; accepted for publication 3 August 1984)

The hyperbolic complex linear groups and the isomorphic relation between these groups and real linear groups are discussed. A local hyperbolic complex gauge symmetry of the hyperbolic complex sesquilinear field is equivalent to some local real gauge symmetry of the real bilinear field.

## I. INTRODUCTION

Although the ring $H$ of hyperbolic complex (or "double") numbers has been studied extensively by mathematicians,[1,2] it has been scarcely applied directly in physics. Kunstatter, Moffat, and Malzan[3] first applied this ring in the theory of gravitation. They have suggested that the metric of the space-time manifold takes their values in $H$, and have proved that this metric has internal GL(4,$R$) gauge symmetry. However, this symmetry, in fact, is a HCLG (i.e., hyperbolic complex linear group) gauge symmetry. (See Ref. 4, and Sec. III.) In this paper we generally discuss the HCLG's and some related problems. In Sec. II, we give briefly definitions of principal HCLG's. These definitions are very similar to the case of CLG's (i.e., complex linear groups). Section III and Sec. IV are the principal parts of this paper. In Sec. III, we prove every principal HCLG to be isomorphic to some RLG (i.e., real linear group) or a direct product of two RLG's. This is an essential distinction between the HCLG's and the CLG's. In Sec. IV, we prove a local HCLG gauge symmetry of the hyperbolic complex sesquilinear fields to be equivalent to some local RLG gauge symmetry of the real bilinear fields. The case concerning the internal symmetry of a hyperbolic complex metric[3] in nonsymmetric gravitational theory is such an example. Finally, Sec. V is conclusions.

## II. GROUP GL($n$,$H$) AND ITS PRINCIPAL SUBGROUPS

Let $\epsilon$ denote the purely hyperbolic imaginary unit of $H$, $\epsilon^2 = +1$. The conjugation number of $z = a + \epsilon b$ ($a$ and $b$ are real) is $\tilde{z} = a - \epsilon b$, and the square of the norm of $z$ is $\|z\|^2 = z\tilde{z} = a^2 - b^2$. Although $z$ does not vanish in $H$, $\|z\|^2$ may vanish (e.g., $z = 1 + \epsilon$). Number $z^{-1}$ exists, if and only if $\|z\|^2 \neq 0$, where

$$z^{-1} = \tilde{z}/\|z\|^2 = (a - \epsilon b)(a^2 - b^2)^{-1}. \tag{1}$$

The above are different from the case in the fields of complex or real numbers.

According to the multiplication of matrices, the set of all $n \times n$ hyperbolic complex matrices obviously forms into a ring on $H$. If $A$ is a hyperbolic complex matrix and $\|\det(A)\|^2$ does not vanish, then we call matrix $A$ "nonsingular." On the contrary, we call $A$ "singular." By Eq. (1) we have the following.

**Theorem 1**: The converse $A^{-1}$ of a hyperbolic complex matrix $A$ exists if and only if $A$ is nonsingular. $A^{-1}$ is calculated in the same way as an ordinary converse matrix.

A hyperbolic complex vector is an $n \times 1$ hyperbolic complex matrix $V = (V^i)$. According to the ordinary way, the set of all $V$'s forms into a linear space $H^n$. We cannot copy the concept about ordinary linear independence for $H^n$. However, the following theorem can be easily proved (in the following, every Latin index takes values 1,2,...,$n$).

**Theorem 2:** $\{V_i\}$ is a basis of $H$ (i.e., any vector $W$ of $H$ can be expressed as $W = \alpha^i V_i, a^i \in H$) if and only if the hyperbolic complex matrix $(V_i^j)$ is nonsingular.

Now, we consider a hyperbolic complex linear mapping $f$: $H^n \to H^n$. $f$ can obviously be expressed in a hyperbolic complex matrix $(f_i^j)$, and the product of mappings corresponds to the product of matrices. If $(f_i^j)$ is nonsingular, we call $f$ nonsingular. Such $f$ has the converse $f^{-1}$, $(f^{-1}) = (f)^{-1}$. A basis is mapped into another basis under the action of a nonsingular mapping. Thus, the set of all nonsingular mappings (or matrices) forms a group GL($n$,$H$), i.e., hyperbolic complex general linear group of degree $n$. The group GL($n$,$H$) has some principal subgroups, the first, of course, is just the group GL($n$,$R$). Since the definitions of other principal subgroups are similar to the case of the ordinary complex linear group GL($n$,$C$), we spread them out directly as follows.

The hyperbolic complex orthogonal group of degree $n$ relative to signature $\eta$ is defined by

$$\mathrm{O}(\eta,H) = \{ f \mid f \in \mathrm{GL}(n,H), (f)\eta(f)^T\eta^{-1} = I \}, \tag{2}$$

where $\eta$ is some signature, $A^T$ is the transpose of $A$, and $I$ is the $n \times n$ unit matrix. When $\eta = I$, $\mathrm{O}(n,H) \equiv \mathrm{O}(\eta,H)$. Obviously, for any $V$ the quadratic form $\sigma = \eta_{ij} V^i V^j$ is invariant under the action of an element of $\mathrm{O}(\eta,H)$. The hyperbolic complex unitary group of degree $n$ relative to signature $\eta$ is defined by

$$\mathrm{U}(\eta,H) = \{ f \mid f \in \mathrm{GL}(n,H), (f)\eta(\tilde{f})^T\eta^{-1} = I \}. \tag{3}$$

When $\eta = I$, $\mathrm{U}(n,H) \equiv \mathrm{U}(\eta,H)$. For any $V$ the square of the norm $\|V\|^2 = \eta_{ij} V^i \tilde{V}^j$ is invariant under the action of an element of $\mathrm{U}(\eta,H)$. Obviously, the ordinary Lorentz group O(3,1) is a real subgroup of U(3,1,$H$). The hyperbolic complex symplectic group of degree $m$ is defined by

$$\mathrm{Sp}(m,H) = \{ f \mid f \in \mathrm{GL}(2m,H), (f)^T H (f) = J \}, \tag{4}$$

$$J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}.$$

According to the ordinary way, all the above HCLG's are Lie groups, and have respective Lie algebras.

 0022-2488/85/030404-03$02.50 © 1985 American Institute of Physics 

The transformation groups, whch can be applied extensively in physics, are the local linear Lie groups. For this kind of group, the Lie group construction exists only in some neighborhood of the unit element, and an "isomorphism," in fact, is a local isomorphism only. In the following, we only consider the local linear group and the word "isomorphism" only means a local isomorphism. According to the Lie group theory, such two groups are isomorphic if and only if their Lie algebras are isomorphic.

## III. ISOMORPHIC RELATION BETWEEN HCLG'S AND RLG'S

There are two special elements in $H$, i.e., $y = \frac{1}{2}(1+\epsilon)$ and $\tilde{y} = \frac{1}{2}(1-\epsilon)$,

$$y^2 = y, \quad \tilde{y}^2 = \tilde{y}, \quad \|y\|^2 = 0. \tag{5}$$

Therefore, $y$ and $\tilde{y}$ play the role of 1 and 0 in a hyperbolic complex matrix, respectively. This is particularly useful in this section.

*Lemma:* Let $n\times n$ matrices $A$ and $B$ be two elements of some real Lie algebra $a$, and a mapping $\rho$ be defined by

$$\rho(A,B) = \tfrac{1}{2}(A+B) + \tfrac{1}{2}\epsilon(A-B), \tag{6}$$

then $\rho$ is a Lie algebra homomorphic mapping.

*Proof:* $(A,B)$ is an element of the direct sum $a+a$. Obviously, $\rho$ is linear. Next,

$$\begin{aligned}\rho[(A,B),(C,D)] &= \rho([A,C],[B,D]) \\ &= \rho(AC-CA, BD-DB) \\ &= \tfrac{1}{2}(AC-CA+BD-DB) \\ &\quad + \tfrac{1}{2}(AC-CA-BD+DB)\epsilon \\ &= [\rho(A,B),\rho(C,D)]. \end{aligned}$$

Q.E.D.

Let $\mathscr{E}_{ij}$ be an $n\times n$ real matrix, and its $k$ th line $l$ th column element be $\delta_{ik}\delta_{jl}$. Then all $\mathscr{E}$'s form into a basis of $\mathrm{gl}(n,R)$ [i.e., the Lie algebra of $\mathrm{GL}(n,R)$].

**Theorem 3:** $\mathrm{GL}(n,R)\times\mathrm{GL}(n,R)$ is isomorphic to $\mathrm{GL}(n,H)$.

*Proof:* $a_{kl} = (\mathscr{E}_{kl}, 0)$ and $a_{\bar{k}\bar{l}} = (0, \mathscr{E}_{kl})$ form a basis of $\mathrm{gl}(n,R)+\mathrm{gl}(n,R)$, where $\bar{k} = k+n$, $\bar{l} = l+n$. $A_{kl} = \mathscr{E}_{kl}$ and $A_{\bar{k}\bar{l}} = \epsilon\mathscr{E}_{kl}$ form the basis of $\mathrm{gl}(n,H)$.

Obviously, $\rho(a_{kl})$ and $\rho(a_{\bar{k}\bar{l}})$ form another basis of $\mathrm{gl}(n,H)$. According to the above Lemma, this theorem is proved.

Q.E.D.

**Theorem 4:** $\mathrm{O}(\eta,R)\times\mathrm{O}(\eta,R)$ is isomorphic to $\mathrm{O}(\eta,H)$.

*Proof:* Any element of $A$ of $\mathrm{o}(\eta,R)$ satisfies

$$A\eta + \eta A^T = 0, \tag{7}$$

$$\rho(A,B)\eta + \eta(\rho(A,B))^T = 0. \tag{8}$$

Therefore, $\rho$ maps $\mathrm{o}(\eta,R)+\mathrm{o}(\eta,R)$ into $\mathrm{o}(\eta,H)$. In addition,

$$a_{kl} = (\mathscr{E}_{kl} - \eta_{kk}\eta_{ll}\mathscr{E}_{lk}, 0),$$

$$a_{\bar{k}\bar{l}} = (0, \mathscr{E}_{kl} - \eta_{kk}\eta_{ll}\mathscr{E}_{lk})$$

form a basis of $\mathrm{o}(\eta,R)+\mathrm{o}(\eta,R)$. Now, $\rho(a_{kl}) = \frac{1}{2}(1+\epsilon)(\mathscr{E}_{kl} - \eta_{kk}\eta_{ll}\mathscr{E}_{lk})$ and $\rho(a_{\bar{k}\bar{l}}) = \frac{1}{2}(1-\epsilon)(\mathscr{E}_{kl} - \eta_{kk}\eta_{ll}\mathscr{E}_{lk})$ form just a basis of $\mathrm{o}(\eta,H)$.

Q.E.D.

**Theorem 5:** $\mathrm{Sp}(2m,R)\times\mathrm{Sp}(2m,R)$ is isomorphic to $\mathrm{Sp}(2m,H)$. The proof is similar to Theorem 4.

**Theorem 6:** For an arbitrary signature $\eta$, $\mathrm{U}(\eta,H)$ is isomorphic to $\mathrm{GL}(n,R)$.

*Proof:* $A$ is an element of $\mathrm{u}(\eta,H)$ if and only if

$$A\eta + \eta\tilde{A}^T = 0. \tag{9}$$

Let

$$\begin{aligned} A_{kl} &= \mathscr{E}_{kl} - \eta_{kk}\eta_{ll}\mathscr{E}_{lk} \quad (k\neq l), \\ A_{\bar{k}\bar{l}} &= \epsilon(\mathscr{E}_{kl} + \eta_{kk}\eta_{ll}\mathscr{E}_{lk}) \quad (k\neq l), \\ A_{kk} &= \epsilon\mathscr{E}_{kk}, \end{aligned} \tag{10}$$

then $n^2$ elements $A_{kl}$, $A_{\bar{k}\bar{l}}$, and $A_{kk}$ form a basis of $\mathrm{u}(\eta,H)$. Let

$$\begin{aligned} \rho(\mathscr{E}_{kl}) &= \tfrac{1}{2}(A_{kl} + A_{\bar{k}\bar{l}}) \quad (k>l), \\ \rho(\mathscr{E}_{kl}) &= \tfrac{1}{2}\eta_{kk}\eta_{ll}(A_{\bar{l}\bar{k}} - A_{lk}) \quad (k<l), \\ \rho(\mathscr{E}_{kk}) &= A_{kk}. \end{aligned} \tag{11}$$

For any $k,l$ Eq. (11), in fact, can be unified by

$$\rho(\mathscr{E}_{kl}) = \tfrac{1}{2}(1+\epsilon)\mathscr{E}_{kl} - \tfrac{1}{2}(1-\epsilon)\eta_{kk}\eta_{ll}\mathscr{E}_{lk}. \tag{12}$$

We can directly examine that $\rho$ is an isomorphic mapping and all elements $\rho(\mathscr{E}_{kl})$ form just a basis of $\mathrm{u}(\eta,H)$.

Q.E.D.

According to Theorems 3 and 6, we can naturally contrast the interal gauge symmetry of the hyperbolic complex metric[3] with the Lorentz symmetry in general relativity. In the following, let symbol "$\rightarrow$" denote some imbedding and a line denote a (local) isomorphism in a graph. Then we obtain

$$\begin{array}{ccccc} \mathrm{GL}(4,R)\times\mathrm{GL}(4,R) & \text{---} & \mathrm{GL}(4,H) & \leftarrow & \mathrm{GL}(4,R) \\ \uparrow & & \uparrow & & \uparrow \\ \mathrm{GL}(4,R) & \text{---} & \mathrm{U}(3,1,H) & \leftarrow & \mathrm{SO}(3,1). \end{array} \tag{13}$$

We see that the local $\mathrm{GL}(4,R)$ gauge symmetry of an expression by real frame corresponds to the local $\mathrm{U}(\eta,H)$ gauge symmetry of an expression by hyperbolic complex frame,[4] where $\eta$ is some signature of degree 4. Corresponding to this, in general relativity the Riemann metric has the SO(3,1) gauge symmetry, and SO(3,1) is a subgroup of $\mathrm{GL}(4,R)$.

In $\mathrm{GL}(n,C)$, $\mathrm{O}(n,C)\cap\mathrm{U}(n,C) = \mathrm{O}(n,R)$. Corresponding to this, we also have the following theorem.

**Theorem 7:** $\mathrm{U}(\eta,H)\cap\mathrm{O}(\eta,H) = \mathrm{O}(\eta,R)$.

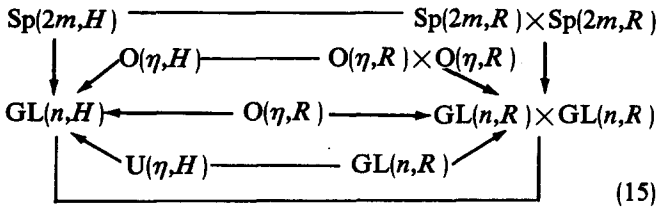
*Proof:* Since any element of $\mathrm{u}(\eta,H)\cap\mathrm{o}(\eta,H)$ must satisfy simultaneously Eq. (7) and Eq. (9), the $\frac{1}{2}n(n-1)$ elements $A_{kl}$ $(k>l)$ of Eq. (10) form a basis of $\mathrm{u}(\eta,H)\cap\mathrm{o}(\eta,H)$. $\{A_{kl}\}$ is also a basis of $\mathrm{o}(\eta,R)$, this means that $\mathrm{U}(\eta,H)\cap\mathrm{O}(\eta,H)$ is locally isomorphic to $\mathrm{O}(\eta,R)$. In addition, any element $a+\epsilon b \in \mathrm{U}(\eta,H)\cap\mathrm{O}(\eta,H)$ must satisfy

$$\begin{aligned} (a+\epsilon b)(a^T + \epsilon b^T) &= I, \\ (a+\epsilon b)(a^T - \epsilon b^T) &= I, \end{aligned} \tag{14}$$

and $(a+\epsilon b)^{-1}$ should exist. Therefore, $a^T + \epsilon b^T = a^T - \epsilon b^T$, i.e., $b = 0$.

Q.E.D.

We can clearly sum up the above discussions by the following graph, which expresses the close relations among every principal HCLG and RLG. These relations do not exist for CLG's.




(15)

## IV. ON LOCAL GAUGE TRANSFORMATION ACTION OF A HCLG

According to Sec. III, the local HCLG gauge symmetry of a hyperbolic complex field, in fact, is a local RLG gauge symmetry. Now, we generaly prove the local HCLG gauge symmetry of a hyperbolic complex sesquilinear form $K$ to be equivalent to the local RLG gauge symmetry of a real bilinear form $K'$. Let $T_x(H)$ and $T_x$ denote the hyperbolic complex and real tangent spaces at a point $x$ of $M$, respectively. A vector $A'$ of $T'_x = T_x \times T_x$ can be written as follows:

$$A' = A^A e'_A = A^\alpha e'_\alpha + A^{\overline{\alpha}} e'_{\overline{\alpha}}, \tag{16}$$

where we use the notation of Ref. 3, $\{e'_A\} = \{e'_\alpha, e'_{\overline{\alpha}}\}$ is a basis of $T'$ in which a hyperbolic complex structure $E$ takes the form as

$$E = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}. \tag{17}$$

A mapping $H: T' \to T(H)$ is defined by

$$H(A') = (A^\alpha + \epsilon A^{\overline{\alpha}}) e_\alpha, \tag{18}$$

where $\{e_\alpha\}$ spans $T_x$. Let $P: \mathrm{GL}(n,R) \times \mathrm{GL}(n,R) \to \mathrm{GL}(\eta,H)$ be an induced mapping of $\rho$ of Theorem 3, which is defined in a neighborhood $U$ of the unit element. Then $P$ is a local isomorphism. For an element $(g,h) \in U$, a transformation $L(g,h): T' \to T'$ is defined by $L(g,h) = H^{-1} P(g,h) H$, i.e.,

$$HL(g,h)(A') = P(g,h) H(A'), \quad \forall A' \in T'_x, (g,h) \in U. \tag{19}$$

Let $K'$ be a real bilinear form on $T'$. According to Kunstatter, Moffat, and Malzan,[3] we define a hyperbolic complex sesquilinear form $K$ on $T(H)$ by

$$K(A,B) = K'(A',B') + \epsilon K'(EA',B'), \tag{20}$$

where $A = H(A')$, $B = H(B')$.

**Theorem 8:** $K$ is invariant under the action of $P(g,h)$ if and only if $K'$ is invariant under the action of $L(g,h)$.

*Proof:* From Eq. (19) we have

$$\begin{aligned} &K(P(g,h)(A), P(g,h)(B)) \\ &\quad = K(P(g,h) H(A'), P(g,h) H(A')) \\ &\quad = K(HL(g,h)(A'), HL(g,h)(B')) \\ &\quad = K'(L(g,h)(A'), L(g,h)(B')) \\ &\qquad + \epsilon K'(EL(g,h)(A'), L(g,h)(A')). \end{aligned}$$

Since

$$EL(g,h) = L(g,h) E, \tag{21}$$

we obtain

$$\begin{aligned} &K(P(g,h)(A), P(g,h)(B)) \\ &\quad = K'(L(g,h)(A'), L(g,h)(B')) \\ &\qquad + \epsilon K'(L(g,h) E(A'), L(g,h)(B')). \end{aligned} \tag{22}$$

This means that the above theorem is true. Q.E.D.

The following Theorems 9, 10, and 11 are direct corollaries of Theorems 6, 4, and 5, respectively.

**Theorem 9:** Let $\{V_\alpha\}$ be an $\mathrm{U}(\eta,H)$ frame[4] on $M$, then the hyperbolic complex metric,[3] which is defined by

$$g_{\alpha\beta} = \eta_{ij} V^i_\alpha \tilde{V}^j_\beta, \tag{23}$$

has a local $\mathrm{GL}(n,R)$ gauge symmetry.

When $n = 4$ and $\eta = \mathrm{diag}(1, -1, -1, -1)$, we obtain the case in the nonsymmetric gravitational theory.[3,4]

**Theorem 10:** Let $\{V_\alpha\}$ be an $\mathrm{O}(\eta,H)$ frame on $M$, then a hyperbolic complex bilinear form

$$K_{\alpha\beta} = \eta_{ij} V^i_\alpha V^j_\beta \tag{24}$$

has a local $\mathrm{O}(\eta,R) \times \mathrm{O}(\eta,R)$ gauge symmetry.

However, $\tilde{K}_{\alpha\beta} \neq K_{\beta\alpha}$, $K$ cannot be explained as a metric.

**Theorem 11:** A hyperbolic complex bilinear form

$$K_{\alpha\beta} = \delta_{ab} (V^a_\alpha V^{m+b}_\beta - V^{m+a}_\alpha B^b_\beta B^b_\beta) \tag{25}$$

has a local $\mathrm{Sp}(2m,R) \times \mathrm{Sp}(2m,R)$ gauge symmetry.

## V. CONCLUSION

Every principal HCLG is (locally) isomorphic to some RLG (or a direct product of two PLG's). Therefore, a hyperbolic complex field usually has the local RLG gauge symmetry. An ordinary complex field generally does not have this property, e.g., the complex metric of Einstein[5] has the local U(3,1) gauge symmetry, however, it is not locally $\mathrm{GL}(4,R)$ gauge symmetric. In short, a discussion about the hyperbolic complex fields may usually be changed into a discussion about the real fields.


[1] I. M. Yaglom, *Complex Numbers in Geometry* (Academic, London, 1968).
[2] I. M. Yaglom, *A Simple Non-euclidean Geometry and Its Physical Basis* (Springer-Verlag, New York, 1979), Supplement B.
[3] G. Kunstatter, J. W. Moffat, and R. Malzan, J. Math. Phys. **24**, 886 (1983).
[4] Zai-Zhe Zhong, J. Math. Phys. **25**, 3538 (1984).
[5] A. Einstein, *The Meaning of Relativity* (Princeton U.P., Princeton, NJ, 1955), Appendix II.

# Infinitesimal null isotropy and Robertson–Walker metrics

Lisa Koch-Sen
*Department of Mathematics, State University of New York at Stony Brook, Stony Brook, New York 11794*

(Received 8 June 1984; accepted for publication 3 August 1984)

The concept of infinitesimal null isotropy is defined for a Lorentz manifold, in terms of null sectional curvature (as defined by Harris). It is shown that infinitesimal null isotropy is equivalent to infinitesimal spatial isotropy (as defined by Karcher), and that a null-isotropic space for which null sectional curvature is infinitesimally spatially constant must have a Robertson–Walker metric.

## I. INTRODUCTION

Astronomical observation has shown that the various sorts of extragalactic objects are distributed about us (approximately) isotropically. This situation is generally interpreted as saying that, for an observer with a certain four-velocity $U$, every direction in his infinitesimal rest space $U^\perp$ is equivalent. Moreover, it seems natural to assume that we do not occupy a special position in the universe; that is, space-time is isotropic at every point, as seen by the members of an appropriate family of observers. Karcher[1] has given the following definition.

Let $U$ be a timelike unit vector field on a Lorentz manifold $M$. Then $M$ is infinitesimally spatially isotropic relative to $U$ if

$$R(X,Y)Z = k[\langle Y,Z\rangle X - \langle X,Z\rangle Y], \quad \forall X,Y,Z\in U^\perp, \text{ (K1)}$$

$$R(X,U)U = \mu X, \quad \forall X\in U^\perp, \text{ (K2)}$$

where $R$ is the Riemann curvature tensor of $M$, and $k$ and $\mu$ are real-valued functions on $M$.

There are a number of theorems which assert that spatial isotropy at each point, together with some assumptions about the matter content of the universe, implies that space-time has a Robertson–Walker metric. For example, see Karcher,[1] Frankel,[2] and Robertson.[3]

Let us reexamine this interpretation of isotropy. What our observers actually see to be isotropic is the light (and other radiation) which has come to them from distant objects. Since light travels along null geodesics, a more accurate formulation of isotropy would be to say that, for our observers, every *null* direction is equivalent. As before, we require that this be the case for each observer in an appropriate family of observers. In order to put this idea in mathematical terms, we shall use the concepts of null sectional curvature and null congruence as defined by Harris.[4]

Given a nonzero null vector $N$ and a null plane $P$ containing $N$, the null sectional curvature with respect to $N$ of the plane $P$ is defined by

$$\kappa_N(P) = \langle R(V,N)N,V\rangle/\langle V,V\rangle,$$

where $V$ is any non-null (and therefore spacelike) vector in $P$. Here, $\kappa_N(P)$ is independent of the choice of $V$ in $P$, but it does depend quadratically on $N$. Therefore, it is best to restrict attention to a set of "normalized" null vectors which contains exactly one representative for each null direction.

Given a timelike unit vector field $U$ on $M$, the null congruence associated with $U$ is the set of null vectors defined by

$$\mathcal{N}(U) = \{N\in TM \mid \langle N,N\rangle = 0 \text{ and } \langle N,U\rangle = 1\},$$

where $TM$ is the tangent bundle of $M$. This set has the property that, for each nonzero null vector $N$, there is a unique $\lambda\in\mathbb{R}$ such that $\lambda N\in\mathcal{N}(U)$.

We shall call $M$ infinitesimally null-isotropic relative to $U$ if null sectional curvature, restricted to $\mathcal{N}(U)$, is a point function; that is, if for each point $p\in M$, $\nu(p) := \kappa_N(P)$ is the same for all null vectors $N$ at $p$ which lie in $\mathcal{N}(U)$ and all null planes $P$ containing $N$.

We shall prove that infinitesimal null isotropy relative to $U$ is equivalent to infinitesimal spatial isotropy relative to $U$, as defined by Karcher. Then we shall prove that if $M$ is null-isotropic relative to $U$ and if, in addition, null sectional curvature is infinitesimally spatially constant [i.e., $X(\nu) = 0$ for all $X\perp U$], then $M$ must be a Robertson–Walker space.

The proof of this second assertion follows closely the proof of the main theorem in Karcher.[1] Karcher's theorem states that $M$ is Robertson–Walker if (i) the matter content of $M$ is a perfect fluid obeying an equation of state, and (ii) $M$ is infinitesimally spatially isotropic relative to $U$, where $U$ is the flow vector field of the fluid. Our assumption that null sectional curvature be infinitesimally spatially constant corresponds to Karcher's assumption that the fluid obey an equation of state, or to the requirement[2] that the pressure of the fluid be spatially constant. Alternately, one could substitute the slightly stronger assumption that $U$ be geodesic ($D_U U = 0$), as in Robertson.[3] Harris[4] shows that $M$ is Robertson–Walker if $M$ is null-isotropic and $U$ is an infinitesimal conformal transformation ($L_U g = 2\alpha g$, where $g$ is the metric on $M$ and $\alpha: M\to\mathbb{R}$).

## II. INFINITESIMAL NULL ISOTROPY

Before showing that infinitesimal null isotropy is equivalent to infinitesimal spatial isotropy, we establish a lemma characterizing infinitesimal spatial isotropy in terms of the principal sectional curvatures of $M$. Recall that the sectional curvature of a non-null plane spanned by vectors $X$ and $Y$ is defined by

$$K(X\wedge Y) = K(X,Y) = \langle R(X,Y)Y,X\rangle/(\langle X,X\rangle\langle Y,Y\rangle - \langle X,Y\rangle^2).$$

We shall use the letter $U$ to denote both a fixed unit timelike vector field, and the value of that vector field at a point $p\in M$; the intent should be clear from context.

*Lemma:* Let $M$ be a Lorentz manifold of signature

 0022-2488/85/030407-04$02.50 © 1985 American Institute of Physics 

$(-,+,+,+)$. $M$ is infinitesimally spatially isotropic relative to $U$ if and only if all planes containing $U$ have sectional curvature $-\mu$, and all planes perpendicular to $U$ have sectional curvature $k$.

*Proof:* First, assume that $M$ is infinitesimally spatially isotropic relative to $U$. Then, for all $X,Y,Z \in U^{\perp}$,

$$R(X,Y)Z = k[\langle Y,Z\rangle X - \langle X,Z\rangle Y], \tag{K1}$$

$$R(X,U)U = \mu X, \tag{K2}$$

where $k$ and $\mu$ are real-valued functions on $M$. Letting $Y = Z$ in (K1) and taking the inner product with $X$ on both sides gives

$$\langle R(X,Y)Y,X\rangle = k(\langle Y,Y\rangle\langle X,X\rangle - \langle X,Y\rangle^2).$$

Dividing both sides by $\|X \wedge Y\|^2 = \langle X,X\rangle\langle Y,Y\rangle - \langle X,Y\rangle^2$ gives

$$K(X,Y) = k, \quad \text{for all } X,Y \in U^{\perp}.$$

Similarly, taking inner products with $X$ on both sides of (K2) and dividing by $\|U \wedge X\|^2$ gives

$$K(U,X) = -\mu, \quad \text{for all } X \perp U,$$

as was to be shown.

Conversely, assume for each $p \in M$ that $K$ is constant on planes containing $U_p$ and that $K$ is constant on planes perpendicular to $U_p$. Define $k: M \to \mathbf{R}$ and $\mu: M \to \mathbf{R}$ by

$$k(p) = K(X,Y), \tag{1a}$$

$$\mu(p) = -K(U_p,X), \tag{1b}$$

where $X$ and $Y$ are arbitrary linearly independent vectors in $U_p^{\perp}$. We must prove that the conditions (K1) and (K2) hold on $M$.

Condition (K1) follows from (1a) by a standard argument.[5] To obtain (K2), let $X$ and $Y$ be unit vectors perpendicular to $U$ with $X \perp Y$. We have

$$\begin{aligned}\langle R(X+Y,U)U,X+Y\rangle &= K(U,X+Y)\cdot\|(X+Y)\wedge U\|^2 \\ &= (-\mu)(-2) \\ &= 2\mu.\end{aligned}$$

On the other hand,

$$\begin{aligned}\langle R(X+Y,U)U,X+Y\rangle &= \langle R(X,U)U,X\rangle \\ &\quad + 2\langle R(X,U)U,Y\rangle \\ &\quad + \langle R(Y,U)U,Y\rangle \\ &= 2\mu + 2\langle R(X,U)U,Y\rangle.\end{aligned}$$

Combining these equations, we find that

$$\langle R(X,U)U,Y\rangle = 0,$$

whenever $X$ and $Y$ are orthogonal vectors in $U^{\perp}$. In addition,

$$\langle R(X,U)U,U\rangle = 0$$

and

$$\langle R(X,U)(U,X\rangle = \mu.$$

The last three equations together yield (K2).

**Theorem 1:** Let $U$ be a timelike unit vector field on $M$. Then $M$ is infinitesimally null-isotropic relative to $U$ if and only if $M$ is infinitesimally spatially isotropic relative to $U$.

*Proof:* Assume that $M$ is null-isotropic relative to $U$. By the lemma, it suffices to show that, at each point, sectional curvature is constant on all planes containing $U$ and on all planes perpendicular to $U$.

Given orthogonal vectors $X$ and $Y$ perpendicular to $U$, let $N = -U + X$. $N$ is a null vector with $\langle N,U\rangle = +1$, so $N \in \mathcal{N}(U)$. Since $\langle N,Y\rangle = 0$, $N$ and $Y$ span a null plane $P$ (see Fig. 1). By the definition of null sectional curvature, and since $\langle Y,Y\rangle = 1$,

$$\kappa_N(P) = \langle R(Y,N)N,Y\rangle.$$

Since $N = -U + X$,

$$\begin{aligned}\kappa_N(P) &= \langle R(Y,U)U,Y\rangle \\ &\quad - 2\langle R(Y,U)X,Y\rangle + \langle R(Y,X)(X,Y\rangle.\end{aligned} \tag{2}$$

On the other hand, let $N_1 = -U - X$. Then $N_1 \in \mathcal{N}(U)$ and $\langle N_1,Y\rangle = 0$; therefore $N_1$ and $Y$ span a null plane $P_1$. As above,

$$\begin{aligned}\kappa_{N_1}(P_1) &= \langle R(Y,U)U,Y\rangle + 2\langle R(Y,U)X,Y\rangle \\ &\quad + \langle R(Y,X)X,Y\rangle.\end{aligned}$$

Since $M$ is null-isotropic relative to $U$, $\kappa_N(P) = \kappa_{N_1}(P_1)$, it follows that $\langle R(Y,U)X,Y\rangle = 0$, so

$$\kappa_N(P) = \langle R(Y,U)U,Y\rangle + \langle R(Y,X)X,Y\rangle.$$

Since $-\langle U,U\rangle = \langle Y,Y\rangle = \langle X,X\rangle = 1$, and $\langle U,X\rangle = \langle X,Y\rangle = 0$, we can rewrite this as

$$\kappa_N(P) = -K(U,X) + K(X,Y);$$

that is, $K(U,X) = K(X,Y) - \nu$, where $\nu: M \to \mathbf{R}$ is the function defined by restricting $\kappa_N(P)$ to $N \in \mathcal{N}(U)$ and null planes $P$ containing $N$. Thus

$$K(Q_2) = K(Q_1) - \nu,$$

whenever $Q_1$ and $Q_2$ are planes with $U \in Q_1$, $U \perp Q_2$, and $Q_1 \wedge Q_2$ a line.

Now suppose that $U \wedge X$ and $U \wedge Y$ are any two planes containing $U$, where $X,Y \in U^{\perp}$ are independent, but not necessarily orthogonal (see Fig. 2). Then the plane $X \wedge Y$ meets both $U \wedge X$ and $U \wedge Y$ in a line, and therefore

$$K(U,X) = K(X,Y) - \nu = K(U,Y).$$

Thus $K$ is constant on all planes at $p$ containing $U$; we denote this constant by $-\mu(p)$. We can conclude also that $K$ has the constant value $k = -\mu + \nu$ on planes at $p$ perpendicular to $U$. Hence, by the lemma, $M$ is infinitesimally spatially isotropic relative to $U$.

Conversely, assume that $M$ is infinitesimally spatially isotropic relative to $U$, so that the conditions

$$R(X,Y)Z = k(\langle Y,Z\rangle X - \langle X,Z\rangle Y), \tag{K1}$$

$$R(X,U)U = \mu X \tag{K2}$$

hold for all $X,Y,Z \in U^{\perp}$, and let $P$ be any null plane. Then

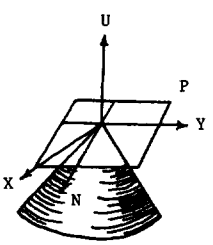



FIG. 1. The null plane $p$ spanned by $N$ and $Y$, where $N = -U + X$.

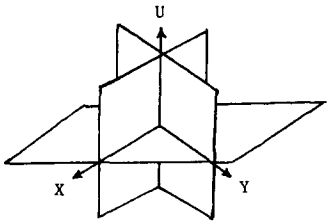



FIG. 2. The plane $X \wedge Y$ meets both $U \wedge X$ and $X \wedge Y$ in a line.

$P = N \wedge Y$, where $N \in \mathcal{N}(U)$ and $\langle Y,N \rangle = \langle Y,U \rangle = 0$. As before, $N = -U + X$ for some unit vector $X$ perpendicular to $U$, and $\langle X,Y \rangle = \langle N + U,Y \rangle = 0$. Recall the expansion of $\kappa_N(P)$ given in Eq. (2):

$$\kappa_N(P) = \langle R(Y,U)U,Y \rangle - 2\langle R(Y,U)X,Y \rangle + \langle R(Y,X)X,Y \rangle. \quad (2)$$

From (K1) and the symmetries of the curvature tensor, it follows that $\langle R(Y,U)X,Y \rangle = 0$, so

$$\begin{aligned}\kappa_N(P) &= \langle R(Y,U)U,Y \rangle + \langle R(Y,X)X,Y \rangle \\ &= -K(U,Y) + K(Y,X) \\ &= \mu + k,\end{aligned}$$

by (K2). Let $\nu = \mu + k$; then $\kappa_N(P) = \nu$ for all $N \in \mathcal{N}(U)$, so $M$ is null-isotropic relative to $U$.

## III. ROBERTSON–WALKER METRICS

Karcher[1] has shown that a perfect fluid space-time is conformally flat if and only if it is infinitesimally spatially isotropic. We show in the proof of Theorem 2 that an infinitesimally spatially isotropic spacetime is necessarily a perfect fluid; this, together with the result of Theorem 1, tells us that the null-isotropic space-times are exactly the conformally flat perfect fluids. These space-times are called Stephani universes[6]; they are a natural generalization of the Robertson–Walker spaces.[7]

**Theorem 2:** Let $M$ be a four-dimensional Lorentz manifold, and let $U$ be a timelike unit vector field on $M$. Suppose that $M$ is infinitesimally null-isotropic relative to $U$, and that null sectional curvature is nonzero and infinitesimally spatially constant. Then $M$ has a Robertson–Walker metric.

*Proof* (following Karcher[1]): By Theorem 1, we know that

$$R(X,Y)Z = k[\langle Y,Z \rangle X - \langle X,Z \rangle Y], \quad \text{(K1)}$$

$$R(X,U)U = \mu X, \quad \text{(K2)}$$

whenever $X,Y,Z \in U^\perp$, where $k$ and $\mu$ are functions on $M$. It follows that

$$\begin{aligned}&\mathrm{Ric}(X,Y) = (2k - \mu)\langle X,Y \rangle, \\ &\mathrm{Ric}(U,U) = 3\mu, \\ &\mathrm{Ric}(U,X) = 0,\end{aligned}$$

whenever $X,Y \in U^\perp$. Raising an index gives

$$\begin{aligned}&\widetilde{\mathrm{Ric}}(U) = -3\mu U, \\ &\widetilde{\mathrm{Ric}}|_{U^\perp} = (2k - \mu)\mathrm{id}|_{U^\perp}.\end{aligned}$$

The Einstein tensor, defined by

$$G := \widetilde{\mathrm{Ric}} - \tfrac{1}{2}\mathrm{tr}(\widetilde{\mathrm{Ric}})\mathrm{id},$$

has the property that $(\mathrm{div}\, G)(W) = 0$ for all $W \in TM$. From the above, we calculate $\frac{1}{2}\mathrm{tr}(\widetilde{\mathrm{Ric}}) = 3k - 3\mu$, so

$$\begin{aligned}&G(U) = -3kU, \\ &G|_{U^\perp} = (2\mu - k)\mathrm{id}|_{U^\perp},\end{aligned}$$

and $G(W) = (2\mu - k)W + 2(\mu + k)\langle U,W \rangle W$, for all $W \in TM$. (If the Einstein equations are assumed, this shows that the matter content of $M$ is a perfect fluid.) Now

$$0 = 2 + d\mu(W) - dk(W) + \langle U,W \rangle \mathrm{div}(2(\mu + k)U) + 2(\mu + k)\langle D_U U,W \rangle, \quad \forall W \in TM. \quad (3)$$

Returning to (K1) and (K2), we calculate

$$\begin{aligned}&D_U(R(X,Y)Z) = dk(U)(\langle Y,Z \rangle X - \langle X,Z \rangle Y) + k(\langle Y,Z \rangle D_U X - \langle X,Z \rangle D_U Y), \\ &R(X,Y)D_U Z = 0, \\ &R(D_U X,Y)Z = -\mu\langle Y,Z \rangle D_U X, \\ &R(X,D_U Y)Z = \mu\langle X,Z \rangle D_U Y.\end{aligned}$$

Thus

$$(D_U R)(X,Y)Z = dk(U)[\langle Y,Z \rangle X - \langle X,Z \rangle Y] + (k + \mu)[\langle Y,Z \rangle D_U X - \langle X,Z \rangle D_U Y].$$

Similarly,

$$(D_X R)(U,Y)Z = -d\mu\langle Y,Z \rangle U - (k + \mu)[\langle Y,Z \rangle D_X U - \langle X,Z \rangle D_Y U].$$

From the Bianchi identity,

$$(D_U R)(X,Y) + (D_X R)(Y,U) + (D_Y R)(U,X) = 0;$$

it follows that

$$[d\mu(X)\langle Y,Z \rangle - d\mu(Y)\langle X,Z \rangle]U = (k + \mu)[\langle Y,Z \rangle D_U X - \langle X,Z \rangle D_U Y]. \quad (4)$$

$$dk(U)[-\langle Y,Z \rangle X + \langle X,Z \rangle Y] = (k + \mu)[\langle Y,Z \rangle D_X U - \langle X,Z \rangle D_Y U - \langle D_X U,Z \rangle Y + \langle D_Y U,Z \rangle X]. \quad (5)$$

From (4) with $Y = Z \perp X$, it follows that

$$d\mu(X) = (k + \mu)\langle D_U X,U \rangle = -(k + \mu)\langle X,D_U U \rangle. \quad (6)$$

On the other hand, from (3) with $X = W \perp U$, we have

$$2\,d\mu(X) - dk(X) + 2(\mu + k)\langle X,D_U U \rangle = 0. \quad (7)$$

Equations (6) and (7) together yield

$$dk(X) = 0, \quad \text{for all } X \perp U. \quad (8)$$

Now let $S := -\frac{1}{2}(DU + DU^{\mathrm{trans}})|_{U^\perp}$ be the symmetric part of $DU|_{U^\perp}$. Setting $Y = Z \perp X$ in (5) and taking the $X$-component of both sides, we find

$$\frac{-dk(U)}{k + \mu} = \langle D_X U,X \rangle + \langle D_Y U,Y \rangle, \quad (9)$$

whenever $X$ and $Y$ are orthogonal vectors in $U^\perp$. ($k + \mu \neq 0$, since $k + \mu = \nu$ is null sectional curvature.) So we see that

$$S = \frac{1}{2} \cdot \frac{dk(U)}{k + \mu}\,\mathrm{id}|_{U^\perp}.$$

Then, letting $X,Y$, and $Z$ be orthogonal in (5) gives

$$0 = (k + \mu)\langle D_Y U,Z \rangle, \quad \text{for } Y \perp Z. \quad (10)$$

This tells us that $(DU - DU^{\text{trans}})|_{U^\perp} = 0$. Moreover, from (6) and (7) we have

$$(k+\mu)\langle D_X U, X\rangle = (k+\mu)\langle D_U U, X\rangle.$$

So $(k+\mu)\text{curl}(U) = 0$. Now $k+\mu = \nu$ where $\nu \neq 0$ by assumption, so $\text{curl}(U) = 0$. Thus we see that $U^\perp$ is integrable, and $S$ is the second fundamental form of the integral manifolds of $U^\perp$. From (8), we know that the integral manifolds of $U^\perp$ are also level sets of $k$. The Gauss equations now give

$$R|_{U^\perp}(X,Y)Z = \left(k + \frac{1}{4}\left(\frac{dk(U)}{k+\mu}\right)^2\right)(\langle Y,Z\rangle X + \langle X,Z\rangle Y),$$

so the integral manifolds of $U^\perp$ have constant curvature.

The condition $X(\nu) = 0$ [i.e., $d\mu(X) = 0$] for $X \perp U$ implies that $\nu$ is constant on the integral manifolds of $U^\perp$, so the gradient of $\nu$, $\text{grad}(\nu)$, is proportional to $U$. Here, $\text{grad}(k)$ is proportional to $U$ by (8); and, since $\nu = \mu + k$, we see that $\text{grad}(\mu)$ is proportional to $U$ also. Putting this information in (3), we can conclude that $D_U U$ is proportional to $U$. This, together with $\text{curl}(U) = 0$, gives $D_U U = 0$. So we see that the integral curves of $U$ are geodesics, whose perpendicular spaces integrate to form three-spaces of constant curvature. Thus the metric of $M$ can be written in the form

$$ds^2 = -dt^2 + f(t)d\sigma^2,$$

where $d\sigma^2$ is a metric of constant curvature. This is a Robertson–Walker metric.


## ACKNOWLEDGMENT

The author would like to thank Professor John A. Thorpe for suggesting the problem, and for his advice and support.



[1] H. Karcher, Arch. Math. **38**, 58 (1982).
[2] T. Frankel, *Gravitational Curvature* (Freeman, San Francisco, 1979), pp. 142–144.
[3] H. P. Robertson, Astrophys. J. **82**, 284 (1935).
[4] S. Harris, "A Characterization of Robertson–Walker Spaces by Null Sectional Curvature," to appear in Gen. Relativ. Gravit.
[5] S. Kobayishi and K. Nomizu, *Foundations of Differential Geometry, Interscience Tracts in Pure and Applied Mathematics, No. 15* (Wiley, New York, 1963), Vol. I, p. 203.
[6] H. Stephani, Commun. Math. Phys. **4**, 137 (1967).
[7] A. Krasinski, Gen. Relativ. Gravit. **15**, 673 (1983).

# High-accuracy approximation techniques for analytic functions

Charles Schwartz
*Department of Physics, University of California, Berkeley, California 94720*

(Received 13 June 1984; accepted for publication 21 September 1984)

A generalization of the familiar mesh point technique for numerical approximation of functions is presented. High accuracy and very rapid convergence may be obtained by thoughtful choice of the reference function chosen for interpolation between the mesh points. In particular, derivative operators are represented by highly nonlocal matrices; but this is no drawback when one has computing machines to perform the algebraic manipulations. Some examples are given from familiar quantum mechanical problems.

## I. INTRODUCTION

The most common approach to numerical approximation of continuous functions involves the representation of the function $f(x)$ by its values on a set of mesh (net, or lattice) points $x_n$. Familiar formulas for the derivative, in the simple case of a uniform mesh $x_n = x_0 + nh$, are

$$f'(x_n) = [f(x_n) - f(x_{n-1})]/h + O(h),$$

or

$$f'(x_n) = [f(x_{n+1}) - f(x_{n-1})]/2h + O(h^2),$$

and for the integral [writing $f(x_n) = f_n$]

$$\int_{x_0}^{x_N} f(x)dx = (x_N - x_0)\frac{1}{N}\sum_{n=1}^{N} f_n + \frac{h(f_0 - f_N)}{2} + O(h^2).$$

These are simple to derive and simple to use but they have a very low order of accuracy in general. This is due to the fact that only *local* information about the function $f$ is used in building the approximation.

The approach presented in this paper is based upon a *global* construction of an approximation for $f(x)$, which is still flexible and easy to use and involves only the values $f_n$ at the selected mesh points. The purpose is to achieve very high-accuracy approximations: with a total of $N$ mesh points it is nice to get errors which are as small as $A^{-N}$ (or even $1/N!$), rather than the $1/N$, $1/N^2$, etc., errors which are characteristic of the usual methods. In this sense the present approach is somewhat reminiscent of Gaussian quadrature; but it is rather more general in its construction and its application. The present method may also be described as a generalization of Lagrange interpolation; and the method of "collocation" is also related.

The general method will be described, along with a formal method for error analysis; then several examples will be given, mostly concerned with solving differential equations familiar in quantum mechanics.

## II. THE GENERAL METHOD

To approximate a given function $f(x)$ we start by choosing a reference function $u(x)$ that has simple zeros at the (real) points $x = x_n$. The construction of an interpolating function $\bar{f}(x)$ to approximate $f(x)$ is

$$\bar{f}(x) \equiv \sum_m f_m \frac{u(x)}{x - x_m}\frac{1}{a_m}, \quad \text{where } a_n = u'(x_n). \tag{1}$$

At the points $x = x_n$, $\bar{f}(x)$ takes on the values $f_n = f(x_n)$. We should choose the reference function $u(x)$ to have analytic properties similar to those of the desired function $f(x)$; the error analysis and examples to follow will help show what this means.

To approximate the derivatives of the function $f$, we take derivatives of the interpolating function (1), evaluated at the mesh points $x_n$. The resulting formulas are [$b_n = u''(x_n)$ and $c_n = u'''(x_n)$]

$$\left.\frac{d\bar{f}}{dx}\right|_{x_n} = \sum_m f_m \begin{cases} m = n: & \dfrac{b_n}{2a_n} \\ & = \displaystyle\sum_m D_1(n, m) f_m, \\ m \neq n: & \dfrac{1}{(x_n - x_m)}\dfrac{a_n}{a_m} \end{cases} \tag{2}$$

$$\left.\frac{d^2\bar{f}}{dx^2}\right|_{x_n} = \sum_m f_m \begin{cases} m = n: \dfrac{c_n}{3a_n} \\ m \neq n: \dfrac{1}{(x_n - x_m)}\dfrac{b_n}{a_m} - \dfrac{2}{(x_n - x_m)^2}\dfrac{a_n}{a_m} \end{cases}$$

$$= \sum_m D_2(n, m) f_m. \tag{3}$$

In case the function $u(x)$ obeys an equation of the form $u''(x) = W(x)u(x)$, then there is a simplification of the above formulas: $b_n = 0$, $c_n = W_n a_n$; and the matrices representing the derivative operators can be put into a symmetric form.

To approximate the integrals of $f$ we get the formulas

$$\int_{x_0}^{x_n} \bar{f}(x)dx = \sum_m Q_1(n, m) f_m,$$

$$\text{where } Q_1(n, m) = \int_{x_0}^{x_n} dx \frac{u(x)}{a_m(x - x_m)}, \tag{4}$$

$$\int_{x_0}^{x_n} dx \int_{x_0}^{x} dy\, \bar{f}(y) = \sum_m Q_2(n, m) f_m,$$

$$\text{where } Q_2(n, m) = \int_{x_0}^{x_n} dx \frac{u(x)(x_n - x)}{a_m(x - x_m)}. \tag{5}$$

In the case where $u(x)$ is an orthogonal polynomial times a weight function and the integral is taken over the entire domain, then (4) yields the usual Gaussian quadrature results.

The above general method is very flexible since one can choose any reference function $u(x)$. The quantities that enter into the matrices for the derivative operators ($D$) or integral operators ($Q$) may be determined by some computational procedure, if not readily expressed in closed form, depending

 0022-2488/85/030411-05$02.50 © 1985 American Institute of Physics 

on this choice of $u$. A practical question is the following: Does one pay a heavy price by having the derivative operator so nonlocal, since one may be forced to invert or otherwise manipulate these matrices in order to solve differential equations? A computing machine can readily carry out such matrix operations numerically for moderate-sized matrices. Furthermore, when one gets into partial differential equations the usual mesh point methods already require working with sizable matrices for the derivative operators. Since the whole point of the present method is to construct approximate but accurate functions $\bar{f}$ in terms of a *small* number of mesh points it is anticipated that the net result should be a general increase in efficiency of computation.

Now we present a general approach for analyzing the error in approximating the function $f$ by $\bar{f}$, once $u$ is chosen. Assume that both $f(x)$ and $u(x)$ are analytic functions in some appreciable domain of the complex plane surrounding the set of mesh points $x_n$. Then, using the contour around $z = x$ [see Fig. 1(a)] we have the identity

$$f(x) = \oint \frac{dz}{2\pi i} \frac{f(z)}{z - x} \frac{u(x)}{u(z)}. \tag{6}$$

One may take the point $x$ to be slightly off the real axis to be assured that there is no difficulty in this integral representation when $x$ approaches one of the mesh points $x_n$, where $u$ vanishes. Now move the contour of integration to the large loop C and the small circles around each of the points $z = x_n$ [see Fig. 1(b)]. Calculating the residues at each $x_n$ we have the exact result

$$f(x) = \sum_n f(x_n) \frac{u(x)}{(x - x_n)} \frac{1}{u'(x_n)} + \epsilon. \tag{7}$$

The first term on the right-hand side of (7) is just the approximation $\bar{f}(x)$ defined in (1); the second term $\epsilon$ is the error and is given by the integral over the contour C of the expression (6). A general argument about the smallness of this error is as follows: Since $u(z)$ has many oscillations along the real axis, one expects it to grow rapidly along the imaginary directions in the $z$ plane; and it is this factor in the denominator that should make the error $\epsilon$ decrease rapidly as the mesh points become more closely spaced. A concrete example will be studied in the next section.

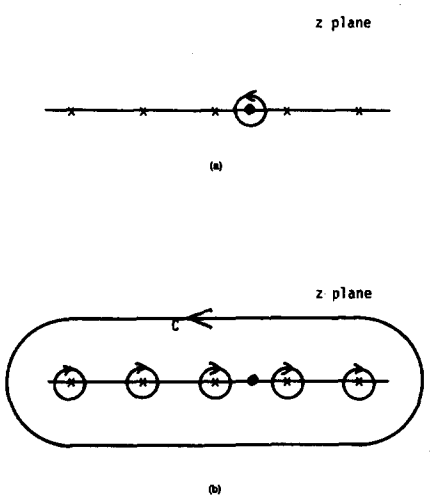



FIG. 1. Integration contours for error analysis. The several crosses × represent the mesh points $x_n$. The solid dots ● represent the point $z = x$.

## III. EXAMPLE I

Consider the infinite line, $-\infty < x < \infty$, and the choice of a uniform sequence of mesh points: $x_n = nh$, $n = 0, \pm 1, \pm 2, \pm 3, \ldots$. Then we take the reference function $u(x) = \sin \pi x/h$; and the matrices for $d/dx$ and $d^2/dx^2$ become

$$D_1(n, m) = \frac{1}{h} \begin{cases} n = m: & 0, \\ n \neq m: & (-1)^{n-m}/(n-m); \end{cases} \tag{8}$$

$$D_2(n, m) = \frac{1}{h^2} \begin{cases} n = m: & -\pi^2/3, \\ n \neq m: & -2(-1)^{n-m}/(n-m)^2. \end{cases} \tag{9}$$

The matrices for the indefinite integrals become

$$Q_1(n, m) = (h/\pi)[S_i((n-m)\pi) + \pi/2], \tag{10}$$

$$Q_2(n, m) = (h^2/\pi^2)[(n-m)\pi(S_i((n-m)\pi) + \pi/2) + (-1)^{n-m}], \tag{11}$$

where

$$S_i(x) = \int_0^x dt \frac{\sin t}{t}. \tag{12}$$

The only familiar result contained here is for the infinite integral

$$Q_1(-\infty, \infty) = h, \quad \int_{-\infty}^{\infty} f(x)dx = \sum_{n=-\infty}^{\infty} hf(nh) + \epsilon. \tag{13}$$

The high accuracy of the trapezoidal rule for the infinite integration of analytic functions has been explored elsewhere.[1] This is the "Gaussian quadrature" formula for the infinite line.

Obviously, if this approach is to be practical, we should be dealing with functions $f(x)$ which decrease very rapidly as $x$ grows large, so that the infinite sums over the mesh points can be truncated effectively. Thus we have two sources of error to analyze: $\epsilon_A$ from (7) due to the analytic approximation and $\epsilon_T$ due to the truncation. A good strategy will be to choose a relation between the mesh spacing $h$ and the truncation at $|n| < N$ so that $\epsilon_A$ and $\epsilon_T$ are approximately equal to each other. This will avoid wasting effort on too small a mesh (when truncation error dominates) or on too large a cutoff (when mesh error dominates).

For illustration, consider that the function $f(x)$ is known to be analytic everywhere in the finite complex plane and is dominated at large distances by the behavior

$$e^{-ax^p}. \tag{14}$$

Then we have

$$\epsilon_T \approx e^{-a(Nh)^p}. \tag{15}$$

For the mesh size error, we see that the error in (7) involves the integral over the large contour C in Fig. 1(b); and with $u(z) = \sin \pi z/h$, we see that this error is given roughly by

$$\epsilon_A \approx \int dz\, e^{\pm i\pi z/h} e^{-az^p}. \tag{16}$$

This integral may be estimated by the stationary phase method (we are interested in the dependence of $\epsilon_A$ on $h$ for small $h$) and we find

$$\epsilon_A \approx e^{-bh^{-q}}, \quad \text{where } q = p/(p-1), \tag{17}$$

and

$$b=\left(\frac{\pi^p}{ap}\right)^{1/(p-1)}\left(\frac{p-1}{p}\right)\sin\left[\pi\frac{1}{2(p-1)}\right].$$

Equating the results (15) and (17) we find the optimum choice of $h$, given $N$,

$$h=(b/aN^p)^{((p-1)/p^2)}, \tag{18}$$

and along with this is the error estimate

$$\epsilon\approx e^{-CN}, \tag{19}$$

where

$$C=b(a/b)^{1/p}.$$

This result—exponential decrease of the error with increasing number of mesh points—is most exciting. Rather than trying to make this rough error analysis more respectable I shall proceed to some numerical experiments.

The one-dimensional Schrödinger equation

$$\left(-\frac{1}{2}\frac{d^2}{dx^2}+\frac{1}{k}x^k\right)\psi(x)=E\psi(x), \tag{20}$$

for $k=2,4,\ldots$, is an eigenvalue problem in which the solution $\psi(x)$ has the characteristics described above: it is an analytic function for all finite $x$ and has the asymptotic behavior for large $x$ given by (14) with

$$p=\frac{k+2}{2},\quad a=\frac{2}{k+2}\left(\frac{2}{k}\right)^{1/2}. \tag{21}$$

Thus we predict the optimum convergence strategy, (18) and (19):

$$h=(\pi/N)^{1/2},\quad \epsilon\approx e^{-1.57N},\quad \text{for } k=2, \tag{22}$$

and

$$h=1.78N^{-2/3},\quad \epsilon\approx e^{-1.32N},\quad \text{for } k=4. \tag{23}$$

Taking account of the symmetry, $\psi(x)=\pm\psi(-x)$, and choosing the mesh points $x_n=(n-1/2)h$, for $n=1,2,\ldots,N$, and using (9) for the second derivative operator, Eq. (20) was represented as an $N\times N$ matrix eigenvalue problem which the computer solved for the sequence $N=1,2,3,\ldots$.

The values of $h$ were chosen according to (22) and (23) with $(N+1/2)$ replacing $N$. The numerical results for the ground state eigenvalue showed very rapid convergence:

$k=2$, $E=0.5$:

$N=1$ error $7\times10^{-2}$, $N=2$ error $3\times10^{-3}$,

$N=3$ error $2\times10^{-4}$, $N=4$ error $1\times10^{-5}$,

with a good fit to the formula $\epsilon\approx e^{-2.9N}$; (24)

$k=4$, $E=0.420\,804\,974\,475$:

errors of $-7\times10^{-2}$, $6\times10^{-3}$, $2\times10^{-4}$,

$7\times10^{-6}$ for $N=1,2,3,4$,

with a good fit to the formula $\epsilon\approx e^{-2.8N}$. (25)

These are very gratifying results: high accuracy at low-order approximation with very rapid improvement as the order of approximation is increased. Indeed, these numerical results for the $x^4$ potential converge even more rapidly than the results of a Rayleigh–Ritz variational calculation that used a harmonic oscillator basis.[2] The predicted exponential form of convergence (19) was well borne out by the numerical results; however, there is some discrepancy between the predicted and observed values of the decay constant $C$. The decay constants in (22) and (23) describe the error in the wave function and perhaps one ought to square these errors for the eigenvalue: the values $2C=3.14$ $(k=2)$ and $2C=2.64$ $(k=4)$ are not so far from the observed results 2.9 and 2.8, respectively.

In the computations described above the matrix eigenvalue was determined by a direct numerical method (which requires computing time proportional to the cube of the dimension of the matrix). For these one-dimensional problems the size of the matrix is so small that this is no problem. However, when one envisions going to multidimensional problems with a much larger dimension for the matrix of the partial differential operators involved then some alternative method of manipulating the matrix may be necessary. There are a variety of iterative techniques commonly used for large matrix manipulations (inversion, diagonalization, etc.) and the critical question is how fast such iterative methods converge. As an experiment I tried solving the above-mentioned Schrödinger equation iteratively by a few different strategies and found convergence that varied from fair (about 1/2 decimal accuracy gained per iteration) to very good (several decimals gained per iteration.) As with all iterative schemes it is valuable to have a good starting guess for the solution; and the attempts I made worked best when I used the resulting eigenvector for the solved $N-1$ problem to get a starting estimate for the $N$-problem eigenvector through use of the basic interpolation formula (1).

## IV. EXAMPLE II

For a problem on the semi-infinite line $0\leqslant r<\infty$ consider the Schrödinger equation for the hydrogen atom:

$$\left[-\frac{1}{2}\frac{d^2}{dr^2}+\frac{1}{2}\frac{l(l+1)}{r^2}-\frac{1}{r}\right]\phi(r)=E\phi(r). \tag{26}$$

At the origin $\phi$ goes to zero as $r^{l+1}$ and at infinity it goes exponentially to zero for bound states ($E$ negative eigenvalues).

To choose a good reference function $u(r)$ we would like a function which has analytic properties similar to $\phi$ for finite $r$ and also has many zeroes. It is known that the solution of (26) for $E=0$ is given in terms of a Bessel function:

$$\phi_{E=0}(r)=r^{1/2}J_{2l+1}((8r)^{1/2}). \tag{27}$$

This leads to the choice

$$u(r)=r^{1/2}J_{2l+1}\left[\left(\frac{8r}{h}\right)^{1/2}\right], \tag{28}$$

with the mesh points

$$r_n=(h/8)y_n^2,\quad J_{2l+1}(y_n)=0,\quad n=1,2,3,\ldots. \tag{29}$$

With the change of variables

$$\phi_m=\chi_m J'_{2l+1}(y_m), \tag{30}$$

we reduce the differential equation (26) to the algebraic form

$$\sum_{m\neq n}\frac{64}{h^2}\frac{1}{(y_n^2-y_m^2)^2}\chi_m+\frac{8}{3h^2}\left[\frac{8l(l+1)}{y_n^4}+\frac{1}{y_n^2}\right]\chi_n-\frac{8}{hy_n^2}\chi_n=E\chi_n. \tag{31}$$

Numerical computations of the ground state eigenvalue ($l = 0$) were carried out for a sequence of mesh scales ($h = 1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}$) and a sequence of truncations ($m, n = 1,2,3,...,N$, for $N = 1,2,3,4,...$). At each $h$ value the error would decrease rapidly with increasing $N$ (about one decimal place improvement per unit step in $N$) until it reached a saturation value; then one would need to decrease $h$ to gain further improvement. Taking the results from these saturation points one could deduce an overall convergence rate which went approximately as

$$\epsilon \approx 10^{-N}. \tag{32}$$

This is an experimental result; I have not tried (as in the previous section) to carry out an analytical estimate of the expected error. This is a very rapid convergence rate, indicating that this is an extremely powerful approximation technique for atomic wave functions.

The major nuisance in this calculation was the need to generate zeroes of a Bessel function. As an alternative, I tried using the reference function

$$u(r) = r^{l+1/2} \sin[\pi(r/h)^{1/2}], \tag{33}$$

with the mesh points given by

$$r_n = hn^2, \quad n = 1,2,\ldots. \tag{34}$$

I will not give details but merely state the results of this approach. The matrix turns out to be unsymmetrical but this poses no serious problem. The ground state eigenvalue computation converges quite well, only slightly slower than the first approach:

$$\epsilon \approx 10^{-(2/3)N}. \tag{35}$$

## V. EXAMPLE III

For problems on a finite interval one usually works with either polynomials or Fourier series as a basis for approximations. I will give a couple of illustrations based upon the latter.

Suppose we want to approximate the function $f(x)$ on the interval $[0, 1]$ with the boundary conditions $f(0) = f(1) = 0$. One choice of the reference function is

$$u(x) = \sin(N+1)\pi x, \tag{36}$$

which satisfies the same boundary conditions as $f(x)$ and has the interior mesh points

$$x_n = n/(N+1), \quad n = 1,2,3,\ldots,N. \tag{37}$$

(Here life is simpler since we do not have to deal with two variables, $h$ and $N$, but only one, $N$.)

If we follow the original prescription for building the approximation (1), then we will have functions that are not simply a finite set of trigonometric functions. An alternative is to divide $u(x)$ by something like $\sin\alpha(x - x_n)$, rather than just $(x - x_n)$. After some experimentation I was able to find the following representation, which is equivalent to a truncated Fourier series:

$$\bar{f}(x) = \sin(N+1)\pi x \sum_{n=1}^{N} f_n \frac{(-1)^n}{2(N+1)} \times [\cot(\pi/2)(x - x_n) - \cot(\pi/2)(x + x_n)]. \tag{38}$$

From this the second derivative was calculated to be

$$\bar{f}''|_{x_n} = \sum_{m=1}^{N} f_m \pi^2 \begin{cases} m = n: & -\frac{1}{3}(N+1)^2 - \frac{1}{6} + \frac{1}{2}\csc^2\pi\frac{n}{N+1}, \\ m \neq n: & \frac{(-1)^{n+m}}{2}\left[-\csc^2\left(\frac{\pi}{2}\frac{n-m}{N+1}\right)\right] + \csc^2\left(\frac{\pi}{2}\frac{n+m}{N+1}\right). \end{cases} \tag{39}$$

The eigenvalues of this matrix (39), in units of $-\pi^2$, are

$$1 \text{ (for } N = 1); \quad 1,4 \text{ (for } N = 2); \quad 1,4,9 \text{ (for } N = 3);\ldots. \tag{40}$$

An alternative problem is one with periodic boundary conditions:

$$f(\varphi + 2\pi) = f(\varphi). \tag{41}$$

For $N$ *odd* we construct the approximate function

$$\bar{f}(\varphi) = \sin\frac{N}{2}\varphi \sum_{n=1}^{N} \frac{f_n}{\sin\frac{1}{2}(\varphi - \varphi_n)} \frac{(-1)^n}{N}, \tag{42}$$

with mesh points

$$\varphi_n = 2\pi n/N; \tag{43}$$

and the second derivative operator is represented by

$$\bar{f}''|_{\varphi_n} = \sum_{m=1}^{N} f_m \begin{cases} m = n: & -\frac{1}{12}(N^2 - 1), \\ m \neq n: & \frac{(-1)^{n-m+1}}{2} \frac{\cos\phi_{nm}}{\sin^2\phi_{nm}}, \end{cases} \tag{44}$$

where

$$\phi_{nm} = (n - m)\pi/N.$$

This matrix has the expected eigenvalues: 0, $-1$(twice), $-4$(twice), etc.

For a numerical application I considered the problem of the Schrödinger pendulum:

$$\left[-\frac{1}{2}\frac{d^2}{d\theta^2} + g^2(1 - \cos\theta)\right]\psi(\theta) = E\psi(\theta). \tag{45}$$

Using (44), the two lowest eigenvalues were computed for a sequence of values of $N$, for two different values of $g$. No account was taken of the reflection symmetry. Results, shown in Table I, exhibit the fastest convergence yet seen. The calculation was repeated shifting the coordinate in (42) by 90° [actually, by changing $\cos\theta$ to $\sin\theta$ in (45)] and these results were even better, by up to two decimal places accuracy at each $N$. For comparison, a variational calculation of (45) using a truncated Fourier series with corresponding number of terms gave results which were in between those of the two computations just described.

Some previous work on trigonometric interpolation of periodic functions[3] bears resemblance to what has been presented here; but the formula (44) appears to be new. I will

TABLE I. Eigenvalues of the Schrödinger pendulum, Eq. (45).

| | $g=1$ | | $g=3$ | |
|---|---|---|---|---|
| $N$ | $E_0$ | $E_1$ | $E_0$ | $E_1$ |
| 3 | 0.29 | 1.71 | 0.33 | 13.7 |
| 5 | 0.457 | 1.382 | 0.89 | 7.09 |
| 7 | 0.464 86 | 1.343 98 | 1.33 | 4.97 |
| 9 | 0.464 934 9 | 1.343 362 9 | 1.455 | 4.43 |
| 11 | 0.464 935 147 34 | 1.343 360 133 | 1.467 3 | 4.345 |
| 13 | 0.464 935 147 7119 | 1.343 360 128 403 | 1.468 031 | 4.337 52 |
| 15 | 0.464 935 147 7122[a] | 1.343 360 128 3991[a] | 1.468 053 5 | 4.337 179 2 |
| 17 | | | 1.468 054 007 | 4.337 170 39 |
| 19 | | | 1.468 054 013 55 | 4.337 170 257 1 |
| 21 | | | 1.458 054 013 609 | 4.337 170 255 64 |

[a] Machine accuracy not reliable after this point.

confess, however, that the formulas (44) and (39) were first obtained by Fourier transform calculation.

## VI. SUMMARY

The general approach presented here should be very powerful in obtaining efficient and accurate numerical computational results in the form of systematic approximations to functions that are very smooth. The high accuracy and rapid convergence usually associated with variational techniques is obtained along with the simplicity of mesh techniques. The key link between these two methods is the judicious choice of the reference function; here is where the human being contributes analytical insights in setting up the problem, while leaving the later computational tedium to the machine.

The numerical examples shown here were restricted to the solution of one-dimensional differential equations (eigenvalue problems); and the results were excellent. There should be many other areas of application for this general method of approximating analytic functions.


## ACKNOWLEDGMENTS

I have benefitted from conversations with Professor K. Miller and Professor A. Chorin.


## APPENDIX: THE SEMI-INFINITE INTERVAL

For the infinite interval $[-\infty, \infty]$ a general interpolation scheme was given, with uniform intervals, based upon trigonometric functions as used in Fourier integrals. For the finite interval problem, alternative schemes were again based upon trigonometric functions, this time as they are used in discrete Fourier series. What follows here is a generalization of the study for the semi-infinite interval $[0, \infty]$ based upon Bessel functions.

Choose the reference function, for unspecified value of $\nu$,

$$u(x) = x^{-\nu/2} J_\nu((x/h)^{1/2}), \tag{A1}$$

which has the mesh points $x_n = hy_n^2$, $n = 1,2,...$, where $y_n$ is the $n$th zero of the Bessel function $J_\nu(y)$ on the positive real axis. Next, construct the identity integral representation, as in (6):

$$f(x) = \oint \frac{dz}{2\pi i} \frac{u(x)}{u(z)} \frac{f(z)}{z-x}, \tag{A2}$$

with the contour a small circle around $z = x$. Now, move the contour following the same general procedure illustrated in Fig. 1. The form of (A1) was chosen so that $u(z)$ is analytic in the domain $\mathrm{Re}(z) \leqslant 0$ as well as $> 0$. Assuming $f(z)$ is analytic in some sizable region around the positive real axis, we expect exponentially small errors to the approximate interpolation function $\bar{f}(x)$ that results from the residues at each of the zeroes of $u(z)$:

$$\bar{f}(x) = \sum_n \frac{f(hy_n^2)}{x - hy_n^2} \frac{u(x)}{J'_\nu(y_n)} 2h^{1+\nu/2} y_n^{1+\nu}. \tag{A3}$$

From this one can calculate the definite integral

$$\int_0^\infty dx\, x^\nu \bar{f}(x) = \sum_n f(hy_n^2) 4h^{\nu+1} \left(\frac{y_n^\nu}{J'_\nu(y_n)}\right)^2. \tag{A4}$$

This is a new "Gaussian quadrature" formula, or rather a family of such for any value of $\nu$. In the special cases $\nu = \pm\frac{1}{2}$ this formula reduces to the trapezoidal rule (13). What is interesting about this formula is the fact that the points $x_n = hy_n^2$ at which one evaluates the function $f(x)$ are spaced farther and farther apart as $n$ increases.


[1] C. Schwartz, J. Comput. Phys. 4, 19 (1969).

[2] C. Schwartz, Ann. Phys. (NY) 32, 277 (1965).

[3] H. Kreiss and J. Olinger, "Methods for the Approximate Solution of Time Dependent Problems," GARP Publications Series No. 10, February 1973, p. 42 ff.

# A new way for solving Laplace's problem (the predictor jump method)

J. M. Vega-Fernández, J. F. Duque-Carrillo, and J. J. Peña-Bernal
*Cátedra de Física Médica, Facultad de Medicina, Universidad de Extremadura, Badajoz, Spain*

(Received 24 January 1984; accepted for publication 8 August 1984)

This paper presents a new method, that we call "the predictor jump," for driving to a faster solution of Laplace's equation. Some results obtained by applying this technique are compared with those that have been obtained by the traditional methods.

## I. INTRODUCTION

Laplace's equation

$$\nabla^2\phi = 0 \tag{1}$$

arises in many physical problems associated to scalar fields (electrostatic and gravitational) and in a large number of steady-state cases (flow of heat, diffusion of a solute, flow of an incompressible fluid, and so on).

The analytic methods for solving this equation are not very useful when the physical system's geometry or the boundary conditions are somewhat complicated. In these cases, some of the numerical methods (Gauss–Siedel, SOR, etc.) are used as alternative ways.[1–5]

In this paper, we present a technique which has been called the predictor jump method (PJM), that can be combined with any of the relaxation methods for driving to a faster solution of Laplace's equation. In particular, the PJM has been used in this paper to calculate the distribution of temperatures in solids with different geometries. The results obtained with PJM are compared with those that have been obtained by using the traditional relaxation methods for showing the advantages of the proposed technique.

## II. NUMERICAL METHODS

To indicate the difference between the PJM and the classical numerical methods used for solving Laplace's equation it is convenient to summarize some basic ideas about them.

The general expression of Laplace's equation for solving problems by applying numerical methods in Cartesian coordinates is[6] at a point

$$\frac{\phi_{i-1,j,k} - 2\phi_{i,j,k} + \phi_{i+1,j,k}}{(\Delta x)^2} + \frac{\phi_{i,j-1,k} - 2\phi_{i,j,k} + \phi_{i,j+1,k}}{(\Delta y)^2} + \frac{\phi_{i,j,k-1} - 2\phi_{i,j,k} + \phi_{i,j,k+1}}{(\Delta z)^2} = 0, \tag{2}$$

$\phi_{i,j,k}$ being the value of the function $\phi$ at a point $(i,j,k)$ of a tridimensional $XYZ$ grid (Fig. 1).

This last equation is obtained by using a Taylor expansion and the central finite difference approximation for the first and second derivatives.

For a grid of $N$ points the problem is reduced to the resolution of $N$ equations which can be expressed by the formula

$$|A|_{N\times N}|\phi|_{N\times 1} = |B|_{N\times 1}, \tag{3}$$

$|A|$, $|B|$, and $|\phi|$ being the matrixes of the coefficients, independent terms, and unknown quantities, respectively. Every term of the matrix $|\phi|$ represents the function's value at a point of the grid. The matrix $|A|$ is banded and diagonally dominant, so that when iterative methods are applied, the convergence to the solution is insured.

Iterative methods start with some arbitrary values for the unknown quantities, and the variables are going to get values more and more approximate to the correct solution by applying the finite difference equations.

In practice, this process is finished when, between two successive iterations ($k-1$ and $k$), it is verified that

$$\text{Error} = E^{(k} = \sum_{i=1}^{N} |\phi_i^{(k} - \phi_i^{(k-1}| < \epsilon, \tag{4}$$

$\epsilon$ being as small as one wants.

One of the first iterative methods used was proposed by Jacobi.[7] The algorithm of this method can be expressed by the following formula:

$$\phi_i^{(k} = \frac{b_i}{a_{ii}} - \sum_{\substack{j=1 \\ j\neq i}}^{N} \frac{a_{ij}}{a_{ii}} \phi_j^{(k-1}, \quad k = 1,2,..., \tag{5}$$

that is to say, the values of the unknown quantities in the generic iteration number $k$, are calculated from the values in the former one.

On the other hand, in the Gauss–Seidel method, whose algorithm is

$$\phi_i^{(k} = \frac{b_i}{a_{ii}} - \sum_{j=1}^{i-1} \frac{a_{ij}}{a_{ii}} \phi_j^{(k} - \sum_{j=i+1}^{N} \frac{a_{ij}}{a_{ii}} \phi_j^{(k-1}, \quad k = 1,2,..., \tag{6}$$

the values of the unknown quantities in the iteration number $k$ are calculated not only from the values obtained in the

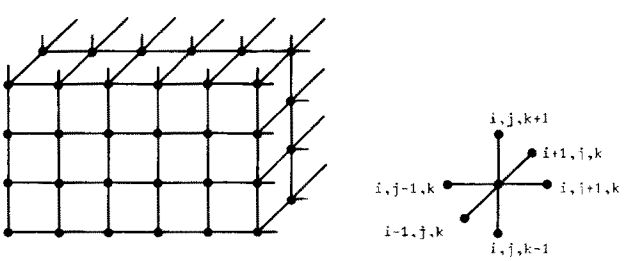



FIG. 1. Tridimensional $XYZ$ grid.

 0022-2488/85/030416-04$02.50 © 1985 American Institute of Physics 

precedent iteration, but also from the previously modified values in the same iteration.

The Gauss–Seidel method presents some advantages over the Jacobi method; its convergence is faster, it uses less computer memory and its programming is easier.

Equation (6) can be rewritten as

$$\phi_i^{(k} = \phi_i^{(k-1} + \frac{1}{a_{ii}} \left\{ b_i - \sum_{j=1}^{i-1} a_{ij} \times \phi_j^{(k} - \sum_{j=i}^{N} a_{ij} \phi_j^{(k-1} \right\}, \tag{7}$$

which is transformed in another algorithm by multiplying the second summand of the last expression by the factor $\omega$, so that

$$\phi_i^{(k} = \phi_i^{(k-1} + \frac{\omega}{a_{ii}} \left\{ b_i - \sum_{j=1}^{i-1} a_{ij} \times \phi_j^{(k} - \sum_{j=i}^{N} a_{ij} \phi_i^{(k-1} \right\}. \tag{8}$$

When $\omega$ has a value between 1 and 2, this last algorithm improves the rate of convergence and the method is called SOR (successive over relaxation).[7,8]

At each particular problem, there is an optimum value for the relaxation factor, named $\omega_{opt}$, with which the number of the necessary iterations to get the correct solution is minimum.[9]

In the bibliography, there are some approximate formulas proposed for $\omega_{opt}$ which have been empirically obtained and used for solving some specific problems.[10,11] Unfortunately, these formulas are very tedious to calculate. Moreover, the number of times that it must be applied to solve a specific problem for studying the influence of each parameter is so high that when one is interested in solving a few times the same model of a problem the method is not practical.

Other methods[12] calculate the value of $\omega_{opt}$ automatically after a certain number of iterations; for instance, SORACO[13] (successive over-relaxation with automatic convergence optimized). Such methods define a convergence criterion previously and according to it, values of $\omega$ are compared until $\omega_{opt}$ is reached. Although a lot of time may be wasted in the research of $\omega_{opt}$, in practice SORACO is one of the most effective iterative methods.

## III. THE PREDICTOR JUMP METHOD

We propose another technique which can be used with any of the other iterative methods, but it starts from a different idea. Instead of finding the solution of Eq. (2) by successive iterations which provide, at the end, the final value of $\phi$, in the PJM we must wait for some conditions to be reached and immediately, by jumping over the possible iterations, the solution is found almost directly.

These conditions are given when the parameter called $EQ^{(k}$ (error quotient between the iterations $k$ and $k-1$) reaches a practically constant value. This parameter is defined by the following expression:

$$EQ^{(k} = E^{(k} / E^{(k-1} \tag{9}$$

where

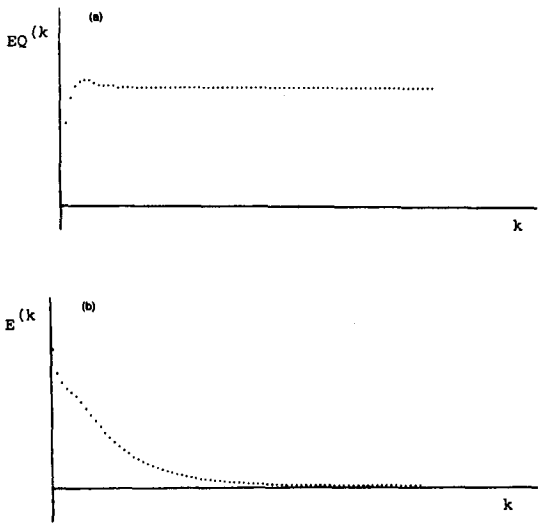



FIG. 2. (a) $EQ_i$ tends to a constant value. (b) The decrease of the error is exponential.

$$E^{(k} = \sum_{i=1}^{N} |\phi_i^{(k} - \phi_i^{(k-1}|$$

and

$$E^{(k-1} = \sum_{i=1}^{N} |\phi_i^{(k-1} - \phi_i^{(k-2}|.$$

Clearly in order to get the convergence of the method, the value of $EQ^{(k}$ must be smaller than unity. Obviously, the closer to zero its value is, the faster the system of equations converges to the solution.

Generally, and depending on the value of $\omega$, $EQ_i$ tends to a constant value as it is shown in Fig. 2(a). This figure has been obtained applying the SOR method for studying some temperature distribution problems. So that, from a certain iteration called $m$, it is verified that

$$|EQ^{(k} - EQ^{(k-1}| < \epsilon', \tag{10}$$

that is to say, by choosing for $\epsilon'$ a sufficiently small value, it can be obtained with good approximation that

$$EQ^{(k-1} \simeq EQ^{(k} \simeq EQ^{(\infty}. \tag{11}$$

This constancy of $EQ^{(k}$ indicates that from the iteration $m$ the decrease of the error is exponential [Fig. 2(b)]. In fact,

$$EQ^{(m} = EQ^{(m+1} = \cdots = \alpha = \text{const.} \tag{12}$$

From (9) and (12), it is proved (when $k \geqslant m+1$) that

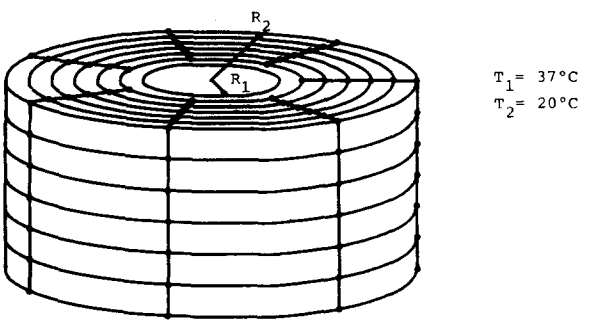



FIG. 3. A cylinder made up of a grid of 294 points inside which is another cylinder at a temperature $T_1 = \text{const.}$

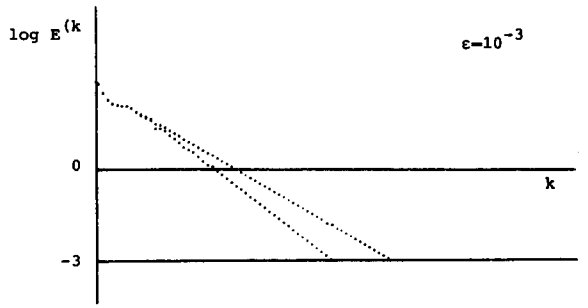



FIG. 4. Log $E$ against the number of iterations for two values of $\omega$ that are different by 0.025.

$$E^{(m+1} = E^{(m}\alpha,$$
$$E^{(m+2} = E^{(m+1}\alpha = E^{(m}\alpha^2,$$
$$\vdots \tag{13}$$
$$E^{(n} = E^{(n-1}\alpha = E^{(m}\alpha^{n-m}.$$

This law of decreasing for $E^{(k}$ is of the same type as for $E_i^{(k}$, which is the error associated to a point $i$ of the system.

By calling

$$E_i^{(k} = |\phi_i^{(k} - \phi_i^{(k-1}| \tag{14}$$

and

$$E_i^{(k-1} = |\phi_i^{(k-1} - \phi_i^{(k-2}|, \tag{15}$$

it is verified that

$$E_i^{(n} = E_i^{(m}\alpha_i^{n-m}, \tag{16}$$

where $\alpha_i = \mathrm{EQ}_i^{(m}$.

The particular values of $\alpha_i$ are very close to the general value of $\alpha$, so that we shall choose $\alpha_i = \alpha$.

If $m$ is the iteration from which $\mathrm{EQ}^{(k} = \alpha = \text{const}$, at a generic point $i$, the values of the function $\phi$ from that iteration can be expressed by the formulas

$$\phi_i^{(m+2} = \phi_i^{(m+1} + \{\phi_i^{(m+1} - \phi_i^{(m}\}\alpha,$$
$$\phi_i^{(m+3} = \phi_i^{(m+1} + \{\phi_i^{(m+1} - \phi_i^{(m}\}\alpha(1+\alpha), \tag{17}$$

that is to say,

$$\phi_i^{(n} = \phi_i^{(m+1} + \{\phi_i^{(m+1} - \phi_i^{(m}\}\alpha[(\alpha^{n-2} - 1)/(\alpha - 1)]. \tag{18}$$

So that, when $n \to \infty$, and because $\alpha < 1$, it is verified that

$$\phi_i^{(\infty} = \phi_i^{(m+1} + \{\phi_i^{(m+1} - \phi_i^{(m}\}[\alpha/(1-\alpha)]. \tag{19}$$

If the value of $\mathrm{EQ}^{(k}$ is constant, this last equation relates the solution of the function $\phi$ in a determined point of the system with the values that this function gets in the iterations $m$ and $m+1$.

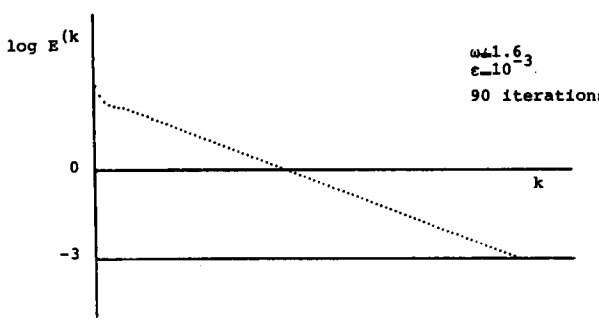



FIG. 5. Log $E^{(k}$ versus the number of iterations applying the SOR method and using $\omega = 1.6$.

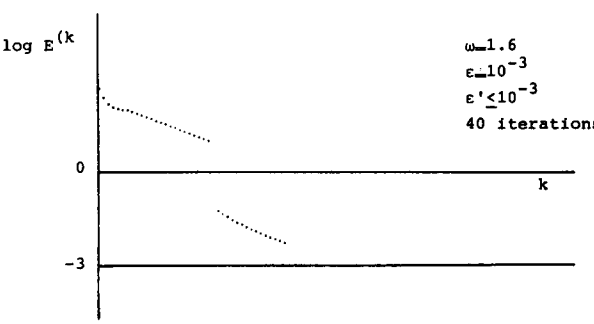



FIG. 6. Same as Filg. 5, but using the PJM when $\epsilon' < 10^{-3}$.

In practice, when the PJM is applied, the value of $\mathrm{EQ}^{(k}$ is not exactly constant, and the solution reached with the jump may not agree accurately with the exact solution, but if it is so close to it that Eq. (4) is verified, we shall consider the calculus process to be finished.

On the contrary, if the solution reached by applying the PJM does not verify Eq. (4), the iterative process continues as usual until it is verified or until a new constant value of $\mathrm{EQ}^{(k}$ is reached and, in this case, the PJM is applied again.

In any way and in all the cases, the number of necessary iterations is substantially lower.

One could think that if instead of applying $\alpha$ to the expression (19) we choose $\alpha_i$, the value for $\phi_i^{(\infty}$ obtained by the PJM would be closer to the exact solution. Nevertheless, we have observed that it does not happen in this manner; it seems as if the constancy of the general value of $\mathrm{EQ}^{(k}$ could be reached faster than the values of $EQ_i^{(k}$.

## IV. RESULTS

We have applied the PJM to the solution of the equations that describes the temperature distribution in solids of different geometries. As an example, the following results correspond to the solid in Fig. 3. It is a cylinder made up of a grid of 294 points inside which is another cylinder at a temperature $T_1 = \text{const}$.

The equations that describe the temperature distribution are

$$\nabla^2 T = 0, \quad T\Big|_{R_1} = T_1, \quad -k\frac{\partial T}{\partial n}\Big|_S = h(T|_S - T_2). \tag{20}$$

This last equation is verified over the external surface.

All the results shown in this paper have been obtained by starting from the initial condition $T_{i,j,k} = T_2 = 20\,^\circ\mathrm{C}$.

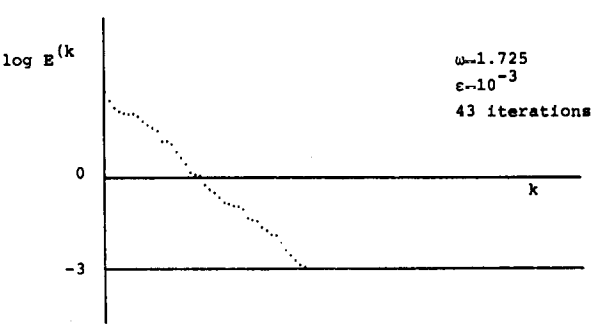



FIG. 7. Log $E^{(k}$ against the number of iterations for $\omega_{\text{opt}} = 1.725$.

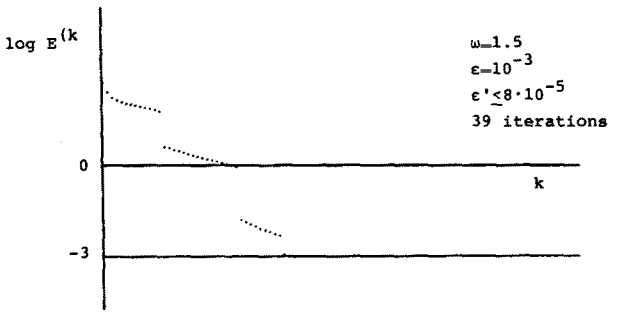



FIG. 8. Applying the PJM to the problem.

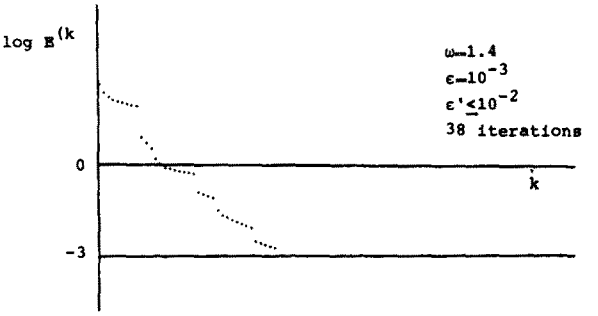



FIG. 9. Applying the PJM to the problem.

Figure 4 shows the logarithm of the error $E^{(k}$ against the number of iterations for two values of $\omega$ that are different by 0.025. The great sensitivity of the convergence for the small variations of $\omega$ can be observed. This example gives an idea of the great advantage of knowing the $\omega_{opt}$ factor and of the high precision necessary to calculate it.

Figure 5 displays the log $E^{(k}$ versus the number of iterations applying the SOR method and using $\omega = 1.6$.

Figure 6 exactly corresponds to the same case but using the PJM when $\epsilon' < 10^{-3}$.

We have calculated $\omega_{opt}$ for this particular problem and it has been obtained $\omega_{opt} = 1.725$.

Figure 7 shows log $E^{(k}$ against the number of iterations for $\omega_{opt} = 1.725$.

Figures 8 and 9 show how by applying the PJM the solution of the problem can be reached with a lower number of iterations. For demonstrating the great advantage of the PJM, we have chosen two values for $\omega$ ($\omega_1 = 1.5$, $\omega_2 = 1.4$) that are rather distant from $\omega_{opt}$.

In following articles some practical applications of this method will be discussed in detail.

[1]E. H. Twizell and P. Smith, Comput. Biol. Med. **12**, 119 (1982).
[2]K. D. Heidtmann, Int. J. Computer **9**, 343 (1981).
[3]J. D. Doss, Med. Phys. **9**, 566 (1982).
[4]C. H. Durney, Proc. IEEE **68**, 33 (1980).
[5]S. J. Bijlsma and R. J. Hoogendorn, Royal Netherlands Meteorological Institute Report W.R. 79-1, 1979.
[6]M. Necati Ozisik, *Heat Conduction* (Wiley, New York, 1980), Chap. 12.
[7]Curtis F. Gerald, *Applied Numerical Analysis* (Addison–Wesley, Reading, MA, 1978), 2nd ed., Chap. 2.
[8]W. F. Ames, *Nonlinear Partial Differential Equations in Engineering* (Academic, New York, 1965), Chap. 7.
[9]D. M. Young, Trans. Am. Math. Soc. **76**, 92 (1954).
[10]M. Distasio and W. C. McHarris, Am. J. Phys. **47**, 440 (1979).
[11]Curtis F. Gerald, *Applied Numerical Analysis* (Wiley, New York, 1980), 2nd ed., Chap. 7.
[12]S. C. Nath, IE(I) J. **CI57**, 153 (1977).
[13]G. F. Vermey, Phys. Med. Biol. **20**, 384 (1975).

# Construction of the second constant of motion for two-dimensional classical systems

R. S. Kaushal
*Department of Physics and Astrophysics, University of Delhi, Delhi-110007, India and* [a] *Department of Physics, Ramjas College, University of Delhi, Delhi-110007, India*

S. C. Mishra
*Department of Physics and Astrophysics, University of Delhi, Delhi-110007, India*

K. C. Tripathy
*Institute of Theoretical Physics, Göteborg, Sweden and* [a] *Department of Physics and Astrophysics, University of Delhi, Delhi-110007, India*

(Received 5 March 1984; accepted for publication 27 July 1984)

A general method for the construction of the second constant of motion of third and fourth orders is given for two-dimensional systems in terms of $z = q_1 + iq_2$, and $\bar{z} = q_1 - iq_2$. Correspondingly, the third- and fourth-order potential equations are obtained whose solutions directly provide the integrable systems. Using the Holt ansatz, the potential equation corresponding to the third-order invariants has been reduced to a pair of potential equations whose solutions yield a large class of integrable systems.

## I. INTRODUCTION

Whittaker[1] first investigated the problem of the construction of an invariant (other than the total energy and which will be called the second constant of motion) for a system described by

$$\ddot{q}_1 = -\frac{\partial v}{\partial q_1}, \quad \ddot{q}_2 = -\frac{\partial v}{\partial q_2}, \quad v = v(q_1,q_2).$$

His studies were, however, restricted to the invariant of first or second order in momenta. Although there have been several attempts[2] in recent years to construct the second-order invariants, not much effort has been made to obtaining the third- or higher-order invariants for such systems. In some cases, no doubt, the system is found to be integrable just by accident[3] (e.g., the Toda lattice case). For an interesting review on the subject, we refer to the work of Hall.[2]

Recently, Holt[4] has introduced a procedure which essentially has bearing on the perturbation theory of McNamara and Whiteman[4] and has obtained the third-order invariants for a number of integrable systems. As an outcome of this method Inozemtsev[5] has shown that the system $v = \lambda (q_1,q_2)^{-2/3}$ is also an integrable one at least classically. In fact, there already exists scarcity of classically integrable systems in two or more degrees of freedom and a test for their corresponding quantum integrability in each case should be carried out[6] separately. The utility of the second or other constants of motion, if they can be constructed for a system, has been noticed[7] in recent years from several points of view particularly, in reducing some nonlinear dynamical problems to a quadrature, in solving several problems of plasma physics and hydrodynamics, and in the study of a classical analog of Yang–Mills field equations with reference to the generation of potentials (both time-dependent and time-independent by choosing suitable gauges).

Earlier, in the light of the work of Katzin and Levine,[8] we have suggested[9] a method for the construction of second-order invariants for time-dependent classical systems in two dimensions. In fact, some new time-dependent integrable systems were found by introducing the complex variable $z = q_1 + iq_2$ and $\bar{z} = q_1 - iq_2$. A lot of simplifications were achieved in the derivation and the analysis turned out to be more transparent. With the same spirit, in the present work, we reexamine the time-independent systems in two dimensions and construct the second constant of motion of third and fourth orders in momenta. We obtain in their most general form, the potential equations of third and fourth orders (corresponding to the third- and fourth-order invariants) whose solutions may directly provide the integrable systems. However, for the third-order case this potential equation reduces to a pair of potential equations each of second order only after making use of the Holt ansatz [cf. Eq. (4.1)]. All the cases discussed by Holt are recovered and a new integrable system $v(q_1,q_2) = \alpha(q_1^2 + q_2^2) + \beta/(q_1^2 + q_2^2)$, is also found. Analytical general results are given for the fourth-order invariants. The plan of our paper is as follows.

In Sec. II, we consider the Lagrangian

$$\mathcal{L} = \tfrac{1}{2}|\dot{z}|^2 - v(z,\bar{z})$$

and requiring that $dI/dt = 0$, we obtain an overdetermined set of partial differential equations involving the coefficients in which the invariant $I$ is already expanded. In Sec. III, we continue with a general solution of these equations in the form of "potential" equations. In Sec. IV, we establish the correspondence between our method and the method of Holt[4] for third-order invariants. Section V deals with the applications of the potential equations obtained in Sec. IV. Various integrable systems are derived and analyzed in Sec. VI.

## II. CONSTRUCTION OF THE POTENTIALS AND CORRESPONDING SECOND CONSTANT OF MOTION

We first give here a general treatment of the construction of invariants up to fourth order in momenta and then[10] discuss separately the third- and fourth-order invariants in the subsequent subsections.

[a] Permanent address.

 0022-2488/85/030420-08$02.50 © 1985 American Institute of Physics 

## A. The method

We consider a dynamical system described by the Lagrangian

$$\mathcal{L} = \tfrac{1}{2}|\dot{z}|^2 - v(z,\bar{z}), \quad z = q_1 + iq_2, \quad \dot{z} = p_1 + ip_2 \tag{2.1}$$

with the concomitant equations of motion

$$\ddot{z} = -2\frac{\partial v}{\partial \bar{z}}, \quad \ddot{\bar{z}} = -2\frac{\partial v}{\partial z}. \tag{2.2}$$

Let us consider the second constant of motion up to fourth order in a general form as

$$I = a_0 + a_i\xi_i + (1/2!)a_{ij}\xi_i\xi_j + (1/3!)a_{ijk}\xi_i\xi_j\xi_k + (1/4!)a_{ijkl}\xi_i\xi_j\xi_k\xi_l, \tag{2.3}$$

where $i,j,k,l = 1,2$; $\xi_1 = \dot{z}$, $\xi_2 = \dot{\bar{z}}$, and the coefficients $a_0$, $a_i$, $a_{ij}$, $a_{ijk}$, and $a_{ijkl}$ are functions of $z$ and $\bar{z}$ only. These coefficients are symmetric with respect to any interchange of their indices. Using $dI/dt = 0$, we find from (2.3),

$$\begin{aligned} &a_{0,i}\dot{\xi}_i + a_{i,j}\xi_i\dot{\xi}_j + a_i\dot{\xi}_i \\ &\quad + \tfrac{1}{2}a_{ij,k}\xi_i\xi_j\dot{\xi}_k + \tfrac{1}{2}a_{ij}(\dot{\xi}_i\xi_j + \dot{\xi}_j\xi_i) + \tfrac{1}{6}a_{ijk,l}\xi_i\xi_j\xi_k\xi_l \\ &\quad + \tfrac{1}{6}a_{ijk}(\dot{\xi}_i\xi_j\xi_k + \xi_i\dot{\xi}_j\xi_k + \xi_i\xi_j\dot{\xi}_k) \\ &\quad + \tfrac{1}{24}a_{ijkl,m}\xi_i\xi_j\xi_k\xi_l\xi_m + \tfrac{1}{24}(\dot{\xi}_i\xi_j\xi_k\xi_l + \xi_i\dot{\xi}_j\xi_k\xi_l \\ &\quad + \xi_i\xi_j\dot{\xi}_k\xi_l + \xi_i\xi_j\xi_k\dot{\xi}_l)a_{ijkl} = 0. \end{aligned} \tag{2.4}$$

After accounting for the proper symmetrization of the coefficients and since (2.4) must hold identically in $\xi$'s, we obtain the following conditions on $a_{ijkl}$, $a_{ijk}$, $a_i$, and $a_0$:

$$a_{ijkl,m} + a_{jklm,i} + a_{klmi,j} + a_{lmij,k} + a_{mijk,l} = 0, \tag{2.5}$$

$$a_{ijk,l} + a_{jkl,i} + a_{kli,j} + a_{lij,k} = 0, \tag{2.6}$$

$$a_{ij,k} + a_{jk,i} + a_{ki,j} + a_{ijkl}\dot{\xi}_l = 0, \tag{2.7}$$

$$a_{i,j} + a_{j,i} + a_{ijk}\dot{\xi}_k = 0, \tag{2.8}$$

$$a_{0,i} + a_{ij}\dot{\xi}_j = 0, \tag{2.9}$$

$$a_i\dot{\xi}_i = 0. \tag{2.10}$$

Equations (2.6), (2.8), and (2.10) yield the following set of partial differential equations:

$$\frac{\partial a_{111}}{\partial z} = 0, \tag{2.11}$$

$$\frac{\partial a_{222}}{\partial \bar{z}} = 0, \tag{2.12}$$

$$\frac{\partial a_{122}}{\partial z} + \frac{\partial a_{112}}{\partial \bar{z}} = 0, \tag{2.13}$$

$$\frac{\partial a_{111}}{\partial \bar{z}} + 3\frac{\partial a_{112}}{\partial z} = 0, \tag{2.14}$$

$$\frac{\partial a_{222}}{\partial z} + 3\frac{\partial a_{122}}{\partial \bar{z}} = 0, \tag{2.15}$$

$$\frac{\partial a_1}{\partial z} = a_{111}\frac{\partial v}{\partial \bar{z}} + a_{112}\frac{\partial v}{\partial z}, \tag{2.16}$$

$$\frac{\partial a_2}{\partial \bar{z}} = a_{122}\frac{\partial v}{\partial \bar{z}} + a_{222}\frac{\partial v}{\partial z}, \tag{2.17}$$

$$\frac{\partial a_1}{\partial \bar{z}} + \frac{\partial a_2}{\partial z} = 2a_{112}\frac{\partial v}{\partial \bar{z}} + 2a_{122}\frac{\partial v}{\partial z}, \tag{2.18}$$

$$a_1\frac{\partial v}{\partial \bar{z}} + a_2\frac{\partial v}{\partial z} = 0, \tag{2.19}$$

whereas Eqs. (2.5), (2.7), and (2.9) yield

$$\frac{\partial a_{1111}}{\partial z} = 0, \tag{2.20}$$

$$\frac{\partial a_{1111}}{\partial \bar{z}} + 4\frac{\partial a_{1112}}{\partial z} = 0, \tag{2.21}$$

$$2\frac{\partial a_{1112}}{\partial \bar{z}} + 3\frac{\partial a_{1122}}{\partial z} = 0, \tag{2.22}$$

$$3\frac{\partial a_{1122}}{\partial \bar{z}} + 2\frac{\partial a_{1222}}{\partial z} = 0, \tag{2.23}$$

$$\frac{\partial a_{2222}}{\partial z} + 4\frac{\partial a_{1222}}{\partial \bar{z}} = 0, \tag{2.24}$$

$$\frac{\partial a_{2222}}{\partial \bar{z}} = 0, \tag{2.25}$$

$$3\frac{\partial a_{11}}{\partial z} = 2a_{1111}\frac{\partial v}{\partial \bar{z}} + 2a_{1112}\frac{\partial v}{\partial z}, \tag{2.26}$$

$$3\frac{\partial a_{22}}{\partial \bar{z}} = 2a_{2222}\frac{\partial v}{\partial z} + 2a_{1222}\frac{\partial v}{\partial \bar{z}}, \tag{2.27}$$

$$\frac{\partial a_{11}}{\partial \bar{z}} + 2\frac{\partial a_{12}}{\partial z} = 2a_{1112}\frac{\partial v}{\partial \bar{z}} + 2a_{1122}\frac{\partial v}{\partial z}, \tag{2.28}$$

$$\frac{\partial a_{22}}{\partial z} + 2\frac{\partial a_{12}}{\partial \bar{z}} = 2a_{1222}\frac{\partial v}{\partial z} + 2a_{1122}\frac{\partial v}{\partial \bar{z}}, \tag{2.29}$$

$$\frac{\partial a_0}{\partial z} = 2a_{11}\frac{\partial v}{\partial \bar{z}} + 2a_{12}\frac{\partial v}{\partial z}, \tag{2.30}$$

$$\frac{\partial a_0}{\partial \bar{z}} = 2a_{12}\frac{\partial v}{\partial \bar{z}} + 2a_{22}\frac{\partial v}{\partial z}. \tag{2.31}$$

Now, we present solutions of these equations for determining various coefficients.

## B. Determination of $a_{ijk}$

From Eqs. (2.11) and (2.12)

$$a_{111} = a_{111}(\bar{z}) = \psi_1(\bar{z}) \quad \text{and} \quad a_{222} = a_{222}(z) = \phi_1(z).$$

Equation (2.14), after using (2.11), yields

$$\frac{\partial^2 a_{112}}{\partial z^2} = 0,$$

whose solution is

$$a_{112} = \psi_2(\bar{z})z + \psi_3(\bar{z}). \tag{2.32}$$

Similarly, Eqs. (2.15) and (2.12) will lead to the solution

$$a_{122} = \phi_2(z)\bar{z} + \phi_3(z). \tag{2.33}$$

With these solutions, Eq. (2.13) implies

$$z\frac{d\psi_2}{d\bar{z}} + \frac{d\psi_3}{d\bar{z}} + \bar{z}\frac{d\phi_2}{dz} + \frac{d\phi_3}{dz} = 0.$$

Let us consider, $\phi_3 = C_1$ and $\psi_3 = C_2$. (Note that here $\phi_i$'s and $\psi_i$'s are the functions of only $z$ and $\bar{z}$, respectively, and $C_i$'s are some arbitrary complex constants.) This reduces the above equation to the form

$$\frac{1}{\bar{z}}\frac{d\psi_2}{d\bar{z}} + \frac{1}{z}\frac{d\phi_2}{dz} = 0, \tag{2.34}$$

which after making use of (2.32) and (2.33) in (2.14) and (2.15), respectively, yields

$$\frac{1}{\bar{z}}\frac{d^2\psi_1}{d\bar{z}^2} = -\frac{1}{z}\frac{d^2\phi_1}{dz^2} = \text{const} \quad (\text{say}, C_3),$$

or

$$\psi_1 = \tfrac{1}{6}C_3\bar{z}^3 + C_4\bar{z} + C_5, \quad \psi_2 = -\tfrac{1}{6}C_3\bar{z}^2 - \tfrac{1}{3}C_4.$$

Similarly, we find $\phi_1$ and $\phi_2$ as

$$\phi_1 = -\tfrac{1}{6}C_3 z^3 + C_6 z + C_7, \quad \phi_2 = \tfrac{1}{6}C_3 z^2 - \tfrac{1}{3}C_6.$$

Finally, the solutions of Eqs. (2.11) to (2.15) become

$$a_{111} = \tfrac{1}{6}C_3\bar{z}^3 + C_4\bar{z} + C_5, \tag{2.35}$$

$$a_{222} = -\tfrac{1}{6}C_3 z^3 + C_6 z + C_7, \tag{2.36}$$

$$a_{112} = -\tfrac{1}{6}C_3 z\bar{z}^2 - \tfrac{1}{3}C_4 z + C_2, \tag{2.37}$$

$$a_{122} = \tfrac{1}{6}C_3\bar{z}z^2 - \tfrac{1}{3}C_6\bar{z} + C_1. \tag{2.38}$$

### C. Determination of $a_{ijkl}$

From Eqs. (2.20) and (2.25) we have $a_{1111} = a_{1111}(\bar{z}) = \sigma_1(\bar{z})$ and $a_{2222} = a_{2222}(z) = \chi_1(z)$. Using (2.20) in Eq. (2.21), we have

$$\frac{\partial^2 a_{1112}}{\partial z^2} = 0,$$

which admits the solution as

$$a_{1112} = \sigma_2(\bar{z})z + \sigma_3(\bar{z}). \tag{2.39}$$

Similarly, Eqs. (2.24) and (2.25) will lead to the solution

$$a_{1222} = \chi_2(z)\bar{z} + \chi_3(z). \tag{2.40}$$

On differentiating (2.22) with respect to $\bar{z}$ and (2.23) with respect to $z$, and correspondingly subtracting the results and making use of (2.39) and (2.40) we obtain

$$z\frac{d^2\sigma_2}{d\bar{z}^2} + \frac{d^2\sigma_3}{d\bar{z}^2} = \bar{z}\frac{d^2\chi_2}{dz^2} + \frac{d^2\chi_3}{dz^2}.$$

Now, we fix $\sigma_3 = D_1$ and $\chi_3 = D_2$. (Note that here $\sigma_i$'s and $\chi_i$'s are the functions of only $\bar{z}$ and $z$, respectively, and $D_i$'s are the arbitrary complex constants.) This reduces the above equation to the form

$$\frac{1}{\bar{z}}\frac{d^2\sigma_2}{d\bar{z}^2} = \frac{1}{z}\frac{d^2\chi_2}{dz^2} = \text{const} \quad (\text{say}, D_3),$$

or

$$\sigma_2 = \tfrac{1}{6}D_3\bar{z}^3 + D_4\bar{z} + D_5,$$

$$\chi_2 = \tfrac{1}{6}D_3 z^3 + D_6 z + D_7.$$

Similarly, we find $\sigma_1$ and $\chi_1$ starting from (2.21) and (2.24) as

$$\sigma_1 = -\tfrac{1}{6}D_3\bar{z}^4 - 2D_4\bar{z}^2 - 4D_5\bar{z} + D_8,$$

$$\chi_1 = -\tfrac{1}{6}D_3 z^4 - 2D_6 z^2 - 4D_7 z + D_9.$$

Equation (2.22) yields

$$\frac{\partial a_{1122}}{\partial z} = -\frac{2}{3}z\frac{d\sigma_2}{d\bar{z}} = -\frac{1}{3}D_3 z\bar{z}^2 - \frac{2}{3}D_4 z,$$

or

$$a_{1122} = -\tfrac{1}{6}D_3\bar{z}^2 z^2 - \tfrac{1}{3}D_4 z^2 + \sigma_4(\bar{z}).$$

We note that $\sigma_4(\bar{z})$ can be determined from Eq. (2.23). Finally, the solutions of Eqs. (2.20)–(2.25) yield

$$a_{1111} = -\tfrac{1}{6}D_3\bar{z}^4 - 2D_4\bar{z}^2 - 4D_5\bar{z} + D_8, \tag{2.41}$$

$$a_{2222} = -\tfrac{1}{6}D_3 z^4 - 2D_6 z^2 - 4D_7 z + D_9, \tag{2.42}$$

$$a_{1112} = \frac{D_3}{6}z\bar{z}^3 + D_4 z\bar{z} + D_5 z + D_1, \tag{2.43}$$

$$a_{1222} = \tfrac{1}{6}D_3\bar{z}z^3 + D_6 z\bar{z} + D_7\bar{z} + D_2, \tag{2.44}$$

$$a_{1122} = -\tfrac{1}{6}D_3 z^2\bar{z}^2 - \tfrac{1}{3}D_4 z^2 - \tfrac{1}{3}D_6\bar{z}^2 + D_{10}. \tag{2.45}$$

Having solved the potential-independent equations (2.11)–(2.15) and (2.20)–(2.25) in terms of arbitrary constants $C_i$'s and $D_i$'s, we now proceed for the solution of potential-dependent equations (2.16)–(2.19) and (2.26)–(2.31).

## III. FORMULATION OF THE GENERAL POTENTIAL EQUATIONS

### A. Derivation

In order to eliminate $a_1$ and $a_2$ from Eqs. (2.16)–(2.18) we differentiate (2.18) with respect to $z$ and using (2.16) obtain

$$\begin{aligned}\frac{\partial^2 a_2}{\partial z^2} = {} & \left(2\frac{\partial a_{112}}{\partial z} - \frac{\partial a_{111}}{\partial \bar{z}}\right)\frac{\partial v}{\partial \bar{z}} \\ & + a_{112}\frac{\partial^2 v}{\partial z\cdot\partial\bar{z}} - a_{111}\frac{\partial^2 v}{\partial \bar{z}^2} \\ & + 2a_{122}\frac{\partial^2 v}{\partial z^2} + \left(2\frac{\partial a_{122}}{\partial z} - \frac{\partial a_{112}}{\partial \bar{z}}\right)\frac{\partial v}{\partial z}.\end{aligned}$$

Now differentiating this equation with respect to $\bar{z}$ and using Eq. (2.17) for $\partial a_2/\partial\bar{z}$ the rearrangement of terms leads to the equation

$$\begin{aligned}& \left(\frac{\partial^2 a_{111}}{\partial\bar{z}^2} + \frac{\partial^2 a_{122}}{\partial z^2} - 2\frac{\partial^2 a_{112}}{\partial z\cdot\partial\bar{z}}\right)\frac{\partial v}{\partial\bar{z}} \\ & + 2\left(\frac{\partial a_{111}}{\partial\bar{z}} - \frac{\partial a_{112}}{\partial z}\right)\frac{\partial^2 v}{\partial\bar{z}^2} \\ & + a_{111}\frac{\partial^3 v}{\partial\bar{z}^3} + \left(\frac{\partial^2 a_{112}}{\partial\bar{z}^2} + \frac{\partial^2 a_{222}}{\partial z^2} - 2\frac{\partial^2 a_{122}}{\partial z\cdot\partial\bar{z}}\right)\frac{\partial v}{\partial z} \\ & + 2\left(\frac{\partial a_{222}}{\partial z} - \frac{\partial a_{122}}{\partial\bar{z}}\right)\frac{\partial^2 v}{\partial z^2} \\ & + a_{222}\frac{\partial^3 v}{\partial z^3} - a_{112}\frac{\partial^3 v}{\partial z\cdot\partial\bar{z}^2} - a_{122}\frac{\partial^3 v}{\partial\bar{z}\cdot\partial z^2} = 0.\end{aligned} \tag{3.1}$$

In the same way we proceed to eliminate $a_{11}$, $a_{12}$, and $a_{22}$ from Eqs. (2.26)–(2.29). On differentiating (2.28) with respect to $z$ and making use of (2.26) for $\partial a_{11}/\partial z$, we find

$$\begin{aligned}\frac{\partial^2 a_{12}}{\partial z^2} = {} & \left(\frac{\partial a_{1122}}{\partial z} - \frac{1}{3}\frac{\partial a_{1112}}{\partial\bar{z}}\right)\frac{\partial v}{\partial z} \\ & + \frac{2}{3}a_{1112}\frac{\partial^2 v}{\partial z\cdot\partial\bar{z}} \\ & + \left(\frac{\partial a_{1112}}{\partial z} - \frac{1}{3}\frac{\partial a_{1111}}{\partial\bar{z}}\right)\frac{\partial v}{\partial\bar{z}} \\ & + a_{1122}\frac{\partial^2 v}{\partial z^2} - \frac{1}{3}a_{1111}\frac{\partial^2 v}{\partial\bar{z}^2}.\end{aligned} \tag{3.2}$$

Similarly, on differentiating (2.29) with respect to $\bar{z}$ and using (2.27) for $\partial a_{22}/\partial\bar{z}$, we find

$$\frac{\partial^2 a_{12}}{\partial \bar{z}^2} = \left(\frac{\partial a_{1222}}{\partial \bar{z}} - \frac{1}{3}\frac{\partial a_{2222}}{\partial z}\right) \times \frac{\partial v}{\partial z} + \frac{2}{3} a_{1222} \frac{\partial^2 v}{\partial z \cdot \partial \bar{z}} + \left(\frac{\partial a_{1122}}{\partial \bar{z}} - \frac{1}{3}\frac{\partial a_{1222}}{\partial z}\right)\frac{\partial v}{\partial \bar{z}} + a_{1122}\frac{\partial^2 v}{\partial \bar{z}^2} - \frac{1}{3} a_{2222}\frac{\partial^2 v}{\partial z^2}. \quad (3.3)$$

Now to eliminate $a_{12}$, we differentiate twice (3.2) with respect to $\bar{z}$ and (3.3) with respect to $z$, and using $\partial^4 a_{12}/(\partial \bar{z}^2 \cdot \partial z^2) = \partial^4 a_{12}/(\partial z^2 \cdot \partial \bar{z}^2)$, we finally obtain

$$\left(\frac{\partial^3 a_{1122}}{\partial z \cdot \partial \bar{z}^2} - \frac{1}{3}\frac{\partial^3 a_{1112}}{\partial \bar{z}^3} - \frac{\partial^3 a_{1222}}{\partial \bar{z} \cdot \partial z^2} + \frac{1}{3}\frac{\partial^3 a_{2222}}{\partial z^3}\right)\frac{\partial v}{\partial z} + \left(\frac{\partial^2 a_{1122}}{\partial \bar{z}^2} - \frac{\partial^2 a_{1222}}{\partial z \cdot \partial \bar{z}} + \frac{\partial^2 a_{2222}}{\partial z^2}\right)\frac{\partial^2 v}{\partial z^2} + \left(\frac{\partial a_{2222}}{\partial z} - \frac{\partial a_{1222}}{\partial \bar{z}}\right)\frac{\partial^3 v}{\partial z^3} + \frac{1}{3} a_{2222} \times \frac{\partial^4 v}{\partial z^4} + \left(\frac{\partial a_{1122}}{\partial \bar{z}} - \frac{\partial a_{1222}}{\partial z}\right)\frac{\partial^3 v}{\partial \bar{z} \cdot \partial z^2} - \frac{2}{3} a_{1222}\frac{\partial^4 v}{\partial \bar{z} \cdot \partial z^3} + \frac{2}{3} a_{1112}\frac{\partial^4 v}{\partial z \cdot \partial \bar{z}^3} + \left(\frac{\partial a_{1112}}{\partial \bar{z}} - \frac{\partial a_{1122}}{\partial z}\right)\frac{\partial^3 v}{\partial z \cdot \partial \bar{z}^2} + \left(\frac{\partial^3 a_{1112}}{\partial z \cdot \partial \bar{z}^2} + \frac{1}{3}\frac{\partial^3 a_{1222}}{\partial z^3} - \frac{\partial^3 a_{1122}}{\partial \bar{z} \cdot \partial z^2} - \frac{1}{3}\frac{\partial^3 a_{1111}}{\partial \bar{z}^3}\right)\frac{\partial v}{\partial \bar{z}} + \left(\frac{\partial^2 a_{1112}}{\partial z \cdot \partial \bar{z}} - \frac{\partial^2 a_{1122}}{\partial z^2} - \frac{\partial^2 a_{1111}}{\partial \bar{z}^2}\right)\frac{\partial^2 v}{\partial \bar{z}^2} + \left(\frac{\partial a_{1112}}{\partial z} - \frac{\partial a_{1111}}{\partial \bar{z}}\right)\frac{\partial^3 v}{\partial \bar{z}^3} - \frac{1}{3} a_{1111}\frac{\partial^4 v}{\partial \bar{z}^4} = 0. \quad (3.4)$$

Equations (3.1) and (3.4) are general "potential equations" corresponding to the third- and fourth-order invariants. These equations involve the potential derivatives and known coefficients $a_{ijk}$ and $a_{ijkl}$ through unknown constants $C_i$'s and $D_i$'s. On substituting these coefficients from (2.35)–(2.38) and (2.41)–(2.45), the potential equations reduce to the forms

$$2C_3\bar{z}\frac{\partial v}{\partial \bar{z}} + \left(\frac{4}{3}C_3\bar{z}^2 + \frac{8}{3}C_4\right)\frac{\partial^2 v}{\partial \bar{z}^2} + \left(\frac{1}{6}C_3\bar{z}^3 + C_4\bar{z} + C_5\right)\frac{\partial^3 v}{\partial \bar{z}^3} - 2C_3 z\frac{\partial v}{\partial z} - \left(\frac{4}{3}C_3 z^2 - \frac{8}{3}C_6\right)\frac{\partial^2 v}{\partial z^2} - \left(\frac{1}{6}C_3 z^3 - C_6 z - C_7\right)\frac{\partial^3 v}{\partial z^3} + \left(\frac{1}{6}C_3 z\bar{z}^2 + \frac{1}{3}C_4 z - C_2\right)\frac{\partial^3 v}{\partial z \cdot \bar{z}^2} - \left(\frac{1}{6}C_3\bar{z}z^2 - \frac{1}{3}C_6\bar{z} + C_1\right)\frac{\partial^3 v}{\partial \bar{z} \cdot \partial z^2} = 0. \quad (3.5)$$

and

$$-\frac{10}{3}D_3 z\frac{\partial v}{\partial z} - \frac{17}{3}\left(\frac{1}{2}D_3 z^2 + D_6\right)\frac{\partial^2 v}{\partial z^2} - 5\left(\frac{1}{6}D_3 z^3 + D_6 z + D_7\right) \times \frac{\partial^3 v}{\partial z^3} + \frac{1}{3}\left(-\frac{1}{6}D_3 z^4 - 2D_6 z^2 - 4D_7 z + D_9\right)\frac{\partial^4 v}{\partial z^4} - \frac{5}{3}\left(\frac{1}{2}D_3\bar{z}z^2 + D_6\bar{z}\right) \times \frac{\partial^3 v}{\partial \bar{z} \cdot \partial z^2} - \frac{2}{3}\left(\frac{1}{6}D_3\bar{z}z^3 + D_6 z\bar{z} + D_7\bar{z} + D_2\right)\frac{\partial^4 v}{\partial \bar{z} \cdot \partial z^3} + \frac{2}{3}\left(\frac{1}{6}D_3 z\bar{z}^3 + D_4\bar{z}\bar{z} + D_5 z + D_1\right)\frac{\partial^4 v}{\partial z \cdot \bar{z}^3} + \frac{5}{3}\left(\frac{1}{2}D_3 z\bar{z}^2 + D_4 z\right)\frac{\partial^3 v}{\partial z \cdot \partial \bar{z}^2} + \frac{10}{3}D_3\bar{z}\frac{\partial v}{\partial \bar{z}} + \frac{17}{3}\left(\frac{1}{2}D_3\bar{z}^2 + D_4\right)\frac{\partial^2 v}{\partial \bar{z}^2} + 5\left(\frac{1}{6}D_3\bar{z}^3 + D_4\bar{z} + D_5\right)\frac{\partial^3 v}{\partial \bar{z}^3} + \frac{1}{3}\left(\frac{1}{6}D_3\bar{z}^4 + 2D_4\bar{z}^2 + 4D_5\bar{z} - D_8\right)\frac{\partial^4 v}{\partial \bar{z}^4} = 0. \quad (3.6)$$

As such the solution of (3.5) and (3.6) for $v(z,\bar{z})$ is a difficult task, but we provide the following recipe for the construction of the invariant. For a given form of $v(z,\bar{z})$, the unknown constants $C_i$'s or $D_i$'s can be determined by rationalizing the potential equation (3.5) or (3.6). Subsequently, the determination of other coefficients $a_i$ for (3.5) and $a_0, a_{ij}$ for (3.6) from Eqs. (2.16)–(2.19) and (2.26)–(2.31), respectively, lead to the final form of the second constant of motion from (2.3). However, for Eq. (3.1) in the next section, we have shown that it reduces into two second-order partial differential equations after making use of the Holt ansatz for $a_1$ and $a_2$.

### B. An example

To demonstrate the method outlined above we consider the example of Inozemtsev,[5] $v(q_1,q_2) = \lambda\,(q_1,q_2)^{-2/3}$ or, equivalently

$$v(z,\bar{z}) = (4i)^{2/3}\lambda\,(z^2 - \bar{z}^2)^{-2/3}. \quad (3.7)$$

On substituting the derivatives of this potential in (3.5) and correspondingly rationalizing the resultant equation, we find that

$$C_1 = C_2 = C_3 = C_5 = C_7 = 0 \text{ and } C_4 = C_6 = C_0 \text{ (say)}.$$

Thus, the coefficients $a_{ijk}$ from (2.35)–(2.38) turn out to be

$$a_{111} = C_0\bar{z}, \quad a_{112} = -\tfrac{1}{3}C_0 z,$$
$$a_{122} = -\tfrac{1}{3}C_0\bar{z}, \quad a_{222} = C_0 z.$$

Now substituting these expressions in (2.16) and (2.17) and integrating the resultant equations we obtain $a_1$ and $a_2$ as

$$a_1 = -C_0 k_0 \lambda z(z^2 - \bar{z}^2)^{-2/3} + k_1,$$
$$a_2 = -C_0 k_0 \lambda\bar{z}(z^2 - \bar{z}^2)^{-2/3} + k_2,$$

where $k_0 = \frac{4}{3}(4i)^{2/3}$, and $k_1, k_2$ are the integration constants. Once the coefficients $a_{ijk}$ and $a_i$ are determined, they can be substituted back in (2.3) to give the invariant $I$. After notic-

ing that $z = q_1 + iq_2, \dot{z} = p_1 + ip_2$, the invariant finally turns out to be

$$I = 2\lambda\,(p_2 q_2 - p_1 q_1)(q_1 q_2)^{-2/3} + p_1 p_2(p_1 q_2 - p_2 q_1), \tag{3.8}$$

with the choice $C_0 = \frac{3}{4}$. In the same way Eqs. (3.5) and (3.6) can be used to find other integrable systems. However, we shall use the Holt's prescription in Sec. V for this purpose.

## IV. CORRESPONDENCE WITH THE HOLT METHOD

In this section we establish the correspondence of our method with that of Holt[4] for the third-order invariant and use his method for reducing the third-order potential equation to a pair of second-order potential equations which are relatively easier to handle. Holt considers the form of the invariant as

$$I = E_1 p_1^{\,3} + E_2 p_1^{\,2} p_2 + E_3 p_1 p_2^{\,2} + E_4 p_2^{\,3} + E_5 p_1 + E_6 p_2,$$

where the coefficients $E_i$'s satisfy a set of partial differential equations. If we identify

$$E_1 = \tfrac{1}{6} a_{111}, \quad E_2 = \tfrac{1}{2} a_{112}, \quad E_3 = \tfrac{1}{2} a_{122},$$
$$E_4 = \tfrac{1}{6} a_{222}, \quad E_5 = a_1, \quad E_6 = a_2,$$

and replace $z$ for $q_1$ and $\bar{z}$ for $q_2$ and $\epsilon = 1$, then a one-to-one correspondence between our equations (2.11)–(2.19) and those of Holt [cf. Eqs. (147)–(155)] of Ref. 4 can be seen except for the fact that the kinetic energy term in the Lagrangian (2.1) is defined differently for notational purposes.

Following Holt,[4] we assume a solution of (2.19) as

$$a_1 = G\frac{\partial v}{\partial z}, \quad a_2 = -G\frac{\partial v}{\partial \bar{z}}, \tag{4.1}$$

where $G$ is some function of $z$ and $\bar{z}$. With the choice of $a_1$ and $a_2$ Eqs. (2.16)–(2.18), respectively, reduce to

$$\frac{\partial G}{\partial z}\cdot\frac{\partial v}{\partial z} + G\frac{\partial^2 v}{\partial z^2} = a_{111}\frac{\partial v}{\partial \bar{z}} + a_{112}\frac{\partial v}{\partial z}, \tag{4.2}$$

$$\left(\frac{\partial G}{\partial \bar{z}} - 2a_{122}\right)\frac{\partial v}{\partial z} - \left(\frac{\partial G}{\partial z} + 2a_{112}\right)\frac{\partial v}{\partial \bar{z}} = 0, \tag{4.3}$$

$$-\frac{\partial G}{\partial \bar{z}}\cdot\frac{\partial v}{\partial \bar{z}} - G\frac{\partial^2 v}{\partial \bar{z}^2} = a_{122}\frac{\partial v}{\partial \bar{z}} + a_{222}\frac{\partial v}{\partial z}. \tag{4.4}$$

Equation (4.3) can be solved for $G$ by introducing a new function $Y$ such that

$$\frac{\partial Y}{\partial \bar{z}} = 2a_{122}, \quad \frac{\partial Y}{\partial z} = -2a_{112}, \tag{4.5}$$

which after using (2.37) and (2.38) leads to the form of $Y$ as

$$Y = Y_0 + 2C_1\bar{z} - 2C_2 z - \tfrac{1}{3}C_6\bar{z}^2 + \tfrac{1}{3}C_4 z^2 + \tfrac{1}{6}C_3 z^2\bar{z}^2, \tag{4.6}$$

where $Y_0$ is some arbitrary constant. The general solution of (4.3) can be written as

$$G = Y + \phi, \tag{4.7}$$

where $\phi$ is an arbitrary function of $v$. The substitution of (4.7) in (4.2) and (4.4) yields the potential equations,

$$Y\frac{\partial^2 v}{\partial z^2} - 3a_{112}\frac{\partial v}{\partial z} - a_{111}\frac{\partial v}{\partial \bar{z}} = -\frac{\partial}{\partial z}\left(\phi\frac{\partial v}{\partial z}\right) \tag{4.8}$$

and

$$Y\frac{\partial^2 v}{\partial \bar{z}^2} + 3a_{122}\frac{\partial v}{\partial \bar{z}} + a_{222}\frac{\partial v}{\partial z} = -\frac{\partial}{\partial \bar{z}}\left(\phi\frac{\partial v}{\partial \bar{z}}\right), \tag{4.9}$$

which are similar to those obtained by Holt [cf. Eqs. (170) and (171) of Ref. 4]. It may be remarked here that Holt has considered only those cases for which the function $\phi = 0$. One case in which $\phi \neq 0$ is considered by Inozemtsev[5] who obtained a new integrable system, $v(q_1,q_2) = \lambda\,(q_1,q_2)^{-2/3}$.

## V. APPLICATIONS OF THE SECOND-ORDER POTENTIAL EQUATIONS (4.8) AND (4.9)

In this section, we look for the integrable systems which admit third-order invariants through the solutions of the potential equations (4.8) and (4.9). We recover here all the cases discussed by Holt for $\phi = 0$ [case (a), (b), and (c)] and also derive new integrable systems for $\phi \neq 0$ [cases (d) and (e)].

*Case (a):* Consider the potential,

$$v(q_1,q_2) = q_1^{\,2} + 4q_2^{\,2} + \delta q_1^{\,-2}$$

or

$$v(z,\bar{z}) = \tfrac{5}{2}z\bar{z} - \tfrac{3}{4}z^2 - \tfrac{3}{4}\bar{z}^2 + 4\delta(z+\bar{z})^{-2}. \tag{5.1}$$

We note here that although $v(q_1,q_2)$ is not symmetric with respect to the interchange of $q_1$ and $q_2$, $v(z,\bar{z})$ is symmetric in $z$ and $\bar{z}$. When this symmetry property of the potential $v(z,\bar{z})$ is used to yield identical solutions of (4.8) and (4.9), we obtain

$$\phi = Y_0 = C_3 = C_4 = C_6 = 0$$

and

$$C_1 = -C_2, \quad C_5 = -C_7 = 3C_2,$$

which lead to

$$G = Y = -2C_2(z+\bar{z}),$$
$$a_{111} = 3C_2, \quad a_{222} = -3C_2,$$
$$a_{112} = C_2, \quad a_{122} = -C_2,$$
$$a_1 = -2C_2(z+\bar{z})\{\tfrac{5}{2}\bar{z} - \tfrac{3}{2}z - 8\delta(z+\bar{z})^{-3}\},$$
$$a_2 = 2C_2(z+\bar{z})\{\tfrac{5}{2}z - \tfrac{3}{2}\bar{z} - 8\delta(z+\bar{z})^{-3}\}.$$

Thus, the invariant $I$ [Eq. (2.3)] reduces to

$$I = 4iC_2[\,p_1^{\,2}p_2 + 2(4q_1q_2p_1 - p_2q_1^{\,2} + \delta p_2 q_1^{\,-2})] \tag{5.2}$$

which coincides with that of Holt on identifying $C_2 = 1/4i$.

*Case (b):* Let us consider the potential[2,4–6]

$$v(q_1,q_2) = (q_1^{\,2} + \tfrac{3}{4}q_2^{\,2} + \delta)q_2^{\,2/3}$$

or

$$v(z,\bar{z}) = (2i)^{2/3}(\tfrac{1}{16}z^2 + \tfrac{1}{16}\bar{z}^2 + \tfrac{7}{8}z\bar{z} + \delta)(z-\bar{z})^{-2/3}. \tag{5.3}$$

On substituting $Y$ from (4.6) and the potential derivatives from (5.3) in (4.8) and (4.9) and rationalizing the latter equations, we find

$$\phi = Y_0 = C_3 = C_4 = C_6 = 0$$

and

$$C_1 = C_2 = -3C_5 = -3C_7,$$

which lead to

$$G = Y = 6C_5(z-\bar{z}), \quad a_{111} = C_5, \quad a_{222} = C_5,$$
$$a_{112} = -3C_5, \quad a_{122} = -3C_5,$$

and correspondingly $a_1$ and $a_2$ can be obtained from (4.1). Finally, the invariant from (2.3) turns out to be

$$I = 4C_5 p_1(3q_2^2 - 2q_1^2 - 2\delta)q_2^{-2/3} - \tfrac{8}{3}C_5 p_1^3 - 4C_5 p_1 p_2^2 - 24C_5 p_2 q_1 q_2^{1/3}, \quad (5.4)$$

which has the same form as that of Holt[4] for $C_5 = -\frac{3}{4}$.

*Case (c):* Consider the Toda lattice potential[3]

$$v(q_1,q_2) = \alpha_+ e^{q_2+\sqrt{3}q_1} + \alpha_- e^{q_2-\sqrt{3}q_1} + \beta e^{-2q_2},$$

or

$$v(z,\bar{z}) = \alpha_+ e^{A_+ z + A_- \bar{z}} + \alpha_- e^{B_+ z + B_- \bar{z}} + \beta e^{C_+ z + C_- \bar{z}}, \quad (5.5)$$

where

$$A_\pm = \frac{\sqrt{3}}{2} \pm \frac{1}{2i}, \quad B_\pm = -\frac{\sqrt{3}}{2} \pm \frac{1}{2i}, \quad C_\pm = \pm i.$$

After substituting the potential derivatives and $Y$ in (4.8) and (4.9) and then rationalizing the resultant equations, we find that

$$Y_0 = 3, \quad C_1 = C_2 = C_3 = C_4 = C_6 = \phi = 0,$$
$$C_5 = C_7 = -3i,$$

which lead to

$$a_{111} = -3i, \quad a_{222} = -3i,$$
$$a_{112} = a_{122} = 0, \quad G = Y = 3,$$

and the corresponding $a_1$ and $a_2$ can be obtained from (4.1). Finally, the invariant from (2.3) can be written as

$$|I| = 3\alpha_+(p_1 - \sqrt{3}p_2)e^{q_2+\sqrt{3}q_1} + 3\alpha_-(p_1 + \sqrt{3}p_2)e^{q_2-\sqrt{3}q_1} - 6\beta p_1 e^{-2q_2} + p_1(p_1^2 - 3p_2^2), \quad (5.6)$$

which has the same structure as given by Hall.[2]

*Case (d):* Now we discuss those cases for which $\phi \neq 0$ in the potential equations (4.8) and (4.9). We assume that the potential $v$ depends on one argument $\eta$ only, where $\eta$ is given by

$$\eta = Y = q_1 q_2 \text{ or } \eta = Y = (4i)^{-1}(z^2 - \bar{z}^2) \quad (5.7)$$

and

$$v = \tau(\eta), \quad \phi = \phi(v) = \phi(\tau). \quad (5.8)$$

A comparison of (5.7) with Eq. (4.6) implies that

$$Y_0 = C_1 = C_2 = C_3 = 0 \text{ and } C_4 = C_6 = 3/4i,$$

and correspondingly the coefficients $a_{ijk}$ can be obtained from (2.35)–(2.38) as

$$a_{111} = C_4\bar{z} + C_5, \quad a_{222} = C_4 z + C_7,$$
$$a_{112} = -\tfrac{1}{3}C_4 z, \quad a_{122} = -\tfrac{1}{3}C_4\bar{z}. \quad (5.9)$$

After making use of Eqs. (5.7)–(5.9) the potential equations (4.8) and (4.9) reduce to the forms

$$(\eta+\phi)[\tau' + (2i)^{-1}z^2\tau''] + [C_4(z^2+\bar{z}^2) + C_5\bar{z}]\tau' + (2i)^{-1}z^2\tau'^2\frac{d\phi}{d\tau} = 0, \quad (5.10)$$

$$-(\eta+\phi)[\tau' - (2i)^{-1}\bar{z}^2\tau''] + [C_4(z^2+\bar{z}^2) + C_7 z]\tau' + (2i)^{-1}\bar{z}^2\tau'^2\frac{d\phi}{d\tau} = 0. \quad (5.11)$$

Now we multiply (5.10) by $\bar{z}^2$ and (5.11) by $z^2$ and subtract the resultant equations to give

$$(\eta+\phi)(z^2+\bar{z}^2) + C_4(\bar{z}^4 - z^4) + C_5\bar{z}^3 - C_7 z^3 = 0. \quad (5.12)$$

For the choice

$$C_5 = C_7 = 0, \quad (5.13)$$

Eq. (5.12) leads to

$$\eta + \phi = C_4(z^2 - \bar{z}^2) = 4iC_4\eta$$

[from Eq. (5.7)] or

$$\phi = (4iC_4 - 1)\eta = 2\eta(\because C_4 = 3/4i).$$

Thus,

$$\frac{d\phi}{d\tau} = \frac{2}{\tau'}. \quad (5.14)$$

Substituting (5.14) in Eq. (5.10), we have

$$3\eta\tau'' + 5\tau' = 0,$$

which yields the solution, $\tau = \lambda\eta^{-2/3}$, thus

$$v(q_1,q_2) = \lambda(q_1q_2)^{-2/3}.$$

We note here that this potential was derived by Inozemtsev[5] using Holt's equations. Substituting for this potential in (4.7) and (4.1), we obtain

$$G = 3\eta, \quad a_1 = i\lambda z\eta^{-2/3}, \quad a_2 = i\lambda\bar{z}\eta^{-2/3},$$

which leads to the same invariant as given earlier in Sec. III [cf. Eq. (3.8)].

*Case (e):* If we choose the single argument $\eta$ on which $v$ depends as

$$\eta = (q_1^2 + q_2^2) \text{ or } \eta = z\bar{z} \quad (5.15)$$

and assume that

$$Y = \eta^2 = z^2\bar{z}^2, \quad v = \tau(\eta), \quad \phi = \phi(v) = \phi(\tau). \quad (5.16)$$

then the comparison of Eq. (5.16) with Eq. (4.6) implies that

$$Y_0 = C_1 = C_2 = C_4 = C_6 = 0 \text{ and } C_3 = 6,$$

which in turn lead to [cf. Eqs. (2.35)–(2.38)]

$$a_{111} = \bar{z}^3 + C_5, \quad a_{222} = -z^3 + C_7,$$
$$a_{112} = -z\bar{z}^2, \quad a_{122} = \bar{z}z^2. \quad (5.17)$$

Now making use of (5.16) and (5.17) in the potential equations (4.8) and (4.9), we obtain

$$(\eta^2+\phi)\bar{z}^2\tau'' + (2z\bar{z}^3 - C_5 z)\tau' + \bar{z}^2\tau'^2\frac{d\phi}{d\tau} = 0, \quad (5.18)$$

$$(\eta^2+\phi)z^2\tau'' + (2\bar{z}z^3 + C_7\bar{z})\tau' + z^2\tau'^2\frac{d\phi}{d\tau} = 0. \quad (5.19)$$

On multiplying (5.18) by $z^2$ and (5.19) by $\bar{z}^2$ and subtracting the results, we find

$$\tau'(C_5 z^3 + C_7\bar{z}^3) = 0.$$

But $\tau' \neq 0$. Therefore, we must have $C_5 = C_7 = 0$. Thus, both (5.18) and (5.19) reduce to ($\eta = z\bar{z}$),

$$(\eta^2+\phi)\tau'' + 2\eta\tau' + \tau'^2\frac{d\phi}{d\tau} = 0. \quad (5.20)$$

Now we discuss the solution of this nonlinear equation with three possibilities.

(1) Let us choose $\tau = \lambda_1 \eta$. Then, Eq. (5.20) reduces to the form

$$2\eta + \frac{d\phi}{d\eta} = 0 \ (\because \lambda_1 \neq 0)$$

which yields the solution as

$$\phi = -\eta^2 + k_1, \tag{5.21}$$

where $k_1$ is the integration constant. For this case, we have $G = Y + \phi = k_1$ and $a_1 = a_2 = 0$, if $k_1 = 0$; otherwise $a_1 = k_1 \bar{z} \lambda_1$, $a_2 = -k_1 z \lambda_1$. Finally, the invariant $I$ from Eq. (2.3) turns out to be

$$|I| = \tfrac{2}{3}(q_1 p_2 - q_2 p_1)[3k_1\lambda_1 - 2(q_1 p_2 - q_2 p_1)^2]. \tag{5.22}$$

(2) If we put

$$\eta^2 \tau'' + 2\eta\tau' = 0 \tag{5.23}$$

in Eq. (5.20), then it reduces to the form

$$\phi\tau'' + \tau'^2 \frac{d\phi}{d\tau} = 0. \tag{5.24}$$

The solution to Eq. (5.23) turns out to be

$$\tau = \lambda_2 \eta^{-1} \tag{5.25}$$

which leads to the solution of (5.24) as

$$\phi = \phi_0 \eta^2 \quad (\phi = \lambda_2 \phi_0 \tau^{-2}), \tag{5.26}$$

where $\phi_0$ is a positive definite integration constant. For this case, the function $G$ and the coefficients $a_1$ and $a_2$ are given by

$$G = Y + \phi = (1 + \phi_0)\eta^2,$$

$$a_1 = -(1 + \phi_0)\lambda_2 \bar{z}, \quad a_2 = (1 + \phi_0)\lambda_2 z.$$

The invariant turns out to be the same as (5.22) except $k_1$ is now replaced by another constant $-(1 + \phi_0)$.

(3) Let us consider a trial solution of (5.20) as

$$\tau = \lambda_1 \eta + \lambda_2 / \eta. \tag{5.27}$$

This provides a consistent solution of the nonlinear Eq. (5.20) provided $\phi = -\eta^2$.

For this case, we obtain

$$G = Y + \phi = 0 \text{ and } a_1 = a_2 = 0,$$

and the invariant $I$ taks a very simple form as

$$|I| = \tfrac{4}{3}(q_1 p_2 - q_2 p_1)^3. \tag{5.28}$$

The cases (1)–(3) correspond to the well-known two-dimensional harmonic oscillator Hamiltonians expressed in cylindrical coordinates and are given by

$$v(q_1, q_2) = \begin{cases} \lambda_1(q_1^2 + q_2^2) & \text{(case 1)} \\ \lambda_2(q_1^2 + q_2^2)^{-1} & \text{(case 2)} \\ \lambda_1(q_1^2 + q_2^2) + \lambda_2(q_1^2 + q_2^2)^{-1} & \text{(case 3).} \end{cases} \tag{5.29}$$

## VI. DISCUSSIONS

Basically, the paper was intended to derive third-order invariants and establish the link between our approach and that of Holt.[4] Many known potentials of physical interest have been rederived. The following additional observations can be made in the light of our analysis.

(1) In general, Hamiltonians which possess invariants of polynomial order 3 in $p_i$, may possess invariants of order $< 3$ in $p_i$ as well, e.g., if we had chosen all $a_{ijk} \equiv 0$, then the invariant $I$ would be expressed as

$$I = a_i \dot{z} + a_2 \dot{\bar{z}} \quad \text{(first-order in } p_i)$$

with $a_i$ satisfying the following equations:

$$\frac{\partial a_1}{\partial z} = 0, \quad \frac{\partial a_2}{\partial \bar{z}} = 0, \quad \frac{\partial a_1}{\partial \bar{z}} + \frac{\partial a_2}{\partial z} = 0,$$

and

$$a_1 \frac{\partial v}{\partial z} + a_2 \frac{\partial v}{\partial \bar{z}} = 0. \tag{6.1}$$

On solving for $a_1$, $a_2$, and $v$, we have

$$a_1 = C_1 \bar{z} + C_2, \quad a_2 = -C_1 z + C_3,$$

$$v = f(C_1 z\bar{z} + C_2 z - C_3 \bar{z}), \tag{6.2}$$

$C_i$ being constants and the potential $v$ is an arbitrary function of the arguments. Correspondingly, $I$ is given as

$$I = C_1(\bar{z}\dot{z} - z\dot{\bar{z}}) + C_2 \dot{z} + C_3 \dot{\bar{z}}. \tag{6.3}$$

Note that if we choose $C_2 = C_3 = 0$, then, the potential is spherically symmetric and the invariant is the angular momentum (first order in $p_i$).

The potential discussed in Sec. III B does not admit invariants of order $< 3$ in $p_i$. Similar conclusions hold good also for potentials discussed in V (a–c), e.g., in V(a), if we choose $C_2 = 0$ (implying $a_{ijk} \equiv 0$), then $a_i \equiv 0$ also. Thus, $I$ vanishes identically. However, for potentials (5.29), it is easily seen that the invariant $I$ involves odd powers of $(q_1 p_2 - q_2 p_1)$ as it should since such systems have angular momentum as a constant of the motion (first-order in $p_i$).

(2) If the Hamiltonian admitting invariants which are polynomials in $p_i$ (or, equivalently in $\dot{z}$ and $\dot{\bar{z}}$) of order 4 should also possess invariants of second order in $p_i$, the following equations are to be satisfied by $a_0$ and $a_{ij}$:

$$\frac{\partial a_{11}}{\partial z} = 0, \quad \frac{\partial a_{22}}{\partial \bar{z}} = 0, \quad \frac{\partial a_{11}}{\partial \bar{z}} + 2\frac{\partial a_{12}}{\partial z} = 0,$$

$$\frac{\partial a_{22}}{\partial z} + 2\frac{\partial a_{12}}{\partial \bar{z}} = 0 \tag{6.4}$$

and

$$\frac{\partial a_0}{\partial z} = 2a_{11}\frac{\partial v}{\partial \bar{z}} + 2a_{12}\frac{\partial v}{\partial z},$$

$$\frac{\partial a_0}{\partial \bar{z}} = 2a_{22}\frac{\partial v}{\partial z} + 2a_{12}\frac{\partial v}{\partial \bar{z}}. \tag{6.5}$$

Solutions of (6.4) yield

$$a_{12} = a_{21} = -\tfrac{1}{2}C_1 z\bar{z} - \tfrac{1}{2}C_4\bar{z} - \tfrac{1}{2}C_2 z + C_6,$$

$$a_{11} = \tfrac{1}{2}C_1\bar{z}^2 + C_2\bar{z} + C_3, \tag{6.6}$$

$$a_{22} = \tfrac{1}{2}C_1 z^2 + C_4 z + C_5.$$

Substituting (6.6) in (6.5), we obtain finally the following condition on the potential so that the system will admit invariant in second order in $p_i$ as

$$(\tfrac{1}{2}C_1\bar{z}^2 + C_2\bar{z} + C_3)^{-1/2}\frac{\partial}{\partial \bar{z}}\left[\left(\frac{1}{2}C_1\bar{z}^2 + C_2\bar{z} + C_3\right)^{3/2}\frac{\partial v}{\partial \bar{z}}\right]$$

$$-(\tfrac{1}{2}C_1 z^2 + C_4 z + C_5)^{-1/2}\frac{\partial}{\partial z}$$

$$\times\left[\left(\frac{1}{2}C_1 z^2 + C_4 z + C_5\right)^{3/2}\frac{\partial v}{\partial z}\right] = 0. \tag{6.7}$$

The solution of (6.7) will yield logarithmic potentials.[2,9]

(3) Alternatively, subsidiary conditions on the coefficients in Eqs. (3.5) and (3.6) can be obtained similar to those outlined by Holt[4] (see the Appendix) such that $v(z,\bar{z})$ does not admit invariants lower than third or fourth order in $p_i$, respectively. The derivation of a general solution to Eq. (3.5) or (3.6) is indeed very much involved. We wish to report more on this problem in a future publication.

## ACKNOWLEDGMENTS

We are extremely grateful to Professor C. R. Holt for many valued suggestions. K.C.T. wishes to thank Professor Jan. S. Nilsson for his kind hospitality at the Institute of Theoretical Physics, Göteborg, Sweden.

[1]E. T. Whittaker, *Analytical Dynamics* (Cambridge U.P., London, 1927).
[2]P. Havas, J. Math. Phys. **16**, 1461 (1975); **16**, 2476 (1975); J. Hietarinta, Phys. Lett. A **96**, 273 (1983); L. S. Hall, Physica D **8**, 90 (1983).
[3]M. Toda, J. Phys. Soc. Jpn. **23**, 501 (1967); J. Ford, S. P. Stoddard, and J. S. Turner, Prog. Theor. Phys. **50**, 1547 (1973); M. Adler, Commun. Math. Phys. **55**, 195 (1977).
[4]C. R. Holt, J. Math. Phys. **23**, 1037 (1982); B. McNamara and K. J. Whiteman, J. Math. Phys. **8**, 2024 (1967).
[5]V. I. Inozemtsev, Phys. Lett. A **96**, 447 (1983).
[6]S. Datta Majundar and M. Englefield, Int. Theor. Phys. **16**, 829 (1977); J. Hietarinta, Phys. Lett. A **93**, 55 (1982); H. J. Korsch, Phys. Lett. A **90**, 113 (1982).
[7]E. S. Nikolaevskii and L. N. Shur, Pis'ma Zh. Eksp. Teor. Fiz. **36**, 176 (1982) [JETP Lett. **36**, 218 (1982)]; G. K. Savvidy, Phys. Lett. B **130**, 303 (1983); H. M. Asatryan and G. K. Savvidy, Phys. Lett. A **99**, 290 (1983); S. J. Chang, Phys. Rev. D **29**, 259 (1984).
[8]G. N. Katzin and J. L. Levine, J. Math. Phys. **23**, 552 (1982); **24**, 1761 (1983).
[9]S. C. Mishra, R. S. Kaushal, and K. C. Tripathy, J. Math. Phys. **25**, 2217 (1984); R. S. Kaushal, S. C. Mishra, and K. C. Tripathy, Phys. Lett. A **102**, 7 (1984).
[10]It is argued by Hall (Ref. 2) and Holt (Ref. 4) and is also clear from Eqs. (2.5)–(2.10) that the third-order invariants involve only odd powers of momenta and fourth-order invariants will involve only even powers. We use this fact all throughout this work. This is not however the case with time-dependent systems, where the couplings between the coefficients, $a_0$, $a_i$, $a_{ij}$, $a_{ijk}$, and $a_{ijkl}$ corresponding to even and odd powers of momenta occur through their time derivatives.

# The anti-self-dual Coulomb field in Minkowski space-time

John R. Porter
*Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260*

(Received 9 January 1984; accepted for publication 19 July 1984)

The twistor encoding of the anti-self-dual Coulomb field is given in terms of the space-time connections pulled back to $\mathscr{I}^+$ and to $\mathscr{I}^-$. This description differs considerably from that of the twistor encoding of transverse or radiation fields, which have been the only fields studied in this fashion to date. A twisted structure results on $\mathscr{I}^+$ and a topologically incomputable one on $\mathscr{I}^-$ and these are identified modulo the null lines intersecting the source world-line.

## I. INTRODUCTION

The twistor description of the anti-self-dual Coulomb field (a left-handed field) was first given by Penrose and Sparling[1] and later amplified by these authors[2] and Penrose and Bailey.[3] In this paper the description is given in terms of structures on $\mathscr{I}$, the null infinity of conformally completed Minkowski space-time. One can see the resulting structures in a different and more conventional fashion and the generalization to the Lienard–Wiechert field at an accelerated charge can also be formulated. This analysis aids one in understanding how longitudinal fields can be incorporated into the twistor description whereas previously only radiation or transverse fields could be directly incorporated in terms of the asymptotic structure of the field.[4] This analysis can also be extended to higher-order multipole moments and to the gravitational case.

## II. THE CONSTRUCTION

Consider Minkowski space-time $M$ in advanced null coordinates $(s, r, \bar{\eta}^A, \eta^{A'})$ related to Cartesian coordinates by

$$x^{AA'} = s\,t^{AA'} + r[\bar{\eta}^A\eta^{A'}/t(\bar{\eta},\eta)], \tag{1}$$

where $t^{AA'}$ is the identity matrix, $t(\bar{\eta},\eta)\equiv t^{AA'}\bar{\eta}_A\eta_{A'}$, and the spinor fields $\bar{\eta}_A,\eta_{A'}$ are defined only mod $\mathbb{C}^*$. The coordinates $x^{AA'}$ are standard Minkowskian coordinates transformed by the Van de Warden symbols so

$$x^{AA'} = \begin{pmatrix} t-z & x+iy \\ x-iy & t+z \end{pmatrix}^{AA'},$$

and thus $t^{AA'}$ represents the components of a timelike vector field on $M$ which is parallel and has norm $\sqrt{2}$ for convenience. Graphically the coordinates of a point are determined as in Fig. 1.

A Coulomb field of a charge $e$ based on the world-line given by $r=0$ is represented by the closed two-form $e(ds\wedge dr)/r^2$ and the anti-self-dual Coulomb field is

$$F = (e\,ds\wedge dr)/r^2 + (e\,\Delta\eta\wedge\Delta\bar{\eta})/t(\bar{\eta},\eta)^2,$$

with $\Delta\eta\equiv\eta^{A'}\,d\eta_{A'}$. The field $F$ pulled back to $\mathscr{I}^+$ (given imprecisely by $r=\infty$), which has coordinates $u,\bar{\pi}_A,\eta_{A'}$ (see Ref. 5), is

$$e(\Delta\pi\wedge\Delta\bar{\pi})/t(\bar{\pi},\pi)^2,$$

and is locally of the form $-\partial\bar{\gamma}_\alpha$, where $\alpha_{A'}$ is an arbitrary spinor and the exterior derivative operator $\partial\equiv(\partial/\partial\pi_{A'})\wedge d\pi_{A'}$ [e.g., $\partial f\Delta\bar{\pi} = (\partial f/\partial\pi_{A'})d\pi_{A'}\Delta\bar{\pi}$],

$$\bar{\gamma}_\alpha = \frac{et(\bar{\alpha},\pi)}{(\bar{\alpha}\cdot\bar{\pi})t(\bar{\pi},\pi)}\Delta\bar{\pi},$$

where $\bar{\alpha}\cdot\bar{\pi}\equiv\bar{\alpha}^A\bar{\pi}_A$. Then

$$F|_{\mathscr{I}^+} = -\partial\bar{\gamma}_\alpha.$$

There is a Hermitian structure defined on $S^2$ sections of $\mathscr{I}^+$ associated with the field on $\mathscr{I}^+$

$$\bar{\gamma}_\alpha = \left[\bar{\partial}\left(\frac{\alpha\cdot\pi\,\bar{\alpha}\cdot\bar{\pi}}{t(\bar{\pi},\pi)}\right)^e\right]\left(\frac{\alpha\cdot\pi\,\bar{\alpha}\cdot\bar{\pi}}{t(\bar{\pi},\pi)}\right)^{-e} \equiv (\bar{\partial}H_\alpha)H_\alpha^{-1},$$

where

$$H_\alpha = (\alpha\cdot\pi\,\bar{\alpha}\cdot\bar{\pi}/t(\bar{\pi},\pi))^e.$$

For another fixed spinor $\beta_{A'}(\neq\alpha_{A'}$ projectively) define similar structures $\bar{\gamma}_\beta,H_\beta$. Then

$$H_\beta = \bar{\phi}_{\beta\alpha}H_\alpha\phi_{\beta\alpha},$$

with

$$\phi_{\beta\alpha} = (\beta\cdot\pi/\alpha\cdot\pi)^e.$$

Since $\bar{\gamma}_\alpha$ is not defined on all of $\mathscr{I}^+$ (being singular for $\pi_{A'}\propto\alpha_{A'}$) and $\bar{\gamma}_\beta$ is singular for $\pi_{A'}\propto\beta_{A'}$, the collection $\{\bar{\gamma}_\alpha,\bar{\gamma}_\beta\}$ covers $\mathscr{I}^+$ and the elements are related by

$$\bar{\gamma}_\beta = (\bar{\partial}\bar{\phi}_{\beta\alpha})\bar{\phi}_{\beta\alpha}^{-1} + \bar{\phi}_{\beta\alpha}\bar{\gamma}_\alpha\bar{\phi}_{\beta\alpha}^{-1}.$$

Thus provided $e\epsilon Z$ (is an integer), a holomorphic $\mathbb{C}^*$ principal bundle is specified on $\mathscr{I}^+$ with a Hermitian structure[6] whose curvature tensor is $i$ times the field on $\mathscr{I}^+$. The bundle is given by the transition function $\phi_{\beta\alpha}$ and $\{H_\alpha,H_\beta\}$ provide the Hermitian structure. The Chern class of the bundle on the $S^2$ section of $\mathscr{I}^+$ is $e$ in contrast to the bundle for a pure radiation field which has a bundle with Chern class

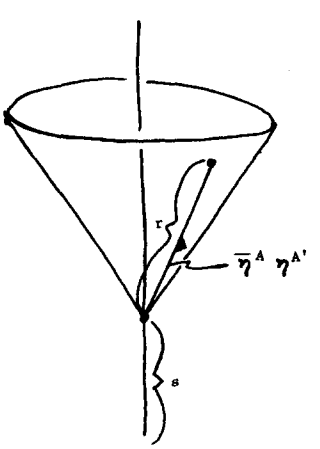



FIG. 1. Null coordinates based on a world-line in Minkowski space.

 0022-2488/85/030428-03$02.50 © 1985 American Institute of Physics 

zero. The structure on $\mathscr{I}^+$ is not sufficient to specify the field on $M$; using standard procedure of twistor theory a zero field is produced—the radiation field of the Coulomb field. The charge $e$ is specified as is the velocity vector of the charge which is given by $t^{AA'}$. This structure locates the charge world-line only up to translation.

To continue we examine the similar structure on $\mathscr{I}^-$. Using retarded null coordinates based on the source world-line $\{v = s + r, \rho = r, \mu^{A'} = t^{AA'}\bar{\eta}_A, \bar{\mu}^A = t^{AA'}\eta_{A'}\}$,

$$F = \frac{e\, dv \wedge d\rho}{p^2} - \frac{e\, \Delta\mu \wedge \Delta\bar{\mu}}{t(\bar{\mu},\mu)^2}$$

pulls back to a field on $\mathscr{I}^-$ given by

$$(-e\, \Delta\pi \wedge \Delta\bar{\pi})/t(\bar{\pi},\pi)^2.$$

Introducing $\bar{\gamma}_\alpha, \bar{\gamma}_\beta$, a twisted $\mathbb{C}^*$ bundle results on $\mathscr{I}^-$ with transistion functions given by

$$\phi_{\hat{\beta}\hat{\alpha}} = (\alpha\cdot\pi/\beta\cdot\pi)^e,$$

where the $\pi$'s label a set of parallel null hyperplanes in $M$ and correspondingly a generator of $\mathscr{I}^+$ and a generator of $\mathscr{I}^-$. Null geodesics, elements of $PN$, can be identified with first jets (7) of sections of $\mathscr{I}^+ \to S^2$, $PN^+$, and also with first jets of sections of $\mathscr{I}^- \to S^2$, $PN^-$, and one identifies $\mathscr{I}^+$ and $\mathscr{I}^-$ and their first jets, $PN^+$ and $PN^-$, to obtain a unique representation of the null geodesics, $PN$. (See Fig. 2.) The bundle over $\mathscr{I}^+$ ($\mathscr{I}^-$) pulls back to a $\mathbb{C}^*$ bundle over $PN^+$ ($PN^-$) of Chern class $+e$ ($-e$) and the identification of $PN^+$ with $PN^-$ produces no bundle over $PN$ as the bundles are topologically inequivalent, having different Chern classes.

Of course, knowledge of the field on $M$ produces a set of null geodesics which intersect the charge world-line and should be regarded as singular. To be more explicit, consider a connection on $M$ based on future null coordinates,

$$\bar{\gamma} = sr^{-e-1}d(r^e) - \frac{et(\bar{\pi},\eta)}{\bar{\pi}\cdot\bar{\eta}\, t(\bar{\eta},\eta)}\Delta\bar{\eta},$$

and a connection based on past null coordinates,

$$\hat{\bar{\gamma}} = v\rho^{-e-1}d(\rho^e) + \frac{et(\bar{\pi},\mu)}{\bar{\pi}\cdot\bar{\mu}\, t(\bar{\mu},\mu)}\Delta\bar{\mu}$$

$$= (s+r)r^{-e-1}d(r^e) - \frac{e\,\bar{\pi}\cdot\bar{\eta}}{t(\bar{\pi},\eta)t(\bar{\eta},\eta)}\Delta\eta.$$

Then $\hat{\bar{\gamma}} - \bar{\gamma}$ [more precisely, $\hat{\bar{\gamma}} = (dH)H^{-1} + H\bar{\gamma}H^{-1}$] is pure gauge and is given by

$$\hat{\bar{\gamma}} - \bar{\gamma} = d\left(-i\frac{r\, t(\bar{\pi},\eta)\bar{\pi}\cdot\bar{\eta}}{t(\bar{\eta}\,\eta)}\right)^e\left(-i\frac{r\, t(\bar{\pi},\eta)\bar{\pi}\cdot\bar{\eta}}{t(\bar{\eta},\eta)}\right)^{-e}.$$

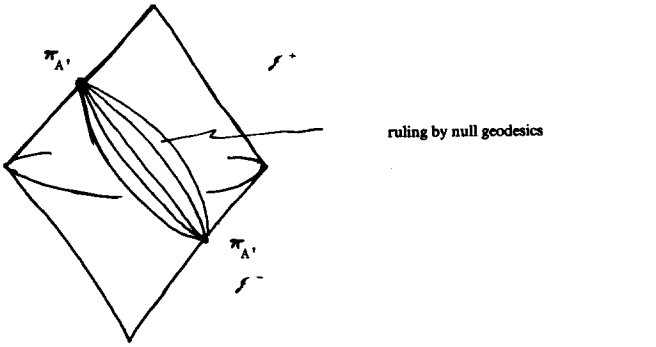



FIG. 2. Null hypersurface in conformally compactified Minkowski space.

The transition function $(-i[r\, t(\bar{\eta},\pi)\pi\cdot\eta/t(\bar{\eta},\eta)])^e$ may be regarded as defined on the primed-spin bundle $P'$ of $M$ with coordinates $(s,\ r,\ \bar{\eta}_A, \eta_{A'}, \pi_{A'})$. Notice that $x^{AA'}t_A{}^{B'}\,\pi_{A'}\pi_{B'} = r\, t(\bar{\eta},\pi)\pi\cdot\eta/t(\bar{\eta},\eta)$ from (1) and thus this transition function is the pullback to $P'$ of a function on $PN$, namely $(\omega\cdot t)^e$, where $\omega^A \equiv ix^{AA'}\pi_{A'}$ and $\omega\cdot t \equiv \omega^A t_A{}^{A'}\pi_{A'}$. The null twistor is given by $Z^\alpha = (\omega^A, \pi_{A'})$ as usual. This function on $PN$, $(\omega\cdot t)^e$, is of homogenity degree $2e$ in $Z^\alpha$ and thus may be regarded as a section of the Hopf bundle to the power $2e$ on $PN$ and is used to define a transition function between the bundle over $PN^+$ and the bundle over $PN^-$ modulo the quadric $Q$ defined by $\omega\cdot t = 0$. But this quadric consists of null geodesics intersecting the charge world-line and thus we can encode the additional information about the field in this manner.

## III. THE BUNDLE

The end result is a $\mathbb{C}^*$ bundle over $PN$, a non-Hausdorff mainfold, given as follows.[2] Construct two Hopf bundles over $PN$. The first has transition functions

$$\phi_{\beta\alpha} = (\beta\cdot\pi/\alpha\cdot\pi)^e$$

and twist $e$. The second has transition function

$$\phi_{\hat{\beta}\hat{\alpha}} = (\beta\cdot\pi/\alpha\cdot\pi)^{-e}$$

and twist $-e$.

Now identify points of $PN - Q$ in the two copies of $PN$ and identify fibers over these points using

$$\phi_{\hat{\alpha}\alpha} = (\omega\cdot t/(\alpha\cdot\pi)^2)^e$$

over the region $\alpha\cdot\pi \neq 0$ and using

$$\phi_{\hat{\beta}\beta} = (\omega\cdot t/(\beta\cdot\pi)^2)^{-e}$$

over the region $\beta\cdot\pi \neq 0$. Do not identify the points of $Q \subset PN$ and do not identify the fibers over these points. The other transition functions $(\phi_{\hat{\beta}\alpha}, \ldots)$ are obtained by composition. The resulting bundle has a holomorphic extension to the complement of $Q$ in $PT$. Thus the bundle is specified by the charge $e$ and the ruled quadric $Q$ with equation $\omega\cdot t = 0$.

If one restricts this bundle over $PN$ to the $\mathbb{C}P(1)$ of null geodesics intersecting at a generic point $p \in M$ and makes a choice of fibers over the two points of $\mathbb{C}P(1)$ taking one fiber from one copy of $PN$ and the other fiber from the other copy of $PN$, then a Hausdorff manifold results with a bundle above it given generically by the transition functions

$$\phi_{\hat{\beta}\alpha} = \phi_{\hat{\beta}\hat{\alpha}}\phi_{\hat{\alpha}\alpha} = \left(\frac{\alpha\cdot\pi}{\beta\cdot\pi}\right)^e\left(\frac{\omega\cdot t}{(\alpha\cdot\pi)^2}\right)^e = \left(\frac{\omega\cdot t}{\alpha\cdot\pi\,\beta\cdot\pi}\right)^e.$$

The resulting field $\phi_{AB}$ at the point is evaluated by considering

$$\frac{1}{2\pi i}\oint\frac{\partial}{\partial\omega^B}\left[\left(\frac{\partial}{\partial\omega^B}\phi_{\hat{\beta}\alpha}\right)\phi_{\hat{\beta}\alpha}{}^{-1}\right]\Delta\pi,$$

restricting the integrand to the $\mathbb{C}P(1)$ and choosing appropriate contours. The resulting field is the 1/2 (advanced + retarded) Coulomb field at the point $p$.

The higher multipole moments do not contribute to the twisting of the bundle over $\mathscr{I}^+$, $\mathscr{I}^-$. The $\phi_{\beta\alpha}$, $\phi_{\hat{\beta}\hat{\alpha}}$ are unchanged and $\phi_{\hat{\alpha}\alpha}$ is replaced by

$$\phi_{\hat{\alpha}\alpha} = \left(\frac{\omega\cdot t}{(\alpha\cdot\pi)^2}\right)^e \exp\left[\frac{a(\pi,\pi)}{\omega\cdot t} + \frac{b(\pi\pi\pi\pi)}{(\omega\cdot t)^2} + \cdots\right],$$

where

$$a(\pi,\pi) = a^{A'B'}\pi_{A'}\pi_{B'},$$

$$b(\pi,\pi,\pi,\pi) = b^{A'B'C'D'}\pi_{A'}\pi_{B'}\pi_{C'}\pi_{D'}$$

represent, respectively, the dipole and quadrupole moments of the source with respect to the charge world-line. Insertion into the appropriate integral produces the field at a point $p\in M$.

## IV. DISCUSSION

The anti-self-dual Coulomb field in Minkowski space-time, the prototype of a longitudinal electromagnetic field, has an encoding in terms of twistor theory which can be obtained from the asymptotic behavior of the field and its connection on $\mathscr{I}^+$ and on $\mathscr{I}^-$. A ruled quadric $Q$ is specified by the source world-line and an $O^*$ bundle results on $PN - Q$ which is inextendible to $Q$. There are two copies of $Q$ in the resulting manifold so that it is non-Hausdorff and a fiber over each of the double points. This structure is specified by considering two copies of $PN$, one arising from $\mathscr{I}^+$ and one from $\mathscr{I}^-$ each with an $O^*$ bundle with twist $+e$ and $-e$, respectively. The corresponding points of $PN - Q$ are identified as are the fibers over $PN - Q$. Over a $\mathbb{C}P(1)$ representing the null geodesics through a point of the Minkowski space-time a Hausdorff bundle can be constructed from the above non-Hausdorff bundle so that the usual twistor construction yields the Coulomb field.

Further insight can be gained by examining the Lienard–Wiechert field of an accelerating charge monopole as the radiation field is nonzero.[7]


[1] R. Penrose and G. A. J. Sparling, *Advances in Twistor Theory*, edited by L. Hughston and R. S. Ward (Pitman, London, 1979), p. 136.
[2] R. Penrose and G. A. J. Sparling, Twistor News Lett. **9**, 18 (1979).
[3] T. N. Bailey and R. Penrose, Twistor News Lett. **14**, 9, 19, 22 (1982).
[4] J. R. Porter, J. Math. Phys. **24**, 1224 (1983).
[5] E. T. Newman and K. P. Tod, *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980), Vol. 2.
[6] K. Kodaira and J. Marrow, *Complex Manifolds* (Holt, Rinehart, and Winston, New York, 1971), p. 83.
[7] J. R. Porter, J. Math. Phys. **26**, 431 (1985).

# A twistor encoding of Lienard–Wiechert fields in Minkowski space-time

John R. Porter
*Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260*

(Received 9 January 1984; accepted for publication 19 July 1984)

The twistor encoding of the anti-self-dual Lienard–Wiechert field on Minkowski space-time yields a considerably richer structure than that of the Coulomb field encoding due to the presence of a nonzero radiation field. The combination of advanced and retarded transverse fields together with the longitudinal field and the individual aspects of these fields provides this structure. Higher-order longitudinal moments can be incorporated so that general longitudinal fields can be given a twistor description.

## I. INTRODUCTION

In an effort to gain a better insight into the twistor construction of the anti-self-dual Coulomb field on Minkowski space and because of considerable interest in its own right, a study of the Lienard–Wiechert field was undertaken. The encoding follows the same general lines as that of the Coulomb field but given the additional presence of a nonzero radiation field is of much intrinsic interest. One can observe how the various combinations of advanced and retarded fields are separately encoded. Additionally, a more complex field with higher-order multipole moments can be easily added so that general longitudinal fields in electromagnetism can be given a twistor description. Penrose and Bailey[1] have also examined this field and give a very elegant cohomological description.

The electromagnetic fields associated with an accelerating charge in Minkowski space-time $M$ are specified by giving the charge world-line and a value for the charge. Denote the source world-line $x^{AA'} = \tau^{AA'}(s)$, which is parametrized by $s$ where $\tau_s^{AA'} \equiv (d/ds)\tau^{AA'} \equiv t^{AA'}$. The parameter $s$ is chosen so that $t^{AA'}t_{AA'} = 2$ and thus $((d/ds)t^{AA'})t_{AA'} = 0$. Null polar coordinates based on the source world-line are chosen so that with respect to the standard Minkowskian coordinates,

$$x^{AA'} = \tau^{AA'}(s) + r\bar{\eta}^A\eta^{A'}/t(\bar{\eta},\eta),$$

where $t(\bar{\eta},\eta) \equiv t^{AA'}\bar{\eta}_A\eta_{A'}$. These are illustrated in Fig. 1. A coordinate system $(s,\bar{\pi},\pi)$ is induced on $\mathscr{I}^+$; the $s$ constant hypersurfaces are the intersection of the null cones from the source world-line with $\mathscr{I}^+$ and on such an intersection the points are labeled by $\pi_{A'} = \eta_{A'}$ (but it is convenient to maintain the distinction between $\pi$ and $\eta$). The retarded field pulled back to $\mathscr{I}^+$, $F$, results in a local connection $\bar{\gamma}_\alpha$ such that $F = -d\bar{\gamma}_\alpha$ with

$$\bar{\gamma}_\alpha = \frac{et(\bar{\alpha},\pi)}{\bar{\alpha}\cdot\bar{\pi}\, t(\bar{\pi},\pi)}\Delta\bar{\pi}$$

for an arbitrary spinor $\alpha_{A'}$ where $\bar{\alpha}\cdot\bar{\pi} \equiv \bar{\alpha}^A\bar{\pi}_A$. Noting that $t = t(s)$,

$$F = \frac{e}{t(\bar{\pi},\pi)^2}\Delta\pi\wedge\Delta\bar{\pi} - \frac{e\dot{t}_A{}^{A'}t^{AB'}\pi_{A'}\pi_{B'}}{t(\bar{\pi},\pi)^2}ds\wedge\Delta\bar{\pi}.$$

The first term represents the $1/r^2$ part of the field pulled back to $\mathscr{I}^+$ while the second term represents the $1/r$ part of the pullback.

Doing the same for another arbitrary spinor $\beta_{A'}$ results in $\bar{\gamma}_\beta$ and the relation

$$\bar{\gamma}_\beta - \bar{\gamma}_\alpha = \left[\bar{\partial}\left(\frac{\bar{\beta}\cdot\bar{\pi}}{\bar{\alpha}\cdot\bar{\pi}}\right)^e\right]\left(\frac{\bar{\beta}\cdot\bar{\pi}}{\bar{\alpha}\cdot\bar{\pi}}\right)^{-e}.$$

[More appropriately $\bar{\gamma}_\beta = d(\ )(\ )^{-1} + (\ )\bar{\gamma}_\alpha(\ )^{-1}$.] Thus the appropriate bundle on $\mathscr{I}^+$ has transition functions

$$\phi_{\beta\alpha} = (\beta\cdot\pi/\alpha\cdot\pi)^e, \tag{1}$$

with $e\epsilon Z$ so that the bundle is the pullback via $\mathscr{I}^+ \to S^2$ of a bundle over $S^2$ with Chern class $e$. A similar construction on $\mathscr{I}^-$ results in a bundle with Chern class $-e$ and transition function

$$\phi_{\hat{\beta}\hat{\alpha}} = (\beta\cdot\pi/\alpha\cdot\pi)^{-e}. \tag{2}$$

The standard twistor construction for a field from $\bar{\gamma}_\alpha$ results in a pure radiation field with zero charge. However, $\bar{\gamma}_\alpha$ is defined on a piece of a twisted bundle on $\mathscr{I}^+$ and is only a local expression for a $\bar{\partial}$ closed (0,1) form pulled back from $PN$.

Examining the source world-line from the perspective of $\mathscr{I}^+$, it corresponds to a selection of unique null geodesic at each point of $\mathscr{I}^+$, namely the null geodesic at that point of $\mathscr{I}^+$ that intersects the source world-line. This is equivalent to specifying at each point of $\mathscr{I}^+$ a first jet of a section of $\mathscr{I}^+ \to S^2 \simeq \mathbb{C}P(1)$. Now $\mathbb{C}P(1)$ has a complex structure given nonprojectively in the coordinates $(\pi_{A'},\bar{\pi}_A)$ by $\partial/\partial\bar{\pi}_A$ as the antiholomorphic vector field [modulo the homogeneity operator $\bar{\pi}_A(\partial/\partial\bar{\pi}_A)$ where the representation is $\mathbb{C}P(1) \simeq \mathbb{C}^2 - \{0\}/\mathbb{C}^*$ and the homogeneity operator is tan-

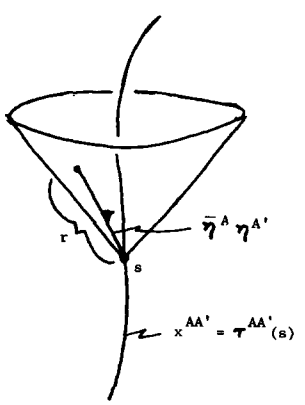



FIG. 1. Null coordinates based on a curve in Minkowski space.

 0022-2488/85/030431-03$02.50 © 1985 American Institute of Physics 

gent to the fibration $\mathbb{C}^*$]. The manifold $\mathscr{I}^+$ has no natural $CR$ structure but a shear-free congruence in $M$ endows $\mathscr{I}^+$ with a $CR$ structure—namely the congruence determines a $CR$ submanifold of $N$ (or $PN$).

Specifically, giving a future null coordinate system based on a straight world-line results in a coordinate system $(\lambda,\bar{\pi}_A,\pi_{A'})$ on $\mathscr{I}^+$ and a timelike parallelly propagated vector field $v^{AA'}$. Let $u=\lambda v(\bar{\pi},\pi)$. This gives a $CR$ structure on $\mathscr{I}^+$ whose antiholomorphic vector field is $(\partial/\partial\bar{\pi}_A)|_{\lambda=\text{const}}$. The $CR$ structure determined by the source world-line is related to a change of the $CR$ structure on $\mathscr{I}^+$: $\partial/\partial\bar{\pi}_A \to \partial/\partial\bar{\pi}_A - \bar{\pi}^A L(\partial/\partial\lambda)$, where $L(\lambda,\bar{\pi},\pi)$ is of degree $(0,-2)$ in $(\pi_{A'},\bar{\pi}_A)$. Given the straight world-line, a null twistor is defined at each point of $\mathscr{I}^+$, $\omega_0^A = i\lambda v^A$, where $v^A \equiv v^{AA'}\pi_{A'}$ and $\omega_0^A\bar{\pi}_A = i\lambda v \equiv iu$. The source world-line gives $\omega_1^A = i\tau^A(s) = i\tau^{AA'}(s)\pi_{A'}$, $\omega_1^A\bar{\pi}_A = iu = i\tau(s)$. Thus

$$u=\lambda v=\tau(s)$$

gives implicitly $s=f(\lambda,\bar{\pi},\pi)$. Writing $\omega_1^A$ in terms of $\bar{\pi}^A$ and $v^A$ gives

$$\omega_1^A = \lambda v^A - Lv\bar{\pi}^A = \omega_0^A - Lv\bar{\pi}^A,$$

with

$$\omega_1^A v_A = Lv^2 = \tau^A(s)v_A = \tau^A(f(\lambda,\bar{\pi},\pi))v_A,$$

so that $Lv^2=\tau^A v_A$. The claim is that

$$\bar{\delta}^A \equiv \frac{\partial}{\partial\bar{\pi}_A} - L\bar{\pi}^A\frac{\partial}{\partial\lambda} \tag{3}$$

is tangent to the sections of $\mathscr{I}^+$ given by $s$ constant. The claim is substantiated by computing the normal one-form to the sections $s=$ const, $ds$, and acting on $\bar{\delta}^A$ with this form. Since $d(\lambda v)=d(\tau(s))$ and $ds=df(\lambda,\bar{\pi},\pi)$, $ds$ can be written as

$$ds = \frac{v}{t}d\lambda + \left(\frac{\lambda v^A - \tau^A}{t}\right)d\bar{\pi}_A + \left(\frac{\lambda v^{A'} - \tau^{A'}}{t}\right)d\pi_{A'}.$$

Applying $\bar{\delta}^A$ from (3) to $ds$ gives

$$\frac{\lambda v^A - \tau^A}{t} - L\bar{\pi}^A\frac{v}{t} = \frac{\lambda v^A - \tau^A}{t} - \frac{\tau^B v_B}{v^2}\bar{\pi}^A\frac{v}{t}.$$

That this is zero can be seen by contracting with $\bar{\pi}_A$ which gives $u-u$ and with $v_A$ which gives $-\tau^A v_A/t + (\tau^B v_B/v^2)v(v/t)$. Also note that $-(v_A/v)\bar{\delta}^A = \eth - L(\partial/\partial\lambda)$.

If $\tau(s)$ is not real analytic but, say $\mathbb{C}^\infty$, the inverse, canonically exists, $s=f(\lambda,\bar{\pi},\pi)$ and is $\mathbb{C}^\infty$ in $\lambda$ and $\omega^A v_A - L(\lambda,\bar{\pi},\pi)v^2$ is defined on real jets of $\mathscr{I}^+$, on $PN$, and the zeros of this function specify the source world-line in terms of null geodesics intersecting that world-line. The zeros also specify a three-dimensional ruled surface in $PN$ (of dimension 5) ruled by complex manifolds each of which consists of all null geodesics through a point of the source world-line. Given the map $PN \to \mathscr{I}^+$, $\bar{\delta}^A$ pulls back (mod homogeneity) to the $CR$ vector field defined on this three-dimensional submanifold of $PN$ given by the intrinsic $CR$ structure of $PN$ restricted to the submanifold.

Additionally, $\bar{\delta}^A = (\partial/\partial\bar{\pi}_A)|_{s=\text{const}}$ and so $\bar{\delta}^A(Lv^2) = (\partial/\partial\bar{\pi}_A)(\tau^A(s)v_A)|_{s=\text{const}} = 0$ and this is proportional to the difference in the asymptotic shears of the two congruences given by $s=$ const and by $\lambda=$ const. The quantity $-(v_{A'}/v)\delta^{A'}L + (v_A/v)\bar{\delta}^A\bar{L}$ is proportional to the difference in the twist of the two congruences and is given by $(-(\tau^{AA'}v_A v_{A'}/v) + (\tau/v)) -$ c.c. and is, of course, zero for the congruences involved.

## II. THE BUNDLE

The construction of the bundle representing the Lienard–Wiechert fields proceeds in a fashion similar to the Coulomb case. The two oppositely twisted bundles on $\mathscr{I}^+$ and on $\mathscr{I}^-$ are each pulled back to $PN^+$ and $PN^-$, respectively, and the identification between them is made on the complement of the ruled quadric $Q$ defined by $w^A v_A - Lv^2 = 0$ using

$$\phi_{\hat{a}\alpha} = (w^A v_A - Lv^2)/(\alpha\cdot\pi)^2)^e \tag{4}$$

and compositions of the transition functions to define the rest.

The radiation fields are obtained by the following construction. The function $L$ satisfies

$$\bar{\delta}^A(Lv^2) = 0,$$

so

$$\frac{\partial(Lv^2)}{\partial\bar{\pi}_A} - L\bar{\pi}^A\frac{\partial(Lv^2)}{\partial\lambda} = 0.$$

Now $eL_\lambda = (Lv^2)^{-e}\eth(Lv^2)^e$, where $\eth = -(v_A/v)(\partial/\partial\bar{\pi}_A)$. Then

$$eL_\lambda\,\Delta\bar{\pi} = (Lv^2)^{-e}\frac{\partial}{\partial\bar{\pi}_A}(Lv^2)^e\,d\bar{\pi}_A = \bar{\gamma} \tag{5}$$

is the connection on $\mathscr{I}^+$ for the radiation field and gives the same $1/r$ part of the field as $\bar{\gamma}_\alpha$ when pulled back to $\mathscr{I}^+$ [conveniently expressed in $(s,\bar{\pi},\pi)$ coordinates on $\mathscr{I}^+$]. But this connection results in a field with zero charge. This is easily seen to be the case by examining the field on $\mathscr{I}^+$, $d\bar{\gamma}$ and pulling it back to a $\lambda=$ const section which is a two-sphere, then integrating the two-form over the two-sphere. The relevant expression is

$$\int_{\lambda=\text{const}} d\bar{\gamma} = \int \frac{\partial\bar{\gamma}}{\partial\pi_{A'}}\wedge d\pi_{A'} = 0$$

by Stokes theorem and the fact that $\bar{\gamma}$ is a globally defined one-form. In addition, $\bar{\gamma}$ represents an element of

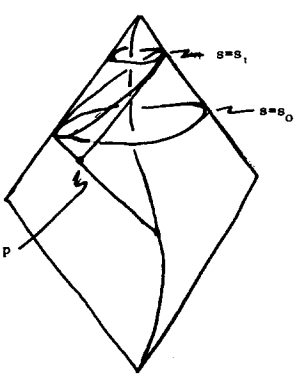



FIG. 2. Extent of $\mathscr{I}$ data for the field at a point $P$ in Minkowski space.

$H^1_{CR}(PN, \mathcal{O}^*)$ as it should in the Dolbeault representation.

One might suspect that (5) gives only advanced information about the field at a point $p \in M$ but in fact it gives the 1/2 (advanced-retarded) field at this point. This can be seen by examining the space-time diagram in Fig. 2, where one sees that the $\mathbb{C}P(1)$ integration that gives the field at $p$ contains the information on the source motion from $s = s_0$ to $s = s_1$ which contains both the advanced and the retarded field information at $p$.

The generalization to higher-order multipoles proceeds as in the Coubomb case and presents no difficulties.

[1]T. N. Bailey and R. Penrose, Twistor News Lett. **14**, 9, 16, 19, 22 (1982).

# Nontrivial zeros of the Wigner (3-*j*) and Racah (6-*j*) coefficients. I. Linear solutions

Simcha Brudno
*Chicago College of Osteopathic Medicine, Chicago, Illinois 60615*

(Received 17 July 1984; accepted for publication 4 October 1984)

Some formuals for nontrivial zeros in the 3-*j* and 6-*j* symbols have been found.

## I. INTRODUCTION

Current work in the fields of algebraic theory and quantum mechanics has underscored the importance of Racah–Wigner algebra and, thereby, of the 3-*j* and 6-*j* symbols. In the ninth volume of the *Encyclopedia of Mathematics and Its Applications*, an entire topic[1] is devoted to the nontrivial zeros of these symbols. The problem of finding such zeros through a non-numerical method is the subject of this paper.

## II. PROCEDURE

The 3-*j* symbol of Wigner[2] has been given explicitly by Racah.[3] As modified by Rotenberg,[4] this symbol has the form

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} = (-1)^{j_1-j_2-m_3}\left(\frac{(j_1+j_2-j_3)!(j_1-j_2+j_3)!(-j_1+j_2+j_3)!(j_1+m_1)!(j_1-m_1)!(j_2+m_2)!(j_2-m_2)!(j_3+m_3)!(j_3-m_3)!}{(j_1+j_2+j_3+1)!}\right)^{1/2}$$
$$\times\sum_k \frac{(-1)^k}{k!(j_1+j_2-j_3-k)!(j_1-m_1-k)!(j_2+m_2-k)!(j_3-j_2+m_1+k)!(j_3-j_1-m_2+k)!}. \tag{1}$$

The 6-*j* symbol has also been explicitly defined by

$$\begin{Bmatrix} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{Bmatrix} = (-1)^{j_1+j_2+l_1+l_2}\Delta(j_1j_2j_3)\Delta(l_1l_2j_3)\Delta(l_1j_2l_3)\Delta(j_1l_2l_3)$$
$$\times\sum_k (-1)^k(j_1+j_2+l_1+l_2+1-k)![k!(j_1+j_2-j_3-k)!(l_1+l_2-j_3-k)!(j_1+l_2-l_3-k)!$$
$$\times(l_1+j_2-l_3-k)!(-j_1-l_1+j_3+l_3+k)!(-j_2-l_2+j_3+l_3+k)!]^{-1}, \tag{2a}$$

where

$$\Delta(abc) = \left[\frac{(a+b-c)!(a-b+c)!(-a+b+c)!}{(a+b+c+1)!}\right]^{1/2}. \tag{2b}$$

In analyzing the 3-*j* and 6-*j* coefficients we note the following: Opposed to the trivial zeros resulting from symmetry conditions (3-*j* symbol) or violations of one or more triangle conditions (3-*j* or 6-*j* symbols), there also exists another class of zeros, called nontrivial zeros. The nontrivial zeros are zeros of the "polynomial part" of Wigner or Racah coefficients. A table of presently known[5,6] nontrivial zeros is given in the ninth volume of the *Encyclopedia of Mathematics*, pp. 420–428. From these tables it is possible to deduce some algebraical formulas, each of which can be used to calculate nontrivial zero solutions.

When the polynomial part of the Wigner and Racah coefficients has only two terms, i.e., when the summation is carried out over $k = 0,1$, we call the resultant expression "linear." Formulas defining the nontrivial zeros of this linear expression have been obtained. For example, a pair of linear formulas for the 3-*j* coefficients are

$$\begin{pmatrix} 3n & 2n+1 & n+1 \\ 3n-1 & -2n & 1-n \end{pmatrix} \tag{3}$$

and

$$\begin{pmatrix} 2n+1 & 2n & 2 \\ n+1 & -n & -1 \end{pmatrix}. \tag{4}$$

For the 6-*j* coefficients, the following linear formulas exist:

$$\begin{Bmatrix} n+2 & n+1 & 2 \\ n & n+1 & n+1 \end{Bmatrix}, \tag{5}$$

$$\begin{Bmatrix} 3n/2+2 & 3n/2+2 & n+2 \\ n+\frac{3}{2} & \frac{3}{2} & 3(n+1)/2 \end{Bmatrix}, \tag{6}$$

and all their Regge symmetries. From (5) we obtain

$$\begin{Bmatrix} n+2 & (1+3n)/2 & (3+n)/2 \\ n & (3+n)/2 & (3+n)/2 \end{Bmatrix}, \tag{5'}$$

and from (6), for $n = 0$, we calculate the first nontrivial zero of the Racah coefficient,

$$W(\tfrac{3}{2}, 2, \tfrac{3}{2}, 2; \tfrac{3}{2}, 2).$$

In all these expressions for the Racah coefficients (3)–(5'), at least two of the six angular momenta are equal. This relationship, however, is not necessary. For example,

$$\begin{Bmatrix} 2 & 7 & 6 \\ 5.5 & 2.5 & 3.5 \end{Bmatrix},$$

 0022-2488/85/030434-02$02.50 © 1985 American Institute of Physics 

which is a specific example of

$$\begin{Bmatrix} J_1 & 4J_1-1 & 3J_1 \\ 2J_1+\frac{3}{2} & J_1+\frac{1}{2} & 2J_1-\frac{1}{2} \end{Bmatrix}, \tag{7}$$

where $J=2$, has a nontrivial zero, but none of the angular momenta are equal.

A more general parametrical formula for the linear 3-$j$ coefficients is given by

$$J_1 = a(b+c)/2, \tag{8a}$$

$$J_2 = d(b+c)/2, \tag{8b}$$

$$J_3 = (b+c)(a+d)/2-1, \tag{8c}$$

$$m_1 = a(c-b)/2, \tag{8d}$$

$$m_2 = d(c-b)/2, \tag{8e}$$

where the parameters $a=1,\ldots,\infty$, $b=1,\ldots,\infty$, $c=1,\ldots,\infty$, and $d=1,\ldots,\infty$.

A more general parametrical formula for the linear 6-$j$ coefficients is given by

$$J_1 = (def+adg+abc)/2-\tfrac{1}{2}, \tag{9a}$$

$$J_2 = (abc+ghi+beh)/2-\tfrac{1}{2}, \tag{9b}$$

$$J_3 = (def+adg+ghi+beh)/2-1, \tag{9c}$$

$$L_1 = (ghi+adg)/2, \tag{9d}$$

$$L_2 = (def+beh)/2, \tag{9e}$$

$$L_3 = (abc+adg+beh)/2-\tfrac{1}{2}, \tag{9f}$$

where the condition

$$abc+def+ghi+adg+beh=cfi \tag{10}$$

must be satisfied and where $a, b, c, d, e, f, g$, and $h$ go from $1,\ldots,\infty$.

The following are two examples. In the first example, for $a=b=d=e=g=h=1$ and $c=f=i=2$, the smallest case arising from (9) and (10) is obtained:

$$\begin{Bmatrix} 2 & 2 & 2 \\ 1.5 & 1.5 & 1.5 \end{Bmatrix} \equiv 0.$$

In the second example, for $a=c=d=h=1$ and $b=4$, $e=3, f=6, g=2$, and $i=9$,

$$\begin{Bmatrix} 11.5 & 16.5 & 24 \\ 10 & 15 & 8.5 \end{Bmatrix} \equiv 0.$$

All the known nontrivial linear zeros given in both tables previously cited[5,6] are obtained by this method; proof that these equations constitute the total solution to the linear nontrivial-zero problem, however, has yet to be demonstrated.

## APPENDIX: DERIVATION OF SOLUTION SETS (8) AND (9)

For the first two terms in the summation of (1) (the polynomial part) to sum to 0, the following sufficient condition exists:

$$F=(j_1+j_2-j_3)(j_1-m_1)(j_2+m_2)$$
$$=(j_3-j_2+m_1+1)(j_3-j_1-m_2+1), \tag{A1}$$

where $F$ can be decomposed into four multiplicands,

$$F=abcd.$$

In order that the summation of (1) consist of only the first two terms ($k=0,1$), one of the first three multiplicands of (A1) must be equal to (1). Without loss of generality,

$$j_1+j_2-j_3=1. \tag{A2a}$$

Taking one particular partition of $F$, namely, $(ab)(cd)$,

$$j_1-m_1=ab, \tag{A2b}$$

$$j_2+m_2=cd, \tag{A2c}$$

$$j_3-j_2+m_1+1=ac, \tag{A2d}$$

$$j_3-j_1-m_2+1=bd. \tag{A2e}$$

The five equations of (A2), with the determinant not equal to zero, yield solution set (8) having 72 symmetries, i.e., 72 nontrivial zeros. The other two remaining partitions, namely $(ac)(bd)$ and $(ad)(bc)$, yield similar solution sets, with each set, again, having 72 symmetries. The combination of these three solution sets yields 216 symmetries, i.e., 216 nontrivial zeros of the 3-$j$ coefficients, all of these derived from a single $a$, $b$, $c$, and $d$. It must be remembered, however, that each $F$ can be decomposed into $abcd$ in many ways and that for each $abcd$ there are 216 nontrivial zeros, all of which may be different.

Solution sets for the 6-$j$ coefficients, e.g., solution set (9), may be derived using an analogous method.


## ACKNOWLEDGMENTS

The author thanks Dennis Bartley for material assistance and Thomas Glonek for critical review.



[1] L. C. Biedenharn and J. D. Louck, "The Racah–Wigner Algebra in Quantum Theory," in *Encyclopedia of Mathematics and its Applications*, edited by G -C. Rota (Addison–Wesley, Reading, MA, 1981), Vol. 9, pp. 415–416.

[2] E. P. Wigner, "On the Matrices which Reduce the Kronecker Products of Representations of Simply-Reducible Groups," in *Quantum Theory of Angular Momentum*, edited by L. C. Biedenharn and H. van Dam (Academic, New York, 1965), pp. 87–133.

[3] G. Racah, Phys. Rev. **62**, 438 (1942).

[4] M. Rotenberg, R. Bivins, N. Metropolis, and J. K. Wooten, Jr., *The 3-j and 6-j Symbols* (Technology, MIT, Cambridge, MA, 1959).

[5] D. A. Varshalovich, A. N. Moskalev, and V. K. Kersonskii, *Quantum Theory of Angular Momentum* (Nauka, Leningrad, 1975) (in Russian).

[6] M. J. Bowick, Thesis, University of Canterbury, Christchurch, New Zealand, 1976.

# Three-dimensional inverse scattering: High-frequency analysis of Newton's Marchenko equation

Margaret Cheney
*Department of Mathematics, Stanford University, Stanford, California 94305*

James H. Rose
*Ames Laboratory—United States Department of Energy, Iowa State University, Ames, Iowa 50011*

(Received 9 August 1984; accepted for publication 4 October 1984)

We obtain a high-frequency asymptotic expansion of Newton's Marchenko equation for three-dimensional inverse scattering. We find that the inhomogeneous term contains the same high-frequency information as does the Born approximation. We show that recovery of the potential via Newton's Marchenko equation plus the "miracle" depends on low-frequency information.

## I. INTRODUCTION

The exact inverse problem for Schrödinger potential scattering in three dimensions has a simple exact solution (even in the presence of bound states) based on the Born approximation.[1] This inversion method uses high-frequency scattering data in the near-forward direction. Considerable effort has been expanded in developing other exact inverse methods which do not depend exclusively on high-frequency data. For example, Newton has recently introduced a method[1-3] which generalizes the one-dimensional Marchenko equation.[1] Newton's Marchenko equation is derived in a way which depends on data at all frequencies, in contrast to the inverse Born approximation, which depends on high-frequency data alone. It is quite natural to ask about the relationship between the two methods. To what extent does Newton's Marchenko method rely on high-frequency data? To answer this question, we have undertaken an asymptotic high-frequency analysis of Newton's Marchenko integral equation. It will be shown that all asymptotic high-frequency information is contained in the inhomogeneous term of the Newton–Marchenko integral equation. This information is sufficient[4] to reconstruct the potential exactly by means of the Radon transform.

The Newton–Marchenko method, however, does not involve the Radon transform. Rather, the wave field everywhere is first recovered from the scattered amplitude using the Newton–Marchenko equation. Finally, the potential is extracted from the high-frequency asymptotics of the wave field using an equation dubbed the "miracle" by Newton. As will be shown, the method just described relies essentially on low- as well as high-frequency components of the scattering amplitude. References 4 and 5 provide a physical discussion of the "miracle." Note that throughout this paper it is assumed that the scattering amplitudes considered were generated from some local potential. That is, the characterization problem is not considered.

The structure of this paper is as follows. In Sec. II we introduce our notation and Newton's Marchenko equation in the frequency and time domains. In Sec. III we complete an asymptotic high-frequency analysis. The determination of the potential from the inhomogeneous term alone via the Radon transform is discussed. Next, we show that Newton's method of recovering the potential relies on the low-frequency content of the data. The final section qualitatively discusses the physical meaning of the inhomogeneous term in the far-field and weak scattering limits. A theorem and lemma on the high-frequency asymptotic are proven in the Appendix.

## II. REVIEW OF NEWTON'S MARCHENKO METHOD

This section states the problem, introduces our notation, and reviews Newton's approach. Both frequency domain results and a time domain interpretation are treated.

Consider the time-independent Schrödinger equation

$$-\Delta\psi(k,\mathbf{x}) + V(\mathbf{x})\psi(k,\mathbf{x}) = k^2\psi(k,\mathbf{x}). \tag{2.1}$$

Here the coordinate $\mathbf{x}$ is a vector in $R^3$, the potential $V(\mathbf{x})$ is real valued and decays at infinity, and $k$ is a scalar. We assume that $V$ induces no bound states. Scattering solutions are defined by the Lippman–Schwinger equation

$$\psi^{\pm}(k,\hat{e},\mathbf{x}) = \exp(ik\hat{e}\cdot\mathbf{x}) - \int (4\pi|\mathbf{x}-\mathbf{y}|)^{-1} \times \exp(\pm ik|\mathbf{x}-\mathbf{y}|)V(\mathbf{y})\psi(k,\hat{e},\mathbf{y})d^3\mathbf{y}. \tag{2.2}$$

Here the incident wave is a plane wave in direction $\hat{e}$, where $\hat{e}$ is a point on the unit sphere. The incoming and outgoing solution $\psi^-$ and $\psi^+$ are related not only by the relation

$$\psi^+(-k,-\hat{e},\mathbf{x}) = \psi^-(k,\hat{e},\mathbf{x}), \tag{2.3}$$

but also by[1]

$$\psi^+(k,\hat{e},\mathbf{x}) - \psi^-(k,\hat{e},\mathbf{x}) = -ik(8\pi^2)^{-1}\int_{S^2}\int \exp(-ik\hat{e}'\cdot\mathbf{y})V(\mathbf{y}) \times \psi^+(k,\hat{e},\mathbf{y})d\mathbf{y}\ \psi^-(k,\hat{e}',\mathbf{x})d\hat{e}'. \tag{2.4}$$

Here we have used Eqs. (10.112) and (10.114) of Ref. 1.

Recall that the scattering amplitude is given by

$$A(k,\hat{e}',\hat{e}) = -(4\pi)^{-1}\int \exp(-ik\hat{e}'\cdot\mathbf{x})V(\mathbf{x}) \times \psi^+(k,\hat{e},\mathbf{x})d\mathbf{x}. \tag{2.5}$$

In what follows, we will use the notation $\psi = \psi^+$ and

$$\beta(k,\hat{e},\mathbf{x}) = \psi(k,\hat{e},\mathbf{x})\exp(-ik\hat{e}\cdot\mathbf{x}). \tag{2.6}$$

The following high-energy asymptotic expansion of $\beta$ is known[5] for $0<\epsilon<\frac{1}{2}$ and for $k$ large:

$$\beta(k,\hat{e},x) = 1 + (ik)^{-1}B(\hat{e},\mathbf{x}) + g(k,\hat{e},\mathbf{x}), \tag{2.7}$$

 0022-2488/85/030436-04$02.50 © 1985 American Institute of Physics 

where

$$B(e,\mathbf{x}) = \tfrac{1}{2}\int_0^\infty V(\mathbf{x} - r\hat{e})dr,$$

and where $|g| < c|k|^{-1-\epsilon}$.

This implies that once $\beta$ is known, the potential can be recovered by the following procedure. We first isolate the $k^{-1}$ coefficient:

$$(2i)^{-1}\int_0^\infty V(\mathbf{x} - r\hat{e})dr = \lim_{k\to\infty} k\,[\beta(k,\hat{e},\mathbf{x}) - 1]. \quad (2.8)$$

The potential can be recovered from (2.8) by noting that

$$\hat{e}\cdot\nabla V(\mathbf{x} - r\hat{e}) = -\frac{\partial}{\partial r}V(\mathbf{x} - r\hat{e}).$$

When we apply $2i\hat{e}\cdot\nabla$ to the left side of (2.8), evaluation of the integral at the lower limit gives us the potential. We have thus obtained the formula

$$V(\mathbf{x}) = 2i\hat{e}\cdot\nabla \lim_{k\to\infty} k\,[\beta(k,\hat{e},\mathbf{x}) - 1]. \quad (2.9)$$

For future reference let us write some of the above equations in the time domain.[4] We use the Fourier transform

$$u(t,\hat{e},\mathbf{x}) = (2\pi)^{-1}\int_{-\infty}^\infty \exp(-ikt)\psi(k,\hat{e},\mathbf{x})dk. \quad (2.10)$$

The Lippmann–Schwinger equation (2.2) implies

$$\psi(k,\hat{e},\mathbf{x}) = \psi^*(-k,\hat{e},\mathbf{x}),$$

where * denotes complex conjugation. The time-domain wave field $u$ is therefore real. If $\psi$ satisfies (2.2), then $u$ defined by (2.10) satisfies[4]

$$\left[\Delta - \frac{\partial^2}{\partial t^2} - V(\mathbf{x})\right]u(t,\hat{e},\mathbf{x}) = 0. \quad (2.11)$$

The Fourier transform of $\beta$ [Eq. (2.6)] is $u(t + \hat{e}\cdot\mathbf{x},\hat{e},\mathbf{x})$. Fourier transformation of the asymptotic expansion (2.7) is then

$$u(t + \hat{e}\cdot\mathbf{x},\hat{e},\mathbf{x}) = \delta(t) - \tfrac{1}{2}B(\hat{e},\mathbf{x})\,\mathrm{sgn}(t) + (\text{smoother terms}). \quad (2.12)$$

However, we know that $\beta$ is well-behaved for small $k$, and is analytic in the upper half $k$ plane. This low-frequency behavior allows us to deduce[5] a more accurate version of (2.12), namely

$$u(t + \hat{e}\cdot\mathbf{x},\hat{e},\mathbf{x}) = \delta(t) - B(\hat{e},\mathbf{x})H(t) + h(t,\hat{e},\mathbf{x}), \quad (2.13)$$

where $H$ is the Heaviside function [$H(t) = 1$ for $t$ positive and zero for $t$ negative] and $h$ is a continuous function that vanishes for $t \leqslant 0$. Equations (2.12) and (2.13) are not contradictory; Eq. (2.13) merely contains more information about the "smoother terms" of (2.12).

From (2.13) it is evident that $u$ satisfied the causality condition

$$u(t,\hat{e},\mathbf{x}) = 0 \text{ for } t < \hat{e}\cdot\mathbf{x}. \quad (2.14)$$

We shall write

$$\eta(t,\hat{e},\mathbf{x}) = u(t + \hat{e}\cdot\mathbf{x},\hat{e},\mathbf{x}) - \delta(t). \quad (2.15)$$

We now return to the frequency domain for a moment to derive the Newton–Marchenko equation of inverse scattering.[2] We multiply (2.4) by $\exp(-ik\hat{e}\cdot\mathbf{x})$ and use (2.3) in (2.4):

$$\begin{aligned}\beta(k,\hat{e},\mathbf{x}) - \beta(-k-\hat{e},\mathbf{x}) &= -k(2\pi i)^{-1}\int_{S^2} A(k,\hat{e}',\hat{e})\exp[ik(\hat{e}' - \hat{e})\cdot\mathbf{x}]d\hat{e}' \\ &\quad - k(2\pi i)^{-1}\int_{S^2} A(k,\hat{e}',\hat{e})\exp[ik(\hat{e}' - \hat{e})\cdot\mathbf{x}] \\ &\quad \times[\beta(-k,-\hat{e},\mathbf{x}) - 1]d\hat{e}'. \end{aligned} \quad (2.16)$$

We shall refer to the first and second terms on the right side of (2.16) as terms I and II, respectively. The Fourier transform (in $k$) of Eq. (2.16) is the Newton–Marchenko equation

$$\begin{aligned}\eta(t,\hat{e},\mathbf{x}) - \eta(-t,-\hat{e},\mathbf{x}) &= \int_{S^2} M(t,\hat{e},\hat{e}',\mathbf{x})d\hat{e}' \\ &\quad + \int_{-\infty}^\infty\int_{S^2} M(t-s,\hat{e},\hat{e}',\mathbf{x})\eta(-s,-\hat{e}',\mathbf{x})d\hat{e}'\,ds, \end{aligned} \quad (2.17)$$

where

$$M(\alpha,\hat{e},\hat{e}',\mathbf{x}) = (2\pi)^{-2}\int_{-\infty}^\infty \exp[-ik(\alpha + (\hat{e} - \hat{e}')\cdot\mathbf{x})] \times ikA(k,\hat{e}',\hat{e})dk, \quad (2.18)$$

and $\eta$, which was defined by (2.15), is the Fourier transform of $\beta - 1$. We consider $t > 0$ only in (2.17), and use causality (2.14) to eliminate $\eta(-t,-\hat{e},\mathbf{x})$ on the left-hand side.

Equation (2.17) can be used to solve the inverse scattering problem as follows. One assumes that the scattering amplitude $A$ is given. One then solves (2.17) or (2.16) for $\beta$, and recovers the potential $V$ from Eq. (2.9).

## III. ASYMPTOTIC HIGH-FREQUENCY POTENTIAL RECONSTRUCTION

In this section we analyze the large $k$ limit of Newton's reconstruction method. A proof of the basic result is contained in the appendix.

Clearly Eq. (2.9) recovers the potential from the large $k$ limit of the reconstructed field $\beta$. The question then arises: is it really necessary to *solve* (2.16) or (2.17) to obtain this high-energy information? Perhaps the desired high-energy information can be extracted more easily. In particular, since $\beta - 1$ decays like $k^{-1}$, it might be suspected that term II of (2.16) decays faster at infinity than term I, and that therefore the high-energy information is contained entirely in term I. In this case, it might not be necessary to solve the integral equation (2.16).

In the Appendix, we show that term II of (2.16) does indeed decay faster than $k^{-1}$, so that all $k^{-1}$ terms on the right side of (2.16) do indeed arise from term I. Moreover, the $n$th term of the Neumann expansion of (2.16) decays at infinity at least as fast as $k^{-n-\epsilon}$ for arbitrarily small positive $\epsilon$. The $k^{-1}$ coefficient, however, is not $(2i)^{-1}\int_0^\infty V(\mathbf{x} - r\hat{e})dr$, but is rather $(2i)^{-1}\int_{-\infty}^\infty V(\mathbf{x} - r\hat{e})dr$. [This follows from (2.7)]. Thus, for each $\mathbf{x}$, term I determines the integral of $V(\mathbf{x})$ over a line which passes through $\mathbf{x}$ and is parallel to the direction of incidence $\hat{e}$. For a single direction of incidence these line integrals are not sufficient to determine the potential. However, if we vary the direction of incidence over 90°

in a plane, we can build from this set of line integrals a set of "plane integrals." In other words, we can determine the Radon transform of $V(\mathbf{x})$. (For a discussion of Radon's transform, see Ref. 6.) We can then invert the Radon transform to recover $V(\mathbf{x})$. Details of the reconstruction method are given in a near-field context in Ref. 4. This reconstruction method is similar to the inversion method using the Born approximation.[1]

The analysis in the appendix gives insight into the Newton–Marchenko equation, which is the Fourier transform of (2.16). The high energies alone do indeed contain information (2.12) about the jump in the solution $u$ across the wave front. However, in the Marchenko equation, the jumps corresponding to both the "outgoing solution" $\beta(k,\hat{e},x)$ and to the "incoming solution" $\beta(-k,-\hat{e},x)$ appear. They occur in such a way that when only high energy information is used, the two jumps add together: the left side of (2.17) can be expanded using (2.12) as

$$-\frac{1}{4}\int_0^\infty V(\mathbf{x}-r\hat{e})dr\,\mathrm{sgn}(t) + \frac{1}{4}\int_0^\infty V(\mathbf{x}+r\hat{e})dr\,\mathrm{sgn}(-t) + (\text{smoother terms}) = -\frac{1}{4}\int_{-\infty}^\infty V(\mathbf{x}-r\hat{e})dr\,\mathrm{sgn}(t) + (\text{smoother terms}). \tag{3.1}$$

This shows that only line integrals of the potential can be recovered from the first term of the high energy expansion.

When low energies are taken into account, we can use expansion (2.13). In this case, the left side of (2.17) is

$$-\frac{1}{2}\int_0^\infty V(\mathbf{x}-r\hat{e})dr\,H(t) + \frac{1}{2}\int_0^\infty V(\mathbf{x}+r\hat{e})H(-t) + (\text{smoother terms}). \tag{3.2}$$

We see that the limit $t\to 0^+$ of (3.2) allows us to recover the function $-\frac{1}{2}\int_0^\infty V(\mathbf{x}-r\hat{e})dr$, from which $V$ can be reconstructed via

$$V(\mathbf{x}) = -2\hat{e}\cdot\nabla\int_0^\infty V(\mathbf{x}-r\hat{e})dr. \tag{3.3}$$

[Equation (3.3) is equivalent to (2.9).] Thus, the low-frequency content of the reconstructed wave field allows us to determine how much of the total jump discontinuity should be apportioned to the incoming wave and how much to the outgoing wave.

## IV. POSSIBLE EXTENSIONS

We have shown that the inhomogeneous term in Newton's Marchenko equation (2.17) dominates at high frequencies. Therefore, if the high-frequency data are good, this term can be used (following the Born approximation method) to invert exactly for the potential. One is consequently led to wonder whether the inhomogeneous term might dominate in other limits as well. Intuitively, four such limits come to mind. First, if the potential is sufficiently weak in some appropriate sense, both the wave field and the scattering amplitude $A$ will be small, and the higher-order terms of (2.17) will be unimportant. Second, the inhomogeneous term should dominate in the far-field limit; this is because the scattered field (which enters into higher-order terms) decays like $|x|^{-1}$ in the far field. Third, the first term of (2.17) may be useful in determining the high-frequency decay of the higher-order terms. Finally, the first term of (2.17) may be useful in determining the potential near the boundaries of its support; this is because of the weakness of the scattered wave (and hence of the higher-order terms) during the time the probing plane wave has barely penetrated the region of interest.


## ACKNOWLEDGMENTS

We would like to thank Joseph Keller for reading the manuscript and making a number of helpful comments.

This work was supported in part by Office of Naval Research Contract No. N0014-83-K0038. This work was also supported by the National Science Foundation, the Army Research Office, and the Air Force Office of Scientific Research.


## APPENDIX: HIGH-ENERGY ANALYSIS

*Lemma:* Suppose that $V$ is bounded and integrable, and suppose that for some $x_0$, the three functions $|V|, |\nabla V|$, and $|\Delta V(x)|$ are all bounded by $F(|x-x_0|)$, where $F$ is a positive function satisfying $\int_0^\infty F(t)\,dt < \infty$. Let $0\leqslant\alpha\leqslant 2$. Then for $k$ sufficiently large, we have the following estimate for the scattering amplitude:

$$|A(k,\hat{e}',\hat{e})| \leqslant c(k|\hat{e}-\hat{e}'|)^{-\alpha}. \tag{A1}$$

*Proof*: We write

$$A(k,\hat{e}',\hat{e}) = -(4\pi)^{-1}\int \exp[ik(\hat{e}-\hat{e}')\cdot\mathbf{x}] \times V(\mathbf{x})\beta(k,\hat{e},\mathbf{x})d\mathbf{x}, \tag{A2}$$

where $\beta$ is defined by (2.6). We multiply (A2) by $ik(\hat{e}-\hat{e}')$ and then integrate by parts, differentiating $V\beta$ and integrating the exponential. The resulting integral is bounded because $\nabla\beta$ is bounded.[5] Therefore, (A1) holds for $\alpha = 1$. We then integrate by parts once again, and use boundedness of $\Delta\beta$. Equation (A1) therefore holds for $\alpha = 2$. Boundedness of $A$ then allows us to interpolate to obtain (A1) for $0\leqslant\alpha\leqslant 2$. Q.E.D.

**Theorem:** Suppose that $V$ satisfies the same hypotheses as above, and suppose that (2.16) holds. Then term II of (2.16) is $O(k^{-\alpha})$, where $1<\alpha<2$.

*Proof:* Equation (2.7) implies that $|\beta(-k,-\hat{e}',\mathbf{x}) - 1| \leqslant ck^{-1}$ for large $k$; we use this fact together with (A1) in term II of (2.16), obtaining

$$\mathrm{II} \leqslant c\int_{S^2}(k|\hat{e}-\hat{e}'|)^{-\alpha}\,d\hat{e}'. \tag{A3}$$

We write (A3) in polar coordinates with the $z$ axis along $\hat{e}'$ and with polar angle $\theta$ and azimuthal angle $\phi$. The resulting integral is independent of $\phi$; accordingly, we carry out the $\phi$ integration. We then split the polar angle integration into pieces corresponding to integration over $\theta < k^{-1}$ and $\theta > k^{-1}$, respectively. We obtain two terms, which we label $\mathrm{II}_1$ and $\mathrm{II}_2$, respectively.

In $\mathrm{II}_1$ we use $\alpha = 0$, which gives us

$$\mathrm{II}_1 \leqslant c \int_0^{k^{-1}} \sin\theta \, d\theta \leqslant c\, k^{-2}.$$

In $\mathrm{II}_2$ we use $1 < \alpha < 2$:

$$\mathrm{II}_2 \leqslant c\, k^{-\alpha} \int_{k^{-1}}^{\pi} (2 - 2\cos\theta)^{-\alpha/2} \sin\theta \, d\theta$$

$$\leqslant c\, k^{-\alpha}[4^{1-\alpha/2} - (2 - 2\cos k^{-1})^{1-\alpha/2}] = O(k^{-\alpha}).$$

Q.E.D.

*Remark 1:* It follows from (2.7) that term I of (2.16) is

$$(2ik)^{-1} \int_{-\infty}^{\infty} V(\mathbf{x} - r\hat{e}) dr + o(k^{-1}).$$

At high frequencies, therefore, term I dominates term II.

*Remark 2:* The proof of the above theorem depends only on estimate (A1); this shows that application of the integral operator appearing in term II gives rise to an asymptotic factor of $k^{-\alpha}$, where $1 < \alpha < 2$.


[1] R. G. Newton, *Scattering Theory of Waves and Particles* (Springer, New York, 1982), 2nd ed.
[2] R. G. Newton, J. Math. Phys. **21**, 1698 (1980).
[3] R. G. Newton, J. Math. Phys. **23**, 594 (1982).
[4] J. H. Rose, M. Cheney, and B. DeFacio, J. Math. Phys. **25**, 2995 (1984).
[5] M. Cheney, J. Math. Phys. **25**, 2988 (1984).
[6] S. R. Deans, J. Math. Phys. **19**, 2346 (1978); D. Ludwig, Commun. Pure Appl. Math. **19**, 49 (1966).

# Scattering theory for extended elementary particles in nonrelativistic stochastic quantum mechanics

R. Gagnon
*Universite de Montreal, Case Postale 6128 "Succ. A," Montreal, Quebec H3C 3J7, Canada*

(Received 2 April 1984; accepted for publication 27 July 1984)

Scattering theory for extended elementary particles in stochastic phase space is studied. It is shown that the interacting Hamiltonian is equivalent to an effective potential in configuration representation. Asymptotic completeness can be studied by investigating the behavior of the effective potential. The sharp-point limit of the extension of these particles is studied. It is also shown that scattering theory can also be studied directly in stochastic phase space in the optimal case.

## I. INTRODUCTION

The mathematically rigorous approach to nonrelativistic scattering theory was initiated in the fifties, and by now it represents an extensively developed mathematical framework. This framework is, however, primarily applicable to pointlike particles. On the other hand, we know that all hadrons (and possibly also the leptons) are extended. We shall see that the theory of stochastic phase space will enable us to adjust the basic concepts of conventional scattering theory to the case of extended particles.

Let us recall that conventional nonrelativistic scattering theory is based on the wave operators

$$\Omega_{\pm} = \underset{\epsilon \to \mp \infty}{\text{s-lim}} \exp(iHt)\exp(-iH_0 t)\,, \tag{1.1a}$$

$$H = H_0 + V, \quad H_0 = \mathbf{P}^2/2m\,, \quad V = V(\mathbf{X}) \tag{1.1b}$$

$$(X^{j}\psi) = x^{j}\psi(\mathbf{x})\,, \quad (P^{j}\psi)\tilde{}\,(\mathbf{k}) = k^{j}\tilde{\psi}(\mathbf{k})\,, \tag{1.1c}$$

which are acting on the Hilbert space of square integrable functions of the relative position of the interacting particles. It turns out[1] that $\Omega_{\pm}$ are defined everywhere and satisfy

$$(E^{H}(B)\Omega_{\pm}\psi)(\mathbf{r}) = \text{l.i.m.}\int_{\Lambda^{-1}(B)} \phi_{\mathbf{k}}^{(\pm)}(\mathbf{r})\tilde{\psi}(\mathbf{k})d\,\mathbf{k}\,, \tag{1.2}$$

$$\Lambda(\mathbf{k}) = \mathbf{k}^2/2m\,, \tag{1.3}$$

under some technical restrictions on the potential. The distorted plane waves $\phi_{\mathbf{k}}^{(\pm)}$ satisfy the Lippmann–Schwinger equation

$$\phi_{\mathbf{k}}^{(\pm)}(\mathbf{r}) = \phi_{\mathbf{k}}(\mathbf{r}) + \int G_0^{(\pm)}\left(\mathbf{r},\mathbf{r}';\frac{\mathbf{k}^2}{2m}\right)V(\mathbf{r}')\phi_{\mathbf{k}}^{(\pm)}(\mathbf{r}')d\mathbf{r}'\,, \tag{1.4}$$

where the plane waves and the advanced and retarded free Green's functions are, respectively, given by

$$\phi_{\mathbf{k}}(\mathbf{r}) = (2\pi)^{-3/2}\exp(i\mathbf{k}.\mathbf{r})\,, \tag{1.5}$$

$$G_0^{(\pm)}\left(\mathbf{r},\mathbf{r}';\frac{\mathbf{k}^2}{2m}\right) = -\frac{m}{2\pi|\mathbf{r}-\mathbf{r}'|}\exp(\pm i|\mathbf{k}|\,|\mathbf{r}-\mathbf{r}'|)\,. \tag{1.6}$$

The operators $\Omega_{\pm}$ are unitary operators with initial domain $L^2(\mathbb{R}^3)$ and final domain $E^{H}(S_c^{H})L^2(\mathbb{R}^3)$, and therefore the $S$ operator

$$S = \Omega_{-}^{*}\,\Omega_{+} \tag{1.7}$$

is unitary.

In adapting this formalism to the scattering of extended particles, we shall show that the potential scattering problem in stochastic phase spaces for such particles can be reduced to a potential scattering problem in the configuration representation of pointlike particles interacting via a new potential and we shall subsequently state sufficient conditions for having $\mathscr{D}_{\Omega_{\pm}} = L^2(\mathbb{R}^3)$ and $R_{\Omega_{\pm}} = E^{H}(S_c^{H})L^2(\mathbb{R}^3)$. We shall also indicate how the three-body problem can be solved. However, while adapting the configuration space formalism to stochastic phase spaces, we shall see that the stochastic center of mass motion turns out not to be separable from the stochastic relative motion. Consequently, we shall have to prove that a unitary mapping between the stochastic phase-space representation and the momentum representation leads us to equations which are the formal analogs of the Lippmann–Schwinger equation. We will find the solutions of these modified Lippmann–Schwinger equations by the Fredholm method, and prove that thereby we get an analog of (1.2). We shall also show that all the basic quantities for scattering theory in stochastic phase space merge into their conventional counterparts in the sharp-point limit, and shall derive a $T$-matrix formula for the scattering of extended particles.

## II. STOCHASTIC QUANTUM MECHANICS

Stochastic quantum mechanics is a recently developed[2–5] framework mathematically related on the Menger–Wald[6,7] concept of statistical metric spaces, and physically based on the idea of stochastic value for observables, which can be traced to the work of Born.[8,9] In this section we shall review those basic concepts and results required for the formulation of the potential scattering of extended particles.

The Galilei group $G$,[10]

$$G = \{(b,\mathbf{a},\mathbf{v},R):b\in\mathbb{R},\ \mathbf{a},\mathbf{v}\in\mathbb{R}^3,\ R\in\mathrm{SO}(3)\}\,, \tag{2.1}$$

is a transformation group which acts on

$$\Gamma\oplus\mathbb{R} = \{(\mathbf{q},\mathbf{p},t):\mathbf{q},\mathbf{p}\in\mathbb{R}^3,\ t\in\mathbb{R}\} \tag{2.2}$$

as follows:

$$\begin{aligned}&\mathbf{q}\to\mathbf{q}' = R\mathbf{q} + \mathbf{v}t + \mathbf{a}\,,\\ &\mathbf{p}\to\mathbf{p}' = R\mathbf{p} + m\mathbf{v}\,,\\ &t\to t' = t + b\,.\end{aligned} \tag{2.3}$$

 0022-2488/85/030440-13$02.50 © 1985 American Institute of Physics 

We define[4]

$$L^2(\Gamma)=\left\{\psi(\mathbf{q},\mathbf{p}):\int_\Gamma|\psi(\mathbf{q},\mathbf{p})|^2\,d\mathbf{q}\,d\mathbf{p}<\infty\right\},\tag{2.4}$$

with inner product

$$\langle\psi_1|\psi_2\rangle=\int_\Gamma\psi_1^*(\mathbf{q},\mathbf{p})\psi_2(\mathbf{q},\mathbf{p})d\mathbf{q}\,d\mathbf{p}\,.\tag{2.5}$$

Let us consider the following reducible unitary ray representation:

$$\begin{aligned}&(U(b,\mathbf{a},\mathbf{v},R)\psi)(\mathbf{q},\mathbf{p},t)\\&=\exp\left\{\frac{i}{\hbar}\left[-\frac{m\mathbf{v}^2}{2}(t-b)+m\mathbf{v}\cdot(\mathbf{q}-\mathbf{a})\right]\right\}\\&\quad\times\psi(R^{-1}[\mathbf{q}-\mathbf{v}(t-b)-\mathbf{a}]\,,R^{-1}(\mathbf{p}-m\mathbf{v});t-b)\,.\end{aligned}\tag{2.6}$$

The operators $X^j$, $P^j$,

$$X^j=q^j+i\hbar\frac{\partial}{\partial p^j},\quad p^j=-i\hbar\frac{\partial}{\partial q^j},\tag{2.7}$$

which are multiples of the infinitesimal generator of velocity boosts and space translation, respectively,

$$\begin{aligned}&U(0,\mathbf{0},\mathbf{v},I)=\exp((i/\hbar)m\mathbf{v}\cdot\mathbf{X})\,,\\&U(0,\mathbf{a},\mathbf{0},I)=\exp(-i\mathbf{a}\cdot\mathbf{P}/\hbar)\,,\end{aligned}\tag{2.8}$$

realize a reducible representation of the canonical commutation relations

$$\begin{aligned}&[X^j,p^k]=i\hbar\delta_{jk}\,,\\&[X^j,X^k]=[P^j,P^k]=0\,,\quad j,k=1,2,3\,.\end{aligned}\tag{2.9}$$

The irreducible subspaces $\eta_\xi$ of physical significance are those possessing a rotationally invariant resolution generator $\xi$,

$$\xi(\mathbf{q},\mathbf{p})=\xi(R^{-1}\mathbf{q},R^{-1}\mathbf{p})\,,\quad\forall R\in SO(3)\,,\tag{2.10a}$$

$$\|\xi\|=h^{-3/2}\,,\tag{2.10b}$$

i.e., an element $\xi$ such that

$$\psi(\mathbf{q},\mathbf{p})=\langle\xi_{\mathbf{q},\mathbf{p}}|\psi\rangle\,,\tag{2.11a}$$

where

$$\xi_{\mathbf{q},\mathbf{p}}=U(0,\mathbf{q},\mathbf{p}/m,I)\xi\,.\tag{2.11b}$$

The resolution generator $\xi$ and the subspace $\eta_\xi$ define each other unambiguously. It is called resolution generator because we have the resolution of identity on $\eta_\xi$

$$\int|\xi_{\mathbf{q},\mathbf{p}}\rangle\langle\xi_{\mathbf{q},\mathbf{p}}|d\mathbf{q}\,d\mathbf{p}=\mathbb{P}_\xi\,,\tag{2.12}$$

where $\mathbb{P}_\xi$ denotes the orthogonal projector of $L^2(\Gamma)$ onto $\eta_\xi$.

The spaces $\eta_\xi$ are of physical importance because they give rise to unitary mappings between $\eta_\xi$ and the configuration representation $L^2(\mathbb{R}^3)$. These mappings can be written in the form

$$\begin{aligned}&(\omega_\xi\psi)(\mathbf{x})=\text{l.i.m.}\int\xi_{\mathbf{q},\mathbf{p}}(\mathbf{x})\psi(\mathbf{q},\mathbf{p})d\mathbf{q}\,d\mathbf{p}\,,\\&(\omega_\xi^{-1}\psi)(\mathbf{q},\mathbf{p})=\int\xi^*_{\mathbf{q},\mathbf{p}}(\mathbf{x})\psi(\mathbf{x})d\mathbf{x}\,,\end{aligned}\tag{2.13}$$

where $\xi_{\mathbf{q},\mathbf{p}}(\mathbf{x})$ is the corresponding configuration space representative of $\xi_{\mathbf{q},\mathbf{p}}$,

$$\begin{aligned}\xi_{\mathbf{q},\mathbf{p}}(\mathbf{x})&=(U(0,\mathbf{q},\mathbf{p}/m,I)\xi)(\mathbf{x})\\&=\exp((i/\hbar)\mathbf{p}\cdot(\mathbf{x}-\mathbf{q}))\xi(\mathbf{x}-\mathbf{q})\,.\end{aligned}\tag{2.14}$$

The variables $\mathbf{q}$ and $\mathbf{p}$ in the stochastic phase space are interpreted as stochastic variables, i.e., measured with imperfectly accurate instruments of confidence functions[2–5]

$$\begin{aligned}&\chi_{\mathbf{q}}(\mathbf{q}')=h^3|\xi(\mathbf{q}'-\mathbf{q})|^2\,,\\&\hat\chi_{\mathbf{p}}(\mathbf{p}')=h^3|\tilde\xi(\mathbf{p}'-\mathbf{p})|^2\,.\end{aligned}\tag{2.15}$$

It turns out[2] that these instruments can be taken to employ test particles which are extended elementary particles of proper wave functions $\xi$. The stochastic phase space is the set of "extended" stochastic points

$$\hat\Gamma=\{(\hat q,\hat p):\hat q=(\mathbf{q},\chi_{\mathbf{q}})\,,\hat p=(\mathbf{p},\hat\chi_{\mathbf{p}})\,,\mathbf{q},\mathbf{p}\in\mathbb{R}^3\}\,.\tag{2.16}$$

On account of (2.15) we have the following relations:

$$(\Delta x_j)(\Delta k_j)\geq\hbar/2\,,\quad j=1,2,3,\tag{2.17}$$

for the spreads of these stochastic points:

$$\begin{aligned}&(\Delta x_j)=\left(\int_{\mathbb{R}^3}(x_j-q_j)^2\chi_{\mathbf{q}}(\mathbf{x})d\mathbf{x}\right)^{1/2},\\&(\Delta k_j)=\left(\int_{\mathbb{R}^3}(k_j-p_j)^2\hat\chi_{\mathbf{p}}(\mathbf{k})d\mathbf{k}\right)^{1/2}.\end{aligned}\tag{2.18}$$

These relations show that we are not violating the uncertainty principle. In the optimal case, i.e., when relation (2.17) is an equality, the only possible choices for the resolution generators depend on a positive parameter $l$ and are given by

$$\xi^{(l)}(\mathbf{x})=(\pi l^2h^2)^{-3/4}\exp(-\mathbf{x}^2/2l^2)\,.\tag{2.19}$$

In fact, in the case where the resolution generator corresponds to optimal localization in phase space, the wave function $h^{3/2}\xi^{(l)}$ coincides with the proper wave functions introduced by Landé in 1939[11] for the description of extended particles.

## III. THE SCATTERING PROBLEM

Let us assume that we are dealing with a system of two particles whose stochastic position and momentum is measured by means of test particles of proper wave functions $\xi^{(i)}$, $i=1,2$. Then probability amplitudes for results of such measurements provide a representation of the states of the system and are elements of the Hilbert space

$$\begin{aligned}&L^2(\Gamma_{12})=\{\psi\in L^2(\mathbb{R}^{12}):\psi(\xi_1,\xi_2)=\langle\xi^{(1)}_{\xi_1}\otimes\xi^{(2)}_{\xi_2}|\psi\rangle\}\,,\\&\xi^{(i)}_{\xi_i}=U(0,\mathbf{q}_i,\mathbf{p}_i/m_i,I)\xi^{(i)}\,,\\&\xi_i=(\mathbf{q}_i,\mathbf{p}_i)\,,\quad i=1,2,\end{aligned}\tag{3.1}$$

where $m_i$ denotes the mass of particles $i$, $i=1,2$, and where $\mathbf{q}_i,\mathbf{p}_i$ represent, respectively, the mean stochastic position and momentum of particle $i$, $i=1,2$. The Hilbert space $L^2(\Gamma_{12})$ is a closed subspace of $L^2(\Gamma)$, $\Gamma=\mathbf{R}^{12}$.

The scattering problem for such a system has two aspects: In the first instance, the system consists of pointlike particles, in which case the potential energy operator acting on $L^2(\Gamma_{12})$ is given by

$$H_I=V(\mathbf{X}_1-\mathbf{X}_2)\,,\quad\mathbf{X}_j=\mathbf{q}_j+i\hbar\nabla_{\mathbf{p}_j}\,,\quad j=1,2\,,\tag{3.2}$$

and the kinetic energy operator is given by

$$H_0 = \mathbf{P}_1^2/2m_1 + \mathbf{P}_2^2/2m_2\,, \quad \mathbf{P}_j = -i\hbar\nabla_{\mathbf{q}_j}\,, \quad j = 1,2\,. \tag{3.3}$$

In that case there is a unitary transformation [cf. (3.7)] which takes us from present stochastic phase-space representation in $L^2(\mathbb{R}^6)$, where we recover the standard potential scattering theory for two pointlike particles.

In the second instance the system consists of extended particles. However, if these extended particles consist in turn of constituents bound together by potential forces, then we are dealing with a multichannel scattering problem in disguise. On the other hand, if we are dealing with extended elementary particles, then their extension might be of a fundamental nature, i.e., due to their possessing proper wave functions. In that case we are faced with a type of scattering problem that has not yet been investigated in literature. It is with the mathematical theory of this problem that we shall concern ourselves in this paper.

In the case of two extended elementary particles of proper wave functions $\xi^{(i)}$, $i = 1,2$, the kinetic energy operator is still given by (3.3), but the potential energy operator is given by

$$H_I = \mathbb{P}(\Gamma_{12})V(\mathbf{Q}_1 - \mathbf{Q}_2)\mathbb{P}(\Gamma_{12})\,, \tag{3.4}$$

where

$$[V(\mathbf{Q}_1 - \mathbf{Q}_2)\psi](\xi_1,\xi_2) = V(\mathbf{q}_1 - \mathbf{q}_2)\psi(\xi_1,\xi_2)\,, \tag{3.5}$$

and $\mathbf{P}(\Gamma_{12})$ is the projection operator from $L^2(\Gamma)$ onto $L^2(\Gamma_{12})$:

$$(\mathbf{P}(\Gamma_{12})\psi)(\xi_1,\xi_2) = \int d\,\xi_1'\,d\,\xi_2'\,\langle \xi^{(1)}_{\xi_1}|\xi^{(1)}_{\xi_1'}\rangle\langle \xi^{(2)}_{\xi_2}|\xi^{(2)}_{\xi_2'}\rangle\psi(\xi_1',\xi_2')\,. \tag{3.6}$$

Naturally, the above problem can be formulated also in configuration representation. Indeed the unitary mapping from $L^2(\Gamma_{12})$ onto the configuration representation space $L^2(\mathbb{R}^6)$ is given by

$$(\omega_{12}\psi)(\mathbf{x}_1,\mathbf{x}_2) = \text{l.i.m.} \int d\,\xi_1\,d\,\xi_2\,\xi^{(1)}_{\xi_1}(\mathbf{x}_1)\xi^{(2)}_{\xi_2}(\mathbf{x}_2)\psi(\xi_1,\xi_2)\,, \tag{3.7}$$

and it has an inverse of the form

$$(\omega_{12}^{-1}\psi)(\xi_1,\xi_2) = \int d\mathbf{x}_1\,d\mathbf{x}_2\,\xi^{(1)*}_{\xi_1}(\mathbf{x}_1)\xi^{(2)*}_{\xi_2}(\mathbf{x}_2)\psi(\mathbf{x}_1,\mathbf{x}_2)\,. \tag{3.8}$$

Therefore, assuming that the potential energy operator $H_I$ is given by (3.4)–(3.6) where the potential $V$ is locally square integrable and bounded at infinity then, by a direct application of the inversion theorem, we see that the potential energy operator $\omega_{12}H_I\omega_{12}^{-1}$ in the configuration representation is given by the operator $V_{\text{eff}}$

$$V_{\text{eff}} = V_{\text{eff}}(\mathbf{X}_1 - \mathbf{X}_2)\,, \quad (\mathbf{X}_i\psi)(\mathbf{x}_1,\mathbf{x}_2) = \mathbf{x}_i\psi(\mathbf{x}_1,\mathbf{x}_2)\,, \tag{3.9}$$

$$V_{\text{eff}}(\mathbf{x}) = \int d\mathbf{q}_1\,d\mathbf{q}_2\,V(\mathbf{q}_1 - \mathbf{q}_2 + \mathbf{x})\chi^{(1)}(\mathbf{q}_1)\chi^{(2)}(\mathbf{q}_2)$$

Therefore, in the configuration representation the potential energy operator acts like an ordinary potential (which we shall call the effective potential) which depends only on the relative position of the particles which act on extended particles and which approaches the original potential $V(\mathbf{x})$, in the sharp-point limit (corresponding to pointlike particles), i.e., when $\chi^{(i)} \to \delta$, $i = 1,2$, at all points where the potential $V(\mathbf{x})$ is continuous.

The Schrödinger operator

$$H = \mathbf{P}_1^2/2m_1 + \mathbf{P}_2^2/2m_2 + \mathbb{P}(\Gamma_{12})V(\mathbf{Q}_1 - \mathbf{Q}_2)\mathbb{P}(\Gamma_{12}) \tag{3.10}$$

is self-adjoint if the potential $V(\mathbf{q})$ is locally square integrable and bounded at infinity. Indeed, in that case the effective potential $V_{\text{eff}}$ is locally square integrable and bounded at infinity.[12] It has also been shown in this last reference that the Schrödinger operator is then essentially self-adjoint on the domain

$$D_0 = \left\{\psi \in L^2(\Gamma) : \psi(\xi_1,\xi_2) = P(\xi_1,\xi_2) \times \exp\left(-\frac{\mathbf{q}_1^2}{2a} - \frac{\mathbf{q}_2^2}{2a} - \frac{\mathbf{p}_1^2}{2b} - \frac{\mathbf{p}_2^2}{2b}\right)\right\}, \tag{3.11}$$

obtained as $P$ varies over all polynomials in the Cartesian components of $\xi_1$ and $\xi_2$.

In the scattering theory of the system described above, two problems are of fundamental importance: the existence of strong asymptotic states and asymptotic completeness of the quantum mechanical theory. Sufficient conditions are known for the scattering of pointlike particles in order to have those two above problems automatically solved. The conditions are imposed on the potential and are the following[1]: $V$ is locally square integrable, it satisfies

$$V(\mathbf{r}) = O(|\mathbf{r}|^{-(2+\epsilon)})\,, \quad r \to \infty\,, \quad \epsilon > 0\,, \tag{3.12}$$

and is of Rollnik class, i.e.,

$$\|V\|_R^2 \overset{\text{def}}{=} \int_{\mathbb{R}^6} \frac{|V(\mathbf{r})V(\mathbf{r}')|}{|\mathbf{r} - \mathbf{r}'|^2}\,d\mathbf{r}\,d\mathbf{r}' < \infty\,. \tag{3.13}$$

Therefore, the above problems reduce to finding sufficient conditions on the potential $V$ and the confidence functions $\chi^{(1)}$ and $\chi^{(2)}$ under which the effective potential satisfies the above restrictions.

It might be possible *a priori* that a long-range potential gives rise to a short-range effective potential. However, a counterexample has been built[12] for the Coulomb potential $V(\mathbf{r}) = 1/|\mathbf{r}|$ in the optimal case (2.19) where the effective potential is then

$$V_{\text{eff}}(\mathbf{x}) = \frac{(2\pi)^{-1/2}}{|\mathbf{x}|}\int_{-\sqrt{2}|\mathbf{x}|/(l_1^2 + l_2^2)^{1/2}}^{\sqrt{2}|\mathbf{x}|/(l_1^2 + l_2^2)^{1/2}} \exp\left(\frac{-y^2}{2}\right)dy\,, \tag{3.14}$$

where $l_1$ and $l_2$ are the parameters $l$ in (2.19) for particle number 1 and particle number 2, respectively. We note that in the sharp-point limit $l_1, l_2 \to 0$, the potential (3.14) indeed approaches $1/|\mathbf{x}|$, i.e., the original Coulomb potential. On the other hand, the effective potential (3.14) is not of short range since the function

$$f(x) = \int_{-\alpha x}^{\alpha x} \exp\left(-\frac{y^2}{2}\right)dy \tag{3.15}$$

does not decrease at infinity as $|\mathbf{x}|^{-\epsilon}$ for some $\epsilon > 0$.

Sufficient conditions for asymptotic completeness are stated in the following theorem.[12]

**Theorem 3.1:** Assume that the potential $V$ and the confidence functions $\chi^{(i)}$, $i = 1,2$, are locally square integrable, that $V$ is uniformly continuous, that

$$|\chi^{(i)}(\mathbf{r})| = O(|\mathbf{r}|^{-4}), \tag{3.16}$$

$$|V(\mathbf{r})| = O(|\mathbf{r}|^{-4}), \tag{3.17}$$

and that

$$\int_{|\mathbf{a}|<R} d\mathbf{a}\, \chi^{(i)}(\mathbf{q}-\mathbf{a})\chi^{(i)}(\mathbf{q}'-\mathbf{a}) \tag{3.18}$$

are almost everywhere uniformly bounded in $\mathbf{q}$ and $\mathbf{q}'$ for any $R>0$, $i = 1,2$. Then the initial domains $M_0$ of the Moller wave operators $\Omega_\pm$ are $L^2(\Gamma_{12})$, the final domains $R_\pm$ of $\Omega_\pm$ are both identical to $E^H(S_c^H)L^2(\Gamma_{12})$, and the $S$ operator is unitary on $M_0$; furthermore, if $\psi(\mathbf{r},\mathbf{R}) = \psi_1(\mathbf{r})\psi_2(\mathbf{R})$, where

$$\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2, \quad \mathbf{R} = (m_1\mathbf{r}_1 + m_2\mathbf{r}_2)/(m_1+m_2), \tag{3.19}$$

then

$$(\Omega_\pm\psi)(\mathbf{r},\mathbf{R}) = \left(\text{l.i.m.}\int_{\mathbf{R}^3}\phi_{\mathbf{k}}^{(\pm)}(\mathbf{r})\tilde{\psi}_1(\mathbf{k})d\,\mathbf{k}\right)\psi_2(\mathbf{R}), \tag{3.20}$$

where $\psi$ and $\tilde{\psi}$ are the configuration and momentum representatives of $\psi\in L^2(\Gamma_{12})$ specified explicitly by (2.13) for the configuration representation, and where $\phi_{\mathbf{k}}^{(\pm)}$ are, respectively, the outgoing and incoming distorted plane waves satisfying (1.4).

We can readily check that the hypothesis of the above theorem on the confidence functions are satisfied in the optimal case.

We see that we cannot readily conclude that we have asymptotic completeness for any extended particles interacting via short-range potentials. There might be, however, a way out of this impass. Indeed Faddeev[13] when investigating the three-body scattering of pointlike particles interacting by pair showed that under some restrictions on the potentials, the theory is asymptotically complete. This theory has a two-body counterpart, and the conditions on the potential are the following:

$$\begin{aligned}&(1)\ |\tilde{V}(\mathbf{k})|\leqslant c(1+|\mathbf{k}|)^{-1-\theta},\\ &(2)\ |\tilde{V}(\mathbf{k})-\tilde{V}(\mathbf{k}+\mathbf{h})|\leqslant c(1+|\mathbf{k}|)^{-1-\theta}|\mathbf{h}|^{u},\quad |\mathbf{h}|\leqslant 1.\end{aligned} \tag{3.21}$$

Using (3.9) it is straightforward to prove the following theorem.

**Theorem 3.2:** Assume that the Fourier transform $\tilde{V}$ of the local potential $V$ satisfies conditions (3.21), and that $\tilde{\chi}^{(i)}$, $i = 1,2$ are bounded and satisfy

$$|\tilde{\chi}^{(i)}(\mathbf{k})-\tilde{\chi}^{(i)}(\mathbf{k}+\mathbf{h})|\leqslant A^i|\mathbf{h}|^{\mu_i}. \tag{3.22}$$

Then the quantum mechanical theory with the interaction Hamiltonian $H_I$ in (3.4) is asymptotically complete.

Let us note that the conditions imposed on the confidence functions in the above theorem are satisfied in the optimal case.

The three-body problem can be solved in the same manner as in Theorem 3.2 if the particles are interacting by pairs if the conditions of Theorem 3.2 are satisfied with $\mu_1 = \mu_2 = \mu_3$. That last restriction is in particular satisfied for the optimal case.

The conditions imposed in Theorem 3.2 are also rather restrictive since the potential must be bounded. Therefore, we will be looking for a more general method (in the optimal case) to solve the scattering problem for extended particles interacting via a potential. We shall see that the framework introduced in the next section is adequate for such a purpose, and that it has the advantage of expressing the states $\Omega_+\psi$ in terms of the stochastic variables $\mathbf{q}_1$, $\mathbf{p}_1$, $\mathbf{q}_2$, $\mathbf{p}_2$.

## IV. EIGENFUNCTION EXPANSIONS IN STOCHASTIC PHASE-SPACE VARIABLES

In this section, we start with the conjecture that in the optimal case the wave operators $\Omega_\pm$, specified by (1.1a), (3.3), (3.4), (3.5), and (3.6) satisfy an asymptotic completeness condition analogous to the one in conventional scattering theory [cf. (1.2)] for pointlike particles, but expressed in terms of stochastic variables. We shall then show that under those circumstances an equation analogous to the Lippmann–Schwinger equation (1.4) is satisfied. We shall subsequently study the sharp-point limit of that theory and derive a $T$-matrix formula. The Lippman–Schwinger equation will be solved in the next section, so that we will be able to prove that the asymptotic completeness is indeed satisfied for the solutions of the aforementioned Lippmann–Schwinger equation.

By (2.13) we have the following unitary mapping between the momentum representation and the stochastic phase-space representation for a system of two particles:

$$\tilde{\psi}(\mathbf{k}_1,\mathbf{k}_2) = \text{l.i.m.}\int d\,\xi_1\, d\,\xi_2\, \phi^*_{\mathbf{k}_1,\mathbf{k}_2}(\xi_1,\xi_2)\psi(\xi_1,\xi_2), \tag{4.1}$$

$$\psi(\xi_1,\xi_2) = \int d\,\mathbf{k}_1\, d\,\mathbf{k}_2\, \phi_{\mathbf{k}_1,\mathbf{k}_2}(\xi_1,\xi_2)\tilde{\psi}(\mathbf{k}_1,\mathbf{k}_2), \tag{4.2}$$

$$\phi_{\mathbf{k}_1,\mathbf{k}_2}(\xi_1,\xi_2) = \tilde{\xi}^{(1)*}_{\xi_1}(\mathbf{k}_1)\tilde{\xi}^{(2)*}_{\xi_2}(\mathbf{k}_2). \tag{4.3}$$

Therefore, we can consider that we have an expansion analogous to the eigenfunction expansion in conventional scattering theory, where the role of the plane waves (1.5) is taken over by the phase-space waves (4.3).

In conventional scattering theory it is possible to separate the center of mass motion from the relative motion, and therefore work in $L^2(\mathbf{R}^3)$ instead of $L^2(\mathbf{R}^6)$. However, this procedure is not possible in stochastic phase space. Indeed, upon performing the change of variables

$$\mathbf{q} = \mathbf{q}_1 - \mathbf{q}_2, \quad \mathbf{Q} = (m_1\mathbf{q}_1 + m_2\mathbf{q}_2)/(m_1+m_2), \tag{4.4a}$$

$$\mathbf{p} = (m_2\mathbf{p}_1 - m_1\mathbf{p}_2)/(m_1+m_2), \quad \mathbf{P} = \mathbf{p}_1 + \mathbf{p}_2, \tag{4.4b}$$

we see that the resolution waves are given by

$$\begin{aligned}\phi_{\mathbf{k}_1,\mathbf{k}_2}(\xi_1,\xi_2) = h^{-3}&\left(\frac{\pi\hbar^2}{l_1l_2}\right)^{-3/2}\exp\left(\frac{i}{\hbar}\mathbf{k}\cdot\mathbf{q}+\frac{i}{\hbar}\mathbf{K}\cdot\mathbf{Q}\right)\\ &\times\exp\left(-\left(\frac{l_1^2}{2\hbar^2}+\frac{l_2^2}{2\hbar^2}\right)(\mathbf{k}-\mathbf{p})^2\right.\\ &-\frac{m_1^2l_1^2+m_2^2l_2^2}{2(m_1+m_2)^2\hbar^2}(\mathbf{K}-\mathbf{P})^2\\ &\left.+\frac{-m_1l_1^2+m_2l_2^2}{(m_1+m_2)\hbar^2}(\mathbf{k}-\mathbf{p})\cdot(\mathbf{K}-\mathbf{P})\right),\end{aligned} \tag{4.5}$$

$$\mathbf{k} = \frac{m_2\mathbf{k}_1 - m_1\mathbf{k}_2}{m_1+m_2}, \quad \mathbf{K} = \mathbf{k}_1 + \mathbf{k}_2. \tag{4.6}$$

In order to derive an equation equivalent to the Lippmann–Schwinger equation we have to introduce the distorted phase-space waves $\phi^{(+)}_{\mathbf{k}_1,\mathbf{k}_2}$ satisfying (as will be proven later)

$$(E^H(B)\Omega_- \psi)(\zeta_1,\zeta_2) = \int_{\Lambda^{-1}(B)} \phi^{(+)}_{\mathbf{k}_1,\mathbf{k}_2}(\zeta_1,\zeta_2)\tilde{\psi}(\mathbf{k}_1,\mathbf{k}_2)d\,\mathbf{k}_1\,d\,\mathbf{k}_2\,, \tag{4.7}$$

where $B$ is any Borel set in $\mathbf{R}$, and where

$$\Lambda(\mathbf{k}_1,\mathbf{k}_2) = \mathbf{k}_1^2/2m_1 + \mathbf{k}_2^2/2m_2\,. \tag{4.8}$$

It will turn out that distorted phase-space waves of the form

$$\phi^{(\pm)}_{\mathbf{k}_1,\mathbf{k}_2}(\zeta_1,\zeta_2) = \exp(i\mathbf{K}\cdot\mathbf{Q})\hat{\phi}^{(\pm)}_{\mathbf{K},\mathbf{k}}(\mathbf{q},\mathbf{p}_1,\mathbf{p}_2) \tag{4.9}$$

will meet our needs, the dependence in $\mathbf{Q}$ being explicitly given by the left factor in the right-hand side of (4.9) when we use the variables (4.4a) and (4.6).

In order to prove that the functions $\phi^{(\pm)}_{\mathbf{K},\mathbf{k}}$ satisfy an equation of the Lippmann–Schwinger type, we first have to introduce the counterpart of the free Green's function in terms of $\zeta\in\mathbb{C}$,

$$G_{0,\mathbf{K}}(\mathbf{q},\mathbf{p}_1,\mathbf{p}_2;\mathbf{q}',\mathbf{p}_1',\mathbf{p}_2';\zeta) \overset{\text{def}}{=} -(2\pi)^3\int d\,\mathbf{k}\left(\frac{\mathbf{k}^2}{2m}-\zeta\right)^{-1} \times\hat{\phi}_{\mathbf{K},\mathbf{k}}(\mathbf{q},\mathbf{p}_1,\mathbf{p}_2)\hat{\phi}^{*}_{\mathbf{K},\mathbf{k}}(\mathbf{q}',\mathbf{p}_1',\mathbf{p}_2')\,, \tag{4.10}$$

where $m$ is the reduced mass

$$m = m_1m_2/(m_1+m_2)\,. \tag{4.11}$$

The above function is defined whenever $0<\arg\zeta<2\pi$.

The function (4.10) is computed explicitly in the Appendix and the result is

$$\begin{aligned} &G_{0,\mathbf{K}}(\mathbf{q},\mathbf{p}_1,\mathbf{p}_2;\mathbf{q}',\mathbf{p}_1',\mathbf{p}_2';\zeta) \\ &= \frac{l_1^3l_2^3m}{4\pi^4(\beta_1^2+\beta_2^2+\beta_3^2)^{1/2}}\exp\Bigg[-\frac{(l_1^2+l_2^2)}{2}(\mathbf{p}^2+\mathbf{p}'^2) \\ &\quad+\frac{(l_1^2m_1-l_2^2m_2)}{m_1+m_2}(\mathbf{p}\cdot(\mathbf{K}-\mathbf{P})+\mathbf{p}'\cdot(\mathbf{K}-\mathbf{P}'))-\frac{(l_1^2m_1^2+l_2^2m_2^2)}{2(m_1+m_2)^2}((\mathbf{K}-\mathbf{P})^2+(\mathbf{K}-\mathbf{P}')^2)-2m\zeta\alpha\Bigg] \\ &\quad\times\Bigg\{\exp\left(i\sqrt{\beta_1^2+\beta_2^2+\beta_3^2}\,\sqrt{2m\zeta}\right)\left[\operatorname{erf}\left(-\frac{\sqrt{\beta_1^2+\beta_2^2+\beta_3^2}}{2\sqrt{\alpha}}-i\sqrt{2m\zeta\alpha}\right)-1\right] \\ &\quad+\exp\left(-i\sqrt{\beta_1^2+\beta_2^2+\beta_3^2}\,\sqrt{2m\zeta}\right)\left[\operatorname{erf}\left(-\frac{\sqrt{\beta_1^2+\beta_2^2+\beta_3^2}}{2\sqrt{\alpha}}+i\sqrt{2m\zeta\alpha}\right)+1\right]\Bigg\}, \end{aligned} \tag{4.12}$$

where

$$\alpha = l_1^2+l_2^2\,, \tag{4.13a}$$

$$\boldsymbol{\beta} = \mathbf{q}-\mathbf{q}'-i\left[(l_1^2-l_2^2)(\mathbf{p}+\mathbf{p}') - \frac{(l_1^2m_1+l_2^2m_2)}{m_1+m_2}(2\mathbf{K}-\mathbf{P}-\mathbf{P}')\right]. \tag{4.13b}$$

We are now ready to derive the Lippmann–Schwinger equation in the context of the following theorem.

**Theorem 4.1:** Let $V(\mathbf{q})$ be locally square integrable and such that $V(\mathbf{q}) = O(|\mathbf{q}|^{-3})$. Suppose that $\hat{\phi}^{(\pm)}_{\mathbf{K},\mathbf{k}}$ exists for $(\mathbf{k},\mathbf{K})\in\mathbb{R}^6\setminus S^6_V$ ($S^6_V$ being a set of Lebesgue measure zero) as a function of $\mathbf{q}$, $\mathbf{p}_1$, $\mathbf{p}_2$, and that

$$\int d\mathbf{p}_1\,d\mathbf{p}_2|\hat{\phi}^{(\pm)}_{\mathbf{k}_1,\mathbf{k}_2}(\mathbf{q},\mathbf{p}_1,\mathbf{p}_2)|^2 \tag{4.14}$$

is uniformly bounded in $\mathbf{q}$ when $\mathbf{k}_1$ and $\mathbf{k}_2$ are restricted to a compact set. Then the functions $\hat{\phi}^{(\pm)}_{\mathbf{K},\mathbf{k}}$, $(\mathbf{k},\mathbf{K})\notin S^6_V$, satisfy the Lippmann–Schwinger equation

$$\hat{\phi}^{(\pm)}_{\mathbf{K},\mathbf{k}}(\mathbf{q},\mathbf{p}_1,\mathbf{p}_2) = \hat{\phi}_{\mathbf{K},\mathbf{k}}(\mathbf{q},\mathbf{p}_1,\mathbf{p}_2) + \int d\mathbf{q}'\,d\mathbf{p}_1'\,d\mathbf{p}_2' \times G^{(\pm)}_{0,\mathbf{K}}\left(\mathbf{q},\mathbf{p}_1,\mathbf{p}_2;\mathbf{q}',\mathbf{p}_1',\mathbf{p}_2';\frac{\mathbf{k}^2}{2m}\right)V(\mathbf{q}')\hat{\phi}^{(\pm)}_{\mathbf{K},\mathbf{k}}(\mathbf{q}',\mathbf{p}_1',\mathbf{p}_2')\,, \tag{4.15}$$

where the advanced and retarded free Green's functions are given by

$$G^{(\pm)}_{0,\mathbf{K}}(\mathbf{q},\mathbf{p}_1,\mathbf{p}_2;\mathbf{q}',\mathbf{p}_1',\mathbf{p}_2';\lambda) = \lim_{\epsilon\to\pm 0} G_{0,\mathbf{K}}(\mathbf{q},\mathbf{p}_1,\mathbf{p}_2;\mathbf{q}',\mathbf{p}_1',\mathbf{p}_2';\lambda+i\epsilon)\,. \tag{4.16}$$

*Proof:* (Ref. 12) The Lippmann–Schwinger equations in Hilbert space [Eq. (3.34) of Chap. V in Ref. 1] are

$$\psi^{\pm} = \psi + \operatorname*{s-lim}_{\epsilon\to\pm\infty}\int_{-\infty}^{\infty}\frac{1}{\lambda-H_0\pm i\epsilon}H_I\,d_\lambda E^H_\lambda\psi^{\pm}\,. \tag{4.17}$$

Hence, by Theorem 5.5,

$$\langle g|f_+\rangle - \langle g|f\rangle = \lim_{\epsilon\to+0}\left\langle g\middle|\int_{-\infty}^{\infty}G_0(\lambda+i\epsilon)H_I\,dE^H_\lambda f_+\right\rangle, \tag{4.18}$$

where by (4.2) and (4.7) (see Theorem 5.8)

$$\langle g|f_+-f\rangle = \int_\Gamma d\zeta_1\,d\zeta_2\,g^*(\zeta_1,\zeta_2)\int_{\mathbb{R}^6}d\,\mathbf{k}_1\,d\,\mathbf{k}_2\tilde{f}(\mathbf{k},\mathbf{K}) \times\left[\phi^{(+)}_{\mathbf{k}_1,\mathbf{k}_2}(\zeta_1,\zeta_2)-\phi_{\mathbf{k}_1,\mathbf{k}_2}(\zeta_1,\zeta_2)\right]. \tag{4.19}$$

Let us assume that

$$g(\zeta_1,\zeta_2)\in C_b^0(\mathbb{R}^6)\cap L^2(\Gamma_{12})\,, \tag{4.20}$$

$$\tilde{f}(\mathbf{k},\mathbf{K})\in C_V^0\subset C_b^0(\mathbb{R}^6)\,, \tag{4.21}$$

where $C_V^0$ is the set of all functions of $C_b^0$ with support disjoint from the exceptional set $S^6_V$. We therefore have for any $a,b\in\mathbb{R}$, $a<b$, by the Fourier–Plancherel inversion theorem

$$\left\langle g \left| \int_{-\infty}^{\infty} G_0(\lambda + i\epsilon) H_I \, d_\lambda E_\lambda^H f_+ \right\rangle = (2\pi)^3 \lim_{\delta \to 0} \sum_{j=1}^{n} \lim_{R \to \infty} \int d\zeta_1 \, d\zeta_2 \, g^*(\zeta_1, \zeta_2) \int_{|k'|<R} d\mathbf{k}' \int d\mathbf{q}' \, d\mathbf{p}_1' \, d\mathbf{p}_2' \, V(\mathbf{q}') \right.$$

$$\times \int_{\lambda_{j-1} < \mathbf{k}^2/2m + \mathbf{K}^2/2M < \lambda_j} d\mathbf{k} \, d\mathbf{K} \exp(-i\mathbf{K} \cdot \mathbf{Q}) \hat{\phi}_{\mathbf{K},\mathbf{k}'}(\mathbf{q}, \mathbf{p}_1, \mathbf{p}_2)$$

$$\times \left( \lambda_j' + i\epsilon - \frac{\mathbf{k}'^2}{2m} - \frac{\mathbf{K}^2}{2M} \right)^{-1} \hat{\phi}^*_{\mathbf{K},\mathbf{k}'}(\mathbf{q}', \mathbf{p}_1', \mathbf{p}_2') \hat{\phi}^{(\pm)}_{\mathbf{K},\mathbf{k}}(\mathbf{q}', \mathbf{p}_1', \mathbf{p}_2') \tilde{f}(\mathbf{k}, \mathbf{K}) \,. \tag{4.22}$$

Since the above integral in $\mathbf{k}'$, $\mathbf{q}'$, $\mathbf{p}_1'$, $\mathbf{p}_2'$, $\mathbf{k}$, $\mathbf{K}$ has an integrand bounded by the integrable function

$$\left| V(\mathbf{q}') \hat{\phi}_{\mathbf{K},\mathbf{k}'}(\mathbf{q}, \mathbf{p}_1, \mathbf{p}_2) \left( \lambda_j' + i\epsilon - \frac{\mathbf{k}'^2}{2m} - \frac{\mathbf{K}^2}{2M} \right)^{-1} \right.$$

$$\left. \times \hat{\phi}^*_{\mathbf{K},\mathbf{k}}(\mathbf{q}', \mathbf{p}_1', \mathbf{p}_2') \hat{\phi}^{(\pm)}_{\mathbf{K},\mathbf{k}}(\mathbf{q}', \mathbf{p}_1', \mathbf{p}_2') \tilde{f}(\mathbf{k}, \mathbf{K}) \right| , \tag{4.23}$$

[cf. (4.14)], we can reverse the order of integration in such a way that we integrate with respect to $\mathbf{k}'$ first, then with respect to $\mathbf{q}'$, $\mathbf{p}_1'$, and $\mathbf{p}_2'$, then with respect to $\mathbf{k}$ and $\mathbf{K}$, and finally with respect to $\zeta_1, \zeta_2$. Now the functions

$$\int_{|\mathbf{k}'|<R} d\mathbf{k}' \, \hat{\phi}_{\mathbf{K},\mathbf{k}'}(\mathbf{q}, \mathbf{p}_1, \mathbf{p}_2) \left( \lambda_j' + i\epsilon - \frac{\mathbf{k}'^2}{2m} - \frac{\mathbf{K}^2}{2M} \right)^{-1} \tag{4.24}$$

$$\times \hat{\phi}^*_{\mathbf{K},\mathbf{k}'}(\mathbf{q}', \mathbf{p}_1', \mathbf{p}_2') \,, \quad R > 0 \,,$$

are bounded by a function $A$ which does not depend on $R$, and is such that the integrand

$$|g^*(\zeta_1, \zeta_2) \exp(-i\mathbf{K} \cdot \mathbf{Q}) \tilde{f}(\mathbf{k}, \mathbf{K}) V(\mathbf{q}') \hat{\phi}^{(\pm)}_{\mathbf{K},\mathbf{k}}(\mathbf{q}', \mathbf{p}_1', \mathbf{p}_2') A| \tag{4.25}$$

is integrable in $\zeta_1$, $\zeta_2$, $\mathbf{k}$, $\mathbf{K}$, $\mathbf{q}'$, $\mathbf{p}_1'$, $\mathbf{p}_2'$. Indeed, applying the following relation:

$$\left| \int_{|\mathbf{k}|<R} d\mathbf{k} \left( \frac{\mathbf{k}^2}{2m} - \zeta \right)^{-1} \exp(-\alpha \mathbf{k}^2 + \boldsymbol{\beta} \cdot \mathbf{k}) \right|$$

$$\leqslant \frac{1}{|\operatorname{Im} \zeta|} \left( \frac{\pi}{\alpha} \right)^{3/2} \exp\left( \frac{(\operatorname{Re} \beta)^2}{4\alpha} \right) = A \,, \tag{4.26}$$

to (4.24), we find that $A$ satisfies

$$\int |A|^2 \, d\mathbf{p}' \, d\mathbf{P}' < \infty \tag{4.27}$$

which is true since $4ad \geqslant c^2$ for

$$a = (l_1^2 + l_2^2) + (l_1^2 - l_2^2)^2 / 2(l_1^2 + l_2^2) \,, \tag{4.28}$$

$$c = -2 \frac{(l_1^2 m_1 - l_2^2 m_2)}{m_1 + m_2} + \frac{(l_1^2 - l_2^2)(l_1^2 m_1 + l_2^2 m_2)}{(l_1^2 + l_2^2)(m_1 + m_2)} \,, \tag{4.29}$$

$$d = \frac{l_1^2 m_1^2 + l_2^2 m_2^2}{(m_1 + m_2)^2} + \frac{(l_1^2 m_1 + l_2^2 m_2)^2}{2(l_1^2 + l_2^2)(m_1 + m_2)^2} \,. \tag{4.30}$$

Therefore, (4.25) is satisfied in view of the Cauchy–Schwartz inequality, (4.20)–(4.21) and (4.14). Therefore we can carry the limit as $R \to \infty$ in our expression for the left-hand side of (4.22). So that, carrying the sum under the integral sign, we obtain an integral in $\zeta_1$, $\zeta_2$, $\mathbf{k}$, $\mathbf{K}$, $\mathbf{q}'$, $\mathbf{p}_1'$, $\mathbf{p}_2'$, for which the integrand is majorized in virtue of inequality (4.26) by an integrable function of $\zeta_1$, $\zeta_2$, $\mathbf{k}$, $\mathbf{K}$, $\mathbf{q}'$, $\mathbf{p}_1'$, $\mathbf{p}_2'$. Therefore, we get by Lebesgue dominated convergence theorem the following expression for the left-hand side of (4.22):

$$\int_\Gamma d\zeta_1 \, d\zeta_2 \, g^*(\zeta_1, \zeta_2) \int d\mathbf{k} \, d\mathbf{K} \exp(-i\mathbf{K} \cdot \mathbf{Q}) \cdot \tilde{f}(\mathbf{k}, \mathbf{K})$$

$$\times \int d\mathbf{q}' \, d\mathbf{p}_1' \, d\mathbf{p}_2' \, V(\mathbf{q}') \hat{\phi}^{(+)}_{\mathbf{K},\mathbf{k}}(\mathbf{q}', \mathbf{p}_1', \mathbf{p}_2')$$

$$\times G_{0,\mathbf{K}}(\mathbf{q}, \mathbf{p}_1, \mathbf{p}_2; \mathbf{q}', \mathbf{p}_1', \mathbf{p}_2'; \mathbf{k}^2/2m + i\epsilon) \,. \tag{4.31}$$

In order to be able to apply the Lebesgue dominated convergence theorem to take the limit as $\epsilon \to 0$, let us note that by (4.12) and (4.20)–(4.21) the integrand in (4.31) is bounded by an integrable function independent of $\epsilon$. Indeed, (4.12) is bounded by

$$\text{const} \frac{1}{|\mathbf{q} - \mathbf{q}'|}$$

$$\times \exp(-a\mathbf{p}'^2 + \mathbf{e} \cdot \mathbf{p}' + c\mathbf{p}' \cdot \mathbf{P}' + \mathbf{f} \cdot \mathbf{P}' - d\mathbf{P}'^2) \,, \tag{4.32}$$

when $\boldsymbol{\beta}$ is restricted to a compact set, where

$$a = \frac{l_1^2 + l_2^2}{2} \,, \quad c = -\frac{l_1^2 m_1 - l_2^2 m_2}{m_1 + m_2} \,,$$

$$d = \frac{(l_1^2 m_1^2 + l_2^2 m_2^2)}{2(m_1 + m_2)^2} \,. \tag{4.33}$$

A straightforward computation shows that $4ad - c^2 = l_1^2 l_2^2 > 0$, which is sufficient to guarantee the integrability of (9.74). Therefore, the integrand of (4.31) is bounded by an integrable function independent of $\epsilon$, when $\boldsymbol{\beta}$ is restricted to a compact set. Now for $\boldsymbol{\beta}$ outside a sufficiently large compact set, the function

$$\overline{G}_\zeta = \left| \exp(-2m\zeta\alpha + i\sqrt{\beta_1^2 + \beta_2^2 + \beta_3^2} \sqrt{2m\zeta}) \right.$$

$$\times \left[ \operatorname{erf}(-(\sqrt{\beta_1^2 + \beta_2^2 + \beta_3^2})/2\sqrt{\alpha} \right.$$

$$\left. \left. - i\sqrt{2m\zeta\alpha}) - 1 \right] \right| \tag{4.34}$$

is smaller than

$$\left| \exp(-(\beta_1^2 + \beta_2^2 + \beta_3^2)/4\alpha) \right| \,. \tag{4.35}$$

Therefore, comparing (4.12) with (4.35) we see that, by (4.13b), (4.28)–(4.30), (4.20)–(4.21), (4.14), the restrictions on the potential and Cauchy–Schwarz inequality, the integral of

$$|g^*(\zeta_1, \zeta_2)| \, |\tilde{f}(\mathbf{k}, \mathbf{K})| \, |V(\mathbf{q}')| \, |\hat{\phi}^{(+)}_{\mathbf{K},\mathbf{k}}(\mathbf{q}', \mathbf{p}_1', \mathbf{p}_2')|$$

$$\times \overline{G}_\zeta(\mathbf{q}, \mathbf{p}_1, \mathbf{p}_2; \mathbf{q}', \mathbf{p}_1', \mathbf{p}_2') \tag{4.36}$$

is convergent. The same arguments apply for

$$\left| \exp(-2m\zeta\alpha - i\sqrt{\beta_1^2 + \beta_2^2 + \beta_3^2} \sqrt{2m\zeta}) \right.$$

$$\left. \times \left[ \operatorname{erf}(-(\sqrt{\beta_1^2 + \beta_2^2 + \beta_3^2})/2\sqrt{\alpha} + i\sqrt{2m\zeta\alpha}) - 1 \right] \right| \,. \tag{4.37}$$

By Cauchy–Schwarz inequality the integral of

$$|g^*(\xi_1,\xi_2)|\,|\tilde{f}(\mathbf{k},\mathbf{K})|\,|V(\mathbf{q}')|\,|\hat{\phi}^{(+)}_{\mathbf{K},\mathbf{k}}(\mathbf{q}',\mathbf{p}_1',\mathbf{p}_2')|$$
$$\times|\exp(-i\sqrt{\beta_1^2+\beta_2^2+\beta_3^2}\sqrt{2m\zeta})| \tag{4.38}$$

is also convergent. Therefore, the limit as $\epsilon\to+0$ in (4.31) can be taken under the integral sign, so that combining (4.18) with (4.19) and the expression (4.31) for the left-hand side of (4.22), we obtain (4.15) since $C_b^0(\mathbb{R}^6)$ and $C_V^0$ are dense sets. Q.E.D.

Let us now investigate the sharp-point limit of the distorted phase-space waves. We see from (4.5) that

$$\lim_{l_1,l_2\to 0}(l_1l_2)^{-3/2}\phi_{\mathbf{k}_1,\mathbf{k}_2}(\xi_1,\xi_2)=(4\pi^3)^{-3/2}\exp(i\mathbf{k}\cdot\mathbf{q}+i\mathbf{K}\cdot\mathbf{Q})\,. \tag{4.39}$$

Furthermore, from (4.12) we see that

$$\lim_{l_1,l_2\to 0}(l_1l_2)^{-3}G_{0,\mathbf{K}}(\mathbf{q},\mathbf{p}_1,\mathbf{p}_2;\mathbf{q}',\mathbf{p}_1',\mathbf{p}_2';\zeta)$$
$$=-(m/2\pi^4|\mathbf{q}-\mathbf{q}'|)\exp(i|\mathbf{q}-\mathbf{q}'|\sqrt{2m\zeta})\,. \tag{4.40}$$

Thus, we can state that upon renormalization with the factor $(l_1l_2/\pi)^{-3/2}$, the phase space wave $\phi_{\mathbf{k}_1,\mathbf{k}_2}$ merges in the sharp-point limit into a configuration representation plane waves from which the center of mass has not been removed [see (1.5)]. Furthermore, the free Green's function $G_{0,\mathbf{K}}$ goes, upon renormalization by the factor $(l_1l_2/\pi)^{-3}$, into the conventional configuration representation free Green's function

$$G_0(\mathbf{r},\mathbf{r}';\rho)=-\frac{m}{2\pi|\mathbf{r}-\mathbf{r}'|}\exp(i|\mathbf{r}-\mathbf{r}'|\sqrt{2m\zeta})\,. \tag{4.41}$$

The same is true for the advanced and retarded free Green's functions, with the same choice of renormalization factor. Let us show that $e^{i\mathbf{K}\cdot\mathbf{Q}}\hat{\phi}^{(+)}_{\mathbf{K},\mathbf{k}}(\xi_1,\xi_2)$ [where $\hat{\phi}^{(+)}_{\mathbf{K},\mathbf{k}}$ are the solutions of (4.15)] go in the limit $l_1,l_2\to 0$ to $\phi_{\mathbf{K}}(\mathbf{Q})\phi^{(+)}_{\mathbf{k}}(\mathbf{q})$ [where $\phi^{(+)}_{\mathbf{k}}$ are the solutions of (1.4)] upon renormalization by the factor $(l_1l_2/\pi)^{-3/2}$. To prove this, let us first find the solutions of the series

$$(a_n,\alpha_n,\beta_n,\gamma_n,\delta_n,\epsilon_n)\,, \tag{4.42}$$

satisfying the following conditions:

$$\frac{l_1^3l_2^3m}{4\pi^4}\int\exp\Big[-\frac{1}{2}(l_1^2+l_2^2)(\mathbf{p}^2+\mathbf{p}'^2)$$
$$+\frac{(l_1^2m_1-l_2^2m_2)}{m_1+m_2}(\mathbf{p}\cdot(\mathbf{K}-\mathbf{P})+\mathbf{p}'\cdot(\mathbf{K}-\mathbf{P}'))$$
$$-\frac{(l_1^2m_1^2+l_2^2m_2^2)}{2(m_1+m_2)^2}((\mathbf{K}-\mathbf{P})^2+(\mathbf{K}-\mathbf{P}')^2)\Big]$$
$$\times f_n(\mathbf{p}',\mathbf{P}')d\mathbf{p}'\,d\mathbf{P}'=f_{n+1}(\mathbf{p},\mathbf{P})\,, \tag{4.43}$$

so that the terms of the series occur in

$$f_n(\mathbf{p},\mathbf{P})=a_n\exp(-\alpha_n\mathbf{p}^2-\beta_n\mathbf{P}^2$$
$$+\gamma_n\cdot\mathbf{p}+\delta_n\cdot\mathbf{P}+\epsilon_n\mathbf{p}\cdot\mathbf{P})\,, \tag{4.44}$$

and for $n=1$ assume the following values:

$$a_1=(2\pi)^{-3}\Big(\frac{\pi}{l_1l_2}\Big)^{-3/2}\exp\Big[-\frac{(l_1^2+l_2^2)}{2}\mathbf{k}^2$$
$$-\frac{(l_1^2m_1^2+l_2^2m_2^2)}{2(m_1+m_2)^2}\mathbf{K}^2+\frac{(-m_1l_1^2+m_2l_2^2)}{m_1+m_2}\mathbf{k}\cdot\mathbf{K}\Big]\,, \tag{4.45}$$

$$\alpha_1=\frac{l_1^2+l_2^2}{2}\,,\quad \beta_1=\frac{(l_1^2m_1^2+l_2^2m_2^2)}{2(m_1+m_2)^2}\,, \tag{4.46}$$

$$\gamma_1=(l_1^2+l_2^2)\mathbf{k}+\frac{(m_1l_1^2-m_2l_2^2)}{m_1+m_2}\mathbf{K}\,, \tag{4.47}$$

$$\delta_1=\frac{(m_1l_1^2-m_2l_2^2)}{m_1+m_2}\mathbf{k}+\frac{(l_1^2m_1^2+l_2^2m_2^2)^2}{(m_1+m_2)^2}\mathbf{K}, \tag{4.48}$$

$$\epsilon_1=\frac{-m_1l_1^2+m_2l_2^2}{m_1+m_2}\,. \tag{4.49}$$

We therefore obtain

$$a_{n+1}=(2\pi)^{-3}\Big(\frac{\pi}{l_1l_2}\Big)^{-3/2}\Big(\frac{l_1^3l_2^3m}{4\pi^4}\Big)^n\Big(\frac{\pi^2}{l_1^2+l_2^2}\Big)^{3n/2}\Big[\frac{(l_1^2m_1^2+l_2^2m_2^2)}{(m_1+m_2)^2}-\frac{(-l_1^2m_1+l_2^2m_2)^2}{(l_1^2+l_2^2)(m_1+m_2)^2}\Big]^{-3n/2}$$
$$\times\exp\Big\{-\frac{1}{4}\Big(l_1^2+l_2^2-\frac{(l_1^2m_1-l_2^2m_2)^2(l_1^2+l_2^2)}{l_1^2l_2^2(m_1+m_2)^2}\Big)\mathbf{k}^2$$
$$-\Big[\frac{(n+\frac{1}{2})(l_1^2m_1^2+l_2^2m_2^2)}{(m_1+m_2)^2}-n\Big[\frac{l_1^2m_1^2+l_2^2m_2^2}{(m_1+m_2)}-\frac{2(l_1^2m_1-l_2^2m_2)^2}{(l_1^2+l_2^2)(m_1+m_2)}\Big]^2\big[4l_1^2l_2^2(m_1+m_2)^2/(l_1^2+l_2^2)\big]^{-1}\Big]\mathbf{K}^2$$
$$-\Big[\frac{(-m_1l_1^2+m_2l_2^2)}{m_1+m_2}-(l_1^2m_1-l_2^2m_2)\Big[\frac{l_1^2m_1^2+l_2^2m_2^2}{m_1+m_2}-\frac{2(l_1^2m_1-l_2^2m_2)^2}{(l_1^2+l_2^2)(m_1+m_2)}\Big]$$
$$\times\Big[\frac{2l_1^2l_2^2(m_1+m_2)^2}{(l_1^2+l_2^2)}\Big]^{-1}\Big]\mathbf{k}\cdot\mathbf{K}\Big\}\,, \tag{4.50}$$

$$\alpha_{n+1}=\frac{l_1^2+l_2^2}{2}\,,\quad \beta_{n+1}=\frac{l_1^2m_1^2+l_2^2m_2^2}{2(m_1+m_2)^2}\,, \tag{4.51}$$

$$\gamma_{n+1}=\frac{(l_1^2m_1-l_2^2m_2)}{m_1+m_2}\mathbf{K}\,,\quad \delta_{n+1}=\frac{(l_1^2m_1^2+l_2^2m_2^2)}{(m_1+m_2)^2}\mathbf{K}\,, \tag{4.52}$$

$$\epsilon_{n+1}=\frac{-l_1^2m_1+l_2^2m_2}{m_1+m_2}\,. \tag{4.53}$$

We see that

$$(l_1l_2)^{-3/2}a_{n+1}\overset{l_1,l_2\to 0}{\to}\Big(\frac{m}{4\pi}\Big)^n(4\pi^3)^{-3/2}\,, \tag{4.54}$$

provided that

$$\frac{(l_1^2 m_1 - l_2^2 m_2)^2 (l_1^2 + l_2^2)}{l_1^2 l_2^2} \overset{l_1, l_2 \to 0}{\to} 0, \tag{4.55}$$

and that

$$\{ (l_1^2 m_1^2 + l_2^2 m_2^2) - 2[(l_1^2 m_1 - l_2^2 m_2)^2/(l_1^2 + l_2^2)] \}^2 \times (l_1^2 + l_2^2)\}/(l_1^2 l_2^2) \overset{l_1, l_2 \to 0}{\to} 0, \tag{4.56}$$

These conditions are satisfied if the radius of one particle does not decrease much faster than the radius of the other one. Thus, upon performing all integrations with respect to the momentum variables (except $\mathbf{p}$ and $\mathbf{P}$) in the expression (5.8), where $D_{\mathbf{k}}^{(+)}$ is given by (5.10), and then letting the parameters $l_1$ and $l_2$ go to zero, we see that

$$\phi_{\mathbf{K},\mathbf{k}}^{(+)}(\xi_1, \xi_2) - \phi_{\mathbf{K},\mathbf{k}}(\xi_1, \xi_2) \tag{4.57}$$

goes upon renormalization by the factor

$$(l_1 l_2/\pi)^{-3/2} \tilde{d}^{(+)}(\mathbf{k}^2/2m)/d^{(+)}(\mathbf{k}^2/2m) \tag{4.58}$$

to the function

$$[\phi_{\mathbf{k}}^{(+)}(\mathbf{q}) - \phi_{\mathbf{k}}(\mathbf{q})]\phi_{\mathbf{K}}(\mathbf{Q}), \tag{4.59}$$

where $\phi_{\mathbf{k}}^{(+)}$ is the solution (4.60) of the conventional Lippmann–Schwinger equation (1.4)[1] for pointlike particles in configuration space, and $\tilde{d}^{(+)}$ and $d^{(+)}$ are, respectively, given by (4.62)–(4.63) and (5.9), (5.11):

$$\phi_{\mathbf{k}}^{(+)}(\mathbf{r}) = \phi_{\mathbf{k}}(\mathbf{r}) + \frac{V^{-1/2}(\mathbf{r})}{\tilde{d}^{(+)}(\mathbf{k}^2/2m)} \times \int_{\mathbf{R}^3} \tilde{D}^{(+)}\left(\mathbf{r}, \mathbf{r}'; \frac{\mathbf{k}^2}{2m}\right) V^{1/2}(\mathbf{r}') \phi_{\mathbf{k}}(\mathbf{r}') d\mathbf{r}', \tag{4.60}$$

$$\tilde{D}^{(+)}(\mathbf{r}, \mathbf{r}'; \lambda) = \sum_{n=0}^{\infty} \tilde{D}_n^{(+)}(\mathbf{r}, \mathbf{r}'; \lambda), \tag{4.61}$$

$$\tilde{d}^{(+)}(\lambda) = \sum_{n=0}^{\infty} \tilde{d}_n^{(+)}(\lambda), \tag{4.62}$$

$$\tilde{d}_n^{(+)}(\lambda) = \frac{(-1)^n}{n!} \int_{\mathbf{R}^{3n}} \begin{vmatrix} 0 & \tilde{K}^{(+)}(\mathbf{r}_1, \mathbf{r}_2; \lambda) & \cdots & \tilde{K}^{(+)}(\mathbf{r}_1, \mathbf{r}_n; \lambda) \\ \tilde{K}^{(+)}(\mathbf{r}_2, \mathbf{r}_1; \lambda) & 0 & \cdots & \tilde{K}^{(+)}(\mathbf{r}_2, \mathbf{r}_n; \lambda) \\ \vdots & \vdots & & \vdots \\ \tilde{K}^{(+)}(\mathbf{r}_n, \mathbf{r}_1; \lambda) & \tilde{K}^{(+)}(\mathbf{r}_n, \mathbf{r}_2; \lambda) & \cdots & 0 \end{vmatrix} d\mathbf{r}_1 \, d\mathbf{r}_2 \cdots d\mathbf{r}_n, \tag{4.63}$$

$$\tilde{D}_n^{(+)}(\mathbf{r}, \mathbf{r}'; \lambda) = \frac{(-1)^n}{n!} \int_{\mathbf{R}^{3n}} \begin{vmatrix} \tilde{K}^{(+)}(\mathbf{r}, \mathbf{r}'; \lambda) & \tilde{K}^{(+)}(\mathbf{r}, \mathbf{r}_1; \lambda) & \cdots & \tilde{K}^{(+)}(\mathbf{r}, \mathbf{r}_n; \lambda) \\ \tilde{K}^{(+)}(\mathbf{r}_1, \mathbf{r}'; \lambda) & 0 & \cdots & \tilde{K}^{(+)}(\mathbf{r}_1, \mathbf{r}_n; \lambda) \\ \vdots & \vdots & & \vdots \\ \tilde{K}^{(+)}(\mathbf{r}_n, \mathbf{r}'; \lambda) & \tilde{K}^{(+)}(\mathbf{r}_n, \mathbf{r}_1; \lambda) & \cdots & 0 \end{vmatrix} d\mathbf{r}_1 \, d\mathbf{r}_2 \cdots d\mathbf{r}_n, \tag{4.64}$$

$$\tilde{K}^{(+)}(\mathbf{r}, \mathbf{r}'; \lambda) = V^{1/2}(\mathbf{r}) G_0^{(+)}(\mathbf{r}, \mathbf{r}'; \lambda) |V(\mathbf{r}')|^{1/2}, \tag{4.65}$$

$$V^{1/2}(\mathbf{r}) = \begin{cases} |V(\mathbf{r})|^{1/2} \operatorname{sgn} V(\mathbf{r}), & \text{if } V(\mathbf{r}) \neq 0, \\ \eta \exp(-\mathbf{r}^2), \quad \eta > 0, & \text{if } V(\mathbf{r}) = 0. \end{cases} \tag{4.66}$$

Hence, in order to obtain for $\phi_{\mathbf{K},\mathbf{k}}^{(+)}$ the same renormalization factor as for $\phi_{\mathbf{K},\mathbf{k}}$ we must have that $\tilde{d}^{(+)}$ converges to $d^{(+)}$ when the parameters $l_1$ and $l_2$ go to zero. Using Eq. (6.35) of Chap. 5 in Ref. 1 we see that

$$D_{\mathbf{K},n}^{(+)}(\mathbf{q}, \mathbf{p}_1, \mathbf{p}_2; \mathbf{q}', \mathbf{p}_1', \mathbf{p}_2'; \lambda) = d_n^{(+)}(\lambda) \tilde{G}_{0,\mathbf{K}}^{(+)}(\mathbf{q}, \mathbf{p}_1, \mathbf{p}_2; \mathbf{q}', \mathbf{p}_1', \mathbf{p}_2'; \lambda) + \int_{\mathbf{R}^9} D_{\mathbf{K},n-1}^{(+)}(\mathbf{q}, \mathbf{p}_1, \mathbf{p}_2; \mathbf{q}'', \mathbf{p}_1'', \mathbf{p}_2''; \lambda) \times \tilde{G}_{0,\mathbf{K}}^{(+)}(\mathbf{q}_1'', \mathbf{p}_1'', \mathbf{p}_2''; \mathbf{q}', \mathbf{p}_1', \mathbf{p}_2'; \lambda) d\mathbf{q}'' \, d\mathbf{p}_1'', d\mathbf{p}_2''. \tag{4.67}$$

And therefore, performing the integration with respect to $\mathbf{p}_1''$ and $\mathbf{p}_2''$ and letting $l_1, l_2 \to 0$ we obtain

$$d_n^{(+)}(\lambda) \overset{l_1, l_2 \to 0}{\to} \frac{\tilde{D}_n^{(+)}(\mathbf{q}, \mathbf{q}'; \lambda)}{\tilde{K}^{(+)}(\mathbf{q}, \mathbf{q}'; \lambda)} - \frac{\int \tilde{D}_{n-1}^{(+)}(\mathbf{q}, \mathbf{q}'') \tilde{K}^{(+)}(\mathbf{q}'', \mathbf{q}') d\mathbf{q}''}{\tilde{K}^{(+)}(\mathbf{q}, \mathbf{q}'; \lambda)} = \tilde{d}_n^{(+)}(\lambda). \tag{4.68}$$

Consequently, in the sharp-point limit, we indeed have the same renormalization factor for free and distorted phase-space waves.

Now, we intend to investigate the $T$-matrix, defined in terms of

$$T = \underset{\epsilon \to +0}{\text{w-lim}} \, T_\epsilon, \tag{4.69}$$

where the operator $T_\epsilon$ can be computed from

$$\langle g | T_\epsilon f \rangle = \frac{1}{\pi} \int_{\mathbf{R}^6} d\mathbf{k}_1 \, d\mathbf{k}_2 \left\langle g | \phi_{\mathbf{k}_1, \mathbf{k}_2} \right\rangle \cdot (\phi_{\mathbf{k}_1, \mathbf{k}_2} | \mathbb{P}(\Gamma_{12}) V \Omega^+ \times \left[ \left( H_0 - \frac{\mathbf{k}_1^2}{2m_1} - \frac{\mathbf{k}_2^2}{2m_2} \right) + \epsilon^2 \right]^{-1} \epsilon f \right\rangle. \tag{4.70}$$

**Theorem 4.2:** Let us assume that the conditions of Theorem 5.4 on the potential $V$ are satisfied. Then the transition operator $T$ satisfies

$$\langle g | Tf \rangle = \lim_{\epsilon \to 0} \left[ \int d\mathbf{k} \, \tilde{g}_1^*(\mathbf{k}) \times \int d\mathbf{k}' \, \delta_\epsilon \left( \frac{\mathbf{k}'^2}{2m} - \frac{\mathbf{k}^2}{2m} \right) \langle \mathbf{k} | T | \mathbf{k}' \rangle \tilde{f}_1(\mathbf{k}') \right] \langle g_2 | f_2 \rangle, \tag{4.71}$$

with $f$ and $g$ related to $f_i$ and $g_i$, $i = 1,2$, by

$$\tilde{f}(\mathbf{k}_1, \mathbf{k}_2) = \tilde{f}_1(\mathbf{k}) \tilde{f}_2(\mathbf{K}), \tag{4.72}$$

$$\tilde{g}(\mathbf{k}_1, \mathbf{k}_2) = \tilde{g}_1(\mathbf{k}) \tilde{g}_2(\mathbf{K}), \tag{4.73}$$

with the $T$-matrix

$$\langle \mathbf{k}|T|\mathbf{k}'\rangle = (2\pi)^3(\hat{\phi}_{\mathbf{K},\mathbf{k}}|V|\hat{\phi}^{(+)}_{\mathbf{K},\mathbf{k}'}) \tag{4.74}$$

that does not depend on $\mathbf{K}$, and with

$$\delta_\epsilon(\mathbf{x}) = \pi^{-1}\epsilon(\mathbf{x}^2+\epsilon^2)^{-1}. \tag{4.75}$$

The proof follows from the formula

$$\text{l.i.m.}\int \tilde{\xi}_{\mathbf{q},\mathbf{p}}(\mathbf{k})\langle \xi_{\mathbf{q},\mathbf{p}}|\xi_{\mathbf{q}',\mathbf{p}'}\rangle d\mathbf{q}\,d\mathbf{p} = \tilde{\xi}_{\mathbf{q}',\mathbf{p}'}(\mathbf{k}), \tag{4.76}$$

which can be derived from (2.12). The independence of (4.74) on $\mathbf{K}$ follows from the fact that $d^{(+)}$ does not depend on $\mathbf{K}$ [in view of (4.12), (5.9), and (5.11)], and from

$$(\hat{\phi}_{\mathbf{K},\mathbf{k}}|V|\hat{\phi}^{(+)}_{\mathbf{K},\mathbf{k}'}) = (\hat{\phi}_{0,\mathbf{k}}|V|\hat{\phi}^{(+)}_{0,\mathbf{k}'}) \tag{4.77}$$

which follows from (4.12), (4.5), (5.8), and (5.10), (5.12).

It can also be checked that

$$(\hat{\phi}_{\mathbf{K},\mathbf{k}}|V|\hat{\phi}^{(+)}_{\mathbf{K},\mathbf{k}'}) \overset{l_1,l_2\to 0}{\to} (\phi_{\mathbf{k}}|V|\phi^{(+)}_{\mathbf{k}'}), \tag{4.78}$$

provided that $l_1$ (or $l_2$) do not decrease much faster than $l_2$ (or $l_1$).

Let us now study the first Born approximation $\langle \mathbf{k}|T|\mathbf{k}'\rangle^{(1)}$ obtained, as usual, by replacing $\phi^{(+)}$ by $\phi$ in the $T$-matrix formula [in our case, formula (4.74)]. We obtain

$$\langle \mathbf{k}|T|\mathbf{k}'\rangle^{(1)} = (\phi_{\mathbf{k}}|V_{\text{eff}}|\phi_{\mathbf{k}'}), \tag{4.79}$$

where $\phi_{\mathbf{k}}$ is the configuration space plane wave, and where $V_{\text{eff}}$ is given in (3.9). Therefore, we have

$$\begin{aligned}\langle \mathbf{k}|T|\mathbf{k}'\rangle^{(1)} &= h^{-3/2}(V_{12})(\mathbf{k}-\mathbf{k}')\\ &= h^{3/2}\tilde{V}(\mathbf{k}-\mathbf{k}')\tilde{\chi}^{(1)}(\mathbf{k}-\mathbf{k}')\tilde{\chi}^{(2)}(\mathbf{k}-\mathbf{k}').\end{aligned} \tag{4.80}$$

In other words, we obtain the form factor $h^{3/2}\tilde{\chi}^{(1)}\tilde{\chi}^{(2)}$ in the resulting first-order Born approximation for the $T$-matrix. Thus, we see that particle extension that manifests itself in terms of nontrivial proper wave functions indeed gives rise to form factors in the scattering theory for such particles.

## V. ASYMPTOTIC COMPLETENESS

In this section we shall solve the Lippmann–Schwinger equation and prove the asymptotic completeness. The Fredholm method can be used for solving that equation, but it cannot be applied directly to Eq. (4.15) since $\hat{\phi}_{\mathbf{K},\mathbf{k}}(\mathbf{q}\cdot\mathbf{p}_1,\mathbf{p}_2)$ is not square integrable in $\mathbf{q},\mathbf{p}_1,\mathbf{p}_2$. Hence, let us perform a Rollnik decomposition

$$\tilde{\phi}_{\mathbf{K},\mathbf{k}}(\mathbf{q},\mathbf{p}_1,\mathbf{p}_2) = V^{1/2}(\mathbf{q})\hat{\phi}_{\mathbf{K},\mathbf{k}}(\mathbf{q},\mathbf{p}_1,\mathbf{p}_2), \tag{5.1}$$

$$\tilde{\phi}^{(+)}_{\mathbf{K},\mathbf{k}}(\mathbf{q},\mathbf{p}_1,\mathbf{p}_2) = V^{1/2}(\mathbf{q})\hat{\phi}^{(+)}_{\mathbf{K},\mathbf{k}}(\mathbf{q},\mathbf{p}_1,\mathbf{p}_2), \tag{5.2}$$

$$\begin{aligned}&\tilde{G}^{(+)}_{0,\mathbf{K}}(\mathbf{q},\mathbf{p}_1,\mathbf{p}_2;\mathbf{q}',\mathbf{p}_1',\mathbf{p}_2';\lambda)\\ &\quad = V^{1/2}(\mathbf{q})G^{(+)}_{0,\mathbf{K}}(\mathbf{q},\mathbf{p}_1,\mathbf{p}_2;\mathbf{q}',\mathbf{p}_1',\mathbf{p}_2';\lambda)|V(\mathbf{q}')|^{1/2},\end{aligned} \tag{5.3}$$

$$V^{1/2}(\mathbf{q}) = |V(\mathbf{q})|^{1/2}\,\text{sgn}\,V(\mathbf{q}), \tag{5.4}$$

and we shall assume that $V(\mathbf{q})\neq 0$ a.e. The Lippmann–Schwinger equation then takes the form

$$\begin{aligned}\tilde{\phi}^{(+)}_{\mathbf{K},\mathbf{k}}(\mathbf{q},\mathbf{p}_1,\mathbf{p}_2) &= \tilde{\phi}_{\mathbf{K},\mathbf{k}}(\mathbf{q},\mathbf{p}_1,\mathbf{p}_2) + \int d\mathbf{q}'\,d\mathbf{p}_1'\,d\mathbf{p}_2'\\ &\quad\times\tilde{G}^{(+)}_{0,\mathbf{K}}\left(\mathbf{q},\mathbf{p}_1,\mathbf{p}_2;\mathbf{q}',\mathbf{p}_1',\mathbf{p}_2';\frac{\mathbf{k}^2}{2m}\right)\\ &\quad\times\tilde{\phi}^{(+)}_{\mathbf{K},\mathbf{k}}(\mathbf{q}',\mathbf{p}_1',\mathbf{p}_2').\end{aligned} \tag{5.5}$$

**Theorem 5.1:** (Ref. 12) Suppose that the potential $V(\mathbf{q})$ satisfying

$$V(\mathbf{q}) = O(|\mathbf{q}|^{-3-\epsilon}),\quad |\mathbf{q}|\to\infty, \tag{5.6}$$

is different from zero almost everywhere and is of Rollnick class, i.e.,

$$\int \frac{|V(\mathbf{q})V(\mathbf{q}')|}{|\mathbf{q}-\mathbf{q}'|^2}\,d\mathbf{q}\,d\mathbf{q}' < \infty. \tag{5.7}$$

Then for $\mathbf{k}\notin S^3_{V,\mathbf{K}}$, where $S^3_{V,\mathbf{K}}$ is a compact set of Lebesgue measure zero for each value of $\mathbf{K}$, the Lippman–Schwinger equation (4.15) has the unique solution

$$\begin{aligned}\phi^{(+)}_{\mathbf{k}_1,\mathbf{k}_2}(\xi_1,\xi_2) &= \phi_{\mathbf{k}_1,\mathbf{k}_2}(\xi_1,\xi_2) + \frac{V^{-1/2}(\mathbf{q})}{d^{(+)}(k^2/2m)}\exp(i\mathbf{K}\cdot\mathbf{Q})\\ &\quad\times\int D^{(+)}_{\mathbf{K}}\left(\mathbf{q},\mathbf{p}_1,\mathbf{p}_2;\mathbf{q}',\mathbf{p}_1',\mathbf{p}_2';\frac{\mathbf{k}^2}{2m}\right)V^{1/2}(\mathbf{q}')\\ &\quad\times\exp(-i\mathbf{K}\cdot\mathbf{Q}')\phi_{\mathbf{k}_1,\mathbf{k}_2}(\xi_1',\xi_2')d\mathbf{q}'\,d\mathbf{p}_1'\,d\mathbf{p}_2',\end{aligned} \tag{5.8}$$

in which

$$d^{(+)}(\lambda) = \sum_{n=0}^{\infty} d^{(+)}_n(\lambda), \tag{5.9}$$

$$D^{(+)}_{\mathbf{K}}(\mathbf{q},\mathbf{p}_1,\mathbf{p}_2;\mathbf{q}',\mathbf{p}_1',\mathbf{p}_2';\lambda) = \sum_{n=0}^{\infty} D^{(+)}_{\mathbf{K},n}(\mathbf{q},\mathbf{p}_1,\mathbf{p}_2;\mathbf{q}',\mathbf{p}_1',\mathbf{p}_2';\lambda), \tag{5.10}$$

$$d^{(+)}_n(\lambda) = \frac{(-1)^n}{n!}\int \begin{vmatrix} 0 & \tilde{G}^{(+)}_{0,\mathbf{K}}(\alpha_1,\alpha_2;\lambda) & \cdots & \tilde{G}^{(+)}_{0,\mathbf{K}}(\alpha_1,\alpha_n;\lambda)\\ \tilde{G}^{(+)}_{0,\mathbf{K}}(\alpha_2,\alpha_1;\lambda) & 0 & \cdots & \tilde{G}^{(+)}_{0,\mathbf{K}}(\alpha_2,\alpha_n;\lambda)\\ \vdots & \vdots & & \vdots\\ \tilde{G}^{(+)}_{0,\mathbf{K}}(\alpha_n,\alpha_1;\lambda) & \tilde{G}^{(+)}_{0,\mathbf{K}}(\alpha_n,\alpha_2;\lambda) & \cdots & 0\end{vmatrix} d\alpha_1\,d\alpha_2\cdots d\alpha_n, \tag{5.11}$$

$$D^{(+)}_{\mathbf{K},n}(\alpha,\alpha';\lambda) = \frac{(-1)^n}{n!}\begin{vmatrix} \tilde{G}^{(+)}_{0,\mathbf{K}}(\alpha,\alpha';\lambda) & \tilde{G}^{(+)}_{0,\mathbf{K}}(\alpha,\alpha_1;\lambda) & \cdots & \tilde{G}^{(+)}_{0,\mathbf{K}}(\alpha,\alpha_n;\lambda)\\ \tilde{G}^{(+)}_{0,\mathbf{K}}(\alpha_1,\alpha';\lambda) & 0 & \cdots & \tilde{G}^{(+)}_{0,\mathbf{K}}(\alpha_1,\alpha_n;\lambda)\\ \vdots & \vdots & & \vdots\\ \tilde{G}^{(+)}_{0,\mathbf{K}}(\alpha_n,\alpha';\lambda) & \tilde{G}^{(+)}_{0,\mathbf{K}}(\alpha_n,\alpha_1;\lambda) & \cdots & 0\end{vmatrix} d\alpha_1\,d\alpha_2\cdots d\alpha_n, \tag{5.12}$$

$$\alpha,\alpha',\alpha_1,\alpha_2,\ldots,\alpha_n\in\mathbb{R}^9. \tag{5.13}$$

The proof of the above theorem[12] consists in showing that the function $\tilde{\phi}_{\mathbf{K},\mathbf{k}}\in L^2(\mathbb{R}^9)$ and that the kernel $\tilde{G}^{(+)}_{0,\mathbf{K}}$ is Hilbert–Schmidt. The fact that $S'_{V,\mathbf{K}} = \{\mathbf{k}^2/2m : \mathbf{k}\in S^3_{V,\mathbf{K}}\}$ is compact and of Lebesgue measure zero is established in Theorem 6.7 of Chap. 5 in Ref. 1. Therefore $S^6_V$ is of measure zero since $d^{(+)}(\lambda)$ do not depend on $\mathbf{K}$.

We now intend to prove that the functions (5.8) are the distorted phase-space waves satisfying (4.7). The proof has to be carried in several stages. Let us note first that the operator $G_0(\zeta)$ can be considered as an integral operator with kernel

$$G_0(\xi_1,\xi_2;\mathbf{k}_1,\mathbf{k}_2;\zeta) = \phi_{\mathbf{k}_1,\mathbf{k}_2}(\xi_1,\xi_2)(\zeta - \mathbf{k}_1^2/2m_1 - \mathbf{k}_2^2/2m_2)^{-1}. \tag{5.14}$$

There might be a similar relation for the kernel of the full Green's function.

**Theorem 5.2:** (Ref. 12) Let us assume that the Green's operator $G(\zeta)$ is an integral operator with kernel $G(\xi_1,\xi_2;\mathbf{k}_1,\mathbf{k}_2;\zeta)$. Then we have the following relation for this kernel (called the full Green's function in stochastic phase space):

$$G(\xi_1,\xi_2;\mathbf{k}_1,\mathbf{k}_2;\zeta) = \text{l.i.m.} \int d\xi_1'\, d\xi_2'\, \phi_{\mathbf{k}_1,\mathbf{k}_2}(\xi_1',\xi_2') \times \int d\mathbf{k}_1'\, d\mathbf{k}_2'\, G^*(\xi_1',\xi_2';\mathbf{k}_1',\mathbf{k}_2';\zeta^*) \times \phi_{\mathbf{k}_1',\mathbf{k}_2'}(\xi_1,\xi_2). \tag{5.15}$$

The proof relies on the fact that $G^*(\zeta) = G(\zeta^*)$.

In view of (5.14), let us set

$$G(\xi_1,\xi_2;\mathbf{k}_1,\mathbf{k}_2;\zeta) = \phi_{\mathbf{k}_1,\mathbf{k}_2}(\xi_1,\xi_2;\zeta)(\zeta - \mathbf{k}_1^2/2m_1 - \mathbf{k}_2^2/2m_2)^{-1}, \tag{5.16}$$

for $\zeta \notin S^H$. Then the next step in the proof of asymptotic completeness consists in proving the following theorem.

**Theorem 5.3:** Let us suppose that the potential $V(\mathbf{q}) = O(|\mathbf{q}|^{-3-\epsilon})$ is of Rollnick class, locally square integrable and different from zero almost everywhere. Then

$$\hat{\psi}_{\pm}(\mathbf{k}_1,\mathbf{k}_2) = \text{l.i.m.} \int \phi^{(\pm)*}_{\mathbf{k}_1,\mathbf{k}_2}(\xi_1,\xi_2)\psi(\xi_1,\xi_2)d\xi_1\, d\xi_2 \tag{5.17}$$

exists for any $(\mathbf{k},\mathbf{K}) \notin S^6_V$,

$$S^6_V = \{(\mathbf{k},\mathbf{K}) \in \mathbf{R}^6 : \mathbf{k} \in S^3_{V,\mathbf{K}}\}, \tag{5.18}$$

if $\phi^{(\pm)}_{\mathbf{k}_1,\mathbf{k}_2}(\xi_1,\xi_2)$ are the solutions (5.8) of (4.15) and (4.9). Furthermore if $B \subset (0,\infty)$ is a compact set and $B \cap S^1_V = \phi$, where $S^1_V = \Lambda^{-1} S^6_V$, with $\Lambda$ given by (4.8), then

$$\|E^H(B)\psi\|^2 = \int_{\mathbf{k}_1^2/2m_1 + \mathbf{k}_2^2/2m_2} |\hat{\psi}_{\pm}(\mathbf{k}_1,\mathbf{k}_2)|^2\, d\mathbf{k}_1\, d\mathbf{k}_2. \tag{5.19}$$

*Proof*: (Ref. 12) Using (5.16) and the second resolvent equation we see that the Green's function is of the form

$$G(\xi_1,\xi_2;\mathbf{k}_1,\mathbf{k}_2;\zeta) = \hat{G}_{\mathbf{K},\mathbf{k}}(\mathbf{q},\mathbf{p}_1,\mathbf{p}_2;\zeta)\exp(i\mathbf{K}\cdot\mathbf{Q}) \tag{5.20}$$

for some $\hat{G}_{\mathbf{K},\mathbf{k}}(\,\cdot\,;\zeta)$ which does not depend on $\mathbf{Q}$. Therefore, writing

$$\phi_{\mathbf{k}_1,\mathbf{k}_2}(\xi_1,\xi_2;\zeta) = \hat{\phi}_{\mathbf{K},\mathbf{k}}(\mathbf{q},\mathbf{p}_1,\mathbf{p}_2;\zeta)\exp(i\mathbf{K}\cdot\mathbf{Q}), \tag{5.21}$$

for some $\hat{\phi}_{\mathbf{K},\mathbf{k}}(\,\cdot\,;\zeta)$ not depending on $\mathbf{Q}$ we obtain

$$\hat{\phi}_{\mathbf{K},\mathbf{k}}(\mathbf{q},\mathbf{p}_1,\mathbf{p}_2;\zeta) = \hat{\phi}_{\mathbf{K},\mathbf{k}}(\mathbf{q},\mathbf{p}_1,\mathbf{p}_2) + \int d\mathbf{q}'\, d\mathbf{p}_1'\, d\mathbf{p}_2' \times G_{0,\mathbf{K}}\left(\mathbf{q},\mathbf{p}_1,\mathbf{p}_2;\mathbf{q}',\mathbf{p}_1',\mathbf{p}_2';\zeta - \frac{\mathbf{K}^2}{2M}\right) \times V(\mathbf{q}')\hat{\phi}_{\mathbf{K},\mathbf{k}}(\mathbf{q}',\mathbf{p}_1',\mathbf{p}_2'). \tag{5.22}$$

Performing a Rollnik decomposition, we obtain an equation which is solvable by the Fredholm method. Therefore, defining

$$\hat{\psi}(\mathbf{k}_1,\mathbf{k}_2;\zeta) = \int \phi^*_{\mathbf{k}_1,\mathbf{k}_2}(\xi_1,\xi_2;\zeta^*)\psi(\xi_1,\xi_2)d\xi_1\, d\xi_2, \tag{5.23}$$

we obtain, by (5.16), Theorem 5.2 and the Fubini and Tonelli theorems

$$(G(\zeta)\psi)^{\sim}(\mathbf{k}_1,\mathbf{k}_2) = (\zeta - \mathbf{k}_1^2/2m_1 - \mathbf{k}_2^2/2m_2)^{-1}\hat{\psi}(\mathbf{k}_1,\mathbf{k}_2;\zeta). \tag{5.24}$$

We get from the first resolvent equation and Theorem 5.12 of Chap. 5 in Ref. 1

$$\langle\psi|(E_b^H - E_a^H)\psi\rangle = \frac{1}{\pi}\lim_{\epsilon\to+0}\int_a^b d\lambda \times \int \frac{\epsilon|\hat{\psi}(\mathbf{k}_1,\mathbf{k}_2;\lambda\pm i\epsilon)|^2}{(\lambda - \mathbf{k}_1^2/2m_1 - \mathbf{k}_2^2/2m_2)^2 + \epsilon^2} d\mathbf{k}_1\, d\mathbf{k}_2, \tag{5.25}$$

since $E^H_\lambda = E^H_{\lambda-0}$ for $\lambda \notin S^1_V$, because $S^H_p$ is disjoint from $[0,\infty)\backslash S^1_V$. Indeed $[0,\infty)\backslash S^1_V \subset S^H_c \subset [0,\infty)$. Furthermore, from the explicit expression for the solution of (5.22), we see that

$$\lim_{\epsilon\to+0}\hat{\psi}(\mathbf{k}_1,\mathbf{k}_2;\mathbf{k}_1^2/2m_1 + \mathbf{k}_2^2/2m_2 \pm i\epsilon) = \hat{\psi}_{\mp}(\mathbf{k}_1,\mathbf{k}_2). \tag{5.26}$$

Q.E.D.

We obtain the following theorem, which is a consequence of the above one.

**Theorem 5.4:** Suppose that the conditions of Theorem 5.3 are satisfied. Then the continuous spectrum of $H$ satisfies

$$[0,\infty)\backslash S^1_V \subset S^H_c \subset [0,\infty), \tag{5.27}$$

and for any Borel set $B \subset S^H_c$,

$$(E^H(B)\psi)(\xi_1,\xi_2) = \text{l.i.m.} \int_{\mathbf{k}_1^2/2m_1 + \mathbf{k}_2^2/2m_2 \in B} \phi^{(\pm)}_{\mathbf{k}_1,\mathbf{k}_2}(\xi_1,\xi_2) \times \hat{\psi}_+(\mathbf{k}_1,\mathbf{k}_2)d\mathbf{k}_1\, d\mathbf{k}_2, \tag{5.28}$$

$$\langle f|E^H(B)g\rangle = \int_{\mathbf{k}_1^2/2m_1 + \mathbf{k}_2^2/2m_2 \in B} \hat{f}^*_+(\mathbf{k}_1,\mathbf{k}_2)\hat{g}_+(\mathbf{k}_1,\mathbf{k}_2)d\mathbf{k}_1\, d\mathbf{k}_2, \tag{5.29}$$

where $\hat{\psi}_+$ is given by (5.17).

The existence of strong asymptotic states has been proven in Ref. 12.

**Theorem 5.5:** Let us assume that the potential $V(\mathbf{q})$ is locally square integrable, and that for some $\epsilon > 0$,

$$V(\mathbf{q}) = O(|\mathbf{q}|^{-1-\epsilon}), \quad |\mathbf{q}| \to \infty, \tag{5.30}$$

then

$$M_0 = M_+ = \left\{ f \in L^2(\Gamma_{12}) : \text{s-lim}_{t\to-\infty}\, \Omega(t) f \text{ exists} \right\} = L^2(\Gamma_{12}). \tag{5.31}$$

The proof[12] consists in showing that there exists a domain $\hat{\mathcal{D}}_1$ dense in $L^2(\Gamma)$ satisfying

$$\exp(-itH_0)\hat{\mathcal{D}}_1 \subset \mathcal{D}_{H_0}, \tag{5.32}$$

$$\int_{-\infty}^{\infty} \|V\exp(-itH_0)\mathbf{P}(\Gamma_{12})\psi\| dt < \infty, \tag{5.33}$$

for all $\psi \in \hat{\mathcal{D}}_1$. The following set

$$\hat{\mathscr{D}}_1 = \Big\{\psi_\rho : \text{l.i.m.}(2\pi)^{-6}\int d\xi_1\, d\xi_2$$
$$\times \exp(-i\xi_1\cdot\xi_1' - i\xi_2 - \xi_2')\psi_\rho(\xi_1,\xi_2)$$
$$= \tilde{\psi}_\rho(\xi_1',\xi_2') = \exp\Big[-\frac{\mathbf{q}_1'^2}{2m_1} - \frac{\mathbf{q}_2'^2}{2m_2} - \frac{\mathbf{p}_1'^2}{a} - \frac{\mathbf{p}_2'^2}{b}\Big]$$
$$\times \exp(-i(\xi_1',\xi_2')\cdot\rho),\ \rho\in\mathbf{R}^{12}\Big\} \tag{5.34}$$

satisfies these requirements.

Before proving the asymptotic completeness, we have to check that the following expression

$$(W_+\psi)(\xi_1,\xi_2) = \text{l.i.m.}\int \phi^{(+)}_{\mathbf{k}_1,\mathbf{k}_2}(\xi_1,\xi_2)\tilde{\psi}(\mathbf{k}_1,\mathbf{k}_2)d\,\mathbf{k}_1\,d\,\mathbf{k}_2 \tag{5.35}$$

determines a well-defined operator $W_+$.

**Theorem 5.6:** Suppose that $\phi^{(+)}_{\mathbf{k}_1,\mathbf{k}_2}(\xi_1,\xi_2)$ are the solutions (5.8) of the Lippmann–Schwinger equation (4.15), where the conditions of Theorem 5.3 are satisfied. Then there exists a unique operator $W_+$, defined by (5.35), and satisfying

$$W_+ W_+^* = E^H(S_c^H), \tag{5.36}$$

$$W_+^* E^H(B) = E^{H_0}(B)W_+^*. \tag{5.37}$$

*Proof:* (Ref. 12) Let us consider the bilinear form

$$(\psi|\psi') = \int \tilde{\psi}^*(\mathbf{k}_1,\mathbf{k}_2)\hat{\psi}'_+(\mathbf{k}_1,\mathbf{k}_2)d\,\mathbf{k}_1\,d\,\mathbf{k}_2\,. \tag{5.38}$$

By Theorem (5.4), this bilinear form is bounded.[12] Consequently there exists a unique bounded linear operator $W_+$ for which

$$\langle W_+\psi|\psi'\rangle = (\psi|\psi'), \tag{5.39}$$

for every $\psi,\psi'\in L^2(\Gamma_{12})$. Assuming that $\tilde{\psi}\in C_V^\infty$ and $\psi'\in C_b^\infty$ we can drop the l.i.m. in (5.35), and (5.38) is equal to

$$(\psi|\psi') = \int d\,\mathbf{k}_1\,d\,\mathbf{k}_2\,\tilde{\psi}^*(\mathbf{k}_1,\mathbf{k}_2)$$
$$\times\int d\,\xi_1\,d\,\xi_2\,\phi^{(+)*}_{\mathbf{k}_1,\mathbf{k}_2}(\xi_1,\xi_2)\psi'(\xi_1,\xi_2)\,. \tag{5.40}$$

On account of the conditions of Theorem 5.3 we can apply Fubini's theorem to (5.40)[12] to obtain (5.35). We can easily show that

$$(W_+^*\psi)(\xi_1,\xi_2) = \text{l.i.m.}\int \phi_{\mathbf{k}_1,\mathbf{k}_2}(\xi_1,\xi_2)\hat{\psi}_+(\mathbf{k}_1,\mathbf{k}_2)d\,\mathbf{k}_1\,d\,\mathbf{k}_2\,, \tag{5.41}$$

for any $\hat{\psi}_+$ given by (5.17). Equation (5.36) is a consequence of (5.28), (5.35), and (5.41); whereas Eq. (5.37) is a consequence of (5.28), (5.29), (5.17), and (5.41). Q.E.D.

We are now left with the task of proving that $W_+ = \Omega_+$, and thus establish asymptotic completeness for $\Omega_+$.

**Theorem 5.7:** Let us assume that a potential satisfies the conditions of Theorem 5.3. Then $W_+ = \Omega_+$.

*Proof:* Since $\psi\in\mathscr{D}_{H_I}$ if $\psi\in\mathscr{D}_{H_0}$, we have by Eq. (2.37) of Chap. 5 in Ref. 1

$$\langle\Omega(\tau)\psi|g\rangle = \langle\psi|g\rangle - i\int_0^\tau \langle\exp(itH)H_I$$
$$\times\exp(-itH_0)\psi|g\rangle dt\,, \tag{5.42}$$

for any $g\in L^2(\Gamma_{12})$. Setting $g = W_+\varphi$, $\varphi\in L^2(\Gamma_{12})$ and using

$$W_+^*\exp(itH) = \exp(itH_0)W_+^*\,, \tag{5.43}$$

which is a consequence of (5.37),[1] we obtain

$$\langle\varphi|W_+^*\Omega(\tau)\psi\rangle = \langle\varphi|W_+^*\psi\rangle + i\int_0^\tau dt$$
$$\times\langle\varphi|\exp(itH_0)W_+^* H_I\exp(-itH_0)\psi\rangle\,. \tag{5.44}$$

Now using Exercise (7.1) in Chap. 5 of Ref. 1 we obtain

$$\langle\varphi|W_+^*\Omega_+\psi\rangle = \langle\varphi|W_+^*\psi\rangle + i\lim_{\epsilon\to 0}\int_0^\infty e^{-\epsilon t}\langle\varphi|$$
$$\times\exp(iH_0t)W_+^* H_I\exp(-itH_0)\psi\rangle dt\,. \tag{5.45}$$

Using (5.41), (4.1), (4.2), and (5.17), we obtain[12] for $\tilde{\varphi}(\mathbf{k}_1,\mathbf{k}_2)\in C_b^\infty$

$$\langle\varphi|W_+^*\Omega_+\psi\rangle$$
$$= \langle\varphi|W_+^*\psi\rangle = i\lim_{\epsilon\to 0}\int_0^\infty dt\,e^{-\epsilon t}\int d\,\mathbf{k}_1\,d\,\mathbf{k}_2\,\tilde{\varphi}^*(\mathbf{k}_1,\mathbf{k}_2)$$
$$\times\int d\,\xi_1\,d\,\xi_2\,\phi^{(+)*}_{\mathbf{k}_1,\mathbf{k}_2}(\xi_1,\xi_2)$$
$$\times(H_I\exp[-i(H_0 - \mathbf{k}_1^2/2m_1 - \mathbf{k}_2^2/2m_2)t\,]\psi)(\xi_1,\xi_2)\,. \tag{5.46}$$

For $\varphi\in C_b^\infty$, supp $\tilde{\varphi}\subset B$, $B$ compact, $B\cap S_V^6 = \phi$, $\psi\subset C_b^\infty$, we can interchange the order of integration in (5.46)[12] to obtain

$$\langle\varphi|W_+^*\Omega_+\psi\rangle$$
$$= \langle\varphi|W_+^*\psi\rangle + i\lim_{\epsilon\to 0}\int d\,\mathbf{k}_1\,d\,\mathbf{k}_2\,\tilde{\varphi}^*(\mathbf{k}_1,\mathbf{k}_2)$$
$$\times\int d\,\xi_1\,d\,\xi_2\,\phi^{(+)*}_{\mathbf{k}_1,\mathbf{k}_2}(\xi_1,\xi_2)\int_0^\infty dt\Big(H_I$$
$$\times\exp\Big[-i\Big(H_0 - \frac{\mathbf{k}_1^2}{2m_1} - \frac{\mathbf{k}_2^2}{2m_2} - i\epsilon\Big)t\,\Big]\psi\Big)(\xi_1,\xi_2)\,. \tag{5.47}$$

Consequently, using Theorem 3.1 in Chap. 4 of Ref. 1, we obtain[12]

$$\langle\varphi|W_+^*\Omega_+\psi\rangle = \langle\varphi|W_+^*\psi\rangle - \lim_{\epsilon\to 0}\int d\,\mathbf{k}_1\,d\,\mathbf{k}_2\,d\,\xi_1'\,d\,\xi_2'\,\tilde{\varphi}^*(\mathbf{k}_1,\mathbf{k}_2)\psi(\xi_1',\xi_2')$$
$$\times\text{l.i.m.}\int d\varphi_1\,d\varphi_2\,\phi^{(+)*}_{\mathbf{k}_1,\mathbf{k}_2}(\xi_1,\xi_2)V(\mathbf{q})\int d\,\mathbf{k}_1'\,d\,\mathbf{k}_2'\,\phi_{\mathbf{k}_1',\mathbf{k}_2'}(\xi_1,\xi_2)$$
$$\times\Big(\frac{\mathbf{k}_1^2}{2m_1} + \frac{\mathbf{k}_2^2}{2m_2} + i\epsilon - \frac{\mathbf{k}_1'^2}{2m_1} - \frac{\mathbf{k}_2'^2}{2m_2}\Big)^{-1}\phi^*_{\mathbf{k}_1',\mathbf{k}_2'}(\xi_1',\xi_2')$$

$$= \text{l.i.m.} \int d\mathbf{k}_1\, d\mathbf{k}_2\, d\zeta_1\, d\zeta_2\, \varphi^*(\mathbf{k}_1,\mathbf{k}_2)\psi(\zeta_1,\zeta_2)$$

$$\times e^{-i\mathbf{K}\cdot\mathbf{Q}}\left[\hat{\phi}^{(+)}_{\mathbf{K},\mathbf{k}}(\mathbf{q},\mathbf{p}_1,\mathbf{p}_2) - \int d\mathbf{q}'\, d\mathbf{p}_1'\, d\mathbf{p}_2'\, G^{(+)}_{0,\mathbf{K}}\left(\mathbf{q},\mathbf{p}_1,\mathbf{p}_2;\mathbf{q}',\mathbf{p}_1',\mathbf{p}_2';\frac{\mathbf{k}^2}{2m}\right)V(\mathbf{q}')\hat{\phi}^{(+)}_{\mathbf{K},\mathbf{k}}(\mathbf{q}',\mathbf{p}_1',\mathbf{p}_2')\right]^*$$

$$= \langle\varphi\,|\psi\rangle\,, \tag{5.48}$$

since $\hat{\phi}^{(+)}_{\mathbf{K},\mathbf{k}}$ satisfy the Lippmann–Schwinger equation. Consequently, since $C_b^\infty$ and $C_V^\infty$ are dense in $L^2(\Gamma_{12})$,

$$W_+^*\,\Omega_+ = I\,. \tag{5.49}$$

Therefore,

$$W_+ = W_+ W_+^*\,\Omega_+ = E^H(S_c^H)\Omega_+ = \Omega_+\,. \tag{5.50}$$

Q.E.D.

We can summarize the results of the last two sections in the following theorem.

**Theorem 5.8:** Suppose that the potential $V(\mathbf{q})$ is different from zero almost everywhere, that it is locally square integrable, of Rollnick class, and that it satisfies (5.6) for some $\epsilon > 0$. Then the initial domain $M_0$ of the Moller wave operators $\Omega_\pm$ is $L^2(\Gamma_{12})$, the final domains $R_\pm$ are both identical to $E^H(S_c^H)L^2(\Gamma_{12})$, and the $S$ operator is unitary on $M_0$. Furthermore, (4.7) is satisfied for the distorted phase-space waves $\phi^{(\pm)}_{\mathbf{k}_1,\mathbf{k}_2}$ [given explicitly by (5.8) for the outgoing case].

## VI. CONCLUSION

The standard nonrelativistic quantum scattering theory for potential interactions between point particles can be transferred, with no fundamental changes, to potential interaction between extended particles. However, for the former the configuration representation is quite satisfactory, in case of the latter stochastic phase space emerges as an essential tool for deriving many of the fundamental results.


## ACKNOWLEDGMENTS

The author wishes to thank Professor Eduard Prugovečki for his constant encouragement and his helpful advice in the course of writing Ref. 12.


## APPENDIX

In this appendix, we intend to evaluate (4.10) explicitly and show that it is given by (4.12). We have the following [since the proper wave functions are the optimal ones given by (2.19)]:

$$G_{0,\mathbf{K}}(\mathbf{q},\mathbf{p}_1,\mathbf{p}_2;\mathbf{q}',\mathbf{p}_1',\mathbf{p}_2';\zeta)$$

$$= -(2\pi)^{-3}\left(\frac{\pi}{l_1 l_2}\right)^{-3}\exp\left[-\frac{(l_1^2+l_2^2)}{2}(\mathbf{p}^2+\mathbf{p}'^2)\right.$$

$$+\frac{(l_2^2 m_1 - l_2^2 m_2)}{m_1+m_2}(\mathbf{p}\cdot(\mathbf{K}-\mathbf{P}) + \mathbf{p}'\cdot(\mathbf{K}-\mathbf{P}'))$$

$$\left.-\frac{(l_1^2 m_1^2 + l_2^2 m_2^2)}{2(m_1+m_2)^2}((\mathbf{K}-\mathbf{P})^2 + (\mathbf{K}-\mathbf{P}')^2)\right]$$

$$\times\int\left(\frac{\mathbf{k}^2}{2m}-\zeta\right)^{-1}\exp(-\alpha\mathbf{k}^2 + i\boldsymbol{\beta}\cdot\mathbf{k})d\mathbf{k}\,, \tag{A1}$$

$$\alpha = l_1^2 + l_2^2\,, \tag{A2}$$

$$\boldsymbol{\beta} = \mathbf{q} - \mathbf{q}' - i\left[(l_1^2 - l_2^2)(\mathbf{p}+\mathbf{p}')\right.$$

$$\left.-\frac{(l_1^2 m_1 + l_2^2 m_2)}{m_1+m_2}(2\mathbf{K}-\mathbf{P}-\mathbf{P}')\right]. \tag{A3}$$

We shall first evaluate the following integral:

$$\int d\mathbf{k}\left(\frac{\mathbf{k}^2}{2m}-\zeta\right)^{-1}\exp(-\alpha\mathbf{k}^2 + i\boldsymbol{\beta}\cdot\mathbf{k})$$

$$= -\frac{4\pi m}{i|\boldsymbol{\beta}|}\int_{-\infty}^{\infty} dr\,\frac{r}{r^2-2m\zeta}\exp(-\alpha r^2 - i|\boldsymbol{\beta}|r) \tag{A4}$$

for vectors $\boldsymbol{\beta}$ with real-valued components. Let us set

$$f(\beta) = \int_{-\infty}^{\infty}\frac{1}{r^2-2m\zeta}\exp(-\alpha r^2 - \beta r)dr\,. \tag{A5}$$

Then we obtain

$$f''(\beta) - 2m\zeta f(\beta) = \left(\frac{\pi}{\alpha}\right)^{1/2}\exp\left(\frac{\beta^2}{4\alpha}\right). \tag{A6}$$

The homogeneous differential equation associated with (9.16) has the solutions

$$f(\beta) = \exp(\pm\sqrt{2m\zeta}\,\beta)\,. \tag{A7}$$

Consequently, solving (A6) by variation of parameters, we set

$$f(\beta) = u_1(\beta)\exp(\sqrt{2m\zeta}\,\beta) + u_2(\beta)\exp(-\sqrt{2m\zeta}\,\beta)\,, \tag{A8}$$

$$0 = u_1'(\beta)\exp(\sqrt{2m\zeta}\,\beta) + u_2'(\beta)\exp(-\sqrt{2m\zeta}\,\beta)\,. \tag{A9}$$

Therefore, combining (A8), (A9), and (A6) we obtain

$$\sqrt{2m\zeta}\,(u_1'(\beta)\exp(\sqrt{2m\zeta}\,\beta) - u_2'(\beta)\exp(-\sqrt{2m\zeta}\,\beta))$$

$$= \sqrt{\pi/\alpha}\exp(\beta^2/4\alpha)\,. \tag{A10}$$

Solving the system consisting of (A9) and (A10), we get

$$u_1'(\beta) = \frac{1}{2}\sqrt{\frac{\pi}{2m\zeta\alpha}}\exp\left(\frac{\beta^2}{4\alpha} - \sqrt{2m\zeta}\,\beta\right), \tag{A11}$$

$$u_2'(\beta) = -\frac{1}{2}\sqrt{\frac{\pi}{2m\zeta\alpha}}\exp\left(\frac{\beta^2}{4\alpha} + \sqrt{2m\zeta}\,\beta\right), \tag{A12}$$

And hence inserting the solutions of (A11) and (A12) into (A8) we obtain

$$f(\beta) = \frac{\pi}{2i\sqrt{2m\zeta}}\exp(\sqrt{2m\zeta}\,\beta - 2m\zeta\alpha)$$

$$\times\operatorname{erf}\left(\frac{i}{2\sqrt{\alpha}}\beta - i\sqrt{2m\zeta\alpha}\right)$$

$$-\frac{\pi}{2i\sqrt{2m\zeta}}\exp(-\sqrt{2m\zeta}\,\beta - 2m\zeta\alpha)$$

$$\times\operatorname{erf}\left(\frac{i}{2\sqrt{\alpha}}\beta + i\sqrt{2m\zeta\alpha}\right)$$

$$+ k_1\exp(\sqrt{2m\zeta}\,\beta) + k_2\exp(-\sqrt{2m\zeta}\,\beta)\,, \tag{A13}$$

in terms of the well-known "error function"

$$\operatorname{erf} x = \frac{2x}{\sqrt{\pi}}\left[1 - \frac{x^2}{1!3} + \frac{x^4}{2!5} - \frac{x^6}{3!7} + \cdots\right]. \tag{A14}$$

Now, since $f'(0) = 0$, we must have that $k_1 = k_2$. To evaluate $k_1$, it is sufficient to evaluate

$$f(0) = \int_{-\infty}^{\infty} \frac{1}{r^2 - 2m\zeta} \exp(-\alpha r^2) dr. \tag{A15}$$

When regarded as a function $g$ of $\alpha$, the above satisfies the differential equation,

$$g'(\alpha) + 2m\zeta g(\alpha) = -\sqrt{\pi/\alpha} \tag{A16}$$

having for solution

$$g(\alpha) = \exp(-2m\zeta\alpha)\left[(i\pi/\sqrt{2m\zeta})\operatorname{erf}(i\sqrt{2m\zeta\alpha}) + \delta\right]. \tag{A17}$$

To evaluate the constant $\delta$ we have to evaluate (A15) for $\alpha = 0$ which lead by the residuum theorem to

$$\delta = \pi i/\sqrt{2m\zeta}. \tag{A18}$$

Therefore, combining (A13) with (A17) we obtain

$$k_1 = (i\pi/2\sqrt{2m\zeta})\exp(-2m\zeta\alpha). \tag{A19}$$

Therefore, since the right-hand side of (A4) is given, for $\beta$ with real valued components, by

$$(4\pi m/i|\beta|)f'(i|\beta|), \tag{A20}$$

then we obtain by analytic continuation the expression (4.12).


[1] E. Prugovecki, *Quantum Mechanics in Hilbert Spaces* (Academic, New York, 1981), 2nd ed.
[2] E. Prugovecki, *Stochastic Quantum Mechanics and Quantum Spacetime* (Reidel, Dordrecht, 1984).
[3] S. T. Ali and E. Prugovecki, J. Math. Phys. **18**, 219 (1977).
[4] E. Prugovecki, Phys. Rev. D **18**, 3655 (1978).
[5] F. E. Schroeck, Jr., J. Math. Phys. **22**, 2652 (1981).
[6] K. Menger, Proc. Nat. Acad. Sci. USA **28**, 535 (1942).
[7] A. Wald, Proc. Nat. Acad. Sci. USA **29**, 196 (1943).
[8] M. Born, Danske Vid. Selsk. Mat. Fys. Medd. **30**, (2) (1955).
[9] M. Born, *Physics in My Generation* (Pergamon, London, 1956).
[10] J. M. Levy-Leblond, in *Group Theory and Its Applications*, edited by E. M. Loeble (Academic, New York, 1971), Vol. II.
[11] A. Landé, Phys. Rev. **56**, 482 486 (1939).
[12] R. Gagnon, "Path Integrals and Scattering Theory in Nonrelativistic Stochastic Quantum Mechanics," Ph.D. thesis, University of Toronto, 1983.
[13] L. D. Faddeev, "Mathematical Aspects of the Three-Body Problem in the Quantum Scattering Theory," Israel program for scientific translation (Jerusalem, 1965).

# Schrödinger semigroups for vector fields

T. A. Osborn[a] and R. A. Corns[a]
*Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742*
Y. Fujiwara
*Department of Physics, University of Michigan, Ann Arbor, Michigan 48109*

(Received 8 May 1984; accepted for publication 17 August 1984)

Suppose $H$ is the Hamiltonian that generates time evolution in an $N$-body, spin-dependent, nonrelativistic quantum system. If $r$ is the total number of independent spin components and the particles move in three dimensions, then the Hamiltonian $H$ is an $r\times r$ matrix operator given by the sum of the negative Laplacian $-\Delta_x$ on the $(d=3N)$-dimensional Euclidean space $\mathbb{R}^d$ plus a Hermitian local matrix potential $W(x)$. Uniform higher-order asymptotic expansions are derived for the time-evolution kernel, the heat kernel, and the resolvent kernel. These expansions are, respectively, for short times, high temperatures, and high energies. Explicit formulas for the matrix-valued coefficient functions entering the asymptotic expansions are determined. All the asymptotic expansions are accompanied by bounds for their respective error terms. These results are obtained for the class of potentials defined as the Fourier image of bounded complex-valued matrix measures. This class is suitable for the $N$-body problem since interactions of this type do not necessarily decrease as $|x|\to\infty$. Furthermore, this Fourier image class also contains periodic, almost periodic, and continuous random potentials. The method employed is based upon a constructive series representation of the kernels that define the analytic semigroup $\{e^{-zH}|\mathrm{Re}\,z>0\}$. The asymptotic expansions obtained are valid for all finite coordinate space dimensions $d$ and all finite vector space dimensions $r$, and are uniform in $\mathbb{R}^d\times\mathbb{R}^d$. The order of expansion is solely a function of the smoothness properties of the local potential $W(x)$.

## I. INTRODUCTION AND SUMMARY

Take $H$ to be the self-adjoint semibounded operator that generates time evolution for the $N$-body problem in nonrelativistic quantum mechanics and let the complex variable $z$ take values in the open right half-plane $D$, then the family of bounded operators

$$\{e^{-zH}|z\in D\} \tag{1.1}$$

constitutes the analytic semigroup induced by $H$. The restriction of $z$ to the positive real axis leads to the one-parameter semigroup associated with the heat transport equation and the partition function of the canonical ensemble. In the heat transport problem the positive value of $z$ is the time variable, whereas in the partition function $z$ is proportional to the inverse temperature of the system. On the other hand, if $z$ belongs to the boundary $\partial D$ and takes on purely imaginary values then the family of operators (1.1) forms the one-parameter unitary group which describes time evolution of the system.

A second, equally basic, family of bounded operators are the resolvent operators

$$\{(H-\lambda)^{-1}|\lambda\in\mathbb{C},\ \mathrm{Im}\,\lambda\neq 0\} \tag{1.2}$$

that appear in the time-dependent formulation of quantum mechanics. This paper studies the uniform asymptotic expansions of the coordinate-space kernels of both operator families (1.1) and (1.2). For the analytic semigroup $\{e^{-zH}|z\in D\}$ the asymptotic expansion variable is $|z|\to 0$. Physically these expansions are applicable for short times or high temperatures. Asymptotic expansions for the resolvent kernels then result from a Laplace transform of the analytic semigroup kernels. The resolvent kernel expansions are valid for $|\lambda|\to\infty$, or convergent for high energies.

Recently the first of these two problems, (1.1), has been discussed at length for the case of scalar fields by Osborn and Fujiwara in Ref. 1 (hereafter OF). Let $x$ be the generic point in Euclidean space $\mathbb{R}^d$ that specifies the location of all $N$ particles in the system. If each individual particle moves in three dimensions then the Euclidean space dimension is $d=3N$. The scalar problem for local potentials is realized if the Hamiltonian $H$ is taken to be the self-adjoint extension in $L^2(\mathbb{R}^d)$ of the quadratic elliptic differential form

$$H_{(x)} = -q\Delta_x + v(x). \tag{1.3}$$

Here $\Delta_x$ denotes the Laplacian in $\mathbb{R}^d$ and $v$: $\mathbb{R}^d\to\mathbb{R}$ is the perturbing local potential. In terms of the rationalized value of Planck's constant $\hbar$, and the particle mass $m$, the quantum scale factor is

$$q=\hbar^2/2m. \tag{1.4}$$

The notationally simplifying device of setting $q=1$ is avoided because it is illuminating to exhibit explicitly the $q$ dependence of the heat-kernel and resolvent-kernel expansions and thereby see the semiclassical content of these expansions.

However, the general $N$-body problem is not described by scalar fields unless all the particles are bosons with spin zero. If the $i$th particle has spin $s_i$, the resulting $N$-body wave function is a vector-valued complex function of dimension

$$r=\sum_{i=1}^{N}(2s_i+1)$$

and an element of the space $L^2(\mathbb{R}^d,\mathbb{C}^r)$. In this circumstance the local perturbation becomes for each $x$ an Hermitian matrix $W(x)$: $\mathbb{C}^r\to\mathbb{C}^r$, and the scalar Laplacian $-\Delta_x$ generalizes to $-\Delta_x I$, where $I$ is the identity on $\mathbb{C}^r$. The matrix-valued analog of (1.3) assumes the form

[a]On leave from the University of Manitoba, Winnipeg R3T 2N2, Canada.

 0022-2488/85/030453-12$02.50 © 1985 American Institute of Physics 

$$H_{(x)} = -q\Delta_x I + W(x). \tag{1.5}$$

The appearance of (1.5) seems to assume that all $N$ particles have common mass $m$. This restriction is apparent rather than real since a scale transformation of the particle coordinates always allows one to write the most general diagonal kinetic energy operator in the form $-q\Delta_x$. The Hamiltonian operator $H$ which defines the physical quantum system is the self-adjoint extension of $H_{(x)}$ in $L^2(\mathbb{R}^d, \mathbb{C}^r)$.

Denote the kernel of the semigroup element $e^{-zH}$ by $U(x,y;z)$: $\mathbb{R}^d \times \mathbb{R}^d \times D \to \mathbb{C}^{r\times r}$. Similarly if $\rho(H)$ is the resolvent set of $H$, we define the kernel of the resolvent operator $(H-\lambda)^{-1}$ by $R(x,y;\lambda)$ for all $\lambda \in \rho(H)$. The basic objective of this paper is to derive the precise forms assumed by the natural asymptotic expansions of $U(x,y;z)$ and $R(x,y;\lambda)$. Specifically, we obtain the existence of the kernels, the analytic form (in $z$) of the asymptotic expansions, explicit closed expressions in terms of $W(x)$ of all the coefficient matrices that enter the asymptotic expansions, and $\mathbb{R}^d \times \mathbb{R}^d$ uniform bounds for the remainder terms. In fact we show that the derivation of these uniform asymptotic expansions requires only continuity and differentiability properties in $x$ of the potential $W(x)$. In particular, there is no necessity to assume that $W(x)$ decays as $|x| \to \infty$. The restriction we do impose on the allowed form of $W$ is that the potential be the Fourier image of a complex bounded $r\times r$ matrix measure, $\mu$, on $\mathbb{R}^d$.

The analytic semigroup (1.1) is characterized uniquely by its associated family of kernels $\{U(x,y;z) | z \in D\}$. Our approach to determining the existence and the properties of $U(x,y;z)$ is constructive. Let $H_0$ denote the free kinetic energy operator (the self-adjoint extension of $-q\Delta_x I$). We establish that the kernel analog of the Dyson series[2,3] for $e^{-zH}$ in terms of time-ordered parametric integrals of $e^{-zH_0}$ and $W$ leads to an absolutely and uniformly (in $\mathbb{R}^d \times \mathbb{R}^d$) convergent series representation of $U(x,y;z)$. As an immediate by-product of this result, it follows that if $W(x)$ has uniformly bounded derivatives of order 2, then $U(x,y;it/\hbar)$, where $t \in \mathbb{R}$ and represents time displacement, constitutes the fundamental solution of the Schrödinger equation

$$i\hbar \frac{\partial}{\partial t} U(x,y;it/\hbar) = H_{(x)} U(x,y;it/\hbar) \quad \forall x,y \in \mathbb{R}^d \tag{1.6}$$

obeying the delta-function initial condition

$$U(x,y;it/\hbar) \to \delta(x-y) I \quad \text{as } t \to 0. \tag{1.7}$$

Parallel conclusions apply to the heat-conduction problem. The heat-conduction or diffusion problem results if one formally replaces the imaginary time variable with $\beta > 0$. Again $U(x,y;\beta)$ is the fundamental solution of the heat problem. In particle physics terminology, $U(x,y;it/\hbar)$ is the propagator of the $N$-body system.

The center of interest in this investigation is the small $z$ uniform asymptotic expansion that $U(x,y;z)$ admits. It is well known that these kernels are highly singular in the limit $z \to 0$. This is evident from the formula for the free heat kernel defined by $H_0$:

$$U_0(x,y;z) = \left[\frac{1}{(4\pi z q)^{d/2}} \exp\left(\frac{-|x-y|^2}{4zq}\right)\right] I. \tag{1.8}$$

Here $|x-y|$ represents the Euclidean distance in $\mathbb{R}^d$ between $x$ and $y$. This convolution kernel has an essential singularity at $z=0$. Thus, asymptotic expansions of $U(x,y;z)$ for small $z$ require that this essential singularity structure be explicitly factored out in order to expose a smooth function of $z$. For this reason it is useful to define a function $F(x,y;z): \mathbb{R}^d \times \mathbb{R}^d \times D \to \mathbb{C}^{r\times r}$ by

$$U(x,y;z) = U_0(x,y;z) F(x,y;z). \tag{1.9}$$

Both $U_0$ and $F$ are $r\times r$ matrices. In (1.9) we adopt the hereafter standard convention that $U_0 F$ implies the matrix product. Now if the Dyson series for $U(x,y;z)$ is written in the appropriate form, one finds an explicit series expansion for $F$. The $F$-series is uniformly convergent on all compact subsets of $\overline{D}$ (the closure of $D$). As a consequence it follows that $F(x,y;z)$ is analytic in $D$, continuous in $\overline{D}$, and $F(x,y;0) = I$ for all $x,y \in \mathbb{R}^d$.

A further restructuring of the $F$-series leads to the asymptotic expansion

$$F(x,y;z) = \sum_{n=0}^{M} \frac{(-z)^n}{n!} P_n(x,y) + E_M(x,y;z). \tag{1.10}$$

Here, $M$ is an integer proportional to the number of continuous bounded partial derivatives that $W(x)$ supports. The error term is of order $O(|z|^{M+1})$ and has a uniform bound in $\mathbb{R}^d \times \mathbb{R}^d$. In addition, identity (1.10) can be differentiated with respect to $x$, $y$, or $z$ as often as desired and the resulting equation is also an asymptotic expansion provided $W(x)$ is sufficiently smooth. This flexible nature of (1.10) permits one to use it as the basis for calculating the small time behavior of correlation functions for an arbitrary pair of observables. The expansion (1.10) has been analyzed extensively in the literature[4-8] for a wide variety of operators $H$. Generally it is known that the coefficients $P_n$ [$P_0(x,y) = I$] are functions of $W(x)$ and its partial derivatives up to order $2(n-1)$. A novel feature of the constructive approach is that one can determine for every $n$ explicit expressions of the coefficient matrices $P_n(x,y)$. These expressions are not only applicable when $x = y$, but valid for all $x,y \in \mathbb{R}^d$. Another useful aspect of the constructive approach is that one can prove that expansion (1.10) is uniform in arg $z$. Thus the short time expansions are on exactly the same analytical footing as the high temperature expansions. In passing we note that the $x,y$ uniform character of expansion (1.10) is a necessary ingredient for the correct description of $N$-body systems that incorporate the wave function symmetrization required by either fermion or boson statistics.

The study of the resolvent kernel proceeds by using the Laplace transform of $U(x,y;z)$ to determine $R(x,y;z)$, namely,

$$R(x,y;z) = \int_0^\infty e^{\beta z} U(x,y;\beta)\, d\beta, \quad x \neq y, \quad \operatorname{Re} z < c, \tag{1.11}$$

where $c$ is any negative lower bound for the spectrum of $H$. The condition $\operatorname{Re} z < c < 0$ ensures that the integral in (1.11) is absolutely convergent. The resolvent kernel is holomorphic in a much larger domain of $z$, namely $z \in \rho(H)$. Identity (1.11) may be analytically continued to a subset of this larger domain in $\mathbb{C}$ by changing the variable of integration so that the integration contour along the positive real axis is rotated until it becomes a complex ray with origin at $\beta = 0$ and having constant $\arg \beta \in (\pi/2, -\pi/2)$. We find, upon using the analytically extended form of (1.11), that the Laplace image

of the heat-kernel asymptotic expansion (1.10) becomes

$$R(x,y;z) = \sum_{n=0}^{M} \frac{(-1)^n}{n!} P_n(x,y) \left(\frac{\partial}{\partial z}\right)^n R_0(x,y;z) + T_M(x,y;z), \quad x \neq y. \tag{1.12}$$

Here, $R_0(x,y;z)$ denotes the kernel defined by $(H_0 - z)^{-1}$. This free resolvent is an analytic function of $z$ in the open cut plane $\mathbb{C}\backslash\mathbb{R}^+$ and for $H_0 = -q\Delta_x I$ it is given by a Bessel function multiplied by $I$.

Although it is not so apparent from the form of (1.12), the small parameter in the expansion is $z^{-1}$. Furthermore if $M + 2 > d/2$ then the error term in (1.12) has a uniform bound in $\mathbb{R}^d \times \mathbb{R}^d$ and is of the order $(|z|^{-1})^{M+2-d/2}$. The boundedness of the error $T_M$ as $x \to y$ implies that the singularities of $R(x,y;z)$, $R_0(x,y;z)$, and the derivatives of $R_0(x,y;z)$ in the neighborhood of the diagonal $x = y$ are identical on both sides of (1.12). The uniformity in $z$ of (1.12) can be characterized as follows. For any $\delta \in (0,\pi/2)$ let $\tilde{V}_\delta$ denote the subset of $\mathbb{C}$ given by

$$\{z | z \in \mathbb{C}, \quad \arg(z + \|\mu\|(\sin\delta)^{-1}) \in (2\delta, 2\pi - 2\delta)\}.$$

Clearly $\tilde{V}_\delta$ is the complement of a wedge symmetric about the positive real axis with its apex at $z = -\|\mu\|(\sin\delta)^{-1}$. The expansion (1.12) is uniform for all $z \in \tilde{V}_\delta$. Although the opening angle of the wedge, $4\delta$, is arbitrary and can be made as small as desired, there will always be a strip about the positive real axis disjoint from $\tilde{V}_\delta$. Thus the values of $R(x,y;z)$ as $z$ converges to points on the spectrum, $\sigma(H)$, are not estimated by (1.12). This is to be expected since stating only the smoothness properties of $W(x)$ is insufficient information to determine the detailed nature of the spectrum. In examples where formula (1.12) has been continued to the spectral boundary (such as Buslaev's treatment[9] of $\mathbb{R}^3$), extensive information about the $|x| \to \infty$ decay of $W(x)$ is required in order to extend the $z$ domain of (1.12) to the boundary of the open domain $\mathbb{C}\backslash\sigma(H)$.

A balanced overview of the results outlined above emerges if we consider their connection to the spectral asymptotics of $H$. Local geometrical spectral asymptotics is the study of the relationships that link three basic structures generated by the differential operator $H_{(x)}$. These structures are (a) the local coefficient functions that define $H_{(x)}$ and the boundary conditions which are obeyed; (b) the Weyl expansion[10,11] predicting the density of eigenvalues $\lambda_i$ satisfying

$$H\Psi_{\lambda_i} = \lambda_i \Psi_{\lambda_i}, \quad \|\Psi_{\lambda_i}\| = 1, \tag{1.13}$$

as $\lambda_i \to \infty$, or the appropriate generalization of the Weyl expansion when the spectrum has a continuous component; and (c) the asymptotic expansions of the integral kernels for the basic operator functions of $H$ such as the semigroup or resolvent operators.

In general, the coefficient functions that enter $H_{(x)}$ come from the functions that define the Laplace–Beltrami operator for a non-Euclidean manifold (either compact or noncompact), the functions appearing in the first-order terms (such as the magnetic vector potential), and the potential $W(x)$. In our $N$-body problem the manifold is flat so the Laplace–Beltrami operator reduces to the Laplacian, the first-order derivative terms are absent, and so the only nontrivial coefficient function is $W(x)$. The boundary condition for $\mathbb{R}^d$ reduces to the requirement that $H_{(x)}$ have a unique self-adjoint extension in $L^2(\mathbb{R}^d, \mathbb{C}^r)$. If the manifold supporting the functions on which $H_{(x)}$ acts is noncompact (as in the $\mathbb{R}^d$ case), characteristically one finds that the spectrum has a continuous part. The appropriate extension of the Weyl problem is to replace the study of the density of eigenvalues by the asymptotic expansion of the spectral kernels $e(x,y;\lambda)$ as $\lambda \to \infty$.

In essence the asymptotic expansions (1.10) for the heat kernel and (1.12) for the resolvent kernel display the relationship between (a) and (c). In particular, the explicit expressions we derive later for $P_n(x,y)$ show how the local form of $H_{(x)}$ controls these two expansions. In this sense this paper is a special application of the local geometrical asymptotics program. The approach emphasized in this work is to determine first (via the Dyson expansion) the detailed properties of the semigroup family of kernels and then by various transforms obtain the other asymptotic expansions of interest.

It is worth observing that the connection (a)→(b), although of fundamental importance, is still not well understood in the noncompact domain problems. Following the technique introduced by Carleman,[12] most investigations of the large $\lambda$ behavior of $e(x,y;\lambda)$ have utilized the Tauberian theorems.[13–16] This approach is capable of predicting only the leading-order behavior of $e(x,y;\lambda)$. An alternate method of investigating the continuum–Weyl problem and obtaining a higher-order asymptotic expansion was developed by Osborn and Wong[17] (hereafter OW). The technique of OW is to obtain the link (a)→(b) by the chain of results (a)→(c) then (c)→(b). In particular, one may prove that the kernel $U(x,y;z)$ has the spectral representation[17,18]

$$U(x,y;z) = \int_{\sigma(H)} e^{-z\lambda}\, de(x,y;\lambda), \quad z \in D. \tag{1.14}$$

In order to implement the stage (c)→(b) the inverse of this transform is required, namely,

$$e(x,y;\lambda) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{e^{z\lambda}}{z} U(x,y;z)dz, \quad c > 0. \tag{1.15}$$

The validity of (1.15) is established in OW. This formula has the feature of transforming semigroup kernels into spectral kernels.

Under the hypothesis that the continuous spectrum contribution to $U(x,y;it/\hbar)$ has nice decay properties as $t \to \pm\infty$ (this assumption just reflects the physically reasonable expectation that particles and stable clusters of particles that belong to the continuum of $H$ will diffuse as $t \to \pm\infty$) then $e(x,y;\lambda)$ admits a higher-order Weyl expansion given by

$$e(x,y;\lambda) = \sum_{n=0}^{M} \frac{(-1)^n}{n!} P_n(x,y) \left(\frac{\partial}{\partial\lambda}\right)^n e_0(x,y;\lambda) + \tilde{E}_M(x,y;\lambda). \tag{1.16}$$

Here $e_0(x,y;\lambda)$ is the spectral kernel of Hamiltonian $H_0$ and is a known analytic function of $\lambda$. This expansion is uniform within compact regions of $\mathbb{R}^d \times \mathbb{R}^d$. The error term is of order $O(|\lambda|^{-N})$, where $N$ depends in a complicated way on the number of bounded derivatives of $W(x)$ and on the nature of the $t$-decay of the continuous spectrum contributions to $U(x,y;it/\hbar)$. A more detailed overview of local geometrical

spectral asymptotics may be found in the excellent review of Fulling.[4]

The construction of this paper is as follows. In Sec. II the complex matrix-valued measure representation of the potential $W(x)$ is introduced. Section III describes the constructive series representation of the kernel $U(x,y;z)$ and obtains the $x,y$-uniform asymptotic expansion associated with Eq. (1.10). Furthermore, under appropriate smoothness restrictions on $W(x)$, it is established that $U(x,y;it/\hbar)$ and $U(x,y;\beta)$ are fundamental solutions of their respective partial differential equations. The explicit formulas for the coefficient matrices $P_n(x,y)$ are determined. Finally in Sec. IV we utilize an analytic continuation of the Laplace transform to find the large $|\lambda|$ expansion of the resolvent kernels $R(x,y;\lambda)$. Formulas bounding the total error in the resolvent asymptotic expansion are derived.

## II. FOURIER IMAGE POTENTIALS

For a particular class of potentials $W(x)$, the operator $H_{(x)}$ is studied. Let $\mathcal{M}(\mathbb{R}^d,\mathbb{C}^{r\times r})$ be the set of all bounded complex matrix-valued measures defined on the Borel field $\mathcal{B}$ on $\mathbb{R}^d$. Each measure $\mu\in\mathcal{M}$, defines a matrix-valued potential function by the Fourier transform of $\mu$,

$$W(x)=\int_{\mathbb{R}^d} e^{ikx}\,d\mu(k), \tag{2.1}$$

where $k\in\mathbb{R}^d$ and $kx$ denotes the scalar product in $\mathbb{R}^d$. The equality above is understood as that appropriate for the space of complex matrices, $\mathbb{C}^{r\times r}$. If $\nu,\gamma=1,2,...,r$ specify the row and column of a matrix, then (2.1) implies

$$W_{\nu\gamma}(x)=\int_{\mathbb{R}^d} e^{ikx}\,d\mu_{\nu\gamma}(k)\quad \text{all } \nu,\gamma. \tag{2.2}$$

Each $\mu_{\nu\gamma}$ is a bounded complex-valued measure on $\mathcal{B}$ and each $\nu\gamma$ component of the matrix $W$ is a complex-valued function of $x$. Hereafter, in order to simplify our notation the integration domain $\mathbb{R}^d$ will be omitted.

We employ the symbol $|\cdot|$ to represent several different norms. If the argument of $|\cdot|$ is in $\mathbb{C}$, then $|\cdot|$ denotes the absolute value; if the argument is in$\mathbb{C}^r$ or $\mathbb{C}^{r\times r}$ then the norm is the appropriate Euclidean vector length. For example, if $E$ is any set in $\mathcal{B}$,

$$|\mu(E)|=\left\{\sum_{\nu=1}^{r}\sum_{\gamma=1}^{r}|\mu_{\nu\gamma}(E)|^2\right\}^{1/2}. \tag{2.3}$$

A somewhat different meaning of the absolute value sign applies to $|\mu|$. Here, $|\mu|$ is defined to be the total variation of $\mu$, namely the non-negative scalar-valued set function on $\mathcal{B}$ given by

$$|\mu|(E)=\sup\sum_{i=1}^{\infty}|\mu(E_i)|, \tag{2.4}$$

where the supremum is taken over all partitions $\{E_i\}$ of $E$. The statement that $\mu\in\mathcal{M}$ is bounded means that $|\mu|(\mathbb{R}^d)<\infty$. In fact, the total variation $|\mu|$ may be used to define a norm $\|\cdot\|$ for $\mathcal{M}$ if we set

$$\|\mu\|=|\mu|(\mathbb{R}^d). \tag{2.5}$$

Equipped with this norm $\mathcal{M}$ is a Banach space.[19] It is clear from the definition of (2.1) that the Fourier image of $\mathcal{M}$ consists of $\mathbb{R}^d\to\mathbb{C}^{r\times r}$ functions that are uniformly bounded and uniformly continuous. In particular,

$$|W(x)|\leqslant\|\mu\|,\quad \forall x\in\mathbb{R}^d. \tag{2.6}$$

The transformation (2.1) defines a class of potentials

$$\mathcal{F}=\left\{W(x)=\int e^{ikx}\,d\mu(k)\,|\,\mu\in\mathcal{M}\right\}. \tag{2.7}$$

The elements of the spaces $\mathcal{F}$ and $\mathcal{M}$ are in a one-to-one correspondence. This is a consequence of the uniqueness[20] of the transformation (2.1) that states $W(x)=0$ if and only if $\mu=0$. Again by adjoining norm $\|\mu\|$ to $\mathcal{F}$ one defines a Banach space.

Consider the subset of potentials in $\mathcal{F}$ that are Hermitian matrices for all $x\in\mathbb{R}^d$. For a set $E\in\mathcal{B}$ the reflected set $-E$ is defined to be $-E=\{k\,|\,k\in\mathbb{R}^d,-k\in E\}$. We say the measure $\mu$ satisfies the reflection property if

$$\mu(-E)=\mu^*(E),\quad \forall E\in\mathcal{B}, \tag{2.8}$$

where $^*$ denotes the adjoint on $\mathbb{C}^{r\times r}$. Then the Fourier transform of a $\mu$ satisfying the reflection property is a Hermitian matrix for all $x$. And, conversely an element $W\in\mathcal{F}$ that is Hermitian for all $x$ has an associated measure $\mu\in\mathcal{M}$ that obeys the reflection property. Define the subset of $\mathcal{M}$ that consists of measures $\mu$ of the type (2.8) as $\mathcal{M}^*$ and let $\mathcal{F}^*$ be the Fourier image of $\mathcal{M}^*$. The potentials $W\in\mathcal{F}^*$ are the physically significant ones since they comprise all the Hermitian potentials in $\mathcal{F}$.

The asymptotic expansions derived in the next several sections are a manifestation of the smoothness of the potentials $W(x)$. For this reason it is convenient to further classify the potentials in $\mathcal{F}^*$ into subclasses in which derivatives up to order $M$ are bounded. We define $\mathcal{F}^*_M\subseteq\mathcal{F}^*$, $M=0,1,2,...$, to be

$$\mathcal{F}^*_M=\left\{W\,|\,W\in\mathcal{F}^*,\int d\,|\mu|(k)|k|^n<\infty,\ n=0,1,...,M\right\}. \tag{2.9}$$

In fact, for $W\in\mathcal{F}^*_M$ there exists a smallest finite positive constant $K(W,M)$ such that

$$\int d\,|\mu|(k)|k|^n\leqslant K(W,M)^n\|\mu\|,\quad n=0,1,...,M. \tag{2.10}$$

We call $K(W,M)$ the bound constant of potential $W$ in the space $\mathcal{F}^*_M$.

If $D_x^L$ represents an arbitrary partial derivative in $\mathbb{R}^d$, multi-indexed by $L=(l_1,l_2,...,l_d)$, $l_i\geqslant 0$, with length $|L|=l_1+l_2+\cdots+l_d$, and given by

$$D_x^L=\left(\frac{\partial}{\partial x_1}\right)^{l_1}\cdots\left(\frac{\partial}{\partial x_d}\right)^{l_d}, \tag{2.11}$$

then $W\in\mathcal{F}^*_M$ implies

$$|(D_x^L W)(x)|\leqslant K(W,M)^{|L|}\|\mu\|,\quad |L|\leqslant M. \tag{2.12}$$

It is useful to decompose the matrix measure $\mu$ into the product of the total variation measure $|\mu|$ and a matrix function with unit Euclidean norm.

*Lemma 1:* Let $\mu\in\mathcal{M}$, then there is a unique $|\mu|$-measurable matrix function $A:\mathbb{R}^d\to\mathbb{C}^{r\times r}$ such that

$$|A(k)|=1,\quad \forall k\in\mathbb{R}^d, \tag{2.13}$$

and

$$\int_E d\mu(k) = \int_E d|\mu|(k) A(k), \quad \forall E \in \mathscr{B}. \tag{2.14}$$

*Proof:* Since $\mu \ll |\mu|$ the Radon–Nikodym theorem implies the existence of a $|\mu|$-measurable $L^1(|\mu|)$ function $A(k)$. The proof that $A(k)$ has Euclidean norm equal to unity follows from a simple modification of the argument Rudin[21] (Theorem 6.12) gives for the scalar case. □

For scalar problem ($r = 1$) the potential class $\mathscr{F}$ was introduced by Ito[22] to study the Feynman path integral representations of $e^{itH/\hbar}$. Later, Albeverio and Høegh-Krohn[23] used $\mathscr{F}$ for the same purpose and in a fashion similar to our treatment of the kernel form of the Dyson series.

## III. ASYMPTOTIC EXPANSIONS FOR $e^{-zH}$

Throughout the remainder of the paper it is always assumed that $W \in \mathscr{F}^*$. For this class of potentials the Hamiltonian $H$ is defined as the self-adjoint extension of $H_{(x)}$ in $L^2(\mathbb{R}^d, \mathbb{C}^r)$. We also let the symbol $W$ stand for the linear operator on $L^2(\mathbb{R}^d, \mathbb{C}^r)$ given by multiplication in $\mathbb{C}^r$ with the potential function $W(x)$. Inequality (2.6) implies that $W$ has the operator norm bound $\|W\| \leqslant \|\mu\|$; this in turn means that $H$ is semibounded with lower bound $H \geqslant -\|\mu\| I$. Because $W$ is bounded, the unbounded operators $H$ and $H_0$ have common domains $\mathscr{D}(H) = \mathscr{D}(H_0) \subset L^2(\mathbb{R}^d, \mathbb{C}^r)$. Take $\Lambda = \sigma(H)$ to be the spectrum of $H$ and $\{E_\lambda \mid \lambda \in \Lambda\}$ to be the unique family of spectral projectors generated by $H$. The semiboundedness estimate above tells us that $\Lambda \subseteq [-\|\mu\|, \infty)$. The analytic semigroup operators are then defined by their spectral integrals

$$e^{-zH} = \int_\Lambda e^{-z\lambda} dE_\lambda, \quad z \in D. \tag{3.1}$$

Restricting $z = \beta > 0$ gives us the heat operator $e^{-\beta H}$. Replacing $H$ by $H_0$ in (3.1) determines the free heat operator $e^{-\beta H_0}$.

Before proceeding with the derivations we introduce a number of the notational conventions that will be employed. The Hilbert space on which the semigroup operators act is $\mathscr{H} = L^2(\mathbb{R}^d, \mathbb{C}^r)$. On this space $\langle \cdot, \cdot \rangle$ represents the inner product and $\|f\| = \langle f, f \rangle^{1/2}$ the associated norm. For example, if $f, g \in \mathscr{H}$ then these functions have $r$ components [i.e., $f = (f_1, f_2, \ldots, f_r)$, where $f_i \in L^2(\mathbb{R}^d)$ and similarly for $g$] and the $\mathscr{H}$-inner product is

$$\langle f, g \rangle = \sum_{i=1}^{r} (f_i, g_i), \tag{3.2}$$

where $(\cdot, \cdot)$ is the inner product on $L^2(\mathbb{R}^d)$. The general $L^p$ norm for $f \in L^p(\mathbb{R}^d, \mathbb{C}^r)$ will be indicated by $\|f\|_p$. The $n$th-order iterated parametric integrals which enter the Dyson expansion will be abbreviated by

$$\int_>^1 d^n\xi \equiv \int_0^1 d\xi_1 \int_0^{\xi_1} d\xi_2 \cdots \int_0^{\xi_{n-1}} d\xi_n \tag{3.3}$$

and the iterated $\mu$ measure integrals

$$\int d^n\mu \equiv \int d\mu(k_n) \int d\mu(k_{n-1}) \cdots \int d\mu(k_1). \tag{3.4}$$

Note that for differing values of $x_1$ and $x_2$, generally

$$W(x_1)W(x_2) \neq W(x_2)W(x_1), \tag{3.5}$$

or equivalently $d\mu(k_1)d\mu(k_2) \neq d\mu(k_2)d\mu(k_1)$. This noncommutivity of the potentials $W(x_1), W(x_2)$ is the most significant structural distinction with the scalar problem. Our choice of the Euclidean norm (2.3) for the matrix measures $\mu$ is made on the basis that this norm definition for $\mu$ is the one that makes the bound estimates for the vector problem closely parallel to those that enter the scalar case. As a result, we are able to adopt many of the proofs for the scalar problem with only minor alterations.

Several simple mathematical functions occur repeatedly in our analysis so it is convenient to introduce an abbreviated notion for them. For $i = 1, \ldots, n$, let $\xi_i \in [0,1]$, set

$$\Theta(\xi_l, \xi_m) = \min\{\xi_l(1 - \xi_m), \xi_m(1 - \xi_l)\}. \tag{3.6}$$

For $k_i \in \mathbb{R}^d$ define the polynomials in $k_i$ by

$$a_n(\xi_1, \ldots, \xi_n; k_1, \ldots, k_n) = \sum_{l,m=1}^{n} \Theta(\xi_l, \xi_m) k_l k_m, \tag{3.7}$$

$$b_n(\xi_1, \ldots, \xi_n; k_1, \ldots, k_n) = \sum_{l=1}^{n} [(1 - \xi_l)x + \xi_l y] k_l, \tag{3.8}$$

where $k_l k_m$ denotes the scalar product in $\mathbb{R}^d$. Furthermore we denote the scalar free diffusion kernel by

$$h(x; z) = (4\pi z q)^{-d/2} \exp\{-|x|^2/4zq\}. \tag{3.9}$$

The Dyson series for $e^{-\beta H}$ is obtained by iterating the identity[24]

$$e^{-\beta H} = e^{-\beta H_0} - \int_0^\beta d\beta_1 \, e^{-\beta_1 H} V e^{-(\beta - \beta_1)H_0}. \tag{3.10}$$

The constructive series representation for $U(x, y; z)$ given in Lemma 2 and Theorem 1 results from the analysis of the $x, y$ kernel analog of the Dyson series (3.10) followed by an analytic extension from the positive real axis $\beta \in (0, \infty)$ to $z \in D$. Except for some minor technical details in handling the vector norm on $\mathbb{C}^r$, Lemma 1 permits one to adopt in an obvious way the proofs given in Ref. 1. We have the following results.

*Definition 1:* Let $W \in \mathscr{F}^*$. For each $z \in \overline{D}$, let $F(\cdot, \cdot; z): \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}^{r \times r}$ be the function defined by the series

$$F(x, y; z) = \sum_{n=0}^{\infty} B_n(x, y; z), \tag{3.11}$$

where $B_0(x, y; z) = I$ and

$$B_n(x, y; z) = (-z)^n \int_>^1 d^n\xi \int d^n\mu \, e^{-zqa_n + ib_n}. \tag{3.12}$$

*Lemma 2:* Let $W \in \mathscr{F}^*$. The function $F(x, y; z)$ has the following properties.

(i) Let $D_c$ be any compact subset of $\overline{D}$. The series (3.11) is absolutely and uniformly convergent for all $(x, y; z) \in \mathbb{R}^d \times \mathbb{R}^d \times D_c$. Furthermore, $F_{\nu\gamma}(x, y; z)$ has the bound

$$|F_{\nu\gamma}(x, y; z)| \leqslant e^{|z| \|\mu\|}. \tag{3.13}$$

(ii) $F(x, y; z)$ is a $\mathbb{C}^{r \times r}$ valued holomorphic function in $D$ and continuous in $\overline{D}$. It is jointly continuous in $\mathbb{R}^d \times \mathbb{R}^d$.

In order to proceed further, let us define the free analytic semigroup kernel by

$$U_0(x, y; z) = h(x - y; z) I, \quad z \in \overline{D} \setminus \{0\}, \tag{3.14}$$

where $h$ is (3.9). With this notation we have the following theorem.

**Theorem 1:** Let $W \in \mathcal{F}^*$. Define the function $U: \mathbf{R}^d \times \mathbf{R}^d \times (\overline{D} \setminus \{0\}) \to \mathbf{C}^{r \times r}$ by

$$U(x,y;z) = U_0(x,y;z) F(x,y;z). \tag{3.15}$$

(i) For all $z \in D$ and all $f \in \mathcal{H}$,

$$(e^{-zH} f)(x) = \int dy\, U(x,y;z) f(y), \quad \text{a.a. } x \in \mathbf{R}^d. \tag{3.16}$$

(ii) Suppose $t \in \mathbf{R}$, $t \neq 0$, and $f \in L^1(\mathbf{R}^d, \mathbf{C}^r) \cap \mathcal{H}$,

$$(e^{-itH} f)(x) = \int dy U(x,y;it) f(y), \quad \text{a.a. } x \in \mathbf{R}^d. \tag{3.17}$$

If $f \in \mathcal{H}$ then

$$(e^{-itH} f)(x) = \text{s-lim}_{Y \to \infty} \int_{|y| < Y} dy\, U(x,y;it) f(y). \tag{3.18}$$

The form of Theorem 1, in particular part (ii), for time-evolution kernels is the best result that can be expected since if $H$ is replaced by $H_0$ then (ii) is the standard[25] representation of the free time evolution kernel. Related results on the existence of time evolution kernels have been obtained by Kitada,[26] Kitada and Kumanogo,[27] Fujiwara,[28] and Zelditch.[29]

Turn now to the problem of demonstrating that $U(x,y;it/\hbar)$ and $U(x,y;\beta)$ are fundamental solutions of their respective Schrödinger and heat partial differential equations.

*Proposition 1:* Let $W \in \mathcal{F}_2^*$, the function $U(x,y;it/\hbar)$ is the fundamental solution of the Schrödinger equation

$$i\hbar \frac{\partial}{\partial t} U(x,y;it/\hbar) = H_{(x)} U(x,y;it/\hbar), \quad t \neq 0, \tag{3.19}$$

that satisfies the delta-function initial condition

$$U(x,y;it/\hbar) \to \delta(x-y) I, \quad \text{as } t \to 0. \tag{3.20}$$

*Proposition 2:* Let $W \in \mathcal{F}_2^*$, the function $U(x,y;\beta)$ is the fundamental solution of the heat equation

$$-\frac{\partial}{\partial \beta} U(x,y;\beta) = H_{(x)} U(x,y;\beta), \quad \beta > 0, \tag{3.21}$$

that satisfies the delta-function initial condition

$$U(x,y;\beta) \to \delta(x-y) I, \quad \text{as } \beta \to 0. \tag{3.22}$$

*Proof:* Propositions 1 and 2 have similar proofs. We shall write out the proof of Proposition 1. First substitute expression (3.15) into Eq. (3.19). So $U(x,y;it/\hbar)$ is a solution of (3.19) if and only if $F(x,y;it/\hbar)$ is a solution of

$$i\hbar \frac{\partial}{\partial t} F(x,y;it/\hbar) = \left\{ H_{(x)} - \frac{i\hbar}{t} (x-y) \nabla_x I \right\} F(x,y;it/\hbar). \tag{3.23}$$

The function $F(x,y;it/\hbar)$ as defined by series (3.11) converges uniformly in $\mathbf{R}^d \times \mathbf{R}^d$ to $I$ as $t \to 0$. Furthermore it is well known that $U_0(x,y;it/\hbar)$ is a delta function in the limit $t \to 0$. Thus $U(x,y;it/\hbar)$ obeys the initial condition (3.20). So it suffices to prove that $F(x,y;it/\hbar)$ satisfies (3.23) for all $t$.

To proceed further, consider the terms $B_n(x,y;it/\hbar)$ in the series expansion (3.11) for $F$. Make the change of variable $\xi_i = (1 - t_i/t)$, $i = 1,\ldots,n$ for the integral expression of $B_n$. One obtains

$$B_n(x,y;it/\hbar) = \left(-\frac{i}{\hbar}\right)^n \int_0^t dt_n \int_0^{t_n} dt_{n-1} \cdots \int_0^{t_2} dt_1 \times \int d^n\mu \exp\{S_n\}, \tag{3.24}$$

where the exponential argument is

$$S_n(x,y;t_1 \ldots t_n;k_1 \ldots k_n) = -iq \sum_{l,m=1}^{n} \left( t_l \wedge t_m - \frac{t_l t_m}{t} \right) k_l k_m + i \sum_{l=1}^{n} \left[ \frac{t_l}{t} x + \left(1 - \frac{t_l}{t}\right) y \right] k_l. \tag{3.25}$$

Here the notation $t_l \wedge t_m$ is

$$t_l \wedge t_m = \text{sgn}(t)\,\text{Min}\{|t_l|, |t_m|\}. \tag{3.26}$$

Note that $S_n$ has the algebraic property

$$S_n(x,y;t_1 \ldots t_n;k_1 \ldots k_n)|_{t_n = t} = S_{n-1}(x,y;t_1 \ldots t_{n-1};k_1 \ldots k_{n-1}) + ixk_n. \tag{3.27}$$

Now assume $W \in \mathcal{F}_2^*$. Using formula (3.24) together with identity (3.27) gives one the recursion relation for $n \geq 1$,

$$i\hbar \frac{\partial}{\partial t} B_n\left(x,y;\frac{it}{\hbar}\right) = [-q\Delta x - (i\hbar/t)(x-y)\nabla_x] B_n(x,y;it/\hbar) + W(x) B_{n-1}(x,y;it/\hbar). \tag{3.28}$$

The assumption $W \in \mathcal{F}_2^*$ is needed in order to justify passing the derivative operators through the multiple integral in (3.24). The last step is to sum (3.28) from $n = 1$ to $\infty$. If $W \in \mathcal{F}_2^*$, all the infinite sums are absolutely and uniformly convergent in $\mathbf{R}^d \times \mathbf{R}^d$. In addition, the $n$-summation may be interchanged with all the differential operators appearing in (3.28). Thus (3.23) is satisfied for all $x,y,t$ by $F(x,y;it/\hbar)$ defined through series (3.11). □

The solution of (3.19) and (3.20) is appropriately termed fundamental. All other forms of solution of the time-dependent Schrödinger equation are implied by Proposition 1. For example, one immediately obtains the following statement of the Cauchy-data problem.

*Corollary 1:* Let $W \in \mathcal{F}_2^*$. Suppose $f$ is any wave packet (element) in $L^1(\mathbf{R}^d, \mathbf{C}^r) \cap \mathcal{H}$. Then

$$\psi(x,t) = \int dy U(x,y;it/\hbar) f(y) \tag{3.29}$$

is a solution of

$$i\hbar \frac{\partial}{\partial t} \psi(x,t) = H_{(x)} \psi(x,t), \tag{3.30}$$

with the Cauchy initial data

$$\psi(x,0) = f(x). \tag{3.31}$$

*Proof:* This follows from an obvious modification of the argument demonstrating Proposition 1. □

The remainder of this section implements a reordering of the series expansion for $F(x,y;z)$ and obtains the higher-order asymptotic expansion (1.10). It is proved that the number of terms $M$ in the expansion is solely a function of the order of the bounded partial derivatives that $W(x)$ supports. Explicit formulas for the diagonal and nondiagonal values of $P_n(x,y)$ for all $n$ are found. The remainder term $E_M(x,y;z)$ is given a bound that is uniform in both $\mathbf{R}^d \times \mathbf{R}^d$ and $\arg z \in [\pi/2, -\pi/2]$. It is observed that the asymptotic expansion (1.10) may be freely differentiated with respect to all the variables of $F(x,y;z)$.

*Proposition 3:* Let $W \in \mathscr{F}^{*}_{2(N+1)}$ and let $K$ be the corresponding bound constant of $W$ in the family $\mathscr{F}^{*}_{2(N+1)}$. For all $z \in \overline{D}$ and $n \geq 1$,

$$B_n(x,y;z) = \sum_{m=0}^{N} \frac{(-z)^{n+m}}{(n+m)!} q^m D_{m,n+m}(x,y) + E_{n,N}(x,y;z). \tag{3.32}$$

The coefficient functions $D_{m,n+m}(x,y)$ are jointly and uniformly continuous in $\mathbf{R}^d \times \mathbf{R}^d$ and are represented by the parametric integrals,

$$D_{m,m+n}(x,y) = n! \binom{m+n}{m} \int_{>}^{1} d^n\xi \int d^n\mu(a_n)^m e^{ib_n}, \quad n>0, \quad m \geq 0, \tag{3.33}$$

where the factor in front of the multiple integral is the binomial coefficient. Furthermore, $D_{0,0}(x,y) = I$, and for $m \geq 1$, $D_{m,m}(x,y) = 0$. The coefficient matrices and remainder term have estimates

$$|D_{m,m+n}(x,y)| \leq \binom{m+n}{m} \|\mu\|^n \left(\frac{nK}{2}\right)^{2m}, \tag{3.34}$$

$$|E_{n,N}(x,y;z)| \leq \frac{(|z|\,\|\mu\|)^n}{n!(N+1)!} \left(\frac{|z|qn^2K^2}{4}\right)^{N+1}, \tag{3.35}$$

for $m \leq N$ and $n \geq 0$.

*Proof:* The argument for Proposition 5, Ref. 1, is easily modified to accommodate the matrix nature of (3.32). □

The asymptotic expansion of the analytic semigroup kernel $U(x,y;z)$ follows from Proposition 3 and Lemma 2 for $F(x,y;z)$. Suppose $W \in \mathscr{F}^{*}_{2(M+1)}$ then $F$-series (3.11) is decomposed into two parts:

$$F(x,y;z) = \sum_{n=0}^{M} B_n(x,y;z) + \sum_{n=M+1}^{\infty} B_n(x,y;z). \tag{3.36}$$

Inserting (3.32) into (3.36) with $N = M - n$ constructs the $M$-term power series in $z$ for $F(x,y;z)$ [Eq. (1.10)]. The definition of the error term in (1.10) is then

$$E_M(x,y;z) = \sum_{n=0}^{M} E_{n,M-n}(x,y;z) + \sum_{n=M+1}^{\infty} B_n(x,y;z). \tag{3.37}$$

It is straightforward to see that $E_M$ is of order $O(|z|^{M+1})$. Concise bounds for $E_M$ follow from the bounds (3.35) and the absolutely convergent integrals that define $B_n(x,y;z)$. By this process it is found that the following theorem holds.

**Theorem 2:** Let $W \in \mathscr{F}^{*}_{2(M+1)}$ and $K$ be the associated bound constant in the space $\mathscr{F}^{*}_{2(M+1)}$. Let $U(x,y;z)$ and $U_0(x,y;z)$ be the integral kernels of operators $e^{-zH}$ and $e^{-zH_0}$, respectively. Then for all $z \in \overline{D} \setminus \{0\}$,

$$U(x,y;z) = U_0(x,y;z) \left\{ \sum_{n=0}^{M} \frac{(-z)^n}{n!} P_n(x,y) + E_M(x,y;z) \right\}, \tag{3.38}$$

where the $\mathbf{C}^{r\times r}$ valued coefficient functions are $[P_0(x,y) = I]$

$$P_n(x,y) = \sum_{m=0}^{n-1} q^m D_{m,n}(x,y), \quad n = 1,\ldots,M. \tag{3.39}$$

The coefficient functions $P_n$ and the remainder $E_M$ have $\mathbf{R}^d \times \mathbf{R}^d$ uniform bounds for $z \in \overline{D}$

$$|P_n(x,y)| \leq (\|\mu\| + qn^2K^2/4)^n \tag{3.40}$$

and

$$|E_M(x,y;z)| \leq \frac{(|z|\,\|\mu\|)^{M+1}}{(M+1)!} \left\{ \left[1 + \frac{qM^2K^2}{4\|\mu\|}\right]^{M+1} + e^{|z|\|\mu\|} \right\}. \tag{3.41}$$

Several comments are in order. Although we do not formulate it as a theorem, it is apparent that if $W \in \mathscr{F}^{*}_{2(M+1+j)}$ then one may differentiate the $F$-asymptotic expansion $j$ times with respect to $z$ (for details see the proof of Theorem 2, Ref. 1). The resultant identity is an asymptotic expansion with uniform (in $x,y$,arg $z$) error term of order $O(|z|^{M+1-j})$. Similar conclusions apply to differentiating (1.10) with respect to the variables $x$ and $y$.

Equation (3.39) contains the general formula for the coefficient matrices that appear in the heat-kernel expansion (3.38). The $x \neq y$ off-diagonal formulas for the coefficients $P_n(x,y)$ have a geometrical character. Denote by $\hat{\xi}_i$ the linear path from $x$ to $y$ parametrized by $\xi_i$,

$$\hat{\xi}_i = (1-\xi_i)x + \xi_i y. \tag{3.42}$$

In terms of this linear path the formulas for $P_1(x,y)$ and $P_2(x,y)$ become

$$P_1(x,y) = \int_0^1 d\xi_1\, W(\hat{\xi}_1), \tag{3.43}$$

$$P_2(x,y) = 2\int_0^1 d\xi_1 \int_0^{\xi_1} d\xi_2\, W(\hat{\xi}_2) W(\hat{\xi}_1) - 2q\int_0^1 d\xi_1\, \xi_1(1-\xi_1)(\Delta W)(\hat{\xi}_1). \tag{3.44}$$

So, $P_1(x,y)$ is just the linear average of $W$ along on the straight line drawn in $\mathbf{R}^d$ between $x$ and $y$. In $P_2(x,y)$, the $q$ part is an average of $\Delta W$ weighted by the polynomial $\xi_1(1-\xi_1)$. Of course, the parametric integrals over $\xi_i$ are inherited from the time-ordered parameters that appear in the Dyson expression (3.11).

The functions simplify markedly on the $x = y$ diagonal. If $x = y$ then $\hat{\xi}_i = x$ and is thus independent of the value of $\xi_i$. So in the integrals (3.43) and (3.44), and in general, the potentials $W$ can be taken outside the parametric integration $d^n\xi$. The parametric integral $d^n\xi$ multiplies a given polynomial in $\xi_1, \xi_2, \ldots, \xi_n$ and yields a numerical coefficient. In this way one is able to determine all the formulas for diagonal values of $P_n$. Specific formulas for $P_1$ through $P_4$ are

$$P_1(x,x) = W(x), \tag{3.45}$$

$$P_2(x,x) = W^2(x) - \tfrac{1}{3}q(\Delta W)(x), \tag{3.46}$$

$$P_3(x,x) = W^3(x) - q\{\tfrac{1}{2}W(\Delta W) + \tfrac{1}{2}(\Delta W)W + \tfrac{1}{2}(\nabla W)^2\}(x) + \tfrac{1}{10}q^2(\Delta^2 W)(x), \tag{3.47}$$

$$\begin{aligned} P_4(x,x) = {} & W^4(x) - q\{\tfrac{3}{3}W^2(\Delta W) \\ & + \tfrac{4}{3}W(\Delta W)W + \tfrac{3}{3}(\Delta W)W^2 \\ & + \tfrac{4}{3}W(\nabla W)^2 + \tfrac{2}{3}(\nabla W)W(\nabla W) + \tfrac{4}{3}(\nabla W)^2 W\}(x) \\ & + q^2\{\tfrac{1}{3}W(\Delta^2 W) + \tfrac{1}{3}(\Delta^2 W)W + \tfrac{1}{3}(\Delta W)^2 \\ & + \tfrac{2}{3}(\nabla W)[\nabla(\Delta W)] + \tfrac{2}{3}[\nabla(\Delta W)](\nabla W) \\ & + \tfrac{4}{15}(\nabla_2\nabla_1)^2 W_2 W_1\}(x) - \tfrac{1}{35}q^3(\Delta^3 W)(x). \end{aligned} \tag{3.48}$$

If these coefficients are compared to the known formulas[8,30] for the $r=1$ case we see the first structural change occurs in the $P_3(x,x)$ coefficient, where the noncommutivity of $W$ and $\Delta W$ lead to the symmetric combination $\frac{1}{2}W\Delta W+\frac{1}{2}\Delta WW$ rather than $W\Delta W$. The result available in the literature that comes closest to (3.45)–(3.48) is the determination by Fulling[31] of the coefficient matrices $P_n$ for $r>1$ but with space dimension $d=1$. In this special case, complete agreement is found with Fulling's coefficient expressions, including the $n=5$ coefficient which we have not given above because of its substantial length.

Finally observe the symmetry that the $P_n$ obeys as a consequence of the self-adjoint nature of $H$. Since $H$ is self-adjoint, so is $e^{-\beta H}$, $\beta>0$. If * denotes the adjoint on $\mathbb{C}^{r\times r}$, then the integral kernel of $e^{-\beta H}$ satisfies

$$U(x,y;\beta)=U(y,x;\beta)^*. \tag{3.49}$$

Inserting (3.15) and series (1.10) in the above identity gives

$$P_n(x,y)=P_n(y,x)^*. \tag{3.50}$$

This follows since $h(x-y;\beta)$ is real and invariant under $x\leftrightarrow y$. If $r=1$, then (3.50) and the symmetry $P_n(x,y)=P_n(y,x)$ requires that the $P_n(x,y)$ be real valued.

The semiclassical facet of the heat-kernel expansion (1.10) resides in the fact that the $P_n(x,y)$ are polynomials in $q$ of order $n-1$. It has been shown[32,33] in the $r=1$ case that if the asymptotic series for (1.10) is exponentiated, then a nonperturbative semiclassical approximation for $U(x,y;z)$ is defined. Furthermore, if $x=y$ the Wigner–Kirkwood semiclassical expansion[34,35] is recovered as a special case.

## IV. RESOLVENT KERNEL EXPANSIONS

This section describes the large $z$ asymptotic expansion of the resolvent kernel $R(x,y;z)$. The technique utilized is to investigate the behavior of $R(x,y;z)$ by using the Laplace transform (1.11) that connects the heat kernel to the resolvent kernel. In the first instance, the Laplace transform (1.11) is defined as a convergent integral only on the restricted set $\mathrm{Re}\,z<-\|\mu\|$. However by exploiting the holomorphic character in $z$ of the kernels $U(x,y;z)$ it is possible to analytically continue the Laplace transform representation of $R(x,y;z)$ to the domain $\tilde{V}_\delta$. With this approach the Laplace image of asymptotic expansion (3.38) for $U(x,y;z)$ becomes the asymptotic expansions for $R(x,y;z)$. Furthermore the error term bound (3.41) for $E_M(x,y;z)$ suffices to provide an $\mathbb{R}^d\times\mathbb{R}^d$ uniform error term bound for the $R(x,y;z)$ asymptotic expansion.

Let $\{E^0_\lambda\,|\,\lambda\geqslant 0\}$ be the family of spectral projectors that is defined by $H_0$. In terms of $E^0_\lambda$, the spectral representation of the free solvent is

$$r_0(z)=(H_0-z)^{-1}=\int_0^\infty \frac{1}{\lambda-z}\,dE^0_\lambda,\quad z\notin[0,\infty). \tag{4.1}$$

Consider first the kernel representation of the free resolvent.

*Lemma 3:* Suppose $\mathrm{Re}\,z<0$, then we have the following.

(i) For each $f\in L^2(\mathbb{R}^d,\mathbb{C}^r)$,

$$(r_0(z)f)(x)=\int_0^\infty e^{\beta z}(e^{-\beta H_0}f)(x)\,d\beta,\quad \text{a.a. } x. \tag{4.2}$$

(ii) Let $C_0=\{z\,|\,z\in\mathbb{C},\mathrm{Re}\,z<0\}$. Define the function $R_0:\mathbb{R}^d\times\mathbb{R}^d\times C_0\to\mathbb{C}^{r\times r}$ by the integral

$$R_0(x,y;z)=\int_0^\infty e^{\beta z}U_0(x,y;\beta)\,d\beta. \tag{4.3}$$

$R_0(x,y;z)$ is translation invariant in $\mathbb{R}^d\times\mathbb{R}^d$ (i.e., it depends only on the variable $x-y$). For each pair $x,y$ $(x\neq y)$ $R_0(x,y;z)$ is an analytic function in domain $C_0$. Finally, $R_0(x,y;z)$ satisfies the integral estimate

$$\int dy\,|R_0(x,y;z)|\leqslant\frac{\sqrt{r}}{-\mathrm{Re}\,z}. \tag{4.4}$$

(iii) For $z\in C_0$, $r_0(z)$ is an integral operator with an $L^1$-convolution kernel $R_0(x,y;z)$. For each $f\in L^2(\mathbb{R}^d,\mathbb{C}^r)$,

$$(r_0(z)f)(x)=\int dy\,R_0(x,y;z)f(y),\quad \text{a.a. } x. \tag{4.5}$$

*Proof:* These results are all elementary but we will write out a detailed proof in a form that allows an easy extension to include the $(H-z)^{-1}$ case. Start with (i). Observe that the right-hand side of (4.2) defines an $L^2$ function of $x$:

$$\begin{aligned}\int dx\left|\int_0^\infty e^{\beta z}(e^{-\beta H_0}f)(x)\,d\beta\right|^2 &\leqslant\int_0^\infty d\beta_1\int_0^\infty d\beta_2\,e^{(\beta_1+\beta_2)\mathrm{Re}\,z}\\ &\quad\times\int dx\,|(e^{-\beta_1 H_0}f)(x)|\,|(e^{-\beta_2 H_0}f)(x)|\\ &\leqslant\|f\|^2\left(\int_0^\infty d\beta\,e^{\beta\,\mathrm{Re}\,z}\right)^2<\infty.\end{aligned} \tag{4.6}$$

The second inequality follows from $\|e^{-\beta H_0}\|\leqslant 1$. Let $g$ be an arbitrary element of $L^2(\mathbb{R}^d,\mathbb{C}^r)$. Then

$$\begin{aligned}A&\equiv\left\langle g,\int_0^\infty e^{\beta z}(e^{-\beta H_0}f)\,d\beta\right\rangle\\ &=\int dx\,g(x)^*\left\{\int_0^\infty e^{\beta z}(e^{-\beta H_0}f)(x)\,d\beta\right\}.\end{aligned} \tag{4.7}$$

Using the Schwartz inequality and $\|e^{-\beta H_0}f\|\leqslant\|f\|$ it follows that the $dx\,d\beta$ integral in (4.7) is absolutely convergent. Fubini's theorem allows us to interchange the order of integration giving

$$\begin{aligned}A&=\int_0^\infty e^{\beta z}\langle g,e^{-\beta H_0}f\rangle\,d\beta\\ &=\int_0^\infty e^{\beta z}\left\{\int_0^\infty e^{-\beta\lambda}\,d\langle g,E^0_\lambda f\rangle\right\}d\beta.\end{aligned} \tag{4.8}$$

The last equality employs the spectral theorem for $H_0$. The spectral measure has finite total variation bounded by $\|g\|\,\|f\|$. Further, $e^{\beta\,\mathrm{Re}\,z}d\beta$ is absolutely integrable, so the order of integration in (4.8) again can be reversed. Using

$$\int_0^\infty e^{\beta(z-\lambda)}\,d\beta=\frac{1}{\lambda-z}, \tag{4.9}$$

(4.8) takes the form

$$A=\int_0^\infty\frac{1}{\lambda-z}\,d\langle g,E^0_\lambda f\rangle=\langle g,r_0(z)f\rangle. \tag{4.10}$$

The equality of (4.10) and (4.7) for all $g\in L^2(\mathbb{R}^d,\mathbb{C}^r)$ implies (4.2).

(ii) The expression (3.14) for $U_0(x,y;\beta)$ shows that the integral (4.3) for a fixed pair $x,y$ $(x\neq y)$ is absolutely convergent for $\mathrm{Re}\, z<\epsilon<0$. Since the integrand of (4.3) is analytic in $C_0$ it follows that the integral defines an analytic function for $\mathrm{Re}\, z<\epsilon$. Since $\epsilon$ may be selected to be as small as desired, the allowed domain for $z$ may be extended to $C_0$. Here, $R_0(x,y;z)$ is translation invariant because $U_0(x,y;\beta)$ is a function of $x-y$. Finally, since $h(x-y;\beta)$ is real and positive,

$$\int dy|R_0(x,y;z)|\leqslant\int_0^\infty e^{\beta\,\mathrm{Re}\,z}\sqrt{r}\left\{\int dy\,h(x-y;\beta)\right\}d\beta. \tag{4.11}$$

The diffusion function $h(x-y;\beta)$ is normalized so that the $dy$ integral is unity. Equation (4.4) follows by carrying out the $d\beta$ integration.

(iii) Identity (3.16) with $W=0$ states for $f\in L^2(\mathbb{R}^d,\mathbb{C}^r)$

$$(e^{-\beta H_0}f)(x)=\int dy\,U_0(x,y;\beta)f(y),\quad \text{a.a. } x. \tag{4.12}$$

Combining this with (4.2) gives us

$$(r_0(z)f)(x)=\int_0^\infty e^{\beta z}\left\{\int dy\,U_0(x,y;\beta)f(y)\right\}d\beta. \tag{4.13}$$

Changing the order of the $d\beta\,dy$ integrals here leads at once to formula (4.5) with $R_0(x,y;\beta)$ defined by (4.3). Now, consider the justification of this interchange of integral order. The iterated integral on the right of (4.13) is majorized by

$$A(x)=\int_0^\infty e^{\beta\,\mathrm{Re}\,z}\left\{\int dy\,h(x-y;\beta)|f(y)|\right\}d\beta. \tag{4.14}$$

All the functions in (4.14) are non-negative so it may be written

$$\begin{aligned}A(x)&=\int dy\left\{\int_0^\infty e^{\beta\,\mathrm{Re}\,z}h(x-y;\beta)d\beta\right\}|f(y)|\\&=\int dy\,r_0(x-y;\mathrm{Re}\,z)|f(y)|,\end{aligned} \tag{4.15}$$

where $r_0(x-y;\mathrm{Re}\,z)$ is just the resolvent kernel $R_0(x,y;\mathrm{Re}\,z)$ in the $r=1$ case. Estimate (4.4) means that (4.15) is an $L^1$-convolution. Now the Hausdorff–Young inequality for convolutions[36] states that if we have a convolution

$$\psi(x)=\int dy\,K(x-y)\phi(y), \tag{4.16}$$

where $K\in L^1(\mathbb{R}^d)$ and $\phi\in L^p(\mathbb{R}^d)$, $p\geqslant 1$, then

$$\|\psi\|_p\leqslant\|K\|_1\|\phi\|_p. \tag{4.17}$$

Applying (4.17) to (4.15) with $p=2$, it follows that $A(x)$ is $L^2(\mathbb{R}^d)$ and thus $A(x)<\infty$ for all but an exceptional set of measure zero in $\mathbb{R}^d$. So, the majorant (4.14) of the iterated integral is absolutely convergent almost everywhere in $x$. Outside this exception set we apply Fubini's theorem in order to justify the change of integral order in (4.13). □

The next step is to extend our analysis to treat the full resolvent $r(z)$. For the self-adjoint operator $H$, we let $\{E_\lambda\,|\lambda\geqslant-\|\mu\|\}$ be the family of spectral projectors. The resolvent then has the spectral representation

$$r(z)=(H-z)^{-1}=\int_{-\|\mu\|}^\infty\frac{1}{\lambda-z}\,dE_\lambda,\quad z\in\rho(H). \tag{4.18}$$

*Lemma 4:* Let $W\in\mathscr{F}^*$ and $\mathrm{Re}\,z<-\|\mu\|$, then we have the following.

(i) For each $f\in L^2(\mathbb{R}^d,\mathbb{C}^r)$,

$$(r(z)f)(x)=\int_0^\infty e^{\beta z}(e^{-\beta H}f)(x)d\beta,\quad \text{a.a. } x. \tag{4.19}$$

(ii) Let $C_\mu=\{z|z\in\mathbb{C},\mathrm{Re}\,z<-\|\mu\|\}$. Define the function $R:\mathbb{R}^d\times\mathbb{R}^d\times C_\mu\to\mathbb{C}^{r\times r}$ by the integral

$$R(x,y;z)=\int_0^\infty e^{\beta z}U(x,y;\beta)d\beta. \tag{4.20}$$

For each fixed pair $x,y$ $(x\neq y)$, $R(x,y;z)$ is an analytic function of $z$ with domain $C_\mu$. Furthermore, for $z\in C_\mu$, $R(x,y;z)$ satisfies the pointwise bound

$$|R(x,y;z)|\leqslant\sqrt{r}|R_0(x,y;\mathrm{Re}\,z+\|\mu\|)|,\quad \forall x,y\in\mathbb{R}^d. \tag{4.21}$$

(iii) If $z\in C_\mu$, $r(z)$ is an integral operator with the kernel $R(x,y;z)$. For each $f\in L^2(\mathbb{R}^d,\mathbb{C}^r)$,

$$(r(z)f)(x)=\int dy\,R(x,y;z)f(y),\quad \text{a.a. } x. \tag{4.22}$$

*Proof:* The estimate

$$\|e^{-\beta H}f\|\leqslant e^{\mathrm{Re}\,\beta\|\mu\|}\|f\|,\quad f\in\mathscr{H},\quad \beta\in\overline{D}, \tag{4.23}$$

and estimate (3.13) for $|F_{\nu\gamma}(x,y;\beta)|$ allow us to follow the same line of argument used to prove Lemma 3 provided that the restriction $\mathrm{Re}\,z<0$ is shifted to $\mathrm{Re}\,z\leqslant-\|\mu\|$. The pointwise bound (4.21) results from

$$\begin{aligned}\left|\int_0^\infty e^{\beta z}U(x,y;\beta)d\beta\right|&\leqslant\int_0^\infty e^{\beta\,\mathrm{Re}\,z}|U(x,y;\beta)|d\beta\\&\leqslant\int_0^\infty e^{\beta(\mathrm{Re}\,z+\|\mu\|)}rh(x-y;\beta)d\beta\\&\leqslant\sqrt{r}|R_0(x,y;\mathrm{Re}\,z+\|\mu\|)|,\end{aligned} \tag{4.24}$$

where the second inequality is a consequence of (3.15) and (3.13). □

The existence of the Laplace transform (4.2) requires the restriction $\mathrm{Re}\,z<0$. However, the domain of analyticity of $R_0(x,y;z)$ is considerably larger. First fix the polar coordinate representation of $z$ by choosing $\arg z\in[0,2\pi)$ with the positive real axis corresponding to $\arg z=0$. With this notation, the transform (4.3) may be evaluated explicitly[37] yielding a modified Bessel function. For $x\neq y$,

$$\begin{aligned}R_0(x,y;z)=&\frac{2}{(4\pi q)^{d/2}}\left(\frac{i|x-y|}{2q^{1/2}z^{1/2}}\right)^{1-d/2}\\&\times K_{(d/2)-1}(-iq^{-1/2}z^{1/2}|x-y|),\end{aligned} \tag{4.25}$$

where $z^{1/2}$ is the square root with positive sign. The right side of (4.25) is an analytic function with domain $\mathbb{C}\setminus[0,\infty)$.

In the subsequent anlaysis it is shown that one may exploit the analyticity of $U(x,y;\beta)$ in order to implement an analytic continuation of $R(x,y;z)$ from the domain $C_\mu$ to a larger subset of $\mathbb{C}$. The method utilizes contour rotation and depends only on the analytic semigroup properties obtained in Sec. III.

We introduce some convenient terminology. Heretofore, $\beta$ has denoted a positive number. Now let $\beta\in\overline{D}$ and specify the polar form to be $\beta=|\beta|e^{i\arg\beta}$ with $\arg\beta\in[\pi/2,-\pi/2]$. For any $\delta\in(0,\pi/2]$ define two linear rays in $D$ by

$$L^{\pm}(\delta)=\{\beta\in\overline{D}\,|\,\pm\arg\beta=\pi/2-\delta\}. \tag{4.26}$$

Here, $L^+(\delta)$ is a ray in the upper right quadrant of the com-

plex $\beta$-plane and $L^-(\delta)$ lies in the lower right quadrant. Let us introduce special notations for several different domains in $\mathbb{C}$ for the $z$-variable appearing in $R(x,y;z)$. For $\delta \in (0,\pi/2]$ define

$$V^+(\delta) = \{z \in \mathbb{C} | 2\delta < \arg z < \pi\}, \tag{4.27}$$

$$V^-(\delta) = \{z \in \mathbb{C} | \pi < \arg z < 2\pi - 2\delta\}. \tag{4.28}$$

Finally indicate by $V^{\pm}(\delta) + (-z')$ the set in $\mathbb{C}$ defined by translating the set $V^{\pm}(\delta)$ by $-z' \in \mathbb{C}$, i.e., $z + z' \in V^{\pm}(\delta)$.

*Proposition 4:* Let $W \in \mathcal{F}^*$ and parameter $\delta \in (0,\pi/2)$. Define $D_\delta^{\pm} = V^{\pm}(\delta) + (-\|\mu\|(\sin\delta)^{-1})$.

(i) If either $z \in D_\delta^+$ or $z \in D_\delta^-$ and $f \in L^2(\mathbb{R}^d,\mathbb{C}^r)$, then

$$(r(z)f)(x) = \int_{L^{\pm}(\delta)} e^{\beta z}(e^{-\beta H}f)(x)d\beta, \quad \text{a.a. } x, \tag{4.29}$$

where the integration path $L^+(\delta)$ is applicable for domain $D_\delta^+$ and $L^-(\delta)$ for $D_\delta^-$.

(ii) Define the functions $R^{\pm}: \mathbb{R}^d \times \mathbb{R}^d \times D_\delta^{\pm} \to \mathbb{C}^{r\times r}$ by the integral

$$R^{\pm}(x,y;z) = \int_{L^{\pm}(\delta)} e^{\beta z}U(x,y;\beta)d\beta. \tag{4.30}$$

For each fixed coordinate pair $x,y$ $(x \neq y)$, $R^+(x,y;z)$ is an analytic function of $z$ in $D_\delta^+$ and $R^-(x,y;z)$ is an analytic function in $D_\delta^-$. The functions $R^{\pm}$ satisfy for all $z \in D_\delta^{\pm}$ the pointwise estimate

$$|R^{\pm}(x,y;z)| \leqslant (\sqrt{r}/(\sin\delta)^{(d/2)-1})|R_0(x,y; -|z_0|\sin^2\delta)|, \tag{4.31}$$

where $z_0 = z + \|\mu\|(\sin\delta)^{-1}$.

(iii) For $z$ in either $D_\delta^+$ or $D_\delta^-$, then $r(z)$ is an integral operator with a kernel given by $R^+(x,y;z)$ or $R^-(x,y;z)$, respectively. For each $f \in L^2(\mathbb{R}^d,\mathbb{C}^r)$ and $z$ in the appropriate domain $D_\delta^+$ or $D_\delta^-$, then

$$(r(z)f)(x) = \int dy\, R^{\pm}(x,y;z)f(y), \quad \text{a.a. } x. \tag{4.32}$$

(iv) $R^{\pm}(x,y;z)$ are analytic continuations of $R(x,y;z)$ from domain $C_\mu$ to the domains $D_\delta^{\pm}$. Specifically, for $x \neq y$,

$$R^+(x,y;z) = R(x,y;z), \quad z \in C_\mu \cap D_\delta^+, \tag{4.33}$$

$$R^-(x,y;z) = R(x,y;z), \quad z \in C_\mu \cap D_\delta^-. \tag{4.34}$$

*Proof:* (i) Define $u: \mathbb{R}^d \to \mathbb{C}^r$ by the integral

$$u(x) = \int_{L^+(\delta)} e^{\beta z}(e^{-\beta H}f)(x)d\beta. \tag{4.35}$$

By employing the Schwartz inequality it is easily shown that the $L^2(\mathbb{R}^d,\mathbb{C}^r)$ norm of $u$ has the bound

$$\|u\| \leqslant \int_{L^+(\delta)} |e^{\beta z}|\, \|e^{-\beta H}f\|\, |d\beta|. \tag{4.36}$$

In order to establish the finiteness of the $L^+(\delta)$ integral in (4.36) it suffices to find an $|d\beta|$ integrable bound. To this end, introduce the following polar coordinates: $\beta = t\exp[i(\pi/2 - \delta)]$, $t = |\beta|$, $\beta \in L^+(\delta)$; $z_0 = z + \|\mu\|(\sin\delta)^{-1} = |z_0|\exp(i\theta)$. If $z_0 \in V^+(\delta)$, then $\theta \in (2\delta,\pi)$. Thus for all $z \in D_\delta^+$ one has

$$|e^{\beta z}| \leqslant \exp(-t|z_0|\sin(\theta - \delta) - t\|\mu\|). \tag{4.37}$$

Furthermore, inequality (4.23) provides the bound

$$\|e^{-\beta H}f\| \leqslant e^{t\|\mu\|\sin\delta}\|f\|, \quad \beta \in L^+(\delta). \tag{4.38}$$

Taken together, (4.37) and (4.38) give the estimate

$$|e^{\beta z}|\, \|e^{-\beta H}f\| \leqslant \exp(-t|z_0|\sin\delta - t\|\mu\|(1 - \sin\delta))\|f\|, \tag{4.39}$$

where we have used the fact that $\sin(\theta - \delta) \geqslant \sin\delta$ for $z_0 \in V^+(\delta)$. So (4.36) acquires the bound

$$\|u\| \leqslant \|f\|/(\|\mu\|(1 - \sin\delta) + |z_0|\sin\delta) < \infty. \tag{4.40}$$

Let $g$ be an arbitrary $L^2(\mathbb{R}^d,\mathbb{C}^r)$ element. Form the inner product

$$\langle g,u\rangle = \int dx\, g(x)^* \left\{\int_{L^+(\delta)} e^{\beta z}(e^{-\beta H}f)(x)d\beta\right\}. \tag{4.41}$$

Inversion of integral order is permitted here since bound (4.40) together with the Schwartz inequality shows that the $dx|d\beta|$ integral is absolutely convergent. Introducing the spectral representation for $\langle g,e^{-\beta H}f\rangle$ gives

$$\langle g,u\rangle = \int_{L^+(\delta)} e^{\beta z}\left\{\int_{-\|\mu\|}^{\infty} e^{-\beta\lambda}d\langle g,E_\lambda f\rangle\right\}d\beta. \tag{4.42}$$

Upon utilizing

$$\int_{L^+(\delta)} e^{\beta(z-\lambda)}d\beta = \frac{1}{\lambda - z} \tag{4.43}$$

and a final inversion of integral order [valid because the measure $d\langle g,E_\lambda f\rangle$ is of finite total variation and $|e^{\beta z}|$ has estimate (4.37)] one obtains

$$\langle g,u\rangle = \int_{-\|\mu\|}^{\infty} \frac{1}{\lambda - z}d\langle g,E_\lambda f\rangle = \langle g,r(z)f\rangle, \tag{4.44}$$

or since $g$ is arbitrary,

$$u = r(z)f. \tag{4.45}$$

This is (4.29) with contour $L^+(\delta)$ and $z$-domain $D_\delta^+$. A similar argument applies to contour $L^-(\delta)$ and domain $D_\delta^-$.

(ii) Given relationship (3.15) and estimate (3.13) it follows for $z \in D_\delta^{\pm}$ that the integral (4.30) is uniformly convergent. Since the integrand is analytic the integral defines a holomorphic function of $z$. The convolution bound (4.31) results from majorizing the integral with estimate (3.13) and inequality (4.37).

(iii) Combining identity (3.16) of Theorem 1 with Eq. (4.29) leads to

$$(r(z)f)(x) = \int_{L^+(\delta)} e^{\beta z}\left\{\int dy\, U(x,y;\beta)f(y)\right\}d\beta, \quad \text{a.a. } x. \tag{4.46}$$

Inverting the order of integration gives (4.32). In view of estimate (4.37) for $|e^{\beta z}|$ and the bound

$$|F(x,y;\beta)f(y)| \leqslant e^{|\beta|\,\|\mu\|}|f(y)|, \tag{4.47}$$

the convolution argument given in (4.14) and (4.15) may be used to show that the integral order in (4.46) may be reversed except possibly on a set of $x$ having zero measure.

(iv) Consider the case with $z \in D_\delta^+$. Let $s$ and $S$ be real parameters $0 < s < S < \infty$. Define a closed contour in the analytic semigroup domain $D$ by joining the four segments: $C_1 = \{\beta | s \leqslant \beta \leqslant S\}$, $C_2 = \{\beta | \beta = S\, e^{i\phi},\ \phi \in [0,(\pi/2) - \delta]\}$, $C_3 = \{\beta | \beta = t\, e^{i((\pi/2)-\delta)},\ s \leqslant t \leqslant S\}$ and $C_4 = \{\beta | \beta = s\, e^{i\phi},\ \phi \in [0,(\pi/2) - \delta]\}$. The kernel $U(x,y;\beta)$ is analytic in $D$.

Cauchy's theorem applied to the contour $C_1 + C_2 + C_3 + C_4$ states

$$\oint e^{\beta z} U(x,y;\beta)d\beta = 0. \tag{4.48}$$

We restrict $z$ so that it lies in a subset of the union of $D_\delta^+$ and $C_\mu$; specifically we take

$$\mathrm{Re}\, z < -\|\mu\|/\sin\delta, \quad z \in D_\delta^+. \tag{4.49}$$

Now estimate the contribution of the $C_2$ integral as $S \to \infty$. Let the variable $z_0$ be that defined after (4.36). Assume $\epsilon > 0$ and note that condition (4.49) is obeyed if

$$z_0 = |z_0| e^{i(\pi/2 + \theta')}, \quad \theta' \in (0,\pi/2), \quad |z_0| > \epsilon. \tag{4.50}$$

Employing estimate (3.13), a computation of the Euclidean norm shows that

$$\left| \int_{C_2} e^{\beta z} U(x,y;\beta)d\beta \right| \leqslant \frac{r(\pi - 2\delta)Se^{-\epsilon a S}}{2(4\pi q S)^{d/2}}, \tag{4.51}$$

where $a > 0$ and is the minimum value of the two numbers $\sin\theta'$ and $\sin(\theta' + \pi/2 - \delta)$. Thus the $C_2$ contribution to (4.48) vanishes as $S \to \infty$. For $x \neq y$, similar reasoning and conclusions apply to the $C_4$ contribution as $s \to 0$. Thus after taking the limits $S \to \infty$ and $s \to 0$, (4.48) becomes

$$\int_{L^+(\delta)} e^{\beta z} U(x,y;\beta)d\beta = \int_0^\infty e^{\beta z} U(x,y;\beta)d\beta. \tag{4.52}$$

This is just equality (4.33) for the $z$ allowed by (4.49). For $z \in D_\delta^-$ the argument proceeds by taking the contour in $D$ that is the conjugate image of $C_1 + C_2 + C_3 + C_4$. □

In view of the fact that $R^\pm(x,y;z)$ and $R(x,y;z)$ represent the same analytic function we will drop the $\pm$ superscripts. Furthermore, we denote by $V(\delta)$ the $z$-plane sector $V^+(\delta) \cup V^-(\delta)$. Observe that inequality (4.31) provides in $V(\delta) + (-\|\mu\|(\sin\delta)^{-1})$ an $L^1$-convolution bound for the resolvent kernel $R(x,y;z)$. Convolution bounds commonly occur for resolvent kernels of elliptic differential operators. For a recent discussion of this topic see Gurarie.[38]

If the potential $W(x)$ is set equal to zero, the the conclusions of Proposition 4 specialize to the following corollary.

*Corollary 2:* Let $\delta \in (0,\pi/2)$, then we have the following.

(i) For $z \in V(\delta)$ the resolvent operator $r_0(z)$ has a kernel $R_0: \mathbb{R}^d \times \mathbb{R}^d \times V^\pm \to \mathbb{C}^{r \times r}$ given by the integral representation

$$R_0(x,y;z) = \int_{L^\pm(\delta)} e^{\beta z} U_0(x,y;\beta)d\beta, \tag{4.53}$$

where the path $L^+(\delta)$ is associated with domain $V^+(\delta)$, and $L^-(\delta)$ with $V^-(\delta)$.

(ii) For all $x,y$ with $x \neq y$ and positive integers $n$,

$$\left(\frac{\partial}{\partial z}\right)^n R_0(x,y;z) = \int_{L^\pm(\delta)} \beta^n e^{z\beta} U_0(x,y;\beta)d\beta, \tag{4.54}$$

for $z$ in the appropriate $V^\pm(\delta)$ domain.

(iii) If $n + 1 > d/2$ and $z \in V(\delta)$,

$$\left| \left(\frac{\partial}{\partial z}\right)^n R_0(x,y;z) \right| \leqslant \frac{\sqrt{r}\,\Gamma(n + 1 - (d/2))}{(4\pi q)^{d/2}} \left(\frac{1}{|z|\sin\delta}\right)^{n+1-(d/2)} \tag{4.55}$$

uniformly in $\mathbb{R}^d \times \mathbb{R}^d$.

*Proof:* (i) This is the statement of Proposition 4 that results if $W = 0$.

(ii) The absolute integrability of the integrand of (4.54) justifies passing the partial differentiation $(\partial/\partial z)^n$ through the integral sign in (4.53).

(iii) The right side of (4.55) is the outcome of taking the modulus of the integrand in (4.54) and completing the $|d\beta|$ integration. Symbol $\Gamma$ denotes the gamma function. □

In the following we set $\tilde{V}_\delta = D_\delta^+ \cup D_\delta^-$. This is the allowed domain for the variation of $z$. Further it is convenient to denote the $n$th $z$-derivative of $R_0(x,y;z)$ by

$$R_0^{(n)}(x,y;z) = \left(\frac{\partial}{\partial z}\right)^n R_0(x,y;z). \tag{4.56}$$

The asymptotic expansion of the resolvent kernel of $r(z)$ for large $z$ is described by the following theorem.

**Theorem 3:** Let $W \in \mathcal{F}^*_{2(M+1)}$ and let $K$ be the associated bound constant of potential $W$ in space $\mathcal{F}^*_{2(M+1)}$. Suppose $\delta \in (0,\pi/2)$. For $z \in \tilde{V}_\delta$ and all $x,y$ $(x \neq y)$ the resolvent kernel of $(H - z)^{-1}$ admits the expansion

$$R(x,y;z) = \sum_{n=0}^{M} \frac{(-1)^n}{n!} P_n(x,y) R_0^{(n)}(x,y;z) + T_M(x,y;z). \tag{4.57}$$

Define the constant $C = C(\|\mu\|, M)$ by

$$C = \frac{\|\mu\|^{M+1}}{(M+1)!} \left[ \left(1 + q \frac{M^2 K^2}{4\|\mu\|}\right)^{M+1} + 1 \right]. \tag{4.58}$$

(i) For $z \in \tilde{V}_\delta$ and all integers $M > 0$ the error term has the $x \neq y$ nonuniform bound

$$|T_M(x,y;z)| \leqslant \frac{C(\sin\delta)^{M+2-(d/2)}}{\sqrt{r}} \times |R_0^{(M+1)}(x,y; -|(\sin^2\delta)z + (\sin\delta)\|\mu\||)|. \tag{4.59}$$

(ii) If $M + 2 > d/2$, then for $z \in \tilde{V}_\delta$ the error term has the $\mathbb{R}^d \times \mathbb{R}^d$ uniform bound

$$|T_M(x,y;z)| \leqslant \frac{C\Gamma(M + 2 - (d/2))}{(4\pi q)^{d/2}} \times \frac{1}{|(\sin\delta)z + \|\mu\||^{M+2-(d/2)}}. \tag{4.60}$$

*Proof:* If $z \in \tilde{V}_\delta$, then either $z \in D_\delta^+$ or $z \in D_\delta^-$. Suppose the first. In this case, the resolvent kernel $R(x,y;z)$ has the integral representation (4.30) with contour $L^+(\delta)$. Since $W \in \mathcal{F}^*_{2(M+1)}$ the semigroup kernel $U(x,y;z)$ obeys the asymptotic expansion (3.38) of Theorem 2. Thus $R(x,y;z)$ may be written

$$R(x,y;z) = \int_{L^+(\delta)} e^{\beta z} U_0(x,y;\beta) \times \left\{ \sum_{n=0}^{M} \frac{(-\beta)^n}{n!} P_n(x,y) + E_M(x,y;\beta) \right\} d\beta. \tag{4.61}$$

The $n$ series is finite and may be passed through the $L^+(\delta)$ integral. Furthermore the individual terms in $n$ are exactly of the form (4.54), so

$$R(x,y;z) = \sum_{n=0}^{M} \frac{(-1)^n}{n!} P_n(x,y) R_0^{(n)}(x,y;z) + T_M(x,y;z), \tag{4.62}$$

where the remainder $T_M$ is defined by

$$T_M(x,y;z) = \int_{L^+(\delta)} e^{\beta z} U_0(x,y;\beta) E_M(x,y;\beta)d\beta. \tag{4.63}$$

If the upper bound (3.41) for $E_M$ is used in (4.63) we have

$$|T_M(x,y;z)| \leqslant C \int_{L^+(\delta)} |e^{\beta z}| \ |h(x-y;\beta)| \times |\beta|^{M+1} e^{|\beta| \|\mu\|} \ |d\beta|, \tag{4.64}$$

where $C$ is (4.58). Formula (4.59) for $|T_M|$ results if $|e^{\beta z}|$ is estimated by (4.37) and $|h|$ is bounded by

$$|h(x-y;\beta)| \leqslant \frac{1}{(4\pi q|\beta|)^{d/2}} \exp\left(-\frac{|x-y|^2}{4q|\beta|} \sin\delta\right). \tag{4.65}$$

Carrying out the $|d\beta|$ integration gives (4.59). Formula (4.60) follows from (4.59) and (4.55). This argument extends to $z \in D_\delta^-$ provided that contour $L^-(\delta)$ is used to represent $R(x,y;z)$. □

A number of comments are in order. An examination of the union of all allowed $z$ domains $\tilde{V}_\delta$ shows that there is a strip in $\mathbb{C}$ parallel to the positive real axis that is forbidden to $z$. All $z$ whose least distance to the positive real axis is less than $\|\mu\|$ are in the complement of all $\tilde{V}_\delta$. Thus Theorem 3 does not allow one to take $z$ arbitrarily close to the real $z$ axis in spite of the fact that $R(x,y;z)$ is analytic for all $z \notin [-\|\mu\|, \infty)$. This domain restriction appears in the analytic continuation technique of this section because of estimate (3.13) for $F(x,y;z)$. An examination of the proof of Lemma 2(see Ref. 1, Proposition 2) shows it has not used the fact that $W(x)$ is Hermitian. Series (3.11) will construct representations of $e^{-zH}$ for non-self-adjoint operators $H$ as well as self-adjoint ones. In the case of non-self-adjoint $H$ the spectrum is not confined to the real axis but may be any point in $\mathbb{C}$ not exceeding a distance $\|\mu\|$ from the positive real axis. Thus the analytic continuation based only on estimate (3.13) cannot penetrate the allowed spectrum of the non-Hermitian operators $H$.

Consider briefly the behavior of identity (4.57). If $n+1 > d/2$, (4.55) shows that the term $R_0^{(n)}(x,y;z)$ has a $\mathbb{R}^d \times \mathbb{R}^d$ uniform $O(|z|^{-n-1+(d/2)})$ estimate. However, the terms with $n+1 \leqslant d/2$ are singular at $x=y$. Since $T_M$ is bounded at $x=y$, one has the conclusion that the $x=y$ singularities of $R(x,y;z)$ and the singularities of the first $n \leqslant (d/2)-1$ terms of (4.57) must cancel. Finally, it should be recalled that the asymptotic expansion (4.57) has been derived assuming no other information about the potential $W(x)$ except the smoothness properties implied by the condition $W \in \mathcal{F}^*_{2(M+1)}$.

In the physics literature an asymptotic procedure similar to the one used in this section but applicable to quantum field theory in curved space-time may be found in DeWitt[39] and Christensen.[40] In the $r=1$ case, formula (4.57) is given in Ref. 30 and obtained heuristically. The treatment in Ref. 30 gave no estimate of the remainder nor a determination of the allowed $z$ domain. In the $d=3$, $r=1$ case, Buslaev[9] found a formula equivalent to (4.57). Specifically for a $C^\infty$ rapidly decreasing potential Buslaev succeeded in finding a bound of $|T_M|$ for $z$ lying on the spectral boundary $(z=\lambda \pm i0)$. This bound decreases as $\lambda \to +\infty$. Agmon and Kannai[41] have obtained a somewhat related expansion for a general class of elliptic differential operators defined on a compact manifold.


## ACKNOWLEDGMENTS

The first two authors (T.A.O. and R.A.C.) would like to thank the members of the Nuclear Theory Group at the University of Maryland for their warm hospitality.

This work has been supported in part by grants from the Natural Sciences and Engineering Research Council of Canada and the U.S. Department of Energy.



[1] T. A. Osborn and Y. Fujiwara, J. Math. Phys. **24**, 1093 (1983).
[2] F. Dyson, Phys.Rev. **75**, (1949) 486; Phys. Rev. **75**, 1730 (1949).
[3] M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1975), Vol. II, Chap. X.
[4] S. A. Fulling, SIAM J. Math. Anal. **13**, 891 (1982).
[5] P. Greiner, Arch. Ration. Mech. Anal. **41**, 163 (1971).
[6] P. B. Gilkey, Compositio Math. **38**, 201 (1979).
[7] M. F. Atiyah, R. Bott, and V. K. Patodi, Invent. Math. **19**, 279 (1973).
[8] A. M. Perelomov, Ann. Inst. H. Poincaré **24**, 161 (1976).
[9] V. S. Buslaev, in *Spectral Theory and Wave Processes*, edited by M. Sh. Birman (Consultants Bureau, London, 1967), Vol. 1, p. 69.
[10] H. Weyl, Math. Ann. **68**, 220 (1910).
[11] H. P. Baltes and E. Hilf, *Spectra of Finite Systems* (Bibliographisches Institute, Mannheim, 1976).
[12] C. Clark, SIAM Rev. **9**, 627 (1967).
[13] J. Odhnoff, Medd. Lunds. Univ. Mat. Sem **14**, 1 (1959).
[14] G. Bergendal, Medd. Lunds. Univ. Mat. Sem. **15**, 1 (1959).
[15] S. Agmon, Commun. Pure Appl. Math. **18**, 627 (1965).
[16] L. Hörmander, Acta.Math **121**, 193 (1968); in *Belfer Graduate School of Science Conference Proceedings: Some Recent Advances in the Basic Sciences*, edited by A. Gelbart (Yeshiva U. P., New York, 1969), Vol. 2, p. 155.
[17] T. A. Osborn and R. Wong, J. Math. Phys. **24**, 1487 (1983).
[18] L. Gårding, K. Fysiogr. Sallsk. Lund. Forh. **24**, 1 (1954);Math. Scand. **1**, 237 (1953).
[19] N. Dunford and J. T. Schwartz, *Linear Operators* (Interscience, New York, 1967), Part 1, pp. 160–162.
[20] W. Rudin, *Fourier Analysis on Groups* (Interscience, New York, 1961), Chap. 1.
[21] W. Rudin, *Real and Complex Analysis* (McGraw-Hill, New York, 1974).
[22] K. Ito, In *Proceedings of the Fourth Berkeley Symposium on Mathematics Statistics and Probability* (Univ. California Press, Berkeley, 1961), Vol. II, p. 227; K. Ito, in *Proceedings of the Fifth Berkeley Symposium on Mathematics Statistics, and Probability* (Univ. California Press, Berkeley, 1967), Vol. 2, part 1, p. 145.
[23] S. A. Albeverio and R. J. Høegh-Krohn, *Mathemetical Theory of Feynman Path Integrals* (Springer-Verlag, Berlin, 1976).
[24] T. Kato, *Perturbation Theory for Linear Operators* (Springer-Verlag,Berlin, 1966), Chap. 9.
[25] W. Amrein, J. Jauch and K. Sinha, *Scattering Theory in Quantum Mechanics* (Benjamin, Reading, MA, 1977), Chap. 3.
[26] H. Kitada, J. Fac. Sci. Univ. Tokyo Sec. 1A **27**, 193 (1980); Osaka J. Math. **19**, 863 (1982).
[27] H. Kitada and H. Kumanogo, Osaka J.Math. **18**, 291 (1981).
[28] D. Fujiwara, J. Math. Soc. Jpn. **28**, 483 (1976); J. Anal. Math. **35**, 41 (1979); Duke Math. J. **47**, 559 (1980).
[29] S. Zelditch, Commun. Math. Phys. **90**, 1 (1983).
[30] S. F. J. Wilk, Y. Fujiwara, and T. A. Osborn, Phys. Rev. A **24**, 2187 (1981).
[31] S. A. Fulling, SIAM J. Math. Anal. **14**, 780 (1983).
[32] Y. Fujiwara, T. A. Osborn, and S. F. J. Wilk, Phys. Rev. A **25**, 14 (1982).
[33] T. A. Osborn, "$\hbar$-Smooth Quantum Solutions: A Semiclassical Method," J. Phys. A (to appear).
[34] E. P. Wigner, Phys. Rev. **40**, 749 (1932).
[35] J. G. Kirkwood, Phys. Rev. **44**, 31 (1933).
[36] S. Mizohata, *The Theory of Partial Differential Equations* (Cambridge U.P., Cambridge, 1973), p. 43.
[37] A. Erdelyi, *Tables of Integral Transforms* (McGraw-Hill, New York, 1954), Vol. 1, p. 146.
[38] D. Gurarie, J. Diff. Eqs. **55**, 1 (1984).
[39] B. DeWitt, in *Relativity Group and Topology*, edited by C.DeWitt (Gordon and Breach, New York, 1964), p. 587–820.
[40] S. M. Christensen, Phys. Rev. D **14**, 2490 (1976).
[41] S. Agmon and Y. Kannai, Israel J. Math. **5**, 1 (1967).

# The gravitational field of a charged, magnetized, accelerating, and rotating mass

Alberto García Díaz[a] and Nora Bretón Baez
*Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Departamento de Física, Apartado Postal 14-740, 07000 México D.F., Mexico*

(Received 25 July 1984; accepted for publication 4 October 1984)

The explicit expression of a Petrov type *G* solution to the Einstein–Maxwell equations is given. This new solution is endowed with eight arbitrary parameters; mass, Newman–Unti–Tamburino (NUT) parameter, angular momentum, acceleration, electric and magnetic charges, and electric and magnetic field parameters.

## I. INTRODUCTION

In this paper we give the explicit form of a new stationary axially symmetric exact solution of the Einstein–Maxwell equations. This solution has been obtained by applying a Harrison transformation[1] to the most general type $\mathscr{D}$ electrovac solution in the coordinate form given by Plebański and Demiański.[2] Since the electrovac $\mathscr{D}$ solution possesses six free parameters and the Harrison transformation incorporates to the seed metric additional electric and magnetic field parameters, the generated new metric is endowed with eight arbitrary continuous parameters corresponding to mass, Newman–Unti–Tamburino (NUT) parameter (magnetic mass), angular momentum (Kerr parameter), acceleration, electric and magnetic charges, and electric and magnetic field parameters.

## II. THE METRIC ELEMENT

The metric can be expressed in the form

$$g=\Omega^2\left\{\frac{q^2+p^2}{(1-pq)^2}\left(\frac{dp^2}{P}+\frac{dq^2}{Q}\right)+f^{-1}\frac{PQ}{(1-pq)^4}\,d\tau^2\right\}-\Omega^{-2}[d\sigma-W\,d\tau]^2. \tag{1}$$

The basic structural functions $P$ and $Q$ are

$$\begin{aligned}P&=\gamma-\nu+2np-\kappa p^2+2mp^3-(\gamma+\nu)p^4,\\ Q&=\gamma+\nu-2mq+\kappa q^2-2nq^3-(\gamma-\nu)q^4,\end{aligned} \tag{2}$$

where $m$ is the mass, $n$ is the NUT parameter, $2\nu:=e^2+g^2$, $e$ is the electric charge, $g$ is the magnetic charge, and $\kappa$ and $\gamma$ are related with the angular momentum and acceleration parameters (see Ref. 2).

The secondary function $f$ is determined by

$$f=-(p^2+q^2)^{-1}(1-pq)^{-2}(q^4P-p^4Q). \tag{3}$$

The factor function $\Omega$ is defined as

$$\Omega^2:=\psi\bar{\psi},\quad \psi=1-2\bar{\alpha}\phi-\alpha\bar{\alpha}\mathscr{E}, \tag{4}$$

where the complex parameter $\alpha=E+iB$, with the constants $E$ and $B$ representing the additional electric and magnetic field correspondingly.

The functions $\phi$ and $\mathscr{E}$ in formula (4) are the complex Ernst potentials[3] of the seed $\mathscr{D}$ metric for the Killing direction $\partial_\sigma$. They are given by

$$\begin{aligned}\mathscr{E}=&f-2\nu p^2q^2/(p^2+q^2)+i(p^2+q^2)^{-1}(1-pq)^{-1}\\&\times[\gamma(1+pq)(p^2+q^2)+2\nu pq(p^2-q^2)\\&+2p^2q^2(nq-mp)],\end{aligned} \tag{5}$$

and

$$\phi=(e+ig)pq(p-iq)^{-1}. \tag{6}$$

At this point it is convenient to introduce the symbol $\delta$; $\delta:=E^2+B^2$.

The function $W$, whose determination entails considerable labor, is

$$\begin{aligned}W=&(q^4P-p^4Q)^{-1}\{(q^2P+p^2Q)+4(Eg-Be)p^3Q\\&-4(Ee+Bg)q^3P+6\nu\delta(q^4P+p^4Q)\\&+4\delta(Eg-Be)[\gamma q^3P+(\gamma+\nu)p^5Q\\&+(1-pq)^{-2}p^3q^2QP]\\&+4\delta(Ee+Bg)[\gamma p^3Q+(\gamma-\nu)q^5P\\&+(1-pq)^{-2}q^3p^2PQ]\\&+\delta^2(1-pq)^{-3}[MP+NQ+S]\},\end{aligned} \tag{7}$$

where the polynomials $M$, $N$, and $S$ are given by

$$\begin{aligned}M=&q^2\{-\gamma^2-\gamma(\gamma+4\nu)pq+4n\gamma q^3+8m\gamma pq^2\\&+(\gamma^2-\nu^2)q^4-4\kappa\gamma pq^3+3\nu(\nu+2\gamma)p^2q^2\\&-4n\gamma pq^4-6m(\gamma+\nu)p^2q^3+(\gamma-\nu)(\gamma-3\nu)pq^5\\&+3\kappa(\gamma+\nu)p^2q^4+[4mn-\nu(\nu+2\gamma)]p^3q^3\\&-4m^2p^4q^2+6n(\gamma-\nu)p^2q^5+2m(\gamma+\nu)p^3q^4\\&+2m\kappa p^4q^3+[4n^2-\kappa(\gamma+\nu)]p^3q^5\},\\N=&p^2\{-\gamma^2-\gamma(\gamma-4\nu)qp-4m\gamma p^3-8\gamma nqp^2\\&+(\gamma^2-\nu^2)p^4+4\kappa\gamma qp^3+3\nu(\nu-2\gamma)p^2q^2+4m\gamma p^4q\\&+6n(\gamma-\nu)q^2p^3+(\gamma+\nu)(\gamma+3\nu)qp^5-3\kappa(\gamma-\nu)q^2p^4\\&+[4mn-\nu(\nu-2\gamma)]p^3q^3-4n^2q^4p^2-6m(\gamma+\nu)q^2p^5\\&-2n(\gamma-\nu)q^3p^4+2\kappa nq^4p^3+[4m^2+\kappa(\gamma-\nu)]q^3p^5\},\\S=&(\gamma+\nu)\{2m(\gamma-\nu)-2m\kappa p^2-4mnp^2q+2n(\gamma+\nu)p^3q\\&-4m^2p^4q+2n\kappa p^3q^3+2m\kappa p^4q^2\}p^4q^3\\&+(\gamma-\nu)\{-2n(\gamma+\nu)-2n\kappa q^2-4mnq^2p\\&-2m(\gamma-\nu)q^3p-4n^2q^4p+2m\kappa p^3q^3+2n\kappa q^4p^2\}q^4p^3\\&+4m^2[2n+2mp^2-\kappa p^2q]p^5q^4\\&-4n^2[2m+2nq^2-\kappa pq^2]q^5p^4.\end{aligned} \tag{8}$$

The electromagnetic field is given by the two-form $\omega$,

$$\begin{aligned}\omega=&\tfrac{1}{2}(f_{\mu\nu}+\check{f}_{\mu\nu})dx^\mu\wedge dx^\nu\\=&[d\sigma-Wd\tau]\wedge d[\psi^{-1}[\phi+(E+iB)\mathscr{E}]]\\&+*\{[d\sigma-Wd\tau]\wedge d[\psi^{-1}[\phi+(E+iB)\mathscr{E}]]\},\end{aligned} \tag{9}$$

[a] Also at Sección de Graduados, Escuela Superior de Ingeniería Mecánica y Eléctrica, Instituto Politécnico Nacional, México D.F., Mexico.

 0022-2488/85/030465-02$02.50 © 1985 American Institute of Physics 

where * denotes the duality Hodge operator (see Ref. 3). Alternatively, the electromagnetic field tensor can be determined according to the formula (2.3) of Ref. 4, or by using the vector potential.[5]

The curvature quantities can be evaluated from the relations given in Ref. 6. Because of their length we do not include them in this report.

## ACKNOWLEDGMENT

One of us (N.B.B.) acknowledges the Consejo Nacional de Ciencias y Tecnología for financial assistance through a doctoral fellowship.

[1]B. K. Harrison, J. Math. Phys. **9**, 1744 (1968).
[2]J. F. Plebański and M. Demiański, Ann. Phys. (NY) **98**, 98 (1976).
[3]F. J. Ernst and J. F. Plebański, Ann. Phys. (NY) **107**, 266 (1977).
[4]F. P. Esposito and L. Witten, Gen. Relativ. Gravit. **6**, 387 (1975).
[5]F. J. Ernst, J. Math. Phys. **15**, 1409 (1974); **17**, 54 (1976).
[6]I. Hauser and F. J. Ernst, J. Math. Phys. **19**, 1316 (1978).

# Static and stationary multiple soliton solutions to the Einstein equations

Patricio S. Letelier
*Departamento de Física, Universidade de Brasília, 70.910-Brasília, Brazil*

(Received 10 May 1984; accepted for publication 3 August 1984)

The application of the Belinsky–Zakharov solution-generating technique, i.e., the inverse scattering method, to generate stationary axially symmetric solutions to the vacuum Einstein equations is reduced to a single quadrature when the seed solution is diagonal. The possibility of having real odd-number soliton solutions is investigated. These solutions represent solitonic perturbations of Euclidean metrics. The possibility of using instantons as seed solutions is also investigated. The one- and two-soliton solutions generated from a diagonal seed solution are studied. As an application, a unified derivation of some well-known static solutions, like the Schwarzschild metric and the Chazy–Curzon metric, as well as other new metrics is presented. By using these metrics as seed solutions, some known stationary solutions, like the Kerr-NUT metric, the double Kerr metric, and the rotating Weyl $C$-metric, as well as other new metrics are also derived in a unified way.

## I. INTRODUCTION

Recently, the inverse scattering method[1–3] (ISM) as well as Bäcklund transformations[4] (BT) have been applied with success to generate new stationary axially symmetric solutions to the vacuum Einstein equations. An example of new metric generated using these techniques is the $N$ Kerr metric.[5,6] Also the ISM and the BT have been used to better understand some known solutions, e.g., the Tomimatsu–Sato (TS) solution[7] with distortion parameter $\delta = 2$ can be interpreted as the coalescence of two Kerr metrics.[6,8]

The purpose of this paper is to use the Belinsky–Zakharov[2] version of the ISM to present a unified derivation of some known stationary solutions, like the Kerr–NUT metric,[9–11] the double Kerr metric,[6,8] and the rotating Weyl $C$-metric,[11,12] as well as other new solutions.[13] To derive the above-mentioned metrics we study a particular case of a diagonal multisoliton solution. Specializations of this solution are some well-known metrics, like the Schwarzschild metric, the Weyl $C$-metric,[14–16] and the Chazy–Curzon metric,[17,18] as well as other new metrics. Then, we use this diagonal $N$-soliton solution as a seed solution to generate new stationary one- and two-soliton solutions. Specializations of these solutions are the above-mentioned known stationary solutions.

In the application of the ISM to generate stationary solutions one finds mainly two difficulties. First, the ISM requires the explicit integration of an overdetermined system of linear partial differential equations. To find integrals that can be expressed in a closed form for the above-mentioned system of equations is not an easy task. Second, the algebraic complexity of the solutions seldom allows us to display their main features. In this paper we also study the overcoming of the above-mentioned difficulties for the special class of diagonal seed solutions.

In Sec. II we present a summary of the main formulas used in the ISM and we study the possibility of having odd-number soliton solutions. We arrive to the conclusion that we can have real solutions that can be interpreted as arising from solitonic perturbations of Euclidean solutions. In Sec. III we study the "Schrödinger equations" for the wave functions $\psi_0$, and in the diagonal case we reduce the integration of these equations, along the poles' trajectories, to a single quadrature. And we point out, since the ISM requires only the knowledge of $\psi_0$ along the poles' trajectories, the problem of finding soliton solutions is reduced to this single quadrature.[19]

In Secs. IV and V we examine the one-and two-soliton solutions generated from a diagonal seed solution, respectively. In particular we give compact formulas for these solutions.

In Sec. VI we study a particular case of diagonal $N$-soliton solution that contains as special cases the above-mentioned static metrics. Also, we find that odd-number solitons can be used to represent accelerated metrics. In Sec. VII we present particular cases of one- and two-soliton solutions generated using a diagonal $N$-soliton as a seed solution. Among these particular cases are all the previously mentioned stationary solutions. In Sec. VIII we briefly discuss the asymptotic and the singular behavior of the solutions generated using the ISM. Finally in the Appendix we study the possibility of using self-dual or anti-self-dual solutions to the Einstein equations on Euclidean space as seed solutions.[20]

## II. THE INVERSE SCATTERING METHOD

In this section we present the main formulas of the ISM used to solve the vacuum Einstein equations for the metric

$$ds^2 = e^{\sigma}(dr^2 + dz^2) + \gamma_{ab}\, dx^a\, dx^b\,, \tag{2.1}$$

where the indices $a$ and $b$ take the values 3 and 4. Here, $\gamma_{ab}$ and $\sigma$ are functions of $z$ and $r$ only $(\theta,t) \equiv (x^3,x^4)$. The Einstein equations for the metric (2.1) tell us that if the determinant of $\gamma_{ab}$ is different from a constant, without losing generality, it can be set equal to $\pm r^2$, i.e.,

$$\det \gamma = -\epsilon r^2\,, \tag{2.2}$$

where $\gamma$ is the $2\times 2$ matrix associated to $\gamma_{ab}$, and $\epsilon$ takes the value $+1$ for the usual axially symmetric metric with Lorentzian signature $(+\ +\ +\ -)$ and $-1$ for the axially symmetric metric with Euclidean signature $(+\ +\ +\ +)$. When (2.2) is satisfied, the rest of the Einstein equations can be cast in the form

$$(r\gamma_r\gamma^{-1})_r + (r\gamma_z\gamma^{-1})_z = 0\,, \tag{2.3}$$

 0022-2488/85/030467-10$02.50 © 1985 American Institute of Physics 

$$\sigma_r = -r^{-1} + (4r)^{-1}\,\mathrm{tr}(U^2 - V^2)\,, \tag{2.4a}$$

$$\sigma_z = (2r)^{-1}\,\mathrm{tr}(UV)\,, \tag{2.4b}$$

$$U \equiv r\gamma_r\gamma^{-1}\,, \quad V \equiv r\gamma_z\gamma^{-1}\,, \tag{2.5}$$

where the subscripts $r$ and $z$ denote partial differentiation. The condition of integrability of $\sigma$, i.e., $\sigma_{rz} = \sigma_{zr}$, is exactly Eq. (2.3), thus any solution to (2.3) will give us a $\sigma$ that can be obtained as a simple quadrature of Eqs. (2.4). Note that Eqs. (2.3)–(2.5) are completely equivalent to the corresponding equations of Ref. 2.

Soliton solutions to (2.3) are obtained by solving the "Schrödinger equations"

$$D_r\psi_0 = [(rU_0 + \lambda V_0)/(\lambda^2 + r^2)]\psi_0\,, \tag{2.6a}$$

$$D_z\psi_0 = [(rV_0 - \lambda U_0)/(\lambda^2 + r^2)]\psi_0\,, \tag{2.6b}$$

$$\psi_0|_{\lambda=0} = \gamma_0\,, \tag{2.6c}$$

$$D_r \equiv \partial_r + [2\lambda r/(\lambda^2 + r^2)]\partial_\lambda\,, \tag{2.7a}$$

$$D_z = \partial_z - [2\lambda^2/(\lambda^2 + r^2)]\partial_\lambda\,, \tag{2.7b}$$

for the wave function $\psi_0$. This wave function is a $2\times 2$ complex matrix function of $z$, $r$, and the spectral parameter $\lambda$. Here, $U_0$ and $V_0$ are obtained by replacing $\gamma$ in (2.5) by a known solution to (2.3), $\gamma_0$. The solution $\gamma_0$ is called the "seed" or "background" solution. The knowledge of $\psi_0$ allows us to find the new solution $\gamma$ to Eq. (2.3), given by[2]

$$\gamma_{ab} = (\gamma_0)_{ab} - \sum_{k,l} \frac{N_a^{(l)}(\Gamma^{-1})_{lk}N_b^{(k)}}{\mu_k\mu_l}\,, \tag{2.8}$$

$$\Gamma_{kl} = m_a^{(k)}(\gamma_0)_{ab}m_a^{(l)}/(r^2 + \mu_k\mu_l)\,, \tag{2.9}$$

$$N_a^{(k)} = m_b^{(k)}(\gamma_0)_{ba}\,, \tag{2.10}$$

$$m_a^{(k)} = m_{0b}^{(k)}M_{ba}^{(k)}\,, \tag{2.11}$$

$$M^{(k)} = \psi_0^{-1}|_{\lambda=\mu_k}\,, \tag{2.12}$$

$$\mu_k = \alpha_k - z + \epsilon_k|[(\alpha_k - z)^2 + r^2]^{1/2}|\,, \tag{2.13}$$

where the sum convention on the indices $a$ and $b$ has been adopted. The indices $k$ and $l$ run from 1 to $N$, $N$ being the number of solitons, i.e., the number of simple poles that appears in the "scattering matrix" used to find[2] (2.8)–(2.13). Here, $m_{0a}^{(k)}$ and $\alpha_k$ are sets of arbitrary constants and $\epsilon_k = \pm 1$. The only restrictions on $\gamma_0$ used to find (2.8) are[2] $\gamma_0 = \gamma_0^T$ and $\det\gamma_0 \neq 0$. Equation (2.8) tells us that $\gamma = \gamma^T$ and

$$\det\gamma = (-1)^N r^{2N}\prod_{k=1}^{N}\mu_k^{-2}\det\gamma_0\,. \tag{2.14}$$

Since the actual space-time metric satisfies the condition $\det\gamma = -r^2$, we can define a new matrix,

$$\gamma^{Ph} = r\gamma/(-\det\gamma)^{1/2}\,, \tag{2.15}$$

that is also a solution of (2.3) and satisfies the conditions $\gamma^{Ph} = (\gamma^{Ph})^T$ and

$$\det\gamma^{Ph} = -r^2\,, \tag{2.16}$$

whenever $\det\gamma < 0$. From (2.14) and the condition (2.2) for $\gamma_0$, we get

$$\det\gamma = \epsilon(-1)^{N+1}r^{2(N+1)}\prod_{k=1}^{N}\mu_k^{-2}\,. \tag{2.17}$$

Thus, we can fulfill the reality condition (2.16) by taking either an odd $N$ with $\epsilon = -1$ or an even $N$ with $\epsilon = 1$ (usual case). Hence, the odd-number soliton solutions can be interpreted as arising from solitonic perturbations of Euclidean metrics.[21,22] We shall return to this point in the next sections and in the Appendix.

The integration of (2.4) can be performed in an explicit way. We get

$$\sigma_N = \sigma_0 + \ln\left[r^{-N^2/2}\left(\prod_{k=1}^{N}\mu_k\right)^{N+1} \times \prod_{\substack{k,l=1\\ k>l}}^{N}(\mu_k - \mu_l)^{-2}\det\Gamma\right] + \ln C_N\,, \tag{2.18}$$

where $C_N$ is an arbitrary constant and the expression

$$\prod_{k,l=1}^{N}(\mu_k - \mu_l)^{-2}$$

should be set equal to 1 for $N = 1$.

## III. THE FUNCTION $\psi_0$

In this section we study the function $\psi_0$ solution to Eqs. (2.6) associated with the special class of diagonal metrics

$$ds^2 = e^{\sigma_0}(dr^2 + dz^2) + r^2e^{-\phi}\,d\theta^2 - \epsilon e^{\phi}\,dt^2\,, \tag{3.1}$$

where $\sigma_0$ and $\phi$ are functions of $z$ and $r$ only, and obey the equations

$$\phi_{rr} + \phi_r/r + \phi_{zz} = 0\,, \tag{3.2}$$

$$\sigma_0[\phi] = -\phi + \frac{1}{2}\int r[(\phi_r^2 - \phi_z^2)dr + 2\phi_r\phi_z\,dz]\,. \tag{3.3}$$

Note that (2.3) and (2.4) for the metric (3.1) are equivalent to (3.2) and (3.3). When $\epsilon = 1$, these solutions to the vacuum Einstein equations are known in the literature as Weyl solutions or Weyl metrics.[14]

Since the metric (3.1) is diagonal, one may assume that its associated function $\psi_0$ is also a diagonal matrix. With these assumptions, Eqs. (2.6) give us

$$(r\partial_r - \lambda\partial_z + 2\lambda\partial_\lambda)\det\psi_0 = 2\det\psi_0\,, \tag{3.4a}$$

$$(r\partial_z + \lambda\partial_r)\det\psi_0 = 0\,, \tag{3.4b}$$

$$\det\psi_0|_{\lambda=0} = -\epsilon r^2\,. \tag{3.4c}$$

A solution to (3.4) is

$$\det\psi_0 = \epsilon(-r^2 + \lambda^2 + 2\lambda z)\,. \tag{3.5}$$

A more general solution to (3.4) can be obtained adding $\epsilon c\lambda$ to the rhs of (3.5), where $c$ is an arbitrary constant. We have omitted such a term because in the final results it will only introduce a redefinition of arbitrary constants.

From the fact that $(\psi_0)_{34} = (\psi_0)_{43} = 0$ and Eq. (3.5), we conclude that there is not loss of generality in setting

$$(\psi_0)_{33} = (r^2 - \lambda^2 - 2\lambda z)\exp F\,, \tag{3.6a}$$

$$(\psi_0)_{44} = -\epsilon\exp(-F)\,. \tag{3.6b}$$

With this parametrization of $\psi_0$ the matrix equations (2.6) for the metric (3.1) reduce to the scalar equations

$$(r\partial_r - \lambda\partial_z + 2\lambda\partial_\lambda)F = r\phi_{,r}\,, \tag{3.7a}$$

$$(r\partial_z + \lambda\partial_r)F = r\phi_{,z}\,, \tag{3.7b}$$

$$F|_{\lambda=0} = \phi\,. \tag{3.7c}$$

The integrability condition of $F$ is just Eq. (3.2). Equations (3.7) are invariant under the transformation $z \to z + c$ and they are also linear. These two facts can be used to generate new solutions from known ones.[23]

In the final formulas (2.8)–(2.12) and (2.18) the matrix $\psi_0$ appears in the form $\psi_0|_{\lambda=\mu_k}$. Thus, to construct the soliton solutions we only need

$$F_k \equiv F|_{\lambda=\mu_k}, \tag{3.8}$$

i.e., the function $F$ along the poles' trajectories. These trajectories obey the equations[2]

$$\mu_{k,r} = 2r\mu_k/(\mu_k^2 + r^2), \quad \mu_{k,z} = -2\mu_k^2/(\mu_k^2 + r^2). \tag{3.9}$$

From (3.7)–(3.9) we get

$$r\,\partial_r F_k - \mu_k\,\partial_z F_k = r\phi_{,r}, \tag{3.10a}$$

$$\mu_k\,\partial_r F_k + r\,\partial_z F_k = r\phi_{,z}. \tag{3.10b}$$

Thus

$$F_k[\phi] = \frac{1}{2}\int \frac{r}{\mu_k}\left[(\mu_{k,r}\phi_{,r} - \mu_{k,z}\phi_{,z})dr + (\mu_{k,r}\phi_{,z} + \mu_{k,z}\phi_{,r})dz\right]. \tag{3.11}$$

The existence of (3.11) is guaranteed by Eq. (3.2) and the fact that $\ln \mu_k$ is also a solution to (3.2). Note that

$$\left.\frac{\mu_{k,r}}{\mu_k}\right|_{\mu_k \to 0} = \frac{2}{r}, \quad \left.\frac{\mu_{k,z}}{\mu_k}\right|_{\mu_k \to 0} = 0. \tag{3.12}$$

Thus, (3.11) is compatible with the initial condition (3.7c). In other words, the overdetermined system of equations (2.6) for diagonal matrices $U_0$, $V_0$, and $\psi_0$ is completely determined along the poles' trajectories; its solution reduces to a single quadrature.[19] Hence, in principle to any Weyl metric we can associate an $N$-soliton solution that, in general, will be stationary.

In the closely related method of BT we have a similar result, i.e., the application of this method, in the case of diagonal seed solutions, reduces to the finding of a single function.[24–26]

## IV. ONE-SOLITON SOLUTION

One-soliton solutions are defined as those solutions obtained using the ISM with a "scattering matrix" with one simple real pole at $\lambda = \mu_1$. When the seed solution is (3.1), the one-soliton solutions can be written in a simple form; from (2.8)–(2.13), (2.15), (2.17), (2.18), (3.1), (3.6), and (3.8), we find

$$\gamma_{33}^{Ph} = -\frac{p_1^2(r/\mu_1)Y_1^2 - q_1^2(\mu_1/r)Y_1^{-2}}{p_1^2 Y_1^2 + q_1^2 Y_1^{-2}}(r^2 e^{-\phi}), \tag{4.1a}$$

$$\gamma_{34}^{Ph} = -\frac{p_1 q_1 r(r/\mu_1 + \mu_1/r)}{p_1^2 Y_1^2 + q_1^2 Y_1^{-2}}, \tag{4.1b}$$

$$\gamma_{44}^{Ph} = -\frac{p_1^2(\mu_1/r)Y_1^2 - q_1^2(r/\mu_1)Y_1^{-2}}{p_1^2 Y_1^2 + q_1^2 Y_1^{-2}}(-e^{\phi}), \tag{4.1c}$$

$$\sigma_1 = \sigma_0 + \ln\left\{r\frac{p_1^2 Y_1^2 + q_1^2 Y_1^{-2}}{[r^2 + (\alpha_1 - z)^2]^{1/2}}\right\} + \ln C_1, \tag{4.2}$$

where

$$Y_k \equiv (r/\mu_k)^{1/2}\exp(F_k - \phi/2) \tag{4.3}$$

and $C_1$, $p_1$, and $q_1$ are arbitrary constants. These last two constants are related to $m_{0a}^{(1)}$ and $\alpha_1$ by

$$p_k \equiv -m_{03}^{(k)}/(2\alpha_k), \quad q_k \equiv m_{04}^{(k)}. \tag{4.4}$$

Equations (4.1) and (4.2) can be written in the more appealing form

$$\gamma_{33}^{Ph} = -[\sinh(y - x + \delta)/\cosh(y + \delta)](r^2 e^{-\phi}), \tag{4.5a}$$

$$\gamma_{34}^{Ph} = -\eta_1 r\cosh x/\cosh(y + \delta), \tag{4.5b}$$

$$\gamma_{44}^{Ph} = -[\sinh(y + x + \delta)/\cosh(y + \delta)](-e^{\phi}), \tag{4.5c}$$

$$\sigma_1 = \sigma_0 + \ln[r^{-1/2}\cosh(y + \delta)/\cosh x] + \ln C_1, \tag{4.6}$$

where the variables $x$ and $y$ are defined as

$$x \equiv \ln(\mu_1/r), \tag{4.7a}$$

$$y \equiv 2F_1 - \phi - \ln(\mu_1/r), \tag{4.7b}$$

and the constant $\delta$ and the sign function $\eta_1$ as

$$\tanh\delta = (1 - K_1^2)/(1 + K_1^2), \tag{4.8}$$

$$K_k \equiv q_k/p_k, \tag{4.9}$$

$$\eta_k \equiv p_k q_k/|p_k q_k|. \tag{4.10}$$

Also we have denoted the "renormalized" integration constant by the same symbol used in (4.2), a practice that we shall follow in this paper.

Note that the structure of the solution does not depend on the special form of the seed solution $\gamma_0$, as long as $(\gamma_0)_{43} = (\gamma_0)_{34} = 0$. In this case the ISM produces new solutions that, in general, will have two new essential parameters, $\delta$ (or $K_1$) and $\alpha_1$. Let me analyze the behavior of (4.5) for the special case of the Euclidean vacuum. In this case we have $\phi = F_1 = x + y = 0$. Thus $\gamma_{44}^{Ph}$ is the only component of $\gamma_{ab}^{Ph}$ that presents a localized behavior, i.e., it looks like a "bump." The name soliton used to describe these solutions can be only justified due to the method used in finding these solutions, i.e., the ISM. It is interesting to point out that the cylindrically symmetric solutions generated using the ISM present a clear soliton behavior.[27]

## V. TWO-SOLITON SOLUTIONS

Two-soliton solutions are defined as those solutions obtained using the ISM with a "scattering matrix" with two simple poles. In this case the poles are either real or complex conjugated. From (2.8)–(2.13), (2.17), (2.18), (3.1), (3.6), and (3.8) we find, after some algebra,

$$\gamma_{33}^{Ph} = \frac{[r(\mu_2 - \mu_1)P_1]^2 - [(r^2 + \mu_1\mu_2)P_2]^2}{[r(\mu_2 - \mu_1)S_1]^2 + [(r^2 + \mu_1\mu_2)S_2]^2}(r^2 e^{-\phi}), \tag{5.1a}$$

$$\gamma_{34}^{Ph} = -\frac{r(r^2 + \mu_1\mu_2)(\mu_2 - \mu_1)}{\mu_1\mu_2} \times \frac{p_2 q_2 \mu_1(r^2 + \mu_2^2)T_1 - p_1 q_1 \mu_2(r^2 + \mu_1^2)T_2}{[r(\mu_2 - \mu_1)S_1]^2 + [(r^2 + \mu_1\mu_2)S_2]^2}, \tag{5.1b}$$

$$\gamma_{44}^{Ph}=\frac{[r(\mu_2-\mu_1)Q_1]^2-[(r^2+\mu_1\mu_2)Q_2]^2}{[r(\mu_2-\mu_1)S_1]^2+[(r^2+\mu_1\mu_2)S_2]^2}(-e^{\phi}), \tag{5.1c}$$

$$\sigma_2=\sigma_0-\ln\{(r^2+\mu_1^2)(r^2+\mu_2^2)(r^2+\mu_1\mu_2)^2 \times(1/\mu_1-1/\mu_2)^2\}+\sigma_2'+\ln C_2, \tag{5.2}$$

where

$$\sigma_2'\equiv\ln\{[r(\mu_2-\mu_1)S_1]^2+[(r^2+\mu_1\mu_2)S_2]^2\}, \tag{5.3}$$

$$S_1\equiv p_1p_2Y_1Y_2-q_1q_2(Y_1Y_2)^{-1}, \tag{5.4a}$$

$$S_2\equiv p_1q_2(Y_1/Y_2)-q_1p_2(Y_2/Y_1), \tag{5.4b}$$

$$P_1\equiv p_1p_2(r^2/\mu_1\mu_2)^{1/2}Y_1Y_2 + q_1q_2(\mu_1\mu_2/r^2)^{1/2}(Y_1Y_2)^{-1}, \tag{5.5a}$$

$$P_2\equiv p_1q_2(\mu_2/\mu_1)^{1/2}(Y_1/Y_2) - q_1p_2(\mu_1/\mu_2)^{1/2}(Y_2/Y_1), \tag{5.5b}$$

$$T_1=(p_1y_1)^2-(q_1y_1^{-1})^2, \tag{5.6a}$$

$$T_2\equiv(p_2y_2)^2-(q_2y_2^{-1})^2, \tag{5.6b}$$

$$Q_1\equiv p_1p_2(\mu_1\mu_2/r^2)^{1/2}Y_1Y_2 + q_1q_2(r^2/\mu_1\mu_2)^{1/2}(Y_1Y_2)^{-1}, \tag{5.7a}$$

$$Q_2\equiv p_1q_2(\mu_1/\mu_2)^{1/2}(Y_1/Y_2) - q_1p_2(\mu_2/\mu_1)^{1/2}(Y_2/Y_1). \tag{5.7b}$$

The fact that the poles $\mu_1$ and $\mu_2$ are either real or complex conjugated can be used to simplify the previous formulas. For real $\mu_1$ and $\mu_2$ and real constants $p_1$, $q_1$, $p_2$, and $q_2$ such that

$$p_1p_2q_1q_2>0, \tag{5.8}$$

we find

$$\gamma_{33}^{Ph}=\frac{[r(\mu_2-\mu_1)\cosh(\xi_1+\delta_+)]^2-[(r^2+\mu_1\mu_2)\sinh(\xi_2+\delta_-)]^2}{[r(\mu_2-\mu_1)\sinh(x_1+\delta_+)]^2+[(r^2+\mu_1\mu_2)\sinh(x_2+\delta_-)]^2}(r^2e^{-\phi}), \tag{5.9a}$$

$$\gamma_{34}^{Ph}=-\frac{r(r^2+\mu_1\mu_2)(\mu_2-\mu_1)}{2\mu_1\mu_2}\frac{\eta_2\mu_1(r^2+\mu_2^2)\sinh(y_1+\omega_1)-\eta_1\mu_2(r^2+\mu_1^2)\sinh(y_2+\omega_2)}{[r(\mu_2-\mu_1)\sinh(x_1+\delta_+)]^2+[(r^2+\mu_1\mu_2)\sinh(x_2+\delta_-)]^2}, \tag{5.9b}$$

$$\gamma_{44}^{Ph}=\frac{[(r(\mu_2-\mu_1)\cosh(\xi_1+\delta_+)]^2-[(r^2+\mu_1\mu_2)\sinh(\xi_2+\delta_-)]^2}{[r(\mu_2-\mu_1)\sinh(x_1+\delta_+)]^2+[(r^2+\mu_1\mu_2)\sinh(x_2+\delta_-)]^2}(-e^{\phi}), \tag{5.9c}$$

$$\sigma_2'=\ln\{[r(\mu_2-\mu_1)\sinh(x_1+\delta_+)]^2+[(r^2+\mu_1\mu_2)\sinh(x_2+\delta_-)]\}, \tag{5.10}$$

where the variables $x_k$, $\xi_k$, $g_k$, and $\zeta_k$ are defined as

$$x_1\equiv\zeta_1-\tfrac{1}{2}\ln(\mu_1\mu_2/r^2), \tag{5.11a}$$

$$x_2\equiv\zeta_2+\tfrac{1}{2}\ln(\mu_2/\mu_1), \tag{5.11b}$$

$$\xi_1=\zeta_1-\ln(\mu_1\mu_2/r^2), \tag{5.12a}$$

$$\xi_2=\zeta_2+\ln(\mu_2/\mu_1), \tag{5.12b}$$

$$y_1=\zeta_1+\zeta_2-\ln(\mu_1/r), \tag{5.13a}$$

$$y_2=\zeta_1-\zeta_2-\ln(\mu_2/r), \tag{5.13b}$$

$$\zeta_1=F_1+F_2-\phi, \tag{5.14a}$$

$$\zeta_2=F_1-F_2, \tag{5.14b}$$

and the constants $\delta_\pm$ and $\omega_1$ and $\omega_2$ as

$$\tanh\delta_\pm=(K_2^{\mp1}-K_1)/(K_2+K_1), \tag{5.15}$$

$$\tanh\omega_k\equiv(1-K_k^2)/(1+K_k^2). \tag{5.16}$$

In the complementary case of (5.8), i.e.,

$$p_1p_2q_1q_2<0, \tag{5.17}$$

we find that the relations (5.9a), (5.9c), (5.10), and (5.15) keep the same form, but changing the hyperbolic functions by their respective cofunctions. The relations (5.11)–(5.14) and (5.16) remain the same. And the component $\gamma_{34}^{Ph}$ now reads

$$\gamma_{34}^{Ph}=-\frac{r(r^2+\mu_1\mu_2)(\mu_2-\mu_1)}{2\mu_1\mu_2}\frac{\eta_2\mu_1(r^2+\mu_2^2)\sinh(y_1+\omega_1)-\eta_1\mu_2(r^2+\mu_1^2)\sinh(y_2+\omega_2)}{[r(\mu_2-\mu_1)\cosh(x_1+\delta_+)]^2+[(r^2+\mu_1\mu_2)\cosh(x_2+\delta_-)]^2}. \tag{5.18}$$

In the case that $\mu_1$ and $\mu_2$ are complex conjugated, to end up with a real metric, we ought to choose constants $p_1$, $p_2$, $q_1$, and $q_2$ as follows:

$$p\equiv p_1=p_2^*, \quad q\equiv q_1=q_2^*. \tag{5.19}$$

It is also convenient to introduce the following notations:

$$\mu\equiv\mu_1=\mu_2^*, \quad F'\equiv F_1=F_2^*. \tag{5.20}$$

From (5.20), (5.21), and (5.1)–(5.7) we find

$$\gamma_{33}^{Ph}=\frac{[2r|\mu|\sin y_2\cosh(\xi_1+\delta_+)]^2-[(r^2+|\mu|^2)\sin(\xi_2+\delta_-)]^2}{[2r|\mu|\sin y_2\sinh(x_1+\delta_+)]^2+[(r^2+|\mu|^2)\sin(x_2+\delta_-)]^2}(r^2e^{-\phi}), \tag{5.21a}$$

$$\gamma_{34}^{Ph}=2r(r^2+|\mu|^2)\sin y_2\frac{(r^2+|\mu|^2)\sin y_1\cos y_2\cosh(x_1+\delta_+)+(r^2-|\mu|^2)\cos y_1\sin y_2\sinh(x_1+\delta_-)}{[2r|\mu|\sin y_2\sinh(x_1+\delta_+)]^2+[(r^2+|\mu|^2)\sin(x_2+\delta_-)]^2}, \tag{5.21b}$$

$$\gamma_{44}^{Ph}=\frac{[2r|\mu|\sin y_2\cosh(\zeta_1+\delta_+)]^2-[(r^2+|\mu|^2)\sin(\zeta_2+\delta_-)]^2}{[2r|\mu|\sin y_2\sinh(x_1+\delta_+)]^2+[(r^2+|\mu|^2)\sin(x_2+\delta_-)]^2}(-e^{\phi}), \tag{5.21c}$$

$$\sigma_2=\sigma_0+\sigma'-2\ln\{|r^2+\mu^2|(r^2+|\mu|^2)(\sin y_2)/|\mu|\}+\ln C_2, \tag{5.22}$$

$$\sigma'\equiv\ln\{[2r|\mu|\sin y_2\sinh(x_1+\delta_+)]^2+[(r^2+|\mu|^2)\sin(x_2+\delta_-)]^2\}, \tag{5.23}$$

where

$$x_1=\zeta_1-\ln(|\mu|/r), \tag{5.24a}$$

$$x_2=\zeta_2-y_2, \tag{5.24b}$$

$$y_1=x_2+\delta_-, \tag{5.25a}$$

$$y_2=\arg\mu, \tag{5.25b}$$

$$\zeta_1=2\,\mathrm{Re}\,F'-\phi, \tag{5.26a}$$

$$\zeta_2=2\,\mathrm{Im}\,F', \tag{5.26b}$$

and

$$\delta_+=(|p|^2-|q|^2)/(|p|^2+|q|^2), \tag{5.27}$$

$$\delta_-=\arg(pq^*). \tag{5.28}$$

Note that $x_1$, $x_2$, $y_1$, $y_2$, $\zeta_1$, and $\zeta_2$ are real variables and that $\delta_+$ and $\delta_-$ are real constants.

As in the one-soliton case the structure of the two-soliton solutions does not depend on the seed solution $\gamma_0$ particular form, as long as $(\gamma_0)_{34}=(\gamma_0)_{43}=0$.

Letting $\mu_1=\mu_2$ in (5.1)–(5.3), in the general case $p_1q_1p_2q_2\neq 0$ we find $\gamma^{Ph}=\gamma_0$ and $\sigma_2=\sigma_0+\ln C_2$, i.e., the two poles cancel out and we end up with the original seed solution $\gamma_0$. This result can be easily proved for a nondiagonal $\gamma_0$ using the formalism of Ref. 2. In the closely related method of Bäcklund transformations we have exactly the opposite behavior, i.e., the coincidence of "poles" can always be used to generate new solutions.[28] The case $\gamma_{34}^{Ph}=0$ will be studied in the next section.

The two-soliton solutions generated by the ISM will have, in general, four independent new real parameters; $\alpha_1$, $\alpha_2$ and any two of $\omega_1$, $\omega_2$, $\delta_+$, and $\delta_-$, because among these last four parameters there are only two independent ones, as a close examination of relations (5.15) and (5.16) indicates. And in the case of complex poles the new parameters are $\mathrm{Im}\,\alpha_1$, $\mathrm{Re}\,\alpha_1$, $\delta_+$, and $\delta_-$.

## VI. DIAGONAL $N$-SOLITON SOLUTIONS

Taking either $p_1=0$ or $q_1=0$ in (4.1) we get the degenerate one-soliton solution

$$\gamma_{33}^{Ph}=-\delta_1\left(\frac{\mu_1}{r}\right)^{-\delta_1}r^2e^{-\phi}, \tag{6.1}$$

$$\gamma_{34}^{Ph}=0,\quad \gamma_{44}^{Ph}=-r^2/\gamma_{33}, \tag{6.2}$$

where $\delta_k=\pm 1$. We introduce the term degenerate soliton solution for the diagonal solutions ($\gamma_{34}^{Ph}=0$) to indicate that these solutions obey a *linear* differential equation. Similarly choosing constants $p_k$ and $q_k$ such that $\gamma_{34}^{Ph}=0$ for the two-soliton solution (5.2), we get the degenerate two-soliton solution

$$\gamma_{33}=\delta_1\delta_2\left(\frac{\mu_1}{r}\right)^{-\delta_1}\left(\frac{\mu_2}{r}\right)^{-\delta_2}r^2e^{-\phi}, \tag{6.3}$$

where $\gamma_{44}^{Ph}$ is given by (6.2). A close examination of (5.1) shows that after doing $\gamma_{34}^{Ph}=0$, we can have, as a limit, $\mu_1=\mu_2$, i.e., the coalescence of two one-soliton solutions. Hence

$$\gamma_{33}^{Ph}=(\mu_1/r)^{-2\delta_1}r^2e^{-\phi}. \tag{6.4}$$

The above-mentioned limit is studied for a special class of stationary four-soliton solutions in Ref. 8.

The solution (6.3) can also be obtained by considering (6.1) as a seed solution for the same one-soliton solution (6.1). After repeating this procedure $N$ times we get

$$\gamma_{33}^{Ph}=\prod_{k=1}^{N}(-\delta_k)\left(\frac{\mu_k}{r}\right)^{-\delta_k}r^2e^{-\phi}, \tag{6.5}$$

where $\gamma_{34}^{Ph}$ and $\gamma_{44}^{Ph}$ are obtained as before.

The metric associated to this solution can be written as (3.1) with $\epsilon=1$ and a function $\phi$ defined by

$$\phi_N=-\ln(\gamma_{33}^{Ph}/r^2). \tag{6.6}$$

Thus

$$\phi_N=\sum_{k=1}^{N}\delta_k\ln\left(\frac{\mu_k}{r}\right)-\sum_k\ln\delta_k+\phi. \tag{6.7}$$

The function $\phi_N$ satisfies (3.2) for any value of the constants $\delta_k$. Strictly speaking, $\phi_N$ can only be considered as a degenerate soliton solution when $\delta_k=\pm 1$ but if one allows coalescence of solutions $\delta_k$ can be taken as an integer number.

Many well-known metrics are special cases of degenerate $N$-soliton solutions generated by the particular seed solution[29]

$$\phi = b \ln r + \sum_{k=1}^{N} \ln \delta_k \,, \tag{6.8}$$

where $b$ is an arbitrary constant. In this case we find

$$\phi_N = \sum_k \delta_k \ln\left(\frac{\mu_k}{r}\right) + b \ln r \,. \tag{6.9}$$

The computation of the metric function $\sigma_0$ associated to (6.9) can be performed directly; from (3.3) we get

$$\sigma_0[\phi_N] + \phi_N = \ln\left\{ C_0 r^{\beta^2/2} \prod_{i=1}^{N} \left[ \mu_i^{(2\delta_i+\beta)\delta_i} (r^2+\mu_i^2)^{-\delta_i^2} \times \prod_{\substack{j=1 \\ j\neq i}}^{N} (\mu_i - \mu_j)^{\delta_i\delta_j} \right] \right\}, \tag{6.10}$$

where

$$\beta = b - \sum_k^N \delta_k \,. \tag{6.11}$$

For $N=1$ the factor $(\mu_i - \mu_j)^{\delta_i\delta_j}$ that appears in (6.10) is taken to be 1. In computing (6.10) we have made use of the relations (3.9).

Since the metric associated to (6.8) is the Levi–Civita metric,[30] also known as the Kasner cylindrical metric,[30] the metric associated to (6.9) can be considered as a solitonic perturbation of this metric. A physical image of the solution associated to (6.9) can be obtained considering that the function $\phi$ can be related to the Newtonian potential $U$ by[31]

$$U = \phi/2 \,, \tag{6.12}$$

and that the Newtonian potential of an infinite wire of linear mass density $\lambda$ located on the $z$ axis is

$$U = 2\lambda \ln r \,. \tag{6.13}$$

Also, the Newtonian potential of semi-infinite wires of linear mass density $\lambda$ lying on the $z$ axis and located along $[\alpha_1, +\infty[$ and $[\alpha_1, -\infty[$ are, respectively,

$$U = \lambda \ln \mu_1^+ \,, \tag{6.14}$$

$$U = \lambda \ln \mu_1^- \,, \tag{6.15}$$

where we have introduced the notation

$$\mu_k^{\pm} = \mu_k|_{\epsilon_k = \pm 1} \,. \tag{6.16}$$

Thus, the solution (6.9) can be interpreted as arising from the superposition of $N$ semi-infinite wires of linear mass densities $\delta_1/2, \delta_2/2, \ldots, \delta_N/2$ located on the $z$ axis along $[\alpha_1, \infty\epsilon_1[$, $[\alpha_2, \infty\epsilon_2[, \ldots, [\alpha_N, \infty\epsilon_N[$, respectively, and another infinite wire of linear mass density $\beta = b - \Sigma_k \delta_k$ lying along the complete $z$ axis. The identity

$$\mu_k^+ \mu_k^- = -r^2 \tag{6.17}$$

can be used to represent a given distribution of wires in many different ways, e.g., an infinite wire can be represented as the superposition of two semi-infinite ones of the same linear mass density. Specific examples of solutions represented in terms of wires can be found in Refs. 31–33. Note that this interpretation is not without pitfalls since it is notorious that the spherically symmetric Schwarzschild solution transforms to the field of a rod in the cylindrical coordinates of the metric (3.1) with $\epsilon = 1$, as we shall see later in this section.

Now we shall study some particular cases of the diagonal metric (3.1) with $\epsilon = 1$ when the metric functions $\phi$ and $\sigma_0$ are given by (6.9) and (6.10), respectively. The special cases are obtained by assigning particular values to the parameters $\delta_k, \alpha_k, \epsilon_k$, and $b$ that appear in the definition of $\phi_N$ and $\sigma_0[\phi_N]$.

*Minkowski metric:* The specializations $b=0$ and $\delta_k = 0\,(k = 1,\ldots,N)$ yield the Minkowski metric in cylindrical coordinates.

*Accelerated metrics:* Either one of the specializations $\delta_k = 0\ (k = 1,\ldots,N)$ and $b=2$ or $\delta_1 = b = 1$ and $\delta_k = 0$ $(k = 2,\ldots,N)$ give uniformly accelerated flat metrics. The first case corresponds to the metric studied by Rindler[34] and the second is studied in Refs. 16 and 35. Note that in this last case the acceleration is represented by a semi-infinite wire of density $\lambda = \frac{1}{2}$.

*Weyl $\delta$-metric:* Either one of the specializations $\delta_1 = -\delta_2 = \delta$, $\delta_k = 0\ (k = 3,\ldots,N)$, $\epsilon_1 = \epsilon_2$, and $b = 0$ or $\delta_1 = \delta_2 = \delta$, $\delta_k = 0\,(k = 3,\ldots,N)$, $\epsilon_1 = -\epsilon_2$, and $b = 0$ produces the Weyl $\delta$-metric[36] also known as the Weyl $\gamma$-metric,[37] the Zipoy metric,[38] or the Zipoy–Voorhees metric.[39,40] This metric is the static limit of the TS solution[41] with distortion parameter $\delta$, that in the case $\delta = 1$ reduces to the Schwarzschild solution. The different "representations" of a given solution are a consequence of (6.17); from now on we shall indicate only one of such representations. Multiple $\delta$-metrics can be obtained choosing the parameters in the following way: $\delta_k = \delta_{k+1} = \delta^{(k)}$, $\epsilon_k = -\epsilon_{k+1}$ $(k = 1,3,5,N-1$; even $N)$ and $b = 0$. The particular case of $n = N/2$ Schwarzschild masses, i.e., $\delta_k = 1\,(k = 1,\ldots,N)$ has been studied by many authors.[2,29,31,32] In this case the mass of each particle is given by $m_k = (\alpha_{k+1} - \alpha_k)/2$ $(k = 1,3,5,\ldots,N-1)$.

*Chazy–Curzon metric:* The specialization $\delta_1 = -\delta_2 = m^1/\alpha$, $\epsilon^1 \equiv \epsilon_1 = \epsilon_2$, $\alpha_1 = \alpha^1 - \alpha$, $\alpha_2 = \alpha^1 + \alpha$, where $m^1$ and $\alpha^1$ are new constants, $b = \delta_k = 0\,(k = 3,\ldots,N)$, and the limit[32,35] $\alpha \to 0$ give the Chazy–Curzon metric that represents a single "particle" of mass $\epsilon^1 m^1$ located on the $z$ axis at $z = \alpha^1$. The potential $\phi_N$ in this case reduces to

$$\phi_N = -2\epsilon^1 m^1 / \sqrt{r^2 + (z-\alpha^1)^2} \,. \tag{6.18}$$

A metric representing $n = N/2$ Chazy–Curzon particles of masses $\epsilon^k m^k$ located on the $z$ axis at $z = \alpha^k$ can be obtained by choosing the parameters that characterize $\phi_N$ as $\epsilon^k \equiv \epsilon_k = \epsilon_{k+1}$, $\alpha_k = \alpha^k - \alpha$, $\alpha_{k+1} = \alpha^k + \alpha$, $\delta_k = -\delta_{k+1} = m^k/\alpha$ $(k = 1,3,5,N-1$, even $N)$, and $b = 0$, and letting $\alpha \to 0$. The case of multiple Chazy–Curzon metrics generated by particles of positive and negative masses $(\epsilon^k = \pm 1)$ has been widely studied.[17,32,42–44] Metrics representing $n = N/3$ accelerated Chazy–Curzon particles can be obtained doing the specialization $\epsilon^k \equiv \epsilon_k = \epsilon_{k+1}$, $\epsilon_{k+2} = \pm 1$, $\alpha_k = \alpha^k - \alpha$, $\alpha_{k+1} = \alpha^k + \alpha$, $\alpha_{k+2} = 0$, $\delta_k = -\delta_{k+1} = m^k/\alpha$, $\delta_{k+2} = 1$ $(k = 1,4,7,\ldots,N-2$; $N/3$ integer), and $b = n$, and letting $\alpha \to 0$ in $\phi_N$. The case $n = 2$ was first studied by Bonnor and Swaminarayan[35] (BS). The accelerated Chazy–Curzon metric can be generalized by the inclusion of a "distortion" parameter in the acceleration term, i.e., taking $\delta_{k+2}$ arbitrary and $b = \Sigma_{k=1}^{n} \delta_{k+2}$ instead of $\delta_{k+2} = 1$ and $b = n$.

*Weyl C-metric:* The specialization $\delta_1 = \delta_2 = \delta_3 = 1$, $\delta_k = 0$ $(k = 4,...,N)$, $b = 1$ yields the Weyl $C$-metric in the Bonnor[33]–Godfrey[45] coordinates whenever the parameters $\alpha_1$, $\alpha_2$, and $\alpha_3$ satisfy the cubic equation

$$2A^4X^3 - A^2X^2 + m^2 = 0 . \tag{6.19}$$

The new parameters $m$ and $A$ are identified as the particle mass and acceleration, respectively. Note that we have identified the Weyl $C$-metric as a three-soliton perturbation of the Euclidean version of the Weyl–Levi–Civita metric with $\phi = \ln r$. A two-soliton identification, however an indirect one, can be found in the paper quoted in Ref. 13. Metrics representing $n = N/3$ uniformly accelerated particles of mass $m_k$ and acceleration $A_k$ $(k = 1,4,7,...,N-2;\ N/3$ integer) can be obtained letting $\delta_k = \delta_{k+1} = \delta_{k+2} = 1$ and $b = n$ in $\phi_N$, and choosing $\alpha_k$, $\alpha_{k+1}$, and $\alpha_{k+2}$ as the roots of the cubic equation

$$2A_k^4X^3 - A_k^2X^2 + m_k^2 = 0 . \tag{6.20}$$

These metrics can be generalized by the inclusion of a distortion parameter $\delta^k$ in the following way: take $\delta_k = \delta_{k+1} = \delta_{k+2} = \delta^{(k)}$ $b = \Sigma_{k=1}^{n} \delta^k$ and $\alpha_k$, $\alpha_{k+1}$, and $\alpha_{k+2}$ as before. The case of integer $\delta^k$ can be thought of as arising from the coalescence of $\delta^k$ equal accelerated particles. Recently Plebański and García D.[46] discussed a metric that contains a distorted Weyl $C$-metric as a particular case, alas the coordinates used by the above-mentioned authors are completely different to ours. Thus the relation between the above-mentioned particular case with our $n = 1$ case cannot be easily established. We have not been able to find in the literature multiple Weyl $C$-metrics.

New metrics can be also obtained doing the above-mentioned specializations at once, i.e., we can have a metric representing the superposition of $n_\delta$ Weyl $\delta$-metrics, $n_{cc}$ Chazy–Curzon metrics, and so on. Also, we have not been able to find in the literature such a mixed system of "particles." Furthermore, taking an arbitrary $b$ we can superimpose an infinite wire on the $z$ axis. Finally, we want to point out that we have not included in our analysis of particular cases the constant $C_0$ that appears in (6.10). In general, for each particular metric this constant has a particular form that depends on the values of $\delta_k$, $\alpha_k$, etc. This is a consequence of the fact that almost all the metrics presented here were first obtained in completely different systems of coordinates. A discussion of this point can be found in Refs. 32 and 33 and in the paper quoted in Ref. 13.

## VII. STATIONARY MULTIPLE-SOLITON SOLUTIONS

The degenerate $N$-soliton solution studied in the preceding section can again be used as a seed solution to generate new one- and two-soliton solutions to the vacuum Einstein equations. To generate these new solutions we only need to compute the functions $Y_k$ associated to $\phi_N$ as the expressions (4.1)–(4.3) and (5.1)–(5.7) indicate.

Equations (6.8) and (6.11) tell us that the seed solution $\phi_N$ can be written as

$$\phi_N = \sum_{k=1}^{N} \delta_k \ln \mu_k + \beta \ln r . \tag{7.1}$$

From (7.1) and (3.11), with the help of (3.9), we obtain

$$F_J = \frac{1}{2}\beta \ln \mu_j + \sum_{k=1}^{N} \delta_k \ln(\mu_j - \mu_k) , \tag{7.2}$$

where the index $j$ takes the values I for the one-soliton solution and I and II for the two-soliton solution. In computing (7.2) we assume $\mu_j \neq \mu_k$. Finally (7.1), (7.2), and (4.3) yield

$$Y_j = \left(\frac{\mu_j}{r}\right)^{(\beta-1)/2} \prod_{k=1}^{N} \left[\, \mu_k^{-\delta_k/2} (\mu_j - \mu_k)^{\delta_k} \right] . \tag{7.3}$$

Thus, changing the index 1 by I and the function $\phi$ by $\phi_N$ in (4.1)–(4.3) we get a one-soliton solution associated to (7.1). Similarly, replacing 1 by I, 2 by II, and $\phi$ by $\phi_N$ in (5.1)–(5.7), we obtain the two-soliton solution associated to (7.1). Of course, in both cases the function $\sigma_0$ that appears in (4.2) and (5.2) is the one associated to (7.1), i.e., the function $\sigma_0[\phi_N]$ given by (6.10). The one- and two-soliton solutions constructed with (7.1) actually are an $(N+1)$- and an $(N+2)$-soliton solution, respectively. Let me introduce the notation $(N+\mathrm{I})$-soliton and $(N+\mathrm{II})$-soliton to indicate solitons that are formed by the superposition of a degenerate (diagonal) $N$-soliton and either a I or a II nondiagonal soliton. We shall study special cases of $(N+\mathrm{I})$- and $(N+\mathrm{II})$-soliton solutions that are obtained by particularizing the function $\phi_N$.

*(0 + I)-soliton solutions:* The one-soliton solution constructed from (7.1)–(7.3) with $\delta_k = 0$ $(k = 1,...,N)$ is a one-solitonic perturbation of the Euclidean version, of the Kasner cylindrical metric. This $(0+\mathrm{I})$-soliton can be also considered as the elliptic version of a metric studied in Ref. 1. Three special cases are particularly interesting, the cases $b = 2$, $b = 1$, and $b = 0$. In the first case, we have a solitonic perturbation to the Euclidean version of the Rindler metric, in the second a solitonic perturbation of the other accelerated metric described in Sec. VI, and in the third a solitonic perturbation to the Euclidean metric. This last particular case may be called a half-Kerr–NUT since the Kerr–NUT metric is a $(0+\mathrm{II})$-soliton with $b = 0$ as we shall see later in this section. In the general case the $(0+\mathrm{I})$-soliton solution has only one new parameter $K_\mathrm{I}$ (the parameter $\alpha_\mathrm{I}$ can be eliminated by a $z$-axis translation). From the discussion of Sec. VI we can say that in the general case $(b \neq 0)$ the $(0+\mathrm{I})$-soliton solution represents an infinite rotating wire formed by two semi-infinite wires of different constant linear mass densities lying on the $z$ axis along $[\alpha_\mathrm{I}, \infty[$ and $[\alpha_\mathrm{I}, -\infty[$. The static limit of this metric was studied by Godfrey[45] in the search of Weyl metrics with homothetic motions.

*(1 + I)-soliton solutions:* The one-soliton solution constructed from (7.1)–(7.3) with $\delta_1 = 1$, $\delta_k = 0$ $(k = 2,...,N)$, $\epsilon_1 = -\epsilon_\mathrm{I}$, and $b = 0$ is the Kerr metric. This identification is studied in some detail in Ref. 2. The present case admits the simple generalization $\delta_+ \equiv \delta$ (arbitrary) and $b = \delta - 1$. Note that this solution is not simple related to the TS class of solutions since the distortion parameter enters only in "half" of the solution.

*(2 + I)-soliton solutions:* The one-soliton solutions constructed from (7.1)–(7.3) admit three important particular cases that are obtained by the following specializations of the parameters that define $\phi_N$.

(a) Take $\delta_1 = \delta_2 = 1, \delta_k = 0 (k = 3,...,N), b = 1$, and $\alpha_1$, $\alpha_2$, and $\alpha_\mathrm{I}$ as the roots of the cubic equation (6.19). This case represents the rotating version of the Weyl $C$-metric studied by Kinnersley.[10,12]

(b) Take $\delta_1 = \delta_2 = \delta$, $\delta_k = 0$ $(k = 3,...,N)$, $\epsilon_1 = -\epsilon_2$,

and $b=1$. This case represents a rotating Weyl $\delta$-metric with acceleration parameter. Note that this metric is not simple related to the TS family of solutions since letting the parameter $\alpha_I \to 0$ and $K_I \to 0$ we do not end up with the Weyl $\delta$-metric. Also it can be considered as a generalization of the precedent case. A similar metric was studied by Hoenselaers, Kinnersley, and Xanthopoulos[24] (HKS) using a HKS transformation.[47]

(c) Take $\alpha_1=\alpha^1-\alpha$, $\alpha_2=\alpha^1+\alpha$, $\epsilon_1=\epsilon_2$, $\delta_1=-\delta_2=m^1/\alpha$, $b=1$, and $\delta_k=0$ $(k=3,...,N)$, and let $\alpha \to 0$. This case represents a rotating Chazy–Curzon metric with acceleration parameter, i.e., a rotating version of a one-center BS metric. By the addition of two more degenerate solitons and doing a similar specialization of the parameters $\alpha_3$, $\alpha_4$, $\delta_3$, $\delta_4$, etc., we shall have a rotating version of the usual two-center BS metric.

*(0 + II)-soliton solutions:* The two-soliton solution (5.1)–(5.3) constructed from (7.1)–(7.3) with $\delta_k=0$ $(k=1,...,N)$ is a two-soliton perturbation of the Kasner cylindrical metric. In the case $b=0$ (Minkowski background) this metric is equivalent with the Kerr–NUT metric. This equivalence can be proved directly, i.e., by change of variables.[2]

*(1 + II)-soliton solutions:* The two-soliton solution constructed from (7.1)–(7.3) with $\delta_1=1$, $\delta_2=0$ $(k=2,...,N)$, $b=1$, and constants $\alpha_1$, $\alpha_I$, and $\alpha_{II}$ solutions to Eq. (6.19) represents a rotating uniformly accelerated particle with usual "electric" mass as well as "magnetic" mass[48] (NUT parameter), i.e., the generalization of the Weyl $C$-metric studied by Kinnersly in a different system of coordinates. This rotating "Weyl–NUT $C$-metric" can be generalized by the addition of a "distortion" parameter in the acceleration term, i.e., taking $\delta_1=b=\delta$ instead of $\delta_1=b=1$.

*(2 + II)-soliton solutions:* The two-soliton solution constructed from (7.1)–(7.3) with $\delta_k=0$ $(k=3,...,N)$ admits three significant particular cases. Two of them can be described as the superposition of a Kerr–NUT metric with either a Weyl $\delta$-metric or a Chazy–Curzon metric and the third as a double Kerr metric, i.e., as a $(1+I)+(1+I)$-soliton solution. The parameters in the first two cases are taken as indicated in Sec. VI. Due to the relation that exists between the double rank-zero HKX transformation and the Belinsky–Zakharov two-soliton transformation[49] we conclude that the two-soliton solution generated from the Weyl $\delta$-metric is closely related to metrics studied by Cosgrove[50] and Dietz and Hoenselaers.[51,52] Similarly the two-soliton solution generated from the Chazy–Curzon metric is closely related to another solution studied in Ref. 52. We shall return to this point at the end of this section.

*(3 + II)-soliton solutions:* The two-soliton solution constructed from (7.1)–(7.3) with $\delta_k=0$ $(k=4,...,N)$ admits a relevant particular case that can be described as the superposition of a "distorted" Weyl $C$-metric[46] and a Kerr–NUT metric. When the distortion parameter $\delta$ is taken to be 1, we have the superposition of a usual Weyl $C$-metric and a Kerr–NUT metric. In this last case we can think of the $(3+II)$-soliton solution as a $(2+I)+(1+I)$-soliton solution, i.e., we have the superposition of a Kerr metric with a rotating Weyl $C$-metric. To have such particular cases we ought to specialize the parameters $\alpha_I$, $\delta_1$, etc., as indicated in the $(2+I)$- and the $(1+I)$-soliton solutions already studied.

*(4 + II)-soliton solutions:* The two-soliton solution constructed from (4.1)–(4.3) with $\delta_k=0$ $(k=5,...,N)$ admits three significant special cases that can be characterized as, a $(2+2+II)$-soliton, a $3+(1+II)$-soliton, and a $(2+I)+(2+I)$-soliton. The first case can represent the superposition of the Kerr–NUT metric with any two of the degenerate two-soliton solutions described in Sec. III, i.e., the Weyl $\delta$-metric and the Chazy–Curzon metric, depending on the particular values assigned to the different parameters that characterize each particular solution. Similarly by choosing the sets of parameters $(\alpha_1,\alpha_2,\alpha_3)$ and $(\alpha_4,\alpha_I,\alpha_{II})$ satisfying Eq. (6.19); $\delta_1=\delta_2=\delta_3=\delta$ and $b=\delta$ we have that the $3+(1+II)$-soliton case represents the superposition of a distorted Weyl $C$-metric with the rotating Weyl–NUT $C$-metric mentioned before. The last and more interesting case is the $(2+I)+(2+I)$-soliton solution. Specializing the parameters that characterize the function $\phi_N$ as the $(2+I)$-soliton case we have that this metric can describe the superposition of two equal or different metrics as the rotating Weyl $C$-metric, the rotating Weyl $\delta$-metric with acceleration, and the one-center rotating BS metric. A space-time described as the superposition of two one-center rotating BS metrics, i.e., a two-center BS metric, was studied by Dietz and Hoenselaers[51] using HKX transformations.

Any of the above-mentioned metrics can be easily generalized by considering an arbitrary $b$, i.e., we can superimpose an infinite wire lying on the $z$ axis. By choosing the parameters that define $\phi_N$ as indicated in Sec. VI, we can also perform a superposition of $n_\delta$ Weyl $\delta$-metrics, $n_{cc}$ Chazy–Curzon metrics, and $n_c$ Weyl $C$-metrics with any of the stationary metrics described in this section.

The solutions generated using the ISM and HKH transformations are closely related, since the space of solutions generated by these two soliton generating techniques are equivalent.[49] Unhappily, this relation does not give information about particular solutions. The actual identification of solutions obtained using the ISM and the HKX transformation is performed in the usual way, i.e., finding a coordinate transformation that shows the desired equivalence. In general to find this coordinate transformation is not an easy task, since in the actual applications of the ISM and HKX transformations one takes advantage of a completely different system of coordinates.

Finally, we want to point out that in this paper we have focused our attention only on solutions generated using the ISM with (7.1) as a seed solution. For solutions obtained using either a different soliton-solution-generating technique or a different seed solution or both, see Refs. 23, 25, and 53.

## VIII. DISCUSSION

Two important aspects of the solutions presented in the paper are the asymptotic behavior and the existence of singularities on the symmetry axis that do not appear in the potential $\phi_N$.

In general, the even-number soliton ISM maps asymp-

totically flat diagonal solutions into asymptotically flat solutions. Cosgrove proved the following theorem[49]: If the seed solution is such that

$$\delta_N|_{R\to\infty} \sim K + M/R\,, \tag{8.1}$$

where $M$ and $K$ are constants and $R^2 = r^2 + z^2$, then the new solution obtained using the ISM is asymptotically flat (provided that certain weak conditions are satisfied). Unhappily the condition (8.1) is too restrictive to include many of the metrics studied in the present paper. In general, the accelerated metrics do not satisfy (8.1), e.g., the Weyl $C$-metric, which is also an odd soliton. Of course, there exist in the literature more sophisticated definitions of asymptotic flatness[54,55] that are satisfied by accelerated metrics, e.g., the Weyl $C$-metric satisfies the Ashtekar AEFANSI definition.[56] Due to the peculiarities of these definitions and the ISM a great amount of work needs to be done before we can formulate a theorem that relates the ISM with a definition of asymptotic flatness like AEFANSI.

Usually, metrics with an infinite wire type singularity are not considered as asymptotically flat metrics, although some of them present asymptotically flat behavior at $r\to\infty$. See, for instance, the first citation in Ref. 53. In general, if one takes a seed solution that is "asymptotically flat" at $r\to\infty$, the metrics obtained using the ISM also have the same behavior at $r\to\infty$. For the metrics studied in this paper we can always remove an infinite line of singularities by adding a wire of opposite mass density, i.e., by choosing a suitable $b$. Studying the odd-number soliton solution presented in this paper we see that an odd-number soliton ISM will generate an infinite line of singularities and in consequence will produce a nonasymptotically flat solution. In the metrics presented in Secs. VI and VII these infinite lines of singularities were eliminated using the already described method, otherwise we noticed the presence of such wires.

The problem of the existence of singularities on the symmetry axis that do not appear in $\phi_N$ goes back to 1936 when Silverstein[42] criticized, erroneously, the general relativity for allowing a solution representing two Chazy–Curzon masses in equilibrium. Einstein and Rosen[43] found that the metric is singular on the line that joins both masses, i.e., there is a strut keeping both masses apart. The condition of regularity on the axis of symmetry, also known as the elementary flatness condition, has been studied by many authors.[32,35,57] In particular, Bondi considered masses of both signs in order to eliminate the struts.[44]

Recently, the problem of regularity on the symmetry axis was examined for metrics representing two[58,59] and $N$ Kerr particles,[60] and two rotating Chazy–Curzon-like particles.[51] In the last case it was found that the spin–spin interaction can keep apart two positive masses rotating in the same direction.[51] For the case of two positive Kerr particles it seems that the spin–spin interaction is not enough to balance the gravitational attraction.[59]

Due to the simple functional form of the one- and two-soliton solutions presented in Secs. IV and V one may use these expressions to compute the curvature invariants and study their singular behavior. We believe that computing these invariants could present an interesting example in checking the practical efficiency of the algebraic computer programs for the Einstein equations.[61]

## APPENDIX: INSTANTONS WITH AXIAL SYMMETRY

In this appendix we study the possibility of using self-dual or anti-self-dual solutions to the vacuum Einstein equations as seed solutions. The metric

$$ds^2 = (\omega^0)^2 + (\omega^1)^2 + (\omega^2)^2 + (\omega^3)^4\,, \tag{A1}$$

$$\omega^0 \equiv (Rf)^{1/2}(dt + W\,d\theta)\,,\quad \omega^1 \equiv (R/f)^{1/2}\,d\theta\,, \tag{A2}$$

$$\omega^2 \equiv e^k\,dr\,,\quad \omega^3 \equiv e^k\,dz\,,$$

where $k$, $R$, $f$, and $W$ are functions of $r$ and $z$, represents a "stationary" axially symmetric Euclidean metric. The condition that the Riemann Christoffel tensor be self-dual or anti-self-dual, as well as one that the Ricci tensor be zero, is implemented by[20]

$$\omega_{01} = \eta\omega_{23}\,,\quad \omega_{02} = \eta\omega_{31}\,,\quad \omega_{03} = \eta\omega_{12}\,, \tag{A3}$$

where $\omega_{01}$, $\omega_{23}$, etc., are the connection one-forms and $\eta = +1$ for self-dual solutions (instantons) and $\eta = -1$ for anti-self-dual solutions (anti-instantons). From (A2) and (A3) we get

$$k_z = -\tfrac{1}{2}\eta f W_r\,,\quad k_r = \tfrac{1}{2}\eta f W_z\,, \tag{A4}$$

$$R_r/R + f_r/f = -\eta f W_z\,,\quad R_z/R - f_z/f = -\eta f W_r\,, \tag{A5}$$

$$R_r/R - f_r/f = \eta f W_z\,,\quad R_z/R + f_z/f = \eta f W_r\,. \tag{A6}$$

From (A5) and (A6) we get

$$R = 1\,, \tag{A7}$$

$$W = \eta\int f^{-2}(f_z\,dr - f_r\,dz)\,, \tag{A8}$$

and from (A8) and (A4) we have

$$k = -\tfrac{1}{2}\ln f\,, \tag{A9}$$

$$(f^{-1})_{rr} + (f^{-1})_{zz} = 0\,. \tag{A10}$$

Note that $f^{-1}$ and $W$ are harmonic conjugated functions, i.e., $f^{-1} = \mathrm{Re}\,H(\tau)$ and $W = \mathrm{Im}\,H(\tau)$ or $f^{-1} = \mathrm{Im}\,H(\tau)$ and $W = \mathrm{Re}\,H(\tau)$, where $H$ is an arbitrary function of $\tau = r + iz$. In (A8) and (A9) we have omitted the integration constants, because in the metric they can be eliminated by a trivial change of variables, by the same reason we have set $R = 1$. Thus,

$$ds^2 = f(dt + W\,d\theta)^2 + (dr^2 + dz^2 + d\theta)^2/f\,. \tag{A11}$$

First we note that this is a special case of the Hawking metric,[62]

$$ds^2 = V^{-1}(\mathbf{x})(d\tau + \boldsymbol{\omega}\cdot d\mathbf{x})^2 + V(\mathbf{x})d\mathbf{x}\cdot d\mathbf{x}\,, \tag{A12}$$

since (A12) reduces to (A11) letting

$$V(\mathbf{x}) = 1/f(x^1,x^2)\,, \tag{A13}$$

$$\omega(\mathbf{x}) = [0,0,W(x^1,x^2)]\,. \tag{A14}$$

From (A11) and (A8) we conclude that the only diagonal axisymmetric instanton is the Euclidean vacuum. One can easily generalize the previous result by considering the metric

$$ds^2 = (A_0\,dx^0)^2 + (A_1\,dx^1)^2 + (A_2\,dx^2)^2 + (A_3\,dx^3)^2\,, \tag{A15}$$

with $A_0, A_1, A_2$, and $A_3$ functions of $x^0, x^1, x^2$, and $x^3$. Also, in this case the only metric that is a solution to (A3) is the Euclidean metric. Since the axisymmetric instantons cannot be cast as (3.1) we cannot use them as seed solutions in the context of the present paper. The nondiagonal metric (A11) can be used to generate an odd number of soliton solutions in the general case (nondiagonal case), since (A11) also obeys the key equation used in the ISM, i.e., Eq. (2.3). The instantons characterized by (A11) have the following "anomalies": $\det \gamma = 1$ and the equations for $k$ cannot be cast as (2.4). These two "anomalies" will force us to change Eqs. (2.17) and (2.18).


[1]V. A. Belinsky and V. E. Zakharov, Zh. Eksp. Teor. Fiz. **75**, 1955 (1978) [Sov. Phys. JETP **48**, 985 (1978)].
[2]V. A. Belinsky and V. E. Zakharov, Zh. Eksp. Teor. Fiz. **77**, 3 (1979) [Sov. Phys. JETP **50**, 1 (1979)].
[3]I. Hauser and F. J. Ernst, Phys. Rev. D **20**, 365 and 1783 (1979); J. Math. Phys. **21**, 1126 and 1418 (1980); **22**, 1051 (1981).
[4]B. K. Harrison, Phys. Rev. Lett. **41**, 1197 (1978); G. Neugebauer, J. Phys. A **12**, L67 (1979); **13**, L19 (1980).
[5]G. A. Alekseev, Dokl. Akad. Nauk SSSR **256**, 827 (1981) [Sov. Phys. Dokl. **26**, 158 (1981)]; M. Yamazaki, Phys. Rev. Lett. **50**, 1027 (1983).
[6]D. Kramer and G. Neugebauer, Phys. Lett. A **75**, 259 (1980).
[7]A. Tomimatsu and H. Sato, Phys. Rev. Lett. **29**, 1344 (1972).
[8]A. Tomimatsu, Prog. Theor. Phys. **63**, 1054 (1980).
[9]M. Demianski and E. T. Newman, Bull. Acad. Pol. Sci. **24**, 653 (1966).
[10]W. Kinnersley, J. Math. Phys. **10**, 1195 (1969).
[11]C. Reina and A. Treves, J. Math. Phys. **16**, 834 (1975), and references therein.
[12]J. F. Pelbański and M. Demianski, Ann. Phys. (N.Y.) **98**, 98 (1976).
[13]A unified derivation of some well-known diagonal metrics using the ISM can be also found in R. T. Jantzen, Gen. Relativ. Gravit. **15**, 115 (1983).
[14]H. Weyl, Ann. Phys. **54**, 117 (1917).
[15]J. Ehlers and W. Kundt, in *Gravitation, an Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962).
[16]W. Kinnersley and M. Walker, Phys. Rev. D **2**, 1359 (1970).
[17]M. J. Chazy, Bull. Soc. Math. France **52**, 17 (1924); H. E. J. Curzon, Proc. London Math. Soc., Ser. 2 **23**, 477 (1924).
[18]D. Kramer, H. Stephani, M. MacCallum, and E. Herlt, *Exact solutions of Einstein's field equations* (Cambridge U.P., Cambridge, 1980), Chap. 18.
[19]A similar study for the hyperbolic case can be found in P. S. Letelier, J. Math. Phys. **25**, 2675 (1984).
[20]For a review of self-dual solutions to Euclidean gravity, see R. Eguchi and A. J. Hanson, Ann. Phys. (N.Y.) **120**, 82 (1979).
[21]A similar extension of the ISM was used in P. S. Letelier, J. Math Phys. **23**, 1175 (1982) in the context of SU(2) gauge fields.
[22]The relation between Euclidean one-solitons ($N = 1$ and $\epsilon = 1$) and Harrison BT can be found in C. M. Cosgrove, J. Math. Phys. **21**, 2417 (1980).
[23]P. S. Letelier, "Multipole stationary soliton solutions to the Einstein equations," Rev. Bras. Fis. (to appear); P. S. Letelier, "Solitary wave solutions to the Einstein equations," J. Math. Phys. (to appear).
[24]C. Hoenselaers, J. Kinnersley, and B. C. Xanthopoulos, J. Math. Phys. **20**, 2530 (1979).
[25]B. K. Harrison, Phys. Rev. D **21**, 1695 (1980).
[26]The relation between the different solution generating techniques can be found in the paper quoted in Ref. 22 and in C. M. Cosgrove, J. Math. Phys. **23**, 615 (1982).
[27]V. Belinsky and V. Zakharov, in *Sources in Gravitational Radiation*, edited by L. Smarr (Cambridge U.P., Cambridge, 1979); V. Belinsky and D. Fargion, Nuovo Cimento B **59**, 143 (1980); J. Ibañez and E. Verdaguer, Phys. Rev. Lett. **51**, 1313 (1983); B. J. Carr and E. Verdaguer, Phys. Rev. **28**, 2995 (1983).
[28]K. Oohara and H. Sato, Prog. Theor. Phys. **61**, 1891 (1981), and references therein.
[29]Diagonal $N$-soliton solutions are studied in G. A. Alekseev and V. A. Belinsky, Zh. Eksp. Teor. Fiz. **78**, 1297 (1980) [Sov. Phys. JETP **51**, 655 (1980)]. See also the paper quoted in Ref. 13 (case $\phi = 0$); P. S. Letelier, Rev. Bras. Fis. **13**, 698 (1983) (case $\phi \neq 0$).
[30]See, for instance, Ref. 18.
[31]See, for instance, H. P. Robertson and T. W. Noonan, *Relativity and Cosmology* (Saunders, Philadelphia, 1969), Sec. 10.4.
[32]W. Israel and K. A. Khan, Nuovo Cimento **33**, 331 (1964).
[33]W. B. Bonnor, Gen. Relativ. Gravit. **15**, 535 (1983).
[34]W. Rindler, Am. J. Phys. **34**, 1174 (1966).
[35]W. B. Bonnor and N. S. Swaminarayan, Z. Phys. **177**, 240 (1964).
[36]R. Bach and H. Weyl, Math. Z. **13**, 134 (1922).
[37]B. W. Stewart, D. Papadopoulos, L. Witten, R. Berezdivin, and L. Herrera, Gen. Relativ. Gravit. **14**, 97 (1983).
[38]D. M. Zipoy, J. Math. Phys. **7**, 1137 (1966).
[39]B. H. Voorhees, Phys. Rev. D **2**, 2119 (1970).
[40]This metric has a long history, part of which is cited in Ref. 37.
[41]A. Tomimatsu and H. Sato, Prog. Theor. Phys. **50**, 95 (1973).
[42]L. Silberstein, Phys. Rev. **49**, 268 (1936).
[43]A. Einstein and N. Rosen, Phys. Rev. **49**, 404 (1936).
[44]H. Bondi, Rev. Mod. Phys. **29**, 423 (1957); B. Hoffman, in *Les Théories Relativistes de la Gravitation* (CNRS, Paris, 1962); D. J. Booth, Int. J. Theor. Phys. **14**, 361 (1975).
[45]B. B. Godfrey, Gen. Relativ. Gravit. **3**, 3 (1972).
[46]J. F. Plebański and A. García D., Phys. Rev. Lett. **48**, 1447 (1982).
[47]C. Hoenselaers, W. Kinnersley, and B. Xanthopoulos, Phys. Rev. Lett. **42**, 481 (1979).
[48]E. Newman, L. Tamburino, and T. Unti, J. Math. Phys. **4**, 915 (1963); C. W. Misner, J. Math. Phys. **4**, 924 (1963); S. Ramaswamy and A. Sen, J. Math. Phys. **22**, 2612 (1981); A. Ashtekar and A. Sen, J. Math. Phys. **23**, 2168 (1982).
[49]See the paper quoted in Ref. 22.
[50]C. M. Cosgrove, in *Proceedings of the Second Marcel Grossmann Meeting in General Relativity*, edited by R. Ruffini (North-Holland, Amsterdam, 1982).
[51]W. Dietz and C. Hoenselaers, Phys. Rev. Lett. **48**, 778 (1982).
[52]W. Dietz and C. Hoenselaers, Proc. R. Soc. London Ser. A **382**, 221 (1982).
[53]P. S. Letelier, Phys. Rev. D **26**, 3728 (1982); E. Kyriapoulos, Lett. Math. Phys. **5**, 311 (1981); Gen. Relativ. Gravit. **12**, 989 (1980). Also, see the last of the quoted papers in Ref. 29.
[54]A. Ashtekar, in *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980), Vol. 2.
[55]R. Geroch, in *Asymptotic Structure of Space-time*, edited by F. P. Esposito and L. Witten (Plenum, New York, 1977).
[56]T. Dray, Gen. Relativ. Gravit. **14**, 109 (1982).
[57]P. Bergman, *Introduction to the Theory of Relativity* (Prentice–Hall, Englewood Cliffs, NJ, 1942), pp. 206–210; J. L. Synge, *Relativity: the General Theory* (North-Holland, Amsterdam, 1966), pp. 313–317. See also the book quoted in Ref. 31.
[58]A. Tomimatsu and M. Kihara, Prog. Theor. Phys. **67**, 1406 (1982); M. Kihara and A. Tomimatsu, Prog. Theor. Phys. **67**, 349 (1982).
[59]A. Tomimatsu, Prog. Theor. Phys. **70**, 385 (1983).
[60]M. Yamazaki, Prog. Theor. Phys. **69**, 503 (1983).
[61]See, for instance, J. P. Fitch, "Camal user's manual," unpublished; R. A. d'Inverno, in *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980).
[62]S. W. Hawking, Phys. Lett. A **60**, 81 (1977).

# Some simple type $D$ solutions to the Einstein equations with sources

G. F. Torres del Castillo
*Departamento de Física Matemática, Instituto de Ciencias de la Universidad Autónoma de Puebla, Puebla, Mexico and Departamento de Física, Centro de Investigación y de Estudios Avanzados del IPN, Apartado Postal 14-740, 07000 México, D. F. Mexico*

J. F. Plebański[a)]
*Departamento de Física, Centro de Investigación y de Estudios Avanzados del IPN, Apartado Postal 14-740, 07000 México, D. F. Mexico*

(Received 1 June 1984; accepted for publication 3 August 1984)

Starting from a particular metric of the Kerr–Schild form, we find solutions to the Einstein equations coupled to a Weyl field, to an electromagnetic field, to a nonlinear (Born–Infeld) electromagnetic field, to a Yang–Mills field, and to a cosmological constant. These solutions can be superposed to construct others with any combination of the sources considered.

## I. INTRODUCTION

This paper is concerned with exact solutions, of the Kerr–Schild form, to the Einstein field equations with sources. Some of the solutions presented here have been previously known; however, our main purpose is to show that, in a certain sense, these solutions can be superposed. All the solutions given here are of type $D$ and have four Killing vectors and a Killing–Yano tensor; they reduce in the case of vacuum to one of the Newman–Unti–Tamburino (NUT) metrics.

Another contribution of this paper is to show that the proposed metric can be considered as produced by a nonlinear electromagnetic field of the Born–Infeld type. It may be remarked that there are quite few known exact solutions to the Einstein–Born–Infeld equations and that, even in flat space-time, the solutions to the Born–Infeld equations are rather scarce.

Most of this paper uses the null tetrad formalism as described in Ref. 1. In the discussion of the Weyl equation we also employ the spinorial formalism; a brief exposition of the necessary background is given here with the purpose of stating the notation and conventions to be used. For a review concerning the general properties of the Einstein–Weyl equations the reader is referred to the article by Kuchowicz.[2] We also give a very concise exposition of the basic facts about the Born–Infeld nonlinear electrodynamics. A more general and detailed discussion on nonlinear electrodynamics in general relativity can be found in Alarcón Gutiérrez *et al.*[3] and the references cited therein.

We shall consider space-times whose metric is of the Kerr–Schild form, that is, space-times whose metric can be written in the form

$$g = 2\,d\zeta\,d\bar{\zeta} + 2\,du\,dv + 2hk^2, \tag{1.1}$$

where $u$ and $v$ are real coordinates, $\zeta$ is a complex coordinate, $\bar{\zeta}$ denotes its complex conjugate, $h$ is a real function, and $k = k_\mu\,dx^\mu$ is null ($g^{\mu\nu}k_\mu k_\nu = 0$). Since $k$ is also null with respect to the flat metric $\eta = 2d\zeta\,d\bar{\zeta} + 2\,du\,dv$, it can be expressed as

$$k = du + \bar{Y}\,d\zeta + Yd\bar{\zeta} - Y\bar{Y}\,dv, \tag{1.2}$$

where $Y$ is a complex function.

The metric (1.1) can be written as $g = 2e^1e^2 + 2e^3e^4$, with

$$e^1 = d\zeta - Y\,dv, \quad e^2 = \overline{e^1}, \quad e^3 = k, \quad e^4 = dv + hk. \tag{1.3}$$

The tangent tetrad $\partial_a$, defined through $e^a(\partial_b) = \delta^a_b$, is then given by

$$\partial_1 = \partial_\zeta - \bar{Y}\partial_u, \quad \partial_2 = \overline{\partial_1}, \quad \partial_3 = \partial_u - h\partial_4,$$
$$\partial_4 = k^\mu\frac{\partial}{\partial x^\mu} = \partial_v + Y\partial_\zeta + \bar{Y}\partial_{\bar{\zeta}} - Y\bar{Y}\partial_u. \tag{1.4}$$

We will restrict ourselves to the case in which

$$Y = \zeta/(v + ia), \tag{1.5}$$

where $a$ is a real constant and $h$ is a function of $v$ only. With this choice for $Y$ the curves which have $\partial_4$ as tangents form a shear-free congruence of null geodesics (with respect to both metrics, $g$ and $\eta$). This congruence, for $a \neq 0$, is what has been called a Robinson congruence; a geometrical description of it has been given by Penrose.[4] This twisting congruence can be considered as a geometrical representation of a twistor.

The connection one-forms for the tetrad (1.2)–(1.4) are then given by[1]

$$\Gamma_{42} = -Ze^1,$$
$$\Gamma_{12} + \Gamma_{34} = [h_{,v} + (\bar{Z} - Z)h\,]e^3, \tag{1.6}$$
$$\Gamma_{31} = Zhe^2,$$

where

$$Z = -\Gamma_{421} = (v + ia)^{-1}. \tag{1.7}$$

The nonvanishing components of the curvature are determined by

$$R_{12} = -(Z + \bar{Z})^2[h/(Z + \bar{Z})]_{,v},$$
$$R_{34} = -Z^2([h_{,v} + (\bar{Z} - Z)h\,]/Z^2)_{,v}, \tag{1.8}$$
$$C^{(3)} = R/6 + 2Z\,[h_{,v} + (\bar{Z} - Z)h\,],$$

[a)] On leave of absence from the University of Warsaw, Warsaw, Poland.

 0022-2488/85/030477-05$02.50 © 1985 American Institute of Physics 

where $R = 2(R_{12} + R_{34})$ is the scalar curvature and $C^{(3)}$ characterizes the conformal curvature. If $C^{(3)} \neq 0$, then the metric is of type $D$. The vector fields $\partial_3$ and $\partial_4$ are geodesic and shear-free and, when $C^{(3)} \neq 0$, they point along the two double principal null directions of the Weyl tensor.

From the contracted Bianchi identities we find

$$R_{34} = \frac{Z\bar{Z}}{Z+\bar{Z}}\left(\frac{R_{12}}{Z\bar{Z}}\right)_{,4}. \tag{1.9}$$

Similarly, if $T_{12}$ and $T_{34}$ are the only nonvanishing independent components of the energy-momentum tensor, then from the identities $T_{ab}{}^{;a} = 0$, it follows that

$$T_{12} = \frac{Z\bar{Z}}{Z+\bar{Z}}\left(\frac{T_{34}}{Z\bar{Z}}\right)_{,4}, \tag{1.10}$$

thus, the equation $R_{12} = 8\pi T_{34}$ implies $R_{34} = 8\pi T_{12}$. Therefore, the Einstein equations reduce to a single ordinary differential equation.

In the cases considered below, it turns out that the components of the energy-momentum tensor of the matter fields are independent of $h$; therefore, $h$ is determined by an ordinary linear differential equation. Furthermore, the field equations of the sources happen to be solvable without specifying $h(v)$, even though $h$ appears in them. This fact, together with the linearity of the Ricci tensor on $h$, implies that by adding the $h$ corresponding to each source we obtain a solution to the Einstein field equations coupled to any combination of that matter fields considered here (neglecting any interaction between the various matter fields). In the general case of the Kerr–Schild geometry, Gürses and Gürsey[5] have shown that, in an appropriate coordinate system, the Einstein tensor takes a linear form in $g_{\mu\nu}$.

Under the present assumptions, independent of the explicit form of $h(v)$, the metric has the four Killing vectors

$$\begin{aligned} K_1 &= \partial_u, \\ K_2 &= v(\partial_\zeta + \partial_{\bar{\zeta}}) - (\zeta + \bar{\zeta})\partial_u + ia(\partial_\zeta - \partial_{\bar{\zeta}}), \\ K_3 &= iv(\partial_\zeta - \partial_{\bar{\zeta}}) + i(\zeta - \bar{\zeta})\partial_u - a(\partial_\zeta + \partial_{\bar{\zeta}}), \\ K_4 &= i(\zeta\partial_\zeta - \bar{\zeta}\partial_{\bar{\zeta}}). \end{aligned} \tag{1.11}$$

Furthermore, the skew-symmetric tensor field $f_{ab}$ whose only nonvanishing independent components are $f_{12} = -iv$ and $f_{34} = a$ is a Killing–Yano tensor[6]; hence, $Q_{ab} = f_{ac}f_b{}^c$ is a Killing tensor.

## II. INTEGRATION OF THE FIELD EQUATIONS

In this section we solve the equations for the gravitational field coupled to various matter fields.

### A. Einstein–Weyl equations

The Weyl equation for the (two-component) neutrino field is given by

$$\nabla^{A\dot{B}}\Psi_A = 0, \tag{2.1}$$

where $\Psi_A$ $(A = 1,2)$ are complex functions which represent the neutrino field and $\nabla^{A\dot{B}}$ is the covariant derivative along the vector field $\nabla^{A\dot{B}} = g^{a A\dot{B}}\partial_a$ $(A = 1,2;\ \dot{B} = \dot{1},\dot{2})$, with $\partial^{A\dot{B}}\partial^{C\dot{D}} = -2\epsilon^{AC}\epsilon^{\dot{B}\dot{D}}$, where $\epsilon^{AC}$ and $\epsilon^{\dot{B}\dot{D}}$ are Levi-Cività symbols, and $\partial^{\overline{A\dot{B}}} = \partial^{B\dot{A}}$. The covariant derivative of a spinor field $\Phi_{A\dot{B}\ldots}$ can be obtained from $\nabla_a\phi_{A\dot{B}\ldots} = \phi_{A\dot{B}\ldots,a} - \omega^C{}_A(\partial_a)\phi_{C\dot{B}\ldots} - \omega^{\dot{C}}{}_{\dot{B}}(\partial_a)\phi_{A\dot{C}\ldots} - \cdots$, where $\omega^C{}_A$ and $\omega^{\dot{C}}{}_{\dot{B}} = \overline{\omega^C{}_B}$ denote the spinorial connection one-forms. We shall choose the functions $g^{a A\dot{B}}$ as $g^{1\dot{1}2} = g^{2\dot{2}1} = g^{3\dot{1}1} - g^{4\dot{2}2} = \sqrt{2}$ and all others equal to zero. Then, the spinorial connection one-forms are given by

$$(\omega^A{}_B) = \begin{pmatrix} \frac{1}{2}(\Gamma_{12} + \Gamma_{34}) & \Gamma_{31} \\ -\Gamma_{42} & -\frac{1}{2}(\Gamma_{12} + \Gamma_{34}) \end{pmatrix}. \tag{2.2}$$

The gravitational field equations for interacting gravitational and matter fields are

$$R_{ab} - \tfrac{1}{2}Rg_{ab} = -8\pi T_{ab}, \tag{2.3}$$

where $T_{ab}$ is the energy-momentum tensor of the matter. In the case of the neutrino field

$$\begin{aligned} T_{ab} = (i\hbar/8)[&g_a{}^{A\dot{B}}(\Psi_{\dot{B}}\nabla_b\Psi_A - \Psi_A\nabla_b\Psi_{\dot{B}}) \\ &+ g_b{}^{A\dot{B}}(\Psi_{\dot{B}}\nabla_a\Psi_A - \Psi_A\nabla_a\Psi_{\dot{B}})], \end{aligned} \tag{2.4}$$

where $\Psi_{\dot{A}} = \overline{\Psi_A}$.

Using Eqs. (1.6) and (2.2) we obtain the Weyl equation in explicit form

$$\begin{aligned} &(\partial_4 + Z)\Psi_2 - \partial_1\Psi_1 = 0, \\ &\partial_2\Psi_2 + [\partial_3 - \tfrac{1}{2}(h_{,v} + (Z + \bar{Z})h)]\Psi_1 = 0. \end{aligned} \tag{2.5}$$

In order to have $T_{44} = T_{42} = T_{22} = 0$, as required by (2.3) and (1.8), we shall take $\Psi_1 = 0$. Then, from (2.5) it follows that

$$\Psi_2 = AZ, \tag{2.6}$$

where $A$ is a complex function such that $A_{,4} = A_{,2} = 0$. Computing the remaining components $T_{ab}$, we find that when $A$ is a complex constant $A_0$, the only nonvanishing components are given by

$$T_{12} = -T_{34} = -(a\hbar/2\sqrt{2})|A_0Z^2|^2. \tag{2.7}$$

Solving now for $h$, from the field equation $R_{12} = 8\pi T_{34}$ we obtain

$$h = \tfrac{1}{2}m(Z + \bar{Z}) + \sqrt{2}\hbar\pi a|A_0Z|^2, \tag{2.8}$$

where $m$ is a real constant.

Setting $A_0 = 0$, the metric defined by (2.8) becomes a solution of the Einstein vacuum field equations which is one of the NUT solutions. In fact, the coordinate transformation

$$\begin{aligned} x^1 &= -u - \zeta\bar{\zeta}v/(v^2 + a^2), \\ x^2 &= v, \\ x^3 + ix^4 &= \sqrt{2}\,\zeta/(v + ia), \end{aligned} \tag{2.9}$$

brings the metric derived above to the form given by Newman *et al.*[7] for their metric with $\mu^0 = 0$.

The constant $a$ is related to the NUT parameter; however, in the present case, by rescaling the coordinates $u$ and $v$ and the constants $m$ and $A_0$, $a$ can be reduced to one of the three values $-1,0,1$. When $a = 0$ but $A_0 \neq 0$, the neutrino field becomes a ghost field [see Eq. (2.7)] and the metric is a vacuum solution, which is a limiting case of the Schwarzschild metric corresponding to infinite mass.[8] In this case there exists a homothetic Killing vector given by

$$H = 2v\partial_v + 4u\partial_u + 3\zeta\partial_\zeta + 3\bar{\zeta}\partial_{\bar{\zeta}} \tag{2.10}$$

and there is a singularity for the metric and the neutrino field along $v = 0$.

The solution (2.6), (2.8) was found by Díaz[9] while searching for solutions to the Einstein–Weyl equations of the Kerr–Schild form. The branch with $a = 1$ was obtained by Kolassis[10] who was searching for stationary axially symmetric solutions. In the form given by him, it is not clear how to make the neutrino field amplitude to vanish.

Due to the symmetry between the two principal null directions, one can expect to have an analogous solution with $\Psi_2 = 0$. Indeed, assuming $\Psi_2 = 0$, from (2.5) it follows that

$$\Psi_1 = A\,(Z\bar{Z}/h)^{1/2}, \tag{2.11}$$

where $A$ is a complex function such that $A_{,3} = A_{,1} = 0$, is a solution of Weyl's equation. If we restrict the function $A$ to be a complex constant $A_0$ then the only nonvanishing components of the energy-momentum tensor are given by Eq. (2.7) and, therefore, the solution of the gravitational field equations is that given by (2.8).

### B. Einstein–Maxwell equations

The energy-momentum tensor of the electromagnetic field is given by

$$4\pi T_{ab} = F_{ac}F_b{}^c - \tfrac{1}{4}F_{cd}F^{cd}g_{ab}, \tag{2.12}$$

where $F_{ab} = -F_{ba}$ represents the electromagnetic field which must fulfill the Maxwell equations. In order to satisfy the gravitational field equations with $R_{ab}$ given in (1.8), $F_{12}$ and $F_{34}$ must be the only nonvanishing independent components of the Maxwell field. Then, the nonzero components of $T_{ab}$ are determined by

$$T_{12} = -T_{34} = (1/8\pi)|F_{12} + F_{34}|^2, \tag{2.13}$$

and the solution of Maxwell's equations is found to be

$$F_{12} + F_{34} = C_0 Z^2, \tag{2.14}$$

where $C_0$ is a complex constant.

Then, from the equation $R_{12} = 8\pi T_{34}$ we find that

$$h = \tfrac{1}{2}m(Z + \bar{Z}) - \tfrac{1}{2}|C_0 Z|^2, \tag{2.15}$$

where $m$ is a real constant. This solution is implicit in the results obtained by Debney *et al.*[1]

When $a < 0$, the solution (2.15) is identical to (2.8). Thus, when $a < 0$, the gravitational field represented by (2.15) can be thought of as produced by a neutrino field [(2.6) or (2.11)] or by an electromagnetic field [(2.14)].

### C. Cosmological constant

A cosmological constant $\lambda$ can also be easily included. A straightforward integration of $R_{12} = -\lambda$ yields

$$h = \tfrac{1}{2}m(Z + \bar{Z}) + \tfrac{1}{2}\lambda Z\bar{Z}(v^4/3 + 2a^2v^2 - a^4). \tag{2.16}$$

This solution to the Einstein vacuum field equations with cosmological constant was previously obtained by Kowalczyński and Plebański.[11]

### D. Einstein–Born–Infeld equations

In the Born–Infeld nonlinear electrodynamics the electromagnetic field is described by two skew-symmetric tensor fields $F_{ab}$ and $P_{ab}$ which, in absence of sources, must fulfill the equations

$$\check{F}^{ab}{}_{;a} = 0, \quad P^{ab}{}_{;a} = 0, \tag{2.17}$$

where $\check{F}^{ab}$ denotes the dual of $F_{ab}$. We shall use the convention $\check{F}^{ab} = \frac{1}{2}\epsilon_{abcd}F^{cd}$, where $\epsilon_{abcd}$ is the Levi-Cività symbol with $\epsilon_{1234} = 1$. Then, the dual of a real tensor is pure imaginary. The fields $F_{ab}$ and $P_{ab}$ are related by the "constitutive equations"

$$F_{ab} = \left(\frac{\partial \mathscr{H}}{\partial P}\right)P_{ab} + \left(\frac{\partial \mathscr{H}}{\partial \check{Q}}\right)\check{P}_{ab}, \tag{2.18}$$

where

$$\mathscr{H} = b^2 - (b^4 - 2b^2P + \check{Q})^{1/2} \tag{2.19}$$

is the Born–Infeld structural function, $b$ is a positive real constant, and

$$P = \tfrac{1}{4}P_{ab}P^{ab}, \quad \check{Q} = \tfrac{1}{4}P_{ab}\check{P}^{ab}, \tag{2.20}$$

are the invariants of $P_{ab}$; $P$ is real and $\check{Q}$ is pure imaginary. In the limit when $b \to \infty$ one recovers the linear theory.

The energy-momentum tensor of the nonlinear electromagnetic field is given by

$$4\pi T_{ab} = \left(\frac{\partial \mathscr{H}}{\partial P}\right)(P_{ac}P_b{}^c - Pg_{ab}) + \left(\mathscr{H} - \left(\frac{\partial \mathscr{H}}{\partial P}\right) - \check{Q}\left(\frac{\partial \mathscr{H}}{\partial \check{Q}}\right)\right)g_{ab}. \tag{2.21}$$

Thus, in order to satisfy the gravitational field equations with the Ricci tensor given by (1.8), $P_{12}$ and $P_{34}$ must be the only nonzero independent components of $P_{ab}$. Expressing these components in the form

$$P_{12} = ib\frac{\sin\phi}{\cosh\psi}, \quad P_{34} = b\frac{\cos\phi}{\sinh\psi}; \quad 0 \leqslant \phi < 2\pi, \quad 0 \leqslant \psi < \infty, \tag{2.22}$$

from Eq. (2.18) it follows that

$$F_{12} = ib\frac{\sin\phi}{\sinh\psi}, \quad F_{34} = b\frac{\cos\phi}{\cosh\psi}. \tag{2.23}$$

Substituting (2.22) and (2.23) into (2.17) we obtain that $\phi$ and $\psi$ are functions of $v$ only which must satisfy the conditions

$$(Z\bar{Z}\sinh\psi)_{,v} = 0, \quad \phi_{,v} = i(Z - \bar{Z})\tanh\psi. \tag{2.24}$$

Hence,

$$\sinh\psi = b/\rho_0 Z\bar{Z}, \tag{2.25}$$

where $\rho_0$ is a real constant and

$$\phi = \phi_0 + 2a\int_0^v \left[\left(\frac{\rho_0}{b}\right)^2 + (s^2 + a^2)^2\right]^{-1/2} ds, \tag{2.26}$$

with $\phi_0 = \text{const}$.

The fields $F_{ab}$ and $P_{ab}$ can be expressed in a simple and invariant way through the (complex) two-form

$$\omega = \tfrac{1}{2}(F_{ab} - \check{P}_{ab})e^a \wedge e^b. \tag{2.27}$$

The real and imaginary parts of $\omega$ correspond, respectively, to $F_{ab}$ and (minus) $\check{P}_{ab}$ and Eqs. (2.17) amount to

$$d\omega = 0. \tag{2.28}$$

Thus, there exists, locally, a potential one-form $\alpha$ such that

$\omega = -d\alpha$. When $a \neq 0$, from (2.24) and (2.25) we find

$$\omega = -d\left(\left(\rho_0 e^{i\phi}/2ia\right)e^3\right) \tag{2.29}$$

while in the case where $a = 0$, we get

$$\omega = -\rho_0 e^{i\phi_0} d\left\{\frac{1}{2}\left[\frac{\bar{\zeta}}{v}\,d\left(\frac{\zeta}{v}\right) - \frac{\zeta}{v}\,d\left(\frac{\bar{\zeta}}{v}\right)\right] - \int_v^\infty \left[\left(\frac{\rho_0}{b}\right)^2 + s^4\right]^{-1/2} ds\, e^3\right\}. \tag{2.30}$$

In the limit when $b \to \infty$ the fields $F_{ab}$ and $P_{ab}$ are equal and they coincide with the solution to the Maxwell equations given in (2.14).

Turning now to the integration of the gravitational field equations we find

$$R_{12} = 2b^2\{1 - [1 + (\rho_0 Z\bar{Z}/b)^2]^{1/2}\}, \tag{2.31}$$

then, by a straightforward computation, we get

$$h = \tfrac{1}{2} m(Z + \bar{Z}) - \rho_0{}^2 Z\bar{Z} v \int_v^\infty \frac{ds/s^2}{1 + \{1 + [\rho_0/b\,(s^2 + a^2)]^2\}^{1/2}}. \tag{2.32}$$

### E. Einstein–Yang–Mills equations

A Yang–Mills field can also be considered as a source for the Ricci tensor (1.8). In fact, taking SU(2) as the gauge group, assuming that the gauge fields depend on $v$ only and the $A_3^i$ $(i = 1,2,3)$ are the only nonvanishing components of the potential [cf., Eq. (2.29)], we find that the solution to the Einstein–Yang–Mills equations is found by multiplying the respective expressions for $A_3$, $F_{12}$, and $F_{34}$ found in the Einstein–Maxwell case above by a constant element of the Lie algebra of the gauge group, while $h$ is of the form (2.15). This solution to the Yang–Mills equations is rather trivial since, in a sense, the nonabelian features of the gauge field have been lost.

## III. DISCUSSION

In all cases considered in Sec. II, after each matter field is suitably restricted so as to satisfy the gravitational field equations, the equations for the matter field can be integrated without knowing the explicit form of the function $h$. Thus, since the Ricci tensor depends linearly on $h$, by adding the $h$ corresponding to each source one can get solutions to the Einstein field equations coupled, e.g., to several neutrino fields of the form (2.6) and/or (2.11), to an electromagnetic field, linear or nonlinear, to a Yang–Mills field and to a cosmological constant. The form of the matter field is unaltered by this superposition.

Except for the neutrino field (2.11), the solutions to the matter field equations obtained above do not involve the function $h$; hence, setting $h = 0$, those expressions are solutions to the corresponding equations in flat space-time, where an interpretation of the solutions can be more easily given. For example, making the identification, $\zeta = (x + iy)/\sqrt{2}$, $u = (z + t)/\sqrt{2}$, $v = (z - t)/\sqrt{2}$, where $x, y, z, t$ are Minkowskian coordinates, one finds that the neutrino field (2.6) corresponds to a "wave packet," whose thickness depends on $a$, extending in the $x$ and $y$ directions and which travels with the speed of light in the $z$ direction. As in the case of a soliton, the shape of this wave does not change as it propagates. It may also be noticed that in the limit when $a$ tends to zero, the second term in (2.8) has a Dirac's $\delta$-behavior.[12]

In the auxiliary Minkowski metric, the Killing vector $K_4$ given in (1.11) corresponds to rotations around the $z$ axis. Hence, if $T_{12}$ and $T_{34}$ are the only nonvanishing independent components of the energy-momentum tensor, then, assuming $h = 0$, the density of the $z$ component of the angular momentum of the matter field is given by $\sqrt{2}a\zeta\bar{\zeta}\,(T_{34} - T_{12})/(v^2 + a^2)$, which vanishes when $a$ does. In the case of the electromagnetic field, linear or nonlinear, found in Sec. II, this density has the opposite sign to that of $a$. For the Weyl field [(2.6)] this density is always non-negative.

The result obtained in the case of the Weyl field, namely that the expression (2.6) represents a solution to the Weyl equation in the metric $g$ and in the auxiliary Minkowski metric, is part of a more general result. For a metric $g$ of the Kerr–Schild form, given by (1.1), a massless spinor field of arbitrary spin, $\Psi_{AB\ldots D}$, whose only nonvanishing component with respect to the tetrad (1.3)–(1.4) is $\Psi_{22\ldots2}$ (hence, the field has only one principal null direction, coincident with $\partial_4$) is a solution to the massless spinor field equations if it is a solution of these equations in the flat metric $\eta$ [obtained by making $h = 0$ in (1.1)–(1.4)], provided that $\partial_4$ is tangent to a geodesic null congruence. The validity of this proposition, and of the similar facts for other fields found in Sec. II, depends on the use of appropriately related bases for each metric, in which the solution has the same form, even though, due to the difference of the corresponding geometries, the field itself may be quite different. It may be noticed, for example, that the electromagnetic field of the Kerr–Newman solution (which is of the Kerr–Schild form) can be expressed in a form which is also a solution to Maxwell's equations in flat space-time [see, e.g., Ref. 1, (7.15)]. It remains as an open question, in the general case of the Kerr–Schild metrics, which conditions are necessary in order to integrate the matter field equations without specifying $h$ and under which conditions the energy-momentum tensor does not depend on $h$.

The metric (1.1)–(1.2) with $Y$ given by (1.5) and $h = h(v)$ can be written as

$$g = 2(v^2 + a^2)dY\,d\bar{Y} - \frac{(dv)^2}{2h} + 2h\,[d\tau + ia(Y\,d\bar{Y} - \bar{Y}\,dY)]^2, \tag{3.1}$$

where $\tau = -u - Y\bar{Y}v - \int dv/2h$. Considering now $Y$ and $\bar{Y}$ as two real (independent) coordinates and replacing $\tau$ by $i\sigma$, with $\sigma$ real, (3.1) represents another type $D$ metric whose principal null directions, being geodesic and shear-free, have vanishing complex expansion. Except for the neutrino fields (2.6) and (2.11), by performing this complex substitution in the expression found in Sec. II we obtain real solutions to the coupled field equations. [Some minor modifications are necessary, e.g., the sign of the last term in Eq. (2.15) must be changed. For a more detailed discussion on this point as well as more general solutions to the Einstein equations with nonlinear electromagnetic sources see Ref. 13.]

The complex substitution given above amounts, essentially, to interchange the spinorial indices $\dot{1}$ and $\dot{2}$, leaving invariant the undotted ones. In order to obtain another real

field, after executing this process, the dotted components must be the complex conjugates of the undotted ones. Since only one of the components of the neutrino fields (2.6) and (2.11) does not vanish, this process can not lead to another real solution.

## ACKNOWLEDGMENT

One of the authors (G.F.T.dC.) acknowledges partial support from the Subsecretaría de Educación Superior e Investigación Científica de la Secretaría de Educación Pública (México).

[1]G. C. Debney, R. P. Kerr, and A. Schild, J. Math. Phys. **10**, 1842 (1969).
[2]B. Kuchowicz, Gen. Relativ. Gravit. **5**, 201 (1974).
[3]S. Alarcón Gutiérrez, A. L. Dudley, and J. F. Plebański, J. Math. Phys. **22**, 2835 (1981).
[4]R. Penrose, J. Math. Phys. **8**, 345 (1967).
[5]M. Gürses and F. Gürsey, J. Math. Phys. **16**, 2385 (1975).
[6]A fairly complete discussion on Killing–Yano tensors can be found in W. Dietz and R. Rüdiger, Proc. R. Soc. London Ser. A **375**, 361 (1981); **381**, 315 (1982).
[7]E. Newman, L. Tamburino, and T. Unti, J. Math. Phys. **4**, 915 (1963).
[8]I. Robinson and A. Trautman, Proc. R. Soc. London Ser. A **265**, 463 (1962).
[9]J. Díaz Reyes, B. S. thesis (Universidad Autónoma de Puebla, 1983) (unpublished).
[10]C. A. Kolassis, J. Phys. A: Math. Gen. **16**, 749 (1983).
[11]J. K. Kowalczyński and J. F. Plebański, Acta Phys. Pol B **8**, 169 (1977).
[12]The authors are grateful to Dr. E. Cantoral for pointing out these facts to them.
[13]H. Salazar I., A. García D., and J. F. Plebański, "Duality rotations in Einstein theory and type *D* solutions," Ann. Phys. (N.Y.) (in press).

# Gauge theory in relativistic Hamiltonian classical mechanics: Tentative unification of electromagnetic and gravitational fields

Bruno Boisseau
*Département de Physique, Faculté des Sciences, Parc de Grandmont, 37200 Tours, France*

(Received 17 August 1984; accepted for publication 28 September 1984)

The infinitesimal canonical transformations are considered as gauge transformations. Using a method similar to that of Yang and Mills, one deduces, from a general formalism on phase space, the evolution equations of an infinite sequence of long-range fields sized by order of decreasing force. Special attention is given to the electromagnetic and gravitational fields which can be isolated by means of a restricted theory. In this last case, a comparison is done with the usual equations. An interpretation of the general formalism in terms of a vacuum is suggested in conclusion.

## I. INTRODUCTION

It is well known that gauge theories are extremely powerful tools in quantum field theory. The generality and elegance of the principles used lead to an attempt to expand the minimal coupling principle to other scopes, especially that of Hamiltonian mechanics. Sternberg,[1] generalizing a result of Souriau[2] concerning the electromagnetic field, has introduced in the symplectic form a connection associated to an arbitrary Lie group, thus defining the motion of a particle in the gauge field of an internal space.

This paper proceeds in an entirely different manner. Previous works[3,4] have shown the possibility of considering the infinitesimal canonical transformations as gauge transformations. Even so, these transformations do not form a Lie group; it is, however, possible to adapt in a natural way the Yang–Mills[5] initial manner to the Hamiltonian mechanics of a particle (in a relativistic form).

The formal invariance of the Hamiltonian, under any infinitesimal canonical transformation, is obtained by introducing "phase space gauge potentials" $K_\alpha(p,x)$, according to a process similar to the usual minimal coupling principle. The variance of these potentials, consequently determined, is similar to that of Yang and Mills.

Expanding $K_\alpha(p,x)$ in powers of $p$, we define a sequence of interaction potentials among which the first two are the potentials of the electromagnetic and gravitational fields.

It is possible to neglect the upper-order potentials, either by doing an additional hypothesis of approximation, or by using a theory restricted to the subgroup of infinitesimal canonical transformations associated to space-time changes. In this last case, the equations of motion of a particle are identical to the usual equations in the presence of electromagnetic and gravitational fields.

The results so presented are incomplete, from a gauge theory point of view, since there are no field equations. In the present paper, we propose, continuing the analogy with Yang–Mills theory, to define "phase space gauge fields" and to give their evolution equations. From these general equations we deduce the evolution equations of the electromagnetic and gravitational fields as well as those of the other hypothetical long-range fields introduced by the theory. We thus obtain an unification of long-range fields.

In order to make this paper self-consistent a summary of the results obtained in Refs. 3 and 4 is given in Sec. II. This will allow us, beyond the introduction of notation, to bring some precision and some small modifications. In Sec. III we introduce the "phase space gauge fields" and their evolution equations. We then indicate how to deduce from them the evolution equations of the long-range fields and give explicitly the electromagnetic and gravitational field equations in the limit of the theory restricted to the canonical transformations associated to the space-time change. Incidentally, applying the approximation hypothesis, we give, to the lowest order, the equations of the field coming immediately after the gravitational field. The two next sections are an application of the restricted theory. In Sec. IV it is deduced, from the electromagnetic field equations, that the light rays in the gravitational field are null geodesics. In Sec. V a static isotropic solution of the gravitational equations is obtained. Its ability to represent the solar system is examined by comparing it with the Schwarzschild solution. In Sec. VI, at last, we try to find an interpretation for the "phase space fields" and for their propagation, by transposing the notion of quantum vacuum to the Hamiltonian mechanics.

## II. EQUATIONS OF MOTION OF A PARTICLE

### A. Gauge transformation in relativistic Hamiltonian classical mechanics ($c = 1$)

Let us consider a Minkowski space $(x^\alpha),\eta_{\alpha\beta} = (1,-1,-1,-1)$, to which is associated an eight-dimensional phase space $(x^\alpha,p_\alpha)$. We can describe the motion of a free particle by means of the Hamiltonian[6] $H_0 = \sqrt{p_\alpha p_\beta \eta^{\alpha\beta}}$ and an evolution parameter $\tau$ independent of the phase space.

The Hamiltonian $H_0$ is a constant of the motion whose value $m$, determined by intitial conditions, is identified with the mass of the particle. So, each value of $H_0$ yields a particle of mass $m$ whose proper time $s$ is automatically equal to the evolution parameter. Indeed

$$\frac{dx^\alpha}{d\tau} = \frac{\partial H_0}{\partial p_\alpha} = \frac{p^\alpha}{m} \Rightarrow \frac{dx^\alpha\, dx^\beta\, \eta_{\alpha\beta}}{d\tau^2} = 1. \tag{2.1}$$

 0022-2488/85/030482-08$02.50 © 1985 American Institute of Physics 

The independence of $\tau$ in relation to the phase space enables us to consider the usual theory of canonical transformations. If we perform any infinitesimal canonical transformation (this transformation being independent of $\tau$), the Hamiltonian behaves like a scalar and is expressed in the new coordinate system as

$$H'_0 = \left[\left(p'_\alpha - \epsilon\frac{\partial G}{\partial x'^\alpha}(p',x')\right)\left(p'_\beta - \epsilon\frac{\partial G}{\partial x'^\beta}(p',x')\right)\eta^{\alpha\beta}\right]^{1/2}, \tag{2.2}$$

where $\epsilon$ and $G(p,x)$ are, respectively, the infinitesimal parameter and the generating function of the transformation. It is obvious that the Hamiltonian of a free particle is not form-invariant.

If we consider this transformation as a gauge transformation, we are going to replace in the Hamiltonian (2.2) the derivatives coming from the transformation by functions $K_\alpha(p',x')$ according to

$$\epsilon\frac{\partial G}{\partial x'^\alpha}(p',x') \to K_\alpha(p',x'). \tag{2.3}$$

This process is similar to the minimal coupling principle. Thus we obtain, after dropping the primes, the new Hamiltonian

$$H = [(p_\alpha - K_\alpha(p,x))(p_\beta - K_\beta(p,x))]^{1/2}. \tag{2.4}$$

It will be convenient to introduce the quantity

$$h_\alpha = p_\alpha - K_\alpha(p,x); \tag{2.5}$$

the Hamiltonian (2.4) is then written in the form

$$H = \sqrt{h_\alpha h_\beta \eta^{\alpha\beta}}. \tag{2.6}$$

Now if we perform any infinitesimal canonical transformations of the generating function $G(p,x)$, this new Hamiltonian is transformed like a scalar and, $(p',x')$ meaning the new coordinates, the quantities

$$h'_\alpha(p',x') = h_\alpha[p(p',x'),x(p',x')] \tag{2.7}$$

can be written in the form

$$h'_\alpha = p'_\alpha - K'_\alpha(p',x'), \tag{2.8}$$

provided that

$$K'_\alpha(p',x') = K_\alpha(p',x') + \epsilon\frac{\partial G}{\partial x'^\alpha}(p',x') + \epsilon[K_\alpha, G]_{x',p'}, \tag{2.9}$$

where

$$[K_\alpha, G]_{x',p'} = \frac{\partial K_\alpha}{\partial x'^\mu}\frac{\partial G}{\partial p'_\mu} - \frac{\partial K_\alpha}{\partial p'_\mu}\frac{\partial G}{\partial x'^\mu} \tag{2.10}$$

is a Poisson bracket. Thus the Hamiltonian (2.4) is form invariant on condition that the functions $K_\alpha$ should be transformed jointly according to (2.9).

Let us remark that the transformation (2.9) is exactly similar to those of gauge potentials in the usual Yang–Mills theory; the Lie bracket is simply replaced by the Poisson bracket.

Moreover, $K_\alpha$ and $h_\alpha$ can be considered as the covariant components of two vectors under constant Lorentz transformations; this is a consequence of the Lorentz invariance of the free Hamiltonian $H_0 = \sqrt{p_\alpha p^\alpha}$ with which we started. The quantities $K_\alpha$ will be named "phase space gauge potentials."

Let us suppose that we can expand $K_\alpha(x,p)$ and $G(x,p)$ in powers of $p$:

$$G = G^{(0)}(x) + G^{(1)\mu}(x)p_\mu + \cdots + \frac{1}{i!}G^{(i)\mu_1\cdots\mu_i}(x) \times p_{\mu_1}\cdots p_{\mu_i} + \cdots, \tag{2.11}$$

$$K_\alpha = K^{(0)}_\alpha(x) + K^{(1)\mu}_\alpha(x)p_\mu + \cdots + \frac{1}{i!}K^{(i)\mu_1\cdots\mu_i}_\alpha(x) \times p_{\mu_1}\cdots p_{\mu_i} + \cdots. \tag{2.12}$$

[$G^{(i)\mu_1\cdots\mu_i}$ and $K^{(i)\mu_1\cdots\mu_i}_\alpha$ are symmetrical.] The different quantities $K^{(i)}_\alpha(x)$ will be interpreted as potentials of the long-range fields, $K^{(0)}_\alpha$ for the electromagnetic field, $K^{(1)\mu}_\alpha$ for the gravitational field, etc.

The variance of these potentials is obtained by inserting expansions (2.11) and (2.12) in the formula of gauge transformation (2.9) and by identifying the same powers of $p'$ in each member:

$$\begin{aligned}
&K'^{(0)}_\alpha(x') = K^{(0)}_\alpha(x) + G^{(0)}_{,\alpha}(x) - \epsilon K^{(1)\mu}_\alpha(x)G^{(0)}_{,\mu}(x),\\
&K'^{(i)\mu_1\cdots\mu_i}_\alpha(x') = K^{(i)\mu_1\cdots\mu_i}_\alpha(x) + \epsilon G^{(i)\mu_1\cdots\mu_i}_{,\alpha}(x)\\
&\quad - \epsilon\left[K^{(1)\mu}_\alpha(x)G^{(i)\mu_1\cdots\mu_i}_{,\mu}(x) + C^1_i K^{(2)\mu_1\mu}_\alpha(x)G^{(i-1)\mu_2\cdots\mu_i}_{,\mu}(x)\right.\\
&\quad + C^2_i K^{(3)\mu_1\mu_2\mu}_\alpha(x)G^{(i-2)\mu_3\cdots\mu_i}_{,\mu}(x) + \cdots + C^{(i-1)}_i K^{(i)\mu_1\cdots\mu_{i-1}\mu}_\alpha(x)G^{(1)\mu_i}_{,\mu}(x)\\
&\quad \left. + K^{(i+1)\mu_1\cdots\mu_i\mu}_\alpha(x)G^{(0)}_{,\mu}(x)\right] \quad [,\mu = \partial_\mu].
\end{aligned} \tag{2.13}$$

### B. Canonical transformations associated with a change of space-time coordinates: Electromagnetic and gravitational interaction

We started this study by performing any infinitesimal canonical transformation on a coordinate system $(x^\alpha, p_\alpha)$ of the phase space in which $x^\alpha$ are space-time coordinates. The physical meaning of such a transformation is not obvious since it mixes the space-time coordinates and the momentum coordinates whose physical nature is different. This incites us, before carrying on the general case, to restrict this study to the subgroup of canonical transformations associated with a change of space-time coordinates, according to

$$x'^\mu = f^\mu(x^\alpha), \quad p'_\mu = g_\mu(x^\alpha, p_\alpha). \tag{2.14}$$

Generating functions of such transformations are in the form

$$G(p,x) = G^{(0)}(x) + G^{(1)\mu}(x)p_\mu. \tag{2.15}$$

The substitution (2.3) becomes in this case

$$\epsilon\frac{\partial G}{\partial x^\alpha} = \epsilon\frac{\partial G^{(0)}}{\partial x^\alpha}(x) + \epsilon\frac{\partial G^{(1)\mu}}{\partial x^\alpha}(x)p_\mu \rightarrow K_\alpha^{(0)}(x) + K_\alpha^{(1)\mu}(x)p_\mu. \tag{2.16}$$

Hence the "phase space gauge potential" is reduced to

$$K_\alpha = K_\alpha^{(0)}(x) + K_\alpha^{(1)\mu}(x)p_\mu. \tag{2.17}$$

Let us consider

$$V_\alpha{}^\mu = \delta_\alpha^\mu - K_\alpha^{(1)\mu}, \tag{2.18}$$

where $V^\alpha{}_\mu$ is the inverse matrix of $V_\alpha{}^\mu$,

$$eA_\mu = V^\alpha{}_\mu K_\alpha^{(0)} \quad (e = \text{const}). \tag{2.19}$$

The variance of these new quantities is deduced from Eqs. (2.13):

$$A'_\mu(x') = \frac{\partial x^\nu}{\partial x'^\mu}\left(A_\nu(x) + \left(\frac{\epsilon}{e}\right)\partial_\nu G^{(0)}(x)\right), \tag{2.20}$$

$$V'^\mu_\alpha(x') = \frac{\partial x'^\mu}{\partial x^\nu}V_\alpha{}^\nu(x). \tag{2.21}$$

Equation (2.20) points out that $A_\mu$ is transformed, up to a gradient, as a covariant vector under a general transformation of space-time, which agrees exactly with the properties of electromagnetic potentials. Equation (2.21) shows that $V_\alpha{}^\mu$ constitute four contravariant vector fields, and $V^\alpha{}_\mu$ four covariant vector fields.

We can construct the symmetrical tensor fields of order 2:

$$g^{\mu\nu} = V_\alpha{}^\mu V_\beta{}^\nu \eta^{\alpha\beta}, \tag{2.22}$$

$$g_{\mu\nu} = V^\alpha{}_\mu V^\beta{}_\nu \eta_{\alpha\beta}, \tag{2.23}$$

verifying

$$g^{\mu\rho}g_{\rho\nu} = \delta^\mu_\nu. \tag{2.24}$$

Then $g_{\mu\nu}$ can be considered as a pseudo-Riemannian metric tensor associated with tetrad fields $V_\alpha{}^\mu(x)$.

Let us determine now the motion of a particle. The quantity $h_\alpha$ is given by

$$h_\alpha = (p_\mu - eA_\mu)V_\alpha{}^\mu, \tag{2.25}$$

hence

$$H = \sqrt{(p_\mu - eA_\mu)(p_\nu - eA_\nu)g^{\mu\nu}}. \tag{2.26}$$

The Hamiltonian $H$ is a constant of the motion $m$ which is identified with the mass of the particle. We have

$$\frac{dx^\mu}{d\tau} = \frac{\partial H}{\partial p_\mu} = \frac{p^\mu - eA^\mu}{m} \Rightarrow \frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau^2}g_{\mu\nu} = 1.$$

The proper time identifies itself with the evolution parameter. By eliminating $p_\mu$ from Hamilton's equations, we obtain the usual equations of a charged particle in electromagnetic and gravitational fields:

$$m\left(\frac{du^\lambda}{d\tau} + \left\{\begin{matrix}\lambda\\ \mu\nu\end{matrix}\right\}u^\mu u^\nu\right) = g^{\lambda\rho}eF_{\rho\sigma}u^\sigma, \tag{2.27}$$

where $u^\lambda = dx^\lambda/d\tau$, $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, $\left\{\begin{matrix}\lambda\\ \mu\nu\end{matrix}\right\}$ = Christoffel's symbols, and $e$ is the electric charge.

### C. General infinitesimal canonical transformations and approximations

If we consider any infinitesimal canonical transformation, we must look at the infinite series of potentials $K_\alpha^{(0)}, K_\alpha^{(1)},\ldots, K_\alpha^{(i)},\ldots$. However, by restricting the study to the transformations associated with a change of space-time coordinates, we have been able to recognize in the first two, sized by order of decreasing force, the electromagnetic field and the gravitational field, the only two known long-range fields.

This suggests that the following fields, not being detected, must be much weaker. This hypothesis can be fulfilled, for example, by assuming

$$|K^{(i+1)}| \ll |K^{(i)}|, \tag{2.28}$$

$$|K^{(i)}|\cdot|K^{(j)}| \sim |K^{(i+j)}|. \tag{2.29}$$

We will say that we do an approximation of the $n$th order if we neglect the potentials $K^{(i)}$ as well as all equivalent products for $i > n$.

(1) Zeroth-order approximation. The Hamiltonian

$$H = \sqrt{(p_\alpha - eA_\alpha)(p_\beta - eA_\beta)\eta^{\alpha\beta}} \tag{2.30}$$

yields the usual equations of a charged particle in an electromagnetic field:

$$m\frac{du^\lambda}{d\tau} = \eta^{\lambda\rho}eF_{\rho\sigma}u^\sigma. \tag{2.31}$$

(2) First-order approximation. We find again the results of Sec. II B in the approximation of weak gravitational fields. Indeed $g^{\mu\nu}$ contains the products $K_\alpha^{(1)\mu}K_\beta^{(1)\nu}\eta^{\alpha\beta}$ which are neglected in this approximation.

(3) Second-order approximation. It is interesting to examine the approximation coming immediately after the known physics by introducing the potential $K_\alpha^{(2)\mu\nu}$.

In order to simplify computations, we exclude the electromagnetic fields (we suppose the particle is not charged). The Hamiltonian

$$H = (g^{\mu\nu}p_\mu p_\nu - \tfrac{1}{3}\epsilon^{\mu_1\mu_2\mu_3}p_{\mu_1}p_{\mu_2}p_{\mu_3})^{1/2}, \tag{2.32}$$

with

$$\epsilon^{\mu_1\mu_2\mu_3} = g^{\mu_1\nu}V^\alpha{}_\nu K_\alpha^{(2)\mu_2\mu_3} + \text{permut (1,2,3)}, \tag{2.33}$$

yields the equations

$$\frac{du^\mu}{d\tau} + \left\{\begin{matrix}\mu\\ \nu_1\nu_2\end{matrix}\right\}u^{\nu_1}u^{\nu_2} + m\left\{\begin{matrix}\mu\\ \nu_1\nu_2\nu_3\end{matrix}\right\}u^{\nu_1}u^{\nu_2}u^{\nu_3} = 0, \tag{2.34}$$

with

$$\left\{\begin{matrix}\nu\\ \nu_1\nu_2\nu_3\end{matrix}\right\} = \frac{1}{3!}g^{\nu\gamma}(\partial_{\nu_1}\epsilon_{\gamma\nu_2\nu_3} + \text{permut (1,2,3)} - \partial_\gamma\epsilon_{\nu_1\nu_2\nu_3}). \tag{2.35}$$

Let us observe that this second-order field has an influence proportional to the mass of the particle. This leads to a search for an eventual illustration of this field among the

celestial objects. In Ref. 4 we considered the Keplerian motion in the Newtonian approximation.

Finally, let us notice that the proper time cannot be identified exactly with the evolution parameter anymore:

$$\frac{dx^\mu\,dx^\nu}{d\tau^2}g_{\mu\nu} = 1 - \frac{2}{3}me^{\mu\nu\rho}u_\mu u_\nu u_\rho. \tag{2.36}$$

## III. FIELD EQUATIONS

Problems of interpretation are carried forward in Sec. VI.

### A. Phase space gauge field

First we must specify what we call covariance under infinitesimal canonical transformations considered as gauge transformations. We observed in Sec. II the perfect similitude betwen the variance of phase space gauge potentials given by

$$K'_\alpha(p',x') = K_\alpha(p',x') + \epsilon\left(\frac{\partial G}{\partial x'^\alpha}\right)(p',x') + \epsilon[K_\alpha,G]_{x',p'} \tag{3.1}$$

and that of the usual gauge potentials. However, in the Yang–Mills theory, the variance of gauge potentials differs from the covariance by the gradient term. [The gradient term is here $\epsilon(\partial G/\partial x'^\alpha)$.] This leads us to define the covariance of a quantity $A(p,x)$ by

$$A'(p',x') = A(p',x') + \epsilon[A,G]_{x',p'}. \tag{3.2}$$

The transformation is the infinitesimal expression of

$$A'(p',x') = A(p(p',x'), x(p',x')). \tag{3.3}$$

So, the quantity $A$ is a scalar.

Now we can introduce, following the analogy with Yang–Mills theory, the covariant derivative

$$D_\alpha A = \left(\frac{\partial A}{\partial x^\alpha}\right)(p,x) + [K_\alpha,A]_{x,p}, \tag{3.4}$$

and the gauge field

$$R_{\alpha\beta} = \left(\frac{\partial K_\beta}{\partial x^\alpha}\right)(p,x) - \left(\frac{\partial K_\alpha}{\partial x^\beta}\right)(p,x) + [K_\alpha,K_\beta]_{x,p}. \tag{3.5}$$

If we remember, according to Eq. (2.7), that $h_\alpha$ are scalar (covariant) quantities and that Poisson brackets are conserved by canonical transformations, it is easy to verify that $D_\alpha A$ and $R_{\alpha\beta}$ are covariant as is immediately proved by the following identities:

$$D_\alpha A = [A,h_\alpha]_{x,p} \tag{3.6}$$

and

$$R_{\alpha\beta} = [h_\alpha,h_\beta]_{x,p} \quad (= D_\beta h_\alpha). \tag{3.7}$$

The gauge field verifies the Bianchi identities,

$$D_\alpha R_{\beta\gamma} + D_\beta R_{\gamma\alpha} + D_\gamma R_{\alpha\beta} = 0, \tag{3.8}$$

which are simply the expression of the Jacobi identities

$$[[h_\beta,h_\gamma]_{x,p},h_\alpha]_{x,p} + \text{permut}\,(\alpha,\beta,\gamma) = 0.$$

All these analogies are reasons to write the equations of the free field as

$$D^\alpha R_{\alpha\beta} = 0 \quad (\eta^{\alpha\gamma}[R_{\alpha\beta},h_\gamma]_{x,p} = 0). \tag{3.9}$$

*Remark:* The quantities $R_{\alpha\beta}$, $D_\gamma R_{\alpha\beta}$ are the components of tensors under a constant Lorentz transformation.

By inserting the expansion (2.12) in the first member of Eq. (3.9) and by identifying the coefficients of each power of $p$ with zero (since the second member is null), we obtain an infinite sequence of second-order partial differential equations. These are the evolution equations of the long-range fields among which we find the equations of the electromagnetic and gravitational fields.

### B. Theory restricted to canonical transformations associated with a change of space-time coordinates

Since the beginning we have implicitly used the indices $\alpha,\beta,\gamma,\delta,\epsilon$ to mark the Lorentz variables and $\gamma,\mu,\nu,\rho,\sigma$ to mark any others. In order not to complicate some formulas, we have not made it an absolute rule. So, for example, Eq. (2.5) would be more correctly written $h_\alpha = \delta^\mu_\alpha p_\mu - K_\alpha(p,x)$. This convention is going to be particularly useful in the sequel of this paper. So, it will systematically be used except where otherwise stated.

In order to facilitate the subsequent exposition, let us next consider the tetrad fields

$$e_\alpha = V_\alpha{}^\mu\partial_\mu. \tag{3.10}$$

The Lie bracket of $e_\alpha$ and $e_\beta$ satisfies

$$[e_\alpha,e_\beta] = C_{\alpha\beta}^\gamma e_\gamma, \tag{3.11}$$

where

$$C_{\alpha\beta}^\gamma = (e_\alpha V_\beta{}^\nu - e_\beta V_\alpha{}^\nu)V^\gamma{}_\nu \tag{3.12}$$

($e_\alpha$ and $e_\beta$ stand for Pfaff derivatives). The electromagnetic potentials $A_\mu = V^\alpha{}_\mu K^{(0)}_\alpha/e$ can be expressed in the base $e_\alpha$:

$$A_\beta = V_\beta{}^\mu A_\mu = K^{(0)}_\beta/e. \tag{3.13}$$

In the same way the components of the electromagnetic field $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ become

$$F_{\alpha\beta} = V_\alpha{}^\mu V_\beta{}^\nu F_{\mu\nu} = e_\alpha A_\beta - e_\beta A_\alpha - A_\gamma C_{\alpha\beta}^\gamma. \tag{3.14}$$

With this preliminary out of the way, let us now turn to the fundamental equations obtained in the preceding section. This study is restricted to the case considered in Sec. II B, where only the electromagnetic and gravitational fields are occurring. Equations (2.17), (2.18), and (3.13) yield immediately

$$h_\alpha = -eA_\alpha + z_\alpha, \tag{3.15}$$

where

$$z_\alpha = V_\alpha{}^\mu p_\mu. \tag{3.16}$$

The Poisson bracket of $z_\alpha$ and $z_\beta$ yields

$$[z_\alpha,z_\beta]_{x,p} = T_{\alpha\beta}^\gamma z_\gamma, \tag{3.17}$$

where

$$T_{\alpha\beta}^\gamma = -C_{\alpha\beta}^\gamma \tag{3.18}$$

will be interpreted as the gravitational field.[7]

The equations (3.15)–(3.17) allow us to obtain the gauge field (3.7) in the form

$$R_{\alpha\beta} = e(e_\alpha A_\beta - e_\beta A_\alpha) + T_{\alpha\beta}^\gamma z_\gamma, \tag{3.19}$$

or, furthermore, by using Eqs. (3.14) and (3.18),

$$R_{\alpha\beta} = eF_{\alpha\beta} - eA_\gamma T_{\alpha\beta}^\gamma + T_{\alpha\beta}^\gamma z_\gamma, \tag{3.20}$$

setting off the electromagnetic field $F_{\alpha\beta}$, the gravitational field $T_{\alpha\beta}^\gamma$, and a coupling term $eA_\gamma T_{\alpha\beta}^\gamma$ between the two. The gauge field equations are obtained by inserting Eq. (3.20) into Eq. (3.9):

$$e(\eta^{\alpha\epsilon}e_\epsilon F_{\alpha\beta}+\eta^{\alpha\epsilon}T_{\alpha\beta}^{\ \gamma}F_{\gamma\epsilon})-eA_\delta(\eta^{\alpha\epsilon}e_\epsilon T_{\alpha\beta}^{\ \delta}+\eta^{\alpha\epsilon}T_{\alpha\beta}^{\ \gamma}T_{\gamma\epsilon}^{\ \delta})$$
$$+(\eta^{\alpha\epsilon}e_\epsilon T_{\alpha\beta}^{\ \delta}+\eta^{\alpha\epsilon}T_{\alpha\beta}^{\ \gamma}T_{\gamma\epsilon}^{\ \delta})z_\delta=0 \tag{3.21}$$

($e_\epsilon = V_\epsilon^{\ \mu}\partial_\mu$ is a Pfaff derivative). These equations must be verified for any $p$ or, which is equivalent, for any $z$. Since (3.21) is a binomial, we deduce two systems of differential equations by equating to zero, respectively, the coefficients of the zeroth and the first powers of $z$.

Observing in Eq. (3.21) that the coefficient of $z_\delta$ is the same as the factor of $-eA_\delta$, we obtain finally the evolution equations

$$\eta^{\alpha\epsilon}e_\epsilon F_{\alpha\beta}+\eta^{\alpha\epsilon}T_{\alpha\beta}^{\ \gamma}F_{\gamma\epsilon}=0. \tag{3.22}$$

For the electromagnetic field, and

$$\eta^{\alpha\epsilon}e_\epsilon T_{\alpha\beta}^{\ \delta}+\eta^{\alpha\epsilon}T_{\alpha\beta}^{\ \gamma}T_{\gamma\epsilon}^{\ \delta}=0. \tag{3.23}$$

for the gravitational field.

Let us examine quickly these equations by comparing them with the usual ones. The usual Maxwell's free equations in presence of gravitation (expressed in the basis $e_\alpha$ for comparison[8])

$$e_\epsilon F^{\epsilon\beta}-T_{\alpha\delta}^{\ \alpha}F^{\delta\beta}-\tfrac{1}{2}T_{\alpha\epsilon}^{\ \beta}F^{\alpha\epsilon}=0 \tag{3.24}$$

are different from the just proposed equations (3.22). But if the gravitational field $T_{\alpha\beta}^{\ \gamma}$ is null, Eqs. (3.22) and (3.24) become identical. Then, both of them are reduced to Maxwell's equations in Minkowski space,

$$\partial_\alpha F^{\alpha\beta}=0. \tag{3.25}$$

Indeed, in this case the Lie brackets $[e_\alpha,e_\beta]$ are null, hence the basis $\{e_\alpha\}$ is a coordinate basis[9]: $e_\alpha=\partial_\alpha$ and $V_\alpha^{\ \mu}$ are the components of a Cartesian frame $\{e_\alpha\}$ expressed in a curvilinear system $(x^\mu)$: $V_\alpha^{\ \mu}=\partial x^\mu/\partial x^\alpha$.

At last let us notice that the principle of equivalence is applied in the same way to Eq. (3.22) as to Eq. (3.24).

Equations (3.23) are different from Einstein's equations. However we can point out that they are nonlinear. We will solve them in the case of spherical symmetry. We shall discuss at that time their ability to represent the gravitation.

### C. General case

As we pointed out at the end of Sec. III A, it is possible, by inserting the expansions (2.12) in the equations (3.9), to obtain an infinite sequence of second-order partial differential equations that we can limit by using the postulates of approximation (2.28) and (2.29). We will not write explicitly the generic equation, having no use of this in this paper. We give only the following results.

To the zeroth-order approximation we find only Maxwell's equations in Minkowski space.

To the first-order approximation we find Eq. (3.22) for the electromagnetic field and the linearized equation (3.23)

$$\eta^{\alpha\epsilon}e_\epsilon T_{\alpha\beta}^{\ \delta}=0 \tag{3.26}$$

for the gravitational field. (In this approximation the gravitational field is weak: we neglect in $T$ the products $K^{(1)}.K^{(1)}$.)

To the approximation of second order, we find more complicated equations than (3.22) and (3.23) for the electromagnetic and gravitational fields (since these can contain, moreover, the second-order potentials $K_\alpha^{(2)\mu\nu}$). Moreover the potential $K_\alpha^{(2)\mu\nu}$ satisfies the linear equations

$$\partial^\alpha(\partial_\alpha K_\beta^{(2)\mu_1\mu_2}-\partial_\beta K_\alpha^{(2)\mu_1\mu_2})=0, \tag{3.27}$$

which can be also written

$$\Box K_\beta^{(2)\mu_1\mu_2}-\partial_\beta\partial^\alpha K_\alpha^{(2)\mu_1\mu_2}=0. \tag{3.28}$$

[In Eqs. (3.27) and (3.28) we do not apply the convention of indices.] These last equations justify *a posteriori* the use of the Newtonian potential that we introduced,[10] by heuristic arguments, for the study of the Keplerian motion in the Newtonian approximation.

We stop here the evocation of the general case and will come back to it later on.

## IV. GEOMETRICAL OPTICS

In the Einstein theory, light rays are null geodesics. These are naturally the trajectories of photons (massless particles). But this property is also a consequence of Maxwell's equation. Let us summarize the principles of such a deduction.[11] In the geometrical optics limit, the wavelength $\lambda$ is small compared both to the scale $L_1$ over which the amplitude of wave changes and to $L_2$, the scale over which the gravitation changes, i.e.,

$$\epsilon=\lambda/L\ll 1,\quad L=\min(L_1,L_2). \tag{4.1}$$

The electromagnetic vector potential can be written in terms of rapidly varying phase and slowly varying amplitude in the form of the expansion

$$A_\mu=(a_\mu+\epsilon b_\mu+\ldots)e^{i\theta/\epsilon}, \tag{4.2}$$

where $\theta$ is a real phase, $a_\mu$ and $b_\mu$ are complex, and $\epsilon$, given by (4.1), is the expansion parameter. In fact the electromagnetic vector potential is the real part of (4.2). We define the wave vector

$$k_\mu=\partial_\mu\theta. \tag{4.3}$$

By inserting the potential vector (4.2) into the Lorentz gauge condition and keeping only the leading term ($1/\epsilon$ order), we obtain

$$k^\mu a_\mu=0. \tag{4.4}$$

Amplitude is perpendicular to wave vector. Next, inserting the potential vector (4.2) into the Maxwell's equations in the Lorentz gauge, retaining only the leading term ($1/\epsilon^2$ order), we obtain

$$k^\mu k_\mu=0. \tag{4.5}$$

The wave vector is null. Taking the gradient of Eq. (4.5) we get the geodesic equation

$$k^\mu\nabla_\mu k^\nu=0. \tag{4.6}$$

Equations (4.5) and (4.6) are the statement that light rays are null geodesics.

Let us turn now to the electromagnetic equations obtained in the last section:

$$\eta^{\alpha\epsilon}e_\epsilon F_{\alpha\beta}+\eta^{\alpha\epsilon}T_{\alpha\beta}^{\ \gamma}F_{\gamma\epsilon}=0. \tag{4.7}$$

We are going to show that, in this case also, light rays are null geodesics in the Riemannian space of metric

$$g_{\mu\nu}=V^{\alpha\mu}V^\beta_{\ \nu}\eta_{\alpha\beta}. \tag{4.8}$$

In the basis $\{e_\alpha\}$ we have the identity

$$\nabla_\epsilon F^{\epsilon\beta}=e_\epsilon F^{\epsilon\beta}-T_{\alpha\delta}^{\ \alpha}F^{\delta\beta}-\tfrac{1}{2}T_{\alpha\epsilon}^{\ \beta}F^{\alpha\epsilon}, \tag{4.9}$$

where $\nabla$ is the Riemanian connection associated with the metric (4.8). [The expression (4.9) was already mentioned in Eq. (3.24).]

By using the identity (4.9), Eq. (4.7) is now written

$$\nabla_\epsilon F^{\epsilon\beta} + T_{\alpha\delta}^{\ \ \alpha} F^{\delta\beta} + \tfrac{1}{2} T_{\alpha\epsilon}^{\ \ \beta} F^{\alpha\epsilon} + \eta^{\beta\delta}\eta^{\alpha\epsilon} T_{\alpha\delta}^{\ \ \gamma} F_{\gamma\epsilon} = 0, \tag{4.10}$$

where the electromagnetic field can be expressed in the form

$$\nabla_\alpha A_\beta - \nabla_\beta A_\alpha. \tag{4.11}$$

Because of the gauge transformation (2.20), it is always possible to choose the Lorentz gauge

$$\nabla_\alpha A^\alpha = 0. \tag{4.12}$$

Therefore, the first term in the left member of Eq. (4.10) becomes

$$\nabla_\epsilon F^{\epsilon\beta} = \nabla_\epsilon \nabla^\epsilon A^\beta + R_\alpha^{\ \beta} A^\alpha, \tag{4.13}$$

where $R_{\alpha\beta}$ is the Ricci tensor. Finally, the equation of an electromagnetic field in the Lorentz gauge can be written

$$\nabla_\epsilon \nabla^\epsilon A^\beta + R_\alpha^{\ \beta} A^\alpha + T_{\alpha\delta}^{\ \ \alpha} F^{\delta\beta} + \tfrac{1}{2} T_{\alpha\epsilon}^{\ \ \beta} F^{\alpha\epsilon} + \eta^{\beta\delta}\eta^{\alpha\epsilon} T_{\alpha\delta}^{\ \ \gamma} F_{\gamma\epsilon} = 0. \tag{4.14}$$

In the conditions where geometric optics is valid we can use the potential vector (4.2) expressed in the basis $\{e_\alpha\}$,

$$A_\alpha = (a_\alpha + \epsilon b_\alpha + \ldots) e^{i\theta/\epsilon}, \tag{4.15}$$

where

$$a_\alpha = V_\alpha^{\ \mu} a_\mu, \quad b_\alpha = V_\alpha^{\ \mu} b_\mu. \tag{4.16}$$

A covariant derivative of the potential vector (4.15) gives

$$\nabla_\beta A_\alpha = [\nabla_\beta a_\alpha + \epsilon \nabla_\beta b_\alpha + (i/\epsilon)(a_\alpha + \epsilon b_\alpha) e_\beta \theta\,] e^{i\theta/\epsilon}. \tag{4.17}$$

Equations (4.17) yield, writing only leading terms,

$$\nabla_\alpha A^\alpha = [(i/\epsilon) a^\alpha e_\alpha \theta + \ldots] e^{i\theta/\epsilon}, \tag{4.18}$$

$$F_{\alpha\beta} = [(i/\epsilon)(a_\beta e_\alpha \theta - a_\alpha e_\beta \theta) + \ldots] e^{i\theta/\epsilon}, \tag{4.19}$$

and, performing another derivative,

$$\nabla_\epsilon \nabla^\epsilon A^\alpha = [(i/\epsilon)^2 a^\alpha \eta^{\beta\gamma} e_\beta \theta e_\gamma \theta + \ldots] e^{i\theta/\epsilon}. \tag{4.20}$$

We can now obtain the equations of geometrical optics. From the Lorentz condition (4.12) and Eq. (4.18), setting the coefficient of each power of $\epsilon$ equal to zero, we obtain, for the leading term ($1/\epsilon$ order),

$$a^\alpha e_\alpha \theta = 0 \tag{4.21}$$

or

$$a^\alpha V_\alpha^{\ \mu} \partial_\mu \theta = a^\mu \partial_\mu \theta = a^\mu k_\mu = 0. \tag{4.22}$$

Amplitude is perpendicular to wave vector. In the same way, form the field equations (4.14) combined with Eqs. (4.15), (4.19), and (4.20), for the leading term ($1/\epsilon^2$ order), we obtain

$$a^\alpha \eta^{\beta\gamma} e_\beta \theta e_\gamma \theta = 0 \tag{4.23}$$

and consequently

$$g^{\mu\nu} k_\mu k_\nu = 0. \tag{4.24}$$

The wave vector is null.

Taking the gradient of (4.24) and noticing that $\nabla_\mu k_\nu = \nabla_\nu k_\mu$ (since $k_\mu = \partial_\mu \theta$), we get the geodesic equation

$$k^\mu \nabla_\mu k^\nu = 0. \tag{4.25}$$

Thus let us emphasize that, for the electromagnetic equation (4.7) in the geometrical optics limit, light rays are null geodesics as usual.

## V. "STATIC ISOTROPIC" SOLUTION OF GRAVITATIONAL FIELD EQUATIONS

A static isotropic metric can be reduced to the following form:

$$ds^2 = B^2(r) dt - A^{-2}(r)\, dr^2 - r^2\, d\theta^2 - r^2 \times \sin^2\theta\, d\varphi^2. \tag{5.1}$$

We propose to determine a static solution of the gravitational field equations obtained in Sec. III,

$$S_\beta^{\ \delta} \equiv \eta^{\alpha\epsilon} e_\epsilon T_{\alpha\beta}^{\ \ \delta} + \eta^{\alpha\epsilon} T_{\alpha\beta}^{\ \ \gamma} T_{\gamma\epsilon}^{\ \ \delta} = 0, \tag{5.2}$$

such that the associated metric

$$g_{\mu\nu} = V^\alpha_{\ \mu} V^\beta_{\ \nu} \eta_{\alpha\beta} \tag{5.3}$$

is the static isotropic metric

$$g_{\mu\nu} = (B^2(r),\ -A^{-2}(r),\ -r^2,\ -r^2 \sin^2\theta). \tag{5.4}$$

That is what we call briefly a static isotropic solution.

The problem is to find the two unknown functions $A(r)$ and $B(r)$. At first we must determine, by means of the symmetry conditions (5.3) and (5.4), the expression of $V_\alpha^{\ \mu}$ and $V^\alpha_{\ \mu}$ in terms of $A$ and $B$. We are going to use, as an intermediate, the diagonal tetrad $k_\alpha^{\ \mu}$, $k^{\alpha'}_{\ \ \mu}$ such that

$$g_{\mu\nu} = k^{\alpha'}_{\ \ \mu} k^{\beta'}_{\ \ \nu} \tilde{\eta}_{\alpha'\beta'}(r,\theta,\varphi), \tag{5.5}$$

where

$$\tilde{\eta}_{\alpha'\beta'} = (1, -1, -r^2, -r^2 \sin^2\theta) \tag{5.6}$$

is the Lorentz metric in spherical coordinates. So, the diagonal tetrad is

$$\{k^{\alpha'}_{\ \ \mu}\} = \{k^0_{\ 0} = B(r),\ k^1_{\ 1} = A^{-1}(r),\ k^2_{\ 2} = 1,\ k^3_{\ 3} = 1\}. \tag{5.7}$$

By inserting

$$\tilde{\eta}_{\alpha'\beta'} = \frac{\partial x^\alpha}{\partial x^{\alpha'}} \frac{\partial x^\beta}{\partial x^{\beta'}} \eta_{\alpha\beta} \tag{5.8}$$

in Eq. (5.5) and by identifying with Eq. (5.3), we obtain immediately

$$V^\alpha_{\ \mu} = \frac{\partial x^\alpha}{\partial x^{\alpha'}} k^{\alpha'}_{\ \ \mu}, \tag{5.9}$$

i.e., explicitly,

$$\begin{aligned}
&V^0_{\ 0} = B, \quad V^0_{\ 1} = 0, \quad V^0_{\ 2} = 0, \quad V^0_{\ 3} = 0,\\
&V^1_{\ 0} = 0, \quad V^1_{\ 1} = \sin\theta\cos\varphi A^{-1}, \quad V^1_{\ 2} = r\cos\theta\cos\varphi,\\
&V^1_{\ 3} = -r\sin\theta\sin\varphi,\\
&V^2_{\ 0} = 0, \quad V^2_{\ 1} = \sin\theta\sin\varphi A^{-1}, \quad V^2_{\ 2} = r\cos\theta\sin\varphi,\\
&V^2_{\ 3} = r\sin\theta\cos\varphi,\\
&V^3_{\ 0} = 0, \quad V^3_{\ 1} = \cos\theta A^{-1}, \quad V^3_{\ 2} = r\sin\theta, \quad V^3_{\ 3} = 0.
\end{aligned} \tag{5.10}$$

We deduce the inverse matrix $V_\alpha^{\ \mu}$:

$$\begin{aligned}
&V_0^{\ 0} = B^{-1}, \quad V_1^{\ 0} = 0, \quad V_2^{\ 0} = 0, \quad V_3^{\ 0} = 0,\\
&V_0^{\ 1} = 0, \quad V_1^{\ 1} = \sin\theta\cos\varphi A,
\end{aligned}$$

$$V_2^1 = \sin\theta \sin\varphi A, \quad V_3^1 = \cos\theta A,$$
$$V_0^2 = 0, \quad V_1^2 = \cos\theta\cos\varphi/r,$$
$$V_2^2 = cos\,\theta \sin\varphi/r,$$
$$V_3^2 = -\sin\theta/r,$$
$$V_0^3 = 0, \quad V_1^3 = -\sin\varphi/(r\sin\theta),$$
$$V_2^3 = \cos\varphi/(r\sin\theta), \quad V_3^3 = 0. \tag{5.11}$$

We can now proceed to the resolution of the differential equations (5.2). For that we have to insert the expressions (5.10) and (5.11) of $V^\alpha{}_\mu$ and $V_\alpha{}^\mu$ in Eq. (5.2). The computation of

$$T_{\alpha\beta}^{\gamma} = (V_\beta{}^\mu \partial_\mu V_\alpha{}^\nu - V_\alpha{}^\mu \partial_\mu V_\beta{}^\nu) V^\gamma{}_\nu \tag{5.12}$$

gives explicitly

$$T_{\alpha\alpha}^{\gamma} = 0, \quad T_{\alpha\beta}^{\gamma} = -T_{\beta\alpha}^{\gamma},$$
$$T_{01}^{0} = \sin\theta\cos\varphi AB\,\partial_r(B^{-1}),$$
$$T_{02}^{0} = \sin\theta\sin\varphi AB\,\partial_r(B^{-1}),$$
$$T_{03}^{0} = \cos\theta AB\,\partial_r(B^{-1}),$$
$$T_{ij}^{0} = T_{0j}^{i} = 0 \quad (i,j \neq 0),$$
$$T_{12}^{1} = \sin\theta\sin\varphi(1-A)/r, \quad T_{12}^{2} = \sin\theta\cos\varphi(1-A)/r,$$
$$T_{12}^{3} = 0, \; T_{13}^{1} = \cos\theta(1-A)/r, \quad T_{13}^{2} = 0,$$
$$T_{13}^{3} = -\sin\theta\cos\varphi(1-A)/r,$$
$$T_{23}^{1} = 0, \quad T_{23}^{2} = \cos\theta(1-A)/r,$$
$$T_{23}^{3} = -\sin\theta\sin\varphi(1-A)/r. \tag{5.13}$$

Let us write

$$K = AB\,\partial_r(B^{-1}), \tag{5.14}$$
$$H = (1-A)/r, \tag{5.15}$$

and

$$N = A\,\partial_r H - H/R + H^2.$$

The equations of motion (5.2) are then written

$$S_0^0 \equiv +A\,\partial_r K + 2K/r + K^2 = 0, \tag{5.16}$$
$$S_0^n \equiv S_n^0 \equiv 0 \quad (n \neq 0);$$
$$S_1^1 \equiv -(\sin^2\theta\cos^2\varphi - 1)N + 2H/r = 0,$$
$$S_2^2 \equiv -(\sin^2\theta\sin^2\varphi - 1)N + 2H/r = 0, \tag{5.17}$$
$$S_3^3 \equiv +\sin^2\theta N + 2H/r = 0;$$
$$S_1^2 \equiv S_2^1 \equiv -\sin^2\theta\sin\varphi\cos\varphi N = 0,$$
$$S_1^3 \equiv S_3^1 \equiv -\cos\theta\sin\theta\cos\varphi N = 0, \tag{5.18}$$
$$S_2^3 \equiv S_3^2 \equiv -\cos\theta\sin\theta\sin\varphi N = 0.$$

Equations (5.17) and (5.18) are verified only for $H = 0$, hence Eq. (5.15) yields

$$A = 1. \tag{5.19}$$

Then Eq. (5.16) can be written

$$\frac{dK}{dr} + 2\frac{K}{r} + K^2 = 0, \tag{5.20}$$

where

$$K = B\frac{d(B^{-1})}{dr}. \tag{5.21}$$

Noticing that

$$\frac{dK}{dr} + 2\frac{K}{r} = \left(\frac{1}{r^2}\right)\frac{d(r^2K)}{dr},$$

We obtain immediately the solution of Eq. (5.20):

$$K = 1/r(ar - 1) \tag{5.22}$$

($a$ is the constant of integration). Then, the solution of (5.21) gives

$$B = br/|ar - 1| \tag{5.23}$$

($b$ is the constant of integration). So, the problem is solved and the static isotropic metric associated with the solution is

$$ds^2 = (b^2/a^2)(1 - 1/ar)^{-2} - dr^2 - r^2\,d\theta^2 - r^2 \times \sin^2\theta\,d\varphi^2. \tag{5.24}$$

Let us suppose that this gravitational field represents the solar system. We are proceeding as in general relativity. At a great distance $r$ from the sun, we must find again, in a first approximation, the Newtonian theory. So that

$$g_{00} \sim 1 + 2\Phi, \tag{5.25}$$

where

$$\Phi = -GM/r. \tag{5.26}$$

Hence, the constants of integration verify

$$b^2/a^2 = 1, \quad a = -1/GM. \tag{5.27}$$

Then the metric (5.24), expanded as a power series in the small parameter $MG/r$, is written

$$ds^2 = [1 - 2GM/r + 3(GM/r)^2 + \cdots]dt^2 - dr^2 - r^2\,d\theta^2 - r^2\sin^2\theta\,d\varphi^2. \tag{5.28}$$

In order to test such a metric and to compare the results with those of the general relativity, we are going to use the expansion of Eddington and Robertson (see Ref. 12)

$$ds^2 = [1 - 2\alpha MG/r + 2(\beta - \alpha\gamma)(MG/r)^2 + \cdots]dt^2 - [1 + 2\gamma MG/r + \cdots]dr^2 - r^2\,d\theta^2 - r^2\sin^2\theta\,d\varphi^2. \tag{5.29}$$

This one is identical (at this approximation) with (5.28) for

$$\alpha = 1, \quad \beta = \tfrac{1}{2}, \quad \gamma = 0, \tag{5.30}$$

whereas the Schwarzschild solution of Einstein equations yields

$$\alpha = \beta = \gamma = 1. \tag{5.31}$$

Let us examine the three classical tests. The gravitational red shift experiment which verifies the principle of equivalence gives $\alpha = 1$, the deflection of light by sun tests $\gamma \simeq 1$, whereas the precession of perihelia verifies that $2\gamma - \beta \simeq 1$. So the test on the red shift is verified but neither the deflection of light

$$\theta = 1.75''(1 + \gamma)/2$$

nor the precession of perihelia

$$\Delta\varphi = (6\pi MG/L)(2 - \beta + 2\gamma)/3 \text{ rad/revolution}$$

($L$ is the semimatus rectum of the ellipse) are correctly obtained.

The test of the principle of equivalence is certainly the most important and its verification gives a certain confidence in the equations of motion. Nevertheless, it is necessary to find an explication at the deficiency of the two other tests.

It is evidently possible that the initial equations (3.9) are not absolutely correct, this deficiency can also proceed from the omission of the upper-order long-range fields. But, more simply, we can also suppose that the spherical symmetry does not represent perfectly the solar system. Indeed we know that the sidereal period of solar rotation is about twenty-five days. This causes us to search for an axially symmetric solution that represents the field of a rotating spherical body. We will reserve for later such a study.

## VI. DISCUSSION

This work constitutes an attempt at unification of the long-range fields and, more particularly, of the gravitational and electromagnetic fields. By transposing the Yang–Mills method in Hamiltonian mechanics for the group of infinitesimal canonical transformations we have been able to deduce these fields from the single quantities $K_\alpha(p,q)$.

Behind this formalism is nevertheless set an interpretation problem that we are going to examine, and to which we will attempt to give some elements of answer. These will have to be considered only as suggestions.

The geometrical frame of the theory is the cotangent bundle $T^*(V_4)$ associated to the space-time manifold $V_4$. Let us consider a system of adapted coordinates, where $x^\mu$ represents local coordinates of $V_4$ and $p_\mu$, the components of a covector in the associated natural frame. The motion of a particle is described by a trajectory of the Hamiltonian flow generated by $H=\sqrt{h_\alpha h^\alpha}$. In this case $x^\mu$ is the space-time position of the particle and $p_\mu$, the conjugated momentum, is related to the motion and to the energy of the particle.

If we now consider the motion of the free fields, ruled by the equations $D_\alpha R^{\alpha\beta}(p,x)=0$, the physical meaning is not so evident. The coordinates $x^\mu$ do represent the space-time position where the field is considered, but we are not able to give any meaning for $p_\mu$. If it is still related to motion and to energy, which motion and which energy is it?

We are going to attempt to answer this question, inspiring ourselves from the quantum theory in which the notions of vacuum and vacuum fluctuations have great importance.

We can imagine, in the classical physics frame, that the vacuum is subject to random-looking fluctuations, whose motion and energy at a given point $x^\mu$ of the space-time are associated with a covector $p_\mu$ called vacuum momentum at this point.

In this way, the notion of a vacuum can enter in the same geometrical frame $T^*(V_4)$ as the notion of motion of a particle. The fiber above $x^\mu$ represents the different possible momentum of the vacuum at $x^\mu$. Thus in a part $D$ of the space-time, a state of vacuum will be a section of the bundle. But a very "chaotic" section must certainly be imagined at the microscopic level, so that if a sufficiently large part is considered, the average $\langle p_\mu \rangle$ is null. Indeed, an energy accumulation cannot be considered on a macroscopic part because vacuum has no observable effects at the classical level.

But the office of the vacuum momentum is essential here because it allows us to consider the gauge field $R_{\alpha\beta}(p,x)$ and its propagation in the vacuum. This momentum is acting as a catalyzer, it is it which allows for example the coupling of the gravitational field with itself.

The presence of a particle in the part $D$ is expressed by a trajectory on the section of the fiber, along which the $p_\mu$ are not "chaotic" anymore but correlated by the Hamiltonian flow.

The physics world described here is that of the ponctual particles flowing in free fields $R_{\alpha\beta}(p,x)$, whose existence and propagation is realized by means of the vacuum previously described. Of course, when in $x$ there is a particle of momentum $p$(part.), the field takes the value $R_{\alpha\beta}(p(\text{part.}),x)$. The particle is considered as a test particle.

But particles are also sources of fields. The source occurs then as a limit condition. Thus, in Sec. V the spherical symmetry is assigned in order to represent the sun, which is considered as the source of the gravitational field.

There are no equations in the matter since it is considered as a discrete set of points. The field is indeed free almost everywhere. But from a practical point of view, it is certainly necessary to replace a discrete set by a continuum and, in this case, the equations must have sources. We can propose

$$D_\alpha R^{\alpha\beta}(p,x)=J^\beta(p,x), \tag{6.1}$$

with

$$D_\beta J^\beta(p,x)=0. \tag{6.2}$$

(In the discrete model, the Dirac distributions occur, the field is then free almost everywhere except on the sources.)

At last, let us mention another problem that the reader has certainly perceived. The electromagnetic field is introduced in this theory by the term $K_\alpha^{(0)}(x)$ of the expansion of $K_\alpha(p,x)$. This term is in fact a potential energy rather than a potential. The charge $e$ of the particle and the potential are then defined subsequently by $K_\alpha^{(0)}(x)=eA_\alpha(x)$. It is due to this construction that $e$ cannot be eliminated from the equations of the fields (3.21). But what does $e$ represent in vacuum? We can try to outline an answer similar to the former one and imagine that $e$ is related to the vacuum polarization. It would be the vacuum polarization, null only on average, which would allow $K_\alpha^{(0)}(x)$ to exist.


[1] S. Sternberg, Proc. Natl. Acad. Sci. USA **74**, 5253 (1977).
[2] J. M. Souriau, *Structure des Systgemes Dynamiques* (Dunod, Paris, 1970).
[3] B. Boisseau and C. Barrabès, J. Math. Phys. **20**, 2058 (1979).
[4] B. Boisseau, J. Math. Phys. **21**, 1372 (1980).
[5] C. N. Yang and R. L. Mills, Phys. Rev. **96**, 191 (1954).
[6] In Refs. 3 and 4 we have used the Hamiltonian $(1/2m)p_\alpha p^\alpha$. The calculus is almost not changed by the use of the Hamiltonian $\sqrt{p_\alpha p^\alpha}$. It is nevertheless more adapted to this study, because it, *a priori*, does not introduce the mass of the particle. For a study of relativistic Hamiltonians see R. Giachetti and E. Sorace, Nuovo Cimento A **43**, 281 (1978); Nuovo Cimento B **56**, 263 (1980).
[7] Such an expression has already been used in another context to represent the gravitational field, see Y. M. Cho, Phys. Rev. D **14**, 2521 (1976); K. Hayashi and T. Shirafuji, Phys. Rev. D **19**, 3524 (1979).
[8] For the use of a noncoordinate basis in a Riemannian manifold see, for example, Y. Choquet-Bruhat, C. De Witt-Morette, and M. Dillard-Bleick, *Analysis, Manifolds and Physics* (North-Holland, Amsterdam, 1977), Chap. 5.
[9] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), p. 239.
[10] See Ref. 4.
[11] C. M. Will, *Theory and Experiment in Gravitational Physics* (Cambridge U. P., Cambridge, 1981), Chap. 3, pp. 74 and 75.
[12] S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972), Chap. 8.

# Fermi states of Bose systems in three space dimensions

Piotr Garbaczewski[a)]
*Institute de Physique Theorique, Universite Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium*

(Received 9 January 1984; accepted for publication 27 July 1984)

Recently an exact spectral solution was constructed by Sudarshan and Tata for the $(N\Theta)$ Fermi version of the Lee model. We demonstrate that it provides a partial solution for the related pure Bose spectral problems. Moreover, the $(N\Theta)$ Bose (Bolsterli–Nelson) version of the Lee model is shown to possess Fermi partners, both exhibiting the partial solubility interplay: finding solutions in the Fermi case would presumably be easier than in the original Bose model. Fermi states of the underlying Bose systems in three space dimensions are explicitly identified.

Let us study a specialized version of the Lee model[1–4] as considered by Sudarshan and Tata.[5] The model consists of two fermions $N$, $\Theta$ interacting with a boson $V$. If compared with the original version of the Lie field theory model, the momentum dependence of $V$ and $N$ is lost due to their (assumed) infinite mass, then $V$ and $N$ play the role of sources, while $\Theta$ is supposed to be massless. We have

$$H = m_0 V^* V + \int d^2k \cdot k \cdot a^*(k) a(k) + \int d^3k f(k)[V^* N a(k) + a^*(k) N^* V] \tag{1}$$

with the commutation rules of Ref. 5

$$\begin{aligned} &[N,N^*]_+ = [V,V^*]_- = 1\,, \\ &[a(k),a^*(p)]_+ = \delta^3(\mathbf{k}-\mathbf{l}), \\ &[N,N]_+ = [V,V]_- = [a(k),a(p)]_+ = 0\,, \\ &[N,a(k)]_+ = [N,a^*(k)]_+ = [N,V]_- = [N,V^*]_- \\ &\qquad = [a(k),V]_- = [a(k),V^*]_- = 0\,. \end{aligned} \tag{2}$$

Let us observe that irrespective of whether quantum objects $V$, $N$, $\Theta$ represent bosons or fermions, and irrespective of whether they mutually commute or anticommute, the following two operators are the constants of motion:

$$\begin{aligned} N_1 &= N_V + N_N\,, \\ N_2 &= N_\Theta - N_N\,, \end{aligned} \tag{3}$$

where

$$N_V = V^* V, \quad N_N = N^* N, \quad N_\Theta = \int d^3k\, a^*(k) a(k)\,, \tag{4}$$

and upon assuming that $N_V, N_N, N_\Theta$ commute with any function of operators belonging to pairs of species $(N,\Theta),(V,\Theta),(V,N)$, respectively, we arrive at

$$\begin{aligned} [N_1,H]_- = \int d^3k f(k)\{ &V^*[V,V^*]_- Na(k) \\ &+ a^*(k) N^*[V^*,V]_- V \\ &+ V^*[N^*,N]_- Na(k) \\ &+ a^*(k) N^*[N,N^*]_- V\}\,, \end{aligned} \tag{5a}$$

$$\begin{aligned} [N_2,H]_- = \int d^3k \int d^3p f(p)\{ &V^* N[a^*(k),a(p)]_- a(k) \\ &+ a^*(k)[a(k),a^*(p)]_- N^* V\} \\ &- \int d^3k f(k)\{V^*[N^*,N]_- Na(k) \\ &+ a^*(k) N^*[N,N^*]_- V\}\,. \end{aligned} \tag{5b}$$

If now to admit that each of the species obeys some canonical (commutation or anticommutation) rules, then the conservation laws

$$[N_1,H]_- = 0 = [N_2,H]_- \tag{6}$$

immediately follow.

The standard ansatz about the form of eigenfunctions for $H$ is [1,2] that they should be superpositions of the bare states, i.e., eigenstates of

$$H_0 = m_0 V^* V + \int d^3k \cdot k \cdot a^*(k) a(k) \Rightarrow H = H_0 + H_{\text{int}}\,.$$

Since we wish to solve a common $(N_1,N_2,H)$ eigenvalue problem, it is natural to look for states $|a,b\rangle$ obeying

$$\begin{aligned} &N_1|a,b\rangle = a|a,b\rangle, \quad N_2|a,b\rangle = b|a,b\rangle\,, \\ &|a,b\rangle = \sum_{\substack{a=m+n \\ b=l-n}} \frac{1}{(m!n!l!)^{1/2}} \int d^3k, \cdots \\ &\qquad \times \int d^3k_l\, \phi^{(m,n,l)}(k_1,\ldots,k_l) \\ &\qquad \times V^{*m} a^*(k_1) \cdots a^*(k_l) N^{*n} |0\rangle\,, \end{aligned} \tag{7}$$

which in the Fermi case (1) are restricted to summations over $n = 0,1$ while ($N,V$ boson, $\Theta$ fermion) or ($N,V,\Theta$ bosons) allow $n = 0,1,2,\ldots$. In case of $\Theta$ fermionic, the coefficient function $\phi(k_1,\ldots,k_l)$ is antisymmetric with respect to the momentum variables, while in case of $\Theta$ bosonic, is symmetric. We demand $|a,b\rangle$ to be an eigenfunction of $H$, to be denoted $|\lambda\rangle = |\lambda,a,b\rangle$,

$$H|\lambda\rangle = \lambda|\lambda\rangle, \quad \lambda = \lambda_{(a,b)}\,. \tag{8}$$

To distinguish between the pure Bose version of (1) and the $(N\Theta)$ Fermi case of (1) we shall use the notation $H_B, |\lambda\rangle_B$ and $H, |\lambda\rangle$, respectively. In the pure Bose case by applying $H^B_{\text{int}}$ to $|\lambda\rangle_B$, as given by (7) we arrive at

[a)] Permanent address: Institute of Theoretical Physics, University of Wroclaw, 50-205 Wroclaw, Poland.

 0022-2488/85/030490-05$02.50 © 1985 American Institute of Physics 

$$H^{B}_{\rm int}|\lambda,a,b\rangle_{B}$$
$$= \sum_{\substack{a=m+n\\ b=l-n}} \left\{ \sqrt{(m+1)n\cdot l} \int d^3k_1 \cdots \int d^3k_{l-1} \left[ \int d^3k f(k)\, \phi^{(m,n,l)}_{B}(k,k_1,\ldots,k_{l-1}) \right] |m+1,n-1,k_1,\ldots,k_{l-1}\rangle_{B} \right.$$
$$\left. + \sqrt{m(n+1)(l+1)} \int d^3k_1 \cdots \int d^3k_{l+1}\, f(k_1)\, \phi^{(m,n,l)}_{B}(k_2,\ldots,k_{l+1}) |m-1,n+1,k_1,\ldots,k_{l+1}\rangle_{B} \right\}, \tag{9}$$

$$|m,n,k_1,\ldots k_l\rangle_{B} = \frac{1}{(m!n!l!)^{1/2}} V^{*m} b^{*}(k_1)\cdots b^{*}(k_l) N^{*n}_{B}|0\rangle .$$

In the case of $(N\Theta)$ fermionic, we must have $n = 0,1$, which implies that $|\lambda,a,b\rangle$ is a superposition of the two types of bare states only: $|m = a-1, n = 1, l = b+1\rangle$ and $|m = a, n = 0, l = b\rangle$ at a fixed choice of $a,b$. Consequently,

$$H_{\rm int}|\lambda,a,b\rangle = \int d^3k f(k) V^{*} N a(k) \int d^3k_1 \cdots \int d^3k_{b+1}\, \phi^{(a-1,1,b+1)}(k_1,\cdots,k_{b+1})\cdot|a-1,1,k_1,\cdots,k_{b+1}\rangle$$
$$+ \int d^3k f(k) a^{*}(k) N^{*} V \int d^3k_1 \cdots \int d^3k_b\, \phi^{(a,0,b)}(k_1,\cdots,k_b)\cdot|a,0,k_1,\cdots,k_b\rangle$$
$$= (-1)^b \sqrt{a(b+1)} \int d^3k_1 \cdots \int d^3k_b\, [f(k)\phi^{(a-1,1,b+1)}(k,k_1,\ldots,k_b)]\, |a,0,k_1,\cdots,k_b\rangle$$
$$+ (-1)^{b\pm 1} \sqrt{a(b+1)} \int d^3k_1 \cdots \int d^3k_{b+1} f(k_1) \phi^{(a,0,b)}(k_2,\ldots,k_{b+1})|a-1,1,k_1,\ldots,k_{b+1}\rangle . \tag{10}$$

The particular structure of the interaction term $H_{\rm int}$ of $H$ as given by (1) has intriguing consequences in the Bose case. Namely, $|\lambda,a,b\rangle_{B}$ for all $b$ but with the value of $a$ restricted not to exceed 1: $a \leqslant 1$, can always be composed as a superposition of vectors taken from pairwise orthogonal Hilbert space sectors, each sector being spanned by vectors of the (shorthand) form

$$|a-k,k,b+k\rangle, \quad |a-k+1,k-1,b+k-1\rangle, \quad k \leqslant a \leqslant 1 . \tag{11}$$

In particular let us consider the following contribution to $|\lambda,a,b\rangle_{B}$:

$$|\lambda,a,b\rangle^{(0)}_{B} := \int d^3k, \cdots \int d^3k_{b+1}\, \phi^{(a-1,1,b+1)}_{B}(k_1,\ldots,k_{b+1})\cdot|a-1,1,k_1,\ldots,k_{b+1}\rangle_{B}$$
$$+ \int d^3k_1 \cdots \int d^3k_b\, \phi^{(a,0,b)}_{B}(k_1,\ldots,k_b)\cdot|a,0,k_1,\ldots,k_b\rangle_{B} . \tag{12}$$

The action of $H^{B}_{\rm int}$ on (12) reads as follows:

$$H^{B}_{\rm int}|\lambda,a,b\rangle^{(0)}_{B} = \sqrt{a(b+1)} \int d^3k_1 \cdots \int d^3k_b \left[ \int d^3k f(k)\cdot \phi^{(a-1,1,b+1)}_{B}(k,k_1,\cdots,k_b) \right] |a,0,k_1,\ldots,k_b\rangle_{B}$$
$$+ \sqrt{a(b+1)} \int d^3k_1 \cdots \int d^3k_{b+1} f(k_1) \phi^{(a,0,b)}_{B}(k_2,\ldots,k_{b+1})\cdot|a-1,1,k_1,\cdots,k_{b+1}\rangle_{B} . \tag{13}$$

Any domain spanned by vectors of the form (11) is in fact left invariant by $H^{B}_{\rm int}$.

*Remark:* Let us observe that if to abandon the restriction $a \leqslant 1$, then the action of $H^{B}_{\rm int}$ on $|\lambda,a,b\rangle^{(0)}_{B}$ would produce an additional additive term in (13) following from the application of $(a^{*}N^{*}V)$ to $|a-1,1,k_1,\ldots,k_{b+1}\rangle$. The resulting $|a-2,2,k_1,\ldots,k_{b+2}\rangle$ contribution can be eliminated from further discussion, but the price paid is the modification of the pure Bose Hamiltonian to the form $PH^{B}_{\rm int}P$ with $P = :\exp(-N^{*}_{B}N_{B}): + N^{*}_{B}:\exp(-N^{*}_{B}N_{B}):N_{B}$.

It corresponds to the replacement in $H^{B}_{\rm int}$ of the pure Bose variable $N_{B}$ by the spin-$\frac{1}{2}$ Pauli operator variable $\sigma^{-} = PN_{B}P$, $(\sigma^{+} = PN^{*}_{B}P)$. Hence, we in fact pass then from the pure Bose model to the $(V\Theta)$ Bose, $N$ Fermi version of the Lee model. It is worth emphasizing that though the whole subsequent analysis is made for the pure Bose model with the restriction $a \leqslant 1$ on state vectors, all the arguments apply without any change (up to minor modifications in $H^{B}_{\rm int}$) to the above-mentioned $(V\Theta)$ Bose $N$ Fermi system, where $a$ and $b$ are completely arbitrary. It means that the spectral solution for the $(N\Theta)$ Fermi case produces this as well for the $(N)$ Fermi case. It is also instructive to mention that the interaction of two static fermions with the scalar boson was studied in Ref. 4 in the $N_1 = N^{*}N + V^{*}V = a = 1$ state (sub) space. The subsequent analysis establishes the $\Theta$ Fermi partner for this case.

On the basis of Ref. 5 we know how to establish the eigenvalues and eigenvectors (i.e., $\phi^{(a-1,1,b+1)}$, $\phi^{(a,0,b)}$) for the $(N\Theta)$ Fermi problem. At this point we are guided by our earlier studies of the $(1+1)$-dimensional models,[6] and the joint Bose–Fermi spectral problems arising there. For the exactly soluble Fermi model of Ref. 5 we wish to establish its pure Bose partner, such as the joint spectral problem makes sense.

Let us make use of Refs. 7–9, where relations between linear spaces of symmetric and antisymmetric functions

were investigated. In application to our problem, the formal realization of the isomorphism, invented in Ref. 7 by means of the Friedrichs–Klauder antisymmetric symbol, is best suited. The symbol reads

$$\sigma(k_1,...,k_n) = \pm 1,0 \tag{14}$$

depending on even (+) or odd (−) permutations of momenta, the value 0 occurring if any two momenta coincide. Then

$$\sigma^3 = \sigma, \quad \sigma(1-\sigma^2) = 0, \tag{15}$$

and any symmetric function $f_s(k_1,...,k_n)$ allows[7] for a decomposition

$$f_s(k_1,...,k_n) = [\sigma^2 f_s + (1-\sigma^2) f_s](k_1,...,k_n)$$
$$:= (\overset{1}{f} + \overset{2}{f})(k_1,...,k_n) \tag{16}$$

with the property that

$$\sigma f_s = \sigma \overset{1}{f} = f_a \tag{17}$$

is an antisymmetric function of $n$-momentum variables. The formula (17) establishes an isomorphism between symmetric functions $\overset{1}{f}_s$ (they respect the Pauli exclusion principle since $\overset{1}{f}_s$ vanishes if any two momenta coincide), and their antisymmetric partners $f_a$.

The above isomorphism has been exploited in Ref. 8 to construct an embedding of the CAR algebra representation with generators $[a(p),a^*(q)]_+ = \delta^3(\mathbf{p}-\mathbf{q})$, $[a(k), a(p)]_+ = 0$ in the representation of the CCR algebra generated by $[b(p),b^*(q)]_- = \delta^3(\mathbf{p}-\mathbf{q})$, $[b(k),b(p)]_- = 0$, provided the representation spaces are constructed about the same (generating in the GNS construction sense) Hilbert space vector. We refer to Ref. 8 for the explicit "bosonization" formulas valid in the Fock case (see also Ref. 9). For our purposes the following identity resulting from the CAR- = CAR(CCR) construction of Ref. 8 is necessary:

$$|k_1,...,k_n\rangle_F = (1/\sqrt{n!})a^*(k_1)\cdots a^*(k_n)|0\rangle$$
$$= \sigma(k_1,...,k_n)(1/\sqrt{n!})b^*(k_1)\cdots b^*(k_n)|0\rangle$$
$$= \sigma(k_1,...,k_n)|k_1,...,k_n\rangle_B . \tag{18}$$

Since in (16) we deal with an object $N^*$

$$N^*|k_1,...,k_n\rangle_F = (-1)^n|1,k_1,...,k_n\rangle_F$$
$$= ((-1)^n/\sqrt{n!})a^*(k_1)\cdots a^*(k_n)N^*|0\rangle , \tag{19}$$

an appropriate realization for $N^* = N_F^*$ is necessary. We define

$$N_F^* = (-1)^{\int d^3k\, a^*(k)a(k)}\, N_B^* : \exp(-N_B^*\, N_B) :, \tag{20}$$

which has all the necessary properties, i.e., $N_F^{*2} = 0$ [notice that $:\exp(-N_B^*\, N_B):$ is a projection on the vacuum state for the boson $[N_B, N_B^*]_- = 1$], and anticommutes with the $a^*(k)$'s in (19). Instead of $(-1)^{\int d^3k\, a^*(k)a(k)}$ one can obviously use $\exp i\pi \int d^3k\, a^*(k)a(k)$.

A nice property of the realization (20) is that a Bose representation for (19) is immediate:

$$|1,k_1,...,k_n\rangle_F = (1/\sqrt{n!})\sigma(k_1,...,k_n)b^*(k_1)\cdots b^*(k_n)\cdot N_B^*|0\rangle$$
$$= \sigma(k_1,...,k_n)|1,k_1,...,k_n\rangle_B . \tag{21}$$

The notion of Fermi states of the Bose system acquires thus a meaning in three space dimensions.

A straightforward application of (18) and (21), if combined with (15)–(17), allows us to rewrite formula (16) as follows:

$$H_{\rm int}|\lambda,a,b\rangle = \sqrt{a(b+1)}\int d^3k_1\cdots\int d^3k_b\Big\{(-1)^b$$
$$\times\Big[\int d^3k f(k)\,\phi^{(a-1,1,b+1)}(k,k_1,...,k_b)\Big]$$
$$\times\sigma(k_1,...,k_b)\Big\}\,|a,0,k_1,...,k_b\rangle_B$$
$$+\sqrt{a(b+1)}\int d^3k_1\cdots\int d^3k_{b+1}$$
$$\times\{(-1)^{b+1}f(k_1)\,\phi^{(a,0,b)}(k_2,...,k_{b+1})$$
$$\times\sigma(k_1,...,k_{b+1})\}\cdot|a-1,1,k_1,...,k_{b+1}\rangle_B . \tag{22}$$

Since in (13) and (22) we deal with superpositions of the Bose (bare) basis vectors, the respective expansion coefficients (with respect to this basis system) can be compared.

The formula (13) implies

$${}_B\langle a,0,k_1,\cdots,k_b|H_{\rm int}^B|\lambda,a,b\rangle_B^{(0)}$$
$$= \sqrt{a(b+1)}\int d^3k f(k)\,\phi_B^{(a-1,1,b+1)}(k,k_1,...,k_b) \tag{23}$$

and

$${}_B\langle a-1,1,k_1,...,k_{b+1}|H_{\rm int}^B|\lambda,a,b\rangle_B^{(0)}$$
$$= \sqrt{a(b+1)}({\rm sym})[f(k_1)\,\phi_B^{(a,0,b)}(k_2,...,k_{b+1})] , \tag{24}$$

where

$$({\rm sym}) = S_{b+1} = \frac{1}{(b+1)!}\sum_P P_{b+1}$$

is a symbol of symmetrization with respect to all momentum variables, $\Sigma_P\, P$ stands for a sum over all permutations.

Quite analogously, from (22) we arrive at

$${}_B\langle a,0,k_1,...,k_b|H_{\rm int}|\lambda,a,b\rangle$$
$$= \sqrt{a(b+1)}(-1)^b\int d^3k f(k)\,\phi^{(a-1,1,b+1)}$$
$$\times(k,k_1,...,k_b)\sigma(k_1,...,k_b) \tag{25}$$

and

$${}_B\langle a-1,1,k_1,...,k_{b+1}|H_{\rm int}|\lambda,a,b\rangle$$
$$= \sqrt{a(b+1)}(-1)^{b+1}$$
$$\times({\rm sym})[f(k_1)\,\phi^{(a,0,b)}(k_2,...,k_{b+1})$$
$$\times\sigma(k_1,...,k_{b+1})] . \tag{26}$$

In addition to (sym), let us introduce the antisymmetrization operation

$$({\rm asym}) = A_{b+1} = \frac{1}{(b+1)!}\sum_P (-1)^P P_{b+1} .$$

Both $S$ and $A$ are examples of the Young's idempotent

operators $Y_n$, allowing for a decomposition of any $n$-point function with respect to different types of symmetry

$$f_n = \sum_Y Y_n f_n .$$

We shall exploit a property (Ref. 7, Theorem 2.7), which connects Young's operators $Y_n$ with their duals $Y_n^d$:

$$Y_n \sigma_n = \sigma_n Y_n^d . \tag{27}$$

In particular $S_n^d = A_n$, $A_n^d = S_n$, hence $S_n \sigma_n = \sigma_n A_n$. It means that (26) acquires the form of

$$\begin{aligned}&\sqrt{a(b+1)}(-1)^{b+1}\sigma(k_1,\ldots,k_{b+1})\\&\quad\times\{(\text{asym})[f(k_1)\phi^{(a,0,b)}(k_2,\ldots,k_{b+1})]\}\\&=\sqrt{a(b+1)}(-1)^{b+1}\sigma^2(k_1,\ldots k_{b+1})\\&\quad\times\{(\text{sym})[f(k_1)\phi_s^{(a,0,b)}(k_2,\ldots,k_{b+1})]\},\end{aligned} \tag{28}$$

where

$$\phi_s^{(a,0,b)}(k_2,\ldots,k_{b+1}) = \sigma(k_2,\ldots,k_{b+1})\cdot\phi^{(a,0,b)}(k_2,\ldots,k_{b+1}) . \tag{29}$$

Let us now make an identification,

$$\phi_{\mathrm{B}}^{(m,n,l)}(k_1,\ldots,k_l) = \sigma(k_1,\ldots,k_l)\,\phi^{(m,n,l)}(k_1,\ldots,k_l) , \tag{30}$$

relating the pure Bose and the ($N\Theta$) Fermi expansion coefficients in the above. By virute of (18) and (21) it implies that the Bose vectors (12) upon (30) satisfy

$$\begin{aligned}|\lambda,a,b\rangle_{\mathrm{B}}^{(0)} &= |\lambda,a-1,1,b+1\rangle_{\mathrm{B}} + |\lambda,a,0,b\rangle_{\mathrm{B}}\\&= |\lambda,a-1,1,b+1\rangle + |\lambda,a,0,b\rangle = |\lambda,a,b\rangle ,\end{aligned} \tag{31}$$

i.e., coincide with the respective Fermi vectors in the Fock space. Furthermore, the pure Bose expression (23) reads

$$\begin{aligned}&\sqrt{a(b+1)}\int d^3k f(k)\,\phi_{\mathrm{B}}^{(a-1,1,b+1)}(k,k_1,\ldots,k_b)\\&=\sqrt{a(b+1)}\int d^3k f(k)\,\phi^{(a-1,1,b+1)}\\&\quad\times(k,k_1,\ldots,k_b)\cdot\sigma(k,k_1,\ldots,k_b)\\&=\sqrt{a(b+1)}\int d^3k f(k)\,\phi^{(a-1,b+1)}\\&\quad\times(k,k_1,\ldots,k_b)\cdot\sigma(k_1,\ldots,k_b),\end{aligned} \tag{32}$$

which by a factor $(-1)^b$ differs from the corresponding ($N\Theta$) Fermi expression $(-1)^b(23) = (25)$.

As a result of (30) and (28) the following formula holds true for the ($N\Theta$) Fermi model expression (26):

$$\begin{aligned}&(-1)^{b+1}\sqrt{a(b+1)}(\text{sym})[f(k_1)\phi^{(a,0,b)}\\&\quad\times(k_2,\ldots,k_{b+1})\sigma(k_1,\ldots,k_{b+1})]\\&=\sqrt{a(b+1)}(-1)^{b+1}\sigma^2(k_1,\ldots,k_{b+1})\\&\quad\times\{(\text{sym})[f(k_1)\phi_B^{(a,0,b)}(k_2,\ldots,k_{b+1})]\},\end{aligned} \tag{33}$$

which upon dropping out a factor $(-1)^{b+1}$ is exactly the $\sigma^2 F$ contribution to the decomposition formula $[\sigma^2 F + (1-\sigma^2)F]$ valid for the pure Bose expression $(25) = F$. By virtue of (15) the decomposition is orthogonal.

Since, because of (31) we have

$$H_0^{\mathrm{B}}|\lambda,a,b\rangle = H_0|\lambda,a,b\rangle ; \tag{34}$$

the relevant information about the relationships between the Bose and Fermi spectral problems comes from the interaction terms.

By virtue of (31) we arrive at

$$\begin{aligned}&H_{\mathrm{int}}^{\mathrm{B}}|\lambda,a,b\rangle_{\mathrm{B}}^{(0)}\\&\quad= H_{\mathrm{int}}^{\mathrm{B}}|\lambda,a,b\rangle\\&\quad=(-1)^b\int d^3k f(k)V^*Na(k)|\lambda,a-1,1,b+1\rangle\\&\qquad+(-1)^{b+1}\int d^3k f(k)a^*(k)\\&\qquad\times N^*V|\lambda,a,0,b\rangle+(-1)^{b+1}|R\rangle ,\end{aligned} \tag{35}$$

where

$$\begin{aligned}|R\rangle = &\int d^3k_1\cdots\int d^3k_{b+1}\sqrt{a(b+1)}\\&\times[1-\sigma^2(k_1,\ldots,k_{b+1})]\\&\times\{(\text{sym})[f(k_1)\phi_{\mathrm{B}}^{(a,0,b)}(k_2,\ldots,k_{b+1})]\}\\&\times|\lambda,a,0,k_1,\ldots,k_{b+1}\rangle_{\mathrm{B}} .\end{aligned} \tag{36}$$

Let us however, recall that because of (30), $\phi_{\mathrm{B}} = \sigma\cdot\phi$ and that

$$(\text{sym}) = \frac{1}{(b+1)!}\sum_P P_{b+1} ,$$

so that in (36) we encounter products of the form

$$[1-\sigma^2(k_1,\ldots,k_{b+1})]\cdot\sigma(k_{i_1},\ldots,k_{i_b}) \tag{37}$$

with $k_i$'s taken from the set $(k_1,\ldots,k_{1+b})$. But (37) either identically vanishes, or gives a nonzero contribution to (36) on the set of measure zero only. Hence, $|R\rangle = 0$.

If we introduce the notation

$$\begin{aligned}H_{\pm}^{\mathrm{B}} &= \int d^3k f(k)[V^*Nb(k) + b^*(k)N^*V] ,\\H_{\pm}^{\mathrm{F}} &= \int d^3k f(k)[V^*Na(k) + a^*(k)N^*V] ,\end{aligned} \tag{38}$$

then (35) appears as an example of a few more relations between Bose and Fermi Hamiltonians

$$\begin{aligned}H_{+}^{\mathrm{B}}|\lambda,a,b\rangle &= (-1)^b H_{-}^{\mathrm{F}}|\lambda,a,b\rangle ,\\H_{-}^{\mathrm{B}}|\lambda,a,b\rangle &= (-1)^b H_{+}^{\mathrm{F}}|\lambda,a,b\rangle , \quad a\leqslant 1 .\end{aligned} \tag{39}$$

After accounting for the contribution of $H_0$, the complete Hamiltonians of the form $H_0 \pm H_{\mathrm{int}}$ become related as follows: $N_1 \leqslant 1$, $N_2 = b$,

$b$ even,

$$\begin{aligned}(H_0^{\mathrm{B}} + H_{+}^{\mathrm{B}})|\lambda,a,b\rangle &= (H_0^{\mathrm{F}} + H_{-}^{\mathrm{F}})|\lambda,a,b\rangle ,\\(H_0^{\mathrm{B}} + H_{-}^{\mathrm{B}})|\lambda,a,b\rangle &= (H_0^{\mathrm{F}} + H_{+}^{\mathrm{F}})|\lambda,a,b\rangle ,\end{aligned} \tag{40}$$

$b$ odd,

$$\begin{aligned}(H_0^{\mathrm{B}} \pm H_{+}^{\mathrm{B}})|\lambda,a,b\rangle &= (H_0^{\mathrm{F}} \mp H_{-}^{\mathrm{F}})|\lambda,a,b\rangle ,\\(H_0^{\mathrm{B}} \pm H_{-}^{\mathrm{B}})|\lambda,a,b\rangle &= (H_0^{\mathrm{F}} \mp H_{+}^{\mathrm{F}})|\lambda,a,b\rangle .\end{aligned} \tag{41}$$

In this number the pure Bose problem $H_B$ of Refs. 3 and 4 is

identified with $H_0^B + H_+^B$ and the $(N\Theta)$ Fermi problem of Ref. 5 with $H_0^F + H_+^F$.

Relations (40) and (41) prove that for the family of four $(N\Theta)$ Fermi models, there is a corresponding family of pure Bose models, with the property that in the state space of the Bose system there exists a projection $\Pi$ such that the eigenvalue problem for $h_F$ can be solved in the range of $\Pi$, and

$$[h_B, \Pi]_- = 0, \quad h_F \equiv \Pi h_B \Pi,$$

$$h_B = \Pi h_B \Pi + (1 - \Pi) h_B (1 - \Pi). \tag{42}$$

Here $h_B$ stands for the Bose, while $h_F$ stands for the respective Fermi Hamiltonian. Complementary studies of $(1 + 1)$-dimensional field theory models sharing the property (42) can be found in Refs. 10, 11, and 6.

The results (40) and (41) mean in particular that the pure Bose model

$$h_B = m_0 V^* V + \int d^3k\, k \cdot b^*(k) b(k) + \int d^3k f(k) [V^* N b(k) - b^*(k) N^* V] \tag{43}$$

has eigenvectors and eigenvalues common with the $(N\Theta)$ Fermi model solved by Sudarshan and Tata[5]: all $b$ even eigenvectors of $h_F$ of (1) are exact eigenvectors with the same eigenvalues for the pure Bose Hamiltonian (43). The odd eigenvectors of $h_F$ are shared with

$$h'_B = m_0 V^* V + \int d^3k\, k \cdot b^*(k) b(k) - \int d^3k f(k) [V^* N b(k) - b^*(k) N^* V]. \tag{44}$$

One should also notice that upon solving the eigenvalue problem for the Fermi Hamiltonians $(H_0^F \pm H_-^F)$ we would have received a partial spectral solution for the pure Bose model of the Bolsterli–Lee type.[1–3] Unfortunately the Bose Hamiltonian $(H_0^B \pm H_+^B)$ is related to the Fermi Hamiltonian $(H_0^F \mp H_-^F)$ and likewise $(H_0^B \pm H_-^B)$ is related to $(H_0^F \mp H_+^F)$. Thus, the spectral solutions of Sudarshan and Tata cannot be used to obtain the solution of the spectrum of the Bolsterli–Lee model: This entails the solution of the problem for $(H_0^F \pm H_-^F)$. For this form of the fermion Hamiltonian, however, the simple form of Eq. (3.2b) in Ref. 5 does not arise since the right-hand side now entails the operator $V^*V - N^*N$ instead of the eigenoperator $V^*V + N^*N = N_1$. (It was the eigenoperator structure that led to the simple solution in Ref. 5.) This is exactly the structure for the corresponding equation that would occur if one were directly dealing with the Hamiltonian for the Bolsterli–Lee model.

One more problem arises in connection with the (formal) non-self-adjointness of operators $H_-^{B,F}$. However, since we relate them to self-adjoint operators via (42) it appears that projections $\Pi$ identify the appropriate (Hermicity) domains.

## ACKNOWLEDGMENTS

I would like to express my thanks to Professor J-P. Antoine for his hospitality in Louvain-la-Neuve, and the referee of this paper for his comments.

[1] T. D. Lee, Phys. Rev. **95**, 1329 (1954).
[2] S. Schweber, *Introduction to Relativistic Quantum Field Theory* (Row, Evanston, IL, 1961).
[3] M. Bolsterli, Phys. Rev. **166**, 1760 (1968); and C. A. Nelson, J. Math. Phys. **13**, 1051 (1972).
[4] D. Fivel, J. Math. Phys. **11**, 699 (1970).
[5] E. C. G. Sudarshan and X. Tata, J. Math. Phys. **24**, 1810 (1983).
[6] P. Garbaczewski, J. Math. Phys. **25**, 862 (1984).
[7] P. Garbaczewski and J. Rzewuski, Rep. Math. Phys. **6**, 431 (1974).
[8] P. Garbaczewski, Commun. Math. Phys. **43**, 131 (1975).
[9] P. Garbaczewski, J. Math. Phys. **19**, 619 (1978).
[10] P. Garbaczewski, J. Math. Phys. **24**, 651 (1983).
[11] P. Garbaczewski, Nucl. Phys. B **218**, 321 (1983).

# Semiclassical statistical mechanics of two-dimensional fluid mixture of hard disks

B. M. Mishra[a] and S. K. Sinha
*Department of Physics, L. S. College, Bihar University, Muzaffarpur 842001, India*

(Received 30 November 1983; accepted for publication 9 August 1984)

High-temperature expansions for the density-independent part of the radial distribution function and the first-order density correction to it are obtained for a two-dimensional binary mixture of hard disks. The "excess" quantum corrections to the second and third virial coefficients and excess free energy are also discussed. It is found that the "excess" quantum effect depends on the concentration and the diameter ratio.

## I. INTRODUCTION

Two-dimensional quantum fluids have been a subject of considerable interest in recent years.[1–5] Ideally a two-dimensional system cannot be achieved. However, a strictly two-dimensional picture has been used in predicting the properties of the adsorbed film.[6,7] Another reason for this interest is to study the effects of reduced dimensionality on the nature of the phase transition.[8,9] Most of the fluids found in the literature have been treated classically, because the quantum effects for them are so small as to be negligible. There are some fluids for which deviations from classical behavior are observed at low temperatures. There are two types of quantum effects: (i) diffraction effects, which are linked to the wave nature of the particles in the fluid, and (ii) exchange effects due to the statistics (Bose–Einstein or Fermi–Dirac) obeyed by the particles. The exchange or symmetry effects are very small for all fluids, except for liquid helium below 5 °K (Ref. 10). The diffraction effects, on the other hand, are appreciable even at high temperatures.

At high temperatures, where quantum effects are small and can be treated as a correction to the classical system, the fluid is treated semiclassically. The usual way of studying the properties of the semiclassical system is to expand them in powers of Planck's constant $\hbar$ (see Refs. 11 and 12). The first term of the series is the classical value and other terms arise due to the quantum effects. We use the Wigner–Kirkwood (WK) expansion method[11] for an analytic potential and the Hemmer–Jancovici (HJ) method[12] for a hard-sphere (or hard-disk) potential. This approach has been extensively used for two-dimensional one-component fluids.[4,5] However a two-dimensional quantum fluid mixture has not yet been investigated systematically.

The classical hard-disk mixture has been studied over a wide range of densities.[13,14] No work is available for the hard-disk mixture in the semiclassical limit. Recently Singh and Sinha[15,16] have used their method to calculate the equilibrium properties of the three-dimensional fluid mixture of hard spheres.

In the present paper we study the equilibrium properties of the hard-disk mixture in the semiclassical limit.

In Sec. II, we describe a basic theory for calculating the equilibrium properties of a two-dimensional binary mixture, the constituent particles of which interact via a hard-core potential. The explicit expressions for the free energy and the radial distribution function (RDF) are reported there. Using this expression for the RDF, we calculate the density-independent RDF and first-order density correction to it for the hard-disk mixture in Sec. III. Section IV is devoted to calculate the second and third virial coefficients for the hard-disk mixture. In Sec. V, we develop another method to calculate the thermodynamic properties of a dense fluid mixture of hard disks.

The exchange effects that arise due to statistics are ignored here.

## II. BASIC THEORY

We consider a two-dimensional fluid mixture of $N_1$ hard-disk molecules of species 1 and $N_2$ hard-disk molecules of species 2, such that the total number of molecules is $N = N_1 + N_2$. We assume that the constituent molecules of both species differ in size. In addition, the interaction between unlike molecules is also assumed to be hard-disk interaction. The Hamiltonian of the system is

$$\hat{H}_N = -\frac{\hbar^2}{2}\sum_{j=1}^{2}\frac{1}{m_j}\sum_{i=1}^{N_j}\nabla_i^2 + \sum_{\alpha,\beta=1}^{2}\sum_{i<k}u_{\alpha\beta}(i,k)\,. \tag{2.1}$$

Here, $u_{\alpha\beta}(i,k)$ is the pair potential between particle $i$ of species $\alpha$ and particle $k$ of species $\beta$.

For the grand canonical ensemble, the density operator $\hat{\rho}$ for a fluid mixture is defined as

$$\hat{\rho} = \exp\left[-\beta\left(\hat{H}_N - \sum_{j=1}^{2}\mu_j\hat{N}_j\right)\right]\Xi^{-1}, \tag{2.2}$$

where $\beta = (kT)^{-1}$, $\mu_j$ and $\hat{N}_j$ are, respectively, the chemical potential and number operator of species $j$, and $\Xi$ is the normalization factor, known as the quantum mechanical grand canonical partition function

$$\Xi = \mathrm{tr}\left\{\exp\left[-\beta\left(\hat{H}_N - \sum_j\mu_j\hat{N}_j\right)\right]\right\}. \tag{2.3}$$

Let $\{\psi_x\}$ be a complete set of (properly symmetrized) orthogonal wave functions of the system, then

[a] Permanent address: Department of Physics, C.M. (Science) College, Darbhanga, India.

 0022-2488/85/030495-10$02.50 © 1985 American Institute of Physics 

$$\Xi = \sum_{\{N_j\}} \sum_{x} \int_{v} \psi_x^*(\bar{r}_1,...,\bar{r}_N) \times \exp\left[ -\beta \left( \hat{H}_N - \sum_j \mu_j \hat{N}_j \right) \right] \times \psi_x(\bar{r}_1,...,\bar{r}_N) \prod_{i=1}^{N} d\bar{r}_i \,. \tag{2.4}$$

It can be rewritten in the form

$$\Xi = \sum_{\{N_j\}} \left( \prod_{j=1}^{2} \frac{z_j^{N_j}}{N_j!} \right) \int \cdots \int W_N(1,2,...,N) \prod_{i=1}^{N} d\bar{r}_i \,, \tag{2.5}$$

where

$$z_j = \lambda_j^{-2} e^{\beta\mu_j} , \tag{2.6a}$$

$$\lambda_j = (2\pi\hbar^2 \beta / m_j)^{1/2} , \tag{2.6b}$$

$$W_N(1,2,...,N) = \prod_{j=1}^{2} N_j! \lambda_j^{2N_j} \sum_x \psi_x^*(\bar{r}_1,...,\bar{r}_N) \times \exp(-\beta \hat{H}_N) \psi_x(\bar{r}_1,...,\bar{r}_N) = \prod_{j=1}^{2} N_j! \lambda_j^{2N_j} \langle 1,2,...,N | \times \exp(-\beta \hat{H}_N) | 1,2,...,N \rangle \,, \tag{2.7}$$

Here $z_j$ and $\lambda_j$ are, respectively, the fugacity and thermal wavelength of species $j$, and $W_N$, defined by Eq. (2.7), is the Slater sum for a two-dimensional fluid mixture.

The quantum mechanical pair correlation function for a binary mixture in the grand canonical ensemble is

$$\rho_\alpha \rho_\beta g_{\alpha\beta}(1,2) = \Xi^{-1} \sum \left[ \prod_{j=1}^{2} \frac{z_j^{N_j}}{(N_j - \delta_{\alpha j} - \delta_{\beta j})!} \right] \times \int \cdots \int W_N(1,2,...,N) \prod_{i=3}^{N} d\bar{r}_i \,. \tag{2.8}$$

In the semiclassical limit (i.e., at high temperature), when the deviation from the classical behavior is small, the Slater sum can be written as[12]

$$W_N = W_N^c \, W_N^m \,, \tag{2.9}$$

where

$$W_N^c(1,2,...,N) = \exp\left[ -\beta \sum_{\alpha,\gamma=1}^{2} \sum_{i<k} u_{\alpha\gamma}(i,k) \right] \tag{2.10}$$

Thus, is the Boltzmann factor and $W_N^m$ is a function which measures the deviation from the classical behavior. We can express $W_N^m$ in terms of the "modified" Ursell function $U^m_{\alpha\beta\gamma\cdots\delta}(1,2,...,l)$. Thus,

$$W_1^m(1;\alpha) = U_\alpha^m(1) = 1 \,, \tag{2.11a}$$

$$W_2^m(1,2;\alpha,\beta) = 1 + U_{\alpha\beta}^m(1,2) \,, \tag{2.11b}$$

$$W_3^m(1,2,3;\alpha,\beta,\gamma) = 1 + U_{\alpha\beta}^m(1,2) + U_{\alpha\gamma}^m(1,3) + U_{\beta\gamma}^m(2,3) + U_{\alpha\beta\gamma}^m(1,2,3) \,, \tag{2.11c}$$

$$\vdots$$

$$W_N^m(1,2,...,N;\alpha,\beta,...,\delta) = 1 + \sum_{\alpha,\beta=1}^{2} \sum_{i<j} U_{\alpha\beta}^m(i,j) + \sum_{\alpha,\beta,\gamma=1}^{2} \sum_{i<j<k} U_{\alpha\beta\gamma}^m(i,j,k) + \sum_{\alpha,\beta,\gamma,\delta=1}^{2} \sum_{i<j\neq k<l} U_{\alpha\beta}^m(i,j) U_{\gamma\delta}^m(k,l) + \cdots \,. \tag{2.11d}$$

The above equations can be solved successively for $U_{\alpha\beta}^m$, $U_{\alpha\beta\gamma}^m$, ...:

$$U_{\alpha\beta}^m(1,2) = W_2^m(1,2;\alpha,\beta) - 1 \,, \tag{2.12a}$$

$$U_{\alpha\beta\gamma}^m(1,2,3) = W_3^m(1,2,3;\alpha,\beta,\gamma) - W_2^m(1,2;\alpha,\beta) - W_2^m(1,3;\alpha,\gamma) - W_2^m(2,3;\beta,\gamma) + 2 \,. \tag{2.12b}$$

By solving the quantum mechanical $l$-body problem, one can obtain, in principle, the $l$-body "modified" Ursell function $U^m_{\alpha\beta\gamma\cdots\delta}(1,2,...,l)$. Unfortunately actual evaluation is too involved to be feasible. For the hard-disk mixture, $U_{\alpha\beta}^m(r)$ is evaluated in the next section.

From Eqs. (2.9) and (2.11d) we can obtain the expression for $W_N$. This is used in Eq. (2.5) to obtain an expression for the free energy. The result is

$$\frac{\beta A}{N} = \frac{\beta A^c}{N} - \frac{1}{2}\rho \sum_{\alpha,\beta=1}^{2} x_\alpha x_\beta \int g_{\alpha\beta}^c(1,2) U_{\alpha\beta}^m(1,2) d\bar{r}_2 - \frac{1}{6}\rho^2 \sum_{\alpha,\beta,\gamma=1}^{2} x_\alpha x_\beta x_\gamma \int g_{\alpha\beta\gamma}^c(1,2,3) U_{\alpha\beta\gamma}^m(1,2,3) \, d\bar{r}_2 \, d\bar{r}_3 - \frac{1}{8}\rho^3 \sum_{\alpha,\beta,\gamma,\delta=1}^{2} x_\alpha x_\beta x_\gamma x_\delta \int [g_{\alpha\beta\gamma\delta}^c(1,2,3,4) - g_{\alpha\beta}^c(1,2) g_{\gamma\delta}^c(3,4)] \, U_{\alpha\beta}^m(1,2) U_{\gamma\delta}^m(3,4) \, d\bar{r}_2 \, d\bar{r}_3 \, d\bar{r}_4 + \frac{\rho}{8\beta} K^c \sum_{\alpha,\beta,\gamma=1}^{2} x_\alpha x_\beta x_\gamma \left\{ \int U_{\alpha\beta}^m(1,2) \frac{\partial}{\partial\rho} [\rho^2 g_{\alpha\beta}^c(1,2)] \, d\bar{r}_2 \right\} \left\{ \int U_{\alpha\gamma}^m(3,4) \frac{\partial}{\partial\rho} [\rho^2 g_{\alpha\gamma}^c(3,4)] \, d\bar{r}_4 \right\} + O(\lambda_{\alpha\beta}^3) \,, \tag{2.13}$$

where

$$K^c = \frac{1}{\rho} \left( \frac{\partial\rho}{\partial P^c} \right)_\beta \tag{2.14}$$

is the isothermal compressibility of the classical fluid mixture and is given by the relation[15]

$$\frac{\rho K^c}{\beta} = 1 + \rho \sum_{\alpha,\beta} x_\alpha x_\beta \int [g_{\alpha\beta}(r) - 1] \, d\bar{r} \,. \tag{2.15}$$

Here $A^c$ and $g^c_{\alpha\beta\cdots\delta}(1,2,\ldots,l)$ are, respectively, the free energy and $l$-particle distribution function of the classical two-dimensional fluid mixture, $\rho$ is the number density, and $x_\alpha$ is the concentration of the species $\alpha$ which is defined as the ratio of the number of particles of species $\alpha$ and the total number of particles in the system, i.e.,

$$x_\alpha = N_\alpha/N = \rho_\alpha/\rho\,.$$

From Eq. (2.8), we find the following expression for the radial distribution function:

$$\begin{aligned} g_{\alpha\beta}(1,2) &= g^c_{\alpha\beta}(1,2)\,[1+U^m_{\alpha\beta}(1,2)] + \rho\sum_{\gamma=1}^{2} x_\gamma \int g^c_{\alpha\beta\gamma}(1,2,3)\,[U^m_{\alpha\beta\gamma}(1,2,3)+U^m_{\alpha\gamma}(1,3)+U^m_{\beta\gamma}(2,3)]\,d\bar{r}_3 \\ &+\frac{1}{2}[1+U_{\alpha\beta}(1,2)]\,\rho^2\sum_{\gamma,\delta=1}^{2} x_\gamma x_\delta \int [g^c_{\alpha\beta\gamma\delta}(1,2,3,4)-g^c_{\alpha\beta}(1,2)\,g^c_{\gamma\delta}(3,4)]\,U^m_{\gamma\delta}(3,4)\,d\bar{r}_3\,d\bar{r}_4 \\ &-\frac{K^c}{4\beta}[1+U_{\alpha\beta}(1,2)]\left[\frac{\partial}{\partial\rho}[\rho^2 g^c_{\alpha\beta}(1,2)]\left(\sum_{\gamma=1}^{2} x_\gamma\left\{\int U^m_{\alpha\gamma}(3,4)\frac{\partial}{\partial\rho}[\rho^2 g^c_{\alpha\gamma}(3,4)]\,d\bar{r}_4\right.\right.\right. \\ &\left.\left.\left.+\int U^m_{\beta\gamma}(3,4)\frac{\partial}{\partial\rho}[\rho^2 g^c_{\beta\gamma}(3,4)]\,d\bar{r}_4\right\}\right)\right] + O(\lambda^2_{\alpha\beta})\,. \end{aligned} \tag{2.16}$$

Equations (2.13) and (2.16) are similar in their functional forms to those obtained for a three-dimensional fluid mixture.[15]

## III. RADIAL DISTRIBUTION FUNCTION OF A DILUTE HARD-DISK MIXTURE

This section is concerned with the evaluation of the radial distribution function of a binary mixture of hard disks. For such a system the pair interaction is given by

$$u_{\alpha\beta}(r) = \begin{cases} \infty\,, & r<d_{\alpha\beta}\,, \\ 0\,, & r>d_{\alpha\beta}\,, \end{cases} \tag{3.1}$$

where $d_{\alpha\beta}$ is the diameter between the hard disks of species $\alpha$ and $\beta$. For, unlike interaction, $d_{\alpha\beta}$ is given by[17]

$$d_{12} = \tfrac{1}{2}(d_{11}+d_{22})(1+\Delta)\,, \tag{3.2}$$

where $\Delta = 0$ for an additive hard-disk mixture, whereas $|\Delta| > 0$ for a nonadditive hard-disk mixture.

### A. Density-independent radial distribution function

The density-independent radial distribution function (RDF) for a two-dimensional mixture is given by

$$g^0_{\alpha\beta}(r) = 2\lambda^2_{\alpha\beta}\langle\bar{r}|\exp(-\beta\hat{H}_{\rm rel})|\bar{r}\rangle\,, \tag{3.3}$$

where $\hat{H}_{\rm rel}$ is the relative Hamiltonian of two particles of species $\alpha$ and $\beta$,

$$\hat{H}_{\rm rel} = -(\hbar^2/m_{\alpha\beta})\nabla^2 + u_{\alpha\beta}(r)\,, \tag{3.4a}$$

and $\lambda_{\alpha\beta}$ is the thermal wavelength associated with the particles of species $\alpha$ and $\beta$. Thus

$$\lambda_{\alpha\alpha} = (2\pi\hbar^2\beta/m_{\alpha\alpha})^{1/2}\,, \tag{3.4b}$$

$$\lambda_{12} = [(\lambda^2_{11}+\lambda^2_{22})/2]^{1/2}\,. \tag{3.4c}$$

In Eq. (3.3) statistics are not taken into account. In the classical limit, Eq. (3.3) reduces to $g^{\rm cl_0}_{\alpha\beta}(r)$, given by

$$g^{\rm cl_0}_{\alpha\beta}(r) = \exp[-\beta u_{\alpha\beta}(r)]\,. \tag{3.5}$$

At high temperature, the density-independent RDF can be written as

$$g^0_{\alpha\beta}(r) = [1+U^m_{\alpha\beta}(r)]\,, \tag{3.6}$$

where $U^m_{\alpha\beta}(r)$ is the two-body "modified" Ursell function, which is given by[4]

$$U^m_{\alpha\beta}(r) = \xi^0_{\alpha\beta}+\xi^1_{\alpha\beta}+\xi^2_{\alpha\beta}+\cdots\,, \quad \text{for } r>d_{\alpha\beta}\,, \tag{3.7}$$

with

$$\xi^0_{\alpha\beta} = -\exp[-Q^2_{\alpha\beta}]\,, \tag{3.8a}$$

$$\xi^1_{\alpha\beta} = (1/2\sqrt{2})(\lambda_{\alpha\beta}/d_{\alpha\beta})Q^2_{\alpha\beta}\,\mathrm{erfc}(Q_{\alpha\beta})\,, \tag{3.8b}$$

$$\xi^2_{\alpha\beta} = (1/8\pi)(\lambda_{\alpha\beta}/d_{\alpha\beta})^2 Q^2_{\alpha\beta}\,[Q^2_{\alpha\beta}\exp(-Q^2_{\alpha\beta}) - \sqrt{\pi}Q_{\alpha\beta}(2+Q^2_{\alpha\beta})\mathrm{erfc}(Q_{\alpha\beta})]\,, \tag{3.8c}$$

where

$$Q_{\alpha\beta} = [(2\pi)^{1/2}/(\lambda_{\alpha\beta}/d_{\alpha\beta})](R-1)\,, \quad R = r/d_{\alpha\beta}\,.$$

### B. First-order density correction to the radial distribution function

The $l$-particle distribution function for a fluid mixture can be expanded in powers of density $\rho$ as[14,18]

$$\begin{aligned} &g_{\alpha\beta}(1,2,\ldots,l) \\ &\quad = \exp\Big[-\beta\sum_{\alpha,\beta=1}^{2}\sum_{i<j}u_{\alpha\beta}(i,j)\Big] \\ &\qquad\times\Big[\sum_{n=0}^{\infty}\rho^n\sum_{\gamma,\delta,\ldots=1}^{2}x_\gamma x_\delta\cdots a^{(n)}_{\alpha\cdots\beta}(1,2,\ldots,l)\Big]\,, \end{aligned} \tag{3.9}$$

where the coefficient $a^{(n)}_{\alpha\cdots\beta}(1,2,\ldots,l)$ is the cluster integral involving $n$ field points and $l$ base points. Equation (3.9) is valid for both classical and quantum fluid mixtures. Substituting Eqs. (2.9) and (3.9) in Eq. (2.11), we get the following expansion coefficients for $g_{\alpha\beta}(1,2)$:

$$a^{(0)}_{\alpha\beta}(1,2) = [1+U^m_{\alpha\beta}(1,2)]\,a^{c^{(0)}}_{\alpha\beta}(1,2)\,, \tag{3.10}$$

$$\begin{aligned} &a^{(1)}_{\alpha\gamma\beta}(1,2) \\ &\quad = a^{c^{(1)}}_{\alpha\gamma\beta}(1,2)\,[1+U^m_{\alpha\beta}(1,2)] \\ &\qquad -[1+U^m_{\alpha\beta}(1,2)]\int\{\exp[-\beta u_{\alpha\gamma}(1,3)] \\ &\qquad\times U^m_{\alpha\gamma}(1,3)+\exp[-\beta u_{\beta\gamma}(2,3)]\,U^m_{\beta\gamma}(2,3)\}\,d\bar{r}_3 \\ &\qquad +\int\exp\{-\beta[u_{\alpha\gamma}(1,3)+u_{\beta\gamma}(2,3)]\}\,[U^m_{\alpha\gamma\beta}(1,2,3) \\ &\qquad +U^m_{\alpha\gamma}(1,3)+U^m_{\beta\gamma}(2,3)]\,d\bar{r}_3\,, \end{aligned} \tag{3.11}$$

and so on. Here

$$a^{c(0)}_{\alpha\beta}(1,2) = 1\,, \tag{3.12a}$$

$$a^{c(1)}_{\alpha\gamma\beta}(1,2) = \int f_{\alpha\gamma}(1,3) f_{\beta\gamma}(2,3)\, d\bar{r}_3\,, \tag{3.12b}$$

where

$$f_{\alpha\beta}(i,j) = \exp\,[\,-\beta u_{\alpha\beta}(i,j)] - 1\,. \tag{3.13}$$

We write the RDF in the form

$$g_{\alpha\beta}(1,2) = g^0_{\alpha\beta}(1,2) + \rho g^1_{\alpha\beta}(1,2) + O(\rho^2)\,, \tag{3.14}$$

where $g^0_{\alpha\beta}(1,2)$ is the density-independent part of the RDF and $g^1_{\alpha\beta}(1,2)$ is the first-order density correction to it.

For a hard-disk mixture, the first-order density correction to the RDF is given by

$$g^1_{\alpha\beta}(1,2) = \sum_{\gamma=1}^{2} x_\gamma a^{(1)}_{\alpha\gamma\beta}(1,2)\,. \tag{3.15}$$

For such a system, $a^{(1)}_{\alpha\gamma\beta}(1,2)$ is obtained from Eq. (3.11) as

$$a^{(1)}_{\alpha\gamma\beta}(1,2) = a^{c(1)}_{\alpha\gamma\beta}(1,2)\,[1 + U^m_{\alpha\beta}(1,2)] + A_{\alpha\gamma\beta}(1,2)\,, \tag{3.16a}$$

where

$$A_{\alpha\gamma\beta}(1,2) = \int [\,U^m_{\alpha\gamma\beta}(1,2,3) - U^m_{\alpha\beta}(1,2)U^m_{\alpha\gamma}(1,3) - U^m_{\alpha\beta}(1,2)U^m_{\beta\gamma}(2,3)]\, d\bar{r}_3\,. \tag{3.16b}$$

For a hard-disk mixture, Eq. (3.12b) is evaluated as

$$\begin{aligned} a^{c(1)}_{\alpha\gamma\beta}(1,2) &= [(\pi/2)\,(d^2_{\alpha\gamma} + d^2_{\beta\gamma}) - d^2_{\alpha\gamma}\,\{\sin^{-1}(\cos\theta_\alpha) \\ &\quad + \tfrac{1}{2}\sin(2\theta_\alpha)\} - d^2_{\beta\gamma}\,\{\sin^{-1}(\cos\theta_\beta) \\ &\quad + \tfrac{1}{2}\sin(2\theta_\beta)\}]\,, \quad \text{for } d_{\alpha\beta} < r_{12} < d_{\alpha\gamma} + d_{\beta\gamma}\,, \\ &= 0\,, \quad \text{for } r_{12} > d_{\alpha\gamma} + d_{\beta\gamma}, \end{aligned} \tag{3.17}$$

where

$$\theta_\alpha = \cos^{-1}\,((r^2_{12} + d^2_{\alpha\gamma} - d^2_{\beta\gamma})/2r_{12}\,d_{\alpha\gamma})\,, \tag{3.18a}$$

$$\theta_\beta = \cos^{-1}((r^2_{12} + d^2_{\beta\gamma} - d^2_{\alpha\beta})/2r_{12}d_{\beta\gamma})\,. \tag{3.18b}$$

For a one-component fluid, where $d_{\alpha\gamma} = d_{\beta\gamma} \equiv d$, Eq. (3.17) reduces to[19]

$$a^{c(1)}(1,2) = \begin{cases} d^2\left[\pi - 2\sin^{-1}\left(\dfrac{r_{12}}{2d}\right) - \dfrac{r_{12}}{2d}\left(4 - \left(\dfrac{r^2_{12}}{d^2}\right)\right)^{1/2}\right], \\ \quad \text{for } d < r_{12} < 2d\,, \\ 0\,, \quad \text{for } r_{12} > 2d\,. \end{cases} \tag{3.19}$$

In order to evaluate $A_{\alpha\gamma\beta}(1,2)$, we split the range of $r_{12}$ into a number of intervals and consider $A_{\alpha\gamma\beta}(1,2)$ in each interval.

*(i) At* $r_{12} = d_{\alpha\beta}$: $U^m_{\alpha\gamma\beta}(1,2,3)$ can be written as[4]

$$U^m_{\alpha\gamma\beta}(1,2,3) = W^m_3(1,2,3;\alpha,\beta,\gamma) - 1 - U^m_{\alpha\beta}(1,2) - U^m_{\alpha\gamma}(1,3) - U^m_{\beta\gamma}(2,3)\,. \tag{3.20}$$

At $r_{12} = d_{\alpha\beta}$, $U^m_{\alpha\beta}(1,2) = -1$ and $W^m_3 = 0$. Thus

$$U^m_{\alpha\beta\gamma}(1,2,3) = -U^m_{\alpha\gamma}(1,3) - U^m_{\beta\gamma}(2,3)\,.$$

Substituting this in Eq. (3.16b), we get

$$A_{\alpha\gamma\beta}(1,2) = 0$$

and

$$a^{(1)}_{\alpha\gamma\beta}(1,2) = 0\,. \tag{3.21}$$

*(ii) For* $d_{\alpha\beta} < r_{12} < d_{\alpha\beta} + \lambda_{\alpha\beta}$: In this configuration, the first-order contribution of $U^m_{\alpha\beta\gamma}(1,2,3)$ comes only from that region in which either $r_{13} > d_{\alpha\gamma} + \lambda_{\alpha\gamma}$ or $r_{23} > d_{\beta\gamma} + \lambda_{\beta\gamma}$. Using the superposition approximation[4,20]

$$\begin{aligned} U^m_{\alpha\beta\gamma}(1,2,3) &= U^m_{\alpha\beta}(1,2)U^m_{\alpha\gamma}(1,3) + U^m_{\alpha\beta}(1,2)U^m_{\beta\gamma}(2,3) \\ &\quad + U^m_{\alpha\gamma}(1,3)U^m_{\beta\gamma}(2,3) + U^m_{\alpha\beta}(1,2)U^m_{\alpha\gamma}(1,3)U^m_{\beta\gamma}(2,3)\,, \end{aligned} \tag{3.22}$$

Eq. (3.16b) can be written as

$$A_{\alpha\gamma\beta}(1,2) = [1 + U^m_{\alpha\beta}(1,2)] \times \int U^m_{\alpha\gamma}(1,3)U^m_{\beta\gamma}(2,3)\, d\bar{r}_3\,. \tag{3.23}$$

For $r_{ij} > d_{\alpha\beta} + \lambda_{\alpha\beta}$, $U^m_{\alpha\beta}(i,j) \sim 0$ and we get

$$A_{\alpha\gamma\beta}(1,2) = 0$$

and

$$a^{(1)}_{\alpha\gamma\beta}(1,2) = [1 + U^m_{\alpha\beta}(1,2)]\, a^{c(1)}_{\alpha\gamma\beta}(1,2)\,. \tag{3.24}$$

*(iii) For* $r_{12} > d_{\alpha\beta} + \lambda_{\alpha\beta}$: For this configuration, $U^m_{\alpha\beta}(1,2) \sim 0$ and Eq. (3.16b) reduces to

$$A_{\alpha\gamma\beta}(1,2) = \int U^m_{\alpha\gamma\beta}(1,2,3)\, d\bar{r}_3\,. \tag{3.25}$$

Thus the leading contribution in the $\lambda_{\alpha\beta}$-expansion of $A_{\alpha\gamma\beta}(1,2)$ comes from the region for which $r_{13} < d_{\alpha\gamma} + \lambda_{\alpha\gamma}$ and $r_{23} < d_{\beta\gamma} + \lambda_{\beta\gamma}$. Consequently, the contribution of Eq. (3.25) is of the order of $\lambda_{\alpha\gamma}\,\lambda_{\beta\gamma}$. Thus

$$A_{\alpha\gamma\beta}(1,2) = O(\lambda_{\alpha\alpha}\,\lambda_{\beta\gamma})$$

and

$$a^{(1)}_{\alpha\gamma\beta}(1,2) = a^{c(1)}_{\alpha\gamma\beta}(1,2)\,. \tag{3.26}$$

Thus the first-order density correction to the RDF, correct to the first-order quantum correction, is

$$g^1_{\alpha\beta}(1,2) = [1 + U^m_{\alpha\beta}(1,2)] \sum_{\gamma=1}^{2} x_\gamma\, a^{c(1)}_{\alpha\gamma\beta}(1,2)\,, \tag{3.27}$$

which leads to

$$\begin{aligned} g^1_{11}(r) &= [1 + U^m_{11}(r)]\left\{X_1 d^2_{11}\left[\pi - 2\sin^{-1}\left(\frac{r}{2d_{11}}\right)\right.\right. \\ &\quad \left. - \frac{r}{2d_{11}}\left(4 - \left(\frac{r}{d_{11}}\right)^2\right)^{1/2}\right] \\ &\quad + X_2 d^2_{12}\left[\pi - 2\sin^{-1}\left(\frac{r}{2d_{12}}\right)\right. \\ &\quad \left.\left. - \frac{r}{2d_{12}}\left(4 - \left(\frac{r}{d_{12}}\right)^2\right)^{1/2}\right]\right\}, \end{aligned} \tag{3.28}$$

$$\begin{aligned} g^1_{22}(2) &= [1 + U^m_{22}(r)]\left\{X_2 d^2_{22}\left[\pi - 2\sin^{-1}\left(\frac{r}{2d_{22}}\right)\right.\right. \\ &\quad \left. - \frac{r}{2d_{22}}\left(4 - \left(\frac{r}{d_{22}}\right)^2\right)^{1/2}\right] \\ &\quad + X_1 d^2_{12}\left[\pi - 2\sin^{-1}\left(\frac{r}{2d_{12}}\right)\right. \end{aligned}$$

$$- \frac{r}{2d_{12}}\left(4 - \left(\frac{r}{d_{12}}\right)^2\right)^{1/2}\Big]\Big\}, \tag{3.29}$$

$$g_{12}^1(r) = [1 + U_{12}^m(r)]\left\{X_1\left[\frac{\pi}{2}(d_{11}^2 + d_{12}^2) - d_{11}^2(\sin^{-1}(\cos\theta_1) + \frac{1}{2}\sin(2\theta_1)) - d_{12}^2(\sin^{-1}(\cos\theta_2) + \frac{1}{2}\sin(2\theta_2))\right] + X_2\left[\frac{\pi}{2}(d_{12}^2 + d_{22}^2) - d_{12}^2\left(\sin^{-1}(\cos\theta') + \frac{1}{2}\sin(2\theta')\right) - d_{22}^2\left(\sin^{-1}(\cos\theta'') + \frac{1}{2}\sin(2\theta'')\right)\right]\right\}, \tag{3.30}$$

where

$$\theta_1 = \cos^{-1}\left(\frac{r + d_{11}^2 - d_{12}^2}{2rd_{11}}\right), \tag{3.31a}$$

$$\theta_2 = \cos^{-1}\left(\frac{r^2 + d_{12}^2 - d_{11}^2}{2rd_{12}}\right), \tag{3.31b}$$

and

$$\theta' = \cos^{-1}\left(\frac{r^2 + d_{12}^2 - d_{22}^2}{2rd_{12}}\right), \tag{3.31c}$$

$$\theta'' = \cos^{-1}\left(\frac{r^2 + d_{22}^2 + d_{12}^2}{2rd_{22}}\right). \tag{3.31d}$$

We may evaluate the RDF of a binary mixture of hard disks using Eqs. (3.28)–(3.30).

## IV. EQUATION OF STATE OF A DILUTE HARD-DISK MIXTURE

Substituting Eqs. (2.13) and (3.9) in the relation[18]

$$P = \frac{\rho^2}{N}\left(\frac{\partial A}{\partial \rho}\right),$$

we obtain an expression for the equation of state in the virial form

$$\beta P = \rho + \sum_{n=2}^{\infty} B_n \rho^n, \tag{4.1}$$

where $B_n$ is the $n$th virial coefficient for the fluid mixture in the semiclassical limit. The first few virial coefficients, which are required to give the equation of state of a dilute gas mixture, can be written as

$$B_2 = B_2^c - \frac{1}{2}\sum_{\alpha,\beta=1}^{2} x_\alpha x_\beta \int \exp[-\beta u_{\alpha\beta}(1,2)] \times U_{\alpha\beta}^m(1,2)\, d\bar{r}_2, \tag{4.2}$$

$$B_3 = B_3^c - \sum_{\alpha,\beta,\gamma=1}^{2} x_\alpha x_\gamma x_\beta \int \exp[-\beta u_{\alpha\beta}(1,2)] \times a_{\alpha\gamma\beta}^{c(1)}(1,2) U_{\alpha\beta}^m(1,2)\, d\bar{r}_2 + \sum_{\alpha,\beta,\gamma=1}^{2} x_\alpha x_\beta x_\gamma \int \exp[-\beta\{u_{\alpha\gamma}(1,3) + u_{\beta\gamma}(2,3)\}] \times U_{\alpha\gamma}^m(1,3) U_{\beta\gamma}^m(2,3)\, d\bar{r}_2\, d\bar{r}_3 - \frac{1}{3}\sum_{\alpha,\beta,\gamma=1}^{2} x_\alpha x_\beta x_\gamma \int \exp\Big[-\beta \sum_{1\le i<j\le 3} u_{\alpha\beta}(i,j)\Big] \times U_{\alpha\gamma\beta}^m(1,2,3)\, d\bar{r}_2\, d\bar{r}_3, \tag{4.3}$$

where $B_2^c$ and $B_3^c$ are, respectively, the second and third virial coefficients for a two-dimensional binary mixture of classical fluid. They are given by[21]

$$B_2^c = -\frac{1}{2}\sum_{\alpha,\beta=1}^{2} x_\alpha x_\beta \int f_{\alpha\beta}(1,2)\, d\bar{r}_2, \tag{4.4}$$

$$B_3^c = -\frac{1}{3}\sum_{\alpha,\beta,\gamma=1}^{2} x_\alpha x_\beta x_\gamma \int f_{\alpha\beta}(1,2) f_{\alpha\gamma}(1,3) \times f_{\beta\gamma}(2,3)\, d\bar{r}_2\, d\bar{r}_3. \tag{4.5}$$

Substituting Eq. (3.7) in Eq. (4.3), we obtain the following expression for the second virial coefficient for a hard-disk mixture in the semiclassical limit:

$$B_2 = \frac{1}{2}\pi \sum_{\alpha,\beta=1}^{2} x_\alpha x_\beta d_{\alpha\beta}^2 \left[1 + \frac{1}{\sqrt{2}}\left(\frac{\lambda_{\alpha\beta}}{d_{\alpha\beta}}\right) + \frac{1}{3\pi}\left(\frac{\lambda_{\alpha\beta}}{d_{\alpha\beta}}\right)^2 + \frac{1}{32\sqrt{2}\pi}\left(\frac{\lambda_{\alpha\beta}}{d_{\alpha\beta}}\right)^3 + \cdots\right], \tag{4.6}$$

where the first term on the right-hand side is the classical value.

For the hard-disk mixture, Eq. (4.3) reduces to

$$B_3 = B_3^c + \sum_{\alpha,\beta,\gamma=1}^{2} x_\alpha x_\beta x_\gamma [B_3^I]_{\alpha\beta\gamma} + O(\lambda_{\alpha\beta}\lambda_{\alpha\gamma}), \tag{4.7}$$

where

$$[B_3^I]_{\alpha\beta\gamma} = -\frac{1}{3}\int [a_{\alpha\gamma\beta}^{c(1)}(1,2) U_{\alpha\beta}^m(1,2) + a_{\alpha\beta\gamma}^{c(1)}(1,2) U_{\alpha\gamma}^m(1,2) + a_{\beta\alpha\gamma}^{c(1)}(1,2) U_{\beta\gamma}^m(1,2)]\, d\bar{r}_2. \tag{4.8}$$

In order to calculate the classical third virial coefficient $B_3^c$, we rewrite Eq. (4.5) in the form[22]

$$B_3^c = \frac{\pi}{6}\sum_{\alpha,\beta,\gamma=1}^{2} x_\alpha x_\beta x_\gamma \times \Big\{\int r_{12}^2 \frac{\partial f_{\alpha\beta}(1,2)}{\partial r_{12}} a_{\alpha\gamma\beta}^{c(1)}(1,2)\, dr_{12} + \int r_{12}^2 \frac{\partial f_{\alpha\gamma}(1,2)}{\partial r_{12}} a_{\alpha\beta\gamma}^{c(1)}(1,2)\, dr_{12} + \int r_{12}^2 \frac{\partial f_{\beta\gamma}(1,2)}{\partial r_{12}} a_{\beta\alpha\gamma}^{c(1)}(1,2)\, d\bar{r}_2\Big\}. \tag{4.9}$$

For hard disks, where

$$\frac{\partial f_{\alpha\beta}(1,2)}{\partial r_{12}} = \delta(r_{12} - d_{\alpha\beta}),$$

Eq. (4.9) can be evaluated as

$$B_3^c = \frac{\pi}{6} \sum_{\alpha,\beta,\gamma=1}^{2} x_\alpha x_\beta x_\gamma \left[ d_{\alpha\beta}^2 a_{\alpha\gamma\beta}^{c(1)}(d_{\alpha\beta}) + d_{\alpha\gamma}^2 a_{\alpha\beta\gamma}^{c(1)}(d_{\alpha\gamma}) + d_{\beta\gamma}^2 a_{\beta\alpha\gamma}^{c(1)}(d_{\beta\gamma}) \right] . \tag{4.10}$$

With the help of Eq. (3.17), Eq. (4.10) can be expressed as

$$B_3^c = \frac{\pi}{3} \sum_{\alpha,\beta,\gamma=1}^{2} x_\alpha x_\beta x_\gamma \left[ d_{\alpha\beta}^2 d_{\alpha\gamma}^2 \left\{ \theta_\alpha^0 - \frac{1}{2}\sin(2\theta_\alpha^0) \right\} + d_{\alpha\beta}^2 d_{\beta\gamma}^2 \left\{ \theta_\beta^0 - \frac{1}{2}\sin(2\theta_\beta^0) \right\} + d_{\alpha\gamma}^2 d_{\beta\gamma}^2 \left\{ \theta_\gamma^0 - \frac{1}{2}\sin(2\theta_\gamma^0) \right\} \right] , \tag{4.11}$$

where

$$\theta_\alpha^0 = \cos^{-1}\left((d_{\alpha\beta}^2 + d_{\alpha\gamma}^2 - d_{\beta\gamma}^2)/2d_{\alpha\beta}d_{\alpha\gamma}\right) , \tag{4.12a}$$

$$\theta_\beta^0 = \cos^{-1}\left((d_{\alpha\beta}^2 + d_{\beta\gamma}^2 - d_{\alpha\gamma}^2)/2d_{\alpha\beta}d_{\beta\gamma}\right) , \tag{4.12b}$$

$$\theta_\gamma^0 = \cos^{-1}\left((d_{\alpha\gamma}^2 + d_{\beta\gamma}^2 - d_{\alpha\beta}^2)/2d_{\alpha\gamma}d_{\beta\gamma}\right) . \tag{4.12c}$$

For a one-component fluid, where $d_{\alpha\beta} = d_{\alpha\gamma} = d_{\beta\gamma} \equiv d$, Eq. (4.11) reduces to[23]

$$B_3^c = \tfrac{1}{4}\pi^2 d^4 (\tfrac{4}{3} - \sqrt{3}/\pi) . \tag{4.13}$$

We now evaluate the leading quantum correction of the order of $(\lambda_{\alpha\beta}/d_{\alpha\beta})$ to the third virial coefficient. Equation (4.8) can be evaluated substituting the values of $U_{\alpha\beta}^m(1,2)$ and $a_{\alpha\gamma\beta}^{c(1)}(1,2)$ from Eqs. (3.7) and (3.17). Thus the final results for the third virial coefficient for the hard-disk mixture, correct to the first-order quantum correction, is

$$\begin{aligned} B_3 = B_3^c + \frac{\pi}{3} \sum_{\alpha,\beta,\gamma=1}^{2} x_\alpha x_\beta x_\gamma \Bigg( & \frac{1}{\sqrt{2}} \left\{ \left(\frac{\lambda_{\alpha\beta}}{d_{\alpha\beta}}\right) \times \left[ d_{\alpha\beta}^2 d_{\alpha\gamma}^2 \left( \theta_\alpha^0 - \frac{1}{2}\sin(2\theta_\alpha^0) \right) + d_{\alpha\beta}^2 d_{\beta\gamma}^2 \left( \theta_\beta^0 - \frac{1}{2}\sin(2\theta_\beta^0) \right) \right] \right. \\ & + \left(\frac{\lambda_{\alpha\gamma}}{d_{\alpha\gamma}}\right) \left[ d_{\alpha\beta}^2 d_{\alpha\gamma}^2 \left( \theta_\alpha^0 - \frac{1}{2}\sin(2\theta_\alpha^0) \right) + d_{\alpha\gamma}^2 d_{\beta\gamma}^2 \left( \theta_\gamma^0 - \frac{1}{2}\sin(2\theta_\gamma^0) \right) \right] \\ & \left. + \left(\frac{\lambda_{\beta\gamma}}{d_{\beta\gamma}}\right) \left[ d_{\alpha\beta}^2 d_{\beta\gamma}^2 \left( \theta_\beta^0 - \frac{1}{2}\sin(2\theta_\beta^0) \right) + d_{\alpha\gamma}^2 d_{\beta\gamma}^2 \left( \theta_\gamma^0 - \frac{1}{2}\sin(2\theta_\gamma^0) \right) \right] \right\} \\ & + O\left( \left( \frac{\lambda_{\alpha\beta}}{d_{\alpha\beta}} \right)^2 \right) \Bigg) . \end{aligned} \tag{4.14}$$

For a one-component fluid, Eq. (4.14) reduces to[4]

$$B_3 = B_3^c \left[ 1 + \sqrt{2}(\lambda/d) + O((\lambda/d)^2) \right] , \tag{4.15}$$

where $B_3^c$ is given by Eq. (4.13).

We are interested in estimating the excess properties of the hard-disk mixture (relative to the pure components). The "excess" second and third virial coefficients of a hard-disk mixture in the semiclassical limit may be obtained from Eqs. (4.6) and (4.14), respectively. Thus,

$$\Delta B_2 = \pi d^2 \left(\frac{d_{12}}{d}\right)^2 x_1 x_2 \left[ 1 + \frac{E}{\sqrt{2}}\left(\frac{\lambda}{d}\right) + \frac{E^2}{3\pi}\left(\frac{\lambda}{d}\right)^2 + \frac{E^3}{32\sqrt{2}\pi}\left(\frac{\lambda}{d}\right)^3 + \cdots \right] \tag{4.16}$$

and

$$\Delta B_3 = \Delta B_3^c + \Delta B_B^I(\lambda/d) + O((\lambda/d)^2) , \tag{4.17}$$

where

$$\Delta B_3^c = \frac{1}{4}\pi^2 d^4 \left\{ x_1 x_2 \left[ \left(\frac{d_{12}}{d}\right)^4 \left(2 - \frac{4}{\pi}e\right) + \left(\frac{d_{12}}{d}\right)^2 \left(4 - \frac{8a}{\pi} - 2b\right) - \frac{1}{\pi}c \right] \right\} , \tag{4.18}$$

$$\Delta B^I = \sqrt{2}\pi d^4 x_1 x_2 \left[ \left(\frac{d_{12}}{d}\right)^4 E\left(\frac{\pi}{2} - e\right) + \left(\frac{d_{12}}{d}\right)^2 \times \left(\frac{\pi}{2} - A + \frac{1}{2}E(\pi - 2a - b)\right) - \frac{1}{4}D \right] , \tag{4.19}$$

where

$$d^2 = x_1 d_{11}^2 + x_2 d_{22}^2 , \tag{4.20a}$$

$$ad^2 = x_1 d_{11}^2 \sin^{-1}\left(\frac{d_{11}}{2d_{12}}\right) + x_2 d_{22}^2 \sin^{-1}\left(\frac{d_{22}}{2d_{12}}\right) , \tag{4.20b}$$

$$bd^2 = x_1 d_{11}\sqrt{4d_{12}^2 - d_{11}^2} + x_2 d_{22}\sqrt{4d_{12}^2 - d_{22}^2} , \tag{4.20c}$$

$$cd^4 = x_1 d_{11}^3\sqrt{4d_{12}^2 - d_{11}^2} + x_2 d_{22}^3\sqrt{4d_{12}^2 - d_{22}^2} , \tag{4.20d}$$

$$e = x_1 \sin^{-1}\left(1 - \frac{d_{11}^2}{2d_{12}^2}\right) + x_2 \sin^{-1}\left(1 - \frac{d_{22}^2}{2d_{12}^2}\right) , \tag{4.20e}$$

and

$$\lambda d = x_1 \lambda_{11} d_{11} + x_2 \lambda_{22} d_{22} , \tag{4.21a}$$

$$A\lambda d = x_1 \lambda_{11} d_{11} \sin^{-1}\left(\frac{d_{11}}{2d_{12}}\right) + x_2 \lambda_{22} d_{22} \sin^{-1}\left(\frac{d_{22}}{2d_{12}}\right) , \tag{4.21b}$$

$$D\lambda d^3 = x_1 \lambda_{11} d_{11}^2\sqrt{4d_{12}^2 - d_{11}^2} + x_2 \lambda_{22} d_{22}^2\sqrt{4d_{12}^2 - d_{22}^2} , \tag{4.21c}$$

$$E(\lambda/d) = (\lambda_{12}/d_{12}) . \tag{4.21d}$$

In Eq. (4.16), the first term is the classical value. From Eqs. (4.20a) and (4.21a), we have

$$\frac{\lambda}{d} = \left[\frac{x_1 + x_2 R(\lambda_{22}/\lambda_{11})}{x_1 + x_2 R^2}\right]\left(\frac{\lambda_{11}}{d_{11}}\right) = \left[\frac{x_1 + x_2 R(m_{11}/m_{22})^{1/2}}{x_1 + x_2 R^2}\right]\left(\frac{\lambda_{11}}{d_{11}}\right) , \tag{4.22}$$

where $R = d_{22}/d_{11}$. For additive hard disks, $E$ is given by

$$E = \sqrt{2}\,\frac{(1+m_{11}/m_{22})^{1/2}}{(1+R)}\,\frac{(x_1+x_2R^2)}{(x_1+x_2R\,(m_{11}/m_{22})^{1/2})}. \quad (4.23)$$

The equation for $E$ is invariant to a labeling of species 1 and 2. For a given mixture, the excess quantum correction can be calculated in term of $(\lambda_{\alpha\alpha}/d_{\alpha\alpha})$ of one of the species. However, in the case of a hard-disk system, we may consider the atomic mass as some function of the diameter of hard disks. If we assume for simplicity that the atomic mass $m_{\alpha\alpha}$ is proportional to $d_{\alpha\alpha}^2$, Eq. (4.23) reduces to

$$E = \sqrt{2}\,[(1+R^{-2})^{1/2}(x_1+x_2R^2)/(1+R)]. \quad (4.24)$$

Equations (4.16) and (4.17) are valid for both additive and nonadditive hard-disk mixtures. From these expressions, it is obvious that the excess virial coefficients depend on $x_1$ and $R$.

## V. THERMODYNAMIC PROPERTIES OF A DENSE HARD-DISK MIXTURE

The free energy of a binary mixture of hard disks, correct to the first-order quantum correction, is obtained from Eq. (2.13):

$$\frac{\beta A}{N} = \frac{\beta A^c}{N} - \frac{1}{2}\rho \sum_{i,j=1}^{2} x_i x_j \times \int U_{ij}^m(r)\, g_{ij}^c(r) d\bar{r} + O(\lambda_{ij}^2). \quad (5.1)$$

Using Eq. (3.7), it can be evaluated as

$$\frac{\beta A}{N} = \frac{\beta A^c}{N} + \frac{\pi}{2\sqrt{2}}\rho \sum_{i,j} x_i x_j\, g_{ij}^c(d_{ij}) d_{ij}\,\lambda_{ij}. \quad (5.2)$$

This expression is valid for both additive and nonadditive hard-disk mixtures.

The van der Waals one (vdW1) fluid theory of mixture, originally developed for the hard-sphere system,[24] has been extended in the case of the classical hard-disk mixture.[15] We adopt this theory to calculate the properties of the classical system. This theory approximates the properties of a mixture by those of a fictitious pure hard-disk fluid with the diameter[14]

$$d_0^2 = \sum_{i,j} x_i\, x_j\, d_{ij}^2. \quad (5.3)$$

In the vdW1 theory of mixture, the free energy and pressure of the classical mixture are written as

$$A^c = A_0^c + NkT \sum_i x_i \ln x_i + \text{second order term}, \quad (5.4)$$

$$P^c = P_0^c + \text{second order term}, \quad (5.5)$$

where $A_0^c$ and $P_0^c$ are, respectively, the free energy and pressure for pure classical fluid containing $N (= N_1 + N_2)$ molecules in volume $V$ at temperature $T$. For a hard-disk model having diameter $d_0$ given by Eq. (2.6), $A_0^c$ and $P_0^c$ are given by[25]

$$\beta A_0^c/N = \tfrac{9}{8}\,[\eta_0/(1-\eta_0)] - \tfrac{7}{8}\,ln(1-\eta_0) \quad (5.6)$$

and

$$\beta P_0^c/\rho = (1 + 0.125\,\eta_0^2)/(1-\eta_0)^2, \quad (5.7)$$

where

$$\eta_0 = \tfrac{1}{4}\pi\rho d_0^2. \quad (5.8)$$

In the vdW1 theory, it is assumed that

$$g_{ij}^c(d_{ij}) = g^c(d_0). \quad (5.9)$$

Further, we assume

$$d_0\lambda_0 = \sum_{i,j} x_i\, x_i\, \lambda_{ij}\, d_{ij}, \quad (5.10)$$

for all $i$ and $j$. Then Eq. (5.2) is written as

$$\frac{\beta A}{N} = \frac{\beta A^c}{N} + \frac{\pi}{2\sqrt{2}}(\rho d_0^2)\, g^c(d_0)\left(\frac{\lambda_0}{d_0}\right). \quad (5.11)$$

Other thermodynamic properties can be derived from Eq. (5.11). Thus the equation of state of the hard-disk mixture in the semiclassical limit is given by

$$\frac{\beta P}{\rho} = \frac{\beta P^c}{\rho} + \frac{\pi}{2\sqrt{2}}(\rho d_0^2) \times \left[ g^c(d_0) + \rho\,\frac{\partial g^c(d_0)}{\partial \rho}\right]\left(\frac{\lambda_0}{d_0}\right). \quad (5.12)$$

Thus the quantum corrections to the thermodynamic properties are expressed in terms of the classical RDF at the contact $g^c(d_0)$, which can be obtained from the relation[14]

$$\frac{\beta P^c}{\rho} = 1 + \frac{1}{2}\pi\rho \sum_{i,j} x_i\, x_j\, d_{ij}^2\, g_{ij}^c(d_{ij}) = 1 + 2\,\eta_0\, g^c(d_0). \quad (5.13)$$

With the help of Eqs. (5.7) and (5.13), $g^c(d_0)$ is given by

$$g^c(d_0) = (1 - \tfrac{7}{16}\,\eta_0)/(1-\eta_0)^2. \quad (5.14)$$

Using Eqs. (5.6)–(5.9) and (5.14), we obtain the following expressions for the free energy and pressure correct to the first order in $\lambda_{ij}$:

$$\frac{\beta A}{N} = \left[1 + \frac{1}{2\sqrt{2}} \sum_{i,j} x_i\, x_j\, \lambda_{ij}\, \frac{\partial}{\partial d_{ij}}\right]\left(\frac{\beta A^c}{N}\right), \quad (5.15)$$

$$\frac{\beta P}{\rho} = \left[1 + \frac{1}{2\sqrt{2}} \sum_{i,j} x_i\, x_j\, \lambda_{ij}\, \frac{\partial}{\partial d_{ij}}\right]\left(\frac{\beta P^c}{\rho}\right). \quad (5.16)$$

Thus the first-order quantum correction to the thermodynamic properties of the hard-disk mixture can be found by replacing the actual diameter $d_{ij}$ by an effective diameter $(d_{ij} + 2^{-3/2}\lambda_{ij})$. Thus the effective diameter method works for the fluid mixture in the vdW1 theory, where $g_{ij}^c(d_{ij}) = g^c(d_0)$.

We use this theory to calculate the thermodynamic properties for both additive and nonadditive hard-disk mixtures.

### A. Binary mixture of additive hard disks

We derive expressions for the thermodynamic properties of the binary mixture of additive hard disks.

Using Eqs. (4.20a) and (4.21a) we get the following relations for the binary mixture of additive hard disks:

$$\eta_0 = \eta\,[1 - \tfrac{1}{2}\,x_1x_2\,(1-R^2)/(x_1+x_2R^2)], \quad (5.17)$$

where

$$\eta = (\pi/4)\rho d^2 = (\pi/4)\rho(x_1 d_{11}^2 + x_2 d_{22}^2) \tag{5.18}$$

and

$$\lambda_0 d_0 = F\lambda d\,, \tag{5.19}$$

where

$$F = 1 + x_1 x_2 \times\left[\frac{2^{-1/2}(1+R)(1+\lambda_{22}/\lambda_{11})^{1/2} - (1+R\,\lambda_{22}/\lambda_{11})}{(x_1 + x_2 R\,\lambda_{22}/\lambda_{11})}\right]. \tag{5.20}$$

Using Eq. (5.14) in Eq. (5.2), we obtain expressions for the free energy and pressure for the additive hard-disk mixture:

$$\frac{\beta A}{N} = \frac{\beta A^c}{N} + A_I^*\left(\frac{\lambda}{d}\right) + O\left(\left(\frac{\lambda}{d}\right)^2\right), \tag{5.21}$$

where

$$A_I^* = \sqrt{2}\,F\eta\,[(1 - \tfrac{7}{16}\eta_0)/(1-\eta_0)^2]\,, \tag{5.22}$$

and

$$\frac{\beta P}{\rho} = \frac{\beta P^c}{\rho} + P_I^*\left(\frac{\lambda}{d}\right) + O\left(\left(\frac{\lambda}{d}\right)^2\right), \tag{5.23}$$

where

$$P_I^* = \sqrt{2}\,F\eta\,[(1 + \tfrac{1}{8}\eta_0)/(1-\eta_0)^2]\,. \tag{5.24}$$

In a theory of mixture the excess properties of the system are of interest. From Eq. (5.2), the excess free energy for the hard-disk mixture is given by

$$\frac{\beta A_E}{N} = \frac{\beta A_E^c}{N} + \frac{\pi}{\sqrt{2}}\,x_1 x_2\,(\rho d_{12}^2)\,g_{12}^c(d_{12})\left(\frac{\lambda_{12}}{d_{12}}\right). \tag{5.25}$$

With the help of Eqs. (5.7) and (5.13), $g_{12}^c(d_{12})$ for the additive hard-disk mixture can be given by

$$g_{12}^c(d_{12}) = \left[\frac{(1+0.125\,\eta)}{(1-\eta)^3} - \frac{25}{8}\,\frac{\eta(1+\frac{2}{25}\eta)}{(1-\eta)^3}\,\mu\right] + O(x_1 x_2)\,, \tag{5.26}$$

$$\mu = \frac{x_1 d_{11}^4 + x_2 d_{22}^4}{(x_1 d_{11}^2 + x_2 d_{22}^2)(d_{11}+d_{22})^2} = \frac{x_1 + x_2 R^4}{(x_1 + x_2 R^2)(1+R)^2}\,. \tag{5.27}$$

Thus the "excess" free energy is given by

$$\frac{\beta A_E}{N} = \frac{\beta A_E^c}{N} + A_E^{I^*}\left(\frac{\lambda}{d}\right), \tag{5.28}$$

where

$$A_E^{I^*} = 2\sqrt{2}x_1 x_2 \eta_a\left[\frac{1+0.125\,\eta}{(1-\eta)^3} - \frac{25}{8}\,\frac{\eta(1+\frac{2}{25}\eta)}{(1-\eta)^4}\,\mu\right] E \tag{5.29}$$

with

$$\eta_a = \frac{\pi}{4}\rho\, d_{12}^2 = \frac{1}{4}\,\eta\,\frac{(1+R)^2}{(x_1 + x_2 R^2)}\,. \tag{5.30}$$

From Eq. (5.29), we find that the value of $A_E^{I^*}$ depends on the concentration $x_1$ and diameter ratio $R$.

### B. Binary mixture of nonadditive hard disks

This section is concerned with the evaluation of the excess properties of the binary mixture of nonadditive hard disks, for which Eq. (3.2) can be written as

$$d_{12} = d_{12}^a(1+\Delta)\,,$$

where

$$d_{12}^a = (d_{11} + d_{12})/2$$

is the effective diameter for additive hard disks. Equation (5.2) can be rewritten in the form

$$\frac{\beta(A - A_a)}{N} = \frac{\beta(A^c - A_a^c)}{N} + \frac{\pi}{\sqrt{2}}\,x_1 x_2 \rho\,[d_{12}\,g_{12}^c(d_{12}) - d_{12}^a\,g_{12}^a(d_{12})]\,\lambda_{12}\,, \tag{5.31}$$

where the subscript "$a$" refers to the properties of the additive hard-disk mixture. Other thermodynamic properties can be obtained from Eq. (5.31). Thus the equation of state for a binary mixture of nonadditive hard disks in the semiclassical limit is given by

$$\frac{\beta(P - P_a)}{\rho} = \frac{\beta(P^c - P_a^c)}{\rho} + \frac{\pi}{\sqrt{2}}\,x_1 x_2 \times\rho\left\{[d_{12}\,g_{12}^c(d_{12}) - d_{12}^a\,g_{12}^c(d_{12}^a)] + \rho\,\frac{\partial}{\partial\rho}\,[d_{12}\,g_{12}^c(d_{12}) - d_{12}^a\,g_{12}^c(d_{12})]\right\}\lambda_{12}\,. \tag{5.32}$$

For a classical binary mixture of nonadditive hard disks, the free energy and equation of state are given by[14]

$$\frac{\beta(A^c - A_a^c)}{N} = 4x_1 x_2\,\eta_a\left[\frac{1 - \frac{7}{16}\eta}{(1-\eta)^3}\right] \times\Delta\,(2+\Delta) + O(x_1^2\,x_2^2)\,, \tag{5.33}$$

$$\frac{\beta(P^c - P_a^c)}{\rho} = 4x_1 x_2\,\eta_a\left[\frac{1+0.125\,\eta}{(1-\eta)^3}\right] \times\Delta\,(2+\Delta) + O(x_1^2\,x_2^2)\,. \tag{5.34}$$

The quantum correction terms of Eqs. (5.31) and (5.32) are expressed in terms of $g_{12}^c(d_{12})$ and $g_{12}^c(d_{12}^a)$. For additive hard-disk mixture, $g_{12}^c(d_{12}^a)$ is given by Eq. (3.2), $g_{12}^c(d_{12})$ is evaluated using Eq. (5.13), and

$$\frac{\beta(P^c - P_a^c)}{\rho} = \pi x_1 x_2 \rho\,[d_{12}^2\,g_{12}^c(d_{12}) - d_{12}^{a^2}\,g_{12}^c(d_{12}^a)]\,. \tag{5.35}$$

Expanding the right-hand side of Eq. (5.35) in the power of $\Delta$, we find

$$\frac{\beta(P^c - P_a^c)}{\rho} = 4x_1 x_2\,\eta_a\,[g_{12}^c(d_{12}) - g_{12}^c(d_{12}^a) + 2\Delta\,g_{12}^c(d_{12}^a) + O(\Delta^2)]\,. \tag{5.36}$$

Comparing Eq. (5.34) and Eq. (5.36) and using Eq. (3.2), we get

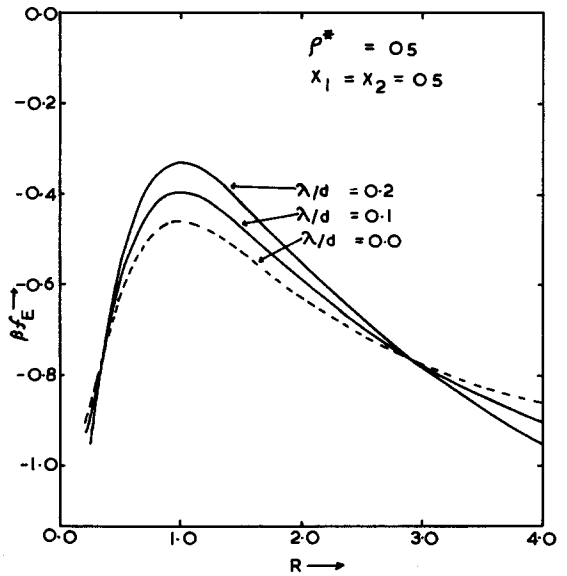



FIG. 1. The excess free energy per particle $\beta f_E$ of a binary mixture of hard disks as a function $R$ for $\rho^* = 0.5$, $x_1 = x_2 = 0.5$, and $\Delta = 0.0$.

$$g_{12}^c(d_{12}) - g_{12}^c(d_{12}^a) = \frac{25}{4}\frac{\eta(1+\frac{2}{25}\eta)}{(1-\eta)}\mu\Delta + O(\Delta^2). \quad (5.37)$$

Substituting Eq. (5.36) in Eqs. (5.31) and (5.32) we can obtain the final results for the free energy and equation of state, correct to the first-order quantum correction. Thus,

$$\frac{\beta(A-A_a)}{N} = \frac{\beta(A^c - A_a^c)}{N} + A_E^\lambda\left(\frac{\lambda}{d}\right) \quad (5.38)$$

and

$$\frac{\beta(P-P_a)}{\rho} = \frac{\beta(P^c - P_a^c)}{\rho} + P_E^\lambda\left(\frac{\lambda}{d}\right). \quad (5.39)$$

Here the quantum coefficients $A_E^\lambda$ and $P_E^\lambda$ are given by

$$A_E^\lambda = 2\sqrt{2}x_1x_2E\eta_a\left[\frac{1+0.125\,\eta}{(1-\eta)^3} + \frac{25}{8}\frac{\eta(1+\frac{2}{25}\eta)}{(1-\eta)^4}\mu\right]\Delta + O(\Delta^2) \quad (5.40)$$

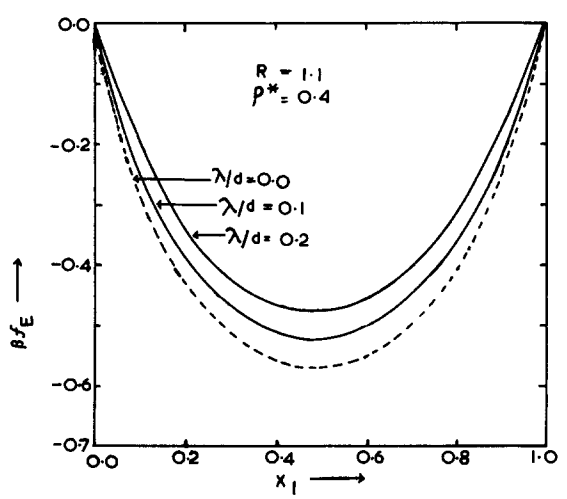



FIG. 2. The excess free energy per particle $\beta f_E$ of a binary mixture of hard disks as a function of $x_1$ for $\rho^* = 0.4$, $R = 1.1$, and $\Delta = 0.0$.

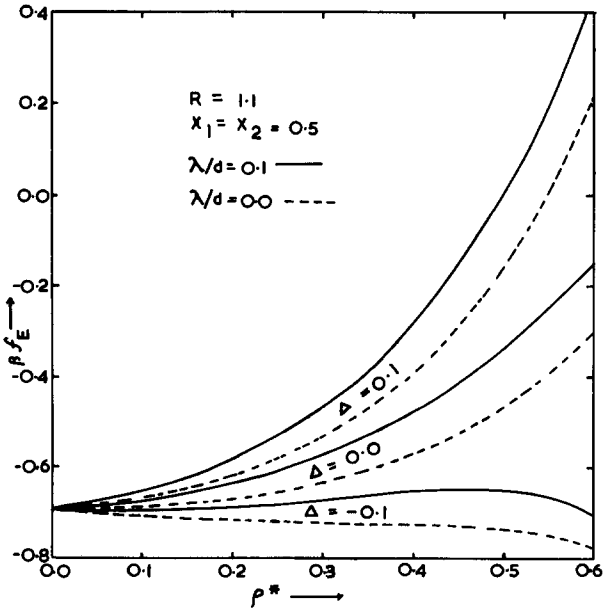



FIG. 3. The excess free energy per particle $\beta f_E$ of a binary mixture of hard disks as a function of $\rho^*$ for $R = 1.1$, $\Delta = 0.0, \pm 0.1$, and $x_1 = x_2 = 0.5$.

and

$$P_E^\lambda = 2\sqrt{2}x_1x_2E\eta_a\left[\frac{(1-\frac{15}{4}\eta-\frac{5}{8}\eta^2)}{(1-\eta)^4} + \frac{25}{4}\eta\left\{\frac{1-\frac{72}{25}\eta-\frac{7}{25}\eta^2}{(1-\eta)^5}\right\}\mu\right]\Delta + O(\Delta^2). \quad (5.41)$$

### C. Results and Discussion

We have used Eq. (5.28) to calculate the excess free energy per particle $\beta f_E (\equiv \beta A_E/N)$ of the additive hard-disk mixture correct to the first order of quantum correction. Values of $\beta f_E$ at $\rho^* = 0.5$ for $x_1 = x_2 = 0.5$ are plotted as a function of diameter ratio $R$ in Fig. 1 for $\lambda/d = 0.0, 0.1$, and 0.2. The quantity $\rho^* = \rho(x_1 d_{11}^2 + x_2 d_{22}^2)$. The excess free energy, both in classical and semiclassical limits, is maximum at $R = 1.0$ and decreases steadily as $R$ moves away from 1.0. The value of $A_E^{I^*}$, which is governed by $E$ and $\mu$, is maximum at $R = 1.0$ and becomes $-\infty$ at $R = 0$ and $\infty$. So the quantum value of $\beta f_E$ decreases faster and becomes negative when $R$ is far away from 1.0.

Figure 2 demonstrate the variation of $\beta f_E$ of the additive hard-disk mixture for $\rho^* = 0.4$ with the concentration $x_1$ for $\lambda/d = 0.0$, 0.1, and 0.2. It is found that the (excess) quantum effect is zero at $x_1 = 0$ and $x_1 = 1.0$ and finite in the intermediate range of $x_1$.

In Fig. 3, the values of $\beta f_E$ for a binary hard-disk mixture with $R = 1.0$ and $x_1 = x_2 = 0.5$ are reported as a function of $\rho^*$ for $\Delta = 0.0$, and $\pm 0.1$ at $\lambda/d = 0.0$ and 0.2. It is found that the quantum effect increases with the increase of $\rho^*$ and decrease of $\Delta$, the nonadditive parameter.

Thus we come to the conclusion that the (excess) quantum effect to the thermodynamic properties of the hard-disk mixture, which depends on the concentration $x_1$ and the diameter ratio $R$, increases with $\rho^*$ and $\Delta$.

[1]J. G. Dash and M. Schick, in *The Physics of Liquid and Solid Helium*, edited by K. H. Bennemann and J. B. Ketterson (Wiley, New York, 1978), Part 2.
[2]R. L. Siddon and M. Schick, Phys. Rev. A **9**, 907 (1974).
[3]W. G. Gibson, preprint, 1983.
[4]S. K. Sinha and Y. Singh, Mol. Phys. **44**, 877 (1981); S. K. Sinha, Y. S. Sainger, and Y. Singh, J. Math. Phys. **23**, 2569 (1982).
[5]J. Ram, Y. S. Sainger, and Y. Singh, Mol. Phys. **45**, 1141 (1982); Y. S. Sainger, S. K. Sinha, and Y. Singh, J. Chem. Phys. **79**, 5088 (1983).
[6]J. G. Dash, *Film on Solid Surfaces* (Academic, New York, 1975).
[7]W. A. Steele, Surf. Sci. **39**, 105 (1973); F. Lado, J. Chem. Phys. **49**, 3092 (1968); D. Henderson, Mol. Phys. **34**, 301 (1977).
[8]S. Toxvaerd, Phys. Rev. Lett. **44**, 1002 (1980); F. F. Abraham, Phys. Rev. Lett. **44**, 483 (1980); J. A. Barker, D. Henderson, and F. F. Abraham, Physica A **106**, 226 (1981).
[9]J. M. Kosterlitz and D. J. Thouless, J. Phys. C **6**, 1181 (1973); V. L. Berezinakii, Sov. Phys. JETP **34**, 610 (1972).
[10]M. E. Boyd, S. Y. Larsen, and J. E. Kilpatrick, J. Chem. Phys. **50**, 4034 (1969).
[11]E. Wigner, Phys. Rev. **40**, 749 (1932); J. G. Kirkwood, Phys. Rev. **44**, 31 (1933).
[12]P. C. Hemmer, Phys. Lett. A **27**, 377 (1968); B. Jancovici, Phys. Rev. **178**, 295 (1969).
[13]E. Dichinson, Mol. Phys. **33**, 1463 (1977).
[14]U. N. Singh and S. K. Sinha, Pramana **20**, 327 (1983).
[15]U. N. Singh and S. K. Sinha, J. Chem. Phys. **77**, 5784 (1982).
[16]U. N. Singh and S. K. Sinha, J. Chem. Phys. **78**, 3191 (1983).
[17]D. J. Adam and I. R. McDonald, J. Chem. Phys. **63**, 1900 (1975).
[18]S. A. Rice and P. Gray, *The Statistical Mechanics of Simple Liquids* (Interscience, New York, 1965).
[19]W. A. Steele, *The Interscience Interaction of Gases with Solid Surfaces* (Pergamon, New York, 1974), Chap. 4.
[20]W. G. Gibson, Mol. Phys. **30**, 1, 13 (1975).
[21]J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids* (Wiley, New York, 1954), Chap. 8.
[22]E. D. Glandt, J. Chem. Phys. **68**, 2952 (1978).
[23]F. H. Ree and W. G. Hoover, J. Chem. Phys. **40**, 939 (1964).
[24]T. W. Leland, J. S. Rowlinson, and G. A. Sather, Trans. Faraday Soc. **64**, 1447 (1968).
[25]D. Henderson, Mol. Phys. **30**, 971 (1975).

# Nonequilibrium statistical thermodynamics of second-order stochastic processes in the limit of large resistance

B. H. Lavenda
*Dipartimento di Scienze Chimiche, Universita' di Camerino, via Sant'Agostino, 1, Camerino 62032 (MC) Italy and TEMA S.p.A., viale Aldo Moro, 38, Bologna 40127, Italy*

R. Serra
*TEMA S.p.A., viale Aldo Moro, 38, Bologna 40127, Italy*

(Received 8 February 1983; accepted for publication 17 August 1984)

The phase-space generalization of the kinetic analog of Boltzmann's principle is derived. A kinetic criterion for the contraction of a second-order (Ornstein–Uhlenbeck) to a first-order (Einstein–Smoluchowski) stochastic process is advanced on the basis of an asymptotic expansion of the joint entropy in the limit of large resistance. The kinetic criterion is shown to be the stationary solution to the first velocity moment of the phase-space Fokker–Planck equation. The Stratonovich criterion for the validity of the Einstein–Smoluchowski description is corroborated from the asymptotic expansion of the joint entropy. Second-order Gaussian processes are used for illustration where it is shown that (i) the principles of least dissipation of energy and maximum joint entropy are equivalent, and (ii) the observed, local velocity arises from an exact balance between rates of growth and decay of velocity fluctuations.

## I. INTRODUCTION

The analysis of dynamical systems is often simplified when their components, say $X(t)$ and $Y(t)$, evolve over two nonoverlapping time scales. Suppose that $X(t)$ is the rapid process and $Y(t)$ is the slow one. The reduction principle[1] states essentially that the $Y(t)$ process can be studied alone by substituting $\dot{X}(t)=0$ into the equation of motion for $Y(t)$.

An analogous phenomenon often occurs for Brownian motion where the inertia and external noise evolve over a time scale much shorter than the variation in the external force. A typical case is that of the Ornstein–Uhlenbeck (OU) process[2]:

$$dY(t)=U(t)dt\,, \tag{1.1}$$

$$dU(t)=-\{RU(t)-F[Y(t)]\}\,dt+\sqrt{2kR}\,dW(t)\,,$$

where $Y(t)$ and $U(t)$ are the displacement and velocity of the Brownian particle, respectively. Here, $F$ denotes the external force and $W(t)$ is a standard Brownian motion. The time scale is set by the magnitude of the resistance parameter $R$. The two well-known limiting cases are small and large resistance.[3] The thermodynamically interesting case is the latter where for a constant force, the system will tend to a limiting velocity, $(1/R)F$ over a time scale greater than the relaxation time $(1/R)$.[4] In this case, over times greater than $(1/R)$, the second-order process (1.1) reduces effectively to the first-order process

$$dY(t)=(1/R)F[Y(t)]\,dt+\sqrt{2k/R}\,dW(t)\,. \tag{1.2}$$

Even when the force is a slowly varying function of the displacement, Eq. (1.2) is assumed to be a valid approximation to (1.1) for times $t\gg(1/R)$ and it is commonly referred to as the Einstein–Smoluchowski (ES) approximation.[5]

There is a long history to the phase-space description of Brownian motion and its subsequent reduction to a configuration space description. Klein[6] generalized the configuration space Fokker–Planck (FP) equation, which we shall refer to as the Smoluchowski (S) equation, to the phase-space Fokker–Planck equation. He even gave an approximate reduction of the FP equation to the S equation which provided the stimulus for Kramers'[3] study. Kramers offered a criterion for the contraction which was later modified by Stratonovich.[7] Other asymptotic expansions have been proposed by Brinkman[8] which was later rediscovered by Landauer and Swanson.[9] Brinkman used a moment generating method which had the defect of choosing the initial velocity distribution as the Maxwell–Boltzmann (MB) distribution. To lowest order, he obtained the Laplace transform of the telegraph equation which would correspond to Eq. (3.28) below, if the second moment were to be replaced by its asymptotic, equilibrium value. However, it is known that the second moment, which is the kinetic contribution to momentum transport, relaxes only slightly faster than the current, or first moment.[10] Chapman–Enskog procedures have also been derived in which the entire time dependence resides in the configuration space transition density.[11] Moreover, the Chapman–Enskog procedure has also been used to obtain a perturbation series for the diffusion operator on the S equation.[12] All these methods come under the heading of "adiabatic elimination procedures" which are reviewed in Ref. 13. The validity of these methods has been questioned in Ref. 14 and it is the purpose of the present paper to offer a thermodynamic criterion for the contraction of a second-order stochastic process to a first-order one, based on a generalization of the configuration space kinetic analog to Boltzmann's principle.[15]

The evolution toward equilibrium of nonequilibrium statistical thermodynamic processes is governed by the wearing off of the statistical correlations between nonequilibrium states for increasing times.[16] In the phase-space description, our aim is to obtain a criterion for the statistical independence of velocity and displacement fluctuations in the large resistance limit for times greater than the relaxa-

 0022-2488/85/030505-09$02.50 © 1985 American Institute of Physics 

tion time ($1/R$). We shall assume that the fast or "driving" process is ergodic and that there is a unique invariant probability distribution $p_\infty$ which is determined in terms of the entropy $S(y,u)$ according to Boltzmann's principle

$$k \ln p_\infty(y,u) = S(y,u) - S(0), \tag{1.3}$$

where $k$ is Boltzmann's constant and $S(0)$ is the equilibrium value of the entropy. Over long time intervals (i.e., assuming that the process is homogeneous), the driving process $U(t)$ will have had ample time to decouple itself from the slow or "driven" process $Y(t)$ so that the latter becomes a Markov process by itself. At this stage, a MB-distribution will have been established at every point in configuration space, implying that the entropy has the form

$$S(y,u) = S(0) - \tfrac{1}{2}u^2 + \int_0^y F(x)\,dx\,. \tag{1.4}$$

On account of the fact that the entropy is invariant under time reversal, no cross term can appear in (1.4). The force $F$ in Eq. (1.2) is thus identified as the space derivative of the entropy and tends to restore the system to thermodynamic equilibrium. In regard to Eq. (1.1), it behaves more like an "external" force: in an isolated system, $F$ would vanish leaving a pure diffusion process in velocity space, which is described by the diffusion operator

$$\mathcal{G}^i := -Ru\left(\frac{\partial}{\partial u}\right) + D^*\left(\frac{\partial^2}{\partial u^2}\right) \quad \text{(isolated process)}, \tag{1.5}$$

where $D^*$ is the velocity space diffusion coefficient which is given by the Einstein formula $D^* = kR$. However, if the system is open thermodynamically in the sense that mechanical work can be done on it (even though it is still thermally isolated), there is a nonvanishing external force $F$ which acts on the particle at each position $y$. Moreover, it induces a drift in velocity space and the diffusion operator now has the form

$$\mathcal{G}^o := u\left(\frac{\partial}{\partial y}\right) + b\left(\frac{\partial}{\partial u}\right) + D^*\left(\frac{\partial^2}{\partial u^2}\right) \quad \text{(open process)}, \tag{1.6}$$

where $b$ is the drift, viz.,

$$b(y,u) := F(y) - Ru\,. \tag{1.7}$$

Hence, for second-order processes, $F$ acts more like an external force rather than an internal force as for first-order stochastic processes (1.2). From (1.6) and (1.7) it is clear that $F$ is responsible for the coupling of the driving process to the driven process. Since the phase-space process is Markov, the statistics are determined completely by the transition density and the invariant distribution (1.3). We shall now turn to the derivation of the former.

## II. PHASE-SPACE KINETIC ANALOG OF BOLTZMANN'S PRINCIPLE

The mathematical technique which we use to generalize our configuration space, kinetic analog of Boltzmann's principle to phase space is again based upon Girsanov's theorem,[17] relating the absolutely continuous substitution of different probability measures on the same probability space for diffusion processes with the same variance but with different drifts. The only difference is that we will now be considering the transformation of the velocity space Wiener measure into a new measure for our nonequilibrium statistical thermodynamic process. Its configuration space analog has been used by Ezawa *et al.*[18] to study the transformation properties of the Wiener measure for quantum mechanical processes. The original idea of transforming the configuration space Wiener measure into a new measure, in which the Onsager–Machlup (OM) potential appeared, is due to Falkoff[19] and has been recently rediscovered in Ref. 20.

The derivation of the transition density will occur in two stages. The first stage consists of the transformation of the isolated process

$$dU(t) = -RU(t)dt + \sqrt{2D^*}\,dW(t)\,, \quad U(0) = u^0 \tag{2.1}$$

with

$$Y(T) - y^0 = \int_0^T U(t)\,dt$$

into the open thermodynamic process (1.1) which is caused by the application of the external field $F$ on the system. The probability measure density for the transformation is given by the Girsanov formula

$$\rho_o[U(t),Y(t)] = \exp\left\{\int_0^T\left[\frac{F}{\sqrt{2D^*}}\,dW(t) - \frac{1}{2}\left(\frac{F}{\sqrt{2D^*}}\right)^2 dt\right]\right\}, \tag{2.2}$$

where the subscript "$o$" stands for open. Eliminating the Wiener process with the aid of (2.1) and noting that

$$\int_0^T F(x)u\,dt = \int_0^y F(x)\,dx = \Delta S^*(y)$$

is the change in the configurational entropy $S^*$ over the interval $[0,T]$, the measure density (2.2) can be written as

$$\rho_o[U(t),Y(t)] = \exp\left\{\left(\frac{1}{2k}\right)[S^*[Y(t)] - S^*(y^0)] + \frac{1}{2D^*}\int_0^T [F[Y(t)]\,dU(t) - \frac{1}{2}F^2[Y(t)]\,dt]\right\}. \tag{2.3}$$

The conditional average of this expression over all paths in velocity space, with respect to the OU probability density, which begin at $u^0$ and terminate at $u$ is

$$\rho_o(u,y,T|u^0,y^0) = E^i_u\{\rho_o[U(t),Y(t)]|U^0(T) = u\}, \tag{2.4}$$

where $U^0(t)$ is the solution to (2.1) and the conditional average is performed with respect to the OU or isolated "$i$" density

$$p_i(u,T|u^0) = \{2\pi k(1 - \exp[-2RT])\}^{-1/2} \times \exp\{-(1/2k)(u - u^0 \times \exp[-RT])^2/(1 - \exp[-2RT])\}\,. \tag{2.5}$$

The relation between the isolated and open transition density is given by

$$p_o(u,y,T\,|u^0,y^0)=\rho_o(u,y,\,T\,|u^0,y^0)\,p_i(u,T\,|u^0)\,, \quad (2.6)$$

which on account of (2.3) can be written in the more instructive form

$$p_o(u,y,T\,|u^0,y^0)$$

$$=\exp\{(1/2k)[S^*(y)-S^*(y^0)]\}\,K_i(u,y,T\,|u^0,y^0)\,, \quad (2.7)$$

where the kernel $K_i$ is

$$K_i(u,y,T\,|u^0,y^0)$$

$$=E^i_u\left\{\exp\left[\left(\frac{1}{2D^*}\right)\int_0^T\left(F\,[Y(t)]\,dU(t)\right.\right.\right.$$

$$\left.\left.\left.-\frac{1}{2}F^2\,[Y(t)]\,dt\right)\right]\,\right|\,U^0(T)=u\Big\}\,p_i(u,T\,|u^0)\,. \quad (2.8)$$

The second stage of the derivation utilizes the fact that expectation values with respect to the velocity space OU, or isolated, process can be converted into expectation values with respect to the velocity space Wiener process whose density is

$$p_W(u,T\,|u^0)=(1/\sqrt{4\pi D^*})\cdot\exp\{-(u-u^0)^2/4D^*T\}\,. \quad (2.9)$$

The probability measure density which relates the two is

$$\rho_i\,[U(t)]$$

$$=\exp\left\{-\frac{1}{2k}\int_0^T\,[U(t)]\,dU(t)+\left(\frac{R}{2}\right)U^2(t)\,dt\right\}\,. \quad (2.10)$$

The conditional expectation of this measure density relates the transition density of the Wiener process (2.9) to the transition density of the isolated process, viz.,

$$p_i(u,T\,|u^0)=\rho_i(u,T\,|u^0)\,p_W(u,T\,|u^0)\,, \quad (2.11)$$

where

$$\rho_i(u,T\,|u^0)$$

$$=E^W_u\left\{\exp\left[-\left(\frac{1}{2k}\right)\int_0^T\left(U(t)dU(t)\right.\right.\right.$$

$$\left.\left.\left.+\left(\frac{R}{2}\right)U^2(t)\,dt\right)\right]\,\right|\,U^0(T)=u\Big\}\,. \quad (2.12)$$

The averaging in (2.12) is performed with respect to the velocity space Wiener measure whose density is (2.9).

Expression (2.12) is in a somewhat inconvenient form due to the presence of the Itô stochastic integral. It can be replaced by the Fisk–Stratonovich integral, which enjoys all the properties of an ordinary integral, by introducing the symmetric "$\circ$" product

$$U(t)\circ dU(t)=U(t)\,dU(t)+D^*\,dt$$

into expression (2.12). We then obtain

$$\rho_i(u,T\,|u^0)$$

$$=\exp\{-(1/4k)[u^2+u^{02}]\}$$

$$\times E^W_u\left\{\exp\left[-\left(\frac{1}{2k}\right)\int_0^T V\,[U(t)]\,dt\right]\,\right|\,U^0(t)=u\Big\}\,, \quad (2.13)$$

where

$$V(u):=(R/2)\,u^2-D^*$$

is the velocity space OM-potential. Introducing (2.13) into (2.11) and the latter into (2.6), we come out with the final expression

$$p_o(u,y,T\,|u^0,y^0)$$

$$=p_\infty^{1/2}(u,y)K_W(u,y,T\,|u^0,y^0)\,p_\infty^{-1/2}(u^0,y^0) \quad (2.14)$$

for the transition density of the open thermodynamic process. The new kernel is the conditional Wiener average

$$K_W(u,y,T\,|u^0,y^0)$$

$$=E^W_u\left\{\exp\left[\frac{1}{2D^*}\int_0^T\left(F\,[Y(t)]\,dU(t)\right.\right.\right.$$

$$\left.\left.\left.-\frac{1}{2}F^2\,[Y(t)]\,dt\right)\right]\right.$$

$$\left.\times\exp\left[-\frac{1}{2k}\int_0^T V\,[U(t)]\,dt\right]\,\right|\,U^0(T)=u\Big\}$$

$$\times p_W(u,T\,|u^0)\,. \quad (2.15)$$

The kernel (2.15) contains information regarding the nature of the statistical correlations between the states $(u^0,y^0)$ and $(u,y)$ provided the time interval $T$ is not very large. In the limit as $T\to\infty$, these states will become statistically independent: a long lapse in time ensures that the process has had ample time to "forget" its past. Provided the invariant probability distribution exists, this means that

$$\lim_{T\uparrow\infty}p_o(u,y,T\,|u^0,y^0)=p_\infty(u,y)=p_\infty(u)\,p_\infty(y)\,, \quad (2.16)$$

independently of the initial conditions. The second equality is due to the time invariancy of the entropy (1.4), since the entropy determines the invariant probability distribution according to (1.3). Taking the asymptotic time limit in (2.14) and using the asymptotic result (2.16), we find that

$$\lim_{T\uparrow\infty}2k\,\ln K_W(u,y,T\,|u^0,y^0)=S(u,y)+S(u^0,y^0)-2S(0)\,, \quad (2.17)$$

where the constant has been chosen so as to satisfy Boltzmann's principle (1.3) in the asymptotic time limit. The limiting relation (2.17) may be taken as a definition of statistical independence. However, if the time lapse is not long, we can expect that the nonequilibrium states will be correlated statistically. As a phase-space generalization of the configuration space joint entropy,[15] we have

$$\sigma_J(u,y,T\,|u^0,y^0):=2k\,\ln K_W(u,y,T\,|u^0,y^0)\,. \quad (2.18)$$

Asymptotic independence (2.17) implies that the joint entropy must reduce the sum of the entropy differences

$$\lim_{T\uparrow\infty}\sigma_J(u,y,T\,|u^0,y^0)=\Delta S(u,y)+\Delta S(u^0,y^0)\,, \quad (2.19)$$

where $\Delta S(u,y)$ is the entropy decrease $S(u,y) - S(0)$.

To prove (2.19), we use the fact tht the transition density of the open process satisfies the pair of Kolmogorov equations

$$-\frac{\partial p_o}{\partial s} = \mathfrak{G}^o p_o \tag{2.20}$$

and

$$\frac{\partial p_o}{\partial t} = \mathfrak{G}^{o\dagger} p_o\,, \tag{2.21}$$

where $\mathfrak{G}^{o\dagger}$ is the formal adjoint of the diffusion operator (1.6) and we have replaced the time interval $T$ by $[s,t]$ to distinguish between initial and final data. The backward equation (2.20) is to be solved subject to the end condition

$$\lim_{s\uparrow t} p_0(u,y,t\,|\,u^0,y^0,s) = \delta(u-u^0)\delta(y-y^0)$$

while the forward or FP equation (2.21) is to be solved with the initial condition

$$\lim_{t\downarrow s} p_0(u,y,t\,|\,u^0,y^0,s) = \delta(u-u^0)\delta(y-y^0)\,.$$

Fortunately enough, we will only need the stationary solution to the Kolmogorov equations (2.20) and (2.21) that arises in the asymptotic time limit.

To facilitate matters, we use the representation (2.14) to convert the FP equation into the self-adjoint diffusion equation

$$\frac{DK_W}{Dt} = D^*\left(\frac{\partial^2 K_W}{\partial u^2}\right) - \left(\frac{1}{2}\right)V(u)K_W\,, \tag{2.22}$$

where $D/Dt$ stands for the Stokes operator

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + u\frac{\partial}{\partial y} + F\frac{\partial}{\partial u}\,.$$

The definition of the joint entropy (2.18) affectuates a logarithmic transformation on the diffusion equation (2.22) giving rise to the nonlinear, generalized Hamilton–Jacobi (HJ) equation

$$-\frac{d\sigma_J}{dt} + \left(\frac{R}{2}\right)\left(\frac{\partial\sigma_J}{\partial u}\right)^2 + D^*\left(\frac{\partial^2\sigma_J}{\partial u^2}\right) = V(u)\,, \tag{2.23}$$

where the total time derivative $d/dt = \partial/\partial t + \{S,\bullet\}$ and

$$\{S,\bullet\} = \left(\frac{\partial S}{\partial y}\right)\left(\frac{\partial \bullet}{\partial u}\right) - \left(\frac{\partial \bullet}{\partial y}\right)\left(\frac{\partial S}{\partial u}\right)$$

are the Poisson brackets. For nonequilibrium statistical thermodynamic processes, the entropy behaves like a classical mechanical Hamiltonian as a generator of the motion. In an analogous way, we obtain

$$\frac{d\sigma_J}{ds} + \left(\frac{R}{2}\right)\left(\frac{\partial\sigma_J}{\partial u^0}\right)^2 + D^*\left(\frac{\partial^2\sigma_J}{\partial u^{02}}\right) = V(u^0) \tag{2.24}$$

by applying the transformations (2.14) and (2.18) to the backward Kolmogorov equation (2.20). We now want to determine the common stationary solution to Eqs. (2.23) and (2.24).

The stationary solution is dictated by the form of the velocity OM potential $V(u)$. Setting the time deterivative equal to zero in Eq. (2.23), we find

$$\left(\frac{\partial}{\partial u}\right)\sigma_J(u,y,T=\infty\,|\,u^0,y^0) = -u\,,$$

provided the Poisson brackets vanish. A similar result is obtained for the backward HJ equation (2.24). Integrating, we obtain

$$\sigma_J(u,y,T=\infty\,|\,u^0,y^0) = -(\tfrac{1}{2})\,[u^2+u^{02}] + C(y,y^0)\,,$$

where the integration constant can depend on the coordinates. If the Poisson brackets are to vanish, it must be a function of the configurational entropy. The simplest choice is

$$C(y,y^0) = S^*(y) + S^*(y^0)\,.$$

This establishes the asymptotic time limit (2.19). We may thus consider

$$2k\ln p_o(u,y,T\,|\,u^0,y^0) = S(u,y) - S(u^0,y^0) + \sigma_J(u,y,T\,|\,u^0,y^0) \tag{2.25}$$

as the phase-space kinetic analog to Boltzmann's principle. The transition density (2.25) together with Boltzmann's principle (1.3) for the invariant distribution determine completely the statistics of the open, second-order thermodynamic process (1.1). Contained in joint entropy is information regarding the velocity and configuration space statistical correlations which we shall investigate in the large resistance limit.

## III. THE ASYMPTOTIC LIMIT OF LARGE RESISTANCE

To introduce the fact that there is a time scale separation between driving and driven processes, we apply the stretching transformation $t^* = \epsilon T$, where $\epsilon = (1/R)$ to the phase-space HJ equations (2.23) and (2.24). We then obtain

$$-\epsilon^2\frac{\partial\sigma_J}{\partial t^*} - \epsilon\left\{u\left(\frac{\partial}{\partial y}\right) + F\left(\frac{\partial}{\partial u}\right)\right\}\sigma_J + \left(\frac{1}{2}\right)\left(\frac{\partial\sigma_J}{\partial u}\right)^2 + k\left(\frac{\partial^2\sigma_J}{\partial u^2}\right) = V(u) \tag{3.1}$$

and

$$-\epsilon^2\frac{\partial\sigma_J}{\partial t^*} + \epsilon\left\{u^0\left(\frac{\partial}{\partial y^0}\right) + F\left(\frac{\partial}{\partial u^0}\right)\right\}\sigma_J + \left(\frac{1}{2}\right)\left(\frac{\partial\sigma_J}{\partial u^0}\right)^2 + k\left(\frac{\partial^2\sigma_J}{\partial u^{02}}\right) = V(u^0)\,. \tag{3.2}$$

We are going to search for an approximate, time-dependent solution to these equations in the form of an asymptotic expansion in the small relaxation parameter $\epsilon$, viz.,

$$\sigma_J(u,y,t^*\,|\,u^0,y^0) = \sum_{n=0}^{\infty}\epsilon^n\sigma_J^{(n)}(u,y,t^*\,|\,u^0,y^0)\,.$$

Substituting this asymptotic expansion into the pair of phase-space HJ equations (3.1) and (3.2) and equating to zero terms in the same power of the relaxation time, we get an infinite set of coupled, nonlinear partial differential equations which are to be solved one after the other beginning with the first. The first three pairs in the infinite hierarchy are

$$\frac{1}{2}\left(\frac{\partial\sigma_J^{(0)}}{\partial u}\right)^2 + k\left(\frac{\partial^2\sigma_J^{(0)}}{\partial u^2}\right) = \frac{1}{2}u^2 - k\,,$$

$$\frac{1}{2}\left(\frac{\partial\sigma_J^{(0)}}{\partial u^0}\right)^2 + k\left(\frac{\partial^2\sigma_J^{(0)}}{\partial u^{02}}\right) = \frac{1}{2}u^{02} - k,$$

$$-\left\{u\left(\frac{\partial}{\partial y}\right) + F\left(\frac{\partial}{\partial u}\right)\right\}\sigma_J^{(0)} + \left(\frac{\partial\sigma_J^{(1)}}{\partial u}\right)\left(\frac{\partial\sigma_J^{(0)}}{\partial u}\right) + k\left(\frac{\partial^2\sigma_J^{(1)}}{\partial u^2}\right) = 0,$$

$$\left\{u^0\left(\frac{\partial}{\partial y^0}\right) + F\left(\frac{\partial}{\partial u^0}\right)\right\}\sigma_J^{(0)} + \left(\frac{\partial\sigma_J^{(1)}}{\partial u^0}\right)\left(\frac{\partial\sigma_J^{(0)}}{\partial u^0}\right) + k\left(\frac{\partial^2\sigma_J^{(1)}}{\partial u^{02}}\right) = 0,$$

$$\frac{-\partial\sigma_J^{(0)}}{\partial t^*} - \left\{u\left(\frac{\partial}{\partial y}\right) + F\left(\frac{\partial}{\partial u}\right)\right\}\sigma_J^{(1)} + \left(\frac{\partial\sigma_J^{(2)}}{\partial u}\right)\left(\frac{\partial\sigma_J^{(0)}}{\partial u}\right) + \frac{1}{2}\left(\frac{\partial\sigma_J^{(1)}}{\partial u}\right)^2 + k\left(\frac{\partial^2\sigma_J^{(2)}}{\partial u^2}\right) = 0,$$

$$\frac{-\partial\sigma_J^{(0)}}{\partial t^*} + \left\{u^0\left(\frac{\partial}{\partial y^0}\right) + F\left(\frac{\partial}{\partial u^0}\right)\right\}\sigma_J^{(1)} + \left(\frac{\partial\sigma_J^{(2)}}{\partial u^0}\right)\left(\frac{\partial\sigma_J^{(0)}}{\partial u^0}\right) + \frac{1}{2}\left(\frac{\partial\sigma_J^{(1)}}{\partial u^0}\right)^2 + k\left(\frac{\partial^2\sigma_J^{(2)}}{\partial u^{02}}\right) = 0.$$

The solution to the first set of equations is easily seen to be

$$\sigma_J^{(0)}(u, y, t^* | u^0, y^0) = -\tfrac{1}{2}[u^2 + u^{02}] + \sigma_J^*(y, t^* | y^0), \tag{3.3}$$

where $\sigma_J^*$ is an (as yet) undetermined integration constant. If we break off the expansion at this order and substitute (3.3) into the phase-space analog of Boltzmann's principle (2.25) we find that the equilibrium MB distribution has been established at every point in configuration space. The higher-order terms in the asymptotic expansion of the joint entropy provide the correction terms.

At next order we find

$$\sigma_J^{(1)}(u, y, t^* | u^0, y^0) = -\left\{u\left[\left(\frac{\partial\sigma_J^*}{\partial y}\right) - F\right] - u^0\left[\left(\frac{\partial\sigma_J^*}{\partial y^0}\right) - F\right]\right\}, \tag{3.4}$$

which involves the spatial derivatives of the unknown integration constant in (3.3). The constant of integration can be determined by substituting in

$$\sigma_J^{(2)}(u, y, t^* | u^0, y^0) = \frac{1}{2}\left\{u^2\left[\left(\frac{\partial^2\sigma_J^*}{\partial y^2}\right) - \frac{\partial F}{\partial y}\right] + u^{02}\left[\left(\frac{\partial^2\sigma_J^*}{\partial y^{02}}\right) - \left(\frac{\partial F}{\partial y^0}\right)\right]\right\} \tag{3.5}$$

into the third set of equations. Our proposed solution will satisfy these equations provided $\sigma_J^*$ satisfies the following pair of configuration space HJ equations:

$$-\frac{\partial\sigma_J^*}{\partial t^*} + \frac{1}{2}\left(\frac{\partial\sigma_J^*}{\partial y}\right)^2 + k\left(\frac{\partial^2\sigma_J^*}{\partial y^2}\right) = \left(\frac{1}{\epsilon}\right)V(y) \tag{3.6}$$

and

$$-\frac{\partial\sigma_J^*}{\partial t^*} + \frac{1}{2}\left(\frac{\partial\sigma_J^*}{\partial y^0}\right)^2 + k\left(\frac{\partial^2\sigma_J^*}{\partial y^{02}}\right) = \left(\frac{1}{\epsilon}\right)V(y^0), \tag{3.7}$$

where $V(y)$ is the configuration space OM potential[15,19]

$$V(y) := \epsilon\left\{\frac{1}{2}F^2 + k\left(\frac{\partial F}{\partial y}\right)\right\}.$$

By introducing the transformation

$$\sigma_J^*(y, t^* | y^0) = 2k \ln p^*(y, t^* | y^0) - S^*(y) + S^*(y^0) \tag{3.8}$$

into the forward HJ equation (3.6) we come out precisely with the configuration space FP or S equation

$$\frac{\partial p^*}{\partial t^*} = -\frac{\partial(Fp^*)}{\partial y} + D\left(\frac{\partial^2 p^*}{\partial y^2}\right), \tag{3.9}$$

where $D$ is the diffusion coefficient in configuration space which is given by the Einstein formula $\epsilon k$. Obviously, the transformation (3.8) converts Eq. (3.7) into the backward, configuration space Kolmogorov equation.

We can now see a pattern being formed in the successive approximations to the joint entropy. At next order, we are led to try

$$\sigma_J^{(3)}(u, y, t^* | u^0, y^0) = -\left(\frac{1}{3!}\right)\left\{u^3\left[\left(\frac{\partial^3\sigma_J^*}{\partial y^3}\right) - \left(\frac{\partial^2 F}{\partial y^2}\right)\right] - u^{03}\left[\left(\frac{\partial^3\sigma_J^*}{\partial y^{03}}\right) - \left(\frac{\partial^2 F}{\partial y^{02}}\right)\right]\right\}. \tag{3.10}$$

This is, however, only approximately valid provided the condition

$$\epsilon^2\left|\frac{\partial F}{\partial y}\right| \ll 1 \tag{3.11}$$

is satisfied. A variance of this criterion has been proposed by Kramers more than forty years ago.[3] Kramers rewrote the phase-space FP equation (2.21) in the divergent form

$$\frac{\partial p_o}{\partial t} = \left\{\left[\left(\frac{1}{\epsilon}\right)\left(\frac{\partial}{\partial u}\right) - \left(\frac{\partial}{\partial y}\right)\right]\left[up_o + k\left(\frac{\partial p_o}{\partial u}\right) - \epsilon\left(Fp_o - k\frac{\partial p_o}{\partial y}\right)\right]\right\} - \epsilon\frac{\partial(Fp_o)}{\partial y} + \epsilon k\left(\frac{\partial^2 p_o}{\partial y^2}\right). \tag{3.12}$$

For time $t > \epsilon$, the displacement and velocity of the Brownian particle will be related by the approximate integral curve

$$y + \epsilon u = y^* = \text{const}. \tag{3.13}$$

Integrating Eq. (3.12) over lines of constant $y^*$, from $y = -\infty$ to $y = \infty$, the term in the curly brackets vanishes and what remains is an equation similar to the S equation (3.9).

Kramers argued that the S approximation should be valid in the limit of large resistance since a MB distribution in velocity space will be established very soon at every point in configuration space. This implies that the transition density should approximately factor into

$$p_o(u,y,t\,|u^0,y^0)$$
$$\approx \hat{p}(y,t\,|\,y^0)\left(\frac{1}{(2\pi k)^{1/2}}\right)\exp\left\{-\left(\frac{1}{2k}\right)[u^2+u^0]\right\}, \tag{3.14}$$

where

$$\hat{p}(y,t\,|\,y^0)=\int_{y+\epsilon u=y^*} p_o(u,y,t\,|u^0,y^0)\,du\,. \tag{3.15}$$

It is evident that $\hat{p}(y,t\,|\,y^0)$ will only approach the true configuration space transition density

$$p^*(y,t\,|\,y^0)=\int_{-\infty}^{\infty} p_o(u,y,t\,|\,u^0,y^0)\,du \tag{3.16}$$

in the limit as $\epsilon\downarrow 0$. Whereas, (3.16) has a clear physical meaning for all values of $\epsilon$, (3.15) is only valid in the limit of large resistance. As we have mentioned in Sec. I, if the external force is relatively constant over distances in which the displacement of the Brownian particle is appreciable, the particle should tend to a limiting velocity $\epsilon F$. The approximate phase-space transition density (3.14) will then be significantly different from zero only in the region where $|u|\leqslant\sqrt{k}$. Moreover, the variations in displacement and velocity are related by the approximate integral curve (3.13) so that the variation in the displacement of the Brownian particle is of $O(\sqrt{k}\,\epsilon)$ for constant $y^*$. Kramers thus concluded that the ES approximation will be valid so long as the inequality

$$\epsilon\sqrt{k}\left|\frac{\partial F}{\partial y}\right| \ll 1 \tag{3.17}$$

is satisfied.

Conditions (3.11) and (3.17) differ merely by constant factors so, at first sight, it would not seem crucial to distinguish between the two. Yet, Boltzmann's constant, relative to the magnitude of the other physical parameters, is a measure of the intensity of random thermal fluctuations[21] so that merely by decreasing $k$, condition (3.17) could always be fulfilled. In fact, condition (3.11) can be justified in the following way.[22] By a change of variables we have

$$\int_{y+\epsilon u=y^*}\left\{\frac{\partial[F(y)p_o]}{\partial y}-k\left(\frac{\partial^2 p_o}{\partial y^2}\right)\right\}du$$
$$=\int_{-\infty}^{\infty}\left(\frac{\partial}{\partial y^*}\right)\left\{\left[F(y^*-\epsilon u)-k\left(\frac{\partial}{\partial y^*}\right)\right]\right.$$
$$\left.\times\, p_o(u,y^*-\epsilon u,t\,|u^0,y^0)\,du\,.\right.$$

Developing $F$ and $p$ in a Taylor series about $y^*$ and using the approximate expression (3.14) for the latter, we get

$$\frac{\partial\hat{p}}{\partial t}=-\epsilon\left(\frac{\partial}{\partial y^*}\right)\left\{F(y^*)-k\left[1-\epsilon^2\left(\frac{\partial F}{\partial y^*}\right)\right]\left(\frac{\partial}{\partial y^*}\right)\right\}\hat{p}\,.$$

Upon comparison with the S equation (3.9), we conclude that (3.11), and not (3.17), is the valid criterion for the validity of the ES approximation.

Kramers' equation (3.12) can be written in the form

$$\frac{\partial p_o}{\partial t}=-\left\{\left[\left(\frac{1}{\epsilon}\right)\left(\frac{\partial}{\partial u}\right)-\left(\frac{\partial}{\partial y}\right)\right][J_V+J_C]\right\}-\left(\frac{\partial J_C}{\partial y}\right) \tag{3.18}$$

by defining the transition current densities

$$J_V:=-\left(up_o+k\frac{\partial p_o}{\partial u}\right)=-\frac{1}{2}\left[u+\frac{\partial\sigma_J}{\partial u}\right]p_o=:v_V p_o \tag{3.19}$$

and

$$J_C:=\epsilon\left(Fp_o-k\frac{\partial p_o}{\partial y}\right)=\frac{1}{2}\epsilon\left[F-\frac{\partial\sigma_J}{\partial y}\right]p_o=:v_C p_o\,. \tag{3.20}$$

On the strength of our asymptotic expansion of the joint entropy, we find that the rate balance condition

$$J_V+J_C=0 \tag{3.21}$$

is satisfied to all orders in the relaxation time $\epsilon$ provided inequality (3.11) is fulfilled. It is precisely this condition which establishes a steady state between the transition current densities in velocity and configuration spaces.

This interpretation can be justified directly from the phase-space FP equation (2.21), which we write in the form

$$\frac{\partial p_o}{\partial t}=-\left\{u\left(\frac{\partial}{\partial y}\right)+F\left(\frac{\partial}{\partial u}\right)\right\}p_o+\left(\frac{1}{\epsilon}\right)\frac{\partial J_C}{\partial u}$$
$$=-\left(\frac{\partial}{\partial y}\right)\left\{u+k\left(\frac{\partial}{\partial u}\right)\right\}p_o$$
$$=\frac{\partial J_V}{\partial y}=-\frac{\partial J_C}{\partial y}\,.$$

Integration over the velocity gives precisely the S equation (3.9).

The kinetic criterion (3.21) for the validity of the ES description, can further be substantiated from an analysis of the velocity moment equations, which are derived from the FP equation (2.21). Integrating it over the velocity variable $u$ gives rise to the continuity equation

$$\frac{\partial p^*}{\partial t}=-\frac{\partial J^*}{\partial y}, \tag{3.22}$$

where the configuration space current density $J^*$ is obtained by integrating the phase-space transition current density (3.19) over the velocity, viz.,

$$J^*(y,t\,|\,y^0)=-\int_{-\infty}^{\infty}J_V(u,y,t\,|u^0,y^0)\,du$$
$$=\int_{-\infty}^{\infty}up_o(u,y,t\,|u^0,y^0)\,du\,.$$

The equation of motion for the first moment of the velocity $J^*$ is obtained by multiplying the FP equation by $u$ and integrating. We then obtain

$$\frac{\partial J^*}{\partial t}=-\left(\frac{1}{\epsilon}\right)J^*+Fp^*-\frac{\partial M}{\partial y}, \tag{3.23}$$

where $M$ is the second velocity moment or the kinetic energy contribution to the momentum current density, i.e.,

$$M(y,t\,|\,y^0):=\int_{-\infty}^{\infty}u^2p_o(u,y,t\,|u^0,y^0)\,du\,.$$

Proceeding further, we find

$$\frac{\partial M}{\partial t}+\frac{\partial Q}{\partial y}-2FJ^*=-\left(\frac{2}{\epsilon}\right)\{M-kp^*\} \tag{3.24}$$

as the equation of motion for the second moment where $Q$ is

the third velocity moment of the kinetic energy current density, i.e.,

$$Q(y,t\,|\,y^0) := \int_{-\infty}^{\infty} u^3 p_o(u,y,t\,|u^0,y^0)\,du\,.$$

Asymptotically, the second moment will tend to $kp^*$ in time so that to leading order in the relaxation time, we have

$$M \approx kp^* + \left\{ FJ_S^* - \left(\frac{1}{2}\right)\frac{\partial Q}{\partial y} + \left(\frac{k}{2}\right)\frac{\partial J_S^*}{\partial y}\right\}, \quad (3.25)$$

where $J_S^*$ is the S current density

$$J_S^* = \epsilon\left[Fp^* - k\frac{\partial p^*}{\partial y}\right]$$

$$= \int_{-\infty}^{\infty} J_c(u,y,t\,|u^0,y^0)\,du\,. \quad (3.26)$$

Differentiating (3.25) with respect to $y$ and using the continuity equation (3.22) we obtain

$$\frac{\partial M}{\partial y} \approx k\frac{\partial p^*}{\partial y} - \frac{\partial J_S^*}{\partial t} + \left(\frac{\epsilon}{2}\right)\left(\frac{\partial}{\partial y}\right)\{3kJ^* - Q\} + \epsilon J_S^*\frac{\partial F}{\partial y}\,.$$

Using this expression to evaluate the last term in Eq. (3.23) we get

$$J^* = \epsilon\left[Fp^* - k\frac{\partial p^*}{\partial y}\right] - \epsilon^2 J_S^*\frac{\partial F}{\partial y}$$

$$= J_S^*\left\{1 - \epsilon^2\left(\frac{\partial F}{\partial y}\right)\right\}, \quad (3.27)$$

where we used the fact that the leading contribution to $Q$ is $3kJ^*$ since $\langle u^3\rangle = 3\langle u^2\rangle\langle u\rangle$ + higher-order terms in $\epsilon$. Expression (3.27) agrees with our previous finding: The S equation (3.9) will be a good approximation to the phase-space FP equation (2.21) when condition (3.11)—and not (3.17)—is satisfied.

Finally, eliminating $J^*$ between the continuity equation (3.22) and the equation of motion of the first moment (3.23) results in

$$\epsilon\frac{\partial^2 p^*}{\partial t^2} + \frac{\partial p^*}{\partial t} - \epsilon\left(\frac{\partial}{\partial y}\right)\left\{\frac{\partial M}{\partial y} - Fp^*\right\} = 0\,. \quad (3.28)$$

Replacing $M$ by its asymptotic form $kp^*$, Eq. (3.28) becomes the telegraph equation which gives the exact result for the average square of a free Brownian particle's displacement with an initial MB distribution. But as Wilemski[10] has pointed out, the second moment $M$ relaxes only slightly faster than the first moment $J^*$ so that this can only be valid for long times where the inertial time has certainly had ample time to die out. Moreover, the rate balance condition (3.21) is simply the stationary solution to the equation of motion of the first moment (3.23) which should be valid for long times provided (3.11) holds.

## IV. THE ONSAGER–MACHLUP PRINCIPLE FOR SECOND-ORDER GAUSSIAN PROCESSES

As an illustration of our kinetic criterion of the ES approximation and the asymptotic expansion of the joint entropy, we treat the special case of Gaussian processes where explicit calculations can be made. For Gaussian processes with inertial effects, the Taylor series expansion of the entropy about the equilibrium state is

$$S(u,y) = S(0) - \tfrac{1}{2}u^2 - \tfrac{1}{2}kQ_\infty^{-1}y^2\,, \quad (4.1)$$

where $Q_\infty$ is the equilibrium second moment in configuration space.

By virtue of the equivalence of means and modes of a Gaussian process, the exact conditional Wiener average for the phase-space transition density (2.14) is equivalent to

$$p_o(u,y,t\,|u^0,y^0) = \max\,.$$

Machlup and Onsager[23] interpreted this maximum likelihood for a transition in terms of the principle of least dissipation of energy. In terms of the phase-space kinetic analog to Boltzmann's principle (2.25), the principle of least dissipation of energy implies that

$$\sigma_J(u,y,t\,|u^0,y^0) \propto -\int_0^t \{\Phi(u) + \Psi(y,\dot u)\}\,ds = \max\,, \quad (4.2)$$

where $\Phi(u) = (\frac{1}{2})Ru^2$ is the Rayleigh–Onsager dissipation function and $\Psi(y,\dot u) = (1/2R)F^{*2}(y,\dot u) = (1/2R)[\dot u + kQ_\infty^{-1}y]^2$ is the generating function.[24] The reason why the joint entropy is only proportional to the negative of the time integral of dissipation functions is due to the fact that we have neglected terms which are proportional to time. These terms arise from the stochastic correction terms to the entropy and joint entropy when they are considered as functionals of the diffusion process and together these terms provide the correct normalization for the transition density.[25]

Denote by $\Omega$ the thermodynamic Lagrangian which is the integrand of (4.2). The condition for an extremum of the joint entropy is

$$\left(\frac{d^2}{dt^2}\right)\left(\frac{\partial}{\partial\dot u}\right)\Omega - \left(\frac{d}{dt}\right)\left(\frac{\partial}{\partial u}\right)\Omega + \frac{\partial\Omega}{\partial y} = 0\,,$$

which is explicitly given by the Euler–Lagrange equation

$$\left(\frac{d^4y}{dt^4}\right) - \{R^2 - 2\beta\}\left(\frac{d^2y}{dt^2}\right) + \beta^2 y = 0\,, \quad (4.3)$$

where $\beta = k/Q_\infty$. Machlup and Onsager[22] observed that Eq. (4.3) can be factored into

$$\left[R\left(\frac{d}{dt}\right) + \beta + \left(\frac{d^2}{dt^2}\right)\right]\left[R\left(\frac{d}{dt}\right) - \beta - \left(\frac{d^2}{dt^2}\right)\right] y = 0\,,$$

which they attributed to a "symmetry in past and future" for the growth and decay of nonequilibrium fluctuations. The average, or most probable, paths for growth and decay are mirror images in time of one another.

After performing several integrations by parts, the joint entropy (4.2) can be written as

$$\sigma_J(u,y,t\,|u^0,y^0)$$

$$\propto \left(\frac{1}{2R}\right)\int_0^t \left\{[R^2 - 2\beta]\left(\frac{d^2y}{ds^2}\right) - \beta^2 y - \left(\frac{d^4y}{ds^4}\right)\right\} y\,ds$$

$$- \frac{1}{2}\left\{Ruy + \left(\frac{1}{R}\right)u\dot u - \left(\frac{1}{R}\right)y\left(\frac{d^2u}{ds^2}\right)\right\}\Bigg|_0^t\,. \quad (4.4)$$

On account of the Euler–Lagrange equation (4.3), the inte-

grand vanishes. With the aid of the general solution to the Euler–Lagrange equation, in the over-damped case say, the joint entropy can be expressed as

$$\begin{aligned}\sigma_J(u,y,t\,|u^0,y^0)&\propto -(1/\Delta)\{(\beta/4)[\,y^2+y^{02}][\sinh Rt+(R/\mu)\sinh\mu t\,]\\&+\tfrac{1}{4}[u^2+u^{02}][\sinh Rt-(R/\mu)\sinh\mu t\,]\\&-uu^0[\sinh(Rt/2)\cosh(\mu t/2)\\&-(R/\mu)\cosh(Rt/2)\sinh(\mu t/2)]\\&-\beta yy^0[\sinh(Rt/2)\sinh(\mu t/2)\\&+(R/\mu)\cosh(Rt/2)\sinh(\mu t/2)]\\&+2[uy^0-u^0y](\beta/\mu)\sinh(Rt/2)\sinh(\mu t/2)\\&-2[uy-u^0y^0](R\beta/\mu^2)\sinh^2(\mu t/2)\}\,,\end{aligned}\tag{4.5}$$

where $\Delta=\sinh^2(Rt/2)-(R/\mu)^2\sinh^2(\mu t/2)$ and $\mu=(R^2-4\beta)^{1/2}$. Expression (4.5) can be arrived at by other methods; for example by the method of characteristic functions. Then retracing our steps, we obtain the principle of least dissipation of energy (4.2) for the phase-space transition density and, apart from a normalization constant, it is exact. This constitutes an explicit proof of the OM principle (4.2) for Gaussian fluctuations.

The symmetry in past and future, which is discernible in the Euler–Lagrange equation (4.3), is a manifestation of the exact balance between the rates of growth and decay of Gaussian fluctuations. A measure of the strength of the statistical correlations in velocity space will be given by

$$\begin{aligned}\frac{\partial\sigma_J}{\partial u}=&-\left(\frac{1}{\Delta}\right)\left\{\sinh\left(\frac{Rt}{2}\right)\right.\\&\times\left[u\cosh\left(\frac{Rt}{2}\right)-u^0\cosh\left(\frac{\mu t}{2}\right)\right]\\&-\left(\frac{R}{\mu}\right)\sinh\left(\frac{\mu t}{2}\right)\left[u\cosh\left(\frac{\mu t}{2}\right)\right.\\&\left.-u^0\cosh\left(\frac{Rt}{2}\right)\right]+\left(\frac{2\beta}{\mu}\right)\sinh\left(\frac{\mu t}{2}\right)\\&\left.\times\left[y^0\sinh\left(\frac{Rt}{2}\right)-\left(\frac{R}{\mu}\right)y\sinh\left(\frac{\mu t}{2}\right)\right]\right\}.\end{aligned}\tag{4.6}$$

In the long time limit, where the velocity correlations have worn off, (4.6) reduces to

$$\lim_{t\uparrow\infty}\left(\frac{\partial}{\partial u}\right)\sigma_J(u,y,t\,|u^0,y^0)=-u\,,$$

independently of the initial conditions. It is precisely this behavior which we predicted in Sec. II [cf., discussion following Eq. (2.24)].

The transition velocity (3.19) is explicitly given by

$$\begin{aligned}v_V(u,y,t\,|u^0,y^0)&=(1/2\Delta)\{[u\exp(-Rt/2)\\&-u^0\cosh(\mu t/2)]\sinh(Rt/2)\\&-(R/\mu)[u\{\cosh(\mu t/2)-(R/\mu)\sinh(\mu t/2)\}\\&-u^0\cosh(Rt/2)]\sinh(\mu t/2)\\&+(2\beta/\mu)[\,y^0\sinh(Rt/2)\\&-(R/\mu)y\sinh(\mu t/2)]\sinh(\mu t/2)\}\,.\end{aligned}\tag{4.7}$$

This transition velocity measures the rate at which fluctuations grow from the most probable path of their growth

$$\begin{aligned}\tilde{y}(t)&=y^0\exp(Rt/2)\{\cosh(\mu t/2)-(R/\mu)\sinh(\mu t/2)\}\\&-(2/\mu)u^0\exp(RT/2)\sinh(\mu t/2)\,.\end{aligned}\tag{4.8}$$

Along (4.8), the transition velocity (4.7) vanishes. The mirror image in time of (4.7) is

$$\begin{aligned}v_V^\dagger(u,y,t\,|u^0,y^0)&=(1/2\Delta)\{[u\exp(Rt/2)-u^0\cosh(\mu t/2)]\sinh(Rt/2)\\&-(R/\mu)[u\{\cosh(\mu t/2)+(R/\mu)\sinh(\mu t/2)\}\\&-u^0\cosh(Rt/2)]\sinh(\mu t/2)\\&+(2\beta/\mu)\sinh(\mu t/2)[\,y^0\sinh(Rt/2)\\&-(R/\mu)y\sinh(\mu t/2)]\}\,.\end{aligned}\tag{4.9}$$

This transition velocity measures the rate of decay of the fluctuations relative to the most probable path of their decay

$$\begin{aligned}\tilde{y}^\dagger(t)&=y^0\exp(-Rt/2)[\cosh(\mu t/2)+(R/\mu)\sinh(\mu t/2)]\\&+(2/\mu)u^0\exp(-Rt/2)\sinh(\mu t/2)\end{aligned}\tag{4.10}$$

and along $\tilde{y}^\dagger(t)$, $v_V^\dagger(t)$ vanishes. The fact that the difference in the nonlocal transition velocities (4.9) and (4.7) reduces to the observed local velocity of the Brownian particle, i.e.,

$$v_V^\dagger(u,y,t\,|u^0,y^0)-v_V(u,y,t\,|u^0,y^0)=u$$

is a consequence of the exact balance between the rates of growth and decay of velocity fluctuations. This embodies the principle of symmetry in past and future for second-order Gaussian processes.

In addition, Gaussian processes allow for an explicit verification of the asymptotic expansion of the joint entropy that was made in Sec. III. In the limit of large resistance, $\mu\approx R-2(\beta/R)$ and in this limit the joint entropy reduces to

$$\begin{aligned}\sigma_J(u,y,t\,|u^0,y^0)&\propto-\tfrac{1}{2}[u^2+u^{02}]+\sigma_J^*(y,t\,|y^0)\\&-[\beta/R\sinh(\beta t/R)]\{(y^0u-yu^0)\\&-(uy-u^0y^0)\exp(-\beta t/R)\}\,,\end{aligned}\tag{4.11}$$

where $\sigma_J^*$ is the configurational joint entropy[15]

$$\begin{aligned}\sigma_J^*(y,t\,|y^0)\propto&-[\beta/2\sinh(\beta t/R)]\\&\times\{(y^2+y^{02})\cosh(\beta t/R)-2yy^0\}\,.\end{aligned}$$

It will now be appreciated that the first two terms in (4.11) comprise the zero-order term (3.3) in the asymptotic expansion of the joint entropy. We calculate the first-order term (3.4) as

$$\begin{aligned}\sigma_J^{(1)}(u,y,t\,|u^0,y^0)&=-2k\{Q^{-1}(-t)[\,y-y^0\exp(\beta t/R)]\,u\\&-Q^{-1}(t)[y-y^0\exp(-\beta t/R)]u\exp(-\beta t/R)\},\end{aligned}\tag{4.12}$$

where $Q(t)$ is the second moment of the distribution, viz.,

$$Q(t) = Q_\infty [1 - \exp(-2\beta t/R)]$$

and $Q(-t)$ is its mirror image in time. A simple rearrangement of (4.12) shows it to be the remaining term in joint entropy expression (4.11). Were we to carry out the asymptotic expansion to second-order, we would find

$$\begin{aligned}\sigma_J^{(2)}(u, y, t \mid u^0, y^0) &= -k(u^2 + u^{02})Q^{-1}(t)\exp(-2\beta t/R) \\ &= -\left(\frac{kR}{2\beta}\right)(u^2 + u^{02})\frac{d \ln Q(t)}{dt} \\ &= -R(u^2 + u^{02})\left(\frac{\partial V^*}{\partial y}\right), \qquad (4.13)\end{aligned}$$

where $V^*$ is the configurational transition velocity[15]

$$\begin{aligned}V^*(y, t \mid y^0) &:= \left(\frac{1}{2R}\right)\left\{\frac{\partial S^*}{\partial y} - \frac{\partial \sigma_J^*}{\partial y}\right\} \\ &= -[RQ(-t)/k]^{-1} \\ &\quad \times \{y - y^0 \exp(\beta t/R)\}.\end{aligned}$$

Expressions (4.13) relate the rate at which the distribution spreads out to the "compressibility" of the fluid motion.[15] Nevertheless, in the limit of large resistance, this will go unobserved since upon multiplying (4.13) by $(1/R)^2$ (to form the second-order term in the asymptotic expansion of the joint entropy), the coefficients combine to give

$$(1/R)^2\beta \ll 1$$

on the strength of criterion (3.11) for the validity of the ES approximation.


## ACKNOWLEDGMENT

This work was supported in part by a contribution from the Consiglio Nazionale delle Richerche (CNR).



[1] I. G. Malkin, *Theory of the Stability of the Motion* (Atomic Energy Commission Translation, Washington, D.C., 1966).
[2] G. E. Uhlenbeck and L. S. Ornstein, Phys. Rev. **36**, 823 (1930).
[3] H. A. Kramers, Physica **7**, 284 (1940).
[4] See, for example, E. Nelson, *Dynamical Theories of Brownian Motion* (Princeton U.P., Princeton, NJ, 1967), pp. 62ff.
[5] S. Chandrasekhar, Rev. Mod. Phys. **15**, 1 (1943).
[6] O. Klein, Ark. Mat. Astron. Fys. **5**, 1 (1922).
[7] R. L. Stratonovich, *Topics in the Theory of Random Noise* (Gordon and Breach, New York, 1967), pp. 110–117.
[8] H. C. Brinkman, Physica **22**, 149 (1956).
[9] R. Landauer and J. A. Swanson, Phys. Rev. **121**, 1668 (1961).
[10] G. Wilemski, J. Stat. Phys. **14**, 153 (1976).
[11] S. R. de Groot and P. Mazur, *Non-Equilibrium Thermodynamics* (North-Holland, Amsterdam, 1962), pp. 191–194.
[12] U. M. Titualer, Physica A **91**, 321 (1978).
[13] F. Marchesoni and P. Grigolini, "Adiabatic elimination procedures" in *Advances in Chemical Physics*, edited by M. W. Evans, P. Grigolini, and G. Pastori-Parravicini (Wiley, New York, 1984).
[14] T. Morita, H. Mori, and K. T. Mashiyama, Prog. Theor. Phys. **64**, 500 (1980).
[15] B. H. Lavenda and E. Santamato, J. Math. Phys. **22**, 2926 (1981).
[16] E. Santamato and B. H. Lavenda, J. Math. Phys. **23**, 2452 (1982).
[17] I. V. Girsanov, Theory Probab. Its Appl. **5**, 285 (1960).
[18] H. Ezawa, J. L. Klauder, and L. A. Shepp, Ann. Phys. (NY) **88**, 588 (1974).
[19] D. Falkoff, Ann. Phys. (NY) **4**, 325 (1958).
[20] K. Yasue, J. Math. Phys. **19**, 1671 (1978); D. Dürr and A. Bach, Commun. Math. Phys. **60**, 153 (1978).
[21] B. H. Lavenda and E. Santamato, J. Stat. Phys. **29**, 345 (1982).
[22] P. A. Lee, J. Appl. Phys. **42**, 325 (1971).
[23] S. Machlup and L. Onsager, Phys. Rev. **91**, 1512 (1953).
[24] B. H. Lavenda, *Thermodynamics of Irreversible Processes* (Macmillan, London/Wiley, New York, 1978), p. 91.
[25] E. Santamato and B. H. Lavenda, Lett. Nuovo Cimento **28**, 189 (1980).

# A multitype random sequential process

B. Mellein and E. E. Mola[a)]
*INIFTA, División Química Teórica, Sucursal 4, Casilla de Correo 16, 1900 La Plata, Argentina*

(Received 25 April 1984; accepted for publication 27 July 1984)

A random sequential process (RSP) can be formulated in two different ways: (I) A linear array of $n$ identical compartments is ceaselessly bombarded by particles which occupy $\beta$ ($\beta \geqslant 2$) contiguous lattice sites. These so-called $\beta$-bell particles are assumed to make contact in a spatially random manner and to stick only if striking $\beta$ (adjacent) vacant compartments. (II) A linear array of $n$ identical compartments is sequentially filled by $\beta$-bell particles, the occupation process being "selective," i.e., always being directed to vacant sequences of compartments of lengths at least $\beta$. Dynamics of RSP's has always been formulated in either way and in the case of lattice spaces all the efforts have been addressed so far to allow in the space filling process *just one kind of particle*. The aim of this paper is to remove that restriction and to consider a lattice space filling problem with *various* kinds of particles involved. The situations of the two above-described models, making no difference in the one-type case, become distinct in the multitype case. In this paper we will be concerned with the generalization of model II.

## I. INTRODUCTION

The solution of problems concerned with the random filling of space with sets of geometrical objects is of considerable interest.[1] However, the exact solution of such problems in more than one dimension presents serious difficulties[1–4] and most of the exact results available are for one-dimensional problems.[2,5–9]

Two related types of problems can be distinguished. In the first, which arises in evaluating partition functions, all possible nonoverlapping configurations of the geometrical objects are assumed to be equally likely,[10–12] while, in the second, the space is filled sequentially and the configurations are not all equally likely (see Fig. 1). The present paper is concerned with a problem of the latter type in one dimension. Some of the above-mentioned problems are as follows.

(i) Adsorption of molecules on a crystal surface[6,13] is a good example of a random sequential process if the temperature is so low that a molecule once adsorbed cannot migrate over the surface. The restriction to one dimension is justified if one is concerned with the adsorption of linear molecules into parallel troughs such as occur on a (110) surface of a face-centered cubic crystal or on a (112) surface of a body-centered cubic crystal.[6]

(ii) Cascade processes[14,15] form another example. At the start of the process there is a particle of specified energy, which is subject to collision and subdivision into particles of smaller energy. This cascade is characterized by the fact that particles below a certain energy level cannot further subdivide.

(iii) There are chemical reactions confined to groups occupying adjacent sites.[16] An example of such a system is the addition of zinc to a solution of polyvinyl chloride,[17] the zinc extracting chlorines in a pairwise manner.

(iv) According to the model proposed by Gornick and Jackson[18] the crystallization of linear polymer chains can be considered as a process involving a random selection of crystallizable sequences from the melt. If, owing to the requirements of thermodynamic stability, such sequences must exceed in length some critical value less than the chain length, then the melt will be increasingly subdivided into uncrystallized sequences of varying length. Some of these may be expected to be less than the critical length, so that the segments comprising them are wasted insofar as participation in further crystallization is concerned.

(v) The kinetics of a reaction in a polymer system where each reacted unit protects its $\alpha$ nearest neighbors completely against reaction.[19] The oxidation of polysaccharides by periodate ions is a good example.

Clearly, in the five above-mentioned examples, after a period of time a saturation situation arises in which the following occur.

(i) The adsorption process stops even though the degree of coverage of the surface is less than 1.

(ii) The cascade reaches a stable terminal state consisting of a finite number of particles.

(iii) The addition of zinc ceases and one wishes to predict the chlorine concentration left in the polymer.

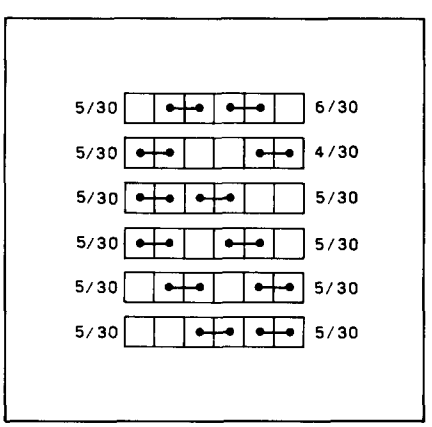



FIG. 1. The total number of configurations of two dimers on a linear array composed of six compartments. Left; probabilities of each configuration in a nonsequential process. Right; probabilities of each configuration in a random sequential process.

[a)] To whom the correspondence related to this article should be sent.

 0022-2488/85/030514-08$02.50 © 1985 American Institute of Physics 

(iv) Further crystallization cannot occur because the length of all created amorphous sequences left is less than the critical value.

(v) Chemical reaction terminates leaving some fraction of unreacted units.

The sequential saturation of a one-dimensional lattice space by identical particles occupying $\beta$ contiguous lattice sites has been extensively studied.[5,6,20,21] The relationship between the sequential filling of a discrete and a continuous one-dimensional space has also been shown.[6,19,22] A random sequential process (RSP) can be formulated in two different ways.

(I) A linear array of $n$ identical compartments is ceaselessly bombarded by particles which occupy $\beta$ ($\beta > 2$) contiguous lattice sites. These so called $\beta$-bell particles (dumbbells or dimers in the case $\beta = 2$, trimers when $\beta = 3$) are assumed to contact in a spatially random manner and to stick only if striking $\beta$ (adjacent) vacant compartments. After some period of time, a saturation situation arises in wich the probability of placing an additional particle on such an array becomes zero since all available space left cannot accommodate further particles.

(II) A linear array of $n$ identical compartments is sequentially filled by $\beta$-bell particles, the occupation process being "selective," i.e., always being directed to vacant sequences of compartments of lengths at least $\beta$. (See Sec. II below.)

Thus while in model I filling is effected by "trial and error" and many trials will not lead to fixation of a particle, in model II all trials are successful.

Dynamics of RSP's has always been formulated in either way and in the case of lattice spaces all the efforts have been addressed so far to allow in the space filling process *just one kind of particle*. The aim of this paper is to remove that restriction and to consider a lattice space filling problem with *various* kinds of particles involved. The situations of the two above-described models, making no difference in the one-type case, become distinct in the multitype case. (See Fig. 2.) This is due to the fact that the "trial and error" occupation procedure of model I will favor, as the lattice space approaches the jammed state, more and more the shorter particles, while the assumptions of model II are such that the likelihood of a given type of particle to get stuck, remains unaltered (as long as such a particle can be accommodated) during the entire filling process. In this paper we will be concerned with the generalization of model II, results on a multitype version of model I will be published elsewhere. In Table I are given the probabilities of observing each configuration of Fig. 2 in either model. Observe that the value of the average number of unoccupied sites is smaller in model II than in model I. This is due to the fact that the conditional probability of placing a trimer, given that a dimer has landed on sites 1 and 2 or sites 4 and 5, is smaller in model I than in model II, thus configurations 5 and 6 being less probable in model I than in model II. Notice also that configuration 4 is equally likely in either model.

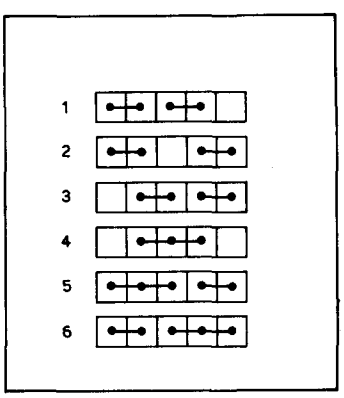



FIG. 2. The six possible ways of saturating a five-sites lattice space with dimers and trimers.

The need of considering the multitype version emerges if one is interested in knowing what would happen when the following occurs.

(i) Linear molecues of *different lengths* are simultaneously adsorbed on a crystal surface.

(ii) In a cascade process in every collision the amount of kinetic energy lost *is not a constant.*

(iii) We consider several chemical reactions each of them confined to groups occupying a *not necessarily constant* number of adjacent sites.

(iv) In a crystallization process the crystallizable sequence is a *random variable* with a given discrete distribution around a given mean value.

(v) Every unit can suffer more than one chemical reaction, each of them protecting *different numbers* of nearest neighbors.

In the five above-mentioned examples of random sequential processes we can distinguish two groups, even in the one-type particle version. One group is formed by examples (i)–(iv); example (v) belongs to a kind of RSP where a partial overlapping between particles[19] is allowed. Models I and II are concerned with the first group of examples. Work on the multitype version of RSP with partial overlapping has also started at La Plata.

## II. THE MODEL

We consider the following sequential process in which particles of varying integral lengths are randomly placed, one by one, onto a one-dimensional lattice space (see Fig. 2) of $n$ equivalent compartments: From a mixture of $\beta$-bell particles, whose concentrations $p_\beta$, $2 \leqslant q \leqslant \beta \leqslant r \leqslant n$, are supposed to remain constant throughout the subsequently described

TABLE I. Probabilities of observing the configurations of Fig. 2 in random sequential processes due to models I and II, and three selected sets of relative frequency ($p_2, p_3$) with $p_2 + p_3 = 1$. The symbol $\bar{m}$ denotes the average number of unoccupied sites.

| Configuration | $p_2 = 0.1$ | | $p_2 = 0.5$ | | $p_2 = 0.9$ | |
|---|---|---|---|---|---|---|
| | I | II | I | II | I | II |
| 1 | 0.0268 | 0.0262 | 0.1625 | 0.1562 | 0.3297 | 0.3263 |
| 2 | 0.0036 | 0.0025 | 0.0750 | 0.0625 | 0.2095 | 0.2025 |
| 3 | 0.0268 | 0.0262 | 0.1625 | 0.1562 | 0.3297 | 0.3263 |
| 4 | 0.3000 | 0.3000 | 0.1666 | 0.1666 | 0.0333 | 0.0333 |
| 5 | 0.3214 | 0.3225 | 0.2167 | 0.2292 | 0.0489 | 0.0558 |
| 6 | 0.3214 | 0.3225 | 0.2167 | 0.2292 | 0.0489 | 0.0558 |
| $\bar{m}$ | 0.6572 | 0.6549 | 0.7332 | 0.7081 | 0.9355 | 0.9217 |

selection and occupation process, a first particle is chosen randomly and placed on the $1\times n$ array at random, i.e., being the particle's length $\beta$, say, its left-hand endpoint has equal probabilities $1/(n-\beta+1)$ of ocupying any of the sites $1,2,\ldots,n-\beta+1$.

In this way arise two random subarrays, a left-hand array consisting of $n_1$ compartments and a right-hand array made up of $n_2$ sites ($n_1+n_2+\beta=n$), whose further occupations will be done independently, in the following manner: If $n_1<q$, the left-hand array will be unoccupied. Otherwise we sample from the mixture until getting (we assume $p_q$ to be positive) a first particle whose length does not exceed $n_1$ and put it randomly onto the left-hand array. We then turn to a *first* further occupation of the right-hand array which is effected in a similar way at random and independently.

In this manner we continue filling the originating random subarrays until no further particle fits. In the final state, known as the "jamming limit," all unoccupied sequences (gaps) between two contiguous particles will be composed of less than $q$ sites.

The total number of unoccupied compartments in the terminal state $A_n$ is a random variable of considerable interest. The study of the asymptotic behavior of its mean and variance, denoted by $a_n$ and $\sigma_n^2$, respectively, is the subject of this paper.

As mentioned in the Introduction, this model constitutes a possible generalization of previously formulated discrete one-dimensional sequential random filling problems[5,6] and consequently provides results which were formerly obtained as special cases.

## III. THE RESULTS

Let $2\leqslant q\leqslant r$ be integer numbers and $\{p_q,p_{q+1},\ldots,p_{r-1},p_r\}$ be a probability distribution on $\{q,q+1,\ldots,r-1,r\}$. Suppose that $p_q>0$ and let $a_n$, as introduced in Sec. II, denote the average number of unoccupied sites of a $1\times n$ array in the jamming limit. Put

$$v_0=0,\quad v_n=\sum_{k=1}^{n}a_k,\quad n=1,2\ldots, \tag{1}$$

$$\xi(s)=\sum_{k=1}^{q-1}\frac{s^k}{k}+\sum_{k=q}^{r-1}\frac{1}{k}s^k\sum_{j=k+1}^{r}p_j \tag{2}$$

(here and in the sequel an empty sum is given the value zero!),

$$\Pi(s)=\sum_{n=1}^{r-1}a_n s^n-2\sum_{k=q+1}^{r-1}s^k\sum_{j=1}^{k-2}\frac{v_j\,p_{k-j}}{(j+1)}, \tag{3}$$

and

$$L(s)=2e^{-2\xi(s)}\int_0^s(1-t)\Pi(t)\,e^{2\xi(t)}\,dt. \tag{4}$$

We shall show in Sec. V that

$$\lim_{n\to\infty}a_n/n=L(1). \tag{5}$$

Effectively, we will derive a more precise result on the asymptotic behavior of $a_n$: As $n$ tends to infinity

$$a_n=(n+M_1)L(1)+o(n^{-k}),\quad \text{for any } k=0,1,\ldots, \tag{6}$$

where

$$M_1=\sum_{k=q}^{r}k\,p_k \tag{7}$$

is the average size of the $\beta$-bell particles taking part in the occupation process.

A similar result is deduced for the variance $\sigma_n^2$ of the random variable $A_n$: As $n$ tends to infinity

$$\sigma_n^2=(n+M_1)[L^2(1)+K(1)]+(M_1^2-M_2)L^2(1)+o(n^{-k}), \tag{8}$$

for any $k=0,1,\ldots$,

where

$$M_2=\sum_{k=q}^{r}k^2p_k \tag{9}$$

is the second moment of the particle size distribution and $K$ the function as given in (36). Observe that the second additive term on the right of (8) vanishes in the case of just one kind of particle, i.e., $p_q=1$, since then $M_1^2=q^2=M_2$.

An immediate consequence of (8) and Chebyshev's inequality[23] is the stochastic convergence of $A_n/n$ to $L(1)$, i.e., for any $\epsilon>0$, the probability that $A_n/n$ differs from $L(1)$ (or equally well from $a_n/n$) by more than $\epsilon$, tends to zero as $n$ tends to infinity. Symbolically,

$$\lim_{n\to\infty}P(|A_n/n-a_n/n|>\epsilon)=0=\lim_{n\to\infty}P(|A_n/n-L(1)|>\epsilon).$$

The proofs of the results (5), (6), and (8) and some alternative representations of the limit in (5) will be given in Sec. V. In the subsequent section we present briefly some special cases which might be of interest or at least illuminating.

## IV. SPECIAL CASES

### A. The most simple three-type model

Let us suppose that the mixture of $\beta$-bell particles contains three types of particles, of lengths 2, 3, and 4, with relative frequencies $p_2>0$, $p_3$, and $p_4$, respectively. Recalling the definitions of the quantities introduced in the foregoing section and noticing that here $q=2$ and $r=4$, we find from (4) and (5) that

$$l(p_2,p_3,p_4)\equiv\lim_{n\to\infty}a_n/n=2\exp\{-2-p_3-\tfrac{5}{3}p_4\}\times\int_0^1(1-t)\left(t+t^3\left[\frac{p_2}{(p_2+p_3)}-p_2\right]\right)\times\exp\{2t+t^2(p_3+p_4)+\tfrac{2}{3}p_4t^3\}\,dt. \tag{10}$$

*(i) The case $p_3=p_4=0$.* We see that (10) reduces to

$$l(1,0,0)=2e^{-2}\int_0^1(1-t)t\,e^{2t}\,dt=e^{-2}\simeq 0.135, \tag{11}$$

the well-known result first established by Page[5] and rediscovered several times.[6,24]

*(ii) The case $p_4=0$.* Equation (10) reduces to

$$l(p_2,p_3,0)=2\exp\{-2-p_3\}\int_0^1(1-t)t\times\exp\{2t+p_3t^2\}\,dt. \tag{12}$$

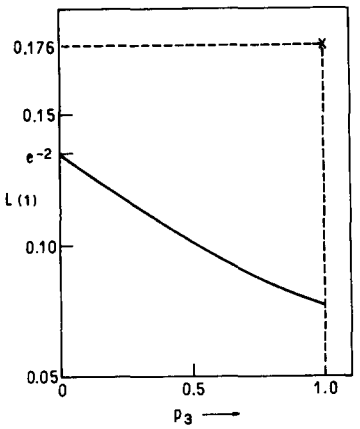



FIG. 3. The average uncovered fraction $L(1)$ in the jamming limit of an infinite lattice space filled sequentially and randomly (due to model II) with dimers and trimers whose respective relative frequencies are $p_2$ and $p_3\times$ average unsaturated coverage in absence of dimers, i.e., $p_2=0$.

from which it is seen that $(\partial/\partial p_3)l(p_2,p_3,0)<0$ and $(\partial^2/\partial p_3^2)l(p_2,p_3,0)>0$. Hence $l(p_2,p_3,0)$ decreases with increasing $p_3$ and as a function of $p_3$ is convex (see Fig. 3). A numerical evaluation of the integral in (12) shows that to four decimal places

$$\lim_{p_3\uparrow 1} l(p_2,p_3,0)\simeq 0.0767.$$

Thus, compared with the case $p_2=1$ (dimers only), the average uncovered portion of an infinite array in the jamming limit diminishes by more than 43%. The corresponding value with relation to the case of trimers is 56%.

*(iii) The general case.* Intuitively, the average saturation coverage in the jamming limit should take on its maximum value when, first, a high percentage of four-bell particles leads to an initial filling of the lattice space almost exclusively by particles of this type, when, second, due to a significant rareness of dumbbells, almost exclusively the three-bell particles take care of the occupation of the segments of exactly three adjacent lattice sites, and when, third, the pure presence of dumbbells provides for a final filling of the gaps consisting of two compartments. Indeed, the function $l$ of (10) takes on its minimum when $0\simeq p_2\ll p_3\simeq 0$ and $p_4\simeq 1$ (see also Fig. 4). More precisely,

$$\lim_{p_3\downarrow 0}\lim_{p_2\downarrow 0} l(p_2,p_3,p_4)) = 2e^{-11/3}\int_0^1 t(1-t)\exp\left\{2t+t^2+\frac{2}{3}t^3\right\}dt \simeq 0.0505.$$

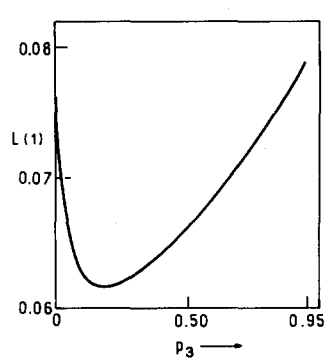



FIG. 4. The average uncovered fraction $L(1)$ in the jamming limit of an infinite lattice space filled sequentially and randomly (due to model II) with particles of lengths 2, 3, and 4 and relative frequencies $p_2=0.05$, $p_3$, and $p_4$, respectively.

*(iv) For $p_2>0$ fixed.* It is interesting to determine the trimer concentration for which the average saturation coverage in the terminal state attains its maximum. As an example we took $p_2=0.05$ and found that this maximum value [the minimum of $l(0.05,p_3,p_4)$] is attained when (see Fig. 4) $p_3\simeq 0.2$ (and hence $p_4\simeq 0.75$; observe that, roughly, $p_3=4p_2$ and $p_4=4p_3$).

### B. Rare dimers in a two-type model

We imagine the mixture of $\beta$-bell particles consisting of just two kinds of particles, dumbbells and $k$-bell particles, $k>3$, the dumbbells being rare, i.e., $p_2\simeq 0$. Then interested in the average fraction of vacant sites in the jammed state, we must determine

$$l(k)\equiv\lim_{p_k\uparrow 1}\left(\lim_{n\to\infty} a_n/n\right). \tag{13}$$

It follows from (2)–(4) that

$$l(k)=2\exp\left\{-2\sum_{j=1}^{k-1}\frac{1}{j}\right\}\int_0^1(1-t)\sum_{j=1}^{k-1}a_j t^j \times\exp\left\{2\sum_{j=1}^{k-1}\frac{t^j}{j}\right\}dt.$$

Some values of $l(k)$ are given in Table II. Utilizing the fact [see Page[5] or observe (6) and (10)] that $a_n\sim e^{-2}n$ as $n\to\infty$, and well-known results[20] on the asymptotic behavior of the average saturation coverage when the particle size becomes large, it is not difficult to show that

$$\lim_{k\to\infty} l(k)=e^{-2}(1-L^*)\simeq 0.033,$$

where

$$L^*=\int_0^\infty\exp\left\{-2\int_0^t\frac{1-e^{-z}}{z}dz\right\}dt\simeq 0.748, \tag{14}$$

the expression gotten by Rényi[2] for the average fraction of occupied space (in the terminal state) of a large one-dimensional parking lot gradually filled at random with cars of unit length.

## V. PROOFS

### A. Mean convergence

Define a pair of random variables $(\beta,\gamma)$ as follows. Let $\beta$ denote the length of the first particle to be placed on the given $1\times n$ array and let $\gamma$ indicate its left-hand position on

TABLE II. The function $l(k)$ introduced in (13).

| $k$ | $l(k)$ |
|---|---|
| 3 | 0.077 |
| 4 | 0.076 |
| 5 | 0.066 |
| 6 | 0.062 |
| 7 | 0.058 |
| 10 | 0.051 |
| 15 | 0.045 |
| . | . |
| . | . |
| . | . |
| ∞ | 0.033 |

the lattice space. It then follows from our model assumptions that, for all $n = r, r+1,\ldots$,

$$\begin{aligned} a_n &= \langle A_n \rangle \\ &= \sum_{k=q}^{r} \sum_{j=1}^{n} \langle A_n | \beta = k, \gamma = j \rangle P(\beta = k, \gamma = j) \\ &= \sum_{k=q}^{r} \sum_{j=1}^{n-k+1} \langle A_n | \beta = k, \gamma = j \rangle P(\gamma = j | \beta = k) \\ &\quad \times P(\beta = k) \\ &= \sum_{k=q}^{r} \frac{p_k}{n-k+1} \left\{ 2a_{n-k} + \sum_{j=2}^{n-k} (a_{j-1} + a_{n-k-j+1}) \right\} \\ &= \sum_{k=q}^{r} \frac{2p_k}{n-k+1} \sum_{j=1}^{n-k} a_j \end{aligned}$$

or, using (1),

$$a_n = \sum_{k=q}^{r} \frac{2p_k}{n-k+1} v_{n-k}, \quad n = r, r+1,\ldots . \tag{15}$$

The initial conditions to be imposed in (15) are

$$a_1 = 1, \quad a_2 = 2,\ldots,a_{q-1} = q-1,$$
$$a_q = 0, \quad a_{q+1} = p_q/(p_q + p_{q+1}),$$

and the additional values $a_{q+2},\ldots,a_{r-1}$ must be calculated successively from (15), redefining $r$ and $p_q,\ldots,p_r$ in every step appropriately.

Before illustrating this procedure let us determine $a_{q+2}$ without making (direct) use of (15): Let $p = p_q + p_{q+1} + p_{q+2}$ and $q = 2$. Since a trimer leaves necessarily one site of a $1\times 4$ array unoccupied and since exactly one (of three possible and equally likely) placement(s) of a first dimer results in two subarrays of length 1 each (which will remain vacant), we see that

$$a_{q+2} = p_{q+1}/p + \tfrac{2}{3} p_q/p, \quad \text{if } q = 2.$$

If $q \geqslant 3$, $q$-bell particles leave two sites (contiguous or not) and $(q+1)$-bell particles leave one site of a $1\times(q+2)$ array unoccupied; thus

$$a_{q+2} = 2p_q/p + p_{q+1}/p, \quad \text{if } q \geqslant 3.$$

Now suppose that $a_1,\ldots,a_{q+t-1}$, $q+t<r$, are known. Set $p = p_q + \cdots + p_{q+t}$ and observe that in the occupation process of a $1\times(q+t)$ array intervene $q$-,..., $(q+t)$-bell particles with relative frequencies $p'_q = p_q/p,\ldots$, $p'_{q+t} = p_{q+t}/p$, respectively. Since this situation is governed by (15) with $n = r = q+t$ and $p_k = p'_k$, $k = q,\ldots,q+t$, $a_{q+t}$ may be determined by means of (15), modified as indicated.

From now on we will assume that $a_{q+3},\ldots,a_{r-1}$ have been determined when dealing with the recursion relation (15).

The next step in evaluating the asymptotic behavior of $a_n$ is to derive a differential equation satisfied by

$$A(s) = \sum_{n=1}^{\infty} a_n s^n, \tag{16}$$

the generating function of the $a_n$, $n = 1,2,\ldots$. Due to the nature of our problem, $|a_n|$ is bounded by $n$; the power series in (16) is therefore uniformly (and absolutely) convergent on compact subsets of the open unit disk. This observation justifies the subsequent interchange of summation and integration and the rearrangement of the series. We proceed from (15). Multiplying both sides of (15) by $s^n$, $0 \leqslant s < 1$, and summing from $r$ to $\infty$ we get

$$\begin{aligned} \sum_{n=r}^{\infty} a_n s^n &= \sum_{k=q}^{r} 2p_k s^{k-1} \sum_{n=r}^{\infty} \frac{s^{n-k+1}}{n-k+1} v_{n-k} \\ &= \sum_{k=q}^{r} 2p_k s^{k-1} \sum_{n=r}^{\infty} v_{n-k} \int_0^s t^{n-k}\, dt \\ &= \sum_{k=q}^{r} 2p_k s^{k-1} \int_0^s \sum_{n=r-k}^{\infty} v_n t^n\, dt. \end{aligned}$$

Now, putting

$$g(s) = \sum_{k=q}^{r} p_k s^{k-1}, \tag{17}$$

and making use of (3) and the fact that

$$\frac{A(t)}{(1-t)} = \sum_{n=1}^{\infty} v_n t^n, \quad 0 \leqslant t < 1,$$

we find that

$$A(s) = 2g(s) \int_0^s \frac{A(t)}{1-t}\, dt + \Pi(s)$$

or

$$A'(s) - A(s) \left[ \frac{g'(s)}{g(s)} + \frac{2g(s)}{1-s} \right] = g(s) \left( \frac{\Pi}{g} \right)'(s). \tag{18}$$

When solving (18) with the initial condition $A(0) = 0$ one uses the easily checked relation [for the definition of $\xi$ see Eq. (2)]

$$\int \frac{g(s)}{1-s}\, ds = -\log(1-s) - \xi(s)$$

after performing an integration by parts, the fact that

$$\lim_{x \downarrow 0} [A(x) - \Pi(x)]/g(x) = 0$$

to obtain

$$\begin{aligned} A(s) = \Pi(s) + \frac{2g(s)}{(1-s)^2} e^{-2\xi(s)} \int_0^s \frac{\Pi(t)}{g(t)} (1-t) \\ \times [1 - (1-t)\xi'(t)]\, e^{2\xi(t)}\, dt \end{aligned}$$

or, on observing (4) and

$$1 - (1-t)\xi'(t) = g(t), \tag{19}$$

$$A(s) = \Pi(s) + g(s) L(s)/(1-s)^2. \tag{20}$$

Since $g(1) = 1$ and $\lim_{s \uparrow 1} L(s)$ exists, it follows from the Tauberian theorem for power series,[23] applied to $A$ in the form (20), that

$$v_{n-1} \sim n^2 L(1)/2, \quad \text{as } n \to \infty. \tag{21}$$

On noticing once more that $0 \leqslant a_n \leqslant n$, (21) together with (1) and (15), yields the desired result (5).

### B. The asymptotic form of the mean

Recall (7) and put

$$C(s) = \sum_{n=1}^{\infty} [a_n - (n + M_1) L(1)]\, s^n, \quad 0 \leqslant s < 1.$$

To prove (6) is then equivalent to showing that $C$ and all its

derivatives $C^{(k)}$ or order $k$, $k = 1,2,...$, converge in 1. It is easily seen that $C$ may be put in the form

$$C(s) = (1-s)^{-2}[g(s)L(s) + \Pi(s)(1-s)^2 + L(1)\{M_1 s^2 - (M_1+1)s\}], \tag{22}$$

which suggests to expand $f(s) \equiv g(s)L(s)$ into a Taylor series around 1. To this end observe that

$$g(1) = 1, \quad g'(1) = M_1 - 1 = \xi'(1), \tag{23}$$

and

$$L'(1) = -2L(1)\xi'(1),$$

which provides

$$f(s) = L(1) + L(1)[1 - M_1](s-1) + \sum_{k=2}^{\infty} f^{(k)}(1)(s-1)^k/k! \tag{24}$$

Hence, on substituting for $g(s)L(s)$ from (24), we find that (22) takes the form

$$C(s) = M_1 L(1) + \sum_{j=2}^{\infty} \frac{f^{(j)}(1)(s-1)^{j-2}}{j!}, \tag{25}$$

proving (6) in the case $k = 0$. Now noticing that derivatives of $L$ (and hence of $f$) of all orders in 1 do in fact exist, (6) follows from (25) by successive derivation of $C$.

### C. The asymptotic form of the variance

First let us introduce some more notations. We set

$$b_n = \langle A_n^2 \rangle, \quad n = 1,2,... \tag{26}$$

$$w_0 = 0, \quad w_n = \sum_{k=1}^{n} b_k, \quad n = 1,2,...$$

$$c_0 = c_1 = 0, \quad c_n = \sum_{k=1}^{n-1} a_k a_{n-k}, \quad n = 2,3,...,$$

$$B(s) = \sum_{n=1}^{\infty} b_n s^n, \tag{27}$$

and

$$\Theta(s) = \sum_{k=1}^{r-1} b_k s^k - 2\sum_{k=q+1}^{r-1} s^k \sum_{j=1}^{k-2} \frac{p_{k-j}(c_j + w_j)}{j+1}.$$

Starting out from the quickly verified recurrence relation

$$b_n = \sum_{k=q}^{r} \frac{2p_k}{n-k+1}(c_{n-k} + w_{n-k}), \quad n = r, r+1, ...,$$

and on observing that

$$A^2(t) = \sum_{n=1}^{\infty} c_n t^n \quad \text{and} \quad \frac{B(t)}{1-t} = \sum_{n=1}^{\infty} w_n t^n,$$

we find, following the line of reasoning indicated in Sec. V A, that $B$ satisfies the first-order linear differential equation

$$B'(s) - B(s)\left[\frac{g'(s)}{g(s)} + \frac{2g(s)}{1-s}\right] = g(s)[2A^2(s) + (\Theta/g)'(s)] \tag{28}$$

subject to the initial condition $B(0) = 0$. Rather than solving (28) we introduce

$$B(s) = 2g(s)T(s)/(1-s)^2 + \Theta(s) \tag{29}$$

in (28) to get

$$T'(s) - [2/(1-s)][g(s) - 1]T(s) = \Theta(s)(1-s) + A^2(s)(1-s)^2. \tag{30}$$

The analytic form of the solution

$$T(s) = e^{-2\xi(s)} \int_0^s [\Theta(t)(1-t) + A^2(t)(1-t)^2] e^{2\xi(t)}\, dt \tag{31}$$

of (30) with initial condition $T(0) = 0$ turns out to be much more convenient for our purposes than that of solution $B$ of (28). It is obvious from (20) that the integral in (31) will diverge in $s = 1$, and our main task will therefore be to split it up into a convergent and a divergent part: Substituting for $A^2(t)$ from (20) we see that (31) may be rewritten in the form

$$T(s) = I_1(s) + I_2(s)e^{-2\xi(s)},$$

where we put

$$I_1(s) = e^{-2\xi(s)} \int_0^s [\Theta(t)(1-t) + \Pi^2(t)(1-t)^2 + 2g(t)\Pi(t)L(t)]\, e^{2\xi(t)}\, dt$$

and

$$I_2(s) = \int_0^s \frac{g^2(t)L^2(t)}{(1-t)^2} e^{2\xi(t)}\, dt. \tag{32}$$

To get rid of the (in $s = 1$) divergent integral on the right of (32), we deduce from (18) and (20) that

$$L'(s) - [2/(1-s)][g(s) - 1]L(s) = 2\Pi(s)(1-s), \tag{33}$$

and on carrying out an integration by parts in (32) and using (19) and (33) we get

$$I_2(s) = I_3(s)e^{2\xi(s)} - I_4(s),$$

with

$$I_3(s) = g^2(s)L^2(s)/(1-s)$$

and

$$I_4(s) = 2\int_0^s \left\{\frac{1}{1-t}[g'(t) - g(t)\xi'(t)]\, g(t)L^2(t) + 2g^2(t)L(t)\Pi(t)\right\} e^{2\xi(t)}\, dt. \tag{34}$$

Due to (23), the integral in (34) converges in $s = 1$, and we arrived at the desired decomposition of the right-hand term in (31),

$$T(s) = I_3(s) + \tfrac{1}{2}K(s), \tag{35}$$

where we set

$$K(s) = 2[I_1(s) - I_4(s)e^{-2\xi(s)}]. \tag{36}$$

To proceed further in the analysis of the asymptotic behavior of $\sigma_n^2$ we recall (7) and (9) and establish that

$$2\xi''(1) = g''(1) = M_2 - 3M_1 + 2$$

and

$$\lim_{t \uparrow 1} [g'(t) - g(t)\xi'(t)]/(1-t) = M_1^2 - (M_1 + M_2)/2.$$

From this, (4), (26), (27), and (34), on introducing (35) in (29) and applying the same method as adopted in Sec. V B, it follows after somewhat lengthy calculations, that, as $n$ tends to infinity

$$b_n = L^2(1)n^2 + [L^2(1)(1+2M_1) + K(1)]n + L^2(1)[2M_1^2 + M_1 - M_2] + K(1)M_1 + o(n^{-k}),$$
$$k = 0,1,\ldots .$$

This, in combination with (6), proves that $\sigma_n^2 = b_n - a_n^2$ behaves as stated in (8).

In the case $q = r = 2$ (dimers only), the function $K$ defined in (36) may be determined explicitly. Particularly, we find that $K(1) = 3e^{-4}$. Since $L^2(1) = e^{-4}$, (8) yields the well-known result[5,6]

$$\sigma_n^2 \sim 4e^{-4}(n+2) \quad \text{as } n \to \infty .$$

In all other cases $K(1)$ must be evaluated numerically. Since the function $L$ given in (4) requires a numerical integration, the calculation of the integral appearing in (34) and hence the determination of $K(1)$ becomes rather tedious. Mackenzie[6] gives some values of $K(1)$ in the one-type case.

### D. Other representations of the limiting average saturation coverage

It has been noticed[20] in the one-type case $q = r$ that a simple modification of the generating function of the $a_n$, $n = 1,2,\ldots$ [e.g., replacing $A$ as given in (16) by $A(s) = \Sigma_{n=q}^{\infty} a_n s^{n-q}$], may lead to a significant alteration of the (integral) representation (4) of the limit in (5). A similar phenomenon appears when one treats the average saturation *coverage* (as, e.g., Mackenzie[6] does in the one-type case) instead of dealing with the *uncovered* fraction in the jamming limit: In the present multitype case, introducing $\alpha_n = n - a_n$, the average number of occupied sites of a $1 \times n$ lattice space in the terminal state, into (15) leads to [$M_1$ is defined in (7)]

$$\alpha_n = M_1 + \sum_{k=q}^{r} \frac{2p_k}{n-k+1} \sum_{j=1}^{n-k} \alpha_j, \quad n = r, r+1, \ldots, \tag{37}$$

the most simple initial values being $\alpha_1 = \cdots = \alpha_{q-1} = 0$, $\alpha_q = q$. Then setting

$$A^*(s) = \sum_{n=q}^{\infty} \alpha_n s^{n-q+1}$$

and proceeding from (37) just as in Sec. V A yields

$$A^*(s) = \frac{g^*(s)}{(1-s)^2} e^{-2\xi(s)} \int_0^s \left[\frac{\Pi^* + h}{g^*}\right]'(t)(1-t)^2 e^{2\xi(t)}\, dt, \tag{38}$$

where $\xi$ is as defined in (2),

$$g^*(s) = s^{1-q} g(s) = \sum_{k=q}^{r} p_k s^{k-q}, \tag{39}$$

$$\Pi^*(s) = \sum_{n=q}^{r-1} \alpha_n s^{n-q+1} - 2 \sum_{k=q+1}^{r-q} s^k \sum_{j=0}^{k-q-1} \frac{p_{q+j}(\alpha_q + \cdots + \alpha_{k-j-1})}{k-j},$$

and

$$h(s) = M_1 s^{r-q+1}/(1-s).$$

Now, in the *one-type case* $q = r$, $g^*(s) \equiv 1$, $\Pi^*(s) \equiv 0$, $M_1 = q$, $h'(s) = q/(1-s)^2$, and (38) reduces to

$$A^*(s) = \frac{qe^{-2\xi(s)}}{(1-s)^2} \int_0^s e^{2\xi(t)}\, dt.$$

Hence, on applying the Tauberian theorem for power series,[23]

$$\lim_{n\to\infty} \frac{a_n}{n} = 1 - qe^{-2\xi(1)} \int_0^1 e^{2\xi(t)}\, dt \equiv L(1), \tag{40}$$

the expression stated by Mackenzie.[6]

In the *multitype case* it seems to be more advantageous to perform an integration by parts in (38). Then, using (39) and once more (19) and the Tauberian theorem for power series, we get

$$\lim_{n\to\infty} a_n/n = 1 - 2e^{-2\xi(1)} \times \int_0^1 t^{q-1}(1-t)[\Pi^*(t) + h(t)]\, e^{2\xi(t)}\, dt, \tag{41}$$

which constitutes, in the case $q = r$, a third representation of $\lim_{n\to\infty} a_n/n$, different from those established in (5) and (40). At least in the one-type case, (41) is easier to handle than (5).

Concerning the integral representations involved in the asymptotic form of the variance, similar modifications may be gotten. However, the main difficulty in computing $K(1)$, described at the end of Sec. V C, cannot be removed.

## VI. FINAL OBSERVATIONS

The exact solution of model II developed in previous sections will enable us to attack problems like those mentioned in the Introduction. The present model is particularly well-adapted for being applied to problems like the model for polymer crystallization, (iv), presented but unsolved by Gornick and Jackson. Even more, the analysis presented in this paper allows the solution of that model *without any kind of restriction* on the crystallizable length which now may be a random variable with any discrete distribution around a given mean value. From the comparison of theoretical results obtained by assuming different types of crystallizable length distributions against experimental results, it will now be possible to find the distribution of crystallizable lengths that thermodynamic stability requires.

From Figs. 3 and 4 we learn that there can be a great difference in the average vacant fraction in the jammed state of a lattice space when (sequentially and randomly) filled by different kinds of particles rather than by just one type of particle, but such that the size of the latter one is equal to the average value of the size distribution of the former ones. As an example, let us consider the size distribution $p_2 = 0.05$, $p_3 = 0.90$, $p_4 = 0.05$ whose average value equals 3. Then (see Fig. 4) the average vacant fraction is less than 8% while (see Fig. 3) the corresponding value in the (one-type) case of a trimer is greater than 17%.

In a more general sense we are in a position to affirm that model II is able to deal particularly well with "space filling" problems mainly determined by "internal" restrictions like examples (ii) and (iv). When the space filling problem is controlled by "external" conditions like in examples (i)

and (iii) we should direct our attention towards model I.

Mackenzie[6] (see also Maltz and Mola[22]) succeeded in obtaining asymptotic expansions of the one-type constants $L(1)$ and $K(1)$ when the particle size grows to infinity, on this occasion establishing a connection with Rényi's continuous model.[2] The complex integral representations of these constants in the present multitype model demand a similar treatment. Interesting relations with work done by Ney[9] can be expected.

## ACKNOWLEDGMENTS

We wish to express our appreciation to A. Maltz for the numerical calculations of this study. Financial support of the Consejo Nacional de Investigaciones Científicas y Técnicas is gratefully acknowledged.

[1]B. E. Blaisdell and H. Solomon, J. Appl. Probab. **7**, 667 (1970).
[2]A. Rényi, Publ. Math. Inst. Hung. Acad. Sci. **3**, 109 (1958).
[3]I. Pálasti, Publ. Math. Inst. Hung. Acad. Sci. **5**, 353 (1960).
[4]H. J. Weiner, J. Appl. Probab. **16**, 706 (1979).
[5]E. S. Page, J. R. Statist. Soc. B **21**, 364 (1959).
[6]J. K. Mackenzie, J. Chem. Phys. **37**, 723 (1962).
[7]A. Dvoretzky and H. Robbins, Publ. Math. Inst. Hung. Acad. Sci. **9**, 209 (1964).
[8]D. Mannion, Publ. Math. Inst. Hung. Acad. Sci. **9**, 143 (1964).
[9]P. E. Ney, Ann. Math. Statist. **33**, 702 (1962).
[10]J. L. Jackson and E. W. Montroll, J. Chem. Phys. **28**, 1101 (1958).
[11]H. S. Green and R. Leipnik, Rev. Mod. Phys. **32**, 129 (1960).
[12]D. Lichtman and R. B. McQuistan, J. Math. Phys. **8**, 2441 (1967).
[13]J. K. Roberts, Proc. R. Soc. London, A **161**, 141 (1937).
[14]P. E. Ney, Ph.D. thesis, Columbia University, 1960.
[15]J. P. Mullooly, J. Appl. Probab. **5**, 427 (1968).
[16]D. A. McQuarrie, J. Appl. Probab. **4**, 413 (1967).
[17]E. R. Cohen and H. Reiss, J. Chem. Phys. **38**, 680 (1963).
[18]F. Gornick and J. L. Jackson, J. Chem. Phys. **39**, 1150 (1963).
[19]J. J. Gonzalez, P. C. Hemmer and J. S. Høye, J. Chem. Phys. **3**, 228 (1974).
[20]B. Mellein, A. Maltz, and E. E. Mola, in *Stochastic Processes Applied to Physics and Other Related Fields*, edited by B. Gómez *et al.* (World Scientific, Singapore, 1983), Vol. 1, p. 199.
[21]B. Mellein, A. Maltz, and E. E. Mola, J. Math. Phys. **25**, 1995 (1984).
[22]A. Maltz and E. E. Mola, J. Chem. Phys. **79**, 5141 (1984).
[23]W. Feller, *An Introduction to Probability Theory and Its Applications 2* (Wiley, New York, 1971), 2nd ed.
[24]R. B. McQuistan and D. Lichtman, J. Math. Phys. **9**, 1680 (1968).

# Exact temporal evolution for some nonlinear diffusion processes

L. Garrido
*Departamento de Física Téorica, Universidad de Barcelona, Diagonal 647, Barcelona-28, Spain*

J. Masoliver
*Departament de Matemàtiques, Escola Tècnica Superior Enginiers Telecomunicació, Universitat Politècnica de Barcelona, Barcelona, Spain*

(Received 24 April 1984; accepted for publication 9 August 1984)

Exact solutions to Fokker–Planck equations with nonlinear drift are considered. Applications of these exact solutions for concrete models are studied. We arrive at the conclusion that for certain drifts we obtain divergent moments (and infinite relaxation time) if the diffusion process can be extended without any obstacle to the whole space. But if we introduce a potential barrier that limits the diffusion process, moments converge with a finite relaxation time.

## I. INTRODUCTION

A time-dependent Fokker–Planck equation (FPE) describes the dynamical evolution of the diffusion processes. Nevertheless, when the dynamics of the process is nonlinear it is very difficult to obtain exact or even approximate solutions of such FPE's. Since at the present time nonlinear processes are of highest interest (instabilities, phase transitions, etc.)[1] many people have tried to find exactly soluble nonlinear models.[2–7]

The most common technique used to solve exactly a FPE consists in separating the temporal from the spatial dependence; this latter one is solved by means of an eigenfunction expansion in the same way as occurs with the Schrödinger equation.[2–4]

Another more direct although more skillful technique is initiated in Ref. 6 and continued in Ref. 8. It consists in separating the part which is most related to the potential of the process (which causes the nonlinearity) from the probability density $P(q,t)$; then, by means of convenient assumptions, the remaining part of $P(q,t)$ is solved separately assuming that it is Gaussian. Concretely in Ref. 8 we have found that the $N$-dimensional FPE

$$\dot{P}(q,t) = -\partial_\mu [f^\mu(q)\cdot P(q,t)] + \tfrac{1}{2}\partial_\mu \partial^\mu P(q,t), \tag{1.1}$$

when $\partial_\mu \equiv \partial/\partial q^\mu$ (sum over repeated Greek indices is assumed) has an exact solution, with the usual initial condition

$$P(q,0) = \delta^n(q - q_0) \tag{1.2}$$

if the drift $f^\mu(q) = f^\mu(q_1,q_2,\ldots,q_N)$ has the form

$$f^\mu(q) = -aq^\mu + \partial_\mu \phi(q)/\phi(q), \tag{1.3}$$

where

$$\phi(q) = \prod_{k=1}^{N} \left| \alpha_k F\left(l_k \left| \frac{1}{z} \right| aq_k^2\right) + \beta_k q_k F\left(l_k + \frac{1}{2} \left| \frac{3}{2} \right| aq_k^2\right)\right|, \tag{1.4}$$

the $\alpha_K$ and $\beta_k$ being arbitrary constants and the $l_k$ arbitrary parameters. The function $F(l_k|\frac{1}{2}|aq_k^2)$ is the hypergeometric confluent function. In this case the exact solution to the FPE (1.1) is

$$P(q,t/q_0) = \left(\frac{a}{2\pi}\right)^{N/2} \frac{\phi(q)}{\phi(q_0)} \times \frac{\exp\{-(b/2)t - a\sum_{k=1}^{N}(q_k - \beta_k(t))^2/\eta(t)\}}{(\sinh at)^{N/2}}, \tag{1.5}$$

with

$$b \equiv \left[4\left(\sum_{k=1}^{N} l_k\right) - N\right]a, \tag{1.6}$$

$$\beta_k(t) \equiv q_{0k}e^{-at}, \tag{1.7}$$

$$\eta(t) \equiv 1 - e^{-2at}. \tag{1.8}$$

In Ref. 8 we have also found solutions of Eqs. (1.1) for a drift with spherical symmetry of the form

$$\mathbf{f}(r) = \left[-ar + \frac{d\phi(r)/dr}{\phi(r)}\right]\frac{\mathbf{r}}{r} \tag{1.9}$$

with

$$\phi(r) = |\alpha F(l|N/2|ar^2)| \tag{1.10}$$

for whatever value of the dimension $N$ of the phase space, and

$$\phi(r) = \left|\alpha F\left(l\left|\frac{N}{2}\right|ar^2\right) + \beta r^{2-N} \times F\left(l+1-\frac{N}{2}\left|2-\frac{N}{2}\right|ar^2\right)\right| \tag{1.11}$$

if $N$ is odd. In these cases the normalized solution of the FPE is given by (1.5) with $\phi(q)$ given by (1.10) or (1.11) and

$$b \equiv (4l - N)a. \tag{1.12}$$

In this article we intend to study some of the applications of these solutions for concrete models.

## II. A FIRST MODEL OF NONLINEAR DIFFUSION

By means of an adequate selection of the constants $\alpha_k$ and $\beta_k$ that appear in (1.4) we can write

$$\phi(q) = \left|\prod_{k=1}^{N} \{e^{|\alpha|2|q_k^2} D_{-2l_k}(\sqrt{2a}q_k)\}\right|, \tag{2.1}$$

 0022-2488/85/030522-06$02.50 © 1985 American Institute of Physics 

where $D_{-l_k}(\sqrt{2a}q_k)$ is the parabolic cylindric function defined by[9]

$$D_\nu(z) = z^{\nu/2}e^{-z^2/4}\left\{\frac{\Gamma(1/2)}{\Gamma((1-\nu)/2)}F\left(\frac{-\nu}{2}\middle|\frac{1}{2}\middle|\frac{z^2}{2}\right) + \frac{z}{\sqrt{2}}\frac{\Gamma(-1/2)}{\Gamma(-\nu/2)}F\left(\frac{1-\nu}{2}\middle|\frac{3}{2}\middle|\frac{z^2}{2}\right)\right\}. \quad (2.2)$$

For $l \geqslant 0$ the functions $D_{-l_k}(\sqrt{2a}q_k)$ do not have real zeros and instead of (2.1) we can write

$$\phi(q) = \prod_{k=1}^{N}\{e^{(a/2)q_k^2}D_{-2l_k}(\sqrt{2a}q_k)\}. \quad (2.3)$$

The characteristic function $\Theta(\mu_1,\ldots,\mu_N)$ associated to a density of probability $P(q,t)$ is given by[1]

$$\theta(\mu_1,\ldots,\mu_N) \equiv \int dq_1\cdots dq_N \times \exp\{i(\mu_1 q_1 + \cdots + \mu_N q_N)\}P(q,t). \quad (2.4)$$

Substituting in (2.4) the probability density given by Eq. (1.5) with $\phi(q)$ given by (2.3) we have

$$\Theta(\mu_1,\ldots,\mu_N) = \left(\frac{a}{2\pi}\right)^{N/2}\frac{e^{-(b/2)t}}{\phi(q_0)(\sinh at)^{N/2}} \times \prod_{k=1}^{N}\int_{-\infty}^{+\infty} dq_k D_{-2l_k}(\sqrt{2a}q_k)\exp\left\{i\mu_k q_k + \frac{1}{2}aq_k^2 - \frac{a(q_k-\beta_k(t))^2}{\eta(t)}\right\}, \quad (2.5)$$

since[9]

$$D_{-2l_k}(\sqrt{2a}q_k) = \frac{e^{-(1/2)aq_k^2}}{\Gamma(2l_k)}\int_0^\infty \exp(-\sqrt{2a}q_k s - \tfrac{1}{2}s^2)s^{2l_k-1}\,ds \quad (l_k \geqslant 0). \quad (2.6)$$

Substituting (2.6) in (2.5) we finally arrive at

$$\theta(\mu_1,\ldots,\mu_N) = \frac{1}{\phi(q_0)}\cdot\prod_{k=1}^{N}\left[\exp\left\{i\mu_k\beta_k(t) - \frac{1}{4a}\eta(t)u_k^2\right\} \times e^{-(1/2)\rho_k^2}D_{-2l_k}(\rho_k)\right], \quad (2.7)$$

where

$$\rho_k \equiv \sqrt{2a}q_{0k} + i(z/\sqrt{2a})(\sinh at)\mu_k. \quad (2.8)$$

Once we have evaluated the characteristic function, moments follow easily[1]:

$$\langle q_1,\ldots,q_r\rangle = \frac{1}{i^r}\frac{\partial^r\theta(\mu_1\cdots\mu_r)}{\partial\mu_1\cdots\partial\mu_r}\bigg|_{\mu_1=\cdots=\mu_r=0}. \quad (2.9)$$

In our case

$$\langle q_k(t)\rangle = q_{0k}e^{-at} - \frac{4l_k}{\sqrt{2a}}(\sinh at)\frac{D_{-2l_k-1}(\sqrt{2a}q_k)}{D_{-2l_k}(\sqrt{2a}q_k)} \quad (l_k \geqslant 0) \quad (2.10)$$

and

$$\langle [q_k(t)]^2\rangle = q_{0k}^2e^{-2at} + \frac{1-e^{-2at}}{2a} - \frac{8l_k}{\sqrt{2a}}q_{0_k}e^{-at}(\sinh at)\frac{D_{-2l_k-1}(\sqrt{2a}q_{0k})}{D_{-2l_k}(\sqrt{2a}q_{0k})} + \frac{4l_k(l_k+1)}{a}(\sinh^2 at)\frac{D_{-2l_k-2}(\sqrt{2a}q_k)}{D_{-2l_k}(\sqrt{2a}q_{0k})} \quad (l_k \geqslant 0). \quad (2.11)$$

In this model the drift may be written in the form

$$f^\mu(q) = -aq^\mu - 2l_\mu\sqrt{a}\left[D_{-2l_\mu-1}(\sqrt{2a}q)/D_{-2l_\mu}(\sqrt{2a}q)\right]. \quad (2.12)$$

The first moment (2.10) as a function of the drift is

$$\langle q_k(t)\rangle = q_{0k}e^{-at} + (\sinh at)[q_{0k} + f_k(q_0)/a]. \quad (2.13)$$

This first moment presents a "boomerang" effect since the average velocity

$$\frac{d\langle q_k(t)\rangle}{dt} = -a\left\{q_{0k}e^{-at} - (\cosh at)\left[q_0 + \frac{f_k(q_0)}{a}\right]\right\} \quad (2.14)$$

becomes equal to zero and changes the sign if $0 < l_k < \frac{1}{4}$ and $-q_{1k} < q_{0k} < q_{1k}$ ($\pm q_{1k}$ are the positions of the maxima of the potential of the drift) for a time $t_b$ (see Ref. 6):

$$(t_b)_k = \frac{1}{2a}\ln\left[\frac{aq_{0_k} - f_k(q_0)}{aq_{0k} + f_k(q_0)}\right]. \quad (2.15)$$

We easily observe that Eqs. (2.10) and (2.11) give

$$\lim_{t\to\infty}|\langle q_k(t)\rangle| = \lim_{t\to\infty}|\langle q_k(t)^2\rangle| = \infty, \quad (2.16)$$

except for $l_k = 0$ that corresponds to the case of linear drift [see Eq. (2.12)].

We can also consider the model with spherical symmetry such that, when $q_0 = 0$, the probability density (1.5) can be written as

$$P(r,t\,|0) = \left(\frac{q}{2\pi}\right)^{N/2}\frac{\phi(r)}{\phi(0)}\frac{\exp\{-(b/2)t - ar^2/\eta(t)\}}{(\sinh at)^{N/2}}, \quad (2.17)$$

where $r \equiv (\Sigma_{k=1}^N q_k^2)^{1/2}$ and $\phi(r)$ is given by (1.10). For $l \geqslant 0$ the function $F(1|N/2|ar^2)$ has no zeros and since $F(a|c|0) = 1$, we have

$$\phi(r)/\phi(0) = F(l\,|N/2|ar^2) \quad (l \geqslant 0). \quad (2.18)$$

For this model the potential $V(r)$ of the drift (1.3) is

$$V(r) = \tfrac{1}{2}ar^2 - \ln F(l\,|N/2|ar^2) \quad (l \geqslant 0).$$

In Fig. 1 we represent this potential when $N = 1$ and $a = 1$, for the cases (a) $l = 0$, (b) $l = 0.1$, and (c) $l = 0.5$. The moments $\langle [r(t)]^m\rangle$ are given by

$$\langle [r(t)]^m\rangle = \frac{(2\pi)^{N/2}}{\Gamma(N/2)}\int_0^\infty r^{N+m-1}P(r,t\,|0)dr \quad (m = 1,2,\ldots). \quad (2.19)$$

Substituting in (2.19) the probability density given by (2.17) and since[9]

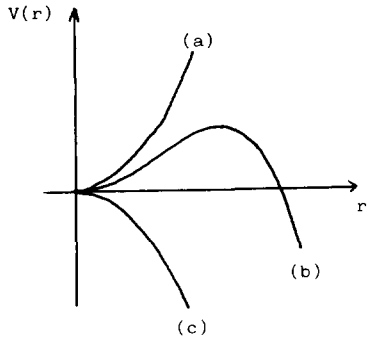



FIG. 1. Representation of the potential $V(r)$ when $N=1$ and $a=1$ for the cases (a) $l=0$, (b) $l=0.1$, and (c) $l=0.5$.

$$\int_0^\infty e^{-\alpha z} z^{V-1} F(a|c|z)dz = \alpha^{-V}\Gamma(V)F\left(a,V|c|\frac{k}{\alpha}\right), \tag{2.20}$$

for $\mathrm{Re}\,\alpha>0$, $\mathrm{Re}\,\alpha>\mathrm{Re}\,k$, $\mathrm{Re}\,V>0$, we arrive easily at

$$\langle[r(t)]^m\rangle = \frac{\Gamma((m+N)/2)}{2^{1-N/2}a^{m/2}\Gamma(N/2)} e^{-2lat}[\eta(t)]^{m/2} \times F\left(l,\frac{m+N}{2}\left|\frac{N}{2}\right|\eta(t)\right). \tag{2.21}$$

In this case the velocity of the moments is given by

$$\frac{d\langle[r(t)]^m\rangle}{dt} = \frac{2^N\Gamma((m+r)/2)}{a^{m/2-1}\Gamma(N|z)} e^{-2lat}[\eta(t)]^{m/2-1} \times\left\{\left[\frac{m}{2}(1-\eta(t))+l\eta(t)\right] \times F\left(l,\frac{m+N}{2}\left|\frac{N}{2}\right|\eta(t)\right) + \frac{l(N+m)}{N}\eta(t)(1-\eta(t))F\left(1+l,1+\frac{m+N}{2}\left|1+\frac{N}{2}\right|\eta(t)\right)\right\}. \tag{2.22}$$

Now we investigate the behavior of the moments $\langle[r(t)]^m\rangle$ for large times ($t\gg 1/2a$). In this case

$$\eta(t)\to 1 \quad (t\gg 1/2a) \tag{2.23}$$

and since[9]

$$F(a,b|c|z) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-b)\Gamma(c-a)} \times F(a,b|a+b-c+1/1-z) + (1-z)^{(-a-b)}\frac{\Gamma(a)\Gamma(a+b-c)}{\Gamma(a)\Gamma(b)} \times F(c-a,c-b|c-a-b+1|1-z) \tag{2.24}$$

and $F(a,b|c|1-n)\simeq 1$ $(t\gg 1/2a)$, we arrive at

$$\langle[r(t)]^m\rangle \simeq \frac{\Gamma(l+m/2)}{2^{1-N/2}a^{m/2}\Gamma(l)} e^{mat} \quad \left(t\gg\frac{1}{2a} \text{ and } l\neq 0\right). \tag{2.25}$$

This expression diverges when $t\to\infty$.

The case $l=0$ corresponds to linear drift and, therefore, presents no difficulty.

We come to the conclusion that both models presented in this section could not be valid for the study of the temporal evolution of physical systems towards equilibrium. In the following section we find a mechanism that yields exactly soluble models that relax towards equilibrium with a finite relaxation time.

## III. ONE-DIMENSIONAL MODEL WITH POTENTIAL BARRIER

For a one-dimensional system the potential $V(x)$ of the drift (1.3) is

$$V(x) = \tfrac{1}{2}ax^2 - \ln\phi(x), \tag{3.1}$$

where $\phi(x)$ is given by Eq. (1.4). By means of an adequate choice of the constants $\alpha_k$ and $\beta_k$, we can write

$$\phi(x) = U(l|\tfrac{1}{2}|ax^2), \tag{3.2}$$

where $U(l|\frac{1}{2}|ax^2)$ is a function of Kummer.[9] Let us suppose now, that, for a certain value of $x_1<x_0$, there exists a potential barrier, that is,

$$V(x) = \begin{cases} \tfrac{1}{2}ax^2 - \ln\phi(x), & x\geqslant x_1, \\ \infty, & x<x_1, \end{cases} \tag{3.3}$$

which is equivalent to the following expression for $\phi(x)$

$$\phi(x) = \begin{cases} U(l|\tfrac{1}{2}|ax^2), & x\geqslant x_1, \\ 0, & x<x_1. \end{cases} \tag{3.4}$$

This is possible since $\phi(x)=0$ is also a solution of the differential equation that satisfies the function (3.2) (see Ref. 8).

In this case, and supposing that $x_0=0$, the probability density is

$$P(x,t|0) = \left(\frac{a}{2\pi}\right)^{1/2}\frac{\phi(x)}{\phi(0)}\frac{\exp\{-(b/2)t - ax^2/\eta(t)\}}{(\sinh at)^{1/2}}. \tag{3.5}$$

In this model the moments are evaluated by

$$\langle[x(t)]^m\rangle = \int_{x_1}^\infty x^m P(x,t|0)dx \quad (m=1,2,\ldots), \tag{3.6}$$

following the procedure described in the Appendix. Expression (3.6) becomes

$$\langle[x(t)]^m\rangle = \frac{\Gamma(l+\frac{1}{2})}{2\pi a^{m/2}}(\eta(t))^{m/2-l} \times\exp[-2lat + ax_1^2/\eta(t)] \times\sum_{n=0}^\infty \psi_n^{(m)}(e,x_1;\eta(t)), \tag{3.7}$$

where

$$\psi_n^{(1)}(l,x_1;\eta(t)) = \frac{(l)_n(l+\frac{1}{2})_n}{n!}\left(1-\frac{1}{\eta(t)}\right)^n \times U\left(l+n\left|-\frac{1}{2}\right|\frac{ax_1^2}{\eta(t)}\right), \tag{3.8a}$$

$$\psi_n^{(2)}(l,x_1;\eta(t)) \equiv \frac{(l)_n(l+\frac{1}{2})_n}{n!}\left(1-\frac{1}{\eta(t)}\right)\left[\frac{ax_1^2}{\eta(t)}U\left(l+n+\frac{1}{2}\left|\frac{1}{2}\right|\frac{ax_1^2}{\eta(t)}\right) + U\left(l+n+\frac{1}{2}\left|-\frac{1}{2}\right|\frac{ax_1^2}{\eta(t)}\right)\right]. \tag{3.8b}$$

When $t \ll 1/2a$, then $\eta(t) \simeq 2at \ll 1$ and the main term of Eq. (3.7) is $\exp(-ax_1^2/\eta(t))$. Therefore,

$$\langle [x(t)]^m \rangle \sim e^{-x_1^2/2t} \quad (t \ll 1/2a). \tag{3.9}$$

When $t \gg 1/2a$, then $\eta(t) \simeq 1$ and the main term of Eq. (3.7) is $\exp(-2lat)$, i.e.,

$$\langle [x(t)]^m \rangle \sim e^{-2lat} \quad (t \gg 1/2a), \tag{3.10}$$

that yields a relaxation time

$$\tau_{\text{relax}} = 1/2la. \tag{3.11}$$

The moments, for small times, are growing functions of time; however, for large times, they are decreasing functions of time. Therefore, in the cases in which (3.7) are continuous functions of time, the moments pass through a maximum for $t \sim 1/2a$ and we again have the "boomerang" effect.

We will finish this section studying the case when the potential barrier is very far away from the origin (that is, our initial state), i.e., when

$$ax_1^2 \gg 1. \tag{3.12}$$

In such a case, as in Ref. 10,

$$U(a|c|z) \sim z^{-a} \quad \text{for} \quad z \to \infty \quad (\text{Re}\, a > 0), \tag{3.13}$$

the functions $\psi_n^{(m)}(l,x_1;\eta(t))$, defined by (3.8), may be written

$$\psi_n^{(m)}(l,x_1;\eta(t)) \simeq \left(\frac{\eta(t)}{ax_1^2}\right)^{l+(1-m)/2} \times \left\{\frac{(l)_n (l+\tfrac{1}{2})_n}{n!} \left(\frac{\eta(t)-1}{ax_1^2}\right)^n\right\} \quad (m = 1,2).$$

Taking into consideration[9]

$${}_2F_0(a,b;z) \equiv \sum_{n=0}^{\infty} (a)_n (b)_n \frac{z^n}{n!},$$

$${}_2F_0(a,b;z^{-1}) = z^a U(a|a-b+1|z),$$

we arrive at

$$\sum_{n=0}^{\infty} \psi_n^{(m)}(l,x;\eta(t)) \simeq \left(\frac{\eta(t)}{ax_1^2}\right)^{1-m/2} \left(\frac{\eta(t)}{1-\eta(t)}\right)^l \times U\left(l \left| \frac{1}{2} \right| \frac{ax_1^2}{1-\eta(t)}\right) \quad (ax_1^2 \gg 1), \tag{3.14}$$

for $m = 1,2$. Substituting (3.14) into Eq. (3.7) we finally get

$$\langle [x(t)]^m \rangle \simeq \frac{\Gamma(l+\tfrac{1}{2})}{2a^{m/2}\pi} \frac{e^{-ax_1^2/\eta(t)}}{(ax_1^2)^{1-m/2}} e^{-4lat} \times U\left(l \left| \frac{1}{2} \right| \frac{ax_1^2}{\eta(t)}\right) \quad (ax_1^2 \gg 1), \tag{3.15}$$

for $m = 1,2$.

When $t \gg 1/2a$, we have $\eta(t) \simeq 1$. With the approximation (3.13) we have

$$\langle [x(t)]^m \rangle \simeq \frac{\Gamma(l+\tfrac{1}{2})}{2\pi a^{m/2}} \frac{e^{-ax_1^2}}{(ax_1^2)^{1-m/2}} e^{-2lat} \quad (ax_1^2 \gg 1 \quad \text{and} \quad t \gg 1/2a). \tag{3.16}$$

When $t \ll 1/2a$ we can approximate

$$\eta(t) \simeq 2at \ll 1, \quad 1-\eta(t) \simeq 1, \quad e^{-4lat} \simeq 1$$

(for $l$ moderate).

Remembering (3.13), Eq. (3.15) becomes

$$\langle [x(t)]^m \rangle \simeq \frac{\Gamma(l+\tfrac{1}{2})}{2\pi a^{m/2}} \frac{e^{-x_1^2/2t}}{(ax_1^2)^{l+(1-m)/2}} \quad (ax_1^2 \gg 1 \quad \text{and} \quad t \ll 1/2a). \tag{3.17}$$

Therefore, even when the potential barrier is very far away from the initial state, the evolution of the system depends on the position $x_1$ of the barrier.

If we compare Eq. (3.16) with Eq. (3.10), and Eq. (3.17) with Eq. (3.9), we observe that the asymptotic temporal evolution of the model is similar to the evolution of the general case.

For $x < x_1$ the potential (3.3) is a hard-core potential. This implies that the barrier is a reflecting barrier. Thus, the probability current $J(x,t)$ must be zero for $x \leqslant x_1$. In our case $J(x,t)$ is given by

$$J(x,t) = K(x,t)[a(1/\eta(t) - 1)x\phi(x) + \tfrac{1}{2}\phi'(x)], \tag{3.18}$$

where

$$K(x,t) \equiv \left(\frac{a}{2\pi}\right)^{1/2} \frac{\exp\{-(b/2)t - ax^2/\eta(t)\}}{(\sinh at)^{1/2}}$$

and[9]

$$\phi'(x) = \begin{cases} -2l^2 x U(l+1|\tfrac{3}{2}|ax^2), & x \geqslant x_1, \\ 0, & x < x_1. \end{cases} \tag{3.19}$$

Thus $J(x,t)$ will be zero at the barrier if $\phi'(x)$ is a continuous function at $x = x_1$. This implies that the potential barrier must be located at the zeroes of the Kummer function.

If, instead of (3.4), we write

$$\phi(x) = \begin{cases} U(l|\tfrac{1}{2}|ax^2), & x \geqslant x_1, \\ \alpha, & x < x_1, \end{cases}$$

where $\alpha \ll 1$ (our potential is not completely hard core), we have

$$P(x,t/0) \cong 0 \quad \text{for } x \leqslant x_1 \tag{3.20}$$

[see Eq. (3.5)] and the barrier may be located anywhere.

In Fig. 2 we have a representation of the potential (3.3) in the case where $l = -0.5$.

## IV. STATIONARY DISTRIBUTIONS

As is well known a one-dimensional FPE

$$\dot{P}(x,t) = -\frac{\partial}{\partial x}[f(x)P(x,t)] + \frac{1}{2}\frac{\partial^2 P(x,t)}{\partial x^2} \tag{4.1}$$

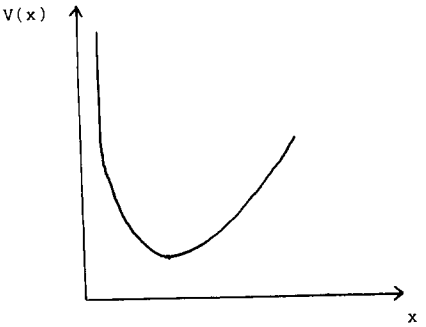



FIG. 2. Representation of the potential (3.3) in the case $l = -0.5$. The potential barrier is located at $x_1 = 0.25\, a^{1/2}$, and the minimum at $x_0 = a^{-1/2}$.

has a stationary distribution of the form

$$P_{St}(x) = N \exp\left\{2 \int^x f(x')dx'\right\}, \tag{4.2}$$

when the probability current

$$J(x,t) = f(x)P(x,t) - \frac{1}{2}\frac{\partial P(x,t)}{\partial x} \tag{4.3}$$

satisfies the boundary condition[10]

$$\lim_{x \to \pm \infty} J(x,t) = 0. \tag{4.4}$$

Both our general model represented by Eq. (1.4) as well as the models represented by Eqs. (2.1) and (3.2) satisfy these boundary conditions. Their stationary solution is

$$P_{st}(x) = N[\phi_i(x)]^2 e^{-ax^2}, \tag{4.5}$$

where

$$\phi_1(x) = \alpha F(l|\tfrac{1}{2}|ax^2) + \beta x F(l+\tfrac{1}{2}|\tfrac{3}{2}|ax^2), \tag{4.6a}$$

$$\phi_2(x) = e^{(a|2)x^2} D_{-2l}(\sqrt{2a}x), \tag{4.6b}$$

$$\phi_3(x) = U(l|\tfrac{1}{2}|ax^2), \tag{4.6c}$$

and

$$N = \frac{a^{1/2}}{\pi}\left[\frac{\alpha^2 l\Gamma(-l)}{\Gamma(\frac{1}{2}-l)} + \frac{\beta^2(l+\frac{1}{2})\Gamma(-l-\frac{1}{2})}{4a\Gamma(1-l)}\right]^{-1} \tag{4.7}$$

[provided that the proper choice of the constants $\alpha$ and $\beta$ extends this normalization to the models (4.6b) and (4.6c)].

Let us study the stability of these stationary distributions. Following the criterion given in Ref. 4 we can affirm that the stochastic process represented by Eq. (4.1) has a stable stationary solution, and all moments $\langle x^m \rangle$ up to the $m$th order exist if the following inequality is satisfied:

$$L \equiv \lim_{x \to \infty} \frac{-2\int^x f(x')dx'}{(m+1)\ln x} > 1. \tag{4.8}$$

In our case we have [see Eq. (1.3) and Ref. 9]

$$L_1 = -\infty, \quad L_2 = L_3 = +\infty, \tag{4.9}$$

whatever the values of $m$ and $l$. We see therefore that the general model is completely unstable (let us remember that in this model, when $\beta = 0$, the model presented in Ref. 6 is included). As a matter of fact, both the general model (4.6a) as well as the model of Ref. 6 do not behave correctly at infinity since $P_{st}(x) \to \infty$ when $x \to \pm\infty$.

Thus, we can affirm that the models presented in Secs. II and III are stable for any value of the parameter $l$.

## V. CONCLUSIONS

Relating the results of Sec. II with those of Sec. III, we observe that the nonlinear diffusion process, represented in general form by the drift (1.3), yields divergent momenta (and infinite relaxation times) if the diffusion process can be extended to the whole physical space. Nevertheless when, due to the introduction of a potential barrier, the diffusion process takes place in a limited part of space, the moments converge with finite relaxation time given by

$$\tau_{\text{nonlinear}} = 1/2la.$$

Comparing this relaxation time with the one that corresponds to a linear drift, $f^{\mu}(q) = aq^{\mu}$, that is,

$$\tau_{\text{linear}} = 1/a,$$

we observe that this process of nonlinear diffusion relaxes quicker than the linear diffusion if

$$l > \tfrac{1}{2}.$$

Let us remark also that the nondivergent model studied in this paper can reach large parts of space since the asymptotic evolution of the process is the same no matter how far away the potential barrier is from the initial state.

The general model represented by Eq. (1.4) is unstable since its stationary distribution $P_{st}(x)$ diverges. The models studied in Secs. II and III are stable.

## APPENDIX: EVALUATION OF INTEGRALS

We have to evaluate the integral

$$I^{(m)} = \int_{x_1}^{\infty} x^m P\left(x, \frac{t}{0}\right)dx \quad (m = 1,2,\ldots). \tag{A1}$$

By means of the change of variables $z = ax^2$, we get

$$I^{(m)} = K(t)[I_1^{(m)} + I_2^{(m)}], \tag{A2}$$

with

$$K(t) \equiv \frac{\Gamma(l+\frac{1}{2})}{2\pi a^{m/2}} e^{-2lat}(\eta(t))^{-1/2}, \tag{A3}$$

$$I_1^{(m)} \equiv \int_{z_1}^{0} z^{(m-1)/2} e^{-z/\eta(t)} U(l|\tfrac{1}{2}|z)dz, \tag{A4}$$

$$I_2^{(m)} \equiv \int_{0}^{\infty} z^{(m-1)/2} e^{-z/\eta(t)} U(l|\tfrac{1}{2}|z)dz. \tag{A5}$$

The evaluation of $I_2^{(m)}$ is immediate taking into account that[9]

$$\int_0^{\infty} e^{-sz} z^{k-1} U(a|c|z)dz = \frac{\Gamma(b)\Gamma(1+b-c)}{\Gamma(1+a+b-c)} s^{-b} \times F\left(a,b|1+a+b-c|1-\frac{1}{s}\right),$$

where $\mathrm{Re}\, s > 1/2$ and $F(a,b|c|z)$ is the hypergeometric function.

The final result is

$$I_2^{(m)} = \frac{\Gamma((1+m/2)/2)\Gamma(1+m/2)}{\Gamma(1+l+m/2)}(\eta(t))^{(1+m)/2} \times F(l,(1+m)/2|1+l+m/2|1-\eta(t)), \tag{A6}$$

valid for $m = 1,2,3,\ldots$.

To evaluate $I_1^{(m)}$ we perform the change of variable $y = z/\eta(t)$. Using the multiplication theorem[9]

$$U(a|c|zz') = (z')^{-a} \sum_{n=0}^{\infty} \frac{(a)_n (1+a-c)_n}{n!} \times \left(1-\frac{1}{z'}\right)^n U(a+n|c|z),$$

we have

$$I_1^{(m)} = (\eta(t))^{(1+m)/2-l} \sum_{n=0}^{\infty} \frac{(l)_n (l+\frac{1}{2})_n}{n!} \left(1-\frac{1}{\eta(t)}\right)^n \times \int_{y_1}^{0} y^{(m-1)/2} e^{-y} U(l+n|\tfrac{1}{2}|y) dy. \tag{A7}$$

For $m = 1$, and recalling that[9]

$$\int e^{-z} U(a|c|z) dz = -e^{-z} U(a|c-1|z) + C, \tag{A8}$$

the expression (A7) becomes

$$I_1^{(1)} = (\eta(t))^{1-l} \cdot \left\{ e^{-ax_1^2/\eta(t)} \left[ \sum_{n=0}^{\infty} \frac{(l)_n (l+\frac{1}{2})_n}{n!} \left(1-\frac{1}{\eta(t)}\right)^n \times U\left(l+n \,\middle|\, -\frac{1}{2} \,\middle|\, \frac{ax_1^2}{\eta(t)}\right)\right] - \frac{\Gamma(\frac{3}{2})}{\Gamma(l+\frac{3}{2})} F\left(l, l+\frac{1}{2} \,\middle|\, 1+\frac{3}{2} \,\middle|\, 1-\frac{1}{\eta(t)}\right)\right\}, \tag{A9}$$

since

$$U(l+n|-\tfrac{1}{2}|0) = \frac{\Gamma(\frac{3}{2})}{\Gamma(l+n+\frac{3}{2})},$$

$$\Gamma(l+n+\tfrac{3}{2}) = \Gamma(l+\tfrac{3}{2})(l+\tfrac{3}{2})_n,$$

$$F(a,b\,|c|z) \equiv \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!}.$$

For $m = 2$, integrating by parts and taking into account (A8), Eq. (A7) becomes

$$I_1^{(2)} = (\eta(t))^{(3/2)-l} \left\{ e^{-ax_1^2/\eta(t)} \left[ \sum_{n=0}^{\infty} \frac{(l)_n (l+\frac{1}{2})_n}{n!} \left(1-\frac{1}{\eta(t)}\right)^n \right] \times \left[ \frac{ax_1^2}{\eta(t)} U\left(l+n+\frac{1}{2} \,\middle|\, \frac{1}{2} \,\middle|\, \frac{ax_1^2}{\eta(t)}\right) + U\left(l+n+\frac{1}{2} \,\middle|\, -\frac{1}{2} \,\middle|\, \frac{ax_1^2}{\eta(t)}\right)\right] - \frac{\Gamma(\frac{3}{2})}{\Gamma(l+2)} F\left(l, l+\tfrac{1}{2} \,\middle|\, l+2 \,\middle|\, 1-\frac{1}{\eta(t)}\right)\right\}. \tag{A10}$$

In general,

$$I_1^{(m)} = (\eta(t))^{(1+m)/2-l} \left\{ e^{-ax_1^2/\eta(t)} \times \left[ \sum_{n=0}^{\infty} \psi_n^{(m)}(l, x_1; \eta(t)) \right] - \frac{\Gamma(\frac{3}{2})}{\Gamma(1+l+m/2)} \times F\left(l, l+\frac{1}{2} \,\middle|\, 1+l+\frac{m}{2} \,\middle|\, 1-\frac{1}{\eta(t)}\right), \right. \tag{A11}$$

with $\psi_n^{(m)}(l, x_1; \eta(t))$ given by (3.8). Expression (A11) is only valid for $m = 1,2$.

Substituting (A11), (A6), and (A3) in Eq. (A1) and with the help of Ref. 9,

$$F(a,b\,|c|z) = (1-z)^{-a} F(a, c-b\,|c|z/(z-1));$$

in this way we obtain Eq. (3.7).


[1] H. Haken, Rev. Mod Phys. **47**, 67 (1975).
[2] N. G. Van Kampen, J. Stat. Phys. **17**, 71 (1977).
[3] M. Mörsch, H. Risken, and H. D. Vollmer, Z. Phys. B **32**, 245 (1979).
[4] A. Schenzle and H. Brand, Phys. Rev. A **20**, 1628 (1979).
[5] M. San Miguel, Z. Phys. B **33**, 307 (1979).
[6] M. O. Hongler, Physica D **2**, 353 (1981).
[7] L. Garrido and J. Masoliver, J. Math. Phys. **23**, 1145 (1982).
[8] J. Masoliver and L. Garrido, Z. Phys. B **47**, 243 (1982).
[9] W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer, Berlin, 1966).
[10] R. L. Stratonovich, *Topics in the Theory of Random Noise* (Gordon and Breach, New York, 1963).

# Chiral fermions in non-Riemannian Kaluza–Klein theory

J. W. Moffat
*Department of Physics, University of Toronto, Toronto, Ontario, Canada M5S 1A7*

(Received 23 July 1984; accepted for publication 31 August 1984)

The problem of obtaining left–right asymmetry of fermion quantum numbers in Kaluza–Klein theories is studied. A non-Riemannian Kaluza–Klein geometry, based on a nonsymmetric fundamental tensor $g_{\Sigma\Lambda}$, is shown to lead to zero modes of the Dirac operator and can possess a nonvanishing chirality index that permits a realistic description of fermions. Some specific models of unification and their group structures are considered.

## I. INTRODUCTION

A higher-dimensional gravity theory can give a unified description of four-dimensional gravity and gauge interactions, if all spacelike dimensions except three are compactified into an internal space with a very small characteristic length of the order of the Planck length. By integrating the action over the internal space the isometries of the compact internal space reduce to gauge symmetries. It is assumed that the ground state has the form $M^4 \times G$, where $M^4$ is the four-dimensional Minkowski space and $G$ is a compact space. The continuous symmetries of $G$ will be the observed gauge symmetries in the four-dimensional world, which should minimally contain the symmetries of $SU(3) \times SU(2) \times U(1)$. This program has been pursued by many authors.[1–3] However, a major difficulty in constructing a realistic model has been pointed out by Witten.[4] By starting with a Riemannian geometry in $4+n$ dimensions that is coupled to spinors and assuming that the internal space is compact, one always ends up in the four-dimensional theory with fermions that belong to vectorlike representations of the gauge group, i.e., such theories do not have the left–right fermion asymmetry observed in nature.

Another problem is that by beginning with a purely Riemannian structure, the Dirac equation cannot have zero mass modes for compact internal spaces. It has been suggested that by including torsion in the space, the problem of not having zero mass solutions to the Dirac equation can be circumvented.[5] However, this does not resolve the problem of left–right fermion asymmetry.

One way out of the problem of obtaining light fermions in realistic chiral representations of a low-energy gauge group is to add extra gauge fields.[4,6] However, it is necessary then to postulate the existence of topologically nontrivial vacuum configurations. If such a program is pursued, then the idea of deriving a unified theory from higher-dimensional gravity is lost, together with much of the appealing simplicity and unity of the theory.

If we retain general coordinate invariance in $4+n$ dimensions as the only local symmetry of the theory, we must either give up the compact nature of the internal space or abandon the notion of purely Riemannian geometry. Giving up the compactness of the internal space complicates considerably the structure of the theory. In the following we shall consider a non-Riemannian extension of Kaluza–Klein theory suggested some time ago,[7–9] based on a nonsymmetric field structure.[10,11] A theory of gravitation in four-dimensional space-time formulated in terms of a nonsymmetric $g_{\mu\nu}$ has been extensively investigated.[12–17] This theory is consistent with all experimental relativity tests in the solar system. Recent new observational results for the eclipsing binary system DI Herculis[18] may in fact provide confirmation of the nonsymmetric gravitational theory (NGT).[19]

## II. GEOMETRICAL STRUCTURE OF THE GENERALIZED THEORY

We adopt the notation that in a $4+n$ dimensional space, capital Greek letters $(\Sigma,\Lambda,\Pi,\ldots)$ denote the base manifold (curved) indices and capital Latin letters $(A,B,C,\ldots)$ denote the flat (tangent space) indices. For a four-dimensional space-time lowercase late Greek letters $(\mu,\nu,\lambda,\ldots)$ will be used for curved space indices, while the late Latin letters $(m,n,p,\ldots)$ denote flat indices. For internal indices taking $n$ values, early lowercase Greek letters are used in the curved case $(\alpha,\beta,\gamma,\ldots)$ and early lowercase Latin letters in the flat case $(a,b,c\ldots)$. The general coordinates $x^\Sigma$ thus consist of space-time coordinates $x^\mu$ and the internal coordinates $y^\alpha$. The internal dimensions are all spacelike.

The four-dimensional form of NGT follows from an algebraic reduction of an eight-dimensional real tangent space to a hypercomplex structure generated by a complex operator $J$ with $J^2 = 1$ (see Refs. 16 and 17). By imposing the hypercomplex structure $J$ on $R^8$ and demanding that $\nabla J = 0$, the group $GL(8,R)$ is reduced to $GL(4,R) \times GL(4,R)$, which in turn is reduced to $GL(4,R)$ in four-dimensional space-time upon introducing a metrically compatible connection in the space. From this follows that in four-dimensional space-time, the fundamental tensor $g_{\mu\nu}$ has the sesquilinear form

$$g_{\mu\nu} = g_{(\mu\nu)} + \epsilon g_{[\mu\nu]}, \tag{1}$$

where $g_{(\mu\nu)}$ denotes the symmetric part and $g_{[\mu\nu]}$ the skew-symmetric part of $g_{\mu\nu}$. Moreover, $g_{\mu\nu}$ is (hypercomplex) Hermitian $\bar{g}_{\mu\nu} = g_{\nu\mu}$ and $\epsilon^2 = 1$. In terms of hypercomplex *vielbeins* $e^m_\mu = e^m_{R_\mu} + \epsilon e^m_{I_\mu}$, we have

$$g_{\mu\nu} = e^m_\mu \bar{e}^n_\nu \eta_{mn}. \tag{2}$$

For this version of NGT, it has been proved that the physical sector of the theory does not possess any ghost poles.[20]

We can generalize the method of algebraic reduction to $2(4+n)$-dimensional space. Let us begin with a real $2(4+n)$-dimensional tangent space with the group $GL(p,R)$, where $p = 2(4+n)$. The (hyper) complex structure $J$ with

 0022-2488/85/030528-04$02.50 © 1985 American Institute of Physics 

$J^2 = 1$ is imposed on this space and we require that $\nabla J = 0$, with the consequence that GL($p$,$R$) is reduced to the semisimple product GL($N$,$R$)$\times$GL($N$,$R$) with $N = 4 + n$. We then demand that the connection on the $p$-dimensional space be metrically compatible, which reduces GL($N$,$R$)$\times$GL($N$,$R$) to GL($N$,$R$). The symmetric metric of the initial (real) $p$-dimensional space is reduced to a nonsymmetric metric in $N$-dimensional space

$$g_{\Sigma\Lambda} = g_{(\Sigma\Lambda)} + \epsilon g_{[\Sigma\Lambda]}, \tag{3}$$

where as before $\tilde{\epsilon} = -\epsilon$ and $\epsilon^2 = 1$. In $4 + n$ dimensions the fundamental tensor $g_{\Sigma\Lambda}$ has the form

$$g_{\Sigma\Lambda} = e^A_\Sigma \tilde{e}^B_\Lambda \eta_{AB}, \tag{4}$$

where the (hyper) complex *vielbeins* $e^A_\Sigma$ obey

$$e^\Sigma_C e^B_\Sigma = \delta^B_C, \quad e^A_\Sigma e^\Pi_A = \delta^\Pi_\Sigma \tag{5}$$

and they satisfy the equation ($e^A_{\Lambda,\Sigma} = \partial_\Sigma e^A_\Lambda$)

$$e^A_{\Lambda,\Sigma} + (\omega_\Sigma)^A_C e^C_\Lambda - W^A_{\Sigma\Lambda} e^A_\Delta = 0. \tag{6}$$

The $\omega$ and $W$ are the spin and the generalized Lorentz connections, respectively. In NGT[12,13] and Kaluza–Klein NGT the torsion is determined by the metric compatibility equation up to an auxiliary vector field. We can solve for $W$ in terms of $e$ and $\omega$:

$$W_{\Sigma\Lambda\Delta} \equiv W^\Pi_{\Sigma\Lambda} g_{\Pi\Delta} = \eta_{AB} (D_\Sigma e^A_\Lambda) e^B_\Delta, \tag{7}$$

where $D_\Sigma$ is the covariant derivative operator, defined by

$$D_\Sigma e^A_\Lambda = e^A_{\Lambda,\Sigma} + (\omega_\Sigma)^A_C e^C_\Lambda. \tag{8}$$

A group of isometries is defined by the transformation

$$e^A_\Sigma = e'^B_\Sigma (U)^A_B, \tag{9}$$

where $U$ is an element of GL($N$,$R$) that leaves the fundamental form (3) invariant. Moreover, $W$ will remain invariant under the transformation (9) provided that

$$(W_\Sigma)^A_B \rightarrow [UW_\Sigma U^{-1} - (\partial_\Sigma U) U^{-1}]^A_B. \tag{10}$$

A curvative tensor can be defined by

$$([D_\Sigma, D_\Lambda])^A_B = (R_{\Sigma\Lambda})^A_B, \tag{11}$$

where

$$(R_{\Sigma\Lambda})^A_B = (\omega_\Lambda)^A_{B,\Sigma} - (\omega_\Sigma)^A_{B,\Lambda} + ([\omega_\Sigma, \omega_\Lambda])^A_B. \tag{12}$$

This equation is invariant under the transformation (9). The scalar curvature in $4 + n$ dimensions is

$$R = \eta^{AC} e^\Lambda_A e^\Sigma_B (R_{\Sigma\Lambda})^B_C. \tag{13}$$

The action of the theory is

$$S = -\frac{1}{4\kappa^2} \int d^4x \int \frac{d^n y}{V(n)} eR, \tag{14}$$

where $e = \det(e^A_\Sigma)$ and $\kappa^2/4\pi = G$ is the Newtonian constant in four dimensions, whereby the invariant volume $V(n)$ of the internal space has dimensions (length)$^n$.

## III. HIGHER-DIMENSIONAL SPINORS AND THE CHIRALITY INDEX FOR FERMIONS

A spinor transforms according to the law

$$\psi \rightarrow \psi + \delta\psi, \quad \delta_\xi \psi = -\xi^\Sigma \partial_\Sigma \psi, \tag{15}$$

where $\xi^\Sigma(x)$ is defined by the infinitesimal transformations

$$x^\Sigma \rightarrow x^\Sigma + \xi^\Sigma(x). \tag{16}$$

The $\psi$ belongs to a fundamental spinor representation of the universal covering group spin($N - 1$,1) of the Lorentz group. Under infinitesimal Lorentz transformations with the coefficients $a_{AB}(x) = -a_{BA}(x)$, a Dirac spinor is a $2^{(N/2)}$ component representation that transforms under infinitesimal Lorentz transformations according to

$$\delta_L \psi = -\tfrac{1}{2} a_{AB} \Sigma^{[AB]} \psi, \tag{17}$$

where the $N(N-1)/2$ generators $\Sigma^{[AB]}$ can be constructed from the $N = 4 + n$ Dirac matrices $\Gamma^A$. These matrices obey the usual rules

$$\{\Gamma^A, \Gamma^B\} = 2\eta^{AB}, \quad \Sigma^{[AB]} = -\tfrac{1}{4}[\Gamma^A, \Gamma^B]. \tag{18}$$

In general, we have for a scalar quantity $\phi$

$$D_\Sigma \phi = [\partial_\Sigma - \tfrac{1}{2}(\omega_\Sigma)_{AB} \Sigma^{AB}] \phi, \tag{19}$$

where $(\omega_\Sigma)_{AB} = (\omega_\Sigma)_{(AB)} + (\omega_\Sigma)_{[AB]}$ and $\Sigma^{AB} = \Sigma^{(AB)} + \Sigma^{[AB]}$. The $\Sigma^{(AB)}$ are $N(N+1)/2$ noncompact generators of GL($N$,$R$). The Dirac covariant derivative that transforms as a (finite component) spinor under Lorentz transformation is

$$\mathscr{D}_\Sigma \psi = [\partial_\Sigma - \tfrac{1}{2}(\omega_\Sigma)_{[AB]} \Sigma^{[AB]}] \psi. \tag{20}$$

We can now construct a scalar Dirac operator

$$\mathscr{D}\psi = \Gamma^\Sigma \mathscr{D}_\Sigma \psi = g^{(\Sigma\Lambda)} E_{\Lambda A} \Gamma^A \mathscr{D}_\Sigma \psi, \tag{21}$$

where the *vielbeins* $E^A_\Sigma$ are defined by the metric tensor

$$g^{(\Sigma\Lambda)} = E^\Sigma_A E^\Lambda_B \eta^{AB}. \tag{22}$$

The physical mass spectrum of our theory is determined by the Dirac operator $\mathscr{D} = \Gamma^A \mathscr{D}_A$.

Witten has discussed the index of chirality for fermions in the context of dimensional reduction.[4] An analysis of the problem for the case of non-Riemannian geometry has been given by Wetterich.[21] Let us assume that

$$[F_z, \mathscr{D}] = 0, \quad \{\Gamma, \mathscr{D}\} = 0, \tag{23}$$

where $F$ denotes the gauge transformation and $\Gamma$ the Dirac matrices. The index number is defined by[4]

$$N_C(\mathscr{D}) = n_c^+ - n_c^- - n_{\bar{c}}^+ + n_{\bar{c}}^-, \tag{24}$$

where $n_c^+$ is the number of zero modes of $\mathscr{D}$ in the Weyl spinor $\psi^+$ associated with a complex representation of the spinor, while $n_c^-$ and $n_{\bar{c}}^+$ denote the corresponding values for $\psi^-$ and for the complex conjugate representation. Thus $N_c$ is the number of four-dimensional left-handed fermion generations, up to a numerical factor.

In the case that the *vielbein* $e^A_\Sigma$ has an inverse everywhere, the operator $\mathscr{D}$ is an elliptic operator, which for compact spaces has the property that $N_c(\mathscr{D})$ remains invariant under continuous changes of the metric. In compact spaces we have for elliptic operators $\mathscr{D}$

$$N_c(\mathscr{D}) = N_c(\mathscr{D} + sB), \tag{25}$$

for some arbitrary parameter $s$, provided that the nonderivative operator $B$ obeys

$$[F_z, B] = 0, \quad \{\Gamma, B\} = 0. \tag{26}$$

For Riemannian geometry, the chirality index vanishes for the Dirac operator for arbitrary compact spaces. For theories with torsion $W^A_{[\Sigma\Lambda]} \neq 0$ the Dirac operator also has zero index because of Eq. (25). This follows because $B$ is a nonder-

ivative operator in the Dirac operator $\Gamma^A D_A + B$ that corresponds to nonvanishing torsion. Thus, although we could have zero modes in the presence of torsion, the chirality index vanishes and we cannot have left- and right-handed chiral asymmetry. Only in the case that $g^{(\Sigma\Lambda)}$ or, equivalently, the *vielbein* $E^A_\Sigma$ fail to have inverses everywhere is it possible for chiral fermions to exist, i.e., to have a nonvanishing chirality index.[21]

## IV. NONELLIPTIC DIRAC OPERATORS IN HIGHER-DIMENSIONAL NGT

Let us write $g_{\Sigma\Lambda}$ and $g^{\Sigma\Lambda}$ in terms of their symmetric parts $h_{\Sigma\Lambda}$ and $p^{\Sigma\Lambda}$ and their skew-symmetric parts $f_{\Sigma\Lambda}$ and $k^{\Sigma\Lambda}$, repsectively,

$$g_{\Sigma\Lambda} = h_{\Sigma\Lambda} + f_{\Sigma\Lambda}, \quad g^{\Sigma\Lambda} = p^{\Sigma\Lambda} + k^{\Sigma\Lambda}. \tag{27}$$

If we assume that $h = \det(h_{\Sigma\Lambda}) \neq 0$, then the metric tensor $h_{\Sigma\Lambda} = h_{\Lambda\Sigma}$ always has an inverse

$$h_{\Sigma\Lambda} h^{\Sigma\Delta} = \delta^\Delta_\Lambda. \tag{28}$$

The following relations can be derived:

$$g = \det(g_{\Sigma\Lambda}) = h + f + (h/2) h^{\Sigma\Lambda} h^{\Delta\Omega} f_{\Sigma\Delta} f_{\Lambda\Omega}, \tag{29}$$

$$g^{-1} = \det(g^{\Sigma\Lambda}) = p + k + (p/2) p_{\Sigma\Lambda} p_{\Delta\Omega} k^{\Sigma\Delta} k^{\Lambda\Omega},$$

and

$$g^2 = h/p = f/k, \tag{30}$$

where $p^{\Sigma\Delta} p_{\Sigma\Lambda} = \delta^\Delta_\Lambda$. Our Dirac operator is

$$\mathscr{D} = p^{\Sigma\Lambda} \Gamma_\Lambda \mathscr{D}_\Sigma = E^\Sigma_A \Gamma^A \mathscr{D}_\Sigma, \tag{31}$$

where the $E^\Sigma_A$ are defined by Eq. (22). This definition of the Dirac operator is not unique in the nonsymmetric theory.

Thus, the physical spinors couple through $p^{\Sigma\Lambda} = p^{\Lambda\Sigma}$. If we assume that $p = \det(p^{\Sigma\Lambda})$ vanishes somewhere in the compact space, then $E^\Sigma_A$ is not invertible everywhere. Thus $\mathscr{D}$ is not an elliptic operator, which is a necessary condition for the existence of chiral fermions: i.e., we can have $N_C(\mathscr{D}) \neq 0$.[21] In the Riemannian case, the metric is just the square of the *vielbein* $e^A_\Sigma$

$$g_{\Sigma\Lambda} = g_{\Lambda\Sigma} = e^A_\Sigma e_{\Lambda A}, \tag{32}$$

and $g$ should be nonvanishing everywhere in the compact space.

Thus in NGT we can preserve $h \neq 0$ and have $p = 0$ at some point in the space such that $\mathscr{D}$ is not an elliptic operator and therefore can possess a nonvanishing chirality index. Since $h \neq 0$ the *metric* properties of the space, defined by $h_{\Sigma\Lambda}$ and $h^{\Sigma\Lambda}$, are preserved in the manifold.

## V. MODELS OF UNIFIED THEORIES BASED ON NONSYMMETRIC KALUZA–KLEIN THEORY

Let us consider examples of a unified theory using our generalized Kaluza–Klein scheme. We choose $p = 28$ with a real tangent space based on the group of transformations GL(28,$R$). By dimensional reduction this is reduced to GL(14,$R$)$\times$GL(14,$R$) and GL(14,$R$). The basic group in $N = 14$-dimensional space is SL(14,$R$) with a 10-dimensional compact space. The symmetry breaking that occurs under compactification could then lead to the scheme

$$\begin{aligned} \mathrm{SL}(14,R) &\to \mathrm{SO}(13,1) \to \mathrm{SO}(10)\times\mathrm{SO}(3,1) \\ &\to \mathrm{SU}(3)\times\mathrm{SU}(2)\times\mathrm{U}(1)\times\mathrm{SO}(3,1) \\ &\to \mathrm{SU}(2)\times\mathrm{U}(1)\times\mathrm{SO}(3,1). \end{aligned} \tag{33}$$

Thus in four-dimensional space-time the compact group SO(10) would describe the grand unified scheme for particle physics. The problem of left- and right-handed fermion asymmetry is avoided since the fermions couple to the nonsymmetric field structure. We choose even $N$ and avoid the case of a single $\frac{1}{2}$ representation of the spin group with $\gamma_5 = +1$ and $\gamma_5 = -1$ parts with the same transformations under SO($N-1$,1).

It is possible to go to more complicated schemes such as one based on quaternion[17,22] and octonion[17] algebraic reductions of the higher-dimensional tangent space. In the case of the quaternion reduction of the theory, the relevant group is GL($p$,$C$), which contains the compact unitary group SU($q$) as well as the group of homogeneous Lorentz transformations SO(3,1). We could envisage the schemes in $p = 24$-dimensional space:

$$\mathrm{SL}(24,C) \to \mathrm{SL}(12,C)\times\mathrm{SL}(12,C) \to \mathrm{SL}(12,C) \tag{34}$$

and

$$\begin{aligned} \mathrm{SL}(12,C) &\to \mathrm{SL}(8,C)\times\mathrm{SL}(4,C)\times\mathrm{GL}(1,C) \\ &\to \mathrm{SU}(8)\times\mathrm{SO}(4,C)\times\mathrm{GL}(1,C) \\ &\to \mathrm{SU}(8)\times\mathrm{SO}(3,1)\times\mathrm{GL}(1,C) \\ &\to \mathrm{SU}(3)\times\mathrm{SU}(2)\times\mathrm{U}(1)\times\mathrm{SO}(3,1)\times\mathrm{GL}(1,C) \\ &\to \mathrm{SU}(2)\times\mathrm{U}(1)\times\mathrm{SO}(3,1)\times\mathrm{GL}(1,C). \end{aligned} \tag{35}$$

Now the relevant grand unified theory is SU(8).

Further work has to be done to obtain detailed predictions of the unified theories of the kind displayed in the above models. In particular, the significance of spontaneous compactification in our non-Riemannian Kaluza–Klein scheme must be studied in detail.

## VI. CONCLUSIONS

We have shown that a non-Riemannian Kaluza–Klein theory, based on a hypercomplex, nonsymmetric $g_{\Sigma\Lambda}$, can have a nonvanishing chirality index, leading to nonequal left- and right-handed zero mass modes of the generalized Dirac operator in the compact space, because the *vielbein* $E^A_\Sigma$ in the Dirac operator $\mathscr{D}$ may no longer have an inverse everywhere. However, the symmetric part of $g_{\Sigma\Lambda}$, which constitutes the metric tensor of the theory, still has an inverse everywhere, thereby retaining the purely geometrical properties of the space. Thus the line element defined in $4+n$-dimensional space by

$$ds^2 = g_{(\Sigma\Lambda)} dx^\Sigma dx^\Lambda \tag{36}$$

is still well-defined in the theory, even though $\det(g^{(\Sigma\Lambda)})$ vanishes somewhere in the manifold, rendering the generalized Dirac operator a nonelliptic operator that permits a nonvanishing chirality index. In this way we realize that the nonsymmetric field structure is related to the existence of physically realizable fermion field operators.

## ACKNOWLEDGMENTS

I thank Dr. G. Kunstatter, Dr. M. Green and Dr. R. B. Mann for helpful discussions.

This work was supported by the Natural Science and Engineering Research Council of Canada.

[1]Th. Kaluza, Sitzungsber. Preuss. Akad. Wiss. Berlin, Math. Phys. Kl., 966 (1921); O. Klein, Z. Phys. **37**, 895 (1926); Ark. Mat. Astron. Fys. B. **34** (1946); Helv. Phys. Acta Suppl. IV, 58 (1956); A. Einstein and P. Bergmann, Ann. Math. **39**, 683 (1938); B. De Witt, in *Relativity, Groups, and Topology*, edited by B. De Witt and C. De Witt (Gordon and Breach, New York, 1964), and *Dynamical Theory of Groups and Fields* (Gordon and Breach, New York, 1965); J. Rayski, Acta Phys. Pol. **27**, 89 (1965); R. Kerner, Ann. Inst. H. Poincaré **9**, 143 (1968); A. Trautman, Rep. Math. Phys. **1**, 29 (1970); Y. M. Cho, J. Math. Phys. **16**, 2029 (1975); Y. M. Cho and P. G. O. Freund, Phys. Rev. D **12**, 1711 (1975); L. N. Chang, K. I. Macrae, and F. Mansouri, Phys. Rev. D **13**, 235 (1976); J. Scherk and R. R. Schwarz, Phys. Lett. B **57**, 463 (1975); E. Cremmer and J. Scherk, Nucl. Phys. B **103**, 393 (1976); E. Cremmer and B. Julia, Nucl. Phys. B **159**, 141 (1979).

[2]E. Witten, Nucl. Phys. B **186**, 412 (1981).

[3]S. Weinberg, Phys. Lett. B **138**, 47 (1984).

[4]E. Witten, Princeton University Preprint, 1983.

[5]Y. S. Wu and Z. Zee, University of Washington Preprint 400 48-25P3, 1983; C. Detri, C. Orzalezi, and P. Rossi, Ann. Phys. **147**, 321 (1983).

[6]S. Randjbar-Daemi, A. Salam, and J. Strathdee, Nucl. Phys. B **214**, 491 (1983).

[7]J. W. Moffat, Acta Phys. Austriaca, Suppl. **23**, 665 (1981).

[8]J. W. Moffat, Proceedings of the Third Annual Montreal–Rochester–Syracuse–Toronto Symposium, University of Rochester, 30 April–1 May, 1981.

[9]J. W. Moffat, Ann. Inst. H. Poincaré **34**, 85 (1981).

[10]M. W. Kalinowski, J. Math. Phys. **24**, 1835 (1983).

[11]G. Kunstatter, J. Math. Phys. **25**, 2691 (1984).

[12]J. W. Moffat, Phys. Rev. D **19**, 3554 (1979).

[13]J. W. Moffat, J. Math. Phys. **21**, 1798 (1980).

[14]J. W. Moffat, Phys. Rev. Lett. **50**, 709 (1983).

[15]R. B. Mann and J. W. Moffat, University of Toronto preprint, 1984.

[16]G. Kunstatter, J. W. Moffat, and J. Malzan, J. Math. Phys. **24**, 886 (1983).

[17]J. W. Moffat, J. Math. Phys. **25**, 347 (1984), **26**, 214 (1985) (erratum).

[18]E. F. Guinan and F. P. Maloney, University of Villanova, Villanova, PA, Preprint, 1984.

[19]J. W. Moffat, Astrophys. J. Lett. (to be published).

[20]R. B. Mann and J. W. Moffat, Phys. Rev. D **26**, 1858 (1982).

[21]C. Wetterich, Nucl. Phys. B **242**, 473 (1984).

[22]R. B. Mann, Nucl. Phys. B **231**, 481 (1984).

# Some static solutions in the general scalar-tensor theory

T. Singh and Tarkeshwar Singh
*Applied Mathematics Section, Institute of Technology, Banaras Hindu University, Varanasi-221 005, India*

(Received 25 May 1984; accepted for publication 3 August 1984)

Exact solutions are found for the static gravitational fields for a matter-free space in the general scalar-tensor theory of Nordtvedt–Barker–Schwinger. The singular behavior of some of the invariants has also been discussed for the solutions in Barker and Schwinger theories.

## I. INTRODUCTION

Static perfect fluid distributions with high symmetries such as spherical, cylindrical, and planar symmetries in general relativity are widely discussed in the literature.[1–6] However, the corresponding problem in scalar-tensor theory still lacks thorough investigation. Recently increasing interest in general scalar-tensor theories seems to stem mainly from attempts to extend the principle of conformal invariance to also include gravitational phenomena (gravity becoming itself the manifestation of a broken symmetry[7,8]). Also Kodama[9] has shown that the use of scalar fields may enable one to construct nonsingular field-theoretical models for elementary particles. Recently Banerjee and Santos[10] have discussed static perfect fluid in Brans–Dicke theory.[11]

In the present paper we have considered gravitational field equations in the general scalar-tensor theory of Nordtvedt[12] in a static nonrotating space-time with two mutually orthogonal spacelike Killing vectors. This metric may be interpreted to represent cylindrical, toroidal, planar, or pseudoplanar symmetry depending on the behavior of the space coordinates.

The scalar-tensor theory proposed by Nordtvedt considers the parameter $\omega$ as a function of the scalar field $\phi$ and, in the special case $\omega = \text{const}$, it reduces to the theory of Brans–Dicke.[11]

Our paper is organized as follows. In Sec. II we have written the field equations and their general solutions. In Sec. III and Sec. IV we have found exact solutions in the Nordtvedt's scalar-tensor theory of gravitation with scalar field $\phi$ being obtained from specific choices of $\omega$ as a function of $\phi$ as proposed by Barker[13] and Schwinger.[8] In Sec. V we study the singular behavior of some of the invariants, viz., Kretschmann curvature invariant and curvature scalar.

## II. FIELD EQUATIONS AND THEIR SOLUTIONS

We consider a static space-time possessing two spacelike Killing vectors which are mutually orthogonal and also orthogonal to the timelike Killing vector. One can choose the coordinates so that the metric has the form[5]

$$ds^2 = e^{2\gamma(x)}dt^2 - e^{2\lambda(x)}dx^2 - e^{2\mu(x)}d\eta^2 - e^{2\beta(x)}d\xi^2. \tag{2.1}$$

It corresponds to cylindrical symmetry if $\xi$ and $\eta$ represent the azimuthal and longitudinal coordinates, respectively, so that $\xi \in (0,2\pi)$ and $\eta \in (-\infty,\infty)$. If both $\xi$ and $\eta$ are angular coordinates, we call the system toroidally symmetric, whereas if both $\xi$ and $\eta$ represent longitudinal coordinates $[\eta \in (-\infty,+\infty), \xi \in (-\infty,+\infty)]$, the symmetry can be called "pseudoplanar" [to obtain the well-known planar symmetry, one should put in addition, $\beta(x) = \mu(x)$].

In the following we attempt to find exact solutions of the field equations in Nordtvedt's general scalar-tensor theory corresponding to the metric (2.1) for a matter-free space. However, one can without loss of generality, use the coordinate condition

$$\lambda = \gamma + \mu + \beta. \tag{2.2}$$

This coordinate condition enables us to write the field equations in a symmetrical form.

The field equations in the Nordtvedt's general scalar-tensor theory[12] are

$$G^i_j = -\frac{\omega}{\phi^2}\left[\phi^{,i}\phi_{,j} - \frac{1}{2}\delta^i_j\phi_{,k}\phi^{,k}\right] - \frac{1}{\phi}\left[\phi^{;i}_{;j} - \delta^i_j\Box\phi\right], \tag{2.3}$$

$$\Box\phi = -\frac{\phi_{,k}\phi^{,k}}{2\omega+3}\frac{d\omega}{d\phi}, \tag{2.4}$$

where $\phi$ is the scalar field and $\omega$ is a function of $\phi$. The field equation (2.3) can be written explicitly as

$$G^1_1 = e^{-2\lambda}U = e^{-2\lambda}\left[\frac{\omega}{2}\left(\frac{\phi'}{\phi}\right)^2 - \frac{\lambda'\phi'}{\phi}\right], \tag{2.5}$$

$$G^2_2 = e^{-2\lambda}[\beta'' + \gamma'' - U] = e^{-2\lambda}\left[-\frac{\omega}{2}\left(\frac{\phi'}{\phi}\right)^2 + \frac{\mu'\phi'}{\phi} - \frac{\phi''}{\phi}\right], \tag{2.6}$$

$$G^3_3 = e^{-2\lambda}[\gamma'' + \mu'' - U] = e^{-2\lambda}\left[-\frac{\omega}{2}\left(\frac{\phi'}{\phi}\right)^2 + \frac{\beta'\phi'}{\phi} - \frac{\phi''}{\phi}\right], \tag{2.7}$$

$$G^4_4 = e^{-2\lambda}[\beta'' + \mu'' - U] = e^{-2\lambda}\left[-\frac{\omega}{2}\left(\frac{\phi'}{\phi}\right)^2 + \frac{\gamma'\phi'}{\phi} - \frac{\phi''}{\phi}\right]. \tag{2.8}$$

Equation (2.4) for the scalar field leads to

$$\Box\phi = e^{-2\lambda}\phi'' = -\frac{e^{-2\lambda}(\phi')^2}{2\omega+3}\frac{d\omega}{d\phi}. \tag{2.9}$$

Here and in what follows, the primes indicate differentiation with respect to $x$ and

$$U = \beta'\gamma' + \beta'\mu' + \gamma'\mu'. \tag{2.10}$$

Now we attempt to solve this system of five equations and five unknown functions $\gamma$, $\mu$, $\beta$, $\omega$, and $\phi$. Adding (2.5) and (2.6) one can immediately obtain the relation

 0022-2488/85/030532-04$02.50 © 1985 American Institute of Physics 

$$\beta'' + \gamma'' = -(\phi'/\phi)(\gamma' + \beta') - \phi''/\phi. \quad (2.11)$$

Subtracting (2.8) from (2.6) and (2.7), we have, respectively,

$$\gamma'' - \mu'' = -(\phi'/\phi)(\gamma' - \mu') \quad (2.12)$$

and

$$\gamma'' - \beta'' = -(\phi'/\phi)(\gamma' - \beta'). \quad (2.13)$$

Adding (2.11) and (2.13), we have

$$\gamma'' + \gamma'\phi'/\phi + \phi''/2\phi = 0. \quad (2.14)$$

On integration Eq. (2.9) yields

$$\phi' = C(2\omega + 3)^{-1/2}, \quad (2.15)$$

$C$ being an integration constant. Once $\omega(\phi)$ is known other equations can be integrated.

## III. SOLUTIONS IN BARKER'S THEORY

Within the framework of Nordtvedt's[12] general scalar-tensor theory Barker[13] has proposed a particular $\omega - \phi$ relationship in the form

$$\omega(\phi) = (4 - 3\phi)/(2\phi - 2). \quad (3.1)$$

Use of (3.1) in (2.15) and integration gives

$$\phi = 1 + (K_1 x + K_2)^2, \quad (3.2)$$

where $K_1$ and $K_2$ are integration constants. Now from (2.14) and (3.2) we have

$$\gamma = M_2 + M_1 \tan^{-1}(K_1 x + K_2) - \tfrac{1}{2}\log[(K_1 x + K_2)^2 + 1], \quad (3.3)$$

where $M_2$ and $M_1$ are constants of integration. Now using (3.2) in (2.12) and (2.13), on integration we have, respectively,

$$\mu = \gamma - (D_4/K_1)\tan^{-1}(K_1 x + K_2) \quad (3.4)$$

and

$$\beta = \gamma - (D_3/K_1)\tan^{-1}(K_1 x + K_2), \quad (3.5)$$

where $D_3$ and $D_4$ are integration constants.

Equations (3.2)–(3.5) constitute the complete solution for the metric (2.1). In order that this solution satisfies all the field equations, we must have a relation between the constants $K_1$, $M_1$, $D_3$, and $D_4$ given by

$$3M_1^2 + D_3 D_4/K_1^2 = 1 + 2M_1(D_3 + D_4)/K_1. \quad (3.6)$$

## IV. SOLUTIONS IN SCHWINGER'S THEORY

Schwinger[8] and Milton and Yee[14] have formulated a scalar-tensor theory (as a mass-varying theory), but it can be put in the form of a standard scalar-tensor theory with a suitable choice of the function $\omega(\phi)$ and after a transformation to "particle units" has been carried out.[14]

Now we consider the $\omega - \phi$ relation as proposed by Schwinger in the form

$$2\omega(\phi) + 3 = 1/\alpha\phi, \quad (4.1)$$

where $\alpha = \text{const}$. Using relation (4.1) in (2.15) and on integration, we have

$$\phi = \alpha(Cx + C_1)^2/4, \quad (4.2)$$

where $C$ and $C_1$ are integration constants.

Now using (4.2) in Eq. (2.14) and integrating, we have

$$\gamma = K_4 - K_3/(Cx + C_1) - \log[Cx + C_1], \quad (4.3)$$

where $K_3$ and $K_4$ are integration constants. With the help of (4.2), Eq. (2.12) and (2.13) yield, respectively, the solutions

$$\mu = \gamma + 4D_2/\alpha C\,[Cx + C_1] \quad (4.4)$$

and

$$\beta = \gamma + 4D_1/\alpha C\,[Cx + C_1], \quad (4.5)$$

where $D_1$ and $D_2$ are integration constants.

Equations (4.2)–(4.5) constitute the complete solution for the metric (2.1). In order that the solution satisfies all the field equations, one must have a relation between the constants $\alpha$, $C$, $D_1$, $D_2$, and $K_3$ given by

$$3K_3^2\alpha^2/4 + 4D_1D_2/C^2 = 1 + 2K_3\alpha(D_1 + D_2)/C. \quad (4.6)$$

## V. SINGULAR BEHAVIOR OF INVARIANTS

We now study the regularity of the solutions from the behavior of some of the invariants, viz., the Kretschmann curvature invariant $\mathcal{L} = R_{hijk}R^{hijk}$ and the curvature scalar $R = g^{ij}R_{ij}$ (given in the Appendix).

In the case of Schwinger's theory the invariants $\mathcal{L}$ and $R$ tend to $\infty$ as $x \to (-C_1/C)$. Therefore, $\mathcal{L}$ and $R$ are singular at $x = -C_1/C$. Hence, there is singularity at $x = -C_1/C$ within the framework of Schwinger's theory. Similar is the situation about the singularities of the solutions within the framework of Barker's theory.

## APPENDIX A

The nonzero components of $R_{hijk}$ for metric (2.1) are

$$R_{1212} = e^{2\mu}[\mu'' + \mu'^2 - \lambda'\mu'],$$
$$R_{1313} = e^{2\beta}[\beta'' + \beta'^2 - \lambda'\beta'],$$
$$R_{1414} = e^{2\gamma}[-\gamma'' - \gamma'^2 + \lambda'\gamma'],$$
$$R_{2323} = e^{-2\gamma}[\mu'\beta'],$$
$$R_{2424} = -\mu'\gamma' e^{-2\beta},$$
$$R_{3434} = -\beta'\gamma' e^{-2\mu}.$$

The curvature invariant in terms of metric coefficients for the metric (2.1) is given by

$$R = -2e^{-2\lambda}[\lambda'' - (\mu'\beta' + \beta'\gamma' + \gamma'\mu')].$$

Thus the expression for Kretschmann scalar $\mathcal{L} = R_{hijk}R^{hijk}$ takes the form

$$\mathcal{L} = 4\{e^{-4(\lambda+\mu)}(R_{1212})^2 + e^{-4(\lambda+\beta)}(R_{1313})^2 + e^{-4(\lambda+\gamma)}(R_{1414})^2 + e^{-4(\mu+\beta)}(R_{2323})^2 + e^{-4(\mu+\gamma)}(R_{2424})^2 + e^{-4(\beta+\gamma)}(R_{3434})^2\}.$$

## APPENDIX B

The nonzero components of Riemann curvature tensor and curvature scalar for the solution in Barker's theory are

$$R_{1212} = \exp\left[2P - \frac{2D_4}{k_1}\tan^{-1}(k_1x+k_2)\right]\left\{Q + \frac{2D_4k_1(k_1x+k_2)}{[(k_1x+k_2)^2+1]^2} - \left(S - \frac{D_4}{[(k_1x+k_2)^2+1]}\right)\left(2S - \frac{D_3}{(k_1x+k_2)^2+1}\right)\right\},$$

$$R_{1313} = \exp\left[2P - \frac{2D_3}{k_1}\tan^{-1}(k_1x+k_2)\right]\left\{Q + \frac{2D_3k_1(k_1x+k_2)}{[(k_1x+k_2)^2+1]^2} - \left(S - \frac{D_3}{(k_1x+k_2)^2+1}\right)\left(2S - \frac{D_4}{(k_1x+k_2)^2+1}\right)\right\},$$

$$R_{1414} = \exp[2P]\left\{-Q + 2S^2 - S\frac{(D_3+D_4)}{(k_1x+k_2)^2+1}\right\},$$

$$R_{2323} = \exp(-2P)\left\{S^2 - \frac{S(D_3+D_4)}{(k_1x+k_2)^2+1} + \frac{D_3D_4}{[(k_1x+k_2)^2+1]^2}\right\},$$

$$R_{2424} = \exp\left[-2P + \frac{2D_3}{k_1}\tan^{-1}(k_1x+k_2)\right]\left\{-S^2 + \frac{SD_4}{(k_1x+k_2)^2+1}\right\},$$

$$R_{3434} = \exp\left[-2P + \frac{2D_4}{k_1}\tan^{-1}(k_1x+k_2)\right]\left\{-S^2 + \frac{SD_3}{(k_1x+k_2)^2+1}\right\},$$

and

$$R = \frac{8k_1^2}{[(k_1x+k_2)^2+1]^2}\exp\left\{3P - \frac{(D_3+D_4)}{k_1}\tan^{-1}(k_1x+k_2)\right\},$$

where $P$, $Q$, and $S$ are

$$P = M_2 + M_1\tan^{-1}(k_1x+k_2) - \tfrac{1}{2}\log[(k_1x+k_2)^2+1],$$

$$Q = -\frac{k_1^2}{(k_1x+k_2)^2+1} - \frac{2M_1k_1^2(k_1x+k_2)}{[(k_1x+k_2)^2+1]^2} + \frac{2K_1^2(k_1x+k_2)^2}{[(k_1x+k_2)^2+1]^2},$$

$$S = \frac{M_1k_1}{(k_1x+k_2)^2+1} - \frac{k_1(k_1x+k_2)}{(k_1x+k_2)^2+1}.$$

## APPENDIX C

The nonzero components of Riemann curvature tensor and curvature scalar for the solution in the Schwinger's theory are

$$R_{1212} = \exp\left[N + \frac{8D_2}{\alpha C[Cx+C_1]}\right]\left\{E + \frac{4k_3C(2D_2+D_1)}{\alpha[Cx+C_1]^4} - \frac{4D_1C}{\alpha[Cx+C_1]^3} - \frac{16D_1D_2}{\alpha^2[Cx+C_1]^4}\right\},$$

$$R_{1313} = \exp\left\{N + \frac{8D_1}{\alpha C[Cx+C_1]}\right\}\left\{E + \frac{4k_3C(2D_1+D_2)}{\alpha[Cx+C_1]^4} - \frac{4D_2C}{\alpha[Cx+C_1]^3} - \frac{16D_1D_2}{\alpha^2[Cx+C_1]^4}\right\},$$

$$R_{1414} = \exp(N)\left\{-E - \frac{4k_3C(D_1+D_2)}{\alpha[Cx+C_1]^4} + \frac{4C(D_1+D_2)}{\alpha[Cx+C_1]^4}\right\},$$

$$R_{2323} = \exp(-N)\left\{T - \frac{4k_3C(D_1+D_2)}{\alpha[Cx+C_1]^4} + \frac{4C(D_1+D_2)}{\alpha[Cx+C_1]^3} + \frac{16D_1D_2}{\alpha^2[Cx+C_1]^4}\right\},$$

$$R_{2424} = \exp\left\{-N - \frac{8D_1}{\alpha C[Cx+C_1]}\right\}\left\{-T + \frac{4D_2k_3C}{\alpha[Cx+C_1]^4} - \frac{4D_2C}{\alpha[Cx+C_1]^3}\right\},$$

$$R_{3434} = \exp\left\{-N - \frac{8D_2}{\alpha C[Cx+C_1]}\right\}\left\{-T + \frac{4D_1k_1C}{\alpha[Cx+C_1]^4} - \frac{4D_1C}{\alpha[Cx+C_1]^3}\right\},$$

and

$$R = \frac{8C^2}{\alpha^2[Cx+C_1]^4}\exp\left\{-3N - \frac{8(D_1+D_2)}{\alpha C[Cx+C_1]}\right\},$$

where

$$N = 2k_4 - \frac{2k_3}{Cx+C_1} - 2\log[Cx+C_1],$$

$$E=\frac{2k_3C^2}{[Cx+C_1]^3}-\frac{C^2}{[Cx+C_1]^2}-\frac{2k_3^2C^2}{[Cx+C_1]^4},$$

$$T=\frac{k_3^2C^2}{[Cx+C_1]^4}+\frac{C^2}{[Cx+C_1]^2}-\frac{2k_3C^2}{[Cx+C_1]^3}.$$


[1]K. Schwarzchild, Sitzungsber. Preuss. Akad. Wiss. Phys. Math. Kl. 424 (1916).
[2]R. C. Tolman, Phys. Rev. **55**, 364 (1939).
[3]A. Marder, Proc. R. Soc. London Ser. A **224**, 524 (1958).
[4]C. W. Misner and H. S. Zapolsky, Phys. Rev. Lett. **12**, 635 (1964).
[5]K. A. Bronnikov, J. Phys. A **12**, 201 (1979).
[6]K. D. Krori and J. Barua, Indian J. Pure Appl. Phys. **12**, 818 (1974).
[7]J. D. Bekenstein and A. Meisels, Phys. Rev. D **22**, 1313 (1980).
[8]J. Schwinger, *Particles, Sources and Fields* (Addison-Wesley, Reading, MA, 1970).
[9]T. Kodama, Phys. Rev. D **18**, 3529 (1978).
[10]A. Banerjee and N. O. Santos, Int. J. Theor. Phys. **20**, 315 (1981).
[11]C. Brans and R. H. Dicke, Phys. Rev. **124**, 925 (1961).
[12]K. Nordtvedt, Astrophys. J. **161**, 1059 (1970).
[13]B. M. Barker, Astrophys. J. **219**, 5 (1978).
[14]K. A. Milton and J. N. Yee, Phys. Rev. D **10**, 420 (1974).

# Fermionic couplings in Kaluza–Klein theories

R. Finkelstein and M. Villasante
*Department of Physics, University of California, Los Angeles, California 90024*

(Received 17 July 1984; accepted for publication 12 October 1984)

The totally symmetric and totally antisymmetric direct couplings of fermions in $d$-dimensional space are investigated. Explicit forms are given for $d \leqslant 11$. A different way of handling the $d$-dimensional Dirac algebra is also described.

## I. INTRODUCTION

Recent interest in Kaluza–Klein theories stems mostly from dimensional extensions of local supersymmetry. In particular, simple Yang–Mills in ten dimensions may be reduced to the maximal extension ($N = 4$) of supersymmetric Yang–Mills in four dimensions[1]; and simple supergravity in 11 dimensions may be reduced to the maximal extension ($N = 8$) of supergravity in four dimensions.[2] The fundamental fermionic fields in the Yang–Mills case are the Weyl–Majorana spinors appropriate to ten dimensions and in the supergravity case the corresponding fundamental fields are the Majorana vector spinors.

There is also a classical solution of the 11-dimensional supergravity field equations that splits the 11-dimensional manifold into a four-dimensional space-time and seven-dimensional internal space.[3] This unforced output of the 11-dimensional theory provides not only a realistic space-time but also a possible internal space since its dimensionality is the minimum compatible with the symmetry of $SU(3) \times SU(2) \times U(1)$.

The Kaluza–Klein idea for unifying gauge interactions with gravity is, of course, not limited to the supergravity approach. In general if the ground state is $M^4 \times B$, where $B$ is a compact space, continuous symmetries of $B$ will be observed as gauge symmetries in $M^4$ (see Ref. 4).

In all of these higher-dimensional theories there is need for efficient ways of dealing with the spinor algebra. For example, one frequently encounters Fierz transformations in the usual on-shell formulations, while the complications become greater if one attempts to construct off-shell theories. On the phenomenological side, one notes that the interaction among four fermions frequently becomes effectively zero range because of the great mass of the intermediate bosons; and the effective interactions therefore reduce to direct quadrilinear forms.

As these direct Fermi interactions appear both actually and potentially in many theoretical schemes, we have thought that it may be useful to examine these couplings in higher-dimensional theories and in a way independent of supergravity. We also describe a different way of handling the Dirac algebra.[5]

## II. NOTATION

The Dirac algebra in $d$ dimensions may be defined by the anticommutation relations

$$(\gamma_j, \gamma_k)_+ = 2\delta_{jk}, \quad j,k = 1,\ldots,d, \tag{2.1}$$

where the $\gamma_i$ are all Hermitian

$$\gamma_k^+ = \gamma_k.$$

Let us introduce the notation

$$\Gamma(\mathbf{g}) = \gamma_1^{g_1}\gamma_2^{g_2},\ldots,\gamma_d^{g_d}, \tag{2.2}$$

where

$$\mathbf{g} = (g_1, g_2, \ldots),$$

$$g_k = (0,1).$$

If $\gamma_k$ transforms like a $d$-dimensional vector under the $d$-dimensional orthogonal group, then the rank of the tensor associated with $\Gamma(\mathbf{g})$ is

$$g = \mathbf{g}^2 = \sum_1^d g_k. \tag{2.3}$$

If $\Gamma(\mathbf{g})$ is written out explicitly, one component of the rank three tensor is, for example,

$$\Gamma(\mathbf{g}) = \gamma_m \gamma_n \gamma_p,$$

where the indices are never equal and always appear in the natural order $m < n < p$. Each component of the rank $g$ tensor is associated with a different vector $\mathbf{g}$.

The $\Gamma(\mathbf{g})$ are unitary and are either Hermitian or anti-Hermitian according to the relation

$$\Gamma^+(\mathbf{g}) = \epsilon(g)\Gamma(\mathbf{g}), \tag{2.4}$$

where

$$\epsilon(g) = (-)^{(g/2)(g-1)}. \tag{2.5}$$

Let the individual $\gamma_k$ be either pure real or pure imaginary and let the number of imaginary ones be $p$. Let $C$ be the product of the imaginary ones only. Then

$$C^+ = C^{-1} = \epsilon(p)C \tag{2.6}$$

and

$$\tilde{C} = (-)^p \epsilon(p) C, \tag{2.7}$$

where $\sim$ means transpose. Then

$$C^{-1}\gamma_k C = (-)^p \tilde{\gamma}_k, \tag{2.8}$$

$$C^{-1}\Gamma(\mathbf{g})C = (-)^{pg}(-)^t \Gamma(\mathbf{g}) = (-)^{pg}\epsilon(g)\tilde{\Gamma}(g), \tag{2.9}$$

where $t$ is the number of imaginary matrices in $\Gamma(\mathbf{g})$. Let

$$E = C^{-1}\Gamma. \tag{2.10}$$

Then if $\psi$ and $\chi$ are two spinors, it follows that $\tilde{\psi}E\chi$ is a tensor of the same rank as $\Gamma$. We may also write this tensor as $\bar{\psi}\Gamma\chi$, where $\bar{\psi} = \tilde{\psi}C^{-1}$. The matrices $E$ are symmetric or antisymmetric according to

$$\tilde{E}(\mathbf{g}) = (-)^p \epsilon(g + p) E(\mathbf{g}). \tag{2.11}$$

 0022-2488/85/030536-06$02.50 © 1985 American Institute of Physics 

Computations that make direct use of (2.1) are manifestly covariant but may become very lengthy if $d$ is large. For this reason we shall use instead of (2.1) the following relation

$$\Gamma(\mathbf{g})\Gamma(\mathbf{g}') = \Gamma(\mathbf{g}')\Gamma(\mathbf{g})(-)^{gg'+\mathbf{gg}'}. \tag{2.12}$$

In this notation the composition law is

$$\Gamma(\mathbf{a})\Gamma(\mathbf{b}) = (-)^{ab}(-)^{\mathbf{bT a}}\Gamma(\mathbf{a}+\mathbf{b}), \tag{2.13}$$

where $T$ is the triangular matrix

$$\begin{aligned} T_{ij} &= \theta(i-j) \\ &= 0, \quad i<j \\ &= 1, \quad i\geqslant j. \end{aligned} \tag{2.14}$$

Then

$$\Gamma(\mathbf{b})\Gamma(\mathbf{a}) = (-)^{\mathbf{bT a}+\mathbf{aT b}}\Gamma(\mathbf{a})\Gamma(\mathbf{b}),$$

and by (2.12)

$$(-)^{\mathbf{bT a}+\mathbf{aT b}} = (-)^{ab+\mathbf{ab}}. \tag{2.15}$$

Also

$$(-)^{\mathbf{aT a}} = (-)^{a}\epsilon(a), \tag{2.16}$$

$$\tfrac{1}{2}(\Gamma(\mathbf{a}),\Gamma(\mathbf{b}))_{\pm} = \tfrac{1}{2}\Gamma(\mathbf{a}+\mathbf{b})(-)^{\mathbf{bT a}}[(-)^{ab}\pm(-)^{\mathbf{ab}}]. \tag{2.17}$$

The $\Gamma$ in these relations are antisymmetric by construction.

If $\Gamma(\mathbf{a})$ is not written in the above notation but is instead written in covariant notation and is also antisymmetrized then formulas like (2.17) are written as follows:

$$\begin{aligned} &\tfrac{1}{2}\{\Gamma^{M_1,\ldots,M_{2k}},\Gamma_{N_1,\ldots,N_m}\} \\ &\quad = \sum_{j=0}^{a}(-)^{j}\frac{m!}{(m-2j)!}\binom{2k}{2j}\delta^{[M_1}_{[N_1,\ldots}\delta^{M_{2j}}_{N_{2j}} \\ &\quad \times\Gamma^{M_{2j+1},\ldots,M_{2k}]}_{N_{2j+1},\ldots,N_m]}, \text{ with } a=\text{Min}\{k,[m/2]\}, \end{aligned} \tag{2.18}$$

$$\begin{aligned} &\tfrac{1}{2}\{\Gamma^{M_1,\ldots,M_{2k+1}},\Gamma_{N_1,\ldots,N_{2p+1}}\} \\ &\quad = \sum_{j=0}^{a}(-)^{j}\frac{(2p+1)!}{[2(p-j)]!}\binom{2k+1}{2j+1}\delta^{[M_1}_{[N_1,\ldots}\delta^{M_{2j+1}}_{N_{2j+1}} \\ &\quad \times\Gamma^{M_{2j+2},\ldots,M_{2k+1}]}_{N_{2j+2},\ldots,N_{2p+1}]}, \quad a=\text{Min}\{k,p\}, \end{aligned} \tag{2.19}$$

$$\begin{aligned} &\tfrac{1}{2}[\Gamma^{M_1,\ldots,M_{2k}},\Gamma_{N_1,\ldots,N_m}] \\ &\quad = \sum_{j=0}^{a}(-)^{j+1}\frac{m!}{(m-2j-1)!}\binom{2k}{2j+1}\delta^{[M_1}_{[N_1,\ldots}\delta^{M_{2j+1}}_{N_{2j+1}} \\ &\quad \times\Gamma^{M_{2j+2},\ldots,M_{2k}]}_{N_{2j+2},\ldots,N_m]}, \quad a=\text{Min}\{k-1,[(m-1)/2]\}, \end{aligned} \tag{2.20}$$

$$\begin{aligned} &\tfrac{1}{2}[\Gamma^{M_1,\ldots,M_{2k+1}},\Gamma_{N_1,\ldots,N_{2p+1}}] \\ &\quad = \sum_{j=0}^{a}(-)^{j}\frac{(2p+1)!}{(2p+1-2j)!}\binom{2k+1}{2j}\delta^{[M_1}_{[N_1,\ldots}\delta^{M_{2j}}_{N_{2j}} \\ &\quad \times\Gamma^{M_{2j+1},\ldots,M_{2k+1}]}_{N_{2j+1},\ldots,N_{2p+1}]}, \quad a=\text{Min}(k,p). \end{aligned} \tag{2.21}$$

*Example:*

To illustrate the use of our abbreviated notation consider a contraction like

$$\gamma^{k}(\gamma_m\gamma_n)\gamma_k,$$

or in general

$$\sum_{\mathbf{g}^2=g}\Gamma(\mathbf{g})\Gamma(\mathbf{m})\Gamma(\mathbf{g}),$$

where the sum is over all components of the tensor $\Gamma(\mathbf{g})$ of rank $g$. Then

$$\begin{aligned} \sum_{\mathbf{g}^2=g}\Gamma(\mathbf{g})\Gamma(\mathbf{m})\Gamma(\mathbf{g}) &= \sum_{\mathbf{g}^2=g}\Gamma(\mathbf{g})\Gamma(\mathbf{g})\Gamma(\mathbf{m})(-)^{gm+\mathbf{gm}} \\ &= (-)^{gm}\epsilon(\mathbf{g})\Gamma(\mathbf{m})\sum_{\mathbf{g}^2=g}(-)^{\mathbf{gm}} \\ &= (-)^{gm}\epsilon(g)W(\mathbf{m},g)\Gamma(\mathbf{m}), \end{aligned} \tag{2.22}$$

where

$$W(\mathbf{m},g) = \sum_{\mathbf{g}^2=g}(-)^{\mathbf{gm}}. \tag{2.23}$$

The function $W(\mathbf{m},g)$ is computed in the next paragraph. It is

$$W(\mathbf{m},g) = W(m,g) = \sum_{\mu=0}^{g} C^{m}_{\mu}C^{d-m}_{g-\mu}(-)^{\mu}. \tag{2.24}$$

## III. FIERZ IDENTITY

Let us next write the Fierz identity in the above notation

$$(\bar{\lambda}M\chi)(\bar{\psi}N\phi) = \frac{1}{\nu}\sum\epsilon(g)\,(\bar{\lambda}\cdot M\Gamma(\mathbf{g})N\cdot\phi)\,(\bar{\psi}\Gamma(\mathbf{g})\chi), \tag{3.1}$$

where

$$\nu = 2^{[d/2]},$$

and

$$M = \Gamma(\mathbf{m}), \quad N = \Gamma(\mathbf{n}).$$

In particular, consider

$$M(\mathbf{k}|\mathbf{n}) = K(\mathbf{k})N(\mathbf{n}), \tag{3.2}$$

with

$$\mathbf{kn} = 0. \tag{3.2a}$$

Then

$$\mathbf{m} = \mathbf{k}+\mathbf{n}, \quad m = k+n. \tag{3.3}$$

In this case the quadrilinear form is a tensor of rank $k$. Then by (2.12)

$$\begin{aligned} K(\mathbf{k})N(\mathbf{n})\Gamma(\mathbf{g})N(\mathbf{n}) &= K(\mathbf{k})N(\mathbf{n})N(\mathbf{n})\Gamma(\mathbf{g})(-)^{ng+\mathbf{ng}} \\ &= K(\mathbf{k})\Gamma(\mathbf{g})\epsilon(n)(-)^{ng+\mathbf{ng}}. \end{aligned} \tag{3.4}$$

Therefore

$$\begin{aligned} (\bar{\lambda}M\chi)(\bar{\psi}N\phi) = \frac{1}{\nu}\sum_{\mathbf{g}}&(\bar{\lambda}\,K(\mathbf{k})\Gamma(\mathbf{g})\phi)\,(\bar{\psi}\Gamma(\mathbf{g})\chi) \\ &\times\epsilon(g)\epsilon(n)(-)^{ng+\mathbf{ng}}. \end{aligned} \tag{3.5}$$

The contracted form of this relationship is

$$\begin{aligned} &\sum_{\mathbf{n}}(\bar{\lambda}M(\mathbf{k}|\mathbf{n})\chi)\,(\bar{\psi}N(\mathbf{n})\phi) \\ &\quad = \frac{1}{\nu}\sum_{g}\epsilon(n+g)\sum_{\mathbf{g}^2=g}\sum_{\substack{\mathbf{n}^2=n \\ \mathbf{nk}=0}}(-)^{\mathbf{ng}} \end{aligned}$$

$$\times[\bar{\lambda}\,K(\mathbf{k})\Gamma(\mathbf{g})\phi\,]\,[\bar{\psi}\Gamma(\mathbf{g})\chi]$$
$$=\frac{1}{\nu}\sum_{g}\epsilon(n+g)\sum_{\mathbf{g}^2=g}W(\mathbf{g},\mathbf{k},n)$$
$$\times[\bar{\lambda}\,K(\mathbf{k})\Gamma(\mathbf{g})\phi\,]\,[\bar{\psi}\Gamma(\mathbf{g})\chi], \tag{3.6}$$

where the following sum is over $\mathbf{n}$:

$$W(\mathbf{g},\mathbf{k},n)=\sum_{\substack{\mathbf{n}^2=n\\ \mathbf{nk}=0}}(-)^{\mathbf{ng}}. \tag{3.7}$$

Then

$$W(\mathbf{g},\mathbf{k},n)=\sum f(\mu)\,(-)^{\mu}, \tag{3.8}$$

where $f(\mu)$ is the number of solutions of the equation

$$\mathbf{ng}=\mu, \tag{3.9}$$

subject to the constraints on $\mathbf{n}$

$$\mathbf{n}^2=n, \quad \text{and } \mathbf{nk}=0.$$

One finds

$$f(\mu)=C_{\mu}^{\bar{g}}\,C_{n-\mu}^{d-k-\bar{g}}, \tag{3.10}$$

where

$$\bar{g}=g-\mathbf{gk}. \tag{3.11}$$

Then

$$W(\mathbf{g},\mathbf{k},n)=\sum_{\mu=0}^{\bar{g}}C_{\mu}^{\bar{g}}\,C_{n-\mu}^{d-k-\bar{g}}(-)^{\mu}. \tag{3.12}$$

The function $W(g,n)$ may also be expressed as a contour integral around the origin

$$W(\mathbf{g},n)=\frac{1}{2\pi i}\oint dz\,z^{-n-1}(1-z)^{\bar{g}}(1+z)^{d-k-\bar{g}}. \tag{3.13}$$

*Example:* In particular, if the quadrilinear is an invariant then $k=0$ and

$$W(\mathbf{g},n)=W(g,n)=\sum_{0}^{g}C_{\mu}^{g}\,C_{n-\mu}^{d-g}(-)^{\mu}, \tag{3.14}$$

or[6]

$$W(g,n)=\frac{1}{2\pi i}\oint dz\,z^{-n-1}(1-z)^{g}(1+z)^{d-g}. \tag{3.15}$$

In these cases $W(g,n)$ depends only on $(g,n)$. [Equation (3.14) is also the equation referred to in (2.24). Compare (2.23) and (3.7) with $\mathbf{k}=0$.]

If $k=0$, (3.6) becomes

$$\sum_{\mathbf{n}^2=\mathrm{n}}(\bar{\lambda}\Gamma(\mathbf{n})\chi)\,(\bar{\psi}\Gamma(\mathbf{n})\phi\,)$$
$$=\frac{1}{\nu}\sum_{g}\epsilon(n+g)W(g,n)\sum_{\mathbf{g}^2=g}(\bar{\lambda}\Gamma(\mathbf{g})\phi\,)\,(\bar{\psi}\Gamma(\mathbf{g})\chi). \tag{3.16}$$

If the quadrilinear is a tensor of rank $k$ then (3.6) holds; but if it is an invariant, then the special result (3.16) holds. In these two cases the appropriate forms of $W(\mathbf{g},n)$ are given by (3.12)–(3.15).

## IV. IDENTITY OF CREMER, JULIA, AND SCHERK[2]

If $k\neq 0$, then $W(\mathbf{g},n)$ depends on $\mathbf{g}$ as well as $g$. To illustrate the usefulness of formula (3.12) we prove the following identity of Cremer, Julia, and Scherk for $d=11$:

$$\tfrac{1}{8}\{\Gamma^{\mu\nu\alpha\delta\beta\gamma}\psi_{\nu}\,\bar{\psi}_{\alpha}\Gamma_{\beta\gamma}-\Gamma_{\beta\gamma}\psi_{\nu}\bar{\psi}_{\alpha}\Gamma^{\mu\nu\alpha\delta\beta\gamma}\}$$
$$+\tfrac{1}{8}\{\Gamma^{\mu\nu\alpha\delta\beta}\psi_{\nu}\bar{\psi}_{\alpha}\Gamma_{\beta}-\Gamma_{\beta}\psi_{\nu}\bar{\psi}_{\alpha}\Gamma^{\mu\nu\alpha\delta\beta}\}$$
$$+\tfrac{1}{4}\{[\Gamma^{\mu\nu\alpha\delta},\Gamma^{\beta}]\,\psi_{\nu}\bar{\psi}_{\alpha}\Gamma_{\beta}+\Gamma_{\beta}\psi_{\nu}\bar{\psi}_{\alpha}\,[\Gamma^{\mu\nu\alpha\delta},\Gamma^{\beta}]\}$$
$$=\tfrac{1}{4}\,[\Gamma^{\mu\nu\alpha\delta},\Gamma^{\beta}]\,(\bar{\psi}_{\alpha}\Gamma_{\beta}\psi_{\nu}). \tag{4.1}$$

*Proof:*

By (3.1) a typical term on the left is Fierz transformed as follows:

$$\sum_{\substack{\mathbf{n}^2=n\\ \mathbf{nk}=0}}M(\mathbf{k}|\mathbf{n})\psi_{\nu}\bar{\psi}_{\alpha}M(\mathbf{n})$$
$$=-\frac{1}{\nu}\sum_{\mathbf{g}}\sum_{\mathbf{n}}(M(\mathbf{k}|\mathbf{n})\Gamma(\mathbf{g})M(\mathbf{n}))\,(\bar{\psi}_{\alpha}\Gamma(\mathbf{g})\psi_{\nu})\epsilon(g), \tag{4.2}$$

where

$$M(\mathbf{k}|\mathbf{n})=\Gamma^{\mu\nu\alpha\delta\beta\gamma},$$
$$M(\mathbf{k})=\Gamma^{\mu\nu\alpha\delta},$$
$$M(\mathbf{n})=\Gamma^{\beta\gamma}.$$

The notation is hybrid, since $(\nu,\alpha)$ are contracted against $M(\mathbf{k}|\mathbf{n})$.

The additional minus sign on the right comes from the anticommutativity of $\bar{\psi}_{\alpha}$ and $\psi_{\nu}$. The $\mathbf{g}$ sum extends only from $g=0$ to $g=5$ since $\Gamma(11)\sim 1$.

Since $\bar{\psi}_{\alpha}$ and $\psi_{\nu}$ are anticommuting Majorana fields

$$\bar{\psi}_{\alpha}\Gamma(g)\psi_{\nu}=(-)^{g}\epsilon(g)\,(\bar{\psi}_{\nu}\Gamma(\mathbf{g})\,\psi_{\alpha}).$$

But by (4.1) $\nu$ and $\alpha$ are contracted against matrices antisymmetric in $\nu$ and $\alpha$. Hence the only nonvanishing terms satisfy

$$(-)^{g}\epsilon(g)=-1,$$

or

$$g=1,2,5, \tag{4.3}$$

since $g<5$. Now define

$$\mathscr{M}(\mathbf{k},m,\mathbf{g})=\sum_{\substack{\mathbf{m}^2=m\\ \mathbf{mk}=0}}[M(\mathbf{k}|\mathbf{m})\Gamma(\mathbf{g})M(\mathbf{m})$$
$$-M(\mathbf{m})\Gamma(\mathbf{g})M(\mathbf{k}|\mathbf{m})], \tag{4.4}$$

and

$$\mathscr{N}(\mathbf{k},n,\mathbf{g})=\sum_{\mathbf{n}^2=n}([M(\mathbf{k}),M(\mathbf{n})]\;\Gamma(\mathbf{g})M(\mathbf{n})$$
$$+M(\mathbf{n})\Gamma(\mathbf{g})\,[M(\mathbf{k}),M(\mathbf{n})]). \tag{4.5}$$

Then Eq. (4.1) becomes after its left side is Fierz transformed by (4.2)

$$\frac{1}{\nu}\sum_{\mathbf{g}}[\mathscr{M}(\mathbf{4},\mathbf{2},\mathbf{g})+\mathscr{M}(\mathbf{4},\mathbf{1},\mathbf{g})+\mathscr{N}(\mathbf{4},\mathbf{1},\mathbf{g})]\,(\bar{\psi}_{\alpha}\Gamma(\mathbf{g})\psi_{\nu})\,\epsilon(g)$$
$$=-(M(\mathbf{4}),\Gamma(\mathbf{1}))\,(\bar{\psi}_{\alpha}\Gamma(\mathbf{1})\psi_{\nu}). \tag{4.6}$$

To pass from (4.1) to the present notation one must double the first term of (4.1) since $(\nu\alpha)$ is counted twice in (4.1). As previously noted the only nonvanishing terms in the $g$ sum are $g=1,2,5$. Therefore (4.6) may be established by proving the following simpler identities:

$$\mathscr{M}(4,2,g)+\mathscr{M}(4,1,g)+\mathscr{N}(4,1,g)=0,\quad g=2,5, \tag{4.7}$$

and

$$\tfrac{1}{32}\,[\mathscr{M}(4,2,1)+\mathscr{M}(4,1,1)+\mathscr{N}(4,1,1)]\,(\bar{\psi}_\alpha\Gamma(1)\psi_\nu)$$
$$=-(M(4),\Gamma(1))\,(\bar{\psi}_\alpha\Gamma(1)\psi_\nu). \tag{4.8}$$

To prove (4.7) and (4.8) note first

$$M(k|m)\Gamma(g)M(m)-M(m)\Gamma(g)M(k|m)$$
$$=(M(k|m)\,M(m),\Gamma(g))\,(-)^{mg+mg}. \tag{4.9}$$

Since, $k$ is even and $km=0$,

$$\mathscr{M}(k,m,g)=\epsilon(m)\,(-)^{mg}W(g,m)\,(M(k),\Gamma(g))=0, \tag{4.10}$$

if $kg$ is also even. By (3.12) one finds

$$W(g,1)=x,\quad W(g,2)=\tfrac{1}{2}(x^2-7), \tag{4.11}$$

where

$$x=7-2\bar{g},\quad \bar{g}=g-gk. \tag{4.11a}$$

One also computes for $n=1$

$$\mathscr{N}(k,n,g)=2(-)^g(k-2gk)\,(M(k),\Gamma(g)). \tag{4.12}$$

Then

$$\tfrac{1}{32}\,[\mathscr{M}(4,2,g)+\mathscr{M}(4,1,g)+\mathscr{N}(4,1,g)]$$
$$=F(g,x)\,(M(4),\Gamma(g)), \tag{4.13}$$

where

$$F(g,x)=\tfrac{1}{32}\,[\tfrac{7}{2}+22(-)^g-4g(-)^g$$
$$-((x^2/2)+(-)^g x)]\,. \tag{4.14}$$

We find the values of $F(g,x)$ given in Table I. From these values one sees that (4.7) and (4.8) are correct and therefore that the identity of Cremer, Julia, and Scherk also holds.

## V. PERMUTATION SYMMETRIES OF FERMI COUPLINGS[7]

Let

$$E_g(abcd)=\sum_{g'=g}(aC^{-1}\Gamma(g)b)\,(cC^{-1}\Gamma(g)d),$$
$$g=0,\ldots,D. \tag{5.1}$$

Let

$$PE_g(abcd)=E_g\,(P(abcd)), \tag{5.2}$$

where $P(abcd)$ is a permutation of $abcd$.

Then

$$PE_g=\sum_{g'}E_{g'}\,P(g',g). \tag{5.3}$$

Let $E$ be a vector in the space spanned by the $E_g$

$$E=\sum h_g E_g. \tag{5.4}$$

The same vector in the new basis $PE_g$ will have new components $(h')$

$$h_g=\sum P(g,g')\,h'_{g'}\,. \tag{5.5}$$

There are 4! permutations in general, but four of these correspond to the unit matrix. Thus there are only 3! different matrices and these provide a representation of $\pi_3$, the permutation group on three objects; that is, $\pi_3$ and its three cosets in $\pi_4$ have the same representation.

Denote the elements of $\pi_3$ as follows: $I,A,B,C,AB,BA$. These fall into three classes: the identity $I$; the transpositions $A, B, C$; and the two elements of period 3, $AB$ and $BA$. Hence the following matrices commute with every member of $\pi_3$:

$$R_1=A+B+C, \tag{5.6a}$$

$$R_2=AB+BA. \tag{5.6b}$$

One also has $C=ABA$. Then all matrices may be generated from the noncommuting matrices $A$ and $B$. Let us take $A=(ab)$ and $B=(bd)$.

In the Hermitian representation that we are using $A$ is diagonal and[7]

$$A(g,g')=(-)^p\epsilon(g+p)\delta(g,g'), \tag{5.7}$$

and $B$ is

$$B(g,g')=2^{-n}\epsilon(g+g')$$
$$\times[\,g!\,g'!(2n-g)!(2n-g)!\,]^{1/2}\sum(g,g'), \tag{5.8}$$

where

$$\sum(g,g')=\sum_0^{2n}[(2n+m-g-g')!(g-m)!$$
$$\times(g'-m)!m!\,]^{-1}(-)^m. \tag{5.8a}$$

In this representation the transposition matrices are real and symmetric.

There are three irreducible representations of $\pi_3$: Two of these, $\Gamma_+$ and $\Gamma_-$, are one dimensional, and the third is two dimensional. Respectively, $\Gamma_+$ and $\Gamma_-$ are completely symmetric and antisymmetric. The projection operators associated with the irreducible representations are, in general,

$$e_i=\frac{n_i}{N}\sum\chi_i^*(P)\,P, \tag{5.9}$$

where $\chi_i(P)$ is the trace of $P$ in the irreducible representation $\Gamma_i$, $n_i$ is the dimensionality of $\Gamma_i$, and the summation is over the $N$ members of the permutation group. Then

$$e_+=\tfrac{1}{6}(1+R_1+R_2),$$
$$e_-=\tfrac{1}{6}(1-R_1-R_2), \tag{5.10}$$
$$e_2=\tfrac{1}{3}(2-R_2).$$

We are particularly interested in the symmetric and antisymmetric representations.

The number of times $c_i$ an irreducible $\Gamma_i$ is contained in any given representation is

$$c_i=\frac{1}{N}\sum\chi_i^*(P)\,\chi(P), \tag{5.11}$$

TABLE I. Values of $F(g,x)$.

| $g$ | $gk$ | $\bar{g}$ | $x$ | $F(g,x)$ |
|---|---|---|---|---|
| 1 | 1 | 0 | 7 | −1 |
| 2 | 1 | 1 | 5 | 0 |
| | 1 | 4 | −1 | 0 |
| 5 | 3 | 2 | 3 | 0 |

where $\chi(P)$ corresponds to the given representation. $P$ is summed over the group.

## VI. CHARACTERS

Since $A$ is diagonal we have simply

$$\chi(A)=\sum_{g=0}^{2n}(-)^p\epsilon(g+p)=\epsilon(-p)(-)^{np}. \tag{6.1}$$

We are working with Euclidean signature so that $p=n-1 \bmod 4$ or $p=n \bmod 4$. In either case

$$\chi(A)=\epsilon(n). \tag{6.2}$$

$B$ is not diagonal but must have the same trace as $A$ since it belongs to the same class. By (5.8)

$$\chi(B)=\frac{1}{2^n}\sum_{g=0}^{2n}(-)^g\sum_{m=0}^{g}(-)^m C_m^g C_{g-m}^{2n-g} \tag{6.3}$$

$$=\frac{1}{2^n}\sum_{g=0}^{2n}(-)^g\oint_c\frac{dz}{2\pi i}\,z^{-g-1}(1+z)^{2n-g}(1-z)^g, \tag{6.4}$$

where $c$ is a small circle around the origin. This character may be computed by interchanging the order of summation and integration. One finds

$$\chi(B)=\frac{1}{2^n}\oint_c\frac{dz}{2\pi i}\frac{(1+z)^{2n}}{z}\sum_{g=0}^{2n}\frac{(-)^g}{z^g}\left(\frac{1-z}{1+z}\right)^g \tag{6.5}$$

$$=\epsilon(n), \tag{6.6}$$

as required. The remaining character is

$$\chi(AB)=\chi(BA)=\sum_{g,g'}A(g,g')\,B(g',g)$$

$$=\sum_{g=0}^{2n}(-)^p\epsilon(g+p)\,B(g,g). \tag{6.7}$$

Let us separate the even and odd values of $p$.

(a) Let $p=2q+1$. Then

$$\epsilon(g+p)=(-)^q\epsilon(-g). \tag{6.8}$$

(b) Let $p=2q$. Then

$$\epsilon(g+p)=(-)^q\epsilon(g). \tag{6.9}$$

Then for $p$ odd

$$\chi(AB)=-(-)^q\sum_{g=0}^{2n}\epsilon(-g)\,B(g,g) \tag{6.10}$$

$$=(-)^{q+1}\left\{\sum_{\text{even}}(-)^{g/2}B(g,g)+\sum_{\text{odd}}(-)^{1/2(g+1)}B(g,g)\right\}$$

$$=(-)^{q+1}\left\{\sum_{\lambda=0}^{n}(-)^\lambda B(2\lambda,2\lambda)+\sum_{\lambda=1}^{n}(-)^\lambda B(2\lambda-1,2\lambda-1)\right\}$$

$$=\frac{(-)^{q+1}}{2^n}\left\{\oint\frac{dz}{2\pi i}\frac{(1+z)^{2n}}{z}\sum_{\lambda=0}^{n}\frac{(-)^\lambda}{z^{2\lambda}}\left(\frac{1-z}{1+z}\right)^{2\lambda}\right.$$

$$\left.-\oint\frac{dz}{2\pi i}\frac{(1+z)^{2n+1}}{1-z}\sum_{\lambda=1}^{n}\frac{(-)^\lambda}{z^{2\lambda}}\left(\frac{1-z}{1+z}\right)^{2\lambda}\right\}$$

$$=\frac{(-)^{q+1}}{2^n}\oint\frac{dz}{2\pi i}\frac{(-)^n}{z^{2n+1}}\frac{(1-z)^{2n+1}}{z^2(1+z)^2+(1-z)^2}$$

$$\times[1-z-(1+z)z]. \tag{6.11}$$

The denominator of the integrand may be factored

$$z^2(1+z)^2+(1-z)^2=[z^2+(1-i)z+i]\,[z^2+(1+i)z-i]. \tag{6.12}$$

The contour about the origin may be deformed to enclose these four roots. Then for $p$ odd

$$\chi(AB)=\frac{(-)^{q+1}}{3}\sqrt{6}\left\{\sin(2n+1)\frac{7\pi}{12}-(-)^n\times\cos(2n+1)\frac{7\pi}{12}\right\}. \tag{6.13}$$

For $p$ even

$$\chi(AB)=\frac{(-)^{q+1}}{3}\sqrt{6}\left\{\cos(2n+1)\frac{7\pi}{12}-(-)^n\times\sin(2n+1)\frac{7\pi}{12}\right\}. \tag{6.14}$$

For either even or odd $p$

$$\chi(AB)=-\epsilon(n)\frac{1}{3}\sqrt{6}\left\{\cos(2n+1)\frac{7\pi}{12}-(-)^n\times\sin(2n+1)\frac{7\pi}{12}\right\}. \tag{6.15}$$

This formula leads to the following results:

$$\chi(AB)=1, \quad n=3m, \tag{6.16a}$$

$$=0, \quad n=3m+1, \tag{6.16b}$$

$$=-1, \quad n=3m+2. \tag{6.16c}$$

## VII. SYMMETRIC AND ANTISYMMETRIC REPRESENTATIONS

According to (5.11) the numbers of symmetric ($c_+$) and antisymmetric ($c_-$) representations are

$$c_\pm=\tfrac{1}{6}[\chi(I)\pm 3\chi(A)+2\chi(AB)]$$

$$=\tfrac{1}{6}[2n+1\pm 3\epsilon(n)+2\chi(AB)]. \tag{7.1}$$

As $\epsilon(x+2m)=(-)^m\epsilon(x)$

(a) $n=3m$, $c_\pm=m+\tfrac{1}{2}[1\pm\epsilon(-m)]$, (7.2a)

(b) $n=3m+1$, $c_\pm=m+\tfrac{1}{2}[1\pm\epsilon(m)]$, (7.2b)

(c) $n=3m+2$, $c_\pm=m+\tfrac{1}{2}[1\mp\epsilon(-m)]$. (7.2c)

The results up to $d=12$ are summarized in Table II.

TABLE II. Numbers of symmetric (antisymmetric) forms.

| $d$ | $n$ | $m$ | $c_+$ | $c_-$ |
|---|---|---|---|---|
| 2 | 1 | 0(b) | 1 | 0 |
| 4 | 2 | 0(c) | 0 | 1 |
| 6 | 3 | 1(a) | 1 | 2 |
| 8 | 4 | 1(b) | 2 | 1 |
| 10 | 5 | 1(c) | 2 | 1 |
| 12 | 6 | 2(a) | 2 | 3 |

TABLE III. Symmetric and antisymmetric forms in (6,7) and (10,11) dimensions.

| $g$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| symmetric | 1 | 0 | 0 | $-\frac{2}{3}\sqrt{5}$ | $\frac{1}{3}\sqrt{5}$ | 0 | 0 | | | | | |
| | | | | | | | | | | | | $d=(6,7)$ |
| anti-symmetric | 0 | 1 | 0 | 0 | 0 | $-1$ | 0 | | | | | $c_+=1$ |
| | | | | | | | | | | | | $c_-=2$ |
| | $-1$ | 0 | 0 | 0 | $\frac{1}{3}\sqrt{15}$ | $\frac{1}{4}\sqrt{6}$ | 0 | | | | | |
| symmetric | 0 | 0 | $\frac{1}{8}\sqrt{35}$ | 0 | 0 | 1 | $\frac{1}{8}\sqrt{30}$ | 0 | 0 | 0 | $\frac{1}{8}\sqrt{63}$ | |
| | 0 | 1 | $-\frac{1}{8}\sqrt{162}$ | 0 | 0 | 0 | $\frac{1}{8}\sqrt{84}$ | 0 | 0 | 1 | $\frac{1}{8}\sqrt{10}$ | $d=(10,11)$ |
| | | | | | | | | | | | | $c_+=2$ |
| anti-symmetric | 1 | 0 | 0 | $\frac{2}{15}\sqrt{30}$ | $-\frac{1}{15}\sqrt{7}$ | 0 | 0 | $\frac{2}{15}\sqrt{30}$ | $-\frac{1}{5}\sqrt{5}$ | 0 | 0 | $c_-=1$ |
| $w$ | $x$ | | $x$ | | $x$ | | $x$ | | $x$ | | $x$ | |

The cases of special interest are the 10- and 11-dimensional Kaluza–Klein theories that lead by dimensional reduction to maximally extended Yang–Mills and supergravity theories in four dimensions. In addition the Euclidean $d=7$ case is also of interest as the internal Kaluza–Klein space corresponding to $d=11$.

The specific symmetric and antisymmetric forms may be obtained with the aid of the corresponding projection operators or alternatively as common eigenvectors of $A$ and $B$. The eigenvectors of $B$ may be obtained by noting

$$A = C^{-1}BC. \tag{7.3}$$

Since $A$ is diagonal the eigenvectors of $B$ are the columns of $C$. But $C$ may be calculated directly as the product $ABA$. One obtains the results in Table III.

The forms described in Table III are all of the type

$$F = \sum_g A_g \sum_{\mathbf{g}} (a\Gamma(\mathbf{g})\mathbf{b})\,(c\Gamma(\mathbf{g})\,d\,). \tag{7.4}$$

The entries in Table III are the values of $A_g$.

For Majorana spinors, which can only exist in 2,3,4,8,9 mod 8 dimensions, we can use Dirac conjugation in our expressions, since for these $\psi^T C^{-1}$ may be replaced by $\psi^+\gamma^0$.

The last line describes the terms that vanish if $(abcd)$ are all Weyl spinors having the same Weyl parity. This condition drastically simplifies all forms in Table III except the first antisymmetric form in six dimensions, which remains unchanged.

In obtaining these forms commuting spinors have been assumed. If anticommuting spinors are used instead, then the symmetric and antisymmetric forms are interchanged. If some of the fermionic factors are vector spinors rather than simple spinors, then the corresponding forms will of course also be antisymmetric in the vector indices.

For supergravity applications the fermionic factors could be anticommuting vector spinors in 11 dimensions or commuting Killing spinors in seven dimensions, but we have not studied possible supersymmetric extensions of these forms.

In an octonionic realization of Englert's compaction of 11-dimensional supergravity, Gürsey and Tze[8] display both the torsion and the field strength as completely antisymmetric forms on the septads. As far as we have been able to discover, however, there is no simple relation between these forms and the corresponding antisymmetric spinor forms of this paper.


[1] L. Brink, J. H. Schwarz, and J. Scherk, Nucl. Phys. B **121**, 77 (1977).
[2] E. Cremmer, B. Julia, and J. Scherk, Phys. Lett. B **76**, 409 (1978).
[3] P. Freund and M. Rubin, Phys. Lett. B **97**, 233 (1980); F. Englert, Phys. Lett. B **119**, 339 (1982).
[4] E. Witten, Nucl. Phys. B **186**, 412 (1981).
[5] R. Finkelstein, Nuovo Cimento **1**, 1104 (1955).
[6] K. Case, Phys. Rev. **97**, 810 (1956).
[7] R. Finkelstein, Phys. Rev. **109**, 1842 (1958).
[8] F. Gürsey and C.H. Tze, Phys. Lett. B **127**, 191 (1983).

# A vector model for electroweak interactions

Frank Reifler
*RCA/Government and Commercial Systems, Moorestown, New Jersey 08057*

(Received 12 June 1984; accepted for publication 4 October 1984)

In this paper we present a vector model for the electroweak interactions. The Cartan map gives an isomorphism between Dirac bispinors and an isotropic class of Yang–Mills vector fields. The isotropic Yang–Mills vector fields $\mathbf{F}_k = \mathbf{E}_k + i\mathbf{H}_k$ with $k = 1,2,3$, satisfy the condition that the matrix of scalar invariants $(\mathbf{F}_j \cdot \mathbf{F}_k)$ equals a scalar multiple of the identity matrix. We show that all the bispinor observables commute with the Cartan isomorphism, including all gauge transformations, as well as Lorentz transformations. We derive the Yang–Mills equivalent Dirac equation. As a consequence of the vector model, we obtain a new Lagrangian for electroweak interactions, which is an alternative to the Weinberg–Salam Lagrangian. Moreover, we show that the vector model predicts that the Weinberg angle $\theta_w$ satisfies $\sin^2 \theta_w = 0.25$, which is close to the measured value of $\sin^2 \theta_w = 0.23$. The vector model accommodates all the lepton and quark flavors. Furthermore, it predicts the conservation of baryon number and lepton number, as well as electric charge in electroweak interactions. The vector model also gives a new interpretation to antiparticles. In the vector model, an antiparticle is characterized by its opposite baryon number, lepton number, and electric charge; yet both particles and antiparticles propagate forward in time with positive energies.

## I. PRESENTATION OF THE MODEL

In a recent paper,[1] we investigated some problems of assigning spinors to physical states; namely, that spinors are tied to a specific Cartesian coordinate frame, that a spinor representation is *not coordinate-free* and consequently can only be defined on a very restricted class of manifolds, and that the relativistic transformations do not act properly on spinors.

The difficulties in the case of spinors were resolved by the Cartan map, which is a locally one to one coordinate map from spinors (that is $C^2$) onto the manifold[2] of isotropic vectors in $C^3$. Unlike spinors, isotropic vector fields are coordinate-free, and hence have curvilinear components in arbitrary coordinate systems. For example, whereas isotropic vector fields can be defined on the manifold $S^4$, spinors cannot. Indeed, $S^4$ is not parallelizable, and therefore the Cartesian frames necessary for "spinor structure" cannot be defined on it.[3] Also, unlike spinors, isotropic vectors are transformed unambiguously under Lorentz transformations.

We further showed that the Cartan map commutes with *all* spinor observables, and leads to a simple Dirac equation which preserves the isotropic vector constraint. Other authors have also discussed relations between antisymmetric tensors and spinors, and in particular, relations between *constrained* tensor systems and spinors, similar to the Cartan map isomorphism.[4]

In this paper, we extend the Cartan map to bispinors. A bispinor $\tilde{\psi} = (\xi, \eta^*)$ consists of a spinor $\xi$ and a conjugated[5] spinor $\eta^*$. The extended Cartan map takes each bispinor $\tilde{\psi}$ to a triplet of Yang–Mills fields $(\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3)$, as depicted in Fig. 1. These Yang–Mills fields ($\mathbf{F}_k = \mathbf{E}_k + i\mathbf{H}_k$ for $k = 1,2,3$) satisfy the isotropic condition that the matrix of scalar invariants $(\mathbf{F}_j \cdot \mathbf{F}_k)$ be a scalar multiple of the identity matrix. That is, by definition,

$$\mathbf{F}_j \cdot \mathbf{F}_k = \lambda \delta_{jk}$$

with $j,k = 1,2,3$ and $\lambda$ a complex scalar field.

We show in Sec. III that the extended Cartan map is locally one to one from $C^4$ onto the manifold of isotropic Yang–Mills vector fields. We also show that the extended Cartan map commutes with all bispinor observables, and gauge and Lorentz transformations. Henceforth we will use the term Cartan map for either the extended or usual map as appropriate.

In addition to the Cartan isomorphism between bispinors and the isotropic Yang–Mills vector triplets $(\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3)$, we also discuss the Weinberg–Salam map from bispinors to *trispinors* depicted in Fig. 2. A trispinor consists of a spinor pair $(\nu, \xi) \in C^4$ and a conjugated spinor $\eta^* \in C^2$. Trispinors were used by Weinberg and Salam to model the electroweak SU(2)×U(1) gauge interactions.[6] We show that the isotropic Yang–Mills vector fields provide a different though similar model for the electroweak interactions. Both the trispinor and vector models obtain their properties from the Dirac bispinors, via the maps shown in Fig. 2.

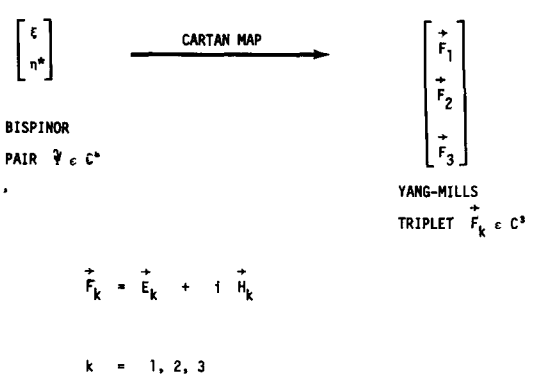



FIG. 1. The Cartan map.

 0022-2488/85/030542-09$02.50 © 1985 American Institute of Physics 

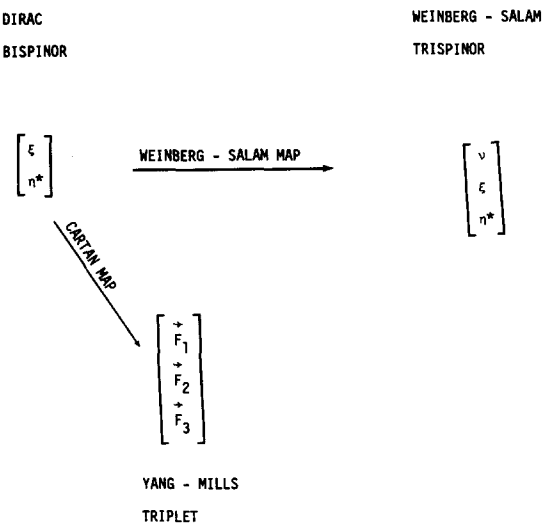



FIG. 2. The Weinberg–Salam map.

The study of isotropic Yang–Mills vector fields reveals the *bispinors* have SL(2,$C$) gauge symmetry. The gauge group SL(2,$C$) acts on Yang–Mills triplets ($\mathbf{F}_1$,$\mathbf{F}_2$,$\mathbf{F}_3$) via the complex orthogonal matrices. An important subgroup of SL(2,$C$) is SU(2), which has the usual three generators. These three generators give rise to "formal rotations" of $\mathbf{F}_1$, $\mathbf{F}_2$, and $\mathbf{F}_3$. By the Cartan map, the electromagnetic gauge transformations become the formal rotations about the "three-axis," that is, they leave $\mathbf{F}_3$ unchanged and formally rotate $\mathbf{F}_1$ and $\mathbf{F}_2$ into each other. We will therefore denote the electromagnetic group of gauge symmetries as $U(1)_3$, which is a subgroup of SL(2,$C$).

Previously, when considering bispinors,[7] physicists were only concerned with the $U(1)_3$ electromagnetic gauge symmetry. Here, we consider the larger group of gauge symmetries, SL(2,$C$). The classification of bispinor invariants under SL(2,$C$) is different than their usual classification under $U(1)_3$.

As shown in Sec. II, there is associated with each isotropic triplet of Yang–Mills vector fields ($\mathbf{F}_1$,$\mathbf{F}_2$,$\mathbf{F}_3$), a unique complex scalar $\rho$, and also a unique quadruplet of orthogonal *real Lorentz* currents ($j_0$,$j_1$,$j_2$,$j_3$). Both the singlet ($\rho$) and the quadruplet of currents ($j_\alpha$) are irreducible representations of the SL(2,$C$) gauge group, and by the Cartan map, *these are all bispinor SL(2,C) invariants.*

The usual bispinor invariants under $U(1)_3$ are those that are unchanged by electromagnetic gauge transformations (i.e., unchanged by formal rotations about the three-axis). These invariants in vector theory notation are $\rho$, $\mathbf{F}_3$, $j_0$, and $j_3$. In particular we will show in Sec. III that $j_0$ is the particle's probability current and $j_3$ is the chiral current. Table I lists those invariants by their usual notation, and gives their equivalent notation in the vector theory.

For the gauge group SL(2,$C$), the three-axis does not have a privileged role, and $\mathbf{F}_1$, $\mathbf{F}_2$ and $j_1$, $j_2$ must be included in the complete set of bispinor invariants shown in Table II. We propose that currents derived from ($j_0$,$j_1$,$j_2$,$j_3$) play a role in the vector model that is similar to the role played by the electroweak currents in the Weinberg–Salam model.

For the wave equation to be SL(2,$C$) gauge invariant, a quadruplet of *real* Higgs scalars ($\phi_0$, $\phi_1$, $\phi_2$, $\phi_3$) is also required. That is, $\phi_\alpha$ are real scalars for Lorentz transformations, but transform as a quadruplet under SL(2,$C$) gauge transformations. In Sec. IV, we extend the usual Lagrangian for bispinors to incorporate these Higgs fields. From this Lagrangian [discussed more fully below—see formula (1.8)] we obtain a complex triple of *conserved Noether currents*, denoted ($J_1$,$J_2$,$J_3$). The Noether currents $J_k$ may be expressed in terms of the real orthogonal currents $j_\alpha$ and the real Higgs scalars $\phi_\alpha$ by the formulas[8]

$$\begin{aligned} \mathrm{Re}\, J_k &= \phi_0 j_k - \phi_k j_0 , \\ \mathrm{Im}\, J_k &= -\epsilon_{kmn}\, \phi_m\, j_n , \end{aligned} \tag{1.1}$$

with $k, m, n = 1, 2, 3$. (See Table III.)

There is, thus, a real conserved Noether current for each of the six generators of SL(2,$C$). In particular, for the electromagnetic generator, we obtain $(e/m)$ Re $J_3$ as the *electric* current (where $e$ is the magnitude of the electric charge and $m$ is the mass). Formula (1) shows that for the conventional choice of Higgs scalars $\phi_\alpha = (0,0,0,m)$, the electric current is just $-ej_0$, where $j_0$ is the probability current (see Table I).

Isotropic Yang–Mills fields ($\mathbf{F}_1$,$\mathbf{F}_2$,$\mathbf{F}_3$) transform under a bigger gauge symmetry group SL(2,$C$)$\times U(1)_0$, which contains SL(2,$C$). The subgroup $U(1)_0$ consists of the "neutral" gauge symmetries.[9] By the Cartan map, the chiral gauge transformations on bispinors become the neutral transformations acting on $\mathbf{F}_k$. These neutral transformations map $\mathbf{F}_k$ to $\mathbf{F}_k\, e^{i\chi}$, where $\chi$ is a phase.

The Lagrangian $L$, given in formula (1.8), is invariant under SL(2,$C$). The kinetic part of $L$ is also invariant under SL(2,$C$)$\times U(1)_0$, however, the mass part of $L$ is *not* invariant under $U(1)_0$.

From the kinetic part of $L$, one can define a (nonconserved) Noether current[10] for $U(1)_0$. This is the *neutral* cur-

TABLE I. Usual bispinor invariants.

| | Usual notation | Vector model notation |
|---|---|---|
| Scalar | $s$ | Re $\rho$ |
| Pseudoscalar | $p$ | Im $\rho$ |
| Probability current | $j$ | $j_0$ |
| Chiral current | $j_5$ | $j_3$ |
| Spin tensor | $S^{\alpha\beta}$ | $\mathbf{F}_3$ |

TABLE II. Complete set of bispinor invariants under SL(2,$C$).

| | |
|---|---|
| $\rho$ | = singlet, complex scalar |
| $\begin{bmatrix}\mathbf{F}_1\\ \mathbf{F}_2\\ \mathbf{F}_3\end{bmatrix}$ | = triplet, Yang–Mills fields |
| $\begin{bmatrix}j_0\\ j_1\\ j_2\\ j_3\end{bmatrix}$ | = quadruplet, *real* orthogonal Lorentz currents |

TABLE III. Additional observables of the vector model.

| | |
|---|---|
| $\begin{bmatrix}\phi_0\\ \phi_1\\ \phi_2\\ \phi_3\end{bmatrix}$ | = quadruplet, *real* Higgs scalars |
| $\begin{bmatrix}J_1\\ J_2\\ J_3\end{bmatrix}$ | = triplet, conserved Noether currents |
| $K$ | = singlet, neutral current |

rent $K$, expressed in terms of the orthogonal currents $j_\alpha$ and the real Higgs scalars $\phi_\alpha$ by the formula

$$K = \phi^\alpha j_\alpha \tag{1.2}$$

(see Table III).

In particular, let us compute $(e_0/m)K$ with the conventional Higgs scalars $\phi_\alpha = (0,0,0,m)$, where $e_0$ is the magnitude of the neutral charge. From (1.2) we see that $(e_0/m)K$ is equal to $-e_0 j_3$, where $j_3$ is the chiral current.[11]

In Sec. IV we show that the gauge-invariant vector equivalent of the Dirac equation is given by

$$iD_\alpha S^\alpha \mathbf{F}_k + (\mathbf{D}\mathbf{F}_m)\cdot\mathbf{F}_n/\rho = \mathbf{J}_k, \tag{1.3}$$

with subscripts $(kmn)$ taken in cyclic order, and where

$D_\alpha$ = Yang–Mills covariant derivatives,

$S^\alpha$ = Proca spin-one matrices,

$\mathbf{F}_m\cdot\mathbf{F}_n = \lambda\delta_{mn}$,

$\rho = (\mathbf{F}_1\times\mathbf{F}_2\cdot\mathbf{F}_3)/\lambda$.

Equation (1.3) is equivalent to the usual Dirac equation for bispinors, with the Higgs scalars $\phi_\alpha = (0,0,0,m)$.

The vector model gives a new interpretation to the mass terms of the Dirac equation. We see from (1.3) that the mass terms are the conserved Noether currents $\mathbf{J}_k$, which appear as sources on the right-hand side of the wave equation.

If we state that $J_1$, $J_2$, $J_3$, and $K$ are the electroweak currents, and that $(\phi_0, \phi_1, \phi_2, \phi_3)$ are the Higgs scalars, then we have a vector model for the electroweak interactions similar to the Weinberg–Salam model. We will always regard the Yang–Mills triple $\mathbf{F}_k$ and the Higgs quadruplet $\phi_\alpha$ as the fundamental fields: $J_k$ and $K$ are determined by them.

The vector model reveals that the Dirac equation has broken SL(2,$C$) symmetry in a way not previously considered. Usually, the Dirac equation is written for bispinors, denoted $\tilde{\psi} = (\xi,\eta^*)$, which are comprised of a spinor $\xi$ and a conjugated spinor $\eta^*$. Associated with $\tilde{\psi}$ is the spinor pair $\psi = (\xi,\eta)$ and its conjugate pair $\psi^* = (\eta^*, -\xi^*)$, where $\eta = -(\eta^*)^*$. We will show in Sec. IV that Dirac's equation has broken SL(2,$C$) gauge symmetry by writing it as an equation for spinor pairs, instead of bispinors, using the bijective map $\tilde{\psi}\rightarrow\psi$ from bispinors to spinor pairs. The following Dirac equation from Sec. IV explicitly includes the Higgs scalars and is equivalent to Eq. (1.3) by the Cartan isomorphism. It is invariant under *both* Lorentz and SL(2,$C$) gauge transformations:

$$D_\alpha\sigma^\alpha\psi = -\phi_\alpha\hat{\tau}^\alpha\psi^*, \tag{1.4}$$

where $D_\alpha$ are the Yang–Mills covariant derivatives, $\sigma_\alpha$ are Pauli matrices, the $\hat{\tau}_\alpha$ are the gauge matrices (see Sec. III), and $\phi_\alpha$ is a real scalar Higgs field, a SL(2,$C$) gauge quadruplet. Solutions of Eq. (1.4) also satisfy the Klein–Gordon equation,[12] which in the case of free particles is given by

$$D^\alpha D_\alpha\psi = -(\phi_\alpha\phi^\alpha)\psi, \tag{1.5}$$

so that the mass is given by

$$M = \sqrt{-\phi_\alpha\phi^\alpha}. \tag{1.6}$$

As previously stated for the vector equation (1.3), Eq. (1.4) is equivalent to the usual Dirac equation when $\phi_\alpha$ is chosen by convention to be

$$\phi_\alpha = (0,0,0,m). \tag{1.7}$$

Equation (1.4) is the Euler–Lagrange equation for the following Lagrangian:

$$L = \mathrm{Re}\{(D_\alpha\sigma^\alpha\psi)\cdot\overline{(\phi^\beta\tau_\beta\psi)} + M^2\psi\cdot\tilde{\psi}^*\}/M \tag{1.8}$$

(see Lemma 4).

This Lagrangian, which is also an invariant scalar under *both* Lorentz transformations and SL(2,$C$) gauge transformations, is different than the Lagrangian used in the Weinberg–Salam[13] model, even though the Higgs field has the same number (four) of real components. Thus, the new Lagrangian will give us new predictions.

Specifically, the vector model predicts that the Weinberg angle $\theta_w$, used in the Weinberg–Salam model for electroweak interactions, satisfies $\sin^2\theta_w = 0.25$, which is close to the measured value[14] of $\sin^2\theta_w = 0.23$, and that the neutral charge is $1/\sqrt{3}$ times the electric charge. If in the vector model, we represent the $\mathbf{F}_k$ by their spinor coordinates $(\xi,\eta)$, then the generators $t_0$ and $t_3$ of the neutral and electromagnetic gauge transformations may be explicitly identified with the following matrices:

$$t_0 = I = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix},$$

$$t_3 = \tau^3 = \begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}. \tag{1.9}$$

In the Weinberg–Salam model the electroweak interaction is represented by trispinors[15]

$$\begin{bmatrix}\nu\\ \xi\\ \eta^*\end{bmatrix}.$$

Let the usual generators $t_0$ and $t_3$ of SU(2)$\times$U(1) for the Weinberg–Salam model be denoted by $t_0'$ and $t_3'$. Acting on trispinors, the generators $t_0'$ and $t_3'$ are represented by the matrices[16]

$$t_0' = \begin{bmatrix}-\frac{1}{2} & 0 & 0\\ 0 & -\frac{1}{2} & 0\\ 0 & 0 & -1\end{bmatrix},$$

$$t_3' = \begin{bmatrix}\frac{1}{2} & 0 & 0\\ 0 & -\frac{1}{2} & 0\\ 0 & 0 & 0\end{bmatrix}.$$

Since these matrices are diagonal, we may restrict them to just the bispinor $(\xi,\eta^*)$. To translate these generators into

the vector model for comparison with (1.9), consider their action on the pair $(\xi,\eta)$. Acting on the spinor pair $(\xi,\eta)$, we have

$$t'_0 = \begin{bmatrix} -\frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix},$$

$$t'_3 = \begin{bmatrix} -\frac{1}{2} & 0 \\ 0 & 0 \end{bmatrix},$$

so that we get from (1.9),

$$t_3 = -t'_3 - t'_0\,, \quad t_0 = -3t'_3 + t'_0\,. \tag{1.10}$$

The two models (vector and Weinberg–Salam) give two versions of the Yang–Mills covariant derivatives[17]:

$$D^{\alpha} = ih\nabla^{\alpha} + eV^{\alpha}_k t^k + e_0 V^{\alpha}_0 t^0\,,$$

$$D'^{\alpha} = ih\nabla^{\alpha} - gW^{\alpha}_k t'^k - g_0 W^{\alpha}_0 t'^0\,,$$

for $\alpha = 0,1,2,3$ and $k = 1,2,3$, where $D^{\alpha}$ is the covariant derivative for the vector model, $D'^{\alpha}$ is the covariant derivative for the Weinberg–Salam model, $e$ and $e_0$ are the absolute values of electric and neutral charges, respectively, $g$ and $g_0$ are the coupling coefficients, the $V^{\alpha}_{\beta}$ are the Yang–Mills potentials in the vector model (directly observable), and the $W^{\alpha}_{\beta}$ are the Yang–Mills potentials in the Weinberg–Salam model (not directly observable). Since they represent the same interaction, $D^{\alpha}$ and $D'^{\alpha}$ must be equal, thus

$$-eV^{\alpha}_3 t_3 - e_0 V^{\alpha}_0 t_0 = gW^{\alpha}_3 t'_3 + g_0 W^{\alpha}_0 t'_0\,. \tag{1.11}$$

Substituting (1.10) into (1.11) and equating coefficients of $t'_3$ and $t'_0$, we have

$$gW^{\alpha}_3 = eV^{\alpha}_3 + 3e_0 V^{\alpha}_0, \quad g_0 W^{\alpha}_0 = eV^{\alpha}_3 - e_0 V^{\alpha}_0\,. \tag{1.12}$$

Furthermore,[18]

$$W^{\alpha}_3 = V^{\alpha}_0 \cos\theta_w + V^{\alpha}_3 \sin\theta_w,$$

$$W^{\alpha}_0 = -V^{\alpha}_0 \sin\theta_w + V^{\alpha}_3 \cos\theta_w\,, \tag{1.13}$$

where $\theta_w$ is the Weinberg angle.

Equation (1.13) shows an essential difference between the vector model and the Weinberg–Salam model. In the vector model, $V^{\alpha}_3$ and $V^{\alpha}_0$ are the directly observed electric and neutral potentials. Whereas, in the Weinberg–Salam model $W^{\alpha}_3$ and $W^{\alpha}_0$ are *not* directly observable. They are related to the observable potentials $V^{\alpha}_3$ and $V^{\alpha}_0$ by the formal Weinberg rotation (1.13). Solving (1.12) and (1.13) we obtain a prediction for the Weinberg angle $\theta_w$ as follows: From (1.12) and (1.13),

$$g\cos\theta_w = 3e_0, \quad g\sin\theta_w = e\,,$$

$$g_0 \sin\theta_w = e_0, \quad g_0\cos\theta_w = e\,.$$

Therefore,

$$\tan\theta_w = \tfrac{1}{3}\cot\theta_w = g_0/g\,,$$

and $\sin^2\theta_w = \frac{1}{4}$. Moreover,

$$e_0 = e\tan\theta_w = (1/\sqrt{3})e\,.$$

Thus $\sin^2\theta_w = 0.25$, and the weak (neutral) charge $e_0$ is $1/\sqrt{3}$ times the electric charge $e$.

By the following argument the vector model also predicts that the electric charge of the neutrino is zero, and that its neutral charge is $2/\sqrt{3}$ times the electric charge $e$. In the vector model we regard the neutrino as having a very small mass $m'$, and the Higgs scalars may be given by $\phi_{\alpha} = (0,m',0,0)$. Then from formula (1.1) it follows that the electric current vanishes: $(e/m')\,\mathrm{Re}\,J_3 = 0$. Also from (1.2), the neutral current $(e_0/m')K = -e_0 j_1$. We substitute $\phi_{\alpha} = (0,m',0,0)$ into Eq. (1.4), and observe that in the massless limit (letting $m'$ approach zero) the spinor pair associated with the neutrino has the form $\psi = (\xi,\xi)$, and hence [see formulas (3.3)]

$$j^{\alpha}_1 = -2\bar{\xi}\sigma^{\alpha}\xi = -j^{\alpha}_0\,.$$

Regarding $j^{\alpha} = \bar{\xi}\sigma^{\alpha}\xi$ as the neutrino current, we see that

$$K = -e_0 j_1 = 2e_0 j = (2/\sqrt{3})ej\,,$$

in agreement with the Weinberg–Salam model. Thus, the electric charge of the neutrino is zero and its neutral charge is given by $(2/\sqrt{3})\,e$ as claimed.

In Table IV we list the predictions of the vector model including predictions for the masses of the gauge fields, $W_1$, $W_2$, and $Z$ $(Z = V_0)$. Except for these masses, which are smaller than predicted by the Weinberg–Salam model, the rest of Table IV agrees with the Weinberg–Salam model predictions, provided that the Weinberg angle satisfies $\sin^2\theta_w = \frac{1}{4}$ as proved above.[19]

When the Higgs scalars $\phi_{\alpha}$ assume conventional values, there is further correspondence between the vector model and the Weinberg–Salam model as shown in Table V. The electric and neutral currents are *identical* for both models, provided that the Weinberg angle $\theta_w$ satisfies $\sin^2\theta_w = \frac{1}{4}$. For example, the neutral current of an electron in the Weinberg–Salam model is given by[20]

$$j_{\text{neutral}} = -(e/\sqrt{3})[\,j_3 + (1 - 4\sin^2\theta_w)j_0\,]\,.$$

Clearly, this agrees with the vector model prediction $-(e/\sqrt{3})j_3$ derived above, when $\sin^2\theta_w = \frac{1}{4}$.

In Table VI we extend the assignment of Higgs scalars $\tilde{\phi}_{\alpha} = \phi_{\alpha}/M$ to include all quark and lepton flavors. An important property of this choice of Higgs scalars is that they are *additive in all electroweak interactions* (such as beta decay). This additivity is *necessary* if we assume that the seven charged Noether currents $(e_0/M)K$, $(e/M)\,\mathrm{Re}\,J_k$, and $(e/M)\,\mathrm{Im}\,J_k$ are additive in interactions. To see this, consider the particles at rest when their $j_0$'s are equal. We may then assume that the probability currents $j_0$ satisfy $\mathbf{j}_0 = 0$ and $j^0_0 = 1$. From Table VI we see that $\tilde{\phi}_0 = 0$; whereas formula (1.1) gives $(e/M)\,\mathrm{Re}\,J^0_k = -e\tilde{\phi}_k$. Hence the additivity of $(e/M)J_k$ implies that the $\tilde{\phi}_{\alpha}$ must be additive. We propose

TABLE IV. Predictions of the vector model.

| Weinberg angle: $\sin^2\theta_w = \frac{1}{4}$ | | |
|---|---|---|
| Charges[a] | Electric | Neutral |
| Electron | $-1$ | $-1/\sqrt{3}$ |
| Neutrino | 0 | $2/\sqrt{3}$ |
| Masses[b] | | |
| $m_w = 40\ \mathrm{GeV}/c^2$ $(80\ \mathrm{GeV}/c^2)$ | | |
| $m_z = 23\ \mathrm{GeV}/c^2$ $(90\ \mathrm{GeV}/c^2)$ | | |

[a] These charges agree with the Weinberg–Salam model predictions provided that $\sin^2\theta_w = \frac{1}{4}$.

[b] Masses predicted by the Weinberg–Salam model are shown in parenthesis.

TABLE V. Electric and neutral currents with conventional Higgs scalars. $\alpha = 1 - 4\sin^2\theta_w$.

| | Vector model | Weinberg–Salam model |
|---|---|---|
| Electric current (electron) | $-ej_0$ | $-ej_0$ |
| Electric current (neutrino) | 0 | 0 |
| Neutral current (electron) | $-(e/\sqrt{3})j_3$ | $-(e/\sqrt{3})(j_3+\alpha j_0)$ |
| Neutral current (neutrino) | $(2e/\sqrt{3})j$ | $(2e/\sqrt{3})j$ |

that electroweak interactions (which change particle identities) will conserve the sum of the Noether currents and, hence, of the Higgs scalars $\tilde{\phi}_\alpha$, giving us four additive "quantum" numbers which are conserved.

For example, consider the beta decay,

$$d \to u + e + \bar{\nu}_e .$$

From Table VI and formula (1.1), the electric current for the particles $u$, $d$, $e$, $\bar{\nu}_e$ is given by $-(e\tilde{\phi}_3)j_0$, so their electric charges are equal to $\frac{2}{3}e$, $-\frac{1}{3}e$, $-e$, 0, respectively (i.e., the Higgs scalars give the correct electric charges for the particles). The additivity of electric charge in all electroweak interactions, follows from the additivity of $\tilde{\phi}_3$ (shown in Table VI).

To obtain the correct charge, it must be true that $\tilde{\phi}_\alpha(\bar{e}) = -\tilde{\phi}_\alpha(e)$. We propose that this is true for all particles and antiparticles, i.e, that the Higgs scalars for the antiparticles (e.g., antineutrino, etc.) equal the *negatives* of the particle Higgs scalars. Then, in the particular example of beta decay, the additivity of $\tilde{\phi}_\alpha$ means that for each $\alpha = 0,1,2,3$,

$$\tilde{\phi}_\alpha(d) = \tilde{\phi}_\alpha(u) + \tilde{\phi}_\alpha(e) + \tilde{\phi}_\alpha(\bar{\nu}_e)$$

(see Table VI).

The additivity of the Higgs scalars $\tilde{\phi}_1$, $\tilde{\phi}_2$, and $\tilde{\phi}_3$ leads directly to three familiar conservation laws. Using Table VI, we derive the following relations:

$$B = -\tfrac{1}{2}\tilde{\phi}_2 = \text{baryon number},$$

$$L = \tilde{\phi}_1 - \tfrac{1}{2}\tilde{\phi}_2 + \tilde{\phi}_3 = \text{lepton number},$$

$$C = -\tilde{\phi}_3 = \text{electric charge}.$$

Since $B$, $L$, and $C$ are linear functions of $\tilde{\phi}_\alpha$ they must be additive also. Thus, the vector model predicts the conservation of baryon number and lepton number, as well as the electric charge, under the proposition that the Noether currents are additive in electroweak interactions.

TABLE VI. Higgs scalars for quarks and leptons. $\tilde{\phi}_\alpha = \phi_\alpha/M$.

| Quark flavors[a] | $\tilde{\phi}_\alpha$ |
|---|---|
| $u,c,t$ | $(0, \frac{1}{3}, -\frac{2}{3}, -\frac{2}{3})$ |
| $d,s,b$ | $(0, -\frac{2}{3}, -\frac{2}{3}, \frac{1}{3})$ |
| **Lepton flavors[a]** | $\tilde{\phi}_\alpha$ |
| $e,\mu,\tau$ | (0,0,0,1) |
| $\nu_e,\nu_\mu,\nu_\tau$ | (0,1,0,0) |

[a] Antiparticles: replace $\tilde{\phi}_\alpha$ with $-\tilde{\phi}_\alpha$.

The proof of this proposition will require a more complete Lagrangian than given in (1.8) which will include the dynamics of the Higgs fields. Such a Lagrangian must be invariant under gauge transformations for the sum of the Noether currents to be conserved. Although it is obvious that such a Lagrangian can be formulated, the dynamics of the Higgs field needs further investigation. This subject will be addressed in a forthcoming paper.

The vector model suggests that the Higgs field extends into the neighborhood of each particle. When two particles approach each other, their Higgs fields must superimpose, and then be redistributed. The redistribution of the Higgs fields causes the identities of the particles to change. However, the redistribution of the Higgs fields is subject to the additivity of the Higgs scalars $\tilde{\phi}_\alpha$, and hence to the additivity of the Noether currents.

Note, however, that there is an essential difference between the vector model and the Weinberg–Salam model. In the vector model, an electron becomes a neutrino when the Higgs field $\phi_\alpha$ changes. However, in the Weinberg–Salam model, this is accomplished introducing an additional neutrino "state." The vector model is, therefore, more economical because it does not require a proliferation of "states" to describe new particles.[21] Instead, the Higgs field $\phi_\alpha$ switches to different mass shells. Methods for switching "mass shells" are currently being studied, and will be addressed in a following paper.

The vector model also gives a new interpretation to the antiparticles. The baryon numbers, lepton numbers, and electric charges of a particle and its antiparticle sum to zero, which as we have seen, implies that their Higgs fields sum to zero. For example, $\phi_\alpha(\bar{e}) = -\phi_\alpha(e)$. Substituting $-\phi_\alpha$ for $\phi_\alpha$ in the Dirac equation (1.4), we see that *antiparticles propagate forward in time with positive energies*, whereas previously they have been regarded as propagating backward in time. (In the Weinberg–Salam model, initial and final states are interchanged for antiparticles. However, this is not required for the vector model.)

In summary, the vector model for electroweak interactions predicts the Weinberg angle, and also the conservation of baryon number, lepton number, and electric charge. However, at present there are some defects. The assignment of Higgs scalars to quark and lepton flavors do not now reflect differences in their families. For example, the muon conservation is not yet predicted. Also, the theory so far does not require the parity violation of weak interactions. Finally, in the vector model, the Higgs fields are dynamic. Therefore, we will be able to predict transition probabilities for the weak interactions as soon as a Higgs field Lagrangian is added to the Lagrangians for the spin-half particles (1.8) and for the gauge fields. The formulation of this Lagrangian is currently being studied.

In the remainder of this paper, we lay the foundation for the mathematical results used in the presentation of the vector model. In Sec. II we introduce isotropic Yang–Mills vector fields and discuss their properties; we present algebraic identities of the Cartan isomorphism in Sec. III; and we derive the isotropic vector wave equation which is equivalent to Dirac's equation for bispinors in Sec. IV.

## II. ISOTROPIC YANG–MILLS FIELDS

Let (E,H) be the real components of a complex three-dimensional vector field F on $R^4$ defined by

$$\mathbf{F} = \mathbf{E} + i\mathbf{H} .$$

It is well known that if E and H are regarded as the components of an electromagnetic field, then Lorentz transformations acting on F are represented by complex orthogonal $3\times3$ matrices.[22] Complex orthogonal matrices leave invariant the Euclidean quadratic form:

$$\mathbf{F}\cdot\mathbf{F} = F_x^2 + F_y^2 + F_z^2 = E^2 - H^2 + 2i\mathbf{E}\cdot\mathbf{H} . \tag{2.1}$$

For example, consider the Lorentz transformation that combines a boost $\eta$ and a rotation $\theta$ about the $x$ axis, given by the real $4\times4$ matrix

$$\begin{bmatrix} \cosh\eta & \sinh\eta & 0 & 0 \\ \sinh\eta & \cosh\eta & 0 & 0 \\ 0 & 0 & \cos\theta & -\sin\theta \\ 0 & 0 & \sin\theta & \cos\theta \end{bmatrix} . \tag{2.2}$$

By including boosts and rotations about all three axes, matrices of the form (2.2) generate the Lorentz group. Acting on complex vectors $\mathbf{F} = \mathbf{E} + i\mathbf{H}$ in $C^3$, the matrix (2.2) becomes the $3\times3$ complex orthogonal matrix with determinant 1,

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta + i\eta) & -\sin(\theta + i\eta) \\ 0 & \sin(\theta + i\eta) & \cos(\theta + i\eta) \end{bmatrix} , \tag{2.3}$$

and hence $\mathbf{E},\mathbf{H} \in R^3$ transform as electromagnetic fields. From (2.1), $E^2 - H^2$ and $\mathbf{E}\cdot\mathbf{H}$ are Lorentz scalars.

A complex vector F is called isotropic if $\mathbf{F}\cdot\mathbf{F}$ vanishes. We see from (1) that for isotropic vectors $\mathbf{F} = \mathbf{E} + i\mathbf{H}$, E and H are orthogonal and have equal lengths, in all reference frames.

More generally, if $\mathbf{F}_1$ and $\mathbf{F}_2$ are any three-dimensional complex vectors, then the complex orthogonal matrices leave invariant the Euclidean inner product $\mathbf{F}_1\cdot\mathbf{F}_2$. Thus, $\mathbf{F}_1\cdot\mathbf{F}_2$ is a scalar invariant under Lorentz transformations.

For three such vector fields $\mathbf{F}_1$, $\mathbf{F}_2$, $\mathbf{F}_3$ there are seven Lorentz scalar invariants, namely, $\mathbf{F}_j\cdot\mathbf{F}_k$ for $j,k = 1,2,3$, and also $\mathbf{F}_1\times\mathbf{F}_2\cdot\mathbf{F}_3$. The triplet $(\mathbf{F}_1,\mathbf{F}_2,\mathbf{F}_3)$ will be called *isotropic* if it satisfies the Lorentz invariant equations

$$\mathbf{F}_j\cdot\mathbf{F}_k = \lambda\delta_{jk} , \tag{2.4}$$

for complex scalar $\lambda$.

For isotropic complex vector triplets $(\mathbf{F}_1,\mathbf{F}_2,\mathbf{F}_3)$, we may define another complex scalar $\rho$ by the formula,

$$\rho = (\mathbf{F}_1\times\mathbf{F}_2\cdot\mathbf{F}_3)/\lambda . \tag{2.5}$$

It is easy from (4) and (5) to show that $\lambda = \rho^2$. Formula (4) then becomes

$$\mathbf{F}_j\cdot\mathbf{F}_k = \rho^2\delta_{jk} . \tag{2.6}$$

**Theorem:** If $(\mathbf{F}_1,\mathbf{F}_2,\mathbf{F}_3)$ is an isotropic complex vector triplet, then there exists a *unique real* orthonormal Lorentz basis

$$e_{(\alpha)} = (e^0_{(\alpha)},\mathbf{e}_{(\alpha)}) ,$$

with $\alpha = 0,1,2,3$ such that for $k = 1,2,3$

$$\mathbf{F}_k/\rho = e^0_{(0)}\mathbf{e}_{(k)} - e^0_{(k)}\mathbf{e}_{(0)} - i\mathbf{e}_{(0)}\times\mathbf{e}_{(k)} . \tag{2.7}$$

*Proof:* Set $\mathbf{f}_k = \mathbf{F}_k/\rho$, then from (5),

$$\mathbf{f}_j\cdot\mathbf{f}_k = \delta_{jk} \quad \text{and} \quad \mathbf{f}_1\times\mathbf{f}_2\cdot\mathbf{f}_3 = 1 . \tag{2.8}$$

The complex vector triplets $(\mathbf{f}_1,\mathbf{f}_2,\mathbf{f}_3)$ satisfying (2.8), comprise a single orbit of the Lorentz group [i.e., the isomorphic group of $SO(3,C)$ matrices]. Similarly, the real orthonormal Lorentz bases $e_{(\alpha)}$ also comprise a single orbit of the Lorentz group. Since by a straightforward derivation from (2.2) and (2.3), the map (2.7) commutes with Lorentz transformations, it suffices to prove that it is one to one.

By an $SO(3,C)$ transformation, any triplet $(\mathbf{f}_1,\mathbf{f}_2,\mathbf{f}_3)$ which satisfies (8) can be made to satisfy $(f_j)_k = \delta_{jk}$. From (2.7), since the $\mathbf{f}_k$ are now real vectors, we see then that $\mathbf{e}_{(0)}\times\mathbf{e}_{(k)} = 0$ for $k = 1,2,3$, which implies $\mathbf{e}_{(0)} = 0$. Since the $e_{(\alpha)}$ are orthonormal, $e_{(0)} = (1,0)$. Then from (2.7) and (2.8), we obtain $e_{(k)} = (0,\mathbf{f}_k)$. We conclude that the $e_{(\alpha)}$ are uniquely determined, so the map (2.7) is one to one. Q.E.D.

$SL(2,C)$ matrices[23] also act as formal gauge transformations on the triplet $(\mathbf{F}_1,\mathbf{F}_2,\mathbf{F}_3)$. These formal transformations do not involve the space-time coordinates; they just permute the vector fields $\mathbf{F}_1$, $\mathbf{F}_2$, and $\mathbf{F}_3$. For example, as discussed in Sec. I, the electromagnetic gauge transformations are elements of $SL(2,C)$ that act as formal "rotations" about the three-axis; i.e., they "rotate" $\mathbf{F}_1$ and $\mathbf{F}_2$ into each other, and leave $\mathbf{F}_3$ unchanged. Also as discussed in Sec. I, the gauge group $SL(2,C)$ may be extended to $SL(2,C)\times U_0(1)$, where elements of $U_0(1)$ map $\mathbf{F}_k$ to $\mathbf{F}_k e^{i\chi}$ for $k = 1,2,3$ ($\chi$ denotes a phase).

Triplets $(\mathbf{F}_1,\mathbf{F}_2,\mathbf{F}_3)$ which transform under Lorentz transformations and formal $SL(2,C)\times U_0(1)$ gauge transformations, we will call Yang–Mills triplets. These triplets may be classified with regard to their seven Lorentz scalar invariants. A Yang–Mills triplet is *isotropic* if as in (2.4)

$$\mathbf{F}_j\cdot\mathbf{F}_k = \lambda\delta_{jk},$$

which is also a gauge-invariant condition.

We conclude from (2.5) and (2.7) that every isotropic Yang–Mills triplet $(\mathbf{F}_1,\mathbf{F}_2,\mathbf{F}_3)$ uniquely determines a scalar $\rho$ and an orthonormal basis from which we define four *real* orthogonal currents $j_\alpha = |\rho|e_{(\alpha)}$. The $SL(2,C)$ gauge transformations leave $\rho$ unchanged and irreducibly act on the quadruplet $(j_0,j_1,j_2,j_3)$ as formal "Lorentz" transformations. The $U_0(1)$ gauge transformations change the phase of $\rho$, but leave the currents $j_\alpha$ unchanged. Thus, the combined gauge group $SL(2,C)\times U_0(1)$ acts irreducibly on the triplet $(\mathbf{F}_1,\mathbf{F}_2,\mathbf{F}_3)$, on the singlet $(\rho)$, and also on the quadruplet $(j_0,j_1,j_2,j_3)$.

We prove in the next section, using the Cartan map, that $\mathbf{F}_k$, $\rho$, and $j_{(\alpha)}$ comprise all possible bilinear invariants which can be associated with bispinors. Consequently, we show that the Cartan map gives an isomorphism between bispinors and isotropic Yang–Mills vector fields.

## III. THE EXTENDED CARTAN MAP

Let $\zeta = \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix} \in C^2$ be a spinor. The conjugate spinor associated with $\zeta$ is

$$\zeta^* = \begin{bmatrix} \bar{\zeta}_2 \\ -\bar{\zeta}_1 \end{bmatrix} \in C^2 ,$$

where the bar denotes complex conjugation. The map $\zeta \to \zeta^*$ is a bijection, since $\zeta = -\zeta^{**}$.

The Cartan map[24] is defined to be a bilinear map $b$ from $C^2 \times C^2$ into $C^4$ given by

$$b^\alpha(\zeta,\eta) = -(\sigma^\alpha\zeta)\cdot\bar{\eta}^*,$$

for $\alpha = 0,1,2,3$, and where $\sigma^\alpha = (I,\boldsymbol{\sigma})$, $\boldsymbol{\sigma} = (\sigma_x,\sigma_y,\sigma_z)$ are the Pauli spin matrices acting on $C^2$, $I$ is the identity matrix, and $\zeta$ and $\eta$ are spinors or conjugate spinors. The Cartan map allows us to construct all the "bilinear" (more precisely quadratic) invariants of a spinor $\zeta$. Since $b^0(\zeta,\zeta) = 0$, the nonvanishing invariants of $\zeta$ are given by

$$\mathbf{F} = i\mathbf{b}(\zeta,\zeta), \quad j = b(\zeta,\zeta^*),$$

where $\mathbf{b} = (b^1,b^2,b^3)$ and $b = (b^0,\mathbf{b})$.

It was shown[24] that $\mathbf{F}\in C^3$ satisfies the isotropic condition $\mathbf{F}\cdot\mathbf{F} = 0$. Moreover $\mathbf{F} = \mathbf{E} + i\mathbf{H}$ transforms under the Lorentz group action as an electromagnetic field $(\mathbf{E},\mathbf{H})$. Also, $j$ equals the probability current,

$$j^\alpha = \bar{\zeta}\sigma^\alpha\zeta,$$

which is a Lorentz four-vector. The isotropic vector $\mathbf{F} = \mathbf{E} + i\mathbf{H}$ determines $j^\alpha = (j^0,\mathbf{j})$ by the formulas

$$j^0 = |E|, \quad \mathbf{j} = (\mathbf{E}\times\mathbf{H})/|E|.$$

We showed that the map $\zeta \to \mathbf{F}$ gives an isomorphism between spinors and isotropic vector fields. Now, we extend this map to Dirac bispinors, mapping bispinors $\tilde{\psi}$ onto the isotropic Yang–Mills triplets $(\mathbf{F}_1,\mathbf{F}_2,\mathbf{F}_3)$ discussed in Sec. II.

A *bispinor* $\tilde{\psi} = (\zeta,\eta^*)\in C^4$ consists of a spinor $\zeta\in C^2$ and a conjugated spinor $\eta^*\in C^2$. Associated with $\tilde{\psi}$ is the *spinor pair* $\psi = (\zeta,\eta)$ and its *conjugate* $\psi^* = (\eta^*, -\zeta^*)$, where $\eta = -(\eta^*)^*$. The maps $\tilde{\psi}\to\psi$ and $\psi\to\psi^*$ are bijections, since $\eta = -\eta^{**}$ and $\psi = \psi^{**}$. Because of these bijections, bispinors $\tilde{\psi}$, spinor pairs $\psi$, and conjugate spinor pairs $\psi^*$ are *all equivalent ways* of expressing a Dirac bispinor. However, whereas SL(2,$C$) gauge transformations may be defined for spinor pairs $(\zeta,\eta)$, and also for the conjugate spinor pairs $(\eta^*, -\zeta^*)$, they are not defined as complex matrices acting on bispinors. We will show in Sec. IV that Dirac's equation has SL(2,$C$) gauge symmetry by writing it as an equation for spinor pairs, instead of bispinors (using the bijective map $\tilde{\psi}\to\psi$ from bispinors to spinor pairs.) [Note that for the spinor pair $\psi$, the electromagnetic gauge generator (via the map $\tilde{\psi}\to\psi$) becomes $\tau^3$. Hence, in the vector model, electromagnetic gauge transformations become the formal "rotations" about the three-axis.]

The extended Cartan map defined below commutes with both Lorentz and SL(2,$C$) gauge transformations. In order to view the gauge symmetry of the extended Cartan map, it is preferable to express it using spinor pairs, instead of using the equivalent bispinors.

*Definition:* The extended Cartan map is defined to be a bilinear map from $C^4\times C^4$ into $C^4\otimes C^4$ given by

$$B^\alpha_\beta(\psi,\chi) = (\sigma^\alpha\psi)\cdot\overline{(\tau_\beta\chi^*)}, \tag{3.1}$$

for $\alpha,\beta = 0,1,2,3$, and where $\sigma^\alpha = (I,\boldsymbol{\sigma})$, $\tau_\beta = (I,-\boldsymbol{\tau})$, $\boldsymbol{\sigma} = (\sigma_x,\sigma_y,\sigma_z)$ are Pauli spin matrices acting on $C^4$, $\boldsymbol{\tau} = (\tau^1,\tau^2,\tau^3)$ are the gauge matrices acting on $C^4$, and $\psi,\chi$ are either spinor pairs or conjugate spinor pairs. Note that whereas the Pauli matrices $\boldsymbol{\sigma} = (\sigma_x,\sigma_y,\sigma_z)$ generate the space rotation group, the gauge matrices $\boldsymbol{\tau} = (\tau^1,\tau^2,\tau^3)$ are the three generators for the gauge subgroup SU(2), which *formally "rotate"* the spinor pair. Moreover, $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ commute. We will adhere to the following notation:

$$\hat{\sigma}^\alpha = (I,-\boldsymbol{\sigma}) = \sigma_\alpha, \quad \hat{\tau}^\alpha = (I,-\boldsymbol{\tau}) = \tau_\alpha,$$
$$\hat{\sigma}_\alpha = (I,\boldsymbol{\sigma}) = \sigma^\alpha, \quad \hat{\tau}_\alpha = (I,\boldsymbol{\tau}) = \tau^\alpha.$$

A complete set of bilinear combinations of the components of the bispinor $\tilde{\psi}$ can be obtained from $B^\alpha_\beta(\psi,\psi)$ and $B^\alpha_\beta(\psi,\psi^*)$. Since,

$$B^0_k(\psi,\psi) = B^k_0(\psi,\psi) = 0,$$

for $k = 1,2,3$, the only bilinear (quadratic) invariants of $\tilde{\psi}$ which do not vanish are (see Table II in Sec. I)

$$\rho = iB^0_0(\psi,\psi), \quad \mathbf{F}_k = i\mathbf{B}_k(\psi,\psi), \quad j_\beta = B_\beta(\psi,\psi^*), \tag{3.2}$$

where $k = 1,2,3$, $\beta = 0,1,2,3$, $\mathbf{B}_\beta = (B^1_\beta,B^2_\beta,B^3_\beta)$, and $B_\beta = (B^0_\beta,\mathbf{B}_\beta)$. Note that whereas $\rho$ and $\mathbf{F}_k$ are complex, the $j_\beta$ are real.

It will be useful to write the four currents $j_\beta$ explicitly in terms of the spinor components of $\tilde{\psi} = (\zeta,\eta^*)$:

$$j^\alpha_0 = \bar{\zeta}\sigma^\alpha\zeta + \bar{\eta}\sigma^\alpha\eta, \quad j^\alpha_1 = -\bar{\zeta}\sigma^\alpha\eta - \bar{\eta}\sigma^\alpha\zeta,$$
$$j^\alpha_2 = i[\bar{\zeta}\sigma^\alpha\eta - \bar{\eta}\sigma^\alpha\zeta], \quad j^\alpha_3 = -\bar{\zeta}\sigma^\alpha\zeta + \bar{\eta}\sigma^\alpha\eta. \tag{3.3}$$

As in Sec. II, one can show that $j_\beta$ with $\beta = 0,1,2,3$ are a real orthogonal Lorentz basis for $R^4$, with scalar lengths equal to $|\rho|$, and with both $\rho$ and the $j_\beta$ determined by the $\mathbf{F}_k$. One may show that the map $\tilde{\psi}\to(\mathbf{F}_1,\mathbf{F}_2,\mathbf{F}_3)$ is an isomorphism from bispinors onto isotropic Yang–Mills triplets as a consequence of the following three lemmas.

*Lemma 1:* As defined by formulas (3.2), we have the following.

(a) $\rho$ is both a Lorentz and an SL(2,$C$) gauge scalar.

(b) The $\mathbf{F}_k$ transform as a Yang–Mills triplet for Lorentz and SL(2,$C$) gauge transformations.

(c) The $j_\beta$ transform as Lorentz four-vectors and as an SL(2,$C$) gauge quadruplet.

As with the SL(2,$C$) gauge transformations, chiral transformations commute with the Cartan map.

*Lemma 2:* For chiral gauge transformations, we have the following.

(a) $\rho$ undergoes a change of phase.

(b) $\mathbf{F}_1,\mathbf{F}_2,\mathbf{F}_3$ undergo a change of phase.

(c) $j_0,j_1,j_2,j_3$ do not change.

*Lemma 3:* Let $\psi = (\zeta,\eta)$ and $\psi' = (\zeta',\eta')$ be two pairs of spinors (or conjugate spinors). Then the following identities are true.

(a) $B^0_0(\psi,\psi)B^0_i(\psi,\psi') = -i\mathbf{B}_j(\psi,\psi)\cdot\mathbf{B}_k(\psi,\psi')$,

with $(ijk)$ taken in cyclic order, where $i,j,k = 1,2,3$ are the subscripts.

(b) $B^0_\alpha(\mathbf{v}\cdot\boldsymbol{\sigma}\psi,\psi') = \mathbf{v}\cdot\mathbf{B}_\alpha(\psi,\psi')$.

(c) $\mathbf{B}_\alpha(\mathbf{v}\cdot\boldsymbol{\sigma}\psi,\psi') = \mathbf{v}B^0_\alpha(\psi,\psi') + i\mathbf{v}\times\mathbf{B}^0_\alpha(\psi,\psi')$.

(d) $B^\alpha_0(\mathbf{v}\cdot\boldsymbol{\tau}\psi,\psi') = \sum_{j=1}^{3} v_j B^\alpha_j(\psi,\psi')$.

(e) $B_i^\alpha(\mathbf{v}\cdot\boldsymbol{\tau}\psi,\psi') = v_i B_0^\alpha(\psi,\psi') + i\epsilon_{ijk}v_j B_k^\alpha(\psi,\psi')$,

where $\mathbf{v}\in R^3$, $\alpha = 0, 1,2,3$ and the subscripts $i,j,k = 1,2,3$.

(f) $B_0^0(\psi',\psi) = B_0^0(\psi,\psi')$,
$\mathbf{B}_0(\psi',\psi) = -\mathbf{B}_0(\psi,\psi')$,
$B_k^0(\psi',\psi) = -B_k^0(\psi,\psi')$,
$\mathbf{B}_k(\psi',\psi) = \mathbf{B}_k(\psi,\psi')$,

for $k = 1,2,3$ (i.e., the bilinear maps, $B_0^0$ and $\mathbf{B}_k$ are symmetric, whereas $\mathbf{B}_0$ and $B_k^0$ are antisymmetric).

**Theorem**: The extended Cartan map $\tilde{\psi}\to(\mathbf{F}_1,\mathbf{F}_2,\mathbf{F}_3)$ from bispinsors onto isotropic Yang–Mills triplets is a *locally one to one (coordinate) map* from $C^4$ onto the four-(complex) dimensional manifold:

$$\bar{M} = \{(\mathbf{F}_1,\mathbf{F}_2,\mathbf{F}_3)\in C^9 | \mathbf{F}_j\cdot\mathbf{F}_k = \lambda\delta_{jk};\ j,k = 1,2,3,\ \lambda\in C\}.$$

The only identifications are $\tilde{\psi}$ with $-\tilde{\psi}$. The coordinate map commutes with both Lorentz and $SL(2,C)\times U(1)_0$ gauge transformations.

## IV. THE DIRAC EQUATION

Dirac's equation is usually stated as an equation for bispinors $\tilde{\psi} = (\zeta,\eta^*)$. However, the symmetry properties of Dirac's equation are more obvious when it is written as an equation for the spinor pair $\psi = (\zeta,\eta)$ obtained via the bijective map $\tilde{\psi}\to\psi$, which sends $(\zeta,\eta^*)\to(\zeta,\eta)$.

The Dirac equation for bispinors is given by

$$(P^0 - \mathbf{P}\cdot\boldsymbol{\sigma})\zeta = m\eta^*,\quad (P^0 + \mathbf{P}\cdot\boldsymbol{\sigma})\eta^* = m\zeta, \tag{4.1}$$

where $\tilde{\psi} = (\zeta,\eta^*)$ is a bispinor field, $P^\alpha = (P^0,\mathbf{P}) = ih\nabla^\alpha +$ potentials, $h$ is Planck's constant, and $m$ is the mass.

By conjugation, we get

$$(P^0 - \mathbf{P}\cdot\boldsymbol{\sigma})\zeta = m\eta^*,\quad (P^0 - \mathbf{P}\cdot\boldsymbol{\sigma})\eta = m\zeta^*, \tag{4.2}$$

and also the conjugate equations

$$(P^0 + \mathbf{P}\cdot\boldsymbol{\sigma})\zeta^* = m\eta,\quad (P^0 + \mathbf{P}\cdot\boldsymbol{\sigma})\eta^* = m\zeta. \tag{4.3}$$

Let us consider the spinor pair $\psi = (\zeta,\eta)$ and also the conjugate spinor pair $\psi^* = (\eta^*, -\zeta^*)$. Equations (4.2) and (4.3) can be written as

$$P^\alpha\sigma_\alpha\psi = -\phi^\alpha\hat{\tau}_\alpha\psi^*,\quad P^\alpha\hat{\sigma}_\alpha\psi^* = \phi^\alpha\tau_\alpha\psi, \tag{4.4}$$

where $\phi^\alpha = (0,0,0,-m)$ is the real quadruplet of Higgs scalars, $\sigma^\alpha = (I,\boldsymbol{\sigma})$ are Pauli matrices, $\hat{\sigma}^\alpha = (I,-\boldsymbol{\sigma})$, $\tau^\alpha = (I,\boldsymbol{\tau})$ are gauge matrices, and $\hat{\tau}^\alpha = (I,-\boldsymbol{\tau})$.

It is evident that Eqs. (4.4) are invariant under Lorentz transformations provided that $\phi_\alpha$ for each $\alpha = 0,1,2,3$ transforms as a Lorentz scalar. Moreover, it is straightforward to show that Eqs. (4.4) are invariant under $SL(2,C)$ gauge transformations provided that $\phi_\alpha$ transforms as a (real) $SL(2,C)$ gauge quadruplet.

Using the fact that $\sigma^\alpha$ and $\tau^\alpha$ commute, we can derive the Klein–Gordon equation, which in the case of free particles is given by

$$p^\alpha p_\alpha\psi = -(\phi^\alpha\phi_\alpha)\psi.$$

Therefore, the mass is given by

$$M = \sqrt{-\phi^\alpha\phi_\alpha}.$$

*Lemma 4.* The Dirac equation,

$$P^\alpha\sigma_\alpha\psi = -\phi^\alpha\hat{\tau}_\alpha\psi^*$$

with

$$P_\alpha = ih\nabla_\alpha + e_0 V_\alpha^0 + e\sum_{k=1}^3 V_\alpha^k\tau_k,$$

may be derived by applying the Euler–Lagrange equations to the Lagrangian

$$L = \mathrm{Re}\{(P^\alpha\sigma_\alpha\psi)\cdot\overline{(\phi^\beta\tau_\beta\psi)} + M^2\psi\cdot\bar{\psi}^*\}/M.$$

Moreover, we have the following.

(a) The Lagrangian $L$ is an invariant scalar under both Lorentz and $SL(2,C)$ gauge transformations.

(b) The conserved Noether currents derived from $L$ [one for each generator of $SL(2,C)$] are given by

$$J_k = \phi_0 j_k - \phi_k j_0 - i\epsilon_{kmn}\phi_m j_n,$$

with $k = 1,2,3$.

(c) The kinetic part of the Lagrangian $L$ is an invariant scalar under neutral (chiral) gauge transformations.

(d) The neutral (chiral) current is given by

$$K = \phi^\alpha j_\alpha,$$

which corresponds to the single generator of $U_0(1)$.

(e) The interaction part of the Lagrangian $L$ is given by

$$L_I = \frac{e_0}{M}V_\alpha^0 K^\alpha + \frac{e}{M}\mathrm{Re}\sum_{k=1}^3 V_\alpha^k J_k^\alpha, \tag{4.5}$$

where $e_0$ and $e$ are the neutral and electric charges, respectively, and the $V_\alpha^\beta$ are the Yang–Mills (complex) potentials.

(f) With the conventional choice of Higgs fields $\phi_\alpha = (0,0,0,m)$, the Lagrangian $L$ reduces to the usual Lagrangian for Dirac's equation (4.1).

Note that in formula (4.5), both $K$ and $V^\circ$ are real. However, both $J_k$ and $V^k$ are complex for $k = 1,2,3$, which is a consequence of the gauge group $SL(2,C)$. Note also that (5) is equivalent to

$$L_I = \frac{e_0}{M}V_\alpha^0 K^\alpha + \frac{e}{M}\sum_{k=1}^3(\mathrm{Re}\,V_\alpha^k)(\mathrm{Re}\,J_k^\alpha) - \frac{e}{M}\sum_{k=1}^3(\mathrm{Im}\,V_\alpha^k)(\mathrm{Im}\,J_k^\alpha). \tag{4.6}$$

Formula (4.6) shows clearly that the interaction is completely described by the *seven* real currents $\mathrm{Re}\,J_k$, $\mathrm{Im}\,J_k$, and $K$ and the *seven* real potentials $\mathrm{Re}\,V^k$, $\mathrm{Im}\,V^k$, and $V^0$ corresponding to the seven generators of $SL(2,C)\times U_0(1)$.

In the remainder of this section, we will derive the vector equivalent of the Dirac equation.

Application of the extended Cartan map $\mathbf{B}_i$ to the first equation of (4.4) gives, using Lemma 3(e),

$$\mathbf{B}_i(\psi,P_\alpha\sigma^\alpha\psi) = -\mathbf{B}_i(\psi,\phi_\alpha\hat{\tau}^\alpha\psi^*) = -\mathbf{J}_i. \tag{4.7}$$

Similarly, the left-hand side of (7) becomes, using Lemma 3, parts (c) and (a),

$$\begin{aligned}\mathbf{B}_i(\psi,P_\alpha\sigma^\alpha\psi) &= -iD_\alpha S^\alpha\mathbf{F}_i + B_i^0(\psi,\mathbf{P}\psi)\\ &= -iD_\alpha S^\alpha\mathbf{F}_i - (i/\rho)\mathbf{F}_j\cdot\mathbf{B}_k(\psi,\mathbf{P}\psi)\\ &= -iD_\alpha S^\alpha\mathbf{F}_i + (\mathbf{D}\mathbf{F}_k)\cdot\mathbf{F}_j/\rho\\ &= -iD_\alpha S^\alpha\mathbf{F}_i - (\mathbf{D}\mathbf{F}_j)\cdot\mathbf{F}_k/\rho,\end{aligned} \tag{4.8}$$

where $S^\alpha = (I,S)$ are the Proca spin-one matrices, $D^\alpha = i(h/2)\nabla^\alpha +$ potentials, $\mathbf{D} = (D_1,D_2,D_3)$, and the $(ijk)$ subscripts are taken in cyclic order. Note that the Cartan map $\mathbf{B}_k(\psi,\chi)$ for $k = 1,2,3$ is symmetric in the variables $\psi$ and $\chi$, and in commuting $P^\alpha$ into $D^\alpha$, Planck's constant $h$ becomes $h/2$. Thus, we have proved the following theorem.

**Theorem**: Via the Cartan map, Dirac's equation for bispinors is equivalent to

$$iD_\alpha S^\alpha \mathbf{F}_i + (\mathbf{D}\mathbf{F}_j)\cdot\mathbf{F}_k/\rho = \mathbf{J}_i\ ,$$

where the subscripts $(ijk)$ are taken in cyclic order.

[1]F. Reifler, J. Math. Phys. **25**, 1088 (1984).
[2]Isotropic vectors $\mathbf{F} \in C^3$ satisfy the condition $\mathbf{F}\cdot\mathbf{F} = 0$.
[3]R. Geroch, J. Math. Phys. **9**, 1739 (1968).
[4]J. Klauder, J. Math. Phys. **5**, 1204 (1964); E. T. Whittaker, Proc. R. Soc. London Ser. A **158**, 38 (1937); Y. Takahashi and K. Okuda, Fortschr. Phys. **31**, 511 (1983).
[5]If $\eta = (\eta_1,\eta_2) \in C^2$ is a spinor, its conjugate is defined by $\eta^* = (\overline{\eta_2}, -\overline{\eta_1})$. Note that $\eta = -\eta^{**}$, so that the conjugation operation has an inverse.
[6]K. Huang, *Quarks, Leptons, and Gauge Fields* (World Scientific, Singapore, 1982).
[7]SU(2) gauge symmetry is used for the spinor pair of a *trispinor* in the Weinberg–Salam model, but *not* for bispinors.
[8]More precisely, the Noether currents equal $(e/M)\,\mathrm{Re}\,J_k$. See Lemma 4(e).
[9]This is by analogy to the Weinberg–Salam model which uses the symmetry group $\mathrm{SU}(2)\times\mathrm{U}(1)$, where U(1) is the "neutral" gauge.
[10]The Lagrangian $L$ may be made invariant under $\mathrm{SL}(2,C)\times\mathrm{U}(1)_0$ by replacing $B^0_0(\psi,\psi)$ in the mass term with its absolute value $|B^0_0(\psi,\psi)|$. However, this change in the Lagrangian slightly alters the Dirac equation.
[11]Also in the Weinberg–Salam model, the neutral current is identical to the chiral current if the Weinberg angle $\theta_w$ satisfies $\sin^2\theta_w = \frac{1}{4}$. (See Table V.)
[12]Note that $\sigma^\alpha$ and $\hat{\tau}^\alpha$ commute.
[13]Ref. 6, p. 107.
[14]C. Quigg, *Gauge Theories of the Strong, Weak, and Electromagnetic Interactions* (Benjamin, London, 1983), page 188.
[15]Ref. 6, p. 106.
[16]Ref. 6, p. 108.
[17]Ref. 6, p. 109.
[18]Ref. 6, p. 109.
[19]Ref. 6, pp. 111 and 114.
[20]Ref. 14, p. 16.
[21]Matter fields are presently described by a state vector consisting of 93 spinors, ranging over all known quarks and leptons. See H. Georgi, *Lie Algebras in Particle Physics* (Benjamin, London, 1982), p. 162.
[22]L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Pergamon, New York, 1971).
[23]Every representation of $\mathrm{SO}(3,C)$ is also a representation of its covering group $\mathrm{SL}(2,C)$. *Either* group may be regarded as the gauge group. We choose to call the gauge group $\mathrm{SL}(2,C)$ to conform with spinors.
[24]F. Reifler, J. Math. Phys. **25**, 1088 (1984); and also E. Cartan, *Theory of Spinors* (MIT, Cambridge, MA, 1966).

# Expansion in condensates

J. Chakrabarti
*Institute of Theoretical Science, University of Oregon, Eugene, Oregon 97403*

M. Sajjad Zahir[a)]
*Department of Computer and Information Science, University of Oregon, Eugene, Oregon 97403*

(Received 5 June 1984; accepted for publication 3 August 1984)

We show that the product of local current operators in quantum chromodynamics (QCD), when expanded in terms of condensates, such as $\bar{\psi}\psi$, $G^a_{\mu\nu}\,G^a_{\mu\nu}$, $\bar{\psi}\Gamma\psi\bar{\psi}\Gamma\psi$, $f_{abc}G^a_{\mu\nu}G^b_{\nu\alpha}G^c_{\alpha\mu}$, etc., yields a series in Planck's constant. This, however, provides no hint that the higher terms in such an expansion may be less significant.

## I. INTRODUCTION

The ground state of QCD has fermions and gauge boson condensates such as $\bar{\psi}\psi$, $G^a_{\mu\nu}\,G^a_{\mu\nu}$, etc. These condensate operators appear in the operator product expansion[1] (OPE) of the products of local currents such as $\bar{\psi}^a(x)\gamma_\mu\psi_a(x)$ (where $a$ is the color index) in addition to the usual perturbative effects[2] due to the presence of the colored gauge field interactions. We present a systematic procedure to be useful in carrying out the above expansion in the momentum space.[3] Alongside, a set of suitable approximation rules is also presented.

It has been found useful, for numerous reasons, to analyze the OPE of the products of local current operators.[4] For one, it is possible to estimate the magnitude of the condensates ($\bar{\psi}\psi$, $G^a_{\mu\nu}\,G^a_{\mu\nu}$,...) from such an expansion.[5] Since this expansion series has an infinite number of terms, its analysis poses some difficulties especially because the expansion parameter cannot be regarded as small. It is therefore difficult to argue that it is meaningful to terminate such an expansion after the first few terms. We show that the expansion in condensates is a series in Planck's constant $\hbar$. Since Planck's constant has dimension, this provides no clue whether the higher-dimensional condensates are less significant. However, such expansions in $\hbar$ frequently have been carried out in the calculations of effective action where it has been presumed to make sense.[6]

## II. CONDENSATE EXPANSION FOR QCD

The fermion part of the QCD Lagrangian in the presence of a background color gauge field is

$$L_F = \bar{\psi}[i\not{D} - m]\psi, \tag{1}$$

where $\not{D}_\mu$ includes a background field $A_\mu$ as

$$D^{ab}_\mu = \partial_\mu\delta^{ab} - i\,(g/2)\cdot\lambda^{ab}_c\cdot A^c_\mu, \tag{2}$$

and $\psi$ is a column matrix. The background field $A_\mu$ corresponds to the presence of gluon condensates. Therefore, the fermion propagator satisfies

$$[i\gamma_\mu\,(\partial_\mu\delta_{ab} - (ig/2)\lambda^c_{ab}A^c_\mu) - m]S_F(x,x') = \delta(x,x'). \tag{3}$$

It is convenient to specialize to coordinate gauge $[x_a A^a(x) = 0]$ and write[7]

$$A^a_\mu(x) = -\tfrac{1}{2}[G^a_{\mu\nu}(0)x^\nu + \cdots]. \tag{4}$$

We go to momentum space by writing

$$S_F(x,x') = \int \frac{d^4k}{(2\pi)^4}\,e^{-ik(x,x')}S_F(k). \tag{5}$$

Substituting (4) and (5) in Eq. (3) and solving for $S_F(k)$ we get (restoring $\hbar$ explicitly)

$$S^{ab}_F(k) = (1/k^2)[\not{k}\delta^{ab} + (ig\hbar/4)\not{k}\lambda^{al}_c\gamma_\sigma\,G^{\sigma\tau}_c\,\partial_\tau S^{lb}_F(k)], \tag{6}$$

where $\partial_\tau$ is a derivative with respect to the momentum variable $k$. Since the second term on the right-hand side has an explicit factor of $\hbar$, it is possible to make successive interations to arrive at a series in Planck's constant. Let us make the lowest approximation and write

$$S^{(0)}_F = \delta_{ab}\not{k}/k^2. \tag{7}$$

Therefore,

$$S^{(1)}_F(k) = \frac{\not{k}}{k^2}\delta^{ab} + \frac{ig\hbar}{4}\frac{\not{k}}{k^2}\lambda^{ab}_c\gamma_\sigma G^{\sigma\tau}_c\,\partial_\tau\left(\frac{\not{k}}{k^2}\right). \tag{8}$$

We are interested in evaluating the polarization tensor $\Pi_{\mu\nu}$ defined as

$$\Pi_{\mu\nu}(q^2) = i\int d^4x\,e^{iq(x/\hbar)}\langle 0|TJ_\mu(x)J_\nu(0)|0\rangle, \tag{9}$$

where, for simplicity we chose

$$J_\mu(x) = \bar{\psi}^a(x)\gamma_\mu\psi_a(x), \tag{10}$$

and $a$ is the color index.

The procedure we outline will go through the other types of currents as well as can be checked by an explicit calculation. It is straightforward to check that the propagators for gluons can also be written as an expansion in Planck's constant in the presence of a background gluonic field.

Since QCD also has a fermionic condensate (chiral symmetry breaking), it is necessary to expand the fermionic field by writing

$$\Psi^a(x) = \phi^a(x) + f^a(x), \tag{11}$$

where $\phi^a(x)$ is regarded as originating from the fermion condensate, and $f^a(x)$ as the fluctuations. Substituting Eq. (11) in (9) we get

[a)] Permanent address: Department of Physics, McGill University, Montreal, Quebec, Canada H3A 2T8.

 0022-2488/85/030551-03$02.50 © 1985 American Institute of Physics 

$$\Pi_{\mu\nu}(q^2)$$
$$= i\int d^4x\, e^{iqx/\hbar}\langle 0|T\bar{f}^{\,a}(x)\gamma_\mu f_a(x)\bar{f}^{\,b}(0)\gamma_\nu f_b(0)|0\rangle$$
$$+ i\int d^4x\, e^{iqx/\hbar}$$
$$\times\langle 0|T\bar{f}^{\,a}(x)\gamma_\mu\phi_a(x)\bar{\phi}^{\,b}(0)\gamma_\mu f_b(0)|0\rangle$$
$$+ i\int d^4x\, e^{iqx/\hbar}$$
$$\times\langle 0|T\bar{\phi}^{\,a}(x)\gamma_\mu f_a(x)\bar{f}^{\,b}(0)\gamma_\nu\phi_b(0)|0\rangle$$
$$\equiv\Pi^1_{\mu\nu}+\Pi^2_{\mu\nu}+\Pi^3_{\mu\nu} \tag{12}$$

where the superscripts in $\Pi$ simply represent the three terms appearing in the preceding expression. The other terms in the expansion can be shown not to contribute in the limit of large $q^2$ (short distance). The first term in this expansion can be easily evaluated by carrying out a Wick expansion of the time-ordered product. Thus,

$$\Pi^1_{\mu\nu}(q^2) = -i\int d^4x\, e^{iqx/\hbar}\langle 0|T\bar{f}^{\,a}(x)f_b\langle 0|0\rangle$$
$$\times\gamma_\mu\langle 0|T\bar{f}_b(0)f_a(x)|0\rangle\gamma_\nu. \tag{13}$$

Since

$$\langle 0|T\bar{f}(y)f(y')|0\rangle = (\hbar/i)S_F(y'-y), \tag{14}$$

we get, by substituting (8) and (14) in (13), an expansion in Planck's constant. The first term (independent of the gluon condensate) is

$$\Pi^1_{\mu\nu} = -\frac{\ln q^2}{4\pi^2\hbar^2}[q_\mu q_\nu - q_{\mu\nu}q^2] + \cdots, \tag{15}$$

where we have disregarded the log-divergent term because it does not depend on $q^2$. Such a term, when differentiated with respect to $q^2$ (as would have happened if we wished to deal with the Borel–transformed series), drops out. In the above calculations, we have also neglected the masses of the quarks. Otherwise, corrections dependent on the quark masses would appear.

When $S_f(k)$ from Eq. (8) is substituted in (13), terms depending on the gluon condensate also appear. The first nonzero contribution is given by (we have carried out the $d^4x$ and one of the momentum integrations)

$$-\frac{ig^2}{16}\int\frac{d^4R}{(2\pi\hbar)^4}e^{igx/\hbar}\left[\frac{\not k}{k^2}\gamma_\sigma\lambda^{\,c}_{ab}\partial_\tau\frac{\not k}{k^2}\right]$$
$$\times\gamma_\mu\left[\frac{(\not q+\not k)}{(q+k)^2}\gamma_k\lambda^{\,d}_{ba}\partial_\xi\left(\frac{\not q+\not k}{(q+k)^2}\right)\right]\gamma_\nu G^{\sigma\tau}_d G^{\kappa\xi}_d. \tag{16}$$

It is tedious but straightforward to evaluate the above object and get

$$\Pi^1_{\mu\nu} = -\frac{\ln q^2}{4\pi^2\hbar^2}(q_\mu q_\nu - q_{\mu\nu}q^2)$$
$$+\frac{g^2}{48\pi^2q^4}G^{c\alpha\beta}G^{c\alpha\beta}(q_\mu q_\nu - q_{\mu\nu}q^2)+\cdots. \tag{17}$$

To evaluate the higher-order terms in gluon condensates it is required to expand the propagator by using the iterative equation (6) by writing

$$S^{(2)}_F(k) = (\not k/k^2)\delta^{ab} + \frac{ig\hbar}{4}\frac{\not k}{k^2}\lambda^{\,ab}_c\gamma_\sigma G^{\sigma\tau c}\partial_\tau S^{(1)}_F(k). \tag{18}$$

It can be checked easily, that the $f_{abc}G^a_{\alpha\beta}G^b_{\beta\delta}G^c_{\delta\alpha}$ condensate term goes with a factor of $\hbar$ and the quartic gluon condensate goes with $\hbar^2$.

Next, we need to expand the expression for $A^a_\mu(x)$ in terms of $G^a_{\mu\nu}(x)$ in the coordinate gauge around $G^a_{\alpha\beta}(0)$. When equations of motion are used for the $G_{\mu\beta}$'s we obtain a series of expansions in $\hbar$ for $\Pi_{\mu\nu}$.

Since $\Pi_{\mu\nu}{}^2$ and $\Pi_{\mu\nu}{}^3$ yield identical results, it is sufficient to consider only one of them:

$$\Pi^{(2)}_{\mu\nu} = i\int d^4x\, e^{iqx/\hbar}$$
$$\times\langle 0|T\bar{f}^{\,a}(x)\gamma_\mu\phi_a(x)\bar{\phi}^{\,b}(0)\gamma_\nu f_b(0)|0\rangle$$
$$= -i\int d^4x\, e^{iqx/\hbar}\langle 0|T\bar{f}^{\,a}(x)f_b(0)|0\rangle$$
$$\times\gamma_\mu\bar{\phi}^{\,b}(0)\phi_a(x)\gamma_\nu. \tag{19}$$

For $\langle 0|T\bar{f}^{\,a}(x)f_b(0)|0\rangle$, we use Eq. (8) and (14) and expand $\phi_a(x)$ around $x=0$:

$$\phi_a(x) = \phi_a(0) + x^\alpha\partial_\alpha\phi + \cdots. \tag{20}$$

It is necessary to retain the second term and integrate over $x$ if the conserved $(q_\mu q_\nu - q_{\mu\nu}q^2)$ structure is to emerge. The first term is easy to evaluate. Note the $\bar{\phi}^{\,b}(0)\phi_a(0)$ is normalized as

$$\bar{\phi}^{\,b}(0)\phi_a(0) = N\delta_{ab}, \tag{21}$$

where $N=\frac{1}{12}$ to take account of three degrees of color and four of spin.

For the second term, $x^\alpha$ may be brought out of the integral by writing it as a derivative with respect to $q$. Inside the integral, the term $\bar{\phi}^{\,b}\partial_\alpha\phi_a$ may be reduced as

$$\bar{\phi}^{\,b}\partial_\alpha\phi_a = \tfrac{1}{4}\bar{\phi}^{\,b}\mathrm{Tr}[\gamma_\alpha\gamma_\beta]\partial^\beta\phi_a. \tag{22}$$

Using the equations of motion and normalization (21), we get

$$\Pi^{(2)}_{\mu\nu}+\Pi^{(3)}_{\mu\nu} = [2m\hbar\langle\bar{\phi}\phi\rangle/q4](q_\mu q_\nu - q_{\mu\nu}q^2). \tag{23}$$

We observe the explicit appearance of $\hbar$ in the numerator.

We can now proceed to include the gluon condensate-dependent term in the propagator and generate higher-dimensional condensates such as $\bar{\phi}\sigma_{\mu\nu}\lambda^{\,a}\phi G^{\mu\nu a}$, which goes like $\hbar^2$.

So far, we have not dealt with the perturbative expansion. The perturbative series is obtained in the usual manner by noting that

$$i\int \mathrm{d}^4x\, e^{iqx/\hbar}\langle 0|TJ_\mu(x)J_\nu(0)|0\rangle_{\text{Heisenberg}}$$
$$= i\int d^4x\, e^{iqx/\hbar}$$
$$\times\langle 0|TJ_\mu(x)J_\nu(0)e^{(i/\hbar)\int L_{\text{int}}\,d^4y}|0\rangle_{\text{interaction}}. \tag{24}$$

Note that $L_{\text{QCD}}$ in our case is arrived at by shifting the gluon field $B^a_\mu \to A^a_\mu + b^a_\mu$ and the fermion field $\psi_a \to \phi_a + f_a$. In the unshifted Lagrangian, $b^a_\mu$ and $f_a$ are the gluonic and fermionic fluctuations, respectively.

We have presented a systematic way of evaluating the condensate terms in momentum space in the operator product expansion of local current operators. We note that such an expansion corresponds to a series in Planck's constant. Such expansions in Planck's constant have been used extensively in computations of effective actions.

## ACKNOWLEDGMENTS

We acknowledge discussions with Professor N. Deshpande, Professor R. C. Hwa, and Professor M. Moravcsik. We are grateful for Dr. L. P. S. Singh and Dr. X. Tata for many valuable suggestions.

The work of J. C. was supported by the U.S. Department of Energy under Contract No. DE-AT06-76ER7004.

[1]K. G. Wilson, Phys. Rev. **179**, 1499 (1969).
[2]M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, Nucl. Phys. B **147**, 385, 448 (1979).
[3]A. V. Smilga, Yad. Fiz. **35**, 473 (1982).
[4]B. L. Ioffe and A. V. Smilga, Phys. Lett. B **114** 353 (1982); V. S. Mathur and M. Yamawaki, Phys. Rev. D **28**, 669 (1983).
[5]J. Chakrabarti, Phys. Lett. B **123**, 339 (1983).
[6]J. Chakrabarti, Phys. Rev. D **28**, 2657 (1983).
[7]This gauge has been used by many authors. See J. Schwinger, *Particles, Sources, and Fields* (Addison-Wesley, Reading, MA 1970), p. 271. See also C. Cronström, Phys. Lett. B **90**, 267 (1980).

# Scaling properties of total energy of heavy positive ions in $d$-dimensions

N. H. March
*Theoretical Chemistry Department, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, England*

(Received 16 May 1984; accepted for publication 3 August 1984)

Using density functional theory in the asymptotic limit of very heavy positive ions, it is demonstrated that the well-known scaling property of the total energy $E_3(Z, N)$ of an ion with $N$ electrons and atomic number $Z$ is a special case of the $d$-dimensional result $E_d(Z, N) = Z^{(4+4d-d^2)/d(4-d)} f_d(N/Z): d \neq 4$.

## I. INTRODUCTION

Interactions between charges in dimensions $d$ other than three have been of interest over a long period. Thus Lenard[1] studied a one-dimensional charged gas with linear $r_{ij}$ interaction between charges $i$ and $j$, while Dyson[2] was concerned with the two-dimensional logarithmic interaction. Much later, Parrinello and March[3] studied the thermodynamics of electron crystallization in $d$-dimensions.

Recently Pucci and March[4] have pointed to the interest in relations between total energy $E$ and chemical potential $\mu$ in apparently complex molecular systems such as the linear polyacenes. Notwithstanding such apparent complexity, they have demonstrated from approximate theories that $E/N\mu$ tends to a value near to $\frac{1}{2}$ as the number of rings tend to infinity, a result they conjecture to be dominated by the $d = 2$ dimensionality. The potential importance of dimensionality in bound-state formation in molecules has also been explored to some extent.[5,6] In Ref. 5, the importance of Hartree self-consistency was emphasized. Since, however, even in the asymptotic limit of large numbers of electrons $N$ to which the present paper is directed, the treatment of molecular systems by analytical, as opposed to numerical, methods remains difficult, we have been motivated by the above considerations to first consider the role of dimensionality in the Hartree self-consistent field theory of heavy positive atomic ions.

## II. CHARACTERISTIC LENGTHS IN ELECTRON DISTRIBUTION IN $d$-DIMENSIONS

As is well known, the length scale of a heavy atom[7] varies proportional to $Z^{-1/3}$. This is derived from the Thomas–Fermi theory, which is known to yield asymptotically correct properties in three dimensions[8] in the limit of large numbers of electrons $N$.

Therefore, to establish the length scale in $d$-dimensions, we turn to the work of Kventsel and Katriel[9] on the $d$-dimensional Thomas–Fermi equation. Writing the maximum local momentum $p_f(r)$ through the relation

$$p_f^2(r)/2m = \mu + e\Phi(r) \equiv e\Psi(r), \tag{2.1}$$

these workers obtain the differential equation obeyed by $\Psi$, which differs from the electrostatic potential $\Phi$ only through a constant

$$\frac{\partial^2\Psi}{\partial r^2} + \frac{(d-1)}{r}\frac{\partial\Psi}{\partial r} = \beta_d \Psi^{d/2}, \tag{2.2}$$

$$\beta_d = \frac{8\pi e(\sqrt{2\pi m e}/h)^d}{\Gamma(d/2+1)}.$$

Taking the Thomas–Fermi density $\rho(r)$, with the boundary condition that as $r \to 0$,

$$\Psi(r) \to [4\pi Ze/(d-2)\Omega_d] r^{2-d}, \quad d \geqslant 3, \tag{2.3}$$

with $\Omega_d = d\pi^{d/2}/\Gamma(d/2+1)$, for a point charge of magnitude $Ze$, Kventsel and Katriel[9] point out that the normalization integral $\int \rho(r) d\tau$ diverges for $d \geqslant 4$, for such singular potentials as given by Eq. (2.3). We assume below that such divergence can be "cured" by introducing finite nuclei, which, because of the totally different length scales of nuclei and atoms, could not significantly influence the length scaling of interest to us here.

Thus, we shall write first

$$\Psi(r) = [4Ze\pi r^{2-d}/(d-2)\Omega_d]\chi_d(r) \equiv D_d \chi_d(r) r^{2-d}, \tag{2.4}$$

where $\chi_d(r)$ tends to unity as $r$ tends to zero from the boundary condition (2.3). The differential equation for $\chi_d(r)$ is readily shown to take the form

$$\frac{r^2\partial^2\chi_d}{\partial r^2} + (3-d) r \frac{\partial\chi_d}{\partial r} = \beta_d D^{(d/2-1)} \chi_d^{\,d/2} r^{(4-d)d/2}. \tag{2.5}$$

As the final step in establishing the length scale of the $d$-dimensional electron cloud, we write

$$r = b_d x. \tag{2.6}$$

Then the choice of $b_d$ as proportional to $Z^{(2-d)/d(4-d)}$ removes the $Z$ dependence completely from the differential equation for $\chi_d$. This establishes the length scale therefore, as desired, and yields of course, the usual three-dimensional result $b_3 \propto Z^{-1/3}$.

## III. CHEMICAL POTENTIAL AND TOTAL ENERGY SCALING

As in three dimensions, $d$-dimensional heavy positive ions have a finite semiclassical radius, say $r_d$. Outside $r_d$ therefore, the ion has an electrostatic potential which is the same as though the charge at the origin were $(Z-N)e$. Since

 0022-2488/85/030554-02$02.50 © 1985 American Institute of Physics 

$\rho(r_d) = 0$, by definition of the semiclassical radius, and $\rho \propto p_f{}^d$ from the usual phase-space arguments, it follows that $p_f(r_d) = 0$. Thus, from Eq. (2.1) we have

$$\mu = -e\Phi(r_d^+) \tag{3.1}$$

and hence, from the point charge form (2.3)

$$\mu = [-4\pi(Z-N)e^2/(d-2)\Omega_d](r_d)^{2-d}, \quad d \geqslant 3. \tag{3.2}$$

This immediately leads to the desired scaling property of the chemical potential as

$$\mu = Z(b_d)^{2-d}F_d(N/Z) \tag{3.3}$$

using the boundary condition dependent only on $N/Z$ as given in Ref. 9. Substituting the result below Eq. (2.6) for the $Z$ dependence of $b_d$ yields

$$\mu = Z^{4/d(4-d)}h_d(N/Z). \tag{3.4}$$

Of course, a full theory of $h_d(N/Z)$ in Eq. (3.4) will eventually require solution by numerical methods of the usually nonlinear equation (2.5).

The final step to arrive at the scaling of the total energy $E_d(Z, N)$ of the $d$-dimensional positive ion can be made in the asymptotic limit under consideration here by using the thermodynamic relation[10]

$$\mu = \left(\frac{\partial E}{\partial N}\right)_Z, \tag{3.5}$$

which yields to leading order

$$E_d(Z, N) = Z^{(4+4d-d^2)/d(4-d)}f_d(N/Z), \quad d \geqslant 4. \tag{3.6}$$

Equation (3.6) is the main result of this paper. The substitution of $d = 3$ reduces Eqs. (3.4) and (3.6) to well-established results in the self-consistent field theory of heavy positive ions.[10] The case $d = 4$ is anomalous,[11] as the virial theorem[3] gives immediately that $E_4(Z, N) = 0$.

Consideration is presently being given to the possible generalization of these scaling properties to some multicenter problems of current interest in molecular physics.

## ACKNOWLEDGMENTS

The opportunity and the stimulus to carry out the work reported here arose as a result of a visit to the University of California at Santa Barbara. It is a pleasure to thank Professor J. R. Schrieffer for his very kind hospitality.

[1] A. Lenard, J. Math. Phys. **2**, 682 (1961).
[2] F. J. Dyson, J. Math. Phys. **3**, 140 (1962).
[3] M. Parrinello and N. H. March, J. Phys. C **9**, L147 (1976).
[4] R. Pucci and N. H. March, Phys. Lett. A **94**, 63 (1983); **96**, 105 (1983).
[5] N. H. March, Phys. Lett. A **99**, 361 (1983).
[6] N. H. March and G. R. Freeman, Phys. Chem. Liq. **13**, 279 (1984).
[7] See, for example, L. I. Schiff, *Quantum Mechanics* (McGraw-Hill, New York, 1949).
[8] E. H. Lieb, Rev. Mod. Phys. **53**, 603 (1981).
[9] G. F. Kventsel and J. Katriel, Phys. Rev. A **24**, 2299 (1981).
[10] See, for example, *Theory of the Inhomogeneous Electron Gas*, edited by S. Lundqvist and N. H. March (Plenum, New York, 1983).
[11] The cases $d = 1$ and 2, corresponding to linear and logarithmic potentials, also appear to require separate treatment. Since however, we regain the scaling properties (3.4) and (3.6) we shall not go into further detail.

# Exact calculation of the lineshape and the scattering operator for a model of a two-level atom interacting with a continuous spectrum of radiation

Russell Davidson
*Department of Economics, Queen's University, Kingston, Ontario K7L 3N6, Canada*
John J. Kozak
*Department of Chemistry and Radiation Laboratory, University of Notre Dame, Notre Dame, Indiana 46556*

(Received 31 May 1984; accepted for publication 21 September 1984)

In this paper we calculate exactly the lineshape for a model of an excited two-level atom in interaction with a continuous spectrum of radiation for the problem of spontaneous emission. Specifically, for the case of a $d = 1$ radiation field, we use the exact results reported in our earlier work [J. Math. Phys. **14**, 414, 423 (1973)] for the probability $\rho(\tau)$ of the atom's being in the excited state at time $\tau$ to obtain an analytic expression for the lineshape of the emitted radiation (i.e., the photon). We also calculate the lineshape using two Wigner–Weisskopf style approximations and the results, for a given choice of coupling function, are compared numerically with the profile generated using the exact solution. These comparisons show convincingly the success with which these two approximations can be expected to reproduce the qualitative and quantitative features of the exact lineshape for values of the coupling constant $\alpha$ ranging between 0.01 and 0.3. Finally, we calculate explicitly the scattering operator corresponding to the above spontaneous emission problem and work out exactly the probability of forward and backward scattering of a photon by the atom at and away from resonance.

## I. INTRODUCTION

The model to be discussed in this paper is that of an excited two-level quantum system interacting with a radiation field. The two-level system will usually be thought of as an atom with two accessible electronic states between which a transition occurs with emission or absorption of radiation. However, the "atom" could just as well be a molecule or a spin and the radiation field could be a set of closely spaced molecular states, a phonon field or any of a variety of things, provided the dimensionality $d$ of the field, the electromagnetic multipolarity of the atomic/molecular transitions, the spin, and parity are properly taken into account.

In a series of papers by the authors (Refs. 1–10, hereafter referred to as I–X), we have worked out the exact dynamics of an excited two- or three-level atomic system in interaction with a $d = 1$ field of electromagnetic radiation, as governed by a certain Hamiltonian (see text below). Specifically, in IV and V we studied the spontaneous emission of a two-level atom and in VII and VIII the induced emission; the further contributions IX and X dealt with the exact dynamics of three-level quantum systems. In these papers, only the time evolution of the atomic state (or states) was considered. For the purposes of seeing what experimental consequences there might be of our theoretical studies, it is more sensible to consider the lineshape of the emitted radiation and to include in the model a more realistic (three-dimensional) geometry. Both of these tasks require no further development of analytic techniques (as was the case in I–X), but do require a fair amount of manipulation of existing expressions and calculation (Sec. II). In this contribution we focus on the first of the above problems, and calculate the lineshape for the problem of spontaneous emission (Sec. III A), as determined from the exact solution to the underlying quantum-statistical problem, presented in IV–V. Then in Sec. III B we introduce two Wigner–Weisskopf style approximations and calculate the lineshape corresponding to these two approximations. The results obtained and their characteristic features are compared numerically in Sec. VI for a specific choice of form factor (which specifies the interaction between the atom and the radiation field) and for various choices of the coupling constant $\alpha$. Finally, we consider in detail the scattering of radiation by the two-level atom and work out exactly the probability of forward and backward scattering of a photon at and away from resonance.

We now specify the Hamiltonian for the model under study in this paper. It should be said at the outset that the restriction to a one-dimensional radiation field (which has been made throughout the series I–X) is not essential: It is made chiefly for ease of presentation and for simplicity of the numerical computations. Provided one counts modes properly, all the formal results go through exactly as for $d = 1$. The Hamiltonian is then

$$H = \epsilon_1 \alpha\alpha^* + \epsilon_2 \alpha^*\alpha + \sum_\lambda \left[\frac{1}{2} h\omega_\lambda (a_\lambda^* a_\lambda + 1)\right] + \sum_\lambda (h_\lambda^* \alpha^* a_\lambda + h_\lambda \alpha a_\lambda^*) ,$$

where $\epsilon_1$ and $\epsilon_2$ are the energies of the ground state $|1\rangle$ and excited state $|2\rangle$ of the two-level atom, and where the operators are defined by

$$\alpha = |1\rangle\langle 2| , \quad \alpha^* = |2\rangle\langle 1| ,$$

$$\langle n_\lambda | a_\lambda | m_\lambda \rangle = [2(n_\lambda + 1)]^{1/2} \delta^{\mathrm{Kr}}(m_\lambda - n_\lambda - 1) = \langle m_\lambda | a_\lambda^* | n_\lambda \rangle .$$

Here the state $|n_\lambda\rangle$ is that which has $n_\lambda (= 0,1,2,...)$ photons in the $\lambda$ th mode of the radiation field, and $\delta^{\mathrm{Kr}}(\cdots)$ is the Kronecker delta. Further, $\hbar\omega_\lambda$ is the energy of a photon in the $\lambda$ th mode, $\hbar E = \epsilon_2 - \epsilon_1$ is the energy separating the two levels of the atom, and the $h_\lambda$ give the coupling between the

 0022-2488/85/030556-10$02.50 © 1985 American Institute of Physics 

atom and the radiation field. A basis for the Hilbert space of the system is given by the product states

$$|i;\{n_\lambda\}\rangle \equiv |i\rangle \prod_\lambda |n_\lambda\rangle ,$$

with $i = 1,2$ and $n_\lambda = 0,1,2,\ldots$ .

An important property of this Hamiltonian is that in the above basis it becomes block diagonal. The blocks or "sectors" of the Hamiltonian are the eigenspaces of the operator

$$N = \alpha^*\alpha + \frac{1}{2}\sum_\lambda a_\lambda^* a_\lambda - 1 ,$$

which commutes with $H$. The eigenvalue of $N$ associated with the state $|i;\{n_\lambda\}\rangle$ is $(i-1) + \Sigma_\lambda n_\lambda$, and so $N$ measures the number of photons present when the atom is deexcited ($i = 1$). Papers I–VI dealt exclusively with the sector where $N = 1$, while in VII and VIII, a solution for $N = 2$ was obtained for a finite and an infinite system, respectively. We remark in passing that the operator $N$ was defined in VII simply as

$$N = \alpha^*\alpha + \frac{1}{2}\sum_\lambda a_\lambda^* a_\lambda .$$

In this paper we consider only the $N = 1$ sector, the sector appropriate for the discussion of spontaneous emission or the scattering of a photon off the deexcited two-level system. It is necessary to work with a continuous spectrum of radiation modes (i.e., an infinitely sized cavity) in order for either of these problems to be well defined, on account of the quasiperiodic behavior (Poincaré recurrences) of a system with a discrete set of modes. It is most convenient to define the Hilbert space of the $N = 1$ sector specifically to take account of this; in other words, to define a space of functions of the mode parameter $\lambda$, rather than a space of sequences indexed by $\lambda$. This will be the first step in the calculation of the spontaneous emission lineshape, to which we now proceed.

## II. FORMULATION

The Hilbert space $\mathscr{H}$ for the problem is defined as follows. Consider the triple $(c_0, c_r, c_l)$, where $c_0$ is a complex number, and $c_r$ and $c_l$ each map the non-negative real line into the complex numbers. Then $c \equiv (c_0, c_r, c_l)$ belongs to $\mathscr{H}$ if its norm $\|c\|$ exists and is finite. The norm is defined by

$$\|c\|^2 = |c_0|^2 + \int_0^\infty d\omega\{|c_r(\omega)|^2 + |c_l(\omega)|^2\} .$$

The inner product of $c^1$ and $c^2$, say, is given by

$$(c^1,c^2) = \bar{c}_0^1 c_0^2 + \int_0^\infty d\omega\{\bar{c}_r^1(\omega)c_r^2(\omega) + \bar{c}_l^1(\omega)c_l^2(\omega)\} ,$$

where bars denote complex conjugates. The interpretation of $c$ is that $c_0$ is the probability amplitude for a state in which a two-level atom is excited in the absence of any excitation of a one-dimensional radiation field; $c_r(\omega)$ is a wave function for a state in which the atom is deexcited and a photon is moving from left to right with probability $|c_r(\omega)|^2\, d\omega$ that the photon energy belongs to the interval $[\omega,\omega + d\omega]$; $c_l(\omega)$ is defined similarly for a photon moving in the other direction, from right to left.

Let $H$ denote the Hamiltonian which governs the evolution of the system described by our Hilbert space. Then if for some $c, g \in \mathscr{H}$, $Hg = c$, we have

$$c_0 = \hbar E\left\{g_0 + \left(\frac{\alpha}{\pi}\right)^{1/2}\int_0^\infty d\omega\,[\bar{h}_r(\omega)g_r(\omega) + \bar{h}_l(\omega)g_l(\omega)]\right\} ; \tag{1a}$$

$$c_r(\omega) = \omega g_r(\omega) + \sqrt{(\alpha/\pi)}\hbar E h_r(\omega)g_0 ; \tag{1b}$$

$$c_l(\omega) = \omega g_l(\omega) + \sqrt{(\alpha/\pi)}\hbar E h_l(\omega)g_0 . \tag{1c}$$

Here $\alpha$ is a dimensionless coupling constant, being a one-dimensional analog of the fine-structure constant of ordinary quantum electrodynamics, $\hbar E$ is the energy difference between the two levels of the atom, and the functions $h_r$ and $h_l$ determine the frequency (or energy) dependence of the interaction between the atom and the radiation field.

If the system is in state $c(0)$ at time $t = 0$, then at later times $t$ the state is given by

$$c(t) = \exp[-iHt/\hbar]c(0) ,$$

or more conveniently in resolvent form

$$c(t) = \frac{1}{2\pi i}\int_{\mathscr{C}} dz\, e^{-izt}\left(\frac{H}{\hbar} - z\right)^{-1} c(0) . \tag{2}$$

Here $\mathscr{C}$ is a Bromwich contour above and parallel to the real axis of $z$, and $(H/\hbar - z)^{-1}$ is the resolvent of $H/\hbar$, defined for all nonreal $z$ (at least), since the spectrum of the self-adjoint operator $H$ is purely real.

It is useful to scale the time by the coupling constant $\alpha$ (this is necessary if any "weak-coupling" approximations are to be well defined) and simultaneously make all variables dimensionless. Accordingly, we make the definition $\tau = \alpha E t$, and we transform the Hilbert space $\mathscr{H}$, so that to any $c \in \mathscr{H}$ there corresponds a new element $\hat{c}$ of a space $\hat{\mathscr{H}}$, where

$$\hat{c}_0 = c_0 , \quad \hat{c}_{r,l}(\lambda) = \sqrt{\hbar E}\, c_{r,l}(\hbar E\lambda) .$$

Note that

$$\|c\|^2 = \|\hat{c}\|^2 = |\hat{c}_0|^2 + \int_0^\infty d\lambda\,[|\hat{c}_r(\lambda)|^2 + |\hat{c}_l(\lambda)|^2] ,$$

and in fact the full Hilbert space structure is preserved by this transformation. We also define

$$\hat{h}_{r,l}(\lambda) = \sqrt{\hbar E}\, h_{r,l}(\hbar E\lambda) .$$

Now, from Eq. (2), if $\hat{c}(\tau) \in \hat{\mathscr{H}}$ corresponds under the Hilbert space transformation to $c(t) \in \mathscr{H}$, we have

$$\hat{c}(\tau) = \frac{1}{2\pi i}\int_{\mathscr{C}} d\xi\, e^{-i\xi\tau}\left(\frac{H}{\alpha\hbar E} - \xi\right)^{-1}\hat{c}(0) . \tag{3}$$

Let

$$(H/\alpha\hbar E - \xi)^{-1}\hat{c}(0) = g_\xi .$$

Then,

$$(H/\alpha\hbar E - \xi)g_\xi = \hat{c}(0) ,$$

and so, from Eqs. (1),

$$[\hat{c}(0)]_0 = (1/\alpha - \xi)(g_\xi)_0 + \left(\frac{1}{\pi\alpha}\right)^{1/2}\int_0^\infty d\lambda\,[\bar{\hat{h}}_r(\lambda)(g_\xi)_r(\lambda) + \bar{\hat{h}}_l(\lambda)(g_\xi)_l(\lambda)] ,$$

and

$$[\hat{c}(0)]_{r,I}(\lambda) = (\lambda/\alpha - \xi)(g_\xi)_{r,I}(\lambda) + \sqrt{(1/\pi\alpha)}\,\hat{h}_{r,I}(\lambda)(g_\xi)_0 .$$

These equations can be solved for $g_\xi$ to yield

$$(g_\xi)_0 = \frac{1}{\hat{H}(\xi)}\left\{[\hat{c}(0)]_0 - \left(\frac{\alpha}{\pi}\right)^{1/2}\int_0^\infty d\lambda \times \frac{[\hat{h}_r(\lambda)[\hat{c}(0)]_r(\lambda) + \bar{\hat{h}}_I(\lambda)[\hat{c}(0)]_I(\lambda)]}{\lambda - \alpha\xi}\right\},$$

$$(g_\xi)_{r,I}(\lambda) = [1/(\lambda - \alpha\xi)]\{\alpha[c(0)]_{r,I}(\lambda) - \sqrt{(\alpha/\pi)}\hat{h}_{r,I}(\lambda)(g_\xi)_0\} . \tag{4}$$

Here the function $\hat{H}$ is defined by

$$\hat{H}(\xi) = \frac{1}{\alpha} - \xi - \frac{2}{\pi}\int_0^\infty d\lambda\,\frac{h(\lambda)}{\lambda - \alpha\xi}, \tag{5}$$

in which we define

$$h(\lambda) = \tfrac{1}{2}[\,|\hat{h}_r(\lambda)|^2 + |\hat{h}_I(\lambda)|^2\,] . \tag{6}$$

In order that $g_\xi$ and $\hat{H}(\xi)$ be well defined we must impose the following regularity condition on the functions $\hat{h}_{r,I}$:

$$\int_1^\infty d\lambda\,\frac{|\hat{h}_{r,I}(\lambda)|^2}{\lambda^2} < \infty \tag{7}$$

and $|\hat{h}_{r,I}(\lambda)|^2$ is bounded for all non-negative $\lambda$. The condition implies essentially that $\hat{h}_{r,I}(\lambda)/(\lambda - \xi)$ is, for all $\xi$ not on the positive real line, a function with bounded Hilbert-space norm.

In addition, in order that the dynamics of the system be "ergodic" in the sense defined in our earlier study, we impose the constraint that $h(\lambda) \to 0$ as $\lambda \to 0$ sufficiently fast that

$$\int_0^\infty \frac{d\lambda\, h(\lambda)}{\lambda} \text{ exists},$$

and further that

$$\alpha < \frac{(\pi/2)}{\int_0^\infty d\lambda\, h(\lambda)/\lambda} .$$

These conditions ensure that $\hat{H}(\xi)$ has no zeros for $\xi$ off the positive real line.

We wish to use Eqs. (3) and (4) to derive the infinite-time limit of the solution $\hat{c}(\tau)$. This limit tells us to what asymptotic state the system will relax if it starts at $\tau = 0$ in state $\hat{c}(0)$. In particular if $[\hat{c}(0)]_0 = 1$ and $[\hat{c}(0)]_{r,I}(\lambda) = 0$, the asymptotic state describes the lineshape for spontaneous emission. It turns out that

$$\lim_{\tau\to\infty} \hat{c}(\tau)$$

does not exist but that

$$\lim_{\tau\to\infty} e^{i\lambda\tau/\alpha}[\hat{c}(\tau)]_{r,I}(\lambda)$$

is well defined for all non-negative $\lambda$. This is all that is needed physically since it is the limit of $|[\hat{c}(\tau)]_{r,I}(\lambda)|^2$ which is of interest. Alternatively, we can view $e^{i\lambda\tau/\alpha}[\hat{c}(\tau)]_{r,I}(\lambda)$ as the wave function in the interaction picture. For present purposes it will be useful to retain the slightly clumsier Schrödinger picture.

First we note that

$$\lim_{\tau\to\infty} [\hat{c}(\tau)]_0 = 0 .$$

To see this and to facilitate future calculations we introduce some definitions:

$$f_1(\tau) = \frac{1}{2\pi i}\int_{\mathscr{C}} d\xi\,\frac{e^{-i\xi\tau}}{\hat{H}(\xi)}, \tag{8}$$

$$\hat{G}(\xi) = \left(\frac{\alpha}{2}\right)^{1/2}\int_0^\infty d\lambda\,\frac{\bar{\hat{h}}_r(\lambda)[\hat{c}(0)]_r(\lambda) + \bar{\hat{h}}_I(\lambda)[\hat{c}(0)]_I(\lambda)}{\lambda - \alpha\xi},$$

$$f_2(\tau) = -\frac{1}{2\pi i}\int_{\mathscr{C}} d\xi\,\frac{e^{-i\xi\tau}\hat{G}(\xi)}{\hat{H}(\xi)} . \tag{9}$$

Here, $\hat{G}(\xi)$ is well defined by condition (7). Observe that $f_1$ does not depend on the initial condition $c(0)$, but that $\hat{G}$ and $f_2$ do. The structure of the definitions $f_1$ and $f_2$ as Laplace transforms permits the following inversion equations:

$$\frac{1}{\hat{H}(\xi)} = i\int_0^\infty d\tau\, e^{i\xi\tau} f_1(\tau) ;$$

$$\frac{\hat{G}(\xi)}{\hat{H}(\xi)} = -i\int_0^\infty d\tau\, e^{i\xi\tau} f_2(\tau) . \tag{10}$$

Then from Eqs. (3) and (4) it is easy to see that

$$[\hat{c}(\tau)]_0 = \frac{1}{2\pi i}\int_{\mathscr{C}} d\xi\, e^{-i\xi\tau}(g_\xi)_0 = [\hat{c}(0)]_0 f_1(\tau) + \sqrt{(2/\pi)}\, f_2(\tau) .$$

Our result follows if we can show that $f_1(\tau)$ and $f_2(\tau)$ tend to zero as $\tau \to \infty$. It is sufficient to check this for $f_1$; the calculation for $f_2$ is similar. From Eq. (5) it is clear that $\hat{H}$ is analytic and nonzero except for a cut along the positive real axis. The contour $\mathscr{C}$ in the definition (8) of $f_1$ can therefore be deformed by Jordan's lemma [note that $|\hat{H}(\xi)|^{-1} \to 0$ as $\xi \to \infty$, $\arg\xi \neq 0$] to a contour $\mathscr{B}$, say, surrounding the positive real axis (see Fig. 1). If $\hat{H}^{\pm}(\xi)$ denotes the limiting values of $\hat{H}$ as its argument tends to real $\xi$ through values with positive/negative imaginary part, then we can conclude that

$$f_1(\tau) = \frac{1}{2\pi i}\int_0^\infty d\xi\, e^{-i\xi\tau}\left[\frac{1}{\hat{H}^+(\xi)} - \frac{1}{\hat{H}^-(\xi)}\right].$$

Now since

$$\hat{H}^{\pm}(\xi) = \frac{1}{\alpha} - \xi - \frac{2}{\pi}\,\mathscr{P}\int_0^\infty \frac{d\lambda\, h(\lambda)}{\lambda - \alpha\xi} \mp 2ih(\alpha\xi) \tag{11}$$

(here $\mathscr{P}$ denotes the Cauchy principal part), it can be seen that $[1/\hat{H}^+ - 1/\hat{H}^-]$ is bounded on the positive real line. Further, since plainly we require that $f_1(0) = 1$, the integral

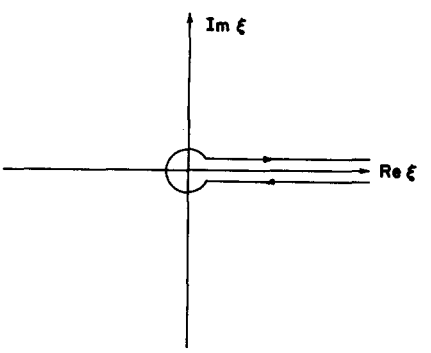



FIG. 1. The contour $\mathscr{B}$.

$$\int_0^\infty d\xi \left[\frac{1}{\hat{H}^+(\xi)} - \frac{1}{\hat{H}^-(\xi)}\right]$$

exists, and so the Riemann–Lebesgue lemma ensures that

$$\lim_{\tau\to\infty} f_1(\tau) = 0 .$$

Again from Eqs. (3) and (4) we find that

$$[\hat{c}(\tau)]_{r,I}(\lambda) = e^{-i\lambda\tau/\alpha}[\hat{c}(0)]_{r,I}(\lambda) - \left(\frac{\alpha}{\pi}\right)^{1/2}\hat{h}_{r,I}(\lambda)\frac{1}{2\pi i}\int_{\mathscr{C}} d\xi \frac{e^{-i\xi\tau}(g_\xi)_0}{\lambda - \alpha\xi} . \tag{12}$$

Now since

$$\frac{1}{2\pi i}\int_{\mathscr{C}} d\xi\, e^{-i\xi\tau}(g_\xi)_0 = [\hat{c}(0)]_0 f_1(\tau) + \left(\frac{2}{\pi}\right)^{1/2} f_2(\tau)$$

and

$$\frac{1}{2\pi i}\int_{\mathscr{C}} d\xi \frac{e^{-i\xi\tau}}{\lambda - \alpha\xi} = \frac{1}{\alpha} e^{-i\lambda\tau/\alpha} ,$$

by the convolution property of Laplace transforms we obtain that

$$\frac{1}{2\pi i}\int_{\mathscr{C}} d\xi \frac{e^{-i\xi\tau}(g_\xi)_0}{\lambda - \alpha\xi} = \frac{i}{\alpha}\int_0^\tau d\tau'\, e^{-i\lambda(\tau-\tau')/\alpha} \times \left[[\hat{c}(0)]_0 f_1(\tau) + \sqrt{(2/\pi)}\, f_2(\tau')\right] .$$

From this result and Eq. (12) we can see that

$$\lim_{\tau\to\infty} e^{i\lambda\tau/\alpha}[\hat{c}(\tau)]_{r,I}(\lambda) = [\hat{c}(0)]_{r,I}(\lambda) - (i/\sqrt{\pi\alpha})\hat{h}_{r,I}(\lambda) \times \int_0^\infty d\tau\, e^{i\lambda\tau/\alpha}\left\{[\hat{c}(0)]_0 f_1(\tau) + \left(\frac{2}{\pi}\right)^{1/2} f_2(\tau)\right\} \tag{13}$$

$$= [\hat{c}(0)]_{r,I}(\lambda) - \frac{\hat{h}_{r,I}(\lambda)}{\sqrt{\pi\alpha}}\left[\frac{[\hat{c}(0)]_0}{\hat{H}^+(\lambda/\alpha)} - \left(\frac{2}{\pi}\right)^{1/2}\frac{\hat{G}^+(\lambda/\alpha)}{\hat{H}^+(\lambda/\alpha)}\right], \tag{14}$$

by Eqs. (10). The limits $\hat{H}^+$ and $\hat{G}^+$ are needed for the real argument $\lambda/\alpha$ since the definitions (8) and (9) use an integration contour $\mathscr{C}$ in the upper half-plane of $\xi$.

## III. CALCULATION OF THE LINESHAPE

### A. The exact result

We now specialize the results presented in the preceding section to obtain the lineshape of a photon emitted spontaneously by the two-level atom. Since the initial state $\hat{c}(0)$ obeys $[\hat{c}(0)]_0 = 1$, $[\hat{c}(0)]_{r,I}(\lambda) = 0$, we see at once that

$$\hat{G}(\xi) = f_2(\tau) = 0 ,$$

and so

$$c^\infty_{r,I}(\lambda) \equiv \lim_{\tau\to\infty} e^{i\lambda\tau/\alpha}[\hat{c}(\tau)]_{r,I}(\lambda) = -\hat{h}_{r,I}(\lambda)/\sqrt{\pi\alpha}\,\hat{H}^+(\lambda/\alpha) .$$

The probability density for the dimensionless energy $\lambda$ of the emitted photon (the *lineshape*) is thus

$$|c^\infty_{r,I}(\lambda)|^2 = \frac{1}{\alpha\pi}\frac{|\hat{h}_{r,I}(\lambda)|^2}{\hat{H}^+(\lambda/\alpha)\hat{H}^-(\lambda/\alpha)}, \tag{15}$$

since

$$\overline{\hat{H}^+(\lambda/\alpha)} = \hat{H}^-(\lambda/\alpha)$$

[see Eq. (11)]. Now, from Eq. (11) we can express Eq. (15) more explicitly:

$$|c^\infty_{r,I}(\lambda)|^2 = \frac{\alpha}{\pi}|\hat{h}_{r,I}(\lambda)|^2\left(\left[1 - \lambda - \frac{2\alpha}{\pi}\mathscr{P}\int_0^\infty \frac{d\mu\, h(\mu)}{\mu - \lambda}\right]^2 + 4\alpha^2[h(\lambda)]^2\right)^{-1} . \tag{16}$$

It is easy to check that

$$\int_0^\infty d\lambda\, [|c^\infty_r(\lambda)|^2 + |c^\infty_I(\lambda)|^2] = 1 .$$

Note that

$$|c^\infty_r(\lambda)|^2 + |c^\infty_I(\lambda)|^2 = \frac{2}{\alpha\pi}\frac{h(\lambda)}{\hat{H}^+(\lambda/\alpha)\hat{H}^-(\lambda/\alpha)} \quad \text{[by Eq. (6)]} , = \frac{1}{2\pi i\alpha}\left[\frac{1}{\hat{H}^+(\lambda/\alpha)} - \frac{1}{\hat{H}^-(\lambda/\alpha)}\right] \quad \text{[by Eq. (11)]} .$$

Then, the above result follows at once from

$$\frac{1}{2\pi i}\int_{\mathscr{C}}\frac{d\xi}{\hat{H}(\xi)} = 1 .$$

It is plain from this normalization and from the form of Eq. (16) that as $\alpha\to 0$, the function

$$|c^\infty_r(\lambda)|^2 + |c^\infty_I(\lambda)|^2$$

approaches the Dirac delta function $\delta(\lambda - 1)$. Thus for very weak coupling, the lineshape, as one would expect, is increasingly concentrated at the resonant frequency.

### B. The Wigner–Weisskopf approximation

It is interesting to compare the exact result, Eq. (16), with the well-known Wigner–Weisskopf (WW) approximation to the lineshape. To obtain this approximation, one starts from Eq. (13) for the case of spontaneous emission:

$$c^\infty_{r,I}(\lambda) = -\frac{i}{\sqrt{\alpha\pi}}\hat{h}_{r,I}(\lambda)\int_0^\infty e^{i\lambda\tau/\alpha} f_1(\tau)d\tau ,$$

but replaces $f_1(\tau)$ by its WW approximation. This approximation is effected by observing that if the function $\hat{H}(\xi)$ is analytically continued from the upper half-plane of $\xi$ through the positive real axis into the lower half-plane, then a zero appears at a point $\xi_0$, say, with $\mathrm{Re}\,\xi_0 > 0$ and $\mathrm{Im}\,\xi_0 < 0$. Since in our dimensionless notation the frequency in resonance with the two-level atom is $\lambda = 1$, the quantity

$$\mathrm{Re}\,\xi_0 - 1/\alpha$$

is the Lamb shift and $(-2\,\mathrm{Im}\,\xi_0)$ is the linewidth (inverse lifetime) for the decay. The WW approximation for $f_1(\tau)$ is simply $e^{-i\xi_0\tau}$, so that for $c^\infty_{r,I}(\lambda)$ the WW approximation is

$$\sqrt{(\alpha/\pi)}[\hat{h}_{r,I}(\lambda)/(\lambda - \alpha\xi_0)] ,$$

and for the lineshape it is

$$\frac{\alpha}{\pi}\frac{|\hat{h}_{r,l}(\lambda)|^2}{(\lambda-\alpha\,\mathrm{Re}\,\xi_0)^2+\alpha^2(\mathrm{Im}\,\xi_0)^2}.$$

Usually $\mathrm{Re}\,\xi_0$ and $\mathrm{Im}\,\xi_0$ are approximated for small $\alpha$ by the formulas:

$$\mathrm{Re}\,\xi_0-\frac{1}{\alpha}=-\frac{2}{\pi}\mathscr{P}\int_0^\infty\frac{d\mu\,h(\mu)}{\mu-1},$$

$$\mathrm{Im}\,\xi_0=-4h(1),$$

and this approximation yields for the lineshape

$$\frac{\alpha}{\pi}|\hat{h}_{r,l}(\lambda)|^2\left(\left[1-\lambda-\frac{2\alpha}{\pi}\mathscr{P}\int_0^\infty\frac{d\mu\,h(\mu)}{\mu-1}\right]^2+4\alpha^2[h(1)]^2\right)^{-1} \tag{17}$$

which looks, at least, very much like the exact expression (16).

Sometimes a still cruder approximation is made in which the numerator $|\hat{h}_{r,l}(\lambda)|^2$ is replaced by $|\hat{h}_{r,l}(1)|^2$. This device may be regarded as just another step in the degradation of the formally exact result, Eq. (16).

In some calculations it may be of interest to consider the proportion of the total energy in the range $[\lambda,\lambda+d\lambda]$ rather than the proportion of total probability. This "energy lineshape" can be found from Eqs. (16) or (17) by multiplying by $\lambda$, since it can be verified that

$$\int_0^\infty d\lambda\,\lambda\,[|c_r^\infty(\lambda)|^2+|c_l^\infty(\lambda)|^2]=1.$$

## IV. PROPERTIES OF THE MAPPING FROM TIME ZERO TO ASYMPTOTIC STATES

We now return to Eq. (14) to establish some of its properties as a preliminary to the discussion of the scattering of radiation by the two-level atom. It is useful to define a mapping $\Omega^+$ from time zero states to the asymptotic states to which they evolve as $\tau\to+\infty$: $\Omega^+(c_0,c_r,c_l)$ denotes a pair $(c_r^\infty,c_l^\infty)$ defined by the equations

$$c_{r,l}^\infty(\lambda)=c_{r,l}(\lambda)-\frac{\hat{h}_{r,l}(\lambda)}{\sqrt{\pi\alpha}}\times\left[\frac{c_0}{\hat{H}^+(\lambda/\alpha)}-\left(\frac{2}{\pi}\right)^{1/2}\frac{\hat{G}^+(\lambda/\alpha)}{\hat{H}^+(\lambda/\alpha)}\right], \tag{18}$$

where

$$\hat{G}(\xi)\equiv\left(\frac{\alpha}{2}\right)^{1/2}\int_0^\infty d\lambda\,\frac{\bar{\hat{h}}_r(\lambda)c_r(\lambda)+\bar{\hat{h}}_l(\lambda)c_l(\lambda)}{\lambda-\alpha\xi}. \tag{19}$$

Our first result is that $\Omega^+$ is a (linear) isometry (and therefore one-to-one) mapping of $\hat{\mathscr{H}}$ on to the Hilbert space of asymptotic states, i.e., the space of pairs $(c_r^\infty,c_l^\infty)$ with obvious norm and scalar product. Formally, the isometry property is

$$\int_0^\infty d\lambda\,[|c_r^\infty(\lambda)|^2+|c_l^\infty(\lambda)|^2]=|c_0|^2+\int_0^\infty d\lambda\,[|c_r(\lambda)|^2+|c_l(\lambda)|^2],$$

and it follows immediately from the fact that the evolution operator $e^{-iHt/\hbar}$ is unitary. To show that $\Omega^+$ is an onto mapping, it is probably best to exhibit its inverse explicitly. This can be done most conveniently if the pairs $(c_r,c_l)$ and $(c_r^\infty,c_l^\infty)$ are transformed:

$$\kappa_s(\lambda)=[1/\sqrt{2h(\lambda)}]\,[\bar{\hat{h}}_r(\lambda)c_r(\lambda)+\bar{\hat{h}}_l(\lambda)c_l(\lambda)],$$

$$\kappa_a(\lambda)=[1/\sqrt{2h(\lambda)}]\,[\hat{h}_l(\lambda)c_r(\lambda)-\hat{h}_r(\lambda)c_l(\lambda)]. \tag{20}$$

Then $\Omega^+$ can be thought of as mapping $(c_0,\kappa_s,\kappa_a)$ into $(\kappa_s^\infty,\kappa_a^\infty)$. One observes at once that $(0,0,\kappa_a)$ is mapped simply into $(0,\kappa_a)$, i.e., that time zero radiation states with only a $\kappa_a$-type component do not interact with the atom. Similarly, states like $(c_0,\kappa_s,0)$ are mapped to states like $(\kappa_s^\infty,0)$. Thus all we need consider is the inverse image of asymptotic states $(\kappa_s^\infty,0)$.

The operator $\Omega^+$ itself can be written explicitly as acting on $(c^0,\kappa_s)$:

$$[\Omega^+(c^0,\kappa_s)](\lambda)=\kappa_s(\lambda)-\left(\frac{2h(\lambda)}{\pi\alpha}\right)^{1/2}\frac{1}{\hat{H}^+(\lambda/\alpha)}\times[c_0-\sqrt{(2/\pi)}\hat{G}^+(\lambda/\alpha)]. \tag{21}$$

It can be verified that the inverse is given by

$$c_0=-\left(\frac{2}{\alpha\pi}\right)^{1/2}\int_0^\infty d\lambda\,\frac{\sqrt{h(\lambda)}\kappa_s^\infty(\lambda)}{\hat{H}^-(\lambda/\alpha)}; \tag{22}$$

$$\kappa_s(\lambda)=-\frac{2}{\pi}\sqrt{h(\lambda)}\int_0^\infty d\mu\,\frac{\sqrt{h(\mu)}\,\kappa_s^\infty(\mu)}{\hat{H}^-(\mu/\alpha)(\mu-\lambda)}+\frac{1}{2}\kappa_s^\infty(\lambda)\frac{\hat{H}^+(\lambda/\alpha)+\hat{H}^-(\lambda/\alpha)}{\hat{H}^-(\lambda/\alpha)}. \tag{23}$$

These equations can be obtained directly by setting up the problem of the inversion of $\Omega^+$ as a singular integral equation and solving it *à la* Muskhelishvili,[11] and can be checked by direct substitution if one uses the Poincaré–Bertrand result:

$$\mathscr{P}\int_0^\infty\frac{d\mu}{\mu-\lambda}\mathscr{P}\int_0^\infty d\mu'\,\frac{\phi(\mu,\mu')}{\mu'-\mu}-\mathscr{P}\int_0^\infty d\mu'\,\mathscr{P}\int_0^\infty d\mu\,\frac{\phi(\mu,\mu')}{(\mu-\lambda)(\mu'-\mu)}=-\pi^2\phi(\lambda,\lambda), \tag{24}$$

for well-behaved functions $\phi$ of two real variables. The calculations that justify the above statements are given in the two Appendices. Appendix A contains the substitution of Eqs. (22) and (23) into Eq. (21) and the verification of the result by that means; Appendix B contains the solution of the appropriate singular integral equation. Note that Eqs. (22) and (23) yield well-defined expressions for all $\kappa_s^\infty$ with finite norm, because of condition (7). This allows us to conclude that the mapping $\Omega^+$ is onto.

The fact that $\Omega^+$ is a one-to-one mapping may seem slightly counterintuitive. After all, if one considers the state that

$$\hat{c}(0)=(c^0,\kappa_s,\kappa_a)$$

evolves into from $\tau=0$ to $\tau=\tau'$, surely this evolved state must map into the same asymptotic state as $\hat{c}(0)$ itself? Physically, of course, this is so:

$$c^{\infty}_{r,l}(\lambda) \equiv \lim_{\tau\to\infty} e^{i\lambda\tau/\alpha}[\hat{c}(\tau)]_{r,l}(\lambda)$$

$$= \lim_{\tau\to\infty} e^{i\lambda(\tau+\tau')/\alpha}[\hat{c}(\tau+\tau')]_{r,l}(\lambda)$$

$$= e^{i\lambda\tau'/\alpha} \lim_{\tau\to\infty} e^{i\lambda\tau/\alpha}[\hat{c}_{\tau'}(\tau)]_{r,l}(\lambda), \tag{25}$$

where $\hat{c}_{\tau'}(\tau)$ is defined as $\hat{c}(\tau+\tau')$ and is the state reached at time $\tau$ from the state that at $\tau = 0$ is just $\hat{c}(\tau')$. In fact, Eq. (25) manifests the rather surprising fact that the entire dynamics of our system, and not just the asymptotic dynamics, can be recovered from the mapping $\Omega^{+}$. If, for some asymptotic state $c^{\infty}$, we have

$$\hat{c}(0) = (\Omega^{+})^{-1}(c^{\infty}),$$

then, if we form another asymptotic state $c^{\infty}_{\tau}$, say, by the rule

$$(c^{\infty}_{\tau})_{r,l}(\lambda) = e^{-i\lambda\tau/\alpha} c^{\infty}_{r,l}(\lambda),$$

we obtain that

$$\hat{c}(\tau) = (\Omega^{+})^{-1}(c^{\infty}_{\tau}).$$

## V. THE SCATTERING PROBLEM

We are now ready to make use of the properties of $\Omega^{+}$ in order to discuss the problem of the scattering of a photon by the two-level atom. In the theory of the $S$ matrix (see, for example, Ref. 12), eigenstates of the system Hamiltonian without interaction are treated as ingoing states and are mapped by the $S$ matrix to other eigenstates of the interactionless Hamiltonian, the outgoing states. These ingoing and outgoing states in our problem must, of course, be states with the atom deexcited. In this context the operator $\Omega^{+}$ and its time-reversed form $\Omega^{-}$ are just the Möller operators of $S$-matrix theory, and $S$ itself can be interpreted as $\Omega^{+}(\Omega^{-})^{-1}$, the operator that acts on an asymptotic state at time $t = -\infty$ and brings it through time zero and out to another asymptotic state at time $t = +\infty$. We shall have fully characterized the scattering problem, then, by calculating $\Omega^{+}(\Omega^{-})^{-1}$.

The only difference between $\Omega^{+}$ and $\Omega^{-}$ is that in $\Omega^{-}$ the expressions $\hat{H}^{+}(\lambda/\alpha)$ and $\hat{G}^{+}(\lambda/\alpha)$ in Eq. (21) must be replaced by $\hat{H}^{-}(\lambda/\alpha)$ and $\hat{G}^{-}(\lambda/\alpha)$. This is easily established by arguments similar to those leading to Eq. (14) but with the $\xi$ integrals in the lower half-plane. Thus let $(c_0,\kappa_s,\kappa_a)$ be any state at time $\tau = 0$. Then $S$ maps $\Omega^{-}(c_0,\kappa_s,\kappa_a)$ into $\Omega^{+}(c_0,\kappa_s,\kappa_a)$. Clearly $\kappa_a$ is unaffected by the whole business. If we denote the image functions under $\Omega^{\pm}$ as $\kappa_s^{\pm\infty}$, then from Eq. (21):

$$\kappa_s^{\pm\infty}(\lambda)\hat{H}^{\pm}(\lambda/\alpha)$$
$$= \kappa_s(\lambda)\hat{H}^{\pm}\left(\frac{\lambda}{\alpha}\right) - c_0\left(\frac{2h(\lambda)}{\pi\alpha}\right)^{1/2}$$
$$+ \frac{2}{\pi}\left(\frac{h(\lambda)}{\alpha}\right)^{1/2}\hat{G}^{\pm}\left(\frac{\lambda}{\alpha}\right).$$

But, from the definitions, Eqs. (11) and (19), it is clear that the right-hand side of this equation is the same with either the $+$ or $-$ sign. Thus

$$\kappa_s^{+\infty}(\lambda)\hat{H}^{+}(\lambda/\alpha) = \kappa_s^{-\infty}(\lambda)\hat{H}^{-}(\lambda/\alpha),$$

and so

$$(S\kappa_s^{-\infty})(\lambda) = [\hat{H}^{-}(\lambda/\alpha)/\hat{H}^{+}(\lambda/\alpha)]\kappa_s^{-\infty}(\lambda).$$

The scattering problem is therefore solved without further calculation. We note that since

$$|\hat{H}^{+}(\kappa)| = |\hat{H}^{-}(\kappa)|,$$

normalization is preserved, as it must be, by the unitary operator $S$. Further $S$ is "diagonal" in $\lambda$, i.e., it conserves energy, again a necessary property.

For a better physical sense for the above result, it is useful to return to the $c_{r,l}$ representation. The inverse of the transformation, Eq. (20), is

$$c_{r,l}(\lambda) = \frac{1}{(2h(\lambda))^{1/2}}\{\hat{h}_{r,l}(\lambda)\kappa_s(\lambda) \pm \bar{\hat{h}}_{l,r}(\lambda)\kappa_a(\lambda)\}. \tag{26}$$

We choose for our $t = -\infty$ state a state with $c_l^{-\infty}(\lambda) = 0$, so that the incident photon travels from left to right. Then,

$$\kappa_s^{-\infty}(\lambda) = [1/(2h(\lambda))^{1/2}]\bar{\hat{h}}_r(\lambda)c_r(\lambda),$$
$$\kappa_a^{-\infty}(\lambda) = [1/(2h(\lambda))^{1/2}]\hat{h}_l(\lambda)c_r(\lambda).$$

Now, $\kappa_a^{-\infty}$ is mapped into itself by $S$ and $\kappa_s^{+\infty}(\lambda)$ is

$$\frac{\hat{H}^{-}(\lambda/\alpha)}{\hat{H}^{+}(\lambda/\alpha)}\frac{1}{(2h(\lambda))^{1/2}}\bar{\hat{h}}_r(\lambda)c_r(\lambda).$$

Use of Eq. (26) yields

$$c_r^{+\infty}(\lambda) = \frac{1}{2h(\lambda)}\left\{|\hat{h}_r(\lambda)|^2\frac{\hat{H}^{-}(\lambda/\alpha)}{\hat{H}^{+}(\lambda/\alpha)} + |\hat{h}_l(\lambda)|^2\right\}c_r(\lambda),$$

$$c_l^{+\infty}(\lambda) = \frac{1}{2h(\lambda)}\bar{\hat{h}}_r(\lambda)\hat{h}_l(\lambda)\left\{\frac{\hat{H}^{-}(\lambda/\alpha)}{\hat{H}^{+}(\lambda/\alpha)} - 1\right\}c_r(\lambda).$$

The probability of forward scattering of a photon of dimensionless energy $\lambda$, then, is just

$$|c_r^{+\infty}(\lambda)|^2/|c_r(\lambda)|^2,$$

i.e.,

$$\frac{1}{4h^2(\lambda)}\left||\hat{h}_r(\lambda)|^2\frac{\hat{H}^{-}(\lambda/\alpha)}{\hat{H}^{+}(\lambda/\alpha)} + |\hat{h}_l(\lambda)|^2\right|^2.$$

For symmetric interactions with $|\hat{h}_r(\lambda)|^2 = |\hat{h}_l(\lambda)|^2 = h(\lambda)$, this simplifies to

$$[1/4|\hat{H}^{+}(\lambda/\alpha)|^2]|\hat{H}^{+}(\lambda/\alpha) + \hat{H}^{-}(\lambda/\alpha)|^2,$$

and if

$$\theta(\lambda/\alpha) = \arg\hat{H}^{+}(\lambda/\alpha) = -\arg\hat{H}^{-}(\lambda/\alpha),$$

it simplifies still further to $\cos^2\theta(\lambda/\alpha)$. Explicitly,

$$\tan\theta\left(\frac{\lambda}{\alpha}\right) = \frac{-2h(\lambda)}{(1/\alpha) - (\lambda/\alpha) - (2/\pi)\,\mathcal{P}\int_0^{\infty} d\mu\, h(\mu)/(\mu-\lambda)}.$$

Note that at resonance, $\lambda = 1$, and $\theta(\lambda/\alpha)$ becomes $\alpha$ independent, with

$$\tan\theta\left(\frac{\lambda}{\alpha}\right) = \frac{h(1)}{(1/\pi)\,\mathcal{P}\int_0^{\infty} d\mu\, h(\mu)/(\mu-\lambda)}.$$

However, if $\lambda = 1$,

$$\lim_{\alpha\to 0}\tan\theta(\lambda/\alpha) = 0,$$

and so, in the weak-coupling, i.e., $\alpha\to 0$, limit, scattering at

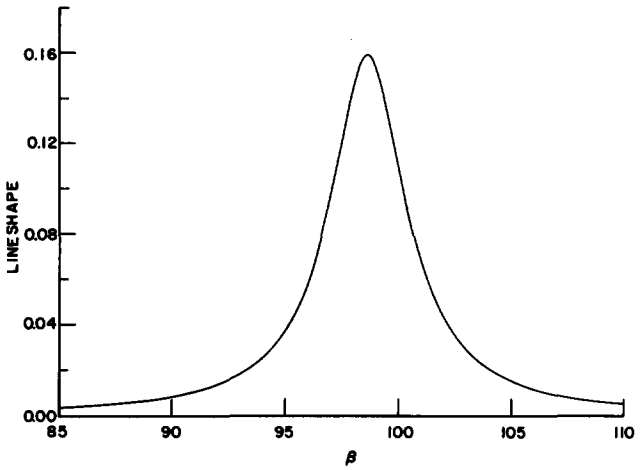



FIG. 2. A plot of the lineshape $|c^{\infty}_{r,l}(\lambda)|^2$ vs $\beta$ (where $\lambda = \alpha\beta$) as determined from the exact expression, Eq. (16), for the choice of coupling constant, $\alpha = 0.01$.

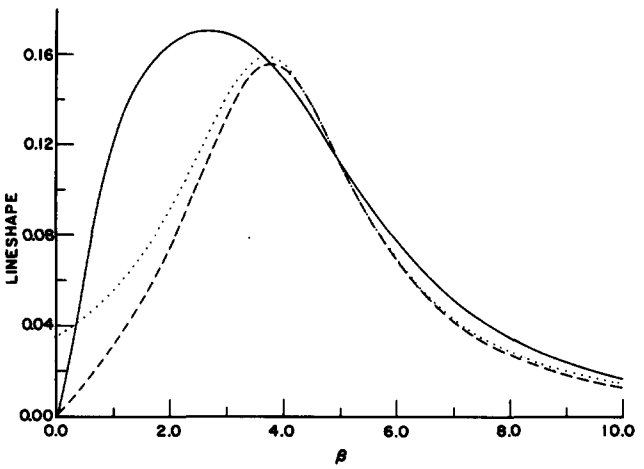



FIG. 4. A plot of the lineshape $|c^{\infty}_{r,l}(\lambda)|^2$ vs $\beta$ for $\alpha = 0.2$. The conventions here are the same as in Fig. 3.

resonance is as with positive $\alpha$, but no backward scattering occurs except exactly at resonance. This result is satisfying to one's physical intuition.

## VI. NUMERICAL RESULTS

In Sec. III an exact expression was derived [viz. Eq. (16)] for the lineshape for spontaneous emission of an excited two-level atom in a (one-dimensional) field of electromagnetic radiation (assumed to be deexcited initially) in the limit where the system size becomes of infinite extent and the mode spectrum becomes continuous. In our earlier studies of the time evolution of the system, calculations were performed for several different choices of the coupling function, in the notation of this paper, $h(x)$. These were

$$h(x) = x^{-1/2}, \quad h(x) = x^{-1/4},$$

and

$$h(x) = 4x/(1+x)^2. \tag{27}$$

It was found in V that the first two choices of coupling function led to "ghost states" and nonergodic behavior in the time evolution of $\rho(\tau)$, the probability that the atom is in the excited state at time $\tau$, for sufficiently large values of the coupling parameter $\alpha$. In this paper, however, we shall confine our attention to results generated using Eq. (16) with the "ergodic" form factor (27). We shall also display results for the $|c_{r,l}(\lambda)|^2$ as calculated using the Wigner–Weisskopf approximation (17) as well as for the case where the numerator in Eq. (17) is replaced by $|\hat{h}_{r,l}(1)|^2$; we shall refer to these two approximations as WWI and WWII.

Given the choice (27), four different values of the coupling constant $\alpha$ were selected for study, viz. $\alpha = 0.01, 0.10, 0.20$, and $0.30$. The results obtained for the $|c_{r,l}(\lambda)|^2$ corresponding to these choices of $\alpha$ are displayed in Figs. 2–5, respectively. From the results recorded in these four figures, it is seen that the correspondence between the exact results for $|c_{r,l}(\lambda)|^2$ and the ones generated using WWI is quite acceptable up to a coupling strength of $\alpha = 0.1$; both lineshapes are essentially Lorentzian. Beyond this coupling,

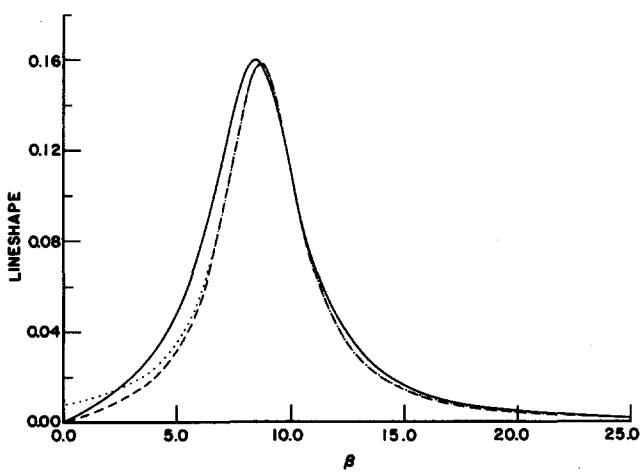



FIG. 3. A plot of the lineshape $|c^{\infty}_{r,l}(\lambda)|^2$ vs $\beta$ for $\alpha = 0.1$. The profile corresponding to the exact lineshape, Eq. (16), is represented by the solid line, the one calculated using the Wigner–Weisskopf approximation, Eq. (17), is represented by the dashed line and the one corresponding to Eq. (17) with the numerator replaced by $|\hat{h}_{r,l}(1)|^2$ (see text) is represented by the dotted line.

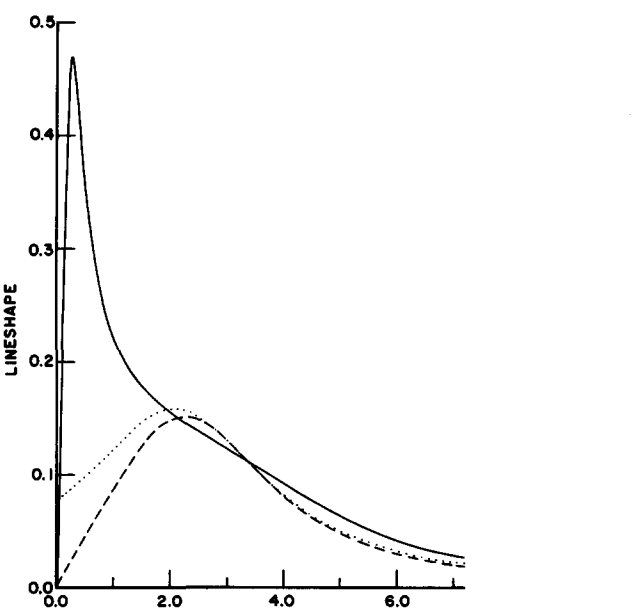



FIG. 5. A plot of the lineshape $|c^{\infty}_{r,l}(\lambda)|^2$ vs $\beta$ for $\alpha = 0.3$. The conventions here are the same as in Fig. 4.

however, the profiles determined from Eqs. (16) and (17) exhibit noticeable and qualitatively significant differences. The exact lineshape is not at all Lorentzian in structure when $\alpha = 0.2$, with the profile generated using (16) tending to skew even more when $\alpha$ is increased to 0.3. The lineshapes calculated using WWI, Eq. (17), remain approximately Lorentzian for all couplings $\alpha$ studied in this paper. Finally, we note that for all $\alpha \lesssim 0.1$, the lineshape generated from the approximation WWII has the wrong qualitative behavior when $\beta \to 0$ (viz., the intercept is finite in this limit) but sensibly describes the results obtained from Eq. (17) in the limit of large $\beta$.

As a further means of quantifying the success with which the profiles generated from the approximations WWI and WWII represent the exact lineshape, one can compare the coordinates of the maxima in the plot of $|c_{r,I}(\lambda)|^2$ vs $\lambda$ for each $\alpha$. As is seen from the data listed in Table I, the matching of the peak height and location for profiles generated using WWI and WWII versus the exact lineshape is quite acceptable for $\alpha \lesssim 0.1$, but becomes seriously in error for $\alpha > 0.1$. One can also check numerically whether the normalization condition

$$\int_0^\infty d\lambda\, \lambda\, [|c_r^\infty(\lambda)|^2 + |c_I^\infty(\lambda)|^2] = 1$$

is satisfied in each of these cases. From Table I it is evident that only the exact solution displays the proper normalization for all values of $\alpha$; the data show that deviations from unity become rather pronounced when either Wigner–Weisskopf approximation is employed in the coupling regime, $\alpha > 0.1$.

## VII. CONCLUSIONS

In this paper we have carried out two exact calculations based on the $N = 1$ sector of the model for a two-level system in interaction with a continuous spectrum of radiation. The first calculation was of the lineshape of a spontaneously emitted photon, and we saw that, for weak coupling, the exact answer was very close to either of the Wigner–Weisskopf approximations. For higher coupling, the exact Lamb shift appears to be substantially higher than that predicted by the approximate procedures, and the actual lineshape itself quite different. The other calculation was that of the scattering of an incident photon from the deexcited atom. Here a very simple result was obtained, permitting easy calculation of the probabilities of forward and backward scattering. It can be remarked here that, in a more realistic three-dimensional calculation, the *only* extra complication would be purely geometrical.

We also derived the explicit Möller operators for the scattering problem and their inverses. Although the inverses were not in fact necessary for the completion of the scattering problem, we found the interesting result that the Möller operators and their inverses—just because the operators were one-to-one and hence invertible—contained implicitly the *full dynamics* of the quantum system. This result is intriguing in view of the usual presumption that restricting attention to asymptotic states entails *loss* of information.


## ACKNOWLEDGMENTS

The research described herein was supported in part by the Division of Basic Energy Sciences of the U. S. Department of Energy. This is Document No. NDRL-2583 from the Notre Dame Radiation Laboratory.


## APPENDIX A: CALCULATION OF THE INVERSE MAPPING OPERATOR BY DIRECT SUBSTITUTION

If $(c_0, \kappa_s)$ are given by Eqs. (22) and (23), we wish to compute $\Omega^+(c_0, \kappa_s)$ and show that it is $(\kappa_s^\infty)$. From Eq. (21) one finds that

$$[\Omega^+(c_0,\kappa_s)]_s(\lambda) = \kappa_s(\lambda) - \frac{(2h(\lambda))^{1/2}}{(\pi/\alpha)^{1/2}} \frac{1}{\hat{H}^+(\lambda/\alpha)} \times \left\{ c_0 - \left(\frac{2\alpha}{\pi}\right)^{1/2} \left[ \mathscr{P} \int_0^\infty d\mu\, \frac{(h(\mu))^{1/2}\kappa_s(\mu)}{\mu - \lambda} + \pi i (h(\lambda))^{1/2}\kappa_s(\lambda) \right] \right\}. \tag{A1}$$

With $\kappa_s(\mu)$ given by Eq. (23), it is necessary first of all to deal with the expression

$$\mathscr{P} \int_0^\infty d\mu\, \frac{h(\mu)}{\mu - \lambda} \mathscr{P} \int_0^\infty d\nu\, \frac{(h(\nu))^{1/2}\kappa_s^\infty(\nu)}{\hat{H}^-(\nu/\alpha)(\nu - \mu)}. \tag{A2}$$

Use of Eq. (24) converts this expression to

$$\mathscr{P} \int_0^\infty d\nu\, \frac{(h(\nu))^{1/2}\kappa_s^\infty(\nu)}{\hat{H}^-(\nu/\alpha)} \mathscr{P} \int_0^\infty d\mu\, \frac{h(\mu)}{(\mu - \lambda)(\nu - \mu)} - \pi^2 \frac{h^{3/2}(\lambda)\kappa_s^\infty(\lambda)}{\hat{H}^-(\lambda/\alpha)}.$$

Now,

$$\mathscr{P} \int_0^\infty d\mu\, \frac{h(\mu)}{(\mu - \nu)(\nu - \mu)} = \frac{1}{\nu - \lambda} \left[ \mathscr{P} \int_0^\infty \frac{d\mu\, h(\mu)}{\mu - \lambda} - \mathscr{P} \int_0^\infty \frac{d\mu\, h(\mu)}{\mu - \nu} \right]$$

TABLE I. Location of lineshape maxima and normalization for various couplings.

| | Exact expression [Eq. (16)] | | | Wigner–Weisskopf approximation I [Eq. (17)] | | | Wigner–Weisskopf approximation II (see text) | | |
|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $\beta_{max}$ | $\lvert c_{r,I}^\infty\rvert^2_{max}$ | Norm. | $\beta_{max}$ | $\lvert c_{r,I}^\infty\rvert^2_{max}$ | Norm. | $\beta_{max}$ | $\lvert c_{r,I}^\infty\rvert^2_{max}$ | Norm. |
| 0.01 | 98.7 | 0.159 158 | 0.995 720 | 98.7 | 0.159 120 | 0.982 940 | 98.7 | 0.159 127 | 0.987 218 |
| 0.1 | 8.50 | 0.160 191 | 0.994 640 | 8.70 | 0.158 358 | 0.860 725 | 8.70 | 0.159 126 | 0.912 854 |
| 0.2 | 2.70 | 0.170 451 | 0.999 314 | 3.80 | 0.155 986 | 0.725 615 | 3.72 | 0.159 153 | 0.836 559 |
| 0.3 | 0.250 | 0.470 024 | 0.979 732 | 2.24 | 0.151 802 | 0.601 075 | 2.06 | 0.159 155 | 0.705 799 |
| 0.375 | 0.0210 | 4.490 135 | 0.981 673 | 1.67 | 0.147 920 | 0.536 036 | 1.39 | 0.159 154 | 0.647 142 |

$$= \frac{1}{\nu-\lambda}\left\{\frac{\pi}{4}\left[\hat{H}^+\left(\frac{\nu}{\alpha}\right)+\hat{H}^-\left(\frac{\nu}{\alpha}\right)-\hat{H}^+\left(\frac{\lambda}{\alpha}\right)\right.\right.$$
$$\left.\left.-\hat{H}^-\left(\frac{\lambda}{\alpha}\right)+\frac{2}{\alpha}(\nu-\lambda)\right]\right\} \quad \text{[see Eq. (11)]}$$
$$= \frac{\pi}{2\alpha}+\frac{\pi}{4}\frac{1}{\nu-\lambda}\left[\hat{H}^+\left(\frac{\nu}{\alpha}\right)+\hat{H}^-\left(\frac{\nu}{\alpha}\right)\right.$$
$$\left.-\hat{H}^+\left(\frac{\lambda}{\alpha}\right)-\hat{H}^-\left(\frac{\lambda}{\alpha}\right)\right].$$

Therefore, expression (A2) equals

$$\frac{\pi}{4}\mathscr{P}\int_0^\infty d\nu\,\frac{(h(\nu))^{1/2}\kappa_s^\infty(\nu)}{\hat{H}^-(\nu/\alpha)}$$
$$\times\left\{\frac{2}{\alpha}+\frac{1}{\nu-\lambda}\left[\hat{H}^+\left(\frac{\nu}{\alpha}\right)+\hat{H}^-\left(\frac{\nu}{\alpha}\right)\right.\right.$$
$$\left.\left.-\hat{H}^+\left(\frac{\lambda}{\alpha}\right)-\hat{H}^-\left(\frac{\lambda}{\alpha}\right)\right]\right\}$$
$$-\pi^2\frac{h^{3/2}(\lambda)\kappa_s(\lambda)}{\hat{H}^-(\lambda/\alpha)}.$$

Consequently expression (A1) becomes

$$\frac{1}{2}\kappa_s^\infty(\lambda)\frac{\hat{H}^+(\lambda/\alpha)+\hat{H}^-(\lambda/\alpha)}{\hat{H}^-(\lambda/\alpha)}-\frac{2}{\pi}(h(\lambda))^{1/2}\mathscr{P}\int_0^\infty d\mu\,\frac{(h(\mu))^{1/2}\kappa_s^\infty(\mu)}{\hat{H}^-(\mu/\alpha)(\mu-\lambda)}+\frac{2}{\pi\alpha}\frac{(h(\lambda))^{1/2}}{\hat{H}^+(\lambda/\alpha)}\int_0^\infty d\mu\,\frac{(h(\mu))^{1/2}\kappa_s^\infty(\mu)}{\hat{H}^-(\mu/\alpha)}$$
$$+\frac{2}{\pi}\frac{(h(\lambda))^{1/2}}{\hat{H}^+(\lambda/\alpha)}\left\{\frac{1}{2}\mathscr{P}\int_0^\infty d\mu\,\frac{(h(\mu))^{1/2}\kappa_s^\infty(\mu)[\hat{H}^+(\mu/\alpha)+\hat{H}^-(\mu/\alpha)]}{\hat{H}^-(\mu/\alpha)(\mu-\lambda)}+\frac{2\pi h^{3/2}(\lambda)\kappa_s^\infty(\lambda)}{\hat{H}^-(\lambda/\alpha)}\right.$$
$$-\frac{1}{\alpha}\int_0^\infty d\mu\,\frac{(h(\mu))^{1/2}\kappa_s^\infty(\mu)}{\hat{H}^-(\mu/\alpha)}-\frac{1}{2}\mathscr{P}\int_0^\infty d\mu\,\frac{(h(\mu))^{1/2}\kappa_s^\infty(\mu)}{\hat{H}^-(\mu/\alpha)(\mu-\lambda)}\left[\hat{H}^+\left(\frac{\mu}{\alpha}\right)+\hat{H}^-\left(\frac{\mu}{\alpha}\right)-\hat{H}^+\left(\frac{\lambda}{\alpha}\right)-\hat{H}^-\left(\frac{\lambda}{\alpha}\right)\right]\Bigg\}$$
$$+\frac{2ih(\lambda)}{\hat{H}^+(\lambda/\alpha)}\left[-\frac{2}{\pi}(h(\lambda))^{1/2}\mathscr{P}\int_0^\infty d\mu\,\frac{(h(\mu))^{1/2}\kappa_s^\infty(\mu)}{\hat{H}^-(\mu/\alpha)(\mu-\lambda)}+\frac{1}{2}\kappa_s^\infty(\lambda)\frac{\hat{H}^+(\lambda/\alpha)+\hat{H}^-(\lambda/\alpha)}{\hat{H}^-(\lambda/\alpha)}\right]$$
$$=\kappa_s^\infty(\lambda)\left\{\frac{\hat{H}^+(\lambda/\alpha)+\hat{H}^-(\lambda/\alpha)}{2\hat{H}^-(\lambda/\alpha)}\left[1+\frac{2ih(\lambda)}{\hat{H}^+(\lambda/\alpha)}\right]+\frac{4h^2(\lambda)}{\hat{H}^+(\lambda/\alpha)\hat{H}^-(\lambda/\alpha)}\right\}$$
$$+\frac{2}{\pi}(h(\lambda))^{1/2}\mathscr{P}\int_0^\infty d\mu\,\frac{(h(\mu))^{1/2}\kappa_s^\infty(\mu)}{\hat{H}^-(\mu/\alpha)(\mu-\lambda)}\left\{-1+\frac{1}{2\hat{H}^+(\lambda/\alpha)}\left[\hat{H}^+\left(\frac{\mu}{\alpha}\right)+\hat{H}^-\left(\frac{\mu}{\alpha}\right)\right.\right.$$
$$\left.\left.-\hat{H}^+\left(\frac{\mu}{\alpha}\right)-\hat{H}^-\left(\frac{\mu}{\alpha}\right)+\hat{H}^+\left(\frac{\lambda}{\alpha}\right)+\hat{H}^-\left(\frac{\lambda}{\alpha}\right)-4ih(\lambda)\right]\right\}.$$

Now,

$$4ih(\lambda)=\hat{H}^-(\lambda/\alpha)-\hat{H}^+(\lambda/\alpha),$$

and so the second term of the above expression vanishes. Similarly, the first term reduces just to $\kappa_s^\infty(\lambda)$, as we wished to show.

## APPENDIX B: CALCULATION OF THE INVERSE MAPPING OPERATOR BY SOLVING THE INTEGRAL EQUATION

The inversion of the operator $\Omega^+$ can also be effected by solving a singular integral equation. If one is given $[\Omega^+(c_0,\kappa_s)]_s(\lambda)$ as a function $y_s(\lambda)$, say, then from Eq. (21) we see that it is necessary to solve for $c_0$ and $\kappa_s(\lambda)$ the equation

$$\kappa_s(\lambda)\left[1+\frac{2ih(\lambda)}{\hat{H}^+(\lambda/\alpha)}\right]+\frac{2(h(\lambda))^{1/2}}{\pi\hat{H}^+(\lambda/\alpha)}$$
$$\times\mathscr{P}\int_0^\infty d\mu\,\frac{(h(\mu))^{1/2}\kappa_s(\mu)}{\mu-\lambda}$$
$$=y_s(\lambda)+\left(\frac{2}{\pi\alpha}\right)^{1/2}(h(\lambda))^{1/2}\frac{c_0}{\hat{H}^+(\lambda/\alpha)}. \tag{B1}$$

Following Muskhelishvili, we define the function

$$X(\xi)=\frac{1}{2\pi i}\int_0^\infty d\mu\,\frac{(h(\mu))^{1/2}\kappa_s(\mu)}{\mu-\xi}.$$

$X$ is holomorphic for all $\xi$ not on the positive real axis and tends to zero as $|\xi|\to\infty$ with $\arg\xi\neq 0$. As $\xi$ tends to a real value $\lambda$ from above or below we get the limiting values

$$X^\pm(\lambda)=\frac{1}{2\pi i}\mathscr{P}\int_0^\infty d\mu\,\frac{(h(\mu))^{1/2}\kappa_s(\mu)}{\mu-\lambda}$$
$$\pm\tfrac{1}{2}(h(\lambda))^{1/2}\kappa_s(\lambda). \tag{B2}$$

Thus Eq. (B1) can be written as

$$[X^+(\lambda)-X^-(\lambda)]\left[1+\frac{2ih(\lambda)}{\hat{H}^+(\lambda/\alpha)}\right]$$
$$+\frac{2ih(\lambda)}{\hat{H}^+(\lambda/\alpha)}[X^+(\lambda)+X^-(\lambda)]$$
$$=(h(\lambda))^{1/2}y_s(\lambda)+\left(\frac{2}{\pi\alpha}\right)^{1/2}\frac{c_0h(\lambda)}{\hat{H}^+(\lambda/\alpha)}.$$

Since

$$4ih(\lambda)=\hat{H}^-(\lambda/\alpha)-\hat{H}^+(\lambda/\alpha),$$

this simplifies to

$$\frac{X^+(\lambda)}{\hat{H}^+(\lambda/\alpha)}-\frac{X^-(\lambda)}{\hat{H}^-(\lambda/\alpha)}$$
$$=\frac{(h(\lambda))^{1/2}y_s(\lambda)}{\hat{H}^-(\lambda/\alpha)}$$
$$+\left(\frac{2}{\pi\alpha}\right)^{1/2}\frac{c_0}{4i}\left[\frac{1}{\hat{H}^+(\lambda/\alpha)}-\frac{1}{\hat{H}^-(\lambda/\alpha)}\right].$$

This equation says that the function

$$\frac{X(\xi)}{\hat{H}(\xi/\alpha)}+\left(\frac{2}{\pi\alpha}\right)^{1/2}\frac{ic_0}{4\hat{H}(\xi/\alpha)},$$

which is holomorphic in $\xi$ except on the positive real axis, has a discontinuity, for positive $\lambda$, of

$$(h(\lambda))^{1/2}y_s(\lambda)/\hat{H}^-(\lambda/\alpha).$$

The above statement is also true for the function

$$\frac{1}{2\pi i}\int_0^\infty d\lambda\,\frac{(h(\lambda))^{1/2}y_s(\lambda)}{\hat{H}^-(\lambda/\alpha)(\lambda-\xi)},$$

and so the difference between the two functions, being holomorphic everywhere, must be an entire function. If this entire function is called $P(\xi)$, we have

$$X(\xi)=\hat{H}(\xi/\alpha)P(\xi)-(2/\pi\alpha)^{1/2}(ic_0/4)+\frac{\hat{H}(\xi/\alpha)}{2\pi i}\int_0^\infty d\lambda\,\frac{(h(\lambda))^{1/2}y_s(\lambda)}{\hat{H}^-(\lambda/\alpha)(\lambda-\xi)}. \quad \text{(B3)}$$

The requirement that as $|\xi|\to\infty$, $\arg\xi\neq 0$, $X(\xi)\to 0$ also implies that $P(\xi)=0$ and also that

$$c_0=-\left(\frac{2}{\pi\alpha}\right)^{1/2}\int_0^\infty d\lambda\,\frac{(h(\lambda))^{1/2}y_s(\lambda)}{\hat{H}^-(\lambda/\alpha)}.$$

This is Eq. (22). But we also have from Eqs. (B2) and (B3) that

$$\begin{aligned}\kappa_s(\lambda)&=\frac{1}{(h(\lambda))^{1/2}}[X^+(\lambda)-X^-(\lambda)]\\&=\frac{[\hat{H}^+(\lambda/\alpha)-\hat{H}^-(\lambda/\alpha)]}{(h(\lambda))^{1/2}}\frac{1}{2\pi i}\\&\quad\times\mathscr{P}\int_0^\infty d\mu\,\frac{(h(\mu))^{1/2}y_s(\mu)}{\hat{H}^-(\mu/\alpha)(\mu-\lambda)}\\&\quad+\left[\hat{H}^+\left(\frac{\lambda}{\alpha}\right)+\hat{H}^-\left(\frac{\lambda}{\alpha}\right)\right]\frac{1}{2}\frac{y_s(\lambda)}{\hat{H}^-(\lambda/\alpha)},\end{aligned}$$

and this simplifies to Eq. (23).


[1] R. Davidson and J. J. Kozak, J. Math. Phys. **11**, 189 (1970).
[2] R. Davidson and J. J. Kozak, J. Math. Phys. **11**, 1420 (1970).
[3] R. Davidson and J. J. Kozak, J. Math. Phys. **12**, 903 (1971).
[4] R. Davidson and J. J. Kozak, J. Math. Phys. **14**, 414 (1973).
[5] R. Davidson and J. J. Kozak, J. Math. Phys. **14**, 423 (1973).
[6] J. J-W. Yang, R. Davidson, and J. J. Kozak, J. Math. Phys. **15**, 491 (1974).
[7] R. Davidson and J. J. Kozak, J. Math. Phys. **16**, 1013 (1975).
[8] R. Davidson and J. J. Kozak, J. Math. Phys. **19**, 1074 (1978).
[9] R. Davidson and J. J. Kozak, J. Math. Phys. **17**, 1692 (1976).
[10] R. Davidson and J. J. Kozak, J. Math. Phys. **24**, 1911 (1983).
[11] N. I. Muskhelishvili, *Singular Integral Equations* (Noordhoff, Groningen, 1953).
[12] J. E. G. Farina, *Quantum Theory of Scattering Processes* (Pergamon, New York, 1973).